BANCO DE ESPAÑA
Eurosistema

ECONOMIC BULLETIN
2023/Q1

Article 18
09/03/2023

A supply-side GDP nowcasting model

https://doi.org/10.53479/29778

Rationale

The recent shocks to the Spanish economy, linked to both COVID-19 and rising energy prices, have had an uneven impact across sectors of activity, underscoring the importance of monitoring the supply side of economic activity.

Takeaways

- This article presents a model for forecasting quarterly GDP using a combination of monthly indicators to estimate the growth of gross value added for each sector of activity.
- The results in terms of forecasting accuracy evidence the usefulness of a sectoral approach, as a complementary tool for monitoring economic activity in the short term.

Keywords

Economic cycle, growth, time series, forecasts, sectors.

JEL classification

E32, E37, C22.

Authors:

Alejandro Fernández Cerezo
Economic Developments Department
Banco de España

Introduction

Analysis of current economic developments is a keystone for decision-making authorities in the economic and financial arena. Short and medium-term macroeconomic projection exercises rest on, first, structural knowledge of the economy and, second, the distillation of the latest economic information, a process supported by econometric models.

The uneven impact of the COVID-19 crisis across sectors of activity, along with the recent shock linked to higher energy prices, has underscored the importance of monitoring economic activity on the supply side. In addition, a sectoral approach allows us to analyse which sectors are behind the changes in short-term GDP forecasts. We can thereby identify whether GDP growth is associated with shocks affecting a specific economic sector,[1] thus complementing the demand-side approach, which is more widely used by the different authorities in their economic analysis.

This article summarises the key aspects of an indirect,[2] supply-side approach to forecasting short-term Spanish GDP growth.[3] Specifically, drawing on a set of monthly indicators, forecasts are made for the current quarter (nowcasting) for each of the supply-side GDP components (i.e. gross value added (GVA) by sector of activity), using a mixed-data sampling (MIDAS) approach. Aggregating the results for these sectoral components yields a GDP forecast. The model's forecasting accuracy is assessed for the period 2020 Q1-2022 Q3. This period was dominated by the health crisis and the subsequent recovery, which posed significant challenges to macroeconomic forecasting.

The rest of this article is structured as follows. The second section describes the set of indicators comprising the database used. The third section describes the modelling strategy employed. The fourth section analyses the predictive power of the proposed model. The final section sets out the main conclusions.

The database

On the supply side of the economy, GDP at market prices is defined as the sum of GVA over all industries plus taxes on products minus subsidies on products. The sectors of activity are grouped and coded as per NACE Rev. 2. Specifically, the Quarterly National Accounts (QNA) published by the National Statistics Institute (INE) provide GVA data in the following 10-industry

1 The dispersion in industry-level GVA growth often widens during recessions, with the sharpest contractions coming in the investment goods and consumer discretionary sectors (Rees, 2020).

2 The available approaches to nowcasting GDP growth can be divided into two types: direct (forecasts drawing on different indicators) and indirect (forecasting the different components of GDP, which are subsequently aggregated).

3 For other models used in the Banco de España's projection exercises, see, for example, Aguilar, Ghirelli, Pacce and Urtasun (2021), Arencibia, Gómez Loscos, De Luis López and Pérez Quirós (2017), Álvarez, Cabrero and Urtasun (2014) and Camacho and Pérez Quirós (2010 and 2011).

breakdown (NACE Rev. 2 sections given in brackets): agriculture, forestry and fishing (A); industry (B-E); construction (F); trade, transportation, accommodation and food services (G-I); information and communication (J); financial and insurance activities (K); real estate activities (L); professional, scientific and technical activities and others (M-N); public administration, education and health (O-Q) and arts, entertainment and recreation and other services (R-T).

In Spain, there is a broad set of higher-frequency economic indicators available at the sectoral level. A given indicator is only included in the database where three criteria are met: (i) a long time series is available, allowing its inclusion in an econometric model; (ii) it is published before the release of QNA data;[4] and (iii) it has a monthly frequency, thus allowing new information to be included over the course of the quarter, although this poses some modelling difficulties (for example, unlike GDP data, some of the data series are not adjusted for calendar and seasonal effects, meaning these adjustments would need to be made).

Typically, to select the most representative indicators the correlation between each indicator and the QNA reference variable is assessed, using quarter-on-quarter rates of the seasonally adjusted series, following the TRAMO-SEATS approach (Gómez and Maravall, 1998). Table 1 presents the indicators considered for each of the productive sectors and the correlation between each one and its corresponding series of real GVA (e.g. the correlation between the industrial production index (IPI) and the industrial sector's GVA)). The indicators ultimately used for each sector are marked in italics. Broadly speaking, both the quantitative indicators (social security registrations and electricity consumption) and the qualitative indicators (survey-based, such as the Purchasing Managers' Indices (PMIs)) have a positive and strong correlation with their respective GVA variables in most of the sectors considered. For instance, in the industrial sector none of the five indicators considered has a correlation coefficient below 0.8. Conversely, the monthly indicators for the primary and financial sectors have a lower correlation with their respective GVA series. In the case of agriculture, the GVA rate of change is highly volatile, which the indicators available struggle to capture, while approximating the GVA of the financial and insurance activities sector using partial indicators is difficult due to the nature of the sector.[5]

As well as showing strong correlation with each sector's GVA, the selected monthly indicators must, as mentioned above, meet a timeliness and availability criterion, since indicators with a publication lag of more than three months are unlikely to provide useful information when it comes to forecasting GDP growth for the current quarter. In some cases, such as the data for social security registrations and for electricity consumption by large firms, the series have a very short publication lag and are available for all sectors (at the NACE Rev. 2 division level), and are therefore considered particularly suitable for the purposes of this study. Chart 1 shows a selection of the time series considered for several representative sectors. The correlation between the indicators

4 The INE publishes flash QNA data 30 days after the reference quarter ends.

5 In the financial intermediation services sector, measuring GVA, defined as output minus inputs, is a challenge. One key difference between financial institutions and non-financial corporations lies in how the value of the output and inputs is defined. For example, as intermediaries between savers and borrowers, banks' revenue is mostly generated through net interest income: the difference between interest received and interest paid. Accordingly, for the agricultural financial services sectors, the series of social security registrations has been used for want of other more precise indicators.

Table 1

Selected indicators in the database and their correlation with the corresponding real GVA series

Indicator (a)	Correlation with real GVA (b)	Series start	Publication lag (days from period end)
AGRICULTURE, FORESTRY AND FISHING (A)			
Social security registrations	*0.02*	*Jan-2009*	*t+3*
Electricity consumption of large firms	-0.18	Jan-2010	t+20
INDUSTRY (B-E)			
Social security registrations	*0.93*	*Jan-2009*	*t+3*
Electricity consumption of large firms	*0.80*	*Jan-2010*	*t+20*
Manufacturing PMI	*0.82*	*Feb-1998*	*t+1*
Industrial production index (IPI)	*0.96*	*Jan-1995*	*t+37*
CONSTRUCTION (F)			
Social security registrations	*0.89*	*Jan-2009*	*t+3*
Electricity consumption of large firms	*0.79*	*Jan-2010*	*t+20*
Cement consumption	*0.65*	*Jan-1995*	*t+15*
TRADE, TRANSPORTATION, ACCOMMODATION AND FOOD SERVICES (G-I)			
Social security registrations	*0.99*	*Jan-2009*	*t+3*
Electricity consumption of large firms	*0.92*	*Jan-2010*	*t+20*
Services PMI	*0.83*	*Aug-1999*	*t+3*
Retail trade index (ICM)	*0.90*	*Jan-2003*	*t+28*
INFORMATION AND COMMUNICATION (J)			
Social security registrations	*0.81*	*Jan-2009*	*t+3*
Electricity consumption of large firms	*0.74*	*Jan-2010*	*t+20*
Services PMI	*0.63*	*Aug-1999*	*t+3*
FINANCIAL AND INSURANCE ACTIVITIES (K)			
Social security registrations	*0.33*	*Jan-2009*	*t+3*
Electricity consumption of large firms	-0.04	Jan-2010	t+20
Stock of lending to other resident sectors	0.28	Jan-1962	t+55
REAL ESTATE ACTIVITIES (L)			
Social security registrations	*0.77*	*Jan-2009*	*t+3*
Electricity consumption of large firms	*0.70*	*Jan-2010*	*t+20*
Services PMI	*0.67*	*Aug-1999*	*t+3*
PROFESSIONAL, SCIENTIFIC AND TECHNICAL ACTIVITIES AND OTHERS (M-N)			
Social security registrations	*0.96*	*Jan-2009*	*t+3*
Electricity consumption of large firms	*0.72*	*Jan-2010*	*t+20*
Services PMI	*0.74*	*Aug-1999*	*t+3*
PUBLIC ADMINISTRATION, EDUCATION AND HEALTH (O-Q)			
Social security registrations	0.36	Jan-2009	t+3
Electricity consumption of large firms	0.07	Jan-2010	t+20
ARTS, ENTERTAINMENT AND RECREATION AND OTHER SERVICES (R-T)			
Social security registrations	*0.90*	*Jan-2009*	*t+3*
Electricity consumption of large firms	*0.88*	*Jan-2010*	*t+20*
Services PMI	*0.79*	*Aug-1999*	*t+3*

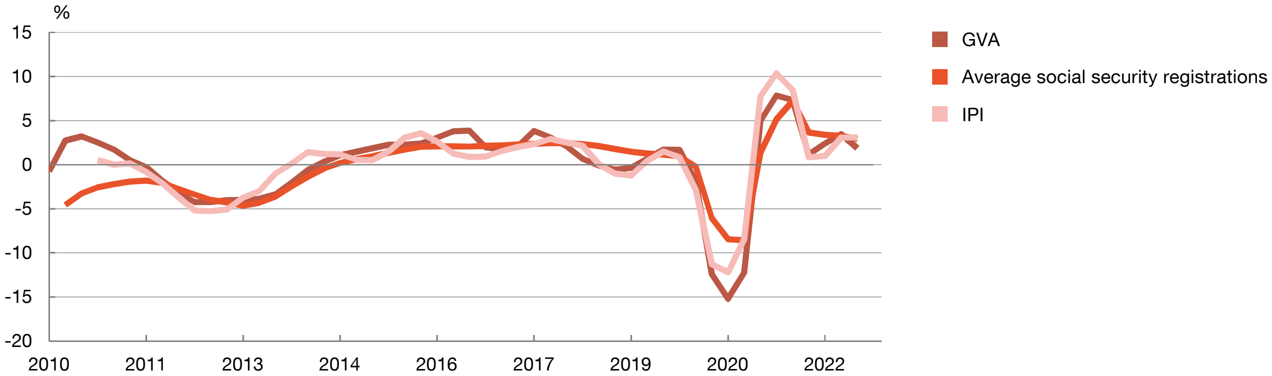
SOURCE: Banco de España.

a The indicators ultimately included in the model are shown in italics.
b Correlation in terms of quarter-on-quarter rates. For the monthly indicators, quarterly rates are found using the arithmetic mean. Unless otherwise indicated, the QNA variables and the indicators are seasonally adjusted series.

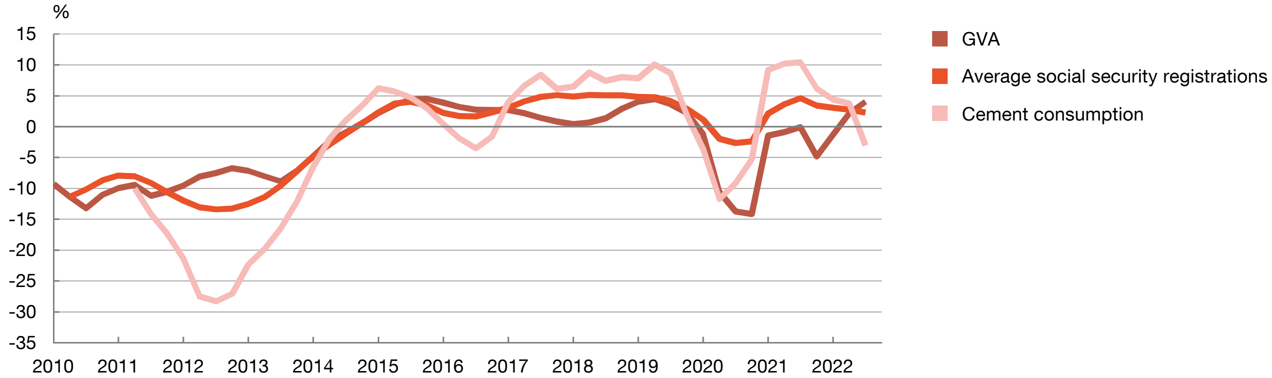
Chart 1

Rate of change of three-quarter moving averages vis-à-vis the previous three quarters. Seasonally adjusted indicators in real terms, unless otherwise stated

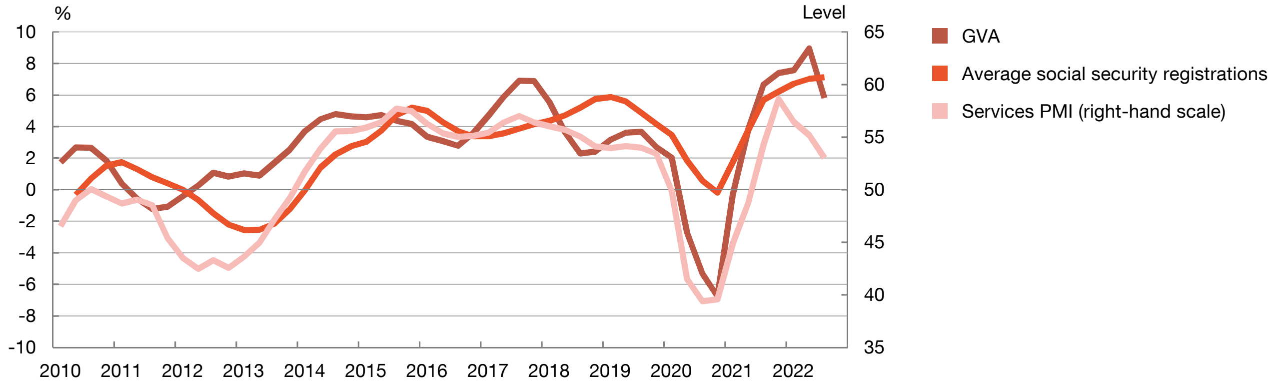
1.a Industry (B-E)



1.b Construction (F)



1.c Information and communication (J)



DOWNLOAD

SOURCES: Ministerio de Inclusión, Seguridad Social y Migraciones, S&P Global, INE, Oficemen and Banco de España.

and real GVA varies over time, which appears to justifies the a priori consideration of a broad set of indicators in the forecasting models.

The modelling strategy

Data being available at different time frequencies (also known as mixed-frequency data) poses a challenge when it comes to incorporating those data into nowcasts of macroeconomic variables. High-frequency variables contain potentially valuable information on recent economic developments. Meanwhile, the traditional estimation models are unsuitable for directly analysing high-frequency data if the main macroeconomic variables (such as GDP) are available at a lower frequency. To address this, the literature has developed econometric techniques for nowcasting using mixed-frequency data.[6]

The modelling strategy followed in this article is based on the MIDAS approach. This is a single-equation model capable of incorporating all of the significant mixed-frequency data that are available at the time of estimation (quarterly GVA and monthly activity indicators). MIDAS is a relatively recent model which was initially applied in the field of finance (Ghysels, Sinko and Valkanov, 2007) and has since expanded into macroeconomics (Clements and Galvao, 2008 and 2009, and Kuzin, Marcelino and Schumacher, 2011). One of the reasons for the growing use of MIDAS models is their flexibility and simplicity, which allows all information available to be incorporated into the GDP estimate in real time. Lastly, as single-equation models, they are less sensitive to specification errors (Bai, Ghysels and Wright, 2013). A more detailed description of the methodology can be found in the annex to this article.

In this study, each sector's quarterly real GVA growth is modelled independently, meaning a MIDAS equation is estimated for each sector of activity. The sector-level forecasts are then aggregated to yield a forecast rate of change for total GVA.[7]

With the modelling strategy adopted, the functional form used to estimate the coefficients when estimating the MIDAS equations must be selected. To this end, the predictive power of the different functional forms considered is assessed, as described below.

Results

To assess the performance of this supply-side GDP nowcasting model, a quasi-real time projection exercise was performed, comparing the model's estimate with the QNA flash estimate in each quarter, resulting in the model being continually re-estimated using an updated dataset. The exercise was conducted for the GVA of the market economy, i.e. excluding the public sector and

6 See Banbura, Giannone, Modugno and Reichlin (2013) for a detailed description.

7 The MIDAS model is estimated for the market economy sectors. The GVA rate of change for the public administration, education, and health (O-Q) and the net taxes on products are drawn from internal Banco de España projections. These projections are essentially based on assumptions relating to developments in public sector employment and real private consumption, respectively.

Table 2
Evaluation of the market economy GVA forecasting model

	Model	2020 Q1-2022 Q3	2021 Q1-2022 Q3
1 Mean squared error relative to an AR(1)	MIDAS - Almon	0.04	0.09
	MIDAS - Exponential Almon	0.08	0.11
	MIDAS - U-MIDAS	0.05	0.12

	Model	2020 Q1-2022 Q3	2021 Q1-2022 Q3
2 Mean squared error relative to a direct forecasting MIDAS model	MIDAS - Almon	0.15	0.54
	MIDAS - Exponential Almon	0.03	0.12
	MIDAS - U-MIDAS	0.10	0.41

SOURCE: Banco de España.

net taxes on production. The dataset comprises observations from 2010 Q1[8] to 2022 Q3. This period is divided into two: the model is estimated up to 2019 Q4, while from 2020 Q1 onwards projections are made using the rolling window method. By distinguishing between training and assessment samples we can evaluate the predictive power during a period marked primarily by the COVID-19 crisis, which posed considerable challenges to preparing macroeconomic projections.[9] The quarter-on-quarter GDP growth forecasting error – defined as the difference between the INE estimate and the projection obtained from the different indicators available on each of these dates – is calculated for each quarter in the out-of-sample period. Following the usual practice in this type of exercise, the accuracy of the model's prediction is compared with that of a benchmark stylised statistical model – specifically, a first-order autoregression for each sector of activity, which is subsequently aggregated to obtain total output.

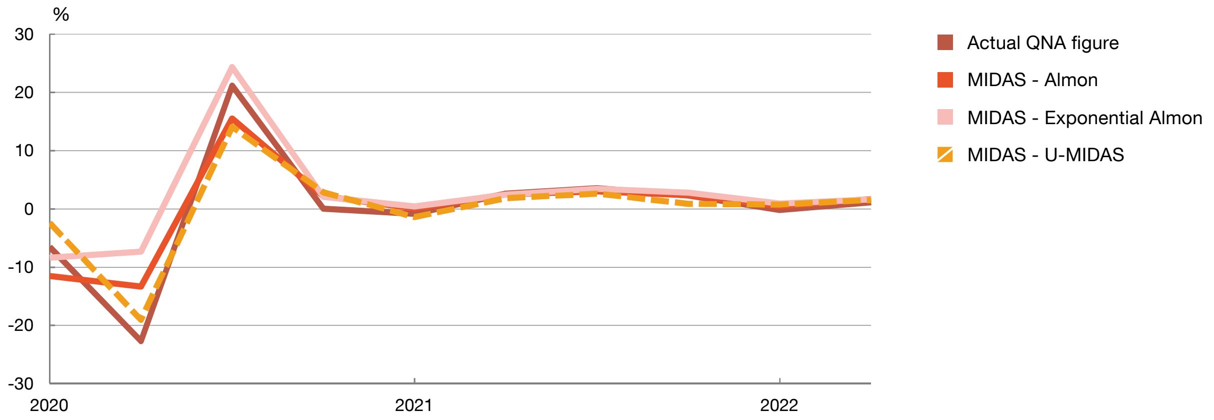
Section 1 of Table 2 shows the mean squared error (MSE) relative to the stylised model for the different specifications of the functional form of the MIDAS equation typically used in the literature (Almon, exponential Almon and U-MIDAS), which are described in the annex to this article. Values greater than one mean that the stylised model is more accurate than the MIDAS-based approach, whereas values less than one mean that this approach has greater predictive power.

The main conclusion from this exercise is that the supply-side forecasting model is considerably more accurate than the stylised model. For the sample as a whole, the error is approximately one-tenth of the autoregressive model's error. The result is unchanged for the period after the collapse and recovery in activity linked to the worst phase of the pandemic. Among the different specifications considered, the Almon functional form stands out, as its MSE is smaller than the others.

8 First date from which information for all variables is available.

9 To isolate the effect of the worst phase of the pandemic, which prompted highly asymmetric rates of change in GVA across sectors, the model's predictive power is also assessed from 2021 Q1 up to the end of the sample, i.e. 2020 is disregarded. The results remain qualitatively unchanged.

Chart 2
Quasi-real time market economy GVA forecasts. Comparison of different forecasting models



SOURCE: Banco de España.

It is also worth examining whether the model estimated indirectly (equations and sectors that are then aggregated to obtain the aggregate output forecast) improves on direct forecasts (estimating the GVA of the market economy directly using the indicators available at this level: total social security registrations, total electricity consumption, IPI and PMI data). Using the same out-of-sample forecasting and estimation windows as in the previous exercise, Section 2 of Table 2 shows the MSE of the indirect model relative to the direct model for the same specifications of the functional form of the MIDAS equation as in Section 1. Similarly, values less than one mean that the indirect approach has greater predictive power than the direct one. In the out-of-sample assessment period as a whole, the MSEs under the indirect model (sectors of activity) are smaller than under the direct model (aggregate). These results are consistent with other papers that point to the potential advantages of forecasting GDP indirectly via its components. For example, using data for the euro area, Foroni and Marcellino (2014) show that aggregating the components makes for more accurate nowcasting of total GDP growth.

Lastly, Chart 2 depicts the different forecasts made for each of the quarters in the assessment period, together with the flash estimates for market economy GVA growth according to the QNA. Except for the first three quarters of 2020 – in which economic activity was highly volatile as a result of the most stringent pandemic restrictions –, the errors are smaller, in both the periods of economic slowdown and those of recovery.

Conclusion

This article presents a GDP growth nowcasting tool, which joins the others regularly used at the Banco de España. Unlike other alternatives, this model offers a comprehensive supply-side macroeconomic forecast, enabling the analysis of which sectors are behind short-term fluctuations in activity. MIDAS equations are estimated for each productive sector, the results of which are subsequently aggregated to obtain a forecast for total output. This methodology allows us to flexibly combine the information available from different short-term economic indicators and useful information on current sectoral developments, with satisfactory results in terms of predictive power. Overall, the results show the usefulness of a sectoral approach as a complementary tool for forecasting changes in GDP in the short term.

REFERENCES

Aguilar, Pablo, Corina Ghirelli, Matías Pacce and Alberto Urtasun. (2021). "Can news help measure economic sentiment? An application in COVID-19 times". *Economics Letters,* Vol. 199(C), Elsevier.

Álvarez, Luis Julián, Alberto Cabrero and Alberto Urtasun. (2014). "A procedure for short-term GDP forecasting". *Economic Bulletin - Banco de España,* October, pp. 29-35. https://www.bde.es/f/webbde/SES/Secciones/Publicaciones/InformesBoletinesRevistas/BoletinEconomico/14/Oct/Files/art3e.pdf

Arencibia, Ana, Ana Gómez Loscos, Mercedes de Luis and Gabriel Pérez Quirós. (2017). "A short-term forecasting model for GDP and its demand components". *Economic Bulletin - Banco de España,* 4/2017, Analytical Articles. https://www.bde.es/f/webbde/SES/Secciones/Publicaciones/InformesBoletinesRevistas/ArticulosAnaliticos/2017/T4/files/beaa1704-art30e.pdf

Bai, J., E. Ghysels and J. Wright. (2013). "State space models and MIDAS regressions". *Econometric Reviews,* Vol. 32(7), pp. 779-813.

Banbura, M., D. Giannone, M. Modugno and L. Reichlin. (2013). "Now-Casting and the Real-Time Data Flow". In G. Elliott, C. Granger and A. Timmermann (eds.). *Handbook of Economic Forecasting,* Vol. 2, Part A, Ch. 4, pp. 195-237, Elsevier.

Camacho, Máximo, and Gabriel Pérez-Quirós. (2010). "Introducing the Euro-Sting: Short-Term INdicator of Euro Area Growth". *Journal of Applied Econometrics,* Vol. 25(4), pp. 663-694.

Camacho, Máximo, and Gabriel Pérez-Quirós. (2011). "Spain-STING: Spain Short-Term INdicator of Growth". *The Manchester School,* Vol. 79, pp. 594-616.

Clements, M. P., and A. B. Galvão. (2009). "Forecasting US output growth using leading indicators: An appraisal using MIDAS models". *Journal of Applied Econometrics,* Vol. 24(7), pp. 1187-1206.

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

Foroni, C., and M. Marcellino. (2014). "A comparison of mixed frequency approaches for nowcasting euro area macroeconomic aggregates". *International Journal of Forecasting,* Vol. 30(3), pp. 554-568.

Ghysels, E., P. Santa-Clara and R. Valkanov. (2004). "The MIDAS touch: MIxed DAta Sampling regression models". Mimeo, Chapel Hill, N. C.

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

Ghysels, E., A. Sinko and R. Valkanov. (2007). "MIDAS regressions: Further results and new directions". *Econometric Reviews,* Vol. 26(1), pp. 53-90.

Gómez, Víctor, and Agustín Maravall. (1998). "Guide for Using the Programs TRAMO and SEATS". Documentos de Trabajo - Banco de España, 9805. https://www.bde.es/f/webbde/SES/Secciones/Publicaciones/PublicacionesSeriadas/DocumentosTrabajo/98/Fic/dt9805e.pdf

Kuzin, V., M. Marcellino and C. Schumacher. (2011). "MIDAS vs mixed-frequency VAR for nowcasting GDP in the Euro area". *International Journal of Forecasting,* Vol. 27, pp. 529-542.

Rees, D. (2020). "What comes next? Recovery from an uneven recession". *BIS Bulletins,* 33, Bank for International Settlements.

Annex
Description of the MIDAS methodology

MIxed-DAta Sampling (MIDAS) is a methodology capable of combining time series with different frequencies. It is therefore a particularly useful tool for measuring economic developments in real time, allowing monthly – or even weekly or daily – indicators to be used to forecast changes in quarterly GDP. This methodology is briefly summarised below.

In a MIDAS equation, γ_t is a dependent variable observed at a low frequency (e.g. gross value added (GVA) each quarter) and $x_t^{(m)}$ is a regressor observed at a higher frequency (e.g. monthly social security registrations or electricity consumption). The MIDAS specification is written as follows:

$$\gamma_t = \beta_0 + \beta_1 \, B\,(L^{1/m};\, \theta)\, x_t^{(m)} + \varepsilon_t$$

For t = 1,…T and where $B\,(L^{1/m};\theta) = \sum_{k=1}^{K} B\,(k;\theta)\, L^{k/m}$, K is the number of lags, $L^{1/m}$ is a lag operator such that $L^{1/m} x_t^{(m)} = x_{t-1/m}^{(m)}$ and θ is a parameter vector to be estimated. If all the variables have the same frequency, then m = 1, whereas if γ_t is observed quarterly while $x_t^{(m)}$ is a monthly indicator, then m = 3.

The unrestricted MIDAS (U-MIDAS) specification is one where the parameters of the lag polynomial are not restricted to a specific functional form (i.e. B(k) does not depend on θ) and all the model's parameters (β_0 and β_1) are estimated using the traditional ordinary least squares without including further restrictions.

However, the number of lags of $x_t^{(m)}$ is likely to be high. For example, if the quarterly observations of γ_t depend on six months of lags of $x_t^{(m)}$, there would be 18 lags (K = 18) of the high-frequency variable. The smaller the sample size and the greater the mix of frequencies (e.g. daily and quarterly data), the more problematic this overparameterisation becomes.

One way of resolving this issue is establishing a known function $B\,(L^{1/m};\, \theta)$ with fewer parameters to be estimated (θ). For example, in the Almon polynomial, which is one of the alternatives commonly used in the literature, the weight of each lag k is calculated as:

$$B\,(k;\, \theta) = \theta_0 K^0 + \theta_1 K^1 \ldots + \theta_Q K^Q$$

where Q is the order of the polynomial. Another widely used alternative polynomial is the exponential Almon:

$$B\,(K;\, \theta) = \frac{e^{\theta_1 k + \ldots + \theta_Q k^Q}}{\sum_{m=1}^{m} e^{\theta_1 k + \ldots + \theta_Q k^Q}}$$

For example, if the order of the exponential Almon polynomial is equal to two, there would be four parameters to be estimated (θ_1, θ_2, β_0 and β_1) in the MIDAS equation. For a more detailed description of this methodology, see Ghysels, Sinko and Valkanov (2007).

How to cite this document

Fernández Cerezo, Alejandro. (2023). "A supply-side GDP nowcasting model". *Economic Bulletin - Banco de España*, 2023/Q1, 18.
https://doi.org/10.53479/29778

ISSN 1695-9086 (online edition)