FORECASTING GDP OF UKRAINE WITH MIXED-FREQUENCY DATA MODEL

L. M. Zomchak, V. T. Marusina
Ivan Franko National University of Lviv,
Lviv, Ukraine

Every macroeconomic model is abstract and based on a set of assumptions and simplifications that limit the scope of its application. At the same time, it is on the basis of macroeconomic modeling that one can get a fairly complete picture of the nature of what is happening in the economy, make a forecast for the development of the economy, justify recommendations on decisions of a macroeconomic scale. Considering the need for strategic and tactical planning for the functioning of the state, it becomes relevant to analyze and forecast the state of the economy in the current and future periods, which in turn requires an approximate calculation of the indicators, has not yet been calculated.

Economic time series often have mixed frequencies. For example, GDP statistics are usually quarterly and with a significant delay and the data of many other economic indicators (average wages, unemployment, exchange rates, etc.) are monthly or have an even higher frequency. Such indicators often carry important information about the current state of the economy and it is important to use the information of data with a high frequency to obtain qualitative short-term forecasts. That is why methods that use mixed frequency data are becoming increasingly popular in predicting current system states and in short-term forecasting. Factor models and vector autoregression (VAR) models are examples of models that can deal with mixed frequency data under the condition of implementing a series of transformations with a time aggregation scheme.

An alternative approach is the mixed data sampling regression (MIDAS), proposed by E. Ghysels for the study of financial processes and phenomena and extended to macroeconomic studies by M. P. Clements and A. B. Galvão [1]. Constructed by E. Ghysels MIDAS regression is an abbreviated form of regression, which includes mixed- data processes [2]. M. Marcellino and C. Schumacher used MIDAS for Nowcasting and Forecasting with Ragged Edge Data (for German GDP) [3]. With coauthors they also used Midas for nowcasting GDP with many predictors or six industrialized countries [4].

Polynomials with a high distribution lag are used for coupling a high frequency variable with a low frequency variable. Polynomials are chosen in such a way as to prevent the growth of parameters simultaneously being very flexible, which makes it possible to construct a model of various forms in which it is necessary to find estimates for only a few parameters.

The base model MIDAS for the prediction of one variable h in future periods is as follows:

$$y_t = \beta_0 + \beta_1 b(L_m;\theta) x_{t_m}^{(m)} + \omega - h_m + \varepsilon_t, \qquad (1)$$

where y – dependent (endogenous) variable;

x – independent (exogenous) variable;

m – frequency of observations (for example, if the variable y is annual, than $x_1^{(4)}$ quarterly);

t – period;

ε – random error;

β_0 and β_1 – parameters;

$b(L^{1/m};\theta)$ – the distribution of log (e.g. Almon or beta distribution);

$b(L^{1/m};\theta) = \sum_{k=0}^{K} c(k;\theta) L_m^k$ and $L_m^k x_{t_m}^{(m)} = x_{t_m - x}^{(m)} \cdot x_{t_m+\omega}^{(m)}$ – the transformation of the indicator;

ω – number of the high frequency observations that explain the variable low frequency.

There are various options parameter calculation. The most common is an exponential Almon lag model:

$$c(k;\theta) = \frac{\exp\left(\theta_1 k + \ldots + \theta_Q k^Q\right)}{\sum_{k=1}^{K} \exp\left(\theta_1 k + \ldots + \theta_Q k^Q\right)}. \qquad (2)$$

Model (1) is quite general representation of MIDAS regression. Below are samples of other popular regression specifications. It is assumed that only one high-frequency variable available (with frequency m).

1. DL-MIDAS(PX):

$$y_{t+1} = \mu + \sum_{r=0}^{PX} \sum_{j=0}^{m-1} \beta_{rm+j} x_{(t-r)m-j} + \varepsilon_{t+1}.$$

2. ADL-MIDAS(PY, PX):

$$y_{t+1} = \mu + \sum_{j=0}^{PY} \mu_j y_{t-j} + \sum_{r=0}^{PX} \sum_{j=0}^{m-1} \beta_{rm+j} x_{(t-r)m-j} + \varepsilon_{t+1}.$$

3. FADL-MIDAS(PF, PY, PX):

$$y_{t+1} = \mu + \sum_{i=0}^{PF} \alpha_i F_{t-i} + \sum_{j=0}^{PY} \mu_j y_{t-j} + \sum_{r=0}^{PX} \sum_{j=0}^{m-1} \beta_{rm+j} x_{(t-r)m-j} + $$
$$+ \varepsilon_{t+1},$$

where F_t – additional variable of the model.

4. ADL-MIDAS-M(PX, PY) or multiplicative MIDAS:

$$y_{t+1} = \mu + \sum_{j=0}^{PY} \mu_j y_{t-j} + \sum_{r=0}^{PX} \alpha_r X_{t-r} + \varepsilon_{t+1},$$

where $X_{t-r} = \sum_{j=0}^{m_i-1} \beta_i x_{(t-r)m-j}$.

5. MIDAS, where there are only $J < m$ samples of high frequency variable x_T for period $t + 1$

$$y_{t+1} = \mu + \sum_{j=0}^{PY} \mu_j y_{t-j} + \sum_{j=1}^{J} \beta_{-j} x_{tm+j} + \sum_{r=0}^{PX} \sum_{j=0}^{m-1} \beta_{rm+j} x_{(t-r)m-j} + + \varepsilon_{t+1}.$$

Pre-selected variables can significantly improve prognosis compared to using large datasets. Quarterly or annual factor models generally lose forecasting models directly relating to low frequency, or using mixed data frequencies. With a simple regression estimation of several variables can be achieved sufficiently reliable performance prediction. Thus, alternative factor model MIDAS can be considered optimal model that passes all tests assess performance.

To construct the model used data of average wages by economic activity for the period from 2002 to 2016, registered unemployment in % of working age population from 2007 to 2016 and gross domestic product at current prices from 2002 to 2016 for Ukraine. Source of data – statistical information from site of State Statistics Service of Ukraine [5]. The model is implemented in the R programming environment using the package midasr [6].

Named above time series are illustrated in figures 1–3.

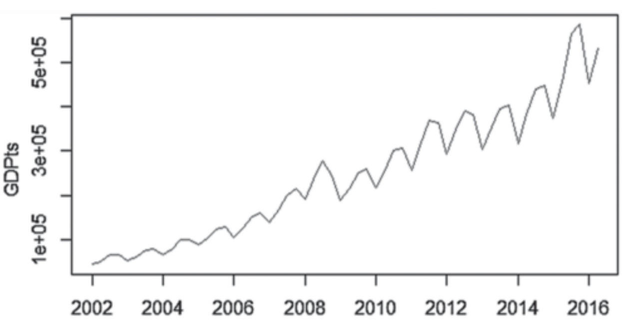

Figure 1 – **Dynamics of Ukrainian quarterly GDP at current prices for 2002–2016 years**

Based on figure 1, we can conclude about the growing trend of Ukraine's GDP. It is also worth noting the presence of seasonal fluctuations, the peak which falls mainly in the third quarter and the recession – at first.

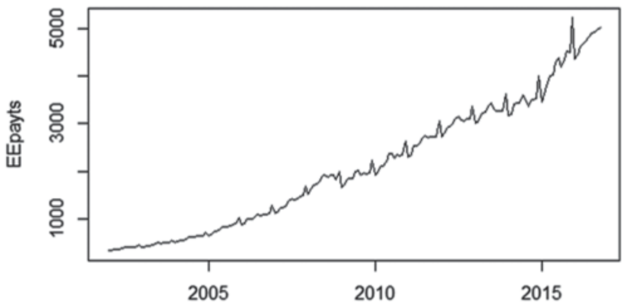

Figure 2 – **Dynamics of average monthly wage in Ukraine for 2002–2016 years**

Figure 2 illustrates the growing dynamics of the average wage in Ukraine, the recovery of growth observed in the period 2010–2016 years compared with preliminary data for 2002–2009. This time series is also characterized by seasonal fluctuations with annual peaks in December and moments of decline in January.

Registered unemployment is represented by a shorter time series (since 2007) and has less pronounced dynamics, can be explained by the fact that this indicator is relative, not absolute. However, it also demonstrates seasonal dependence with sheer annual growth in values in January – February and declines in September – October.

Given the time series under study (registered unemployment in %, average monthly salary in UAH, and quarterly GDP in mln. UAH) and their dynamics, to ensure the comparability and stationarity of time series, a conversion is performed by calculating the difference of logarithms for each time series. The stationarity test was carried out using one of the most common criteria – the extended Dickey-Fuller test, which tests the null hypothesis of the presence of a unit root in the time series ($\gamma = 0$). The alternative hypothesis differs depending on the version of the test used, but, as a rule (and in this example in particular), is stationarity or trend-stationarity ($\gamma > 0$).The value of the test statistics is calculated as:

$$DF_T = \frac{\hat{\gamma}}{SE(\hat{\gamma})}.$$

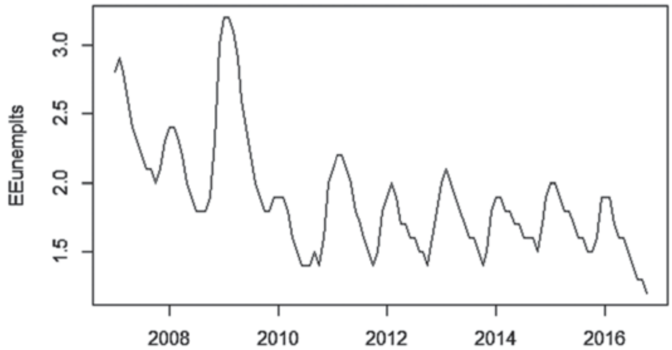

Figure 3 – **Registered unemployment rate as % of the population of working age in Ukraine for 2007–2016**

If the test statistics are less than the critical value, then the null hypothesis is rejected and the time series are considered to be stationary.

Based on the results of testing the logged time series, the hypothesis of their stationarity is confirmed.

General regression equation with one exogenous variable can be written:

$$y_{t+1} = \alpha + \rho y_t + \sum_{j=0}^{8} \theta_j x_{3t-j} + \varepsilon_t ,$$

where y_t – quarterly GDP;
x_t – average monthly salary.

$$y_{t+1} = \alpha + \rho y_t + \sum_{j=0}^{8} \varpi_j z_{3t-j} + \varepsilon_t ,$$

z_t – registered unemployment rate in % of working age population.

Multifactor regression respectively will look like:

$$y_{t+1} = \alpha + \rho y_t + \sum_{j=0}^{8} \theta_j x_{3t-j} + \sum_{j=0}^{8} \varpi_j z_{3t-j} + \varepsilon_t .$$

Models are practically realized by the software package midasr, environment R. To correctly specify the model function midasr rewrite them in the following equivalent form:

$$y_t = \alpha + \rho y_{t-1} + \sum_{j=3}^{11} \theta_j x_{3t-j} + \varepsilon_t ,$$

$$y_t = \alpha + \rho y_{t-1} + \sum_{j=0}^{8} \varpi_j z_{3t-j} + \varepsilon_t ,$$

$$y_1 = \alpha + \rho y_{t-1} + \sum_{j=0}^{8} \theta_j x_{3t-j} + \sum_{j=0}^{8} \varpi_j z_{3t-j} + \varepsilon_t .$$

To build the first model (the dependence of the current value of the quarterly GDP values of the quarterly GDP and average wages of previous periods) use sampling for 2002–2016 years. Because of small volume statistics for the study the U-MIDAS (MIDAS unlimited model) model is selected. Explicitly model can be written:

$$y_t = -5{,}43 + 0{,}05y_{t-1} - 0{,}08x_{3t-3} + 0{,}99x_{3t-4} - 0{,}32x_{3t-5} + 1{,}74x_{3t-6} + 0{,}79x_{3t-7} - 0{,}86x_{3t-8} - 0{,}01x_{3t-9} - 0{,}76x_{3t-10} - 0{,}13x_{3t-11}.$$

To build a second model (the dependence of the current value of the quarterly GDP values of the quarterly GDP and monthly unemployment rate of previous periods) use sampling for 2007–2016:

$$y_t = 2{,}1 + 0{,}2y_{t-1} - 1{,}08z_{3t-3} - 0{,}25z_{3t-4} + 0{,}57z_{3t-5} + 0{,}47z_{3t-6} + 0{,}1z_{3t-7} - 0{,}25z_{3t-8} + 0{,}39z_{3t-9} - 0{,}09z_{3t-10} - 0{,}42z_{3t-11}.$$

To build the next model (dependence of the current value of the quarterly GDP values of the quarterly GDP, the average monthly wage and unemployment rate of previous periods) use sampling for 2007–2016:

$$y_t = 1{,}75 + 1{,}14y_{t-1} - 0{,}42x_{3t-3} + 0{,}52x_{3t-4} + 1{,}18x_{3t-5} + 1{,}2x_{3t-6} + 1{,}45x_{3t-7} - 0{,}62x_{3t-8} - 0{,}4x_{3t-9} - 1{,}96x_{3t-10} - 0{,}63x_{3t-11} - 0{,}49z_{3t-3} + 0{,}16z_{3t-4} + 0{,}37z_{3t-5} + 0{,}58z_{3t-6} + 0{,}35z_{3t-7} + 0{,}45z_{3t-8} - 0{,}13z_{3t-9} - 0{,}69z_{3t-10} - 0{,}33z_{3t-11}.$$

Will assess constructed models with built-in software package MIDAS regression test – hAh restriction test.

On the basis of analysis models can conclude the highest degree of reliability of the first model, which describes the dependence of the current value of the quarterly GDP values of the quarterly GDP and average wages prior periods. Based on this model make prediction on the importance of Ukraine's GDP in the next quarter so logged predicted value of GDP in Q3 2016 is 17,6494. After inverse transformation we get the predicted value of GDP for the 3rd quarter of 2016 at current prices – UAH 551 652,1 mln.

Analyzing the value could argue about sustainable development of economy of Ukraine, as predicted value falls within the general trend of quarterly GDP in Ukraine.

Total built MIDAS model allows virtually combined quarterly GDP statistics and the average salary in Ukraine, causing possible to analyze the relationship dynamics and performance.

References

1. **Clements, M.-P.** Macroeconomic forecasting with mixed-frequency data: Forecasting output growth in the United States / M.-P. Clements, A.-B. Galvão // Journal of Business & Economic Statistics. – 2008.
2. **Ghysels, E.** The MIDAS touch: Mixed data sampling regression models / E. Ghysels, P. Santa-Clara, R. Valkanov // Finance. – 2004.
3. **Marcellino, M.** Factor MIDAS for Nowcasting and Forecasting with Ragged-Edge Data: A Model Comparison for German GDP / M. Marcellino, C. Schumacher // Oxford Bulletin of Economics and Statistics. – 2010.
4. **Kuzin, V.** Pooling versus model selection for nowcasting GDP with many predictors: Empirical evidence for six industrialized countries / V. Kuzin, M. Marcellino, C. Schumacher // Journal of Applied Econometrics. – 2013.
5. State Statistics Service of Ukraine [Electronic resource]. – Mode of access: http://www.ukrstat.gov.ua/.
6. **Ghysels, E.** Mixed frequency data sampling regression models: The R package midasr / E. Ghysels, V. Kvedaras, V. Zemlys // Statistical Software. – 2016.