INSTITUT DE HAUTES ÉTUDES INTERNATIONALES ET DU DÉVELOPPEMENT
GRADUATE INSTITUTE OF INTERNATIONAL AND DEVELOPMENT STUDIES

Graduate Institute of International and Development Studies

International Economics Department

Working Paper Series

Working Paper No. HEIDWP02-2022

Nowcasting real GDP in Tunisia using large datasets and mixed-frequency models

Hagher Ben Rhomdhane
Central Bank of Tunisia

Brahim Mehdi Benlallouna
Central Bank of Tunisia

March 2022

Chemin Eugène-Rigot 2
P.O. Box 136
CH - 1211 Geneva 21
Switzerland

Nowcasting real GDP in Tunisia using large datasets and mixed-frequency models

Hagher Ben Rhomdhane & Brahim Mehdi Benlallouna[1]

Central Bank of Tunisia

March 2022

[1] This working paper was drawn up by the team in charge of the Central Bank of Tunisia's strategic project "Leading Economic Indicators and Business Survey": Hager Ben Romdhane, Brahim Mehdi Benlallouna, Aida Guellouz, Amani Jebali, Rafik Touaiti, Raoudha El Wafi, Ramzi Salem and Kamel Dridi.

The authors would like to express their gratitude to Professors Massimiliano Marcellino and Seton Leonard for their guidance and support, in the framework of the technical assistance under the Bilateral Assistance and Capacity Building for Central Banks (BCC). In addition, we would like to thank State Secretariat for Economic Affairs (SECO) and the management team of BCC- Professor Cedric Till and Naël Shehadeh for their support and assistance during the strategic project. Many thanks go to Mr. Mohamed El Hadhri, sponsor of the project, Mr. Mourad Abdessalam (beneficiary), and Professor Rached Bouaziz (research unit in the CBT).

Abstract

This study aims to construct a new monthly leading indicator for Tunisian economic activity and to forecast Tunisian quarterly real GDP (RGDP) using several mixed-frequency models. These include a mixed dynamic factor model, unrestricted mixed-data sampling (UMIDAS), and a three-pass regression filter (3PRF) developed at the Central Bank of Tunisia, based on a monthly/quarterly set of economic and financial indicators as predictors.

Our methodology is based on direct and indirect approaches, and the direct approach nowcasts aggregate RGDPs . The indirect approach is a disaggregated approach based on the output side of GDP (manufacturing, non-manufacturing, and services) using a set of available monthly indicators by sector.

Furthermore, mixed-frequency dynamic factor models and unrestricted MIDAS perform well in terms of root mean squared errors compared to the benchmark model VAR (2). The forecast errors derived from the disaggregated approach during the recent COVID period are smaller than those derived from classical models such as VAR (2).

In our model, we used indicators such as electricity consumption by sector, stock market index detailed by sector, and international economic surveys to capture the pandemic effect. The financial variables improve forecasting for all horizons.

Additionally, we find that it is better to employ several UMIDAS-ARs by each component of GDP at constant prices and to pool the results rather than relying on aggregated GDP, specifically in volatile times.

***Key words*: Mixed-Frequency Data Sampling, Nowcasting, short-term forecasting.**

JEL Classification Code : E37, C55, F17, O11.

1. Introduction

Policymakers often face the problem of assessing the current state of the economy with incomplete statistical information because important economic variables are released with considerable time lags and low frequencies.

Especially in times of crisis, nowcasting is important because timely forecasts of the gross domestic product's (GDP) growth are useful summaries of recent news on the economy and are commonly used as inputs for structural forecasting.

COVID-19 has raised the issue of nowcasting and short-run forecasting owing to heightened uncertainties. In fact, the pandemic has led to a sudden halt in economic activity worldwide. The supply disruptions due to containment measures were magnified by large-scale demand destruction from employment and income losses and contraction in global trade and tourism. Tunisia's economy also took a severe hit, with GDP for Q2: 2020 declining by 21.7% year-on-year (y-o-y).

Econometric models that consider the information in unbalanced datasets have recently been developed. These datasets are unbalanced because of two features: the different sampling frequency and the "ragged-edge" issue, as publication delays cause missing observations for some of the variables at the end of the sample.

The Tunisian National Statistical Institute (INS) releases an estimate of GDP approximately 45 days after the end of the quarter. Furthermore, many leading and coincident indicators are available monthly or more frequently based on financial and business cycle indicators, such as financial and monetary variables. These helps monitor the current state of the economy, nowcasting, and short-run forecasting.

Usually, the simplest way to handle unbalanced data is to aggregate them to obtain balanced data at the same frequency and work with a "frozen" final vintage dataset; thus, the left-and right-hand side variables are sampled at the same frequency. However, this aggregation process destroys potentially useful information and can lead to misspecifications.

Accordingly, central banks rely on continuously flowing information from leading and coincident activity indicators to gauge the underlying state of the economy in real-time.

Currently, many central banks use a variety of models, such as bridge and factor models for nowcasting, by introducing survey variables. However, central banks such as the European, the Bank of England and the Federal Reserve Bank of America increasingly use Mixed-data sampling (MIDAS) models for short-term forecasting of GDP. Every month, they publish the national activity index.

In Tunisia, periods such as the 2011 revolution and the COVID-19 crisis are characterised by a significant decline in GDP. To address this issue, we remove outliers from the series used to estimate the model parameters for the entire sample period (2000M01-2021M04). Then, we add these outliers to generate predictions for nowcasting and forecasting the GDP.

Following the theoretical and empirical background, we constructed and evaluated the nowcasts of real GDP growth for the Tunisian economy. We collected data from January 2000 to June 2021 to forecast real GDP growth in the second quarter of 2021. Further, we focused on the multivariate unrestricted MIDAS approach proposed by Forroni et al. (2015) and cast real GDP growth several times each quarter. The explanatory variables included in the models were selected so that their single most recent observations were used, considering the delay in data releases. Real GDP nowcasts were aggregated from the sectoral value-added at constant prices (manufacturing, non-manufacturing, and services).

We also considered a large mixed-frequency dynamic factor model in the state-space approach of Mariano and Murassawa (2003) to summarise and exploit large mixed-frequency sets for nowcasting. In addition, we used a mixed-frequency factor model to construct a monthly index of economic activity that is fully consistent with quarterly data in terms of time aggregation.

Moreover, to assess the performance of forecasts, short-term forecasts, and the relevance of models' uncertainty, we considered alternative benchmark model vector autoregressive (VAR (2)) and mean to-date.

As a preview of the results, mixed-frequency dynamic factor models performed well in root mean squared errors concerning a benchmark model VAR (2), with the multivariate unrestricted MIDAS being the second best. The gains were larger during crisis periods. However, the magnitudes of predictions in these crises were significantly less than the realised values. Therefore, we think that an extension of this project will appropriately track the COVID period, which will be based on surveys.

The remainder of this paper is organised as follows. Section (2) briefly reviews the literature. Section (3) explains methodology used in this study. Section (4) gives an overview of Tunisian economic growth. Section (5) presents the data selection. Sections (6) to (9) give the main findings and results, and Section (10) concludes the paper.

2. Literature Review

In this section, we review the empirical literature using mixed-data sampling models to forecast real GDP growth. Various econometric approaches have been developed when ragged-edge data issues arise, such as MIDAS regression and factor models, including the Kalman filter (Giannone et al., Reichlin and Small, 2008) and Clements and Galvao (2008). These researchers introduced the use of MIDAS regressions in forecasting macroeconomic data. They examined whether a mixed-data sampling approach, including an autoregressive term, can improve U.S. real output growth forecasts. They conducted a real-time forecasting exercise that exploits the monthly vintages of the indicators and the quarterly vintages of output growth, consistent with the time of the release of the different data vintages. The authors found that using within-quarter information on monthly indicators can result in a marked reduction in Root Mean Square Error (RMSE) compared with the more traditional quarterly frequency VAR or AR distributed lag models.

Similarly, Marcellino, Foroni, and Schumacher (2012) compared the performance of the MIDAS with functionally distributed lags estimated with non-linear squared technic (NLS) to the performance of the UMIDAS, the unrestricted version of MIDAS. Monte Carlo experiments showed that UMIDAS generally performed better than MIDAS when mixing quarterly and monthly data. However, distributed lag functions outperformed unrestricted polynomials with larger differences in sampling frequencies. In an empirical application of out-of-sample nowcasting GDP in the Euro Area and the U.S. using monthly predictors, UMIDAS performed well.

Mariano and Murasawa (2003) proposed a new coincident index of the business cycle that relied on monthly and quarterly indicators. In addition, Mariano and Murasawa (2010) applied the mixed-frequency VAR method to construct a new coincident indicator, that is, an estimate of monthly real GDP. They found that the coincident index based on the VAR model, relative to that obtained by a factor model, tracks quarterly real GDP well, although they were quite volatile.

Furthermore, Marcellino and Shumacher (2010) proposed merging factor models with the MIDAS approach, allowing them to forecast low-frequency variables, such as GDP, exploiting information in a large set of higher-frequency indicators. They found that all factor–MIDAS nowcasts can improve quarterly factor forecasts based on time-aggregated data.

Recent applications using mixed-frequency factor models include Banbura and Modugno (2014), who discussed maximum likelihood estimation of factor models on datasets with arbitrary patterns of missing data.

Unfortunately, the reliability of forecasts decreases during crisis times and steep recovery periods. The main reason for this pattern is the failure of the model to capture these tail events. Therefore, we can cite the empirical work developed by Marcellino, Foroni, and Stevanovic (2020), who used mixed-frequency MIDAS and UMIDAS models and then adjusted the original nowcasts and forecasts by an amount similar to the nowcast and forecast errors made during the financial crisis. Their main findings show that the adjusted growth forecasts for 2020 Q2 become closer to the actual value. Further, the adjusted growth forecasts based on alternative indicators become much more similar, indicating a slower recovery than without adjustment.

3. Mixed-Frequency Models

This section presents Mixed-Frequency Models: Mixed-Frequency Small Factor Models, UMIDAS Model and Mixed-Frequency 3-Pass Regression Filter.

3.1 Mixed-Frequency Dynamic Factor Models (MFDFM)

A bridge equation relates a quarterly variable to the three-month average of monthly variables. This implicitly restricts the parameters for the months of the quarter, which introduces asymptotic biases and inefficiencies (Ghysels, 2004). By contrast, MIDAS estimates a monthly regression of GDP on monthly indicators using distributed lags. This approach has been extended to mixed dynamic factor models, unrestricted MIDAS, and 3PRF.

Factor models have also been employed to handle data with different frequencies. These models have been used to extract an unobserved state of the economy, create a new coincident indicator, exploit more information, and obtain more precise forecasts. Factor models have a long tradition of use in econometrics. Stock and Watson (1989), (2002) introduced a simple algorithm for estimating DFMs by maximum likelihood. By modelling the driving process behind (multivariate) observed data as latent (unobserved), DFMs can incorporate missing observations without falsified imputed data and model noisy observations due to measurement errors.

Mariano and Murasawa (2003) presented an mixed-frequency dynamic factor model (MFDFM) framework for mixed-frequency data, allowing practitioners to incorporate, for example, monthly and quarterly data without having to aggregate observations to the lowest frequency in the data.

Mariano and Murasawa (2003) applied a maximum likelihood factor analysis to a mixed-frequency series of quarterly GDP and monthly business cycle indicators to construct an index related to monthly real GDP. Further, Giannone et al. (2008) pioneered the application of DFMs to nowcasting, with a specific emphasis on using real-time data flow to update predictions as new information became available.

3.1.1. The Kalman Filter and Smoother

Kalman filtering and smoothing allowed us to model the observations as a function of unobserved latent factors. The measurement equation relating observations to factors is defined as $y_t = Hz_t + \varepsilon_t$ (1)

And the transition equation, governing the evolution of unobserved factors, is

$$z_t = Az_{t-1} + e_t \quad (2)$$

Where ε_t and e_t are normally distributed error terms with covariance matrix:

$$Cov \begin{bmatrix} e_t \\ \varepsilon_t \end{bmatrix} = \begin{bmatrix} Q & 0 \\ 0 & R \end{bmatrix}$$

In the above y_t, are noisy observations, z_t stacked factors $z_t = [x_t, x_{t-1} \dots, x_{t-p}]$. with p lags and n_t predetermined exogenous variables.

Given the parameters $H, A, Q, and\ R,$ estimates of factors or estimates of missing series in y_t are derived from the Kalman filter and are smoother. Our Kalman filter is:

$$z_{t/t-1} = A\, z_{t-1/t-1}$$

$$P_{t/t-1} = A\, P_{t-1/t-1} A' + Q$$

$$y_{t/t-1} = \widetilde{H} z_{t/t-1}$$

$$S_t = \widetilde{H}\, P_{t/t-1}\, \widetilde{H}' + R$$

$$C_t = P_{t/t-1}\, \widetilde{H}'$$

$$z_{t/t} = z_{t/t-1} + C_t S_t^{-1} (y_{t/t} - y_{t/t-1})$$

$$P_{t/t} = P_{t/t-1} - C_t S_t^{-1} C_t{}'$$

Matrix $\widetilde{H}$ in the above incorporates a helper matrix J to extract, in the simplest example, contemporaneous factors from z_t. That is, $J = [I_m \quad 0 \quad 0 \; ...]$ and $\widetilde{H} = HJ$.

In the above notation $x_{t/t-1}$ refers to our estimates of x_t given observations through t, and $x_{t/T}$ is our estimate of x_t conditional on available data through period T. Note that in the above $K_t = C_t S_t^{-1}$ is the Kalman gain, $\vartheta_t = (y_{t/t} - y_{t/t-1})$ is the prediction error and thus $K_t \vartheta_t$ is our forecast update.

3.1. 2 State-Space in Mixed-Frequency Models

Because state-space models are apt to handle missing data, they are particularly well suited to mixed-frequency data sets in which, for example, a quarterly variable will not be observed for two out of three months.

First, suppose that our model is in log levels, and as a concrete example, the frequencies are either monthly or quarterly. Denote y_t^q as the log of a quarterly observation in month t, Then

$$e^{y_t^q} = e^{y_t^m} + e^{y_{t-1}^m} + e^{y_{t-2}^m} \quad (3)$$

The difficulty is that equation (1) is linear in the log variables, while equation (3) is not. To overcome this issue, we simply take a linear approximation of (3), which yields

$$y_t^q = \frac{1}{3}(y_t^m + y_{t-1}^m + y_{t-2}^m) \quad (4)$$

Plugging equation (8) into the above yields the linear state-space structure:

$$y_t^q = \frac{1}{3}Hx_t + \frac{1}{3}Hx_{t-1} + \frac{1}{3}Hx_{t-2} + \varepsilon_t \quad (5)$$

Note that this requires that the model includes at least three lags of factors, although one does not need to estimate the coefficients on factors with more than one lag in the transition equation.

To put the model into log differences, we begin with equation (4) and note that the observed Δy_t^q is

$$\begin{aligned} y_t^q - y_{t-3}^q &= \frac{1}{3}(y_t^m - y_{t-3}^m) + \frac{1}{3}(y_t^m - y_{t-4}^m) + \frac{1}{3}(y_{t-2}^m - y_{t-5}^m) \\ &= \frac{1}{3}\Delta y_t^w + \frac{2}{3}\Delta y_{t-1}^w + \Delta y_{t-2}^w + \frac{2}{3}\Delta y_{t-3}^w + \frac{1}{3}\Delta y_{t-4}^w \quad (6) \end{aligned}$$

This is the result of Mariano and Murassawa's (2003) study. Unlike in the level case, we now need to include at least four lags of the factors.

We used the expectation–maximization (EM) algorithm, which is based on two steps:

- **Expectation step** that computes the expected values of the factors given parameters at the current iteration
- **The maximisation step** consists of finding the parameters that maximise the likelihood of the factors and the observed data y_t for the current iteration.

We repeat this process until the likelihood function converges.

As with principal components, there is an identification issue, and a possible solution to avoid this issue is to use blocks. In fact, the blocks impose zero restrictions on our estimate H. For example, we can use the following over-identified model:

$$H = \begin{bmatrix} h_{1,1} & h_{1,2} & h_{1,3} \\ h_{2,1} & 0 & 0 \\ 0 & h_{3,2} & 0 \\ 0 & h_{4,2} & 0 \\ 0 & 0 & h_{5,3} \end{bmatrix}$$

In this example, our variable of interest (RGDP) is variable 1, which loads all factors. The factors may be nominal, supply, or demand factors. This framework allows us to examine how each economic sector contributes to our variable of interest (RGDP). Deviations in the observed value from the fitted values may be interpreted as a "gap".

3.2 Multivariate UMIDAS Models

Foroni et al. (2015) studied the performance of a variant of MIDAS which does not resort to functionally distributed lag polynomials. They discussed how an unrestricted MIDAS (UMIDAS) regression could be derived in a general linear dynamic framework, under which conditions the parameters of the underlying high-frequency model can be identified.

The UMIDAS is the model when $\boldsymbol{y_{tm}}$ (low-frequency) is regressed on its quarterly lags and on lags $\boldsymbol{x_{itm}}$ (high-frequency N variables); then, the model can be written as

$$\boldsymbol{c(L^m)\, y_{tm} = \delta_1(L) x_{1tm} + \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots + \delta_N(L) x_{jtm} + \varepsilon_{tm}} \quad \textit{(7)}$$

Where $\quad c(L^m) = (1 - c_1 L^m - \ldots\ldots\ldots\ldots - c_c L^{mc}),$

$$\delta_j(L) = \left(\delta_{j,0} + \delta_{j,1} L + \cdots \ldots + \delta_{j,v} L^v\right),$$

This model is estimated at low-frequency, uses high-frequency regressors, and can be re-estimated for each month within the quarter. UMIDAS is linear; therefore, it can be estimated by Ordinary Least Squares (OLS), where t=1 …T and m represents the months of the quarter.

We used a form of direct estimation and constructed the forecast as:

$$\bar{\mathbf{y}}_{\mathbf{T_M^X+m/\,T_M^X}}=\bar{\mathbf{c}}(\mathbf{L^k})\mathbf{y}_{\mathbf{T_M^X}}+\overline{\boldsymbol{\delta}_1}(\mathbf{L})\mathbf{x}_{\mathbf{1T_M^X}}+\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots+\overline{\boldsymbol{\delta}_N}(\mathbf{L})\mathbf{x}_{\mathbf{NT_M^X}} \quad \mathbf{(8)}$$

Where the polynomials $\bar{C}(z)=\bar{c}_1 L^{\,m}+\ldots\ldots\ldots\ldots+\bar{c}_c L^{mc}$ and $\bar{\delta}_i(L)$ are obtained by projecting y_{tm} on information dated $mtm-m\ or\ earlier, for\ t=1,2,\ldots\ldots.T_m^X$.

In general, the direct approach can also be extended to construct $h_m-\ step\ ahead$ forecasts given T_M^X:

$$\bar{\boldsymbol{y}}_{\boldsymbol{T_M^X+hm/\,T_M^X}}=\bar{\boldsymbol{c}}\left(\boldsymbol{L^k}\right)\boldsymbol{y}_{\boldsymbol{T_M^X}}+\overline{\boldsymbol{\delta}_1}(\boldsymbol{L})\boldsymbol{x}_{\boldsymbol{1T_M^X}}+\cdots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots+\overline{\boldsymbol{\delta}_N}(\boldsymbol{L})\boldsymbol{x}_{\boldsymbol{NT_M^X}} \quad \textbf{\textit{(9)}}$$

Where the polynomials $\bar{C}(z)$ and $\bar{\delta}_i(L)$ are obtained by projecting y_{tm} on information dated $mtm-hm\ or\ earlier, for\ t=1,2,\ldots\ldots.T_m^X$.

Additionally, in the case of UMIDAS, an autoregressive term can be easily included without any common factor restrictions.

3.3 Mixed-Frequency Three-Pass Regression Filter

An (OLS) approach is the Mixed 3–Pass Regression Filter of Hepenstrick and Marcellino (2015):

$$y_{t+h}=\beta_0+\beta' F_t+\eta_{t+1} \qquad (10)$$

$$Z_t=\lambda_0+\boldsymbol{\Lambda F_t}+\omega_{\boldsymbol{t}} \qquad \mathbf{(11)}$$

$$x_t=\phi_0+\boldsymbol{\phi F_t}+\varepsilon_{\boldsymbol{t}} \qquad \mathbf{(12)}$$

where $\boldsymbol{y_t}$ is the target variable of interest, $F_t=(f_t', g_t')$ are the $K=K_f+K_g$ common driving forces of all variables, the unobservable factors $\beta=\left(\beta_f', 0'\right)$, so that y only depends on f; Z_t is a small set of L proxies driven by the same underlying forces as y, such that $\Lambda=\left(\Lambda_f, 0\right)$ and Λ_f is nonsingular, x_t is a large set of N variables driven by both f and g and *t=1,……T.*

We can estimate the model by a three-step algorithm:

0) Aggregate monthly dataset to quarterly frequency
1) For each variable in the quarterly dataset, x_i, runs a (time series) regression of x_i on the proxy z: $x_t^{(i)}=\boldsymbol{\phi}_0^{(i)}+z'\boldsymbol{\phi_i}+\boldsymbol{\varepsilon}_{\boldsymbol{t}}^{(i)}$ **(13)**
2) Using the OLS estimates $\widehat{\boldsymbol{\phi_i}}$ obtained in the previous step, we run a cross-sectional regression of $x_t^{(i)}$ on $\widehat{\boldsymbol{\phi_i}}$ for each month in the monthly dataset:

$$x_t^{(i)} = \boldsymbol{\phi}_0^{(t)} + \widehat{\boldsymbol{\phi}_\iota}'^{\boldsymbol{F_t}} + \boldsymbol{\varepsilon}_{\boldsymbol{t}}^{(\boldsymbol{i})} \quad \textbf{(14)}$$

3) Mixed-frequency techniques with OLS estimate $\widehat{F_t}$ to forecast y_{t+h}. Here, the time series regression of $\boldsymbol{y_{t+h}}$ **on** $\widehat{F_t}$:

$$y_{t+h} = \beta_0 + \beta' F_t + \eta_{t+1}$$

4. Tunisian Economic Growth: Narrative Analysis

This section follows the same procedure that was developed before in nowcasting and forecasting aggregate RGDP, but we replace aggregate GDP with its sectoral added value. Therefore, we will take each disaggregated component of GDP and nowcast them based on a monthly dataset of release indicators by using mixed-frequency data sampling as unrestricted autoregressive MIDAS (UMIDAS-AR), three-pass regression filter, and mixed dynamic factor models (MDFM).

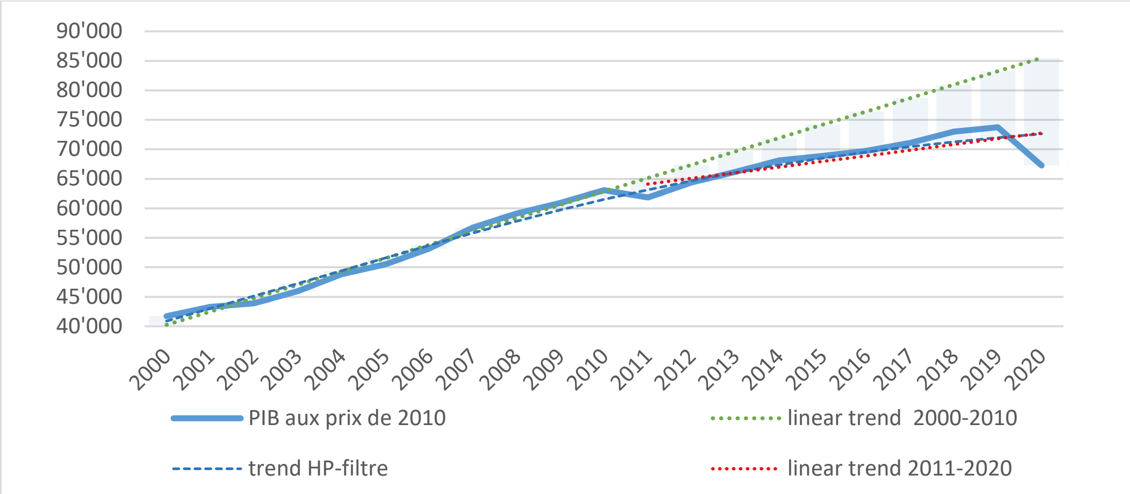
4.1 Stylized Facts of the Tunisian Economy:

- **Decrease in growth from 2011 and decline in potential growth reflecting structural weaknesses**:

In the last decade, the Tunisian economy has been subjected to significant internal and external shocks that have largely impacted its potential growth. The COVID-19 pandemic hit Tunisia hard, leading to an unprecedented economic downturn. Real GDP was estimated to have contracted by 8.8 percent in 2020, the largest economic downturn since the country's independence.

The pandemic aggravated Tunisia's long-standing vulnerabilities, stemming from persistent internal instability, loss of export market share, and decline in productivity.

Figure (1): Evolution of GDP in level (in millions of dinars)

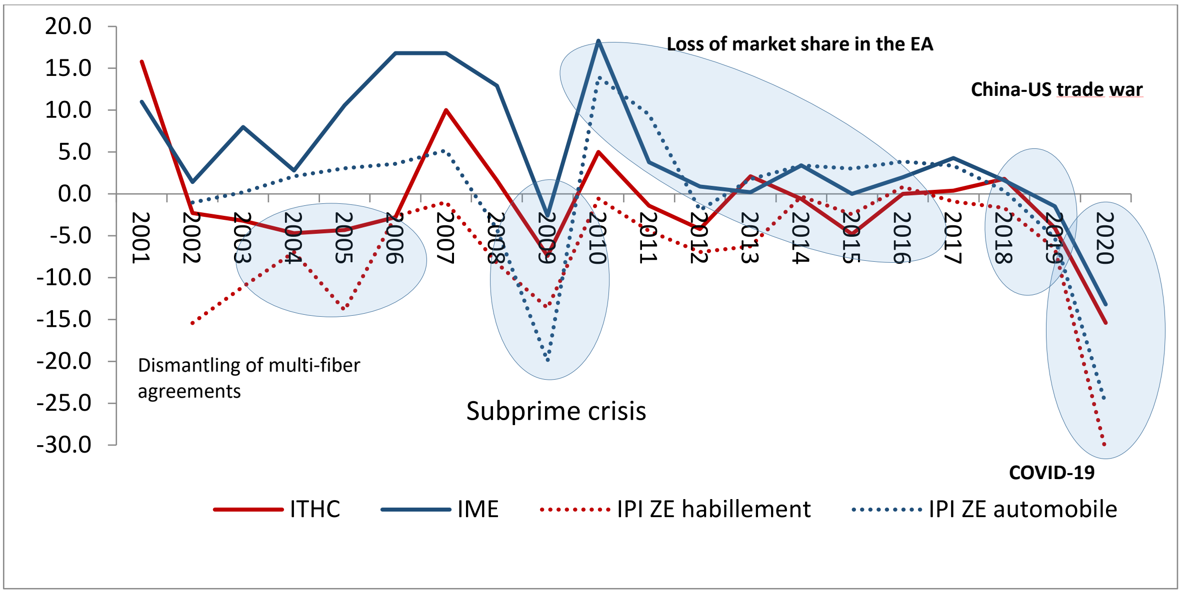


- **Dependence of the eurozone economic cycle:**

The Tunisian economy depends closely on the European cycle since this partner represents more than three-quarters of Tunisian exports, particularly in the mechanical and electrical industries and textiles and clothing, which are the key sectors of the Tunisian industry with a share of 45% and 21% of goods exports, respectively, in 2020.

Thus, it is useful to use indicators that provide an idea of external demand, such as the production index or eurozone business climate indicator.

Figure (2): Tunisian textile and mechanical and electrical industries with Euro Area industrial production index



- **Multitudes of shocks:**
  - ✓ **2004: Dismantling of multi-fibre agreements**
  - ✓ **2009: Subprime crisis**
  - ✓ **2011: revolution**

  Political and social instability affects many sectors, especially phosphates, derivatives, and energy (in addition to drawing on natural reserves).

  - ✓ **2015: terrorist attacks**
  - ✓ **2020: Impact of COVID-19 including supply and demand chocs**
  - ✓ High volatility of agricultural production dependent on climatic conditions

- **Growth drivers:**
  - ✓ Since 2011, growth has been driven mainly by consumption, fuelled by increases in public and private sector wages, against a downward trend in investment and underperformance of exports.
  - ✓ Consumption decline, especially from 2018: tightening of monetary and fiscal policy.

- **Post COVID forecasting difficulties (2020):**
  - ✓ Impact of the health crisis, containment measures, and uncertainties on:
    Exporting sectors: estimated through external demand
    - sectors linked to local demand (construction, trade, and other market services) which are not captured by the available explanatory variables
  - ✓ Negative contribution from the public sector (-1.2 pp in 2020) linked to the application of exceptional administrative timing

- **Estimation of quarterly GDP by the National Statistical Institute:**

Depending on the availability of economic indicators relating to the various economic sectors included in GDP, the National Statistical Institute used both calibration and smoothing techniques.

  - ✓ Estimated added value by calibration on a cyclical indicator: industry except building and civil engineering (industrial production index), hotels (overnight stays), etc.
  - ✓ Estimation of value added (VA) using other techniques in the absence of indicators:
    - o Smoothing: Agriculture, Café and Restaurant, Telecommunications, Other market services

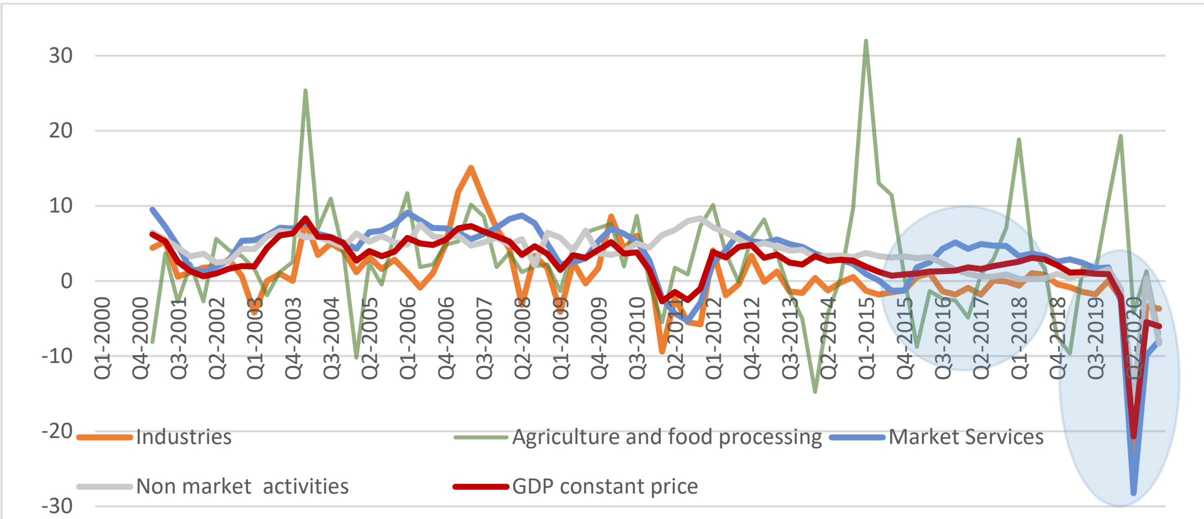
- Adjustment concerning the VA of other sectors: Trade (VA of agriculture and industry), indirect taxes net of subsidies (total VA)

Based on the GDP construction method, variables such as the industrial production index and overnight stays correlate with GDP. However, publication times led us to search for other variables with shorter publication times.

4.2 Sectoral growth analysis

The analysis of quarterly growth shows an important relationship with industrial activity and market services, while the added value of agriculture is experiencing significant volatility linked to dependence on climatic hazards and the olive tree production cycle.

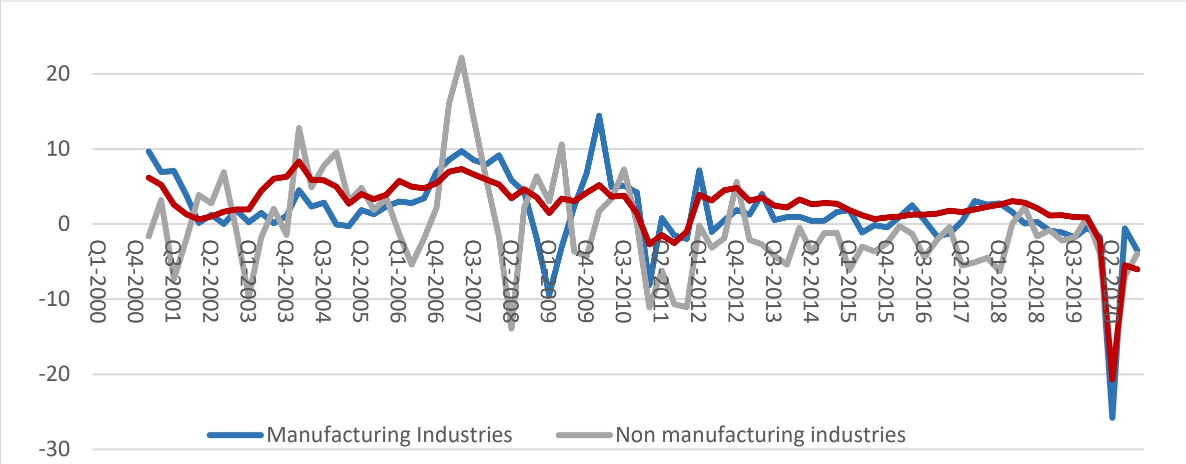
Figure (3): Growth rates by sector



Thus, the analysis of the origins of growth shows that it is the most volatile sector and is dependent on the economic cycle, such as manufacturing industries, which drive the pace of growth. This explains the importance of the indicators in these sectors for modelling short-term growth.

Regarding the change in the trend since 2017 (lower correlation between the growth of industry and services with that of GDP), this should be explained by internal factors (social and political instability, loss of competitiveness, etc.), which cause growth to move away from the European economic cycle (first customers for manufacturing industries and tourism).

__Figure (4): GDP, Manufacturing and non-manufacturing industries__



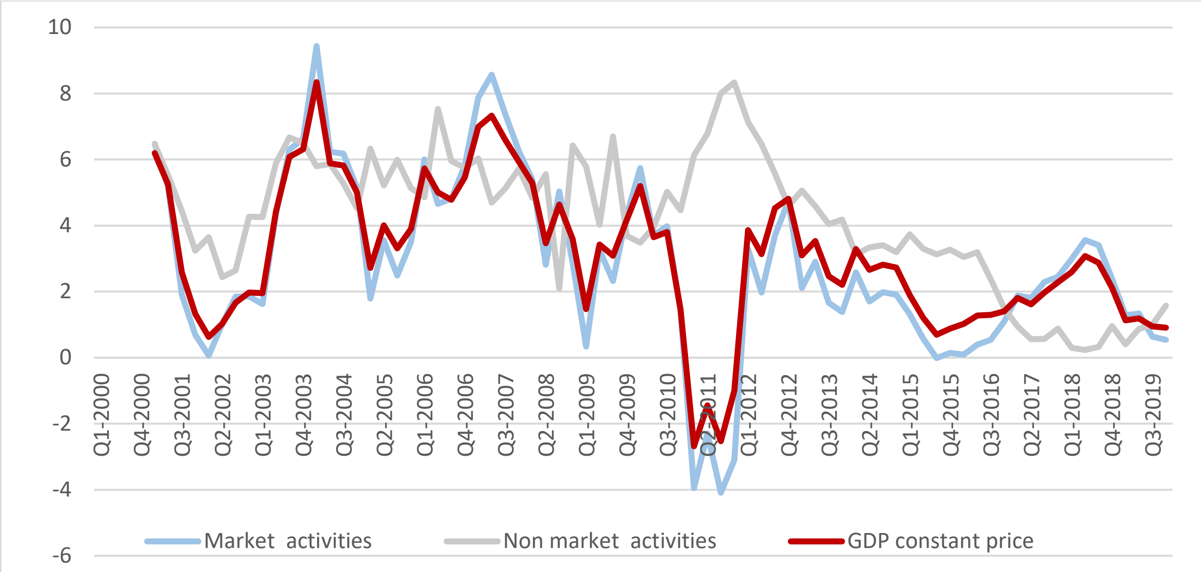
- **Sectoral growth contribution:**

__Table (1): Growth contributions__

Agriculture	Very Volatile: 1.2 percentage points in 2018 vs. -0.9 points in 2016
Manufacturing industries	Weakening of the contribution from 2011: dependence on Euro area and availability of phosphates (chemical industries)
Non-Manufacturing industries	Negative contribution (structural and cyclical problems)
Market services	Usually positive and important except for years of external shock: 2011 (-1,5 pp), 2015 (-0,2 pp), 2020 (-5,7 pp)
Non-market services activities	Significant contribution of the public sector, especially during the period 2011-2013 (1 pp on average) then deceleration (0.1 pp over 2017-2019): public wage bill

- **2015-2016:** Given the growth weakness (terrorist attacks in 2015, political instability, etc.), agriculture's contribution to growth was very significant (1.1 percentage points and -0.9 points respectively; 94% and 79% of growth). On the other hand, the contribution of public administration was around 0.6% in 2015 and 2016 (51% and 42.1% growth, respectively). It should be noted that the contribution of public administration added value weakened from 2017 (tightening of fiscal policy after a very expansionary policy between 2011 and 2016).

Figure (5): GDP, market activities and non-market activities



- **2016-2017:** The performance of market services (gradual recovery of tourism after the 2015 terrorist attacks) was greater than that of GDP (growth in industry dragged down overall growth), particularly in non-manufacturing industries (social protests and continuous depletion of nature reserves and lack of investment).

- **2018-2019**: continued underperformance of the industry compared to overall growth: problems in the extractive activity (Energy and phosphate).

- **2020:** The decline in growth is largely attributable to market services (tourism and transport, trade, etc.) and, to a lesser extent, to manufacturing industries (mechanical industries, textiles, and clothing) and construction. In addition, the growth of non-market services has experienced a significant decline (accounting for exceptional hours and remote work).

5. Data Selection

The selection of indicators was based on the calculated correlation matrix between the aggregated (to quarterly frequency) monthly indicators, aggregated GDP, and disaggregated components of GDP. However, indicators must also have a strong economic relationship and a reasonable publication time.

We collected a dataset with values from January 2000 up to the most recent observations at the time of writing (June 2021). The dataset contains 33-quarterly frequency observations for components of GDP at constant prices by output side, which are defined as the value-added of the main branches of Tunisian economic activity compiled by INS. The data cover 85 monthly indicators of economic and financial variables spanning hard indicators such as electricity consumption and industrial production index by activity sectors, the service sector as Air Transports and Tourists Nights; Natural resources production as phosphate production, Crude oil production; the Financial and monetary sector as TUNINDEX (stock market) by sector, credits to the economy, credit card payments, financial services, aggregate money in the sense of M3, Central Bank balance sheet, Net foreign assets; the International sector as Industrial Production Manufacturing Index of Eurozone, Energy prices and Manufacturing Confidence Indicator, and Employment as job offer and job demand. However, only a few monthly indicators are available for the agricultural sector.

Concerning data processing and seasonal adjustment, before modelling we need to:

- Ensure data are stationary; we have to difference/log differences and remove low-frequency trends.
- Possibly standardise data.
- Seasonally Adjusted data (adding National Calendar Dates).

Tables (1)–(4) in the annexes present the ragged-edge structure of our dataset, including both quarterly and monthly variables.

Currently, we have a mixed-frequency database characterised by publication delays that differ from one indicator to another and a significant number of missing observations.

Our objective is to develop mixed-frequency models that allow us to manage the “ragged-edge” structure of the data because of the publication delays of monthly and quarterly predictors. In fact, MDFM constitute an adapted approach to solve this issue and have several advantages:

- Analyse the intra-quarter dynamics of economic fluctuations conditional on quarterly releases of RGDP.
- Solve the issue of ragged edges and missing observations of data using the prediction routine.
- Construct a monthly index of economic activity.

6. The Monthly Index Of Economic Activity Using MFDFM

6.1 Selection of monthly indicators by sector

As for the variable selection, a wide set of monthly predictors are selected by each sector as the manufacturing, non-manufacturing, and service components of GDP. This selection is based on the calculated correlation matrix between the monthly indicators and disaggregated components of GDP. We chose indicators that were superior to 0.17.

For manufacturing GDP:

We selected 9 monthly indicators:

- Industrial production index (IPI) -manufacturing industries
- IPI -2010
- Manufacturing confidence indicator Zone Euro
- TUNINDEX
- Export agriculture and food industry
- Consumer goods
- Industries automobiles and equipment
- Basic materials
- Tourists nights

<u>Figure (6): growth rates of GDP manufacturing and the long history monthly indicator</u>

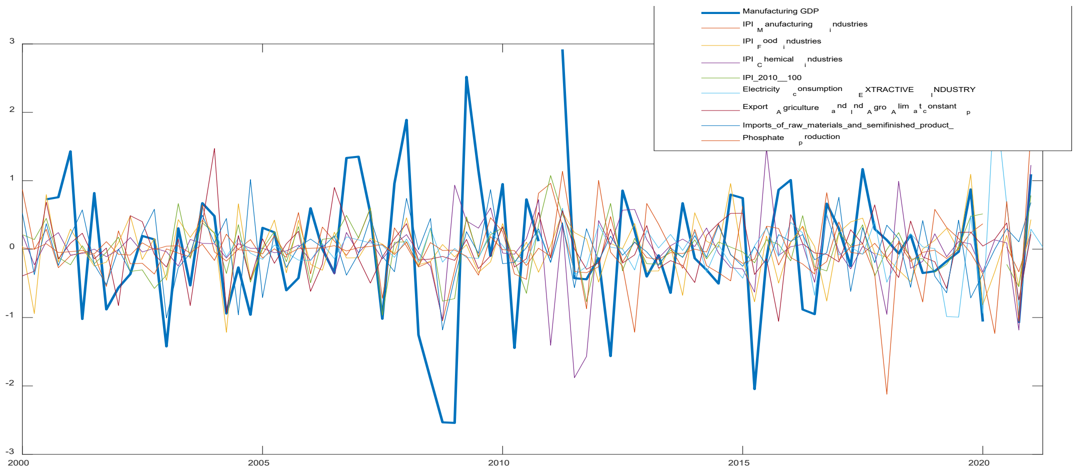
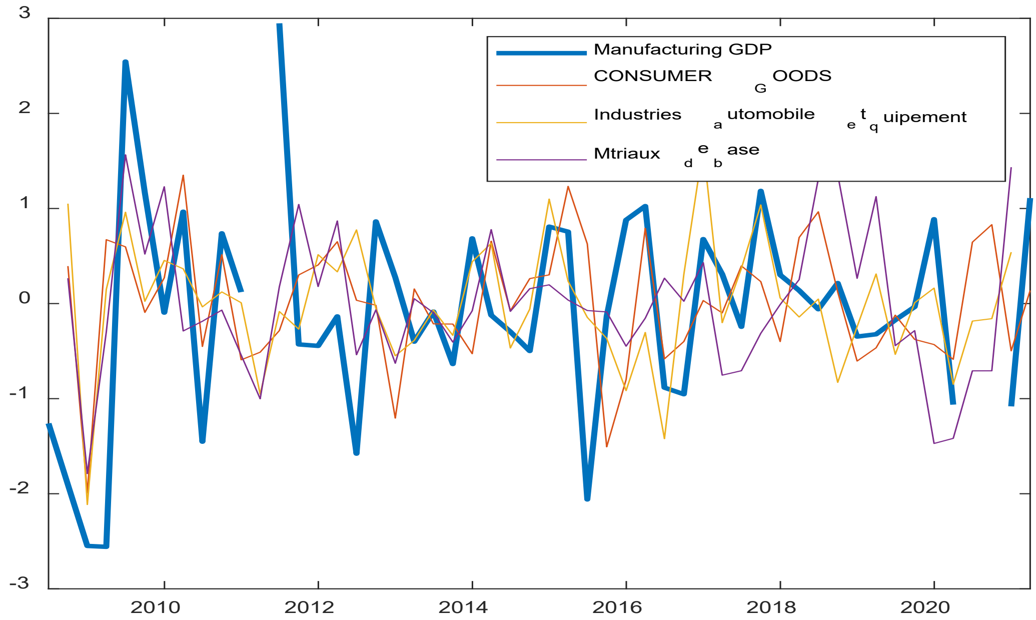


Figure (7): growth rates of manufacturing GDP and the short history monthly indicators



The blue line is the growth rate of manufacturing GDP, and the other colours are the different monthly predictors. Figure 7 shows that the IPI index by sector and exports by sector follow the manufacturing GDP fluctuations and trend, reflecting its significant correlation.

For non-manufacturing GDP:

we selected 11 monthly indicators:

- IPI-energy
- IPI-crude oil natural gaz
- Electricity consumption extractive industry
- Electricity consumption pumping
- Electricity consumption tourism
- Electricity consumption services
- Export other manufacturing industries
- Natural gas production
- Products and personal care
- Industries
- Buildings and materials

Figure (8): growth rates of GDP non-manufacturing and the long history monthly indicators

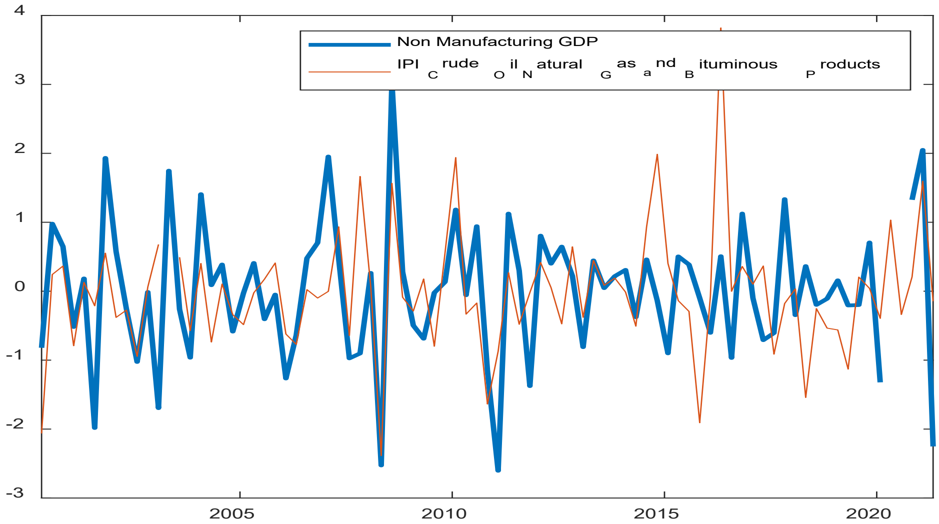


Figure (9): growth rates of GDP non-manufacturing and Consumption electricity monthly indicators

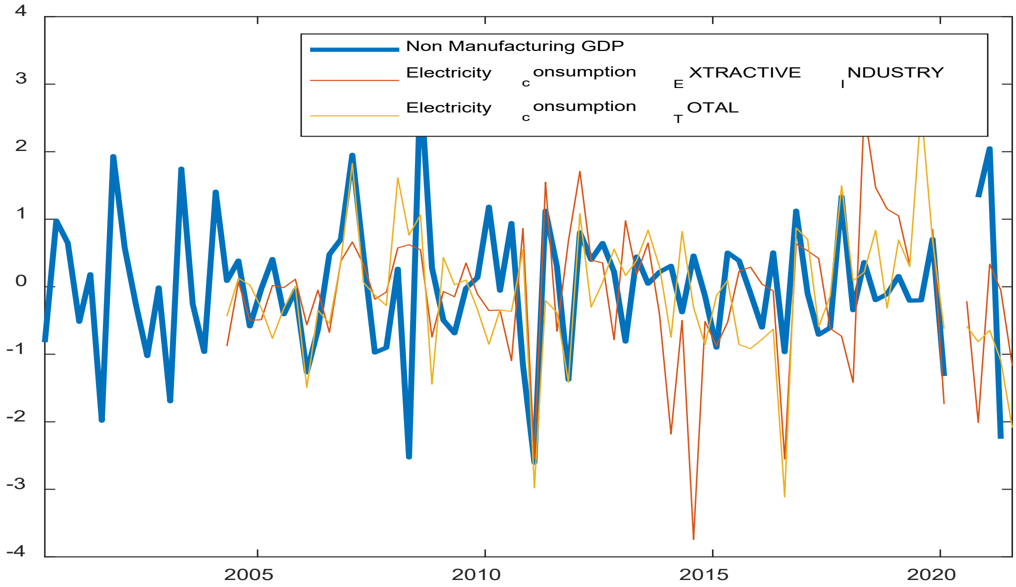
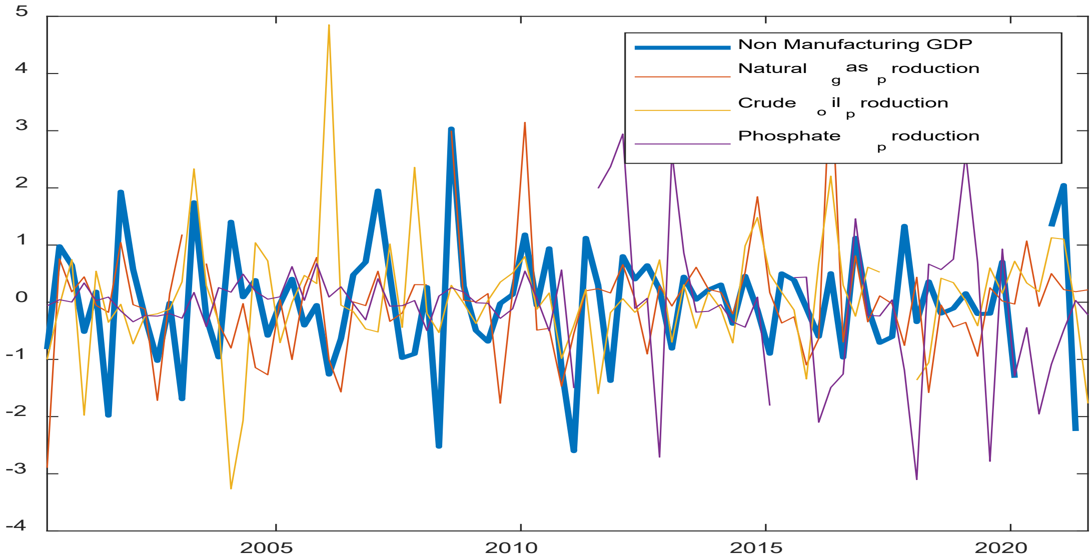


Figure (10): growth rates of GDP non-manufacturing and the long history monthly indicators

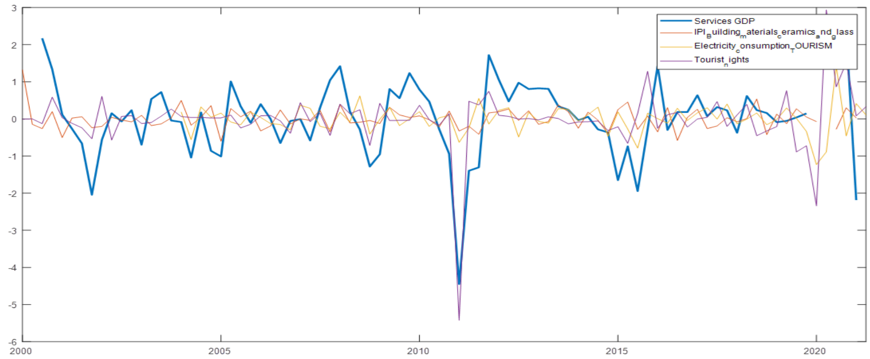


The blue line is the growth rate of non-manufacturing GDP, and the other colours are the different monthly predictors. As shown in figure 10, the natural gas production and electricity consumption by sector follow the fluctuations and the trend of manufacturing GDP, reflecting its significant correlation with it.

For GDP services, we selected these indicators:

- IPI building ceramics,
- Electricity consumption tourism,
- Tourists nights,
- Ipi-manufacturing euro area,
- Consumer services,
- Distribution,
- Industries,
- Banks,
- Insurance,
- Services

Figure (11): growth rates of GDP services and the short history monthly indicators



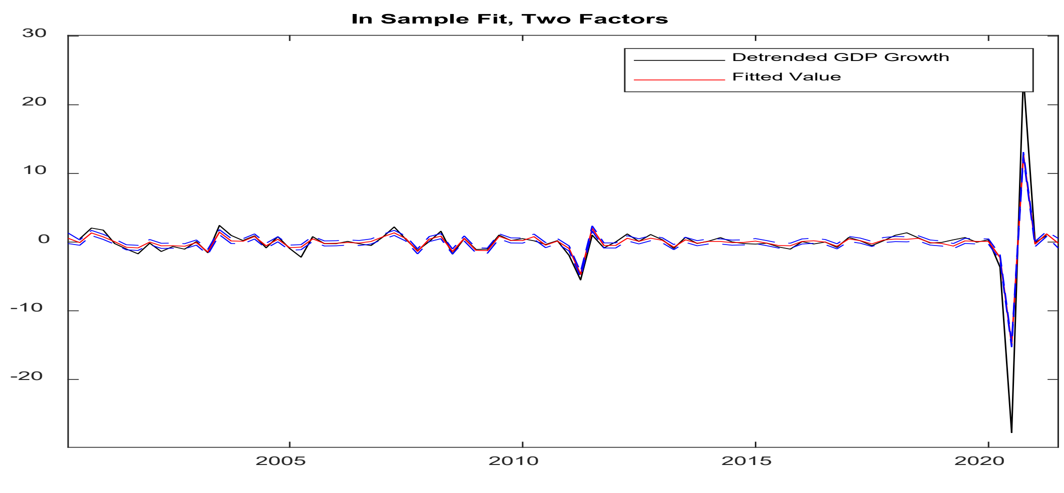
In figure 11, the blue line corresponds to the growth rate of GDP services, and the purple line is the growth rate of tourists' nights. This indicator follows the fluctuations in GDP, especially the significant decline in the crisis periods of 2011 and 2020.

6.2 Specify blocks and estimate the model: 2 factor model

We may want to have only one factor for a single economic activity index. Alternatively, we can have several factors and divide them into blocks. Furthermore, the two-factor model was the most precise among the models with three or one global factor. We identify these two blocks as real and nominal, respectively. (Figure 5 in annexes).

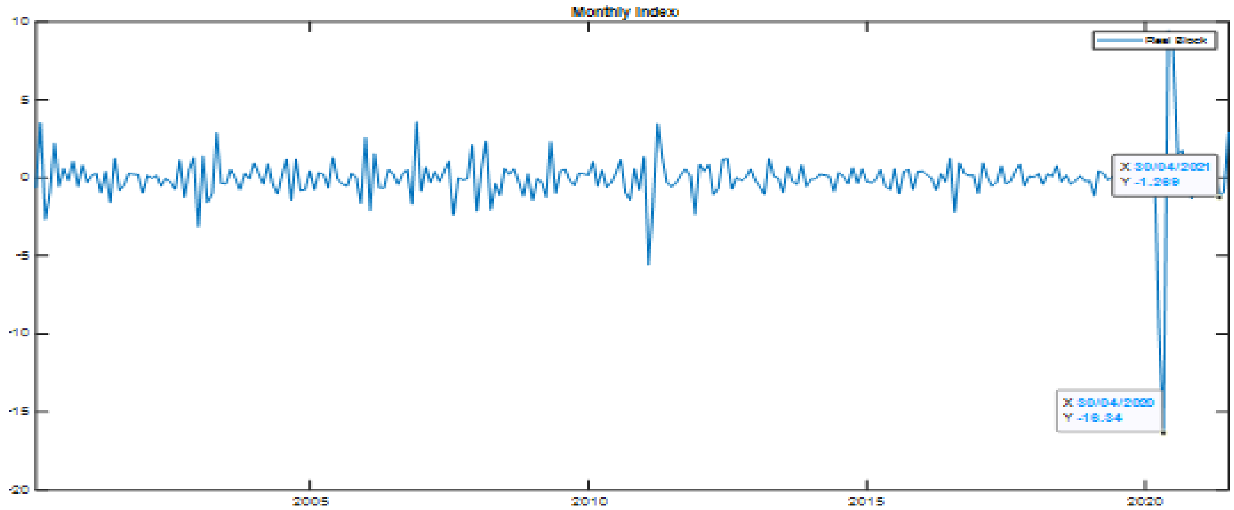
We applied the EM algorithm to estimate monthly GDP values. Furthermore, if the GDP release is not yet available, but some monthly indicators are, the EM algorithm can be employed to obtain an estimate of the corresponding quarterly GDP.

Figure (12): The DFM estimation with two factors



This figure displays the time series used to estimate the dynamic factor model together with the estimate of the common factor; the estimated factor is reported in standardised units. GDP and the factor are reported in annualised growth units in the bottom panel.

Figure (13): calculation of the monthly activity index

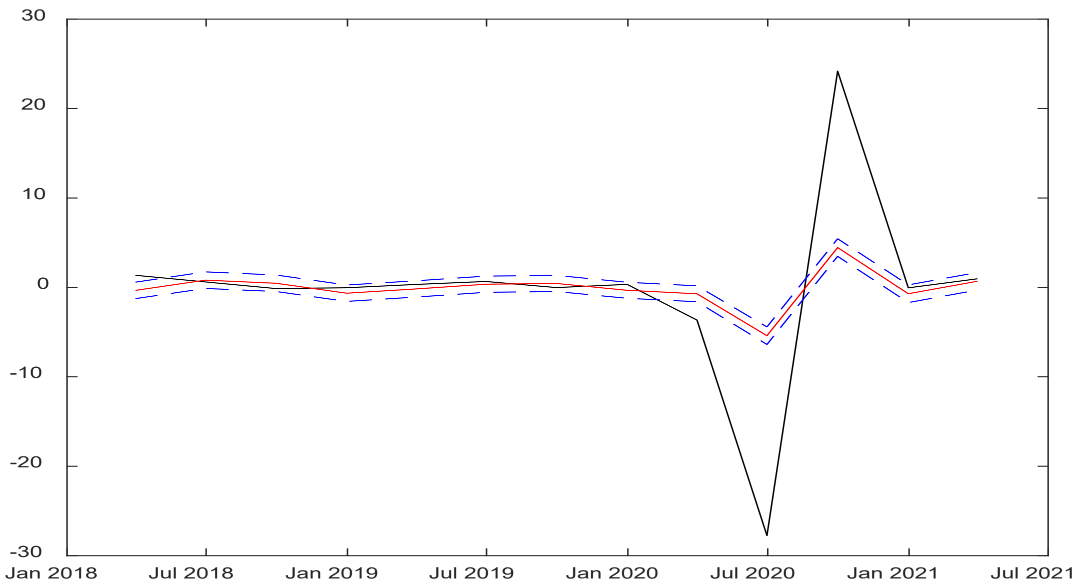


Taking only the real activity block, we have a standardised index where the value zero corresponds to the average growth rate of real economic activity observed from January 2000 to June 2021. When the index level is above (below) the zero value, economic conditions progress (weaken) relative to the average growth rate of the conditions in the Tunisian economy.

6.3 Out-of-sample evaluation of DFM model:

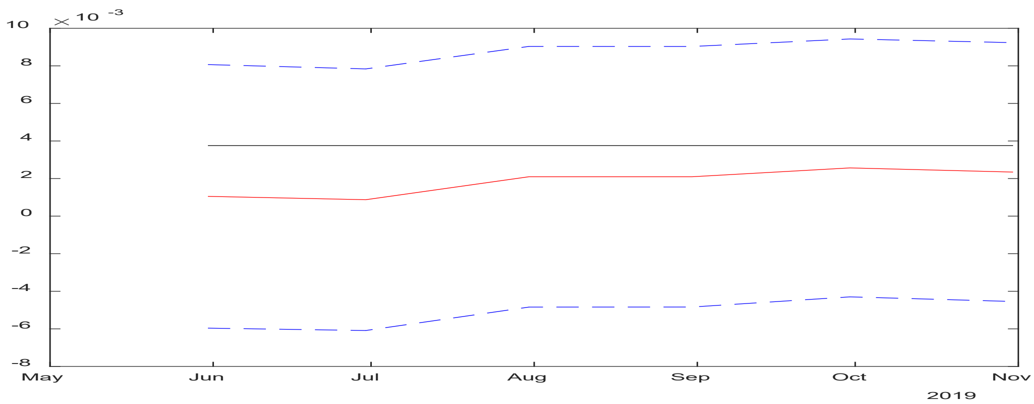
We attempted to consider the backtest from March 2019 to April 2020. Ideally, backtesting should use vintage data to control for publication dates and data revisions. Unfortunately, we did not have vintage data on these dates. Therefore, we used the pattern of missing observations in the tails of the data. Our objective was to recreate this pattern of missing observations in the tail of the data for each backtest date. The question is, what do the backtest GDP values look like from one month after the reference date, which means around one month before the true GDP values are published?

Figure (14): Backtest DFM



COVID 19 had very large volatility. From Figure 14, it seems that the backtest values move in the correct direction, but the magnitude over COVID is too small. We are interested in how the Q2 2019 estimate evolves over time.

Figure (15): six predictions for GDP in Q2 2019



From Figure (15), the black line is the true GDP for Q2 2019 and the nowcast of GDP. In fact, we started further away in the long-run mean, which has detrended data, and as we obtained more data, we became closer to the true GDP.

7. Forecasting Disaggregated GDP Via UMIDAS

In this section, we regress each output side component of real GDP on the stacked monthly indicators released in the first, second, and third months and compare their predictive ability by computing the RMSE:

We stacked the variables of ${xm_{yq1}}^{(1)} = x_{ym01}$

$${xm_{yq2}}^{(2)} = x_{ym02}$$

$$(15) xm_{yq3} = x_{ym03}$$

The UMIDAS regression is defined as follow:

$$RGDP_{tq} = \alpha_1 + \beta_1 xm_{tq}^{(1)} + \beta_2 xm_{tq}^{(2)} + \beta_3 xm_{tq}^{(3)} + u_{tq} \quad (16)$$

7.1 Modelling manufacturing for GDP estimating

This section uses a multivariate unrestricted MIDAS model incorporating the relevant indicators that have predictive power in nowcasting manufacturing GDP. Nineteen right hand side (RHS) variables were included. An increase in RHS variables required more shrinking of parameter estimates toward zero to avoid overfitting; then, the model was estimated by ridge regression.

We specify two models–the first contained variables with a long history. The second model used financial variables with a short history.

In the first model, we regress real GDP manufacturing on the stacked monthly indicators MMP1, which contains one and two months' worth of real indicators. In the second model, we stacked the financial indicators on MMP2, which contains three months, and released financial indicators with term lag 1:

We define:

$${xm_{yq1}}^{(1)} = x_{ym01}$$

$${xm_{yq2}}^{(2)} = x_{ym02}$$

$$xm_{yq3} = x_{ym03}$$

$$\text{MMP1} = [\ lag(1), \underbrace{X2,,}_{IPI\ manufacturing\ industries}, \underbrace{X3,}_{IPI\_2010}, \underbrace{X4,}_{Exports\ \_agroalimentaire}, \underbrace{X5,}_{tourists\ nights},, \underbrace{X6, X7,}_{manufacturing\ confidence\ indicators} \underbrace{X8, X9}_{Tunindex},].$$

$$\text{MMP2} = [\ lag(1), \underbrace{X2, X3, X4}_{consumer\ goods}, \underbrace{X5, X6, X7}_{industries\ automobiles}, \underbrace{X8, X9, X10}_{mtriaux\ de\ base},].$$

Then, we pooled the fitted values of manufacturing GDP. Table (2) reports the regression results on real indicators:

Table (2): Multivariate AR-UMIDAS Estimation using real long history monthly indicators

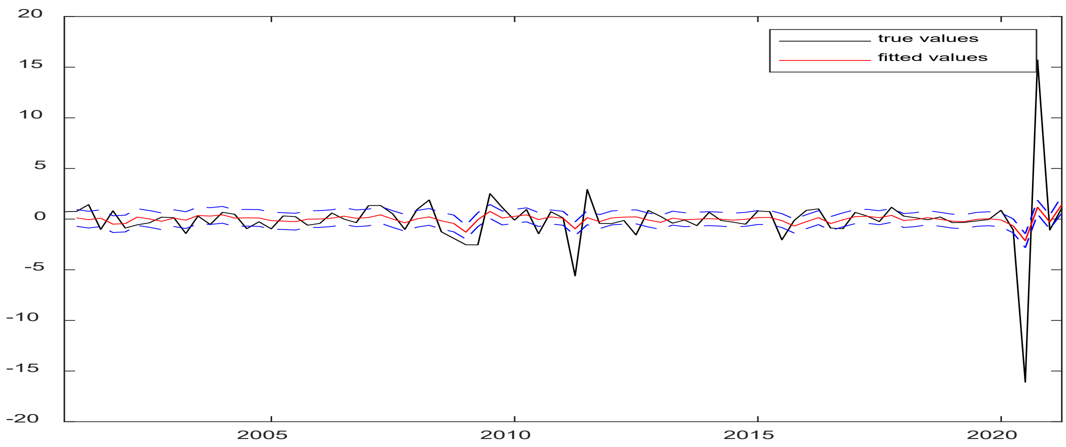
Variables	Coefficients	Std	t-student
AR(1)	-0.28	0.05	-5.86
X2*	0.19	0.11	1.71*
X3	0.09	0.12	0.77
X4	-0.007	0.12	-0.05
X5	0.497	0.059	8.30
X6*	0.30	0.10	3.02
X7	-0.005	0.10	-0.04
X8	0.18	0.08	2.25*
X9	0.08	0.12	0.70
R-squared: 0,6636			
Adjusted R-Squared:0,5917			

The **multivariate AR-UMIDAS** yields an R^2 which is on the order of 66%.

Table (3): Multivariate AR-UMIDAS Estimation using short-horizon financial monthly indicators

Variables	Coefficients	Std	t-student
AR(1)	-0.44	0.08	-4.98
X2*	0.20	0.20	0.99
X3	0.05	0.22	0.26
X4	0.19	0.12	1.59
X5	-0.05	0.20	-0.24
X6*	0.019	0.22	0.08
X7	0.44	0.21	2.04
X8	-0.34	0.22	-1.52
X9	0.32	0.22	1.44
X10	-0.05	0.21	-0.26
R-squared: 0,6236			
Adjusted R-Squared:0,5817			

Figure (16): Multivariate UMIDAS fitted Manufacturing RGDP



The dynamics align well with the Tunisian RGDP Manufacturing growth, which declined sharply in 2011 (revolution: recession of the economy) and 2020 (covid_19). However, note that the magnitudes of the predictions in these crises are significantly less than the realised values.

7.2 Modelling non-manufacturing for GDP estimation

We specified three models, the first contained variables with long histories. The second model used electricity consumption by sector, and the third used financial variables with a short history.

In the first model, we regressed quarterly non-manufacturing GDP on NM1, which contained the first month's releases of long history real indicators with autoregressive term lag1:

$$\mathrm{NM1} = [\ lag(1),\ \underbrace{X2,}_{IPI\ energy},\ \underbrace{X3,}_{IPI\_crude\ oil},\ \underbrace{X4,}_{Exports\ \_other\ manufacturing\ industries},\ \underbrace{X5,}_{Natural\ gaz\ production}\ ].$$

Table (4): Multivariate AR-UMIDAS Estimation using long-history real indicators

Variables	Coefficients	Std	t-student
AR(1)	-0.08	0.07	-1.05
X2*	0.09	0.09	1.07
X3	0.08	0.10	0.82
X4	0.28	0.07	3.98
X5	0.14	0.09	1.55

R-squared:0,5736

Adjusted R-Squared:0,5417

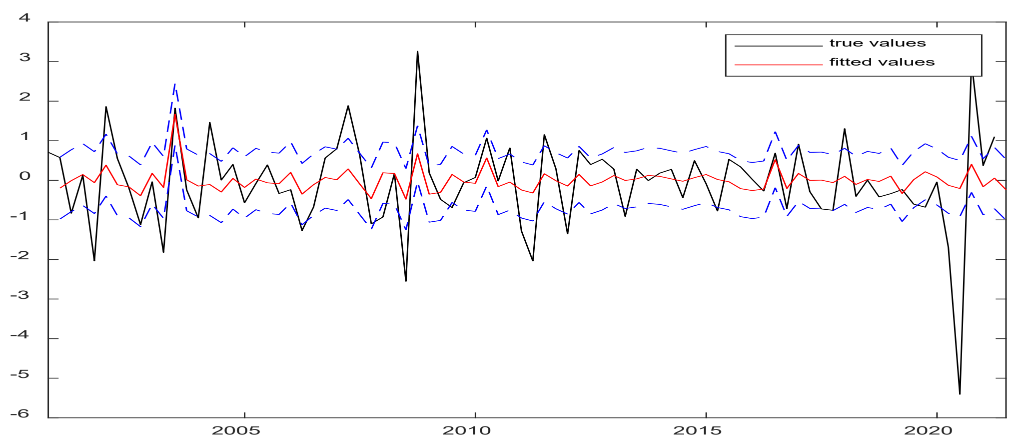
The second equation includes the updated series of electricity consumption by sector on the RHS. Since the data are from different vintages, we must select how many months are currently observed for electricity consumption by sector:

$$\text{NM2}=[\ lag(1),\ \underbrace{X2,,}_{electricity\ consumption\ extractive\ industry},\ \underbrace{X3,}_{electricity\ consumption\ pumping},\ \underbrace{X4,}_{electricity\ consumption\ tourism},\ \underbrace{X5,}_{electricity\ consumption\ services}\ ].$$

In the third model, we regressed quarterly non-manufacturing GDP on NM3, which contained the first month's releases of short history financial indicators with autoregressive term lag1:
For financial indicators, we selected the three months:

$$\text{NM3}=[\ lag(1),\ \underbrace{X2,X3,X4}_{industrries},\ \underbrace{X5,X6,X7}_{Batiment\ de\ constructions},\ ].$$

Figure (17): Multivariate UMIDAS fitted Non- Manufacturing RGDP



7.3 Modelling services for GDP estimation

From the services sector, we specified two models. The first model contained variables with a long history. The second model used financial variables with a short history.

We regressed quarterly services GDP on S1, which contained the long history months, and released real indicators with autoregressive term lag1:

$$\text{S1}=[\ lag(1),\ \underbrace{X2,,}_{IPI-ceramics},\ \underbrace{X3,}_{tourists\ nights},\ \underbrace{X4,X5}_{Tunindex},\ \underbrace{X6,}_{manufacturing\ index\ indicator}\ ].$$

Figure (18): Multivariate UMIDAS fitted services RGDP

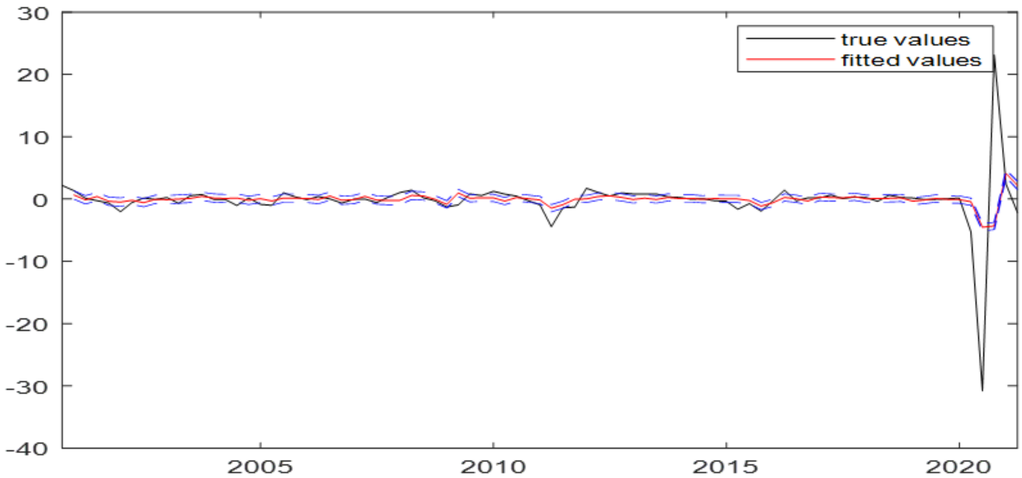


Table (5): Multivariate AR-UMIDAS Estimation using long-history real indicators

Variables	Coefficients	Std	t-student
AR(1)	0.19	0.06	3.166
X2*	0.13	0.06	2.16
X3	0.20	0.06	3.12
X4	0.11	0.05	2.043
X5	0.05	0.06	0.84
X6	0.06	0.066	1.036
R-squared :0,7763			
Adjusted R-Squared :0,6417			

We regressed quarterly services GDP on S2, which contained the monthly releases of financial indicators with autoregressive term lag1:

$$S2 = [\ lag(1),\ \underbrace{X2,,}_{Electricity\ consumption\ tourism},\ \underbrace{X3, X4}_{Consumer\ services},\ \underbrace{X5, X6}_{Industries}\ \underbrace{X6,}_{Banques}\ ].$$

7.4 Modelling GDP from its components:

The different weights of the main components of GDP were determined via an OLS regression of aggregated GDP on components by the output side given in Table (9):

Table (6): Weights of the different components of GDP

By weighting the different estimations of the components of GDP at constant prices (manufacturing GDP, non-manufacturing GDP, and services GDP).

Figure (19): Multivariate UMIDAS fitted RGDP

8. Backtest Out-of-Sample

Sample performance is not indicative of how the model will perform going forward in time; out-of-sample nowcasting is required to test the performance of the model.

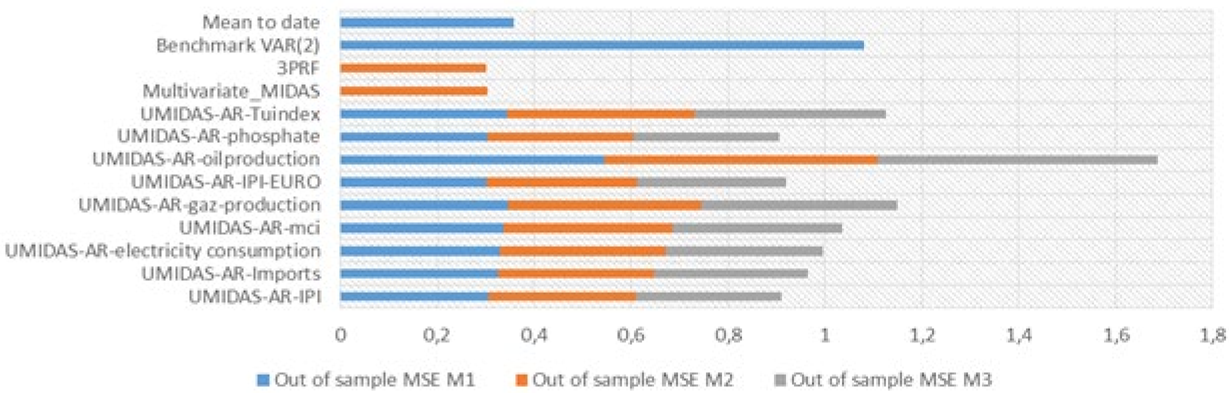
Thus, we backtested the model over the last 30 periods of the data (30/06/2012 to 30/09/2019), using the data up to the current backtesting date to estimate the parameters.

Components of GDP	***Coefficients***	***SE***	***T value***
Manufacturing	***0.3488***	***0.060819***	***5.7365***
Non-Manufacturing	***0.55792***	***0.065737***	***8.4872***
Services	***0.4362***	***0.069598***	***6.2679***

Table (7) Mean squared error (*MSE) performance of UMIDAS-AR models (out-of-sample)*

Models	Out-of-sample MSE M1	M2	M3
UMIDAS-AR-IPI	0,3071	0,3022	0,3211
UMIDAS-AR-Imports	0,3258	0,3228	0,317
UMIDAS-AR-electricity consumption	0,3306	0,3413	0,3041
UMIDAS-AR-mci	0,3371	0,3482	0,3497
UMIDAS-AR-gaz-production	0,3457	0,3998	0,4048
UMIDAS-AR-IPI-EURO	0,3026	0,3084	0,3078
UMIDAS-AR-oilproduction	0,545	0,565	0,5775
UMIDAS-AR-phosphate	0,3036	0,3011	0,301
UMIDAS-AR-Tuindex	0,3441	0,3262	0,3049
Multivariate_MIDAS		**0,3034**	
Mixed Dynamic factor model		**0.13654**	
3PRF		**0,3012**	
Benchmark VAR(2)		1,0808	
Mean to-date		0,3579	

Figure (20): Out of sample MSE



Our results indicate that MFDFM and pooled models (mean of the different models as individual unrestricted MIDAS, Multivariate MIDAS and MDFM) had the lowest MSE, showing that these models have the potential for nowcasting and forecasting with either higher volatility in GDP (particularly during COVID19) and with more limited data availability.

Forecasting RGDP for further quarters of 2021 using MDFM, 3-PRF, and pooled forecasts is reported in Table 8.

Table (8): Nowcasting third quarter of RGDP growth in year 2021 based on the releases data

	Estimations T3-2021	
	MFDFM	**UMIDAS (AR)**
RGDP at constant prices (MDT) (base2015)	22 160	22 240
Quarterly growth (en %)	-0,48	-0,12
Growth (Q-Q) in year(en %)	**-1,4**	**-1**

9. Synthesis

In this study, we set up a large dataset containing 118 potentially relevant indicators to monitor the evolution of the Tunisian economy. In addition to real sector variables, we used relevant indicators to capture the pandemic effect, such as electricity consumption by sectors, stock market index detailed by sectors, and international economic surveys. Financial variables seem to improve the performance of forecasting models at all forecast horizons. However, due to the short history of these variables, we could not thoroughly support this conclusion.

We considered different approaches to nowcasting and forecasting short-term GDP:

The **first approach** uses a mixed dynamic factor model, which can also be used to construct a monthly index of Tunisian economic activity. These models have been used to extract an unobserved state of the economy and create a new coincident indicator and exploit more information and obtain more precise forecasts. We applied the EM algorithm to estimate monthly GDP values. Furthermore, if the GDP release is not yet available, but some monthly indicators are, the EM algorithm can be employed to obtain an estimate of the corresponding quarterly GDP.

-The **second approach** performs estimations based on UMIDAS-AR equations by GDP sector. For each sector, we specified two or three models (due to an unbalanced number of observations) using the ridge method: a model containing variables with a **long history** (IPI, exports, electricity consumption, etc.) and models using financial variables with a **short history** (TUNINDEX by sectors).

Then, we combined the forecasts to form a final GDP forecast by weighting the different estimations of the components of GDP at constant prices.

The main findings of this study are that mixed-frequency models improve forecast models in the disaggregated GDP (vs. aggregated GDP in a previous study).

For all approaches, the dynamics align well with Tunisian RGDP growth, which declined sharply in 2011 (revolution: economy recession) and 2020 (COVID-19). However, note that the magnitudes of the predictions in these crises are significantly less than the realised values.

In addition, MFDFM and unrestricted MIDAS performed well in terms of root mean squared errors concerning a benchmark model VAR (2). The forecast errors derived from the disaggregated approach during the recent COVID period are smaller than those derived from classical models such as VAR (2) and mean to-date.

10. Conclusion

Obtaining reliable nowcasts and short-term forecasts of economic conditions is very relevant for policymaking, especially during crises when the economy witnesses large fluctuations.

Overall, we conclude that the mixed-frequency nowcasting model is particularly useful for volatile times. Additionally, it is better to employ several UMIDAS-ARs by each component of GDP at constant prices and to pool the results rather than relying on aggregated GDP, specifically in crisis periods.

Using the mixed-frequency dynamic factor model also enables estimating the GDP growth rate at a monthly level based on the movement of many available monthly indicators. This can be useful to economic policymakers as a representation of real GDP dynamics and real activity at a monthly frequency.

Nevertheless, the models used could be further investigated to improve nowcasting performance. The assumption of constant parameters could be adjusted in further research by introducing time-varying parameter models to track structural changes in the Tunisian economy. Another improvement of the models might be to use higher-frequency data, such as weekly or even daily data, and to include other variables such as fiscal indicators.

Survey indicators should be added to the forecasts because they are released with a short publication lag and are more informative. In this context, the Central Bank of Tunisia intends to implement a monthly economic survey to obtain qualitative information that can be used to monitor the present economic situation and forecast short-term development.

References

[1] Banbura.M and Modugno.M (2014): "Maximum likelihood estimation on factor models on datasets with arbitrary pattern of missing data". Journal of applied econometrics, 29 (1), 133-160".

[2] Clements, M. and Galvão, A. B. (2008)"Macroeconomic forecasting with mixed-frequency data: Forecasting output growth in the United States". Journal of Business and Economic Statistics, 26.

[3] Foroni, C., Marcellino, M., and Schumacher, C. (2015).: "U-MIDAS regressions with unrestricted lag polynomials. Journal of the Royal Statistical Society - Series A, 178(1):57–82.

[4] Foroni, C., Marcellino, M., and Stevanovic, D. (2019):" Mixed-frequency models with ma components". Journal of Applied Econometrics, 34(5):688–706.

[5] Ghysels, E., Santa-Clara, P., and Valkanov, R. (2004): "The MIDAS touch: Mixed data sampling regression". Technical report, Department of Economics, University of North Carolina, Chapel-Hill, North Carolina.

[6] Giannone, L. Reichlin, D. Small (2008), "Nowcasting: the real time information content of macroeconomic data". Journal of Monetary Economics, volume 55, Issue 4, May (2008) pages 665-676.

[7] Hepenstrick, C. and M. Marcellino (2016): "Forecasting with Large Unbalanced Datasets: The Mixed Frequency Three-Pass Regression Filter", SNB WP.

[8] Marcellino, M., M. Porqueddu, and F. Venditti (2012): "Short-term GDP forecasting with a mixed frequency dynamic factor model with stochastic volatility", mimeo.

[9] Marcellino, M., and C. Schumacher (2010): "Factor-MIDAS for now- and forecasting with ragged-edge data: A model comparison for German GDP", Oxford Bulletin of Economics and Statistics, 72, 518-550.

[10] Mariano, R., and Y. Murasawa (2003): "A new coincident index of business cycles based on monthly and quarterly series", Journal of Applied Econometrics, 18(4), 427-443

[11] Mariano, R., and Y. Murasawa (2010): "A coincident index, common factors, and monthly real GDP", Oxford Bulletin of Economics and Statistics, 72(1), 27-46.

[12] Stock, J., and M.Watson (1989): "New indexes of coincident and leading indicators," NBER Macroeconomics Annual, 4, 351-394.

[13] Stock, J., and M.Watson (2002): "Forecasting using principal components from a large number of predictors", Journal of the American Statistical Association, 97(460), 1167-1179

Appendix 1 : Ragged-edge database

Table 1: Monthly indicators for real sector

data_span =

50×4 table

Series Name	Start Date	End Date	Observations
{'IPI_Manufacturing_industries' }	31-Jan-2000	31-Mar-2021	255
{'IPI_Food_industries' }	31-Jan-2000	31-Mar-2021	255
{'IPI_Building_materials_ceramics_and_glass' }	31-Jan-2000	31-Mar-2021	255
{'IPI_Mechanical_and_electrical_industry' }	31-Jan-2000	31-Mar-2021	255
{'IPI_Chemical_industries' }	31-Jan-2000	31-Mar-2021	255
{'IPI_Textile_clothing_and_leathers' }	31-Jan-2000	31-Mar-2021	255
{'IPI_Miscellaneous_manufacturing_industries' }	31-Jan-2000	31-Mar-2021	255
{'IPI_Mines' }	31-Jan-2000	31-Mar-2021	255
{'IPI_Energy' }	31-Jan-2000	31-Mar-2021	255
{'IPI_Crude_Oil_Natural_Gas_and_Bituminous_Products' }	31-Jan-2000	31-Mar-2021	255
{'IPI_Refined_petroleum_products_and_coking' }	31-Jan-2000	31-Mar-2021	255
{'IPI_Electricity_and_gas' }	31-Jan-2000	31-Mar-2021	255
{'IPI_Distributed_water' }	31-Jan-2000	31-Mar-2021	255
{'IPI_2010__100' }	31-Jan-2000	31-Mar-2021	255
{'Electricity_consumption_EXTRACTIVE_INDUSTRY' }	31-Jan-2004	30-Jun-2021	210
{'Electricity_consumption_METALLURGICAL_INDUSTRY' }	31-Jan-2004	30-Jun-2021	210
{'Electricity_consumption_CHEMICAL_INDUSTRY' }	31-Jan-2004	30-Jun-2021	210
{'Electricity_consumption_IMCCV' }	31-Jan-2004	30-Jun-2021	210
{'Electricity_consumption_PAPER_INDUSTRY' }	31-Jan-2004	30-Jun-2021	210
{'Electricity_consumption_TEXTILE_AND_CLOTHING_INDUS'}	31-Jan-2004	30-Jun-2021	210
{'Electricity_consumption_FOOD_INDUSTRY' }	31-Jan-2004	30-Jun-2021	210
{'Electricity_consumption_VARIOUS_INDUSTRY' }	31-Jan-2004	30-Jun-2021	210
{'Electricity_consumption_AGRICULTURE' }	31-Jan-2004	30-Jun-2021	210
{'Electricity_consumption_PUMPING' }	31-Jan-2004	30-Jun-2021	210
{'Electricity_consumption_TOURISM' }	31-Jan-2004	30-Jun-2021	210
{'Electricity_consumption_TRANSPORT_AND_TELECOMMUNIC'}	31-Jan-2004	30-Jun-2021	210
{'Electricity_consumption_SERVICES_AND_OTHERS' }	31-Jan-2004	30-Jun-2021	210
{'Electricity_consumption_TOTAL' }	31-Jan-2004	30-Jun-2021	210
{'Electricity_consumption_Industries' }	31-Jan-2004	30-Jun-2021	210
{'Electricity_consumption_Manufacturing_industries' }	31-Jan-2004	30-Jun-2021	210
{'Electricity_consumption_Services' }	31-Jan-2004	30-Jun-2021	210
{'Export_Agriculture_and_Ind_Agro_Alim_at_constant_p'}	31-Jan-2000	31-Mar-2021	255

{'Export_Energy_and_Lubricants_at_constant_prices' }	31-Jan-2000	31-Mar-2021	255
{'Export_Mines_Phosphates_and_Derivatives_at_constan'}	31-Jan-2000	31-Mar-2021	255
{'Export_Textiles_Clothing_and_Leather_at_constant_p'}	31-Jan-2000	31-Mar-2021	255
{'Export_Mechanical_and_Electrical_Industries_at_con'}	31-Jan-2000	31-Mar-2021	255
{'Export_Other_Manufacturing_Industries_at_constant_'}	31-Jan-2000	31-Mar-2021	255
{'Export_All_Products_at_constant_prices' }	31-Jan-2000	31-Mar-2021	255
{'Imports_All_Products_at_constant_prices' }	31-Jan-2000	31-Mar-2021	255
{'Imports_of_raw_materials_and_semifinished_product_'}	31-Jan-2000	31-Mar-2021	255
{'Crude_oil_production' }	31-Jan-2000	31-Mar-2021	255
{'Natural_gas_production' }	31-Jan-2000	31-Mar-2021	255
{'Phosphate_production' }	31-Jan-2000	31-Mar-2021	255
{'Local_sales_of_cement' }	31-Jan-2000	31-Mar-2021	255
{'Entries_of_nonresidents' }	31-Jan-2000	30-Jun-2021	258
{'Tourist_nights' }	31-Jan-2000	30-Jun-2021	258
{'air_passenger_traffic' }	31-Jan-2000	31-Mar-2021	255
{'Job_demand_in_thousands' }	31-Jan-2000	31-Oct-2020	250
{'Job_offer_in_thousands' }	31-Jan-2000	31-Oct-2020	250
{'recruitement_in_thousands' }	31-Jan-2000	31-Oct-2020	250

<u>Table 2: Monthly indicators for monetary sector</u>

data_span =

21×4 table

Series Name	Start Date	End Date	Observations
{'CPI' }	31-Jan-2000	31-Mar-2021	255
{'Inflation' }	31-Jan-2000	31-Mar-2021	255
{'Industrial_selling_price_index'}	31-Jan-2000	31-Mar-2021	255
{'Money_market_rate' }	31-Jan-2000	31-Mar-2021	255
{'Policy_rate' }	31-Jan-2000	31-Mar-2021	255
{'M4' }	31-Jan-2000	31-Mar-2021	255
{'M3' }	31-Jan-2000	31-Mar-2021	255
{'M2' }	31-Jan-2000	31-Mar-2021	255
{'M1' }	31-Jan-2000	31-Mar-2021	255
{'Credit_to_the_economy' }	31-Jan-2000	31-Mar-2021	255
{'Tuindex' }	31-Jan-2000	31-Mar-2021	255
{'EURTND' }	31-Jan-2000	31-Mar-2021	255
{'USDTND' }	31-Jan-2000	31-Mar-2021	255
{'relative_price_index_EURTND' }	31-Jan-2000	31-Mar-2021	255
{'relative_price_index_USDTND' }	31-Jan-2004	30-Jun-2021	210
{'EURTND_real' }	31-Jan-2004	30-Jun-2021	210
{'USDTND_real' }	31-Jan-2004	30-Jun-2021	210
{'REER' }	31-Jan-2004	30-Jun-2021	210
{'NEER' }	31-Jan-2004	30-Jun-2021	210
{'credit_card_payments' }	31-Jan-2004	30-Jun-2021	210
{'Net_foreign_currency_assets' }	31-Jan-2004	30-Jun-2021	210

Table 3: Monthly indicators for financial sector

data_span =

14×4 table

Series Name	Start Date	End Date	Observations
{'CONSUMER_GOODS' }	30-Jun-2008	31-May-2021	156
{'Industries_automobile_et_quipement' }	30-Jun-2008	31-Dec-2020	151
{'Industries_agroalimentaires_et_boissons'}	31-Aug-2011	31-Dec-2020	113
{'Produits_mnagers_et_soins_personnels' }	30-Apr-2014	31-Dec-2020	81
{'CONSUMER_SERVICES' }	30-Jun-2008	31-May-2021	156
{'Distribution' }	31-Jan-2010	31-Dec-2020	132
{'Industries' }	30-Jun-2008	31-May-2021	156
{'Btiments_et_matriaux_de_construction' }	30-Jun-2008	31-Dec-2020	151
{'Mtriaux_de_base' }	31-Jul-2008	31-Dec-2020	150
{'FINCIALS_INSTITUTIONS' }	30-Jun-2008	31-May-2021	156
{'services_financiers' }	30-Jun-2008	31-Dec-2020	151
{'Banques' }	30-Jun-2008	31-Dec-2020	151
{'Assurance' }	31-Jan-2011	31-Dec-2020	120
{'TUNINDEX' }	30-Jun-2008	31-May-2021	156

Table 4: Disaggregated RGDP

33×4 table

Series Name	Start Date	End Date	Observations
{'GDP_Agriculture_and_Fishing' }	30-Jun-2000	31-Mar-2021	84
{'Industries' }	30-Jun-2000	31-Mar-2021	84
{'GDP_Manufacturing' }	30-Jun-2000	31-Mar-2021	84
{'Agriculture_and_food_processing' }	30-Jun-2000	31-Mar-2021	84
{'Textiles_leather_and_clothing' }	30-Jun-2000	31-Mar-2021	84
{'Miscellaneous_industries' }	30-Jun-2000	31-Mar-2021	84
{'oil_refining' }	30-Jun-2000	31-Mar-2021	84
{'Chemical_Industries' }	30-Jun-2000	31-Mar-2021	84
{'Construction_materials_Ceramics_and_glass'}	30-Jun-2000	31-Mar-2021	84
{'Mechanical_and_electrical_industries' }	30-Jun-2000	31-Mar-2021	84
{'GDP_non_manufacturing' }	30-Jun-2000	31-Mar-2021	84
{'Gas_and_petroleum_product' }	30-Jun-2000	31-Mar-2021	84
{'Mining' }	30-Jun-2000	31-Mar-2021	84
{'Electricity_and_gaz' }	30-Jun-2000	31-Mar-2021	84
{'Water' }	30-Jun-2000	31-Mar-2021	84
{'Building_and_civil_engineering' }	30-Jun-2000	31-Mar-2021	84
{'Services' }	30-Jun-2000	31-Mar-2021	84
{'Maintenance_and_repair' }	30-Jun-2000	31-Mar-2021	84
{'Trade' }	30-Jun-2000	31-Mar-2021	84
{'Hotels_cafe_restaurant' }	30-Jun-2000	31-Mar-2021	84
{'Transports' }	30-Jun-2000	31-Mar-2021	84
{'Telecommunications' }	30-Jun-2000	31-Mar-2021	84
{'Financial_services' }	30-Jun-2000	31-Mar-2021	84
{'Other_services' }	30-Jun-2000	31-Mar-2021	84
{'Imputed_financial_services' }	30-Jun-2000	31-Mar-2021	84
{'Market_services_activities' }	30-Jun-2000	31-Mar-2021	84
{'Non_market_services_activities' }	30-Jun-2000	31-Mar-2021	84
{'Public_administration_services' }	30-Jun-2000	31-Mar-2021	84
{'Associative_organisation_services' }	30-Jun-2000	31-Mar-2021	84
{'Domestic_services' }	30-Jun-2000	31-Mar-2021	84
{'Total_ADDED_VALUES' }	30-Jun-2000	31-Mar-2021	84
{'Indirect_taxes_net_of_subsidies' }	30-Jun-2000	31-Mar-2021	84
{'GDP_constant_price' }	30-Jun-2000	31-Mar-2021	84

Appendix 2: Covariance matrix between high and low frequencies variables

***Table 5: Covariance matrix between long history high-frequency indicators and Real Manufacturing GDP growth* For manufacturing GDP**

High Frequency_indicators (xit)	***Corr (xit, Manufacturing_RGDPt)***
(1) IPI_Manufacturing_industries	0,19385
(2) IPI_Food_industries	0.16579
(3) IPI_Building_materials_ceramics_and_glass	0.0021226
(4) IPI_Mechanical_and_electrical_industry	-0.081974
(5) IPI_Chemical_industries	0.20914
(6) IPI_Textile_clothing_and_leathers	0.037611
(7) IPI_Miscellaneous_manufacturing_industries	0.063001
(8) IPI_Mines	-0.0077592
(9) IPI_Energy	0.089197
(10) IPI_Crude_Oil_Natural_Gas_and_Bituminous_Products	0.021175
(11) IPI_Refined_petroleum_products_and_coking	0.16268
(12) IPI_Electricity_and_gas	0.0026341
(13) IPI_Distributed_water	-0.00046465
(14) IPI_2010__100	0.29356
(15) Electricity_consumption_EXTRACTIVE_INDUSTRY	0.21528
(16) Electricity_consumption_METALLURGICAL_INDUSTRY	0.03237
(17) Electricity_consumption_CHEMICAL_INDUSTRY	-0.028973
(18) Electricity_consumption_IMCCV	0.14052
(19) Electricity_consumption_PAPER_INDUSTRY	-0.021778
(20) Electricity_consumption_TEXTILE_AND_CLOTHING_INDUS	0.077413
(21) Electricity_consumption_FOOD_INDUSTRY	0.14974
(22) Electricity_consumption_VARIOUS_INDUSTRY	0.12335
(23) Electricity_consumption_AGRICULTURE	-0.0091815
(24) Electricity_consumption_PUMPING	0.043891
(25) Electricity_consumption_TOURISM	0.21668
(26) Electricity_consumption_TRANSPORT_AND_TELECOMMUNIC	-0.011684
(27) Electricity_consumption_SERVICES_AND_OTHERS	0.038011
(28) Electricity_consumption_TOTAL	0.1468
(29) Electricity_consumption_Industries	0.12979
(30) Electricity_consumption_Manufacturing_industries	0.11777
(31) Electricity_consumption_Services	0.073564
(32) Export_Agriculture_and_Ind_Agro_Alim_at_constant_p	0.17232
(33) Export_Energy_and_Lubricants_at_constant_prices	-0.15603
(34) Export_Mines_Phosphates_and_Derivatives_at_constan	0.12992
(35) Export_Textiles_Clothing_and_Leather_at_constant_p	-0.078629
(36) Export_Mechanical_and_Electrical_Industries_at_con	-0.00023294
*(37) Export_Other_Manufacturing_Industries_at_constant_*	0.1036
(38) Export_All_Products_at_constant_prices	-0.042204
(39) Imports_All_Products_at_constant_prices	0.11841
*(40) Imports_of_raw_materials_and_semifinished_product_*	0.098182

(41) Crude_oil_production	0.030896
(42) Natural_gas_production	-0.09574
(43) Phosphate_production	0.17121
(44) Local_sales_of_cement	-0.19415
(45) Entries_of_nonresidents	0.20776
(46) Tourist_nights	0.23961
(47) air_passenger_traffic	0.057817
(48) Job_demand_in_thousands	-0.15241
(49) Job_offer_in_thousands	0.25097
(50) recruitement_in_thousands	0.19323

Table 6: Covariance matrix between short history high-frequency indicators and Real Manufacturing GDP growth

High-Frequency Financial indicators (xit)	***Corr(xit, Manufacturing_RGDPt)***
(1) CONSUMER_GOODS	0.19561
(2) Industries_automobile_et_quipement	*0.26188*
(3) Industries_agroalimentaires_et_boissons	*-0.069281*
(4) Produits_mnagers_et_soins_personnels	*0.12397*
(5) CONSUMER_SERVICES	*0.12688*
(6) Distribution	-0.08064
(7) Industries	0.11123
(8) Btiments_et_matriaux_de_construction	0.11576
(9) Mtriaux_de_base	0.20966
(10) FINCIALS_INSTITUTIONS	0.085281
(11) services_financiers	0.12612
(12) Banques	0.16004

Table 7: Factor loadings

Table 1: Factor Loadings

Series	Var1	Var2
{'IPI_Manufacturing_industries' }	8.566	0
{'IPI_2010__100' }	10.04	0
{'Export_Agriculture_and_Ind_Agro_Alim_at_constant_p'}	2.9387	0
{'CONSUMER_GOODS' }	0	7.3982
{'Industries_automobile_et_quipement' }	0	4.8105
{'Mtriaux_de_base' }	0	4.216
{'ManufacturingConfidenceIndicator' }	1.4833	0
{'GDP_Manufacturing' }	9.0402	0
{'IPI_Energy' }	5.0074	0
{'IPI_Crude_Oil_Natural_Gas_and_Bituminous_Products' }	3.0373	0
{'Electricity_consumption_EXTRACTIVE_INDUSTRY' }	3.5652	0
{'Electricity_consumption_PUMPING' }	1.6304	0
{'Electricity_consumption_Services' }	3.3846	0
{'Export_Other_Manufacturing_Industries_at_constant_'}	3.5079	0
{'Natural_gas_production' }	2.6687	0
{'Produits_mnagers_et_soins_personnels' }	0	6.3639
{'Industries' }	0	5.2959
{'Btiments_et_matriaux_de_construction' }	0	4.2817
{'GDP_non_manufacturing' }	7.268	0
{'IPI_Building_materials_ceramics_and_glass' }	4.5484	0
{'Electricity_consumption_TOURISM' }	3.0377	0
{'Tourist_nights' }	1.8849	0
{'CONSUMER_SERVICES' }	0	6.5522
{'Distribution' }	0	6.0786
{'Banques' }	0	8.3118
{'Assurance' }	0	3.6303
{'Tuniindex' }	0	9.6812
{'IPIManufEuroArea' }	0.65954	0
{'Services' }	5.3635	0
{'GDP_constant_price' }	9.1746	0

Blocks in DFM

For identification:

As with principal components, we have an identification issue:

$$y_t = H\,x_t + \varepsilon_t$$

$$x_t = B\,x_{t-1} + e_t$$

Is equivalent to the model:

$$y_t = H\,\theta^{-1}\,\theta x_t + \varepsilon_t$$

$$\theta x_t = \theta B \theta^{-1} \theta x_t + \theta \varepsilon_t$$

Orthogonal factors are possible solutions. The chart below decomposes observations in terms of common components (factors), autoregressive errors, and ID errors.