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. HEIDWP11-2021

Nowcasting in Tunisia using large datasets and mixed frequency models

Hager Ben Romdhane
Central Bank of Tunisia

June 2021

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

Nowcasting in Tunisia using large datasets and mixed frequency models

Hager Ben Romdhane[1]
Central Bank of Tunisia
(January 2021)

[1] I would like to express my sincere gratitude to my supervisor, **Prof. Luca Gambetti** for his guidance and support. I extend my thanks **to Prof Massimiliano Marcellino and Dr.Seton Leonard** for their valuable input, guidance and suggestions. I am grateful to my institution, the **Central Bank of Tunisia and to the BCC program,** financed by SECO, to give me this opportunity to participate in this interesting coaching program .Many thanks to **Prof Rached Bouaziz, Mr.Mourad Abdessalem and** to my colleagues **Mr. Mehdi Bellalouna** as lead of strategic project “leading and advanced indicators at the Central Bank of Tunisia” and **Ms.Aida Guellouz** for her guidance during the data processing. All errors are my own.

Abstract:

The object of this paper is to nowcast, forecast and track changes in Tunisian economic activity during normal and crisis times. The main target variable is quarterly real GDP (RGDP) and we have collected a large and varied set of monthly indicators as predictors. We use several mixed frequency models, such as unrestricted autoregressive MIDAS (UMIDAS-AR), three pass regression filter (3PRF) and mixed dynamic factor models (MDFM). We evaluate these models by comparing them with benchmarking low frequency models including vector autoregressive (VAR) and ARMA models. The dynamic factor and the 3PRF forecasts are more accurate in terms of mean squared errors (MSE) than other alternatives models both in-sample and out of sample in normal times, meaning before the COVID19 period. Forecast errors derived from low frequency models including crisis periods are larger than errors from mixed data sampling approaches including autoregressive terms due mainly to the failure of the low frequency models to capture these tail events. Fortunately, the reliability of nowcasts and forecasts increase when using the mixed frequency dynamic factor model based on information at both monthly and quarterly frequencies.

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

Author. Email: Hager.BenRomdhane@bct.gov.tn

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

Abbreviations

NLS: Non -linear Least Squared

ARMA: Autoregressive Moving Average

MDFM = Mixed Dynamic Factor models

MIDAS = Mixed Data Sampling

RGDP = Real Gross Domestic Product

UMIDAS-AR: Unrestricted Mixed Data Sampling Autoregressive term.

3PRF = Three Pass Regression Filter

VAR = Vector Autoregressive.

1. Introduction

Policy makers 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 at low frequencies.

Especially in times of crisis, nowcasting is important because timely forecasts of RGDP growth are useful summaries of recent news on the economy and commonly used as inputs to structural forecasting.

In the recent period, Covid-19 has raised the issue of nowcasting and short run forecasting due to heightened uncertainties. In fact, the actual pandemic crisis led to a sudden stop in economic activity all over the world. 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. The economy took a severe hit Tunisia as well with RGDP for Q2: 2020 declining by 21.3 percent year-on- year(y-o-y).

Recently, econometric models that consider the information in unbalanced datasets have been developed. Unbalanced of datasets arise due to two features: different sampling frequency and the "ragged-edge" issue as publication delays cause missing observations of some of the variables at the end of the sample.

The Tunisian National Statistical Institute (INS) releases an estimate of RGDP about 45 days after the end of the quarter. Furthermore, many leading and coincident indicators are available at a monthly or even high frequency such as financial and monetary variables which might help in monitoring the current state of the economy as well as 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 to work with a "frozen" final vintage dataset so the left and right hand side variables are sampled at the same frequency. However, this aggregation process destroys a lot of potentially useful information and can lead to misspecification.

Due to the issue of releases delays for national accounts, central banks rely on continuously flowing information from leading and coincident activity indicators to gauge the underlying state of the economy on a real-time basis.

Accordingly, various econometric approaches have been developed to address the case of ragged-edge data such as MIDAS regressions and factor models (Giannone, Reichlin and Small (2008)). Clements and Galvao (2008) introduce the use of MIDAS regressions in forecasting macroeconomic data. They look at whether a mixed-data sampling approach including an autoregressive term can improve forecasts of US real output growth. They conduct a real-time forecasting exercise that exploits monthly vintages of the indicators and the quarterly vintages of output growth, consistent with the time of the releases of the different data vintages. The authors find that the use of within-quarter information on monthly indicators can result in a marked reduction in RMSE compared with the more traditional quarterly-frequency VAR or AR distributed lag models.

In the same line, Marcellino, Foroni and Schumacher (2012) compare the performance of a MIDAS with functional distributed lags estimated with NLS to a U-MIDAS, the unrestricted version of MIDAS. In Monte Carlo experiments, they show that the U-MIDAS generally performs better than the MIDAS when mixing quarterly and monthly data. One the other hand, with larger differences in sampling frequencies, distributed lag-functions outperform unrestricted polynomials. In an empirical application for out of sample nowcasts of RGDP in the Euro Area and the US using monthly predictors, they find the U-MIDAS performs well.

Mariano and Murasawa (2003) propose a new coincident index of the business cycle that relies on both monthly and quarterly indicators. Also, Mariano and Murasawa (2010) apply a mixed frequency VAR method to construct a new coincident indicator that is, an estimate of monthly real GDP. What they find is that the coincident index based on the VAR model to the one obtained by a factor model track well quarterly real GDP, although they are quite volatile.

Furthermore, Marcellino and Schumacher (2010) propose to merge factor models with the MIDAS approach, which allows them to now- and forecast low frequency variables such as RGDP exploiting information in a large set of higher frequency indicators. They found that all Factor-MIDAS nowcasts can improve over quarterly factor forecast based on time aggregated data.

Recent applications using Mixed frequency factor models are, for example, ***Banbura and Modugno (2014)*** who discuss the maximum likelihood estimation of factor models on datasets with arbitrary pattern of missing data.

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

In our paper, we consider several models such as univariate UMIDAS-AR, multivariate UMIDAS-AR, and a mixed frequency factor models to nowcast the first quarter of 2021. Then, we forecast until the three quarters ahead of 2021, given the monthly information at the end of April 2021, before the first official release of Tunisian RGDP for 2021:Q1. At that point in time, we observe 2020:Q4 for RGDP, and the first three months of 2021 for some indicators we use. Since estimation of the multivariate dynamic factor model can be numerically complex, computational efficiency is achieved by maximum likelihood estimation following Watson and Engle (1983).

In Tunisia there are many severe periods characterized by a significant and deep decline in RGDP such as the revolution period in 2011 and the Covid-19 crisis. To address this issue, we remove outliers from series used to estimate model parameters for the whole sample period (2000M01-2021M04). In a second step for no and forecasting RGDP we add these outliers to generate predictions.

The mixed dynamic factor and pooled (mean of the different models as individual U-MIDAS, multivariate MIDAS and MDFM) nowcasts perform better than individual U-Midas and multivariate UMIDAS-AR in both in-sample and out of -sample in normal times, but during the crisis forecast errors are large and models do not capture the deep decrease of RGDP in **2020:Q2**.

The remainder of this paper is organized as follows. Section (2) contains a brief description of models we used. Section (3) explains the data selection and methodology used in this paper. Section (4) presents empirical results of mixed sampling models estimation both in sample and out of sample, the main findings and results and section (5) concludes.

2. Models

We describe theoretical foundation of the mixed frequency data sampling as: Unrestricted Mixed Frequency, Mixed Dynamic Factor and Mixed Frequency three-pass regression.

2.1 MODEL OF UNRESTRICTED MIXED FREQUENCY (UMIDAS)

Foroni, Marcellino and Shumacher (2015) study the performance of a variant of MIDAS which does not resort to functional distributed lag polynomials. They discuss how an unrestricted MIDAS (UMIDAS) regression can be derived in a general linear dynamic framework and under which conditions the parameters of the underlying high frequency model can be identified:

The U-MIDAS can be written as:

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

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

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

This model is estimated at low frequency, uses the high frequency regressors and can be re-estimated each month within the quarter. As the U-MIDAS is linear it could be estimated by OLS, where $t=1 \ldots T$ and m is months of the quarter.

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

$$\mathbf{\bar{y}_{T_M^X+m/\,T_M^X} = \bar{c}(L^k)y_{T_M^X} + \overline{\delta_1}(L)x_{1T_M^X} + \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots + \overline{\delta_N}(L)x_{NT_M^X}} \qquad \textbf{(2)}$$

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 $mtm - m\ or\ earlier, t = 1,2, \ldots\ldots. T_m^X$.

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

$$\boldsymbol{\bar{y}_{T_M^X+hm/\,T_M^X} = \bar{c}\left(L^{k}\right)y_{T_M^X} + \overline{\delta_1}(L)x_{1T_M^X} + \cdots \ldots\ldots\ldots\ldots\ldots\ldots\ldots.. + \overline{\delta_N}(L)x_{NT_M^X}} \qquad \textbf{(3)}$$

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

In addition, in the case of U-Midas, an autoregressive term can be included easily without any common factor restriction.

2.2 MODEL OF MIXED FREQUENCY (MF-VAR)

The MF-VAR model is represented in the form of the State-Space model. Equations of the state variables are given by a monthly frequency model and the measurement equations link the observable series to the unobservable variables,

$$x_t = \emptyset_1 x_{t-1} + \cdots \emptyset_p x_{t-p} + \emptyset_c + u_t, \qquad u_t \sim iid(0, \boldsymbol{\Sigma}) \qquad \textbf{(4)}$$

The vector of macroeconomic variables x_t is of the size (n×1) and consists of $x_t = [x'_{mt}, \quad x'_{qt}]$ where the x_{mt} dimension $(n_m \times 1)$, includes variables that are observed at monthly frequency, such as electricity consumption, industrial production index, etc., and the vector of x_{qt} $((n_q \times 1)$ includes unobservable variables that are published only on a quarterly frequency such as real RGDP. For all t_m, the monthly variable RGDP that is not observable (y^*_{tm}) and the monthly indicators (x_{tm}) follow a biased VAR (p) process:

$$\emptyset(L_m) \begin{pmatrix} y^*_{tm} - \mu^*_y \\ x_{tm} - \mu^*_x \end{pmatrix} = u_{tm} \qquad (5)$$

Where $u_{tm} \sim N(0, \Sigma)$

For p ≤ 4, we define:

$$s_{tm} = \begin{pmatrix} z_{tm} \\ \vdots \\ z_{tm-4} \end{pmatrix}, \qquad z_{tm} = \begin{pmatrix} y^*_{tm} - \mu^*_y \\ x_{tm} - \mu^*_x \end{pmatrix}$$

The MF- VAR state space representation is defined as follows:

$$s_{tm} = F s_{tm-1} + G\, V_{tm} \quad (6)$$

$$\begin{pmatrix} y_{tm} - \mu_y \\ x_{tm} - \mu_x \end{pmatrix} = H\, s_{tm} \qquad (7)$$

Equations (4) and (5) describe respectively the state and measurement equations. The signal or measurement equation links observable series to unobservable variables where the monthly variables are stacked in the state vector. In order to cope with high dimension of the parameter space, the MF-VAR is enriched with Minnesota priors and it is estimated using a Bayesian approach.

The MIDAS (Mixed data sampling) developed and popularized by Eric Ghysels (2016) permits the mixing of sampling frequencies. The starting point of Bayesian inference for the MF-VAR model is the joint distribution of the observable variables $Y_{1:T}$, the latent variables of States $Z_{0:T}$ and the parameters (ϕ, $\boldsymbol{\Sigma}$), conditional on a pre-sample $Y_{-p+1:0}$ to initialize the delays. The distribution of state variables and observable variables conditional on the parameters is determined from the state-space representation of MF-VAR. However, for the marginal distribution of the parameters (ϕ, $\boldsymbol{\Sigma}$), we use the Minnesota conjugate priors. The- priori go back to the work of Litterman (1980) and Doan, Litterman and Sims (1984).

For this purpose, we use the Minnesota version described in the chapter by Del Negro and Schorfheide (1998). The main idea is to center the distribution of ϕ to a value that implies random walk behavior for each component of x_t in equation (4). The implementation of Minnesota priori is done by mixing the artificial data into the estimate sample. The artificial observations allow us to generate the a priori plausible correlations between the VAR parameters. The distribution variance is controlled by a low dimensional hyper parameter vector. We produce draws from the posterior distributions of (ϕ, $\boldsymbol{\Sigma}$)\$Z_{0:T}$ and $Z_{0:T}/(\phi, \boldsymbol{\Sigma})$ by Gibbs sampling. Based on these draws, we can use the future trajectories of $\boldsymbol{y_t}$ **to characterize the predictive distribution associated with MF-VAR and to compute point and density predictions.**

2.3 Mixed-frequency small Factor models

Factor models have also been employed in the literature to handle data with different frequencies. These models have been utilized to extract an unobserved state of the economy and create a new coincident indicator as well as to exploit more information and obtain more precise forecasts. In fact, factor models have a long tradition in econometrics. Watson, and Engle (1983) introduced a simple algorithm for estimating dynamic factor models (DFMs) by maximum likelihood. By modeling the driving process behind (multivariate) observed data as latent (unobserved), DFMs can incorporate missing observations without falsified imputed data and allow the modeling of noisy observations due to measurement error.

Mariano and Murasawa (2003) present a DFM 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. Giannone et al, (2008) pioneered applications of DFM for nowcasting with a specific emphasis on using the real time data flow to update predictions as new information becomes available. In the estimation, Mariano and Murrasawa (2003) apply a Maximum-likelihood factor analysis to a mixed frequency series of quarterly RGDP and monthly business cycle indicators to construct an index that is related to a monthly real RGDP.

2.3.1. The Kalman Filter and Smoother

Dynamic factor models consist fundamentally of two equations. The measurement equation is defined as:

$$y_t = Hx_t + \varepsilon_t \quad (8)$$

And the transition equation

$$z_t = Az_{t-1} + e_t \quad (9)$$

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, n_t predetermined exogenous variables.

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

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

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

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

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

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

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

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

The matrix $\widetilde{H}$ in the above incorporates a helper matrix J to extract, in the simplest example, contemporaneous factors from z_t. That is $J=[I_m \quad 0 \quad 0 \quad ...]$ 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 $K_t=C_tS_t^{-1}$ 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.

2.3.2 State Space in Mixed Frequency Models

Because state space models are so apt at handling 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.

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

$$e^{y_t^q}=e^{y_t^m}+e^{y_{t-1}^m}+e^{y_{t-2}^m} \quad (10)$$

The difficulties lie in the fact that equation (8) is linear in the log variables while equation (10) is not. To overcome this issue simply take a linear approximation of (10) yielding:

$$y_t^q=\frac{1}{3}(y_t^m+y_{t-1}^m+y_{t-2}^m) \quad (11)$$

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

$$y_t^q=\frac{1}{3}Hx_t+\frac{1}{3}Hx_{t-1}+\frac{1}{3}Hx_{t-2}+\varepsilon_t \quad (12)$$

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

To put the model into log differences we begin with equation (11) and note that we observe $\Delta y_t^q$is:

$$y_t^q - y_{t-3}^q = \frac{1}{3}(y_t^m - y_{t-3}^m) + \frac{1}{3}(y_t^m - y_{t-4}^m) + \frac{1}{3}(y_{t-2}^m - y_{t-5}^m)$$
$$= \frac{1}{3}\Delta y_t^w + \frac{2}{3}\Delta y_{t-1}^w + \Delta y_{t-2}^w + \frac{2}{3}\Delta y_{t-3}^w + \frac{1}{3}\Delta y_{t-4}^w \quad (13)$$

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

2.4 Mixed Frequency 3-Pass Regression Filter

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

$$y_{t+h} = \beta_0 + \beta' F_t + \eta_{t+1} \qquad (14)$$
$$Z_t = \lambda_0 + \boldsymbol{\Lambda F_t} + \omega_t \qquad (\mathbf{15})$$
$$x_t = \phi_0 + \boldsymbol{\phi F_t} + \varepsilon_t \qquad (\mathbf{16})$$

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 $\beta = (\beta_f', 0')$, so that y only depends on f; Z_t is a small set of L proxies that are driven by the same underlying forces as y, so that $\boldsymbol{\Lambda} = (\boldsymbol{\Lambda_f}, \mathbf{0})$ **and** $\boldsymbol{\Lambda_f}$ **is nonsingular,** x_t is a large set of N variables driven by both f and g and *t=1,…....T.*

One can estimate the model by three-step algorithm:

0) Aggregate monthly dataset to quarterly frequency,

1) For each variable in the quarterly dataset, x_i, run a (time series) regression of x_i on the proxy z: $x_t^{(i)} = \boldsymbol{\phi}_0^{(i)} + z'\boldsymbol{\phi_i} + \boldsymbol{\varepsilon}_t^{(i)}$ **(17)**

2) With the OLS estimates $\widehat{\boldsymbol{\phi_i}}$ obtained in the previous step, run a cross section regression of $x_t^{(i)}$ on $\widehat{\boldsymbol{\phi_i}}$ **for each month in the monthly dataset:**

$$x_t^{(i)} = \boldsymbol{\phi}_0^{(t)} + \widehat{\boldsymbol{\phi_i}}'^{\boldsymbol{F_t}} + \boldsymbol{\varepsilon}_t^{(i)} \qquad (\mathbf{18})$$

3) Use mixed frequency techniques with OLS estimates $\widehat{F_t}$ to forecast y_{t+h}. Here we use a time series regression of $\boldsymbol{y_{t+h}}$ **on** $\widehat{F_t}$:

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

3. Data selection

We use a dataset collected on 28 January 2021, with values dating back to January 2001. The dataset consists of 23 monthly variables spanning hard indicators related to: (1) ***real Sector*** such as electricity consumption total, local sales of cement, personal savings; (2) **external *Sector*** such as imports and exports, real effective exchange Rate; (3) ***service sector*** as air transports and tourists nights; (4) ***natural resources production*** as phosphate production, crude oil production; (5) ***financial and monetary sector*** as TUNINDEX (stock market), credits to economy, credit card payments, financial services, aggregate money M3, central bank balance sheet, net foreign assets; (6) ***international sector*** as industrial production manufacturing index of Euro zone, energy prices and manufacturing confidence indicator and (7) **employment** as job offer and job demand.

3.1 Data processing and seasonal adjustment

Prior to modelling, we need to:

- Ensure data is stationary by differencing/log differencing and low frequency trends.
- Possibly standardize data.
- Seasonally adjusted data (adding national calendar dates).

All these transformations of the data are described in table (1) below together with the publication delays.

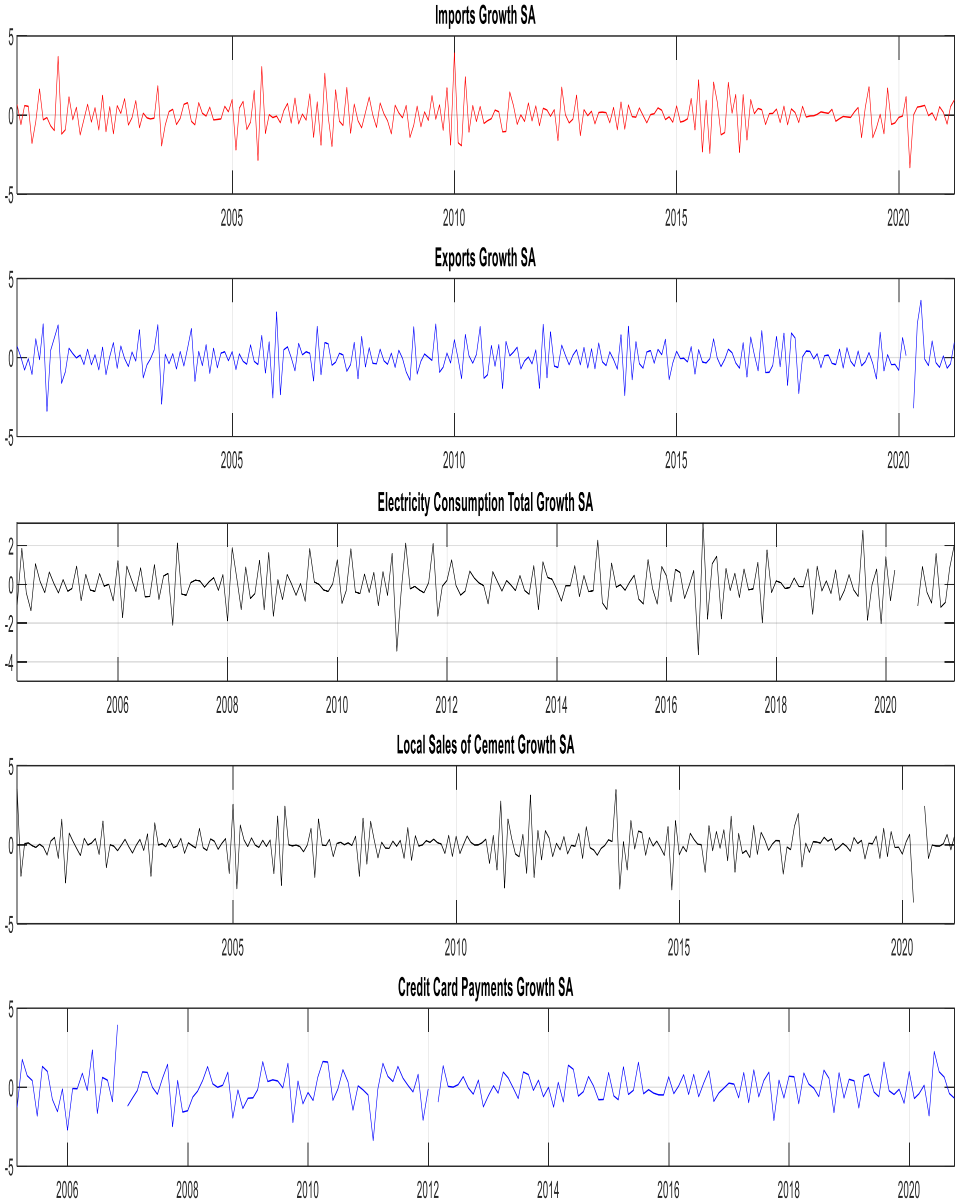
Table 1: Data of high frequency (monthly)

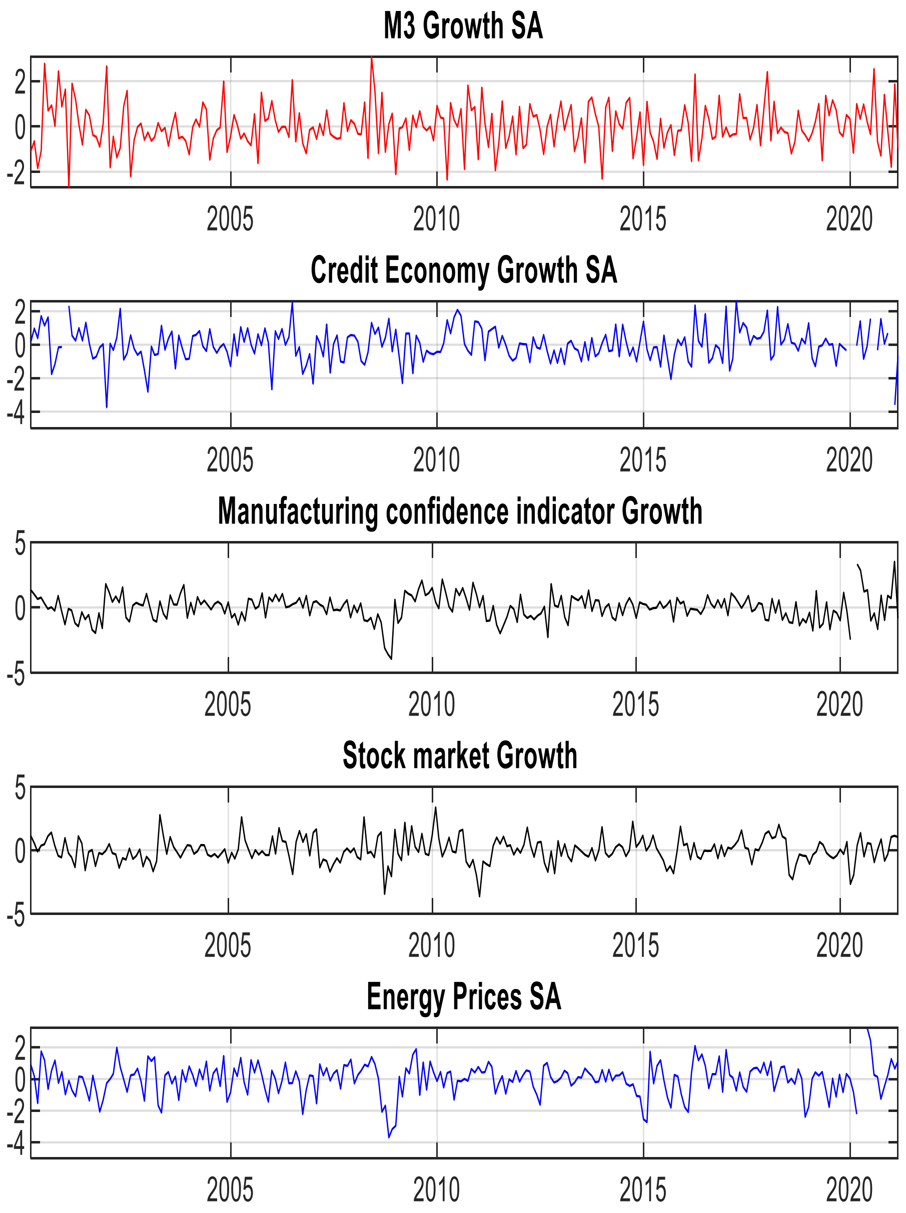
Series	Start Day	End Date	Delays	Observations
'Industrial Production Index	*31-janv-2000*	31-janv-2021	*6 weeks(publication has became irregular)*	253
'Imports'	*31-Jan-2000*	31-Mar-2021	*1 week delay*	255
'Exports'	*31-Jan-2000*	31-Mar-2021	1 week delay	255
'ElectricityConsumptionTotal'	*31-Jan-2004*	28-Feb-2021	*less than one week*	206
'CrudeOilProduction'	31-Jan-2000	28-Feb-2021	*90 days delay*	254
'NaturalGasProduction'	31-Jan-2000	28-Feb-2021	*90 days delay*	254
'PhosphateProduction'	*31-Jan-2000*	*30-Nov-2020*	*60 days delay*	251
'EntriesNonResidents'	31-Jan-2000	28-Feb-2021	15 days delays	254
'ManufacturingConfidenceIndicator'	31-Jan-2000	28-Feb-2021	*0 delay*	254
'Tuniindex'	31-Jan-2000	31-Mar-2021	*0 delay*	255
'M3'	31-Jan-2000	28-Feb-2021	4 weeks delays	254
'Local Sales Of Cement'	31-Janv-2001	28-Feb-2021	20 days delays	242
'CentralBankBalanceSheet'	31-Jan-2001	28-Feb-2021	4 weeks delays	242
'CreditEconomy'	31-Jan-2000	28-Feb-2021	4 weeks delays	254

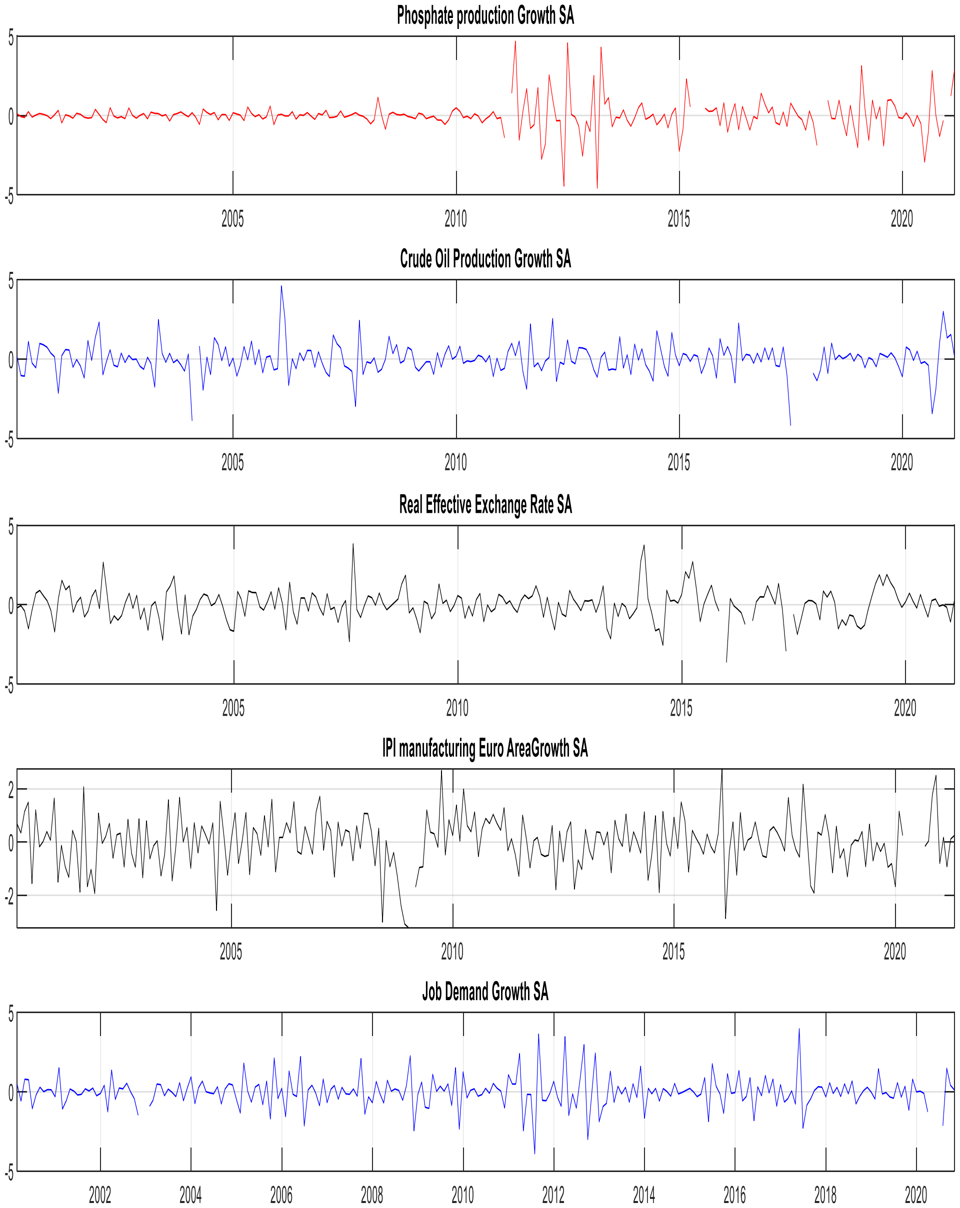
'NetForeignAssets'	31-Jan-2000	28-Feb-2021	4 weeks delays	254
'TermsOfTrade'	31-Jan-2000	28-Feb-2021	4 weeks delays	254
'EnergyPrices'	31-Jan-2000	28-Feb-2021	0 days delays	254
'REER'	31-Jan-2000	31-Jan-2021	0 days delays	253
'IPI_ManufEuroArea'	31-Jan-2000	31-Jan-2021	0 days delays	253
'JobDemand'	31-Jan-2000	31-Oct-2020	60 days delays	250
'CreditCardPayments'	31-Jan-2005	30-Sep-2020	4 weeks delays	189
'JobOffer'	31-Jan-2000	31-Dec-2020	60 days delays	252
'servicefinanciers'	31-May-2008	31-Dec-2020	0 delays	152
'RGDP'	31-Mar-2000	31-Dec-2020	45 days delays	84

We include graphs of indicators below. All indicators are seasonal adjusted and log differenced.

Graph 1: Monthly indicators evolution (Differenced log)

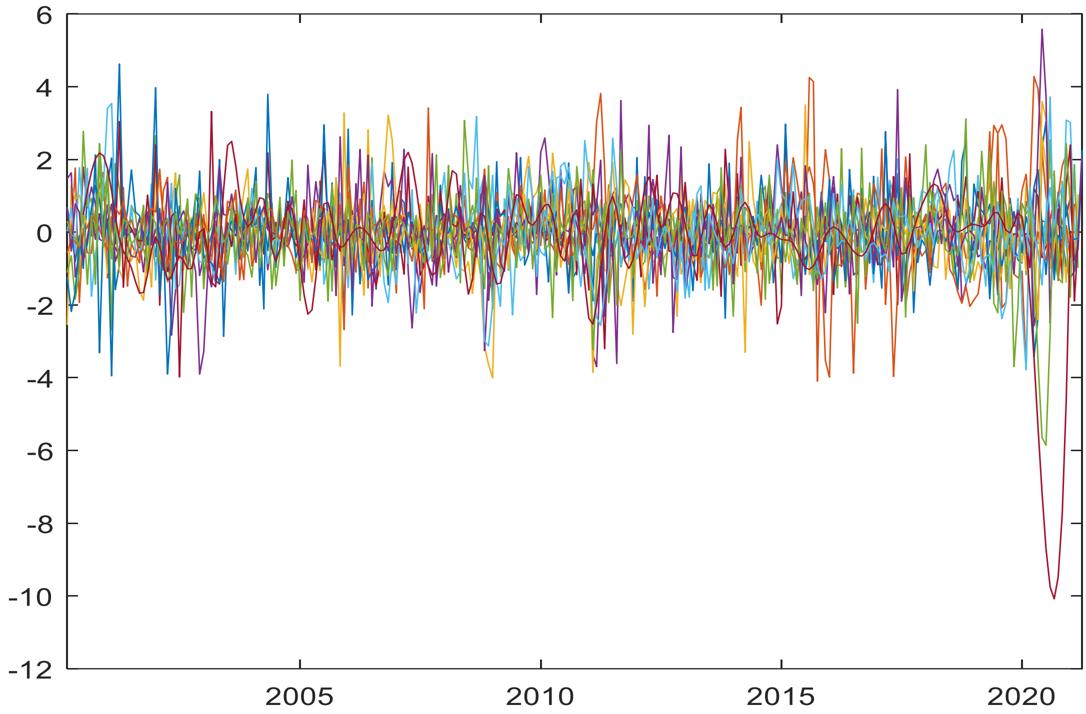



M3 Growth SA
2
0
-2
2005
2010
2015
2020
Credit Economy Growth SA
2
0
-2
-4
2005
2010
2015
2020
Manufacturing confidence indicator Growth
5
0
-5
2005
2010
2015
2020
Stock market Growth
5
0
-5
2005
2010
2015
2020
Energy Prices SA
2
0
-2
-4
2005
2010
2015
2020


Phosphate production Growth SA
5
0
-5
2005
2010
2015
2020
Crude Oil Production Growth SA
5
0
-5
2005
2010
2015
2020
Real Effective Exchange Rate SA
5
0
-5
2005
2010
2015
2020
IPI manufacturing Euro AreaGrowth SA
2
0
-2
2005
2010
2015
2020
Job Demand Growth SA
5
0
-5
2002
2004
2006
2008
2010
2012
2014
2016
2018
2020

Finally, we can look at all the data to see if it looks reasonable after processing as we can see in the figure (2) as below.

Graph 2: Monthly Indicators and Real GDP growth (Seasonal Adjusted)



An important question for our analysis is which series are correlated with RGDP? To answer this question, we will need to aggregate monthly data to quarterly.

The selection of monthly predictors is based on the calculated correlation matrix between standardized monthly indicators and real GDP as shown in table (2). As well as looking at the correlations in the ***Corr(xit, RGDPt)***, we can look at how well data are correlated with an AR(1) residual of GDP. Because an AR (1) model cannot explain RGDP, we will also look at an ARMA model shown in column ***Corr (xit, uit).*** We have highlighted the correlations which are greater than 0.10.

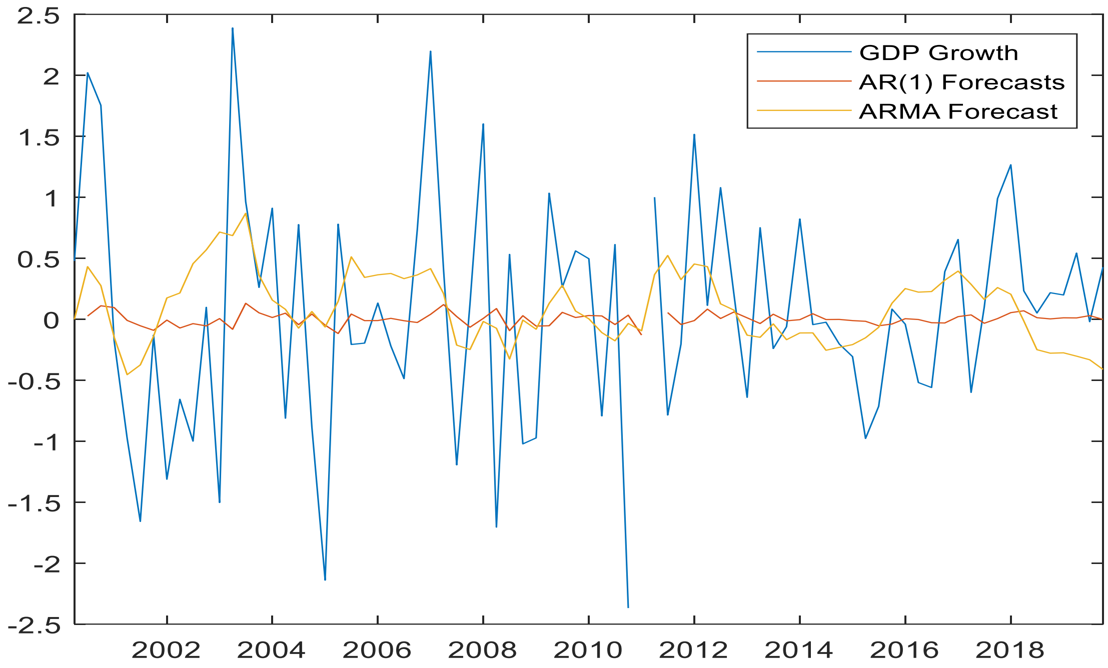
Table (2): Covariance matrix between high frequency indicators and Real GDP growth

High Frequency_indicators (xit)	*Corr(xit, RGDPt)*	*Corr(xit,uit)*
'Industrial Production Index'	0,15786	0,21234587
'Imports'	0.1043156	0, 216006206750171
'Exports'	0.15453287	**0,317356189083783**
'ElectricityConsumptionTotal'	0,26201999	**0,111767749770406**
'CrudeOilProduction'	0,11071908	**0,171123317186058**
'NaturalGasProduction'	0,13189603	**0,206141940558881**
'PhosphateProduction'	0,00374933	**0,0346164203146861**
'TouristsNights'	0,17920095	**0,0154847686662484**
'EntriesNonResidents'	-0,07390975	**-0,00124537191241958**
'ManufacturingConfidenceIndicator'	0,20043047	**0,118849084187689**
'Tuniindex'	0,35383772	**0,321846148025490**
'M3'	0,04789035	**0,0699541101128850**
'Local Sales of Cement'	0,051245	**0.012345**
'CreditEconomy'	0,05161915	-0,0186665586111620
'CentralBankBalanceSheet'	-0,00798818	0,0177618314592436
'NetForeignAssets'	-0,03695034	0,0997966841740152
'TermsOfTrade'	-0,19909429	-0,203452620221471
'REER'	-0,01332977	0,0572514320785606
'EnergyPrices'	0,15939911	0,00758207278137772
'IPI_ManufEuroArea'	0,06947402	0,102560136442251
'CreditCardPayments'	-0,04022047	-0,0377228613747517
'JobDemand'	0,04641359	0,0147183938608800
'JobOffer'	0,11548025	0,130010906738261
'servicefinanciers'	-0,01006134	0,117654374518865
'RGDP'	1	0,950406280432888

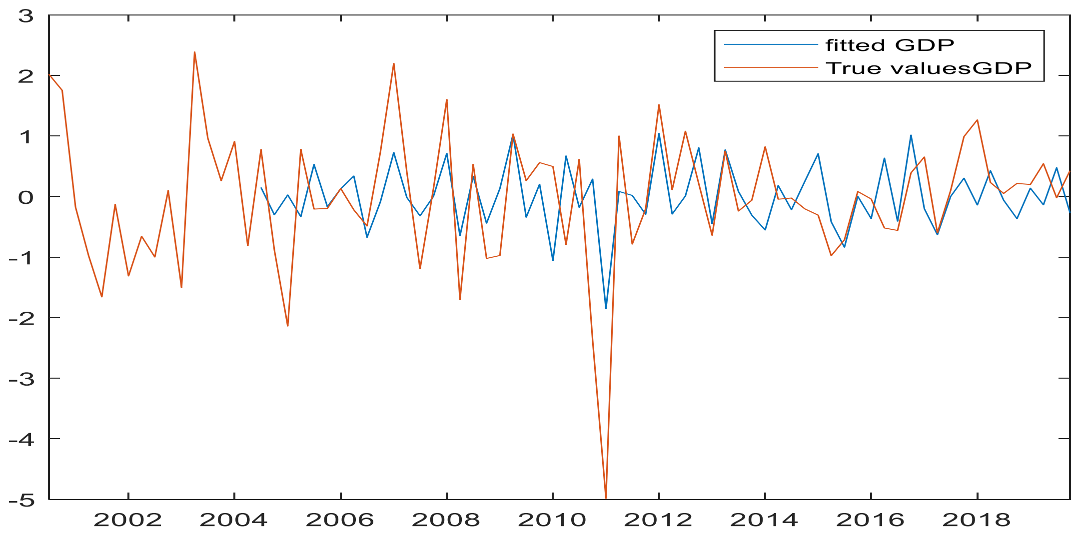
3.2 Benchmark model (uniform frequency VAR)

In the first step for this section, we used an initial ARMA baseline and an AR(1) model to estimate RGDP. As illustrated in graph (3), results are very modest.

Graph (3): Growth rate RGDP and estimated RGDP with AR(1) and ARMA(2,1)



Graph 4: **VAR (2) (In-Sample Fit) RGDP Growth**



For this reason, we have used a simple VAR (2), to see if using multiple right-hand side series including industrial production index, crude oil production, natural gaz production, imports, exports, electricity consumption, IPI_Euro manufacturing confidence indicator, tourist's nights, local sales of cement and stock market will improve our estimates. At least in-sample, this seems to be the case. We restrict this model to data published before RGDP releases and use only indicators that exhibit a high correlation with RGDP and residuals. Our initial analysis runs through the last quarter of 2019. In-sample results look reasonable, but not good, particularly towards the end of the sample as shown in figure (4).

4. Empirical results

4.1 In sample fit Estimation

4.1.1 In sample Individual AR-UMIDAS :

The Univariate Unrestricted MIDAS is estimated using monthly individual indicators such as industrial production index from 2000m01 until 2019m12.

Define the quarterly variables $xm_{tq}^{(i)},\ i \in \{1,2,3\}$ containing the i-month in the quarter, i.e.:

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

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

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

The U-Midas regression is defined as follow:

$$RGDP_{tq} = \alpha_1 + \beta_1 xm_{tq}^{(1)} + \beta_2 xm_{tq}^{(2)} + \beta_3 xm_{tq}^{(3)} + u_{tq} \qquad (14)$$

Table (3) U-Midas-AR (In-Sample Fit) IMPORTS

Variables	*Coefficients*	*st error*	*t-student*
Intercept	0.054766	0.0922	0.59397
AR(1)	0.34083	0.1022	3.33
AR(2)	0.25702	0.0958	2.68
X_1	0.128	0.054	2.37
R-squared :0,2209			
Adjusted R-Squared:0,199			

The first month of the industrial production index and two lags of RGDP within the quarter are very significant.

Figure 5: **U-Midas (In-Sample Fit) Imports**

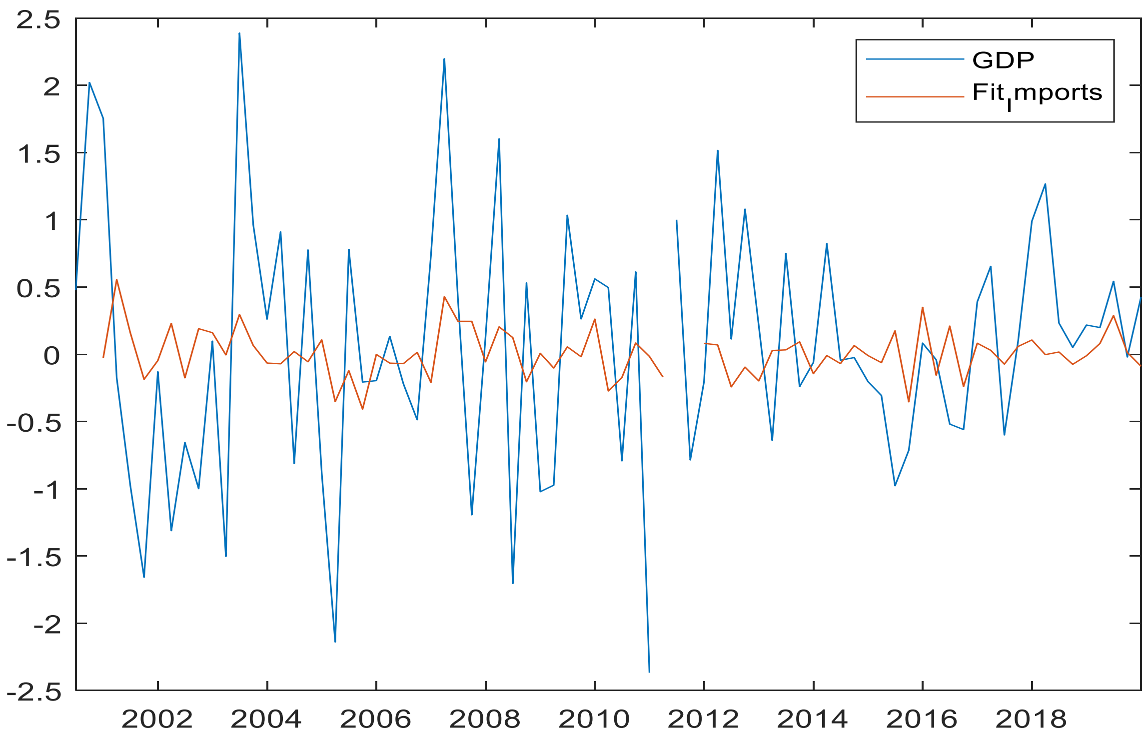


Table (4) U-Midas (In-Sample Fit) Industrial Production index

Variables	*Coefficients*	*st error*	*Student's t*
intercept	0.008	0.07	0.11
AR(1)	0.030	0.078	0.39
AR(2)	0.030	0.078	0.38
X1	0.217	0.08	2.6
X2	0.175	0.07	2.23
R-squared :0,2421			
Adjusted R-Squared:0,240			

Figure 6: U-Midas (In-Sample Fit) IPI

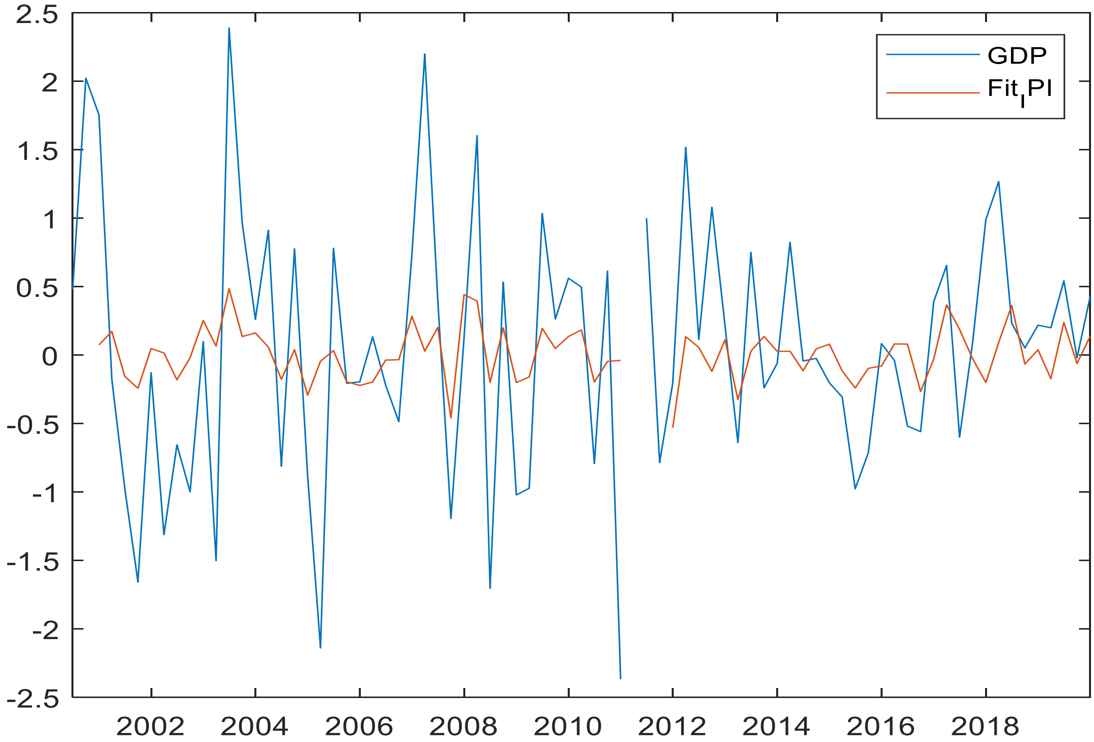


Table (5) U-Midas (In-Sample Fit) Stock market

Variables	*Coefficients*	*st error*	*Student's t*
AR(1)	0.0382	0.095	0.399
X1	0.0926	0.085	1.08
X2	0.205	0.110	1.85
X3	0.142	0.096	1.47
R-squared :0,1422			
Adjusted R-Squared:0,141			

4.1.2. In Sample Multivariate AR-UMIDAS

In this section, we used a multivariate Unrestricted Midas model, incorporating the relevant indicators which have predictive power in nowcasting RGDP. We included 18 RHS variables. An increase of RHS variables required more shrinking of parameter estimates toward zero to avoid overfitting.

Then, we regress the real GDP on the stacked monthly indicators released on the first, second a third month and we compare them in terms of the predictive ability by computing RMSE:

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

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

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

We regress quarterly RGDP on the monthly data and two autoregressive lags:

MMP= $[intercept, lag(1), lag(2), \underbrace{X1,}_{Industrial\ production\ index}, \underbrace{X2,}_{crude\ oil production}, \underbrace{X3,}_{Natural\ Gaz\ Production}, \underbrace{X4,}_{phosphate\ production}, \underbrace{X5,}_{TouristsNights}, \underbrace{X6,}_{Local\ sales\ of\ cement}, \underbrace{X5, X6,}_{Electricity\ manufact} \underbrace{X7, X8}_{Electricity\ services}, \underbrace{X9, X10,}_{IPI\_euro} \underbrace{X11, X12,}_{Imports} \underbrace{X13, X14, X15}_{stock\ market}, \underbrace{X16, X17, X18,}_{confidence\ marketEuro}]$.

Table (5) : Multivariate AR-Umidas Estimation

Variables	Coefficients	Std	Student's t
intercept	0.063	0.40	0.155
AR(1)	-0.055	0.40	-0.134
AR(2)	-0.032	0.40	-0.07
X1*	0.217	0.10	2.1*
X2	0.09	0.052	1.73
X3	0.03	0.109	0.27
X4	0.042	0.091	0.46
X5	0.044	0.134	0.33

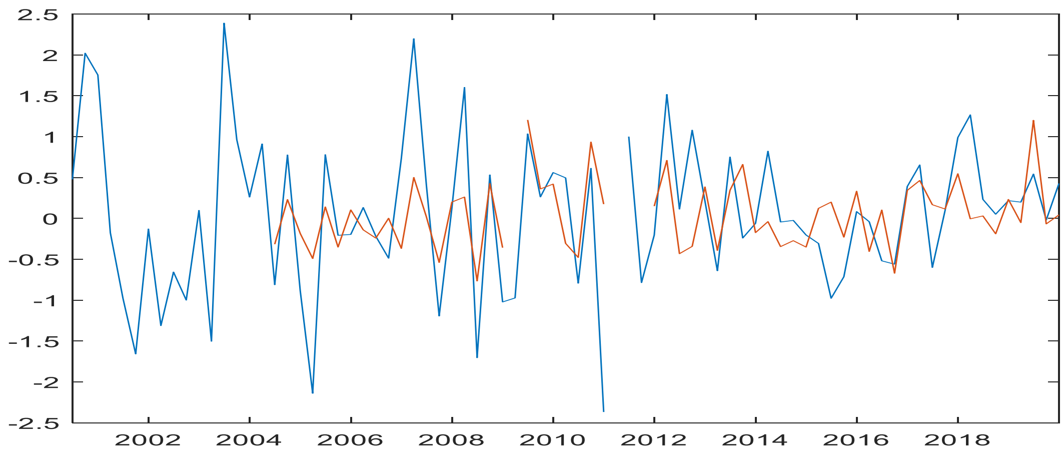
X6	0.063	0.406	0.15
X7*	0.166	0.0864	1.86*
X8	0,10392499	0,05906171	1,7596*
X9*	0.263	0.077	3.41*
X10	0.255	0.103	2.55*
X11	0.107	0.39	0.27
X12	0.17	0.14	1.20
X13	0.101	0.10	0.94
X14	0.22	0.11	2.00*
X15	0,25471364	0,1202614	2,118
X16	0.27	0.115	2.37
X17	0.137	0.126	1.08
X18	0.03	0.12	0.28

R-squared:0,6636

Adjusted R-Squared:0,5917

The **Multivariate AR-UMIDAS** yields an R^2, higher than Univariate Unrestricted Midas which is on the order of 66%. Although, the in-sample fit for this larger MIDAS looks good in the figure below, it is susceptible to overfitting.

Figure 7: Multivariate UMIDAS fitted RGDP



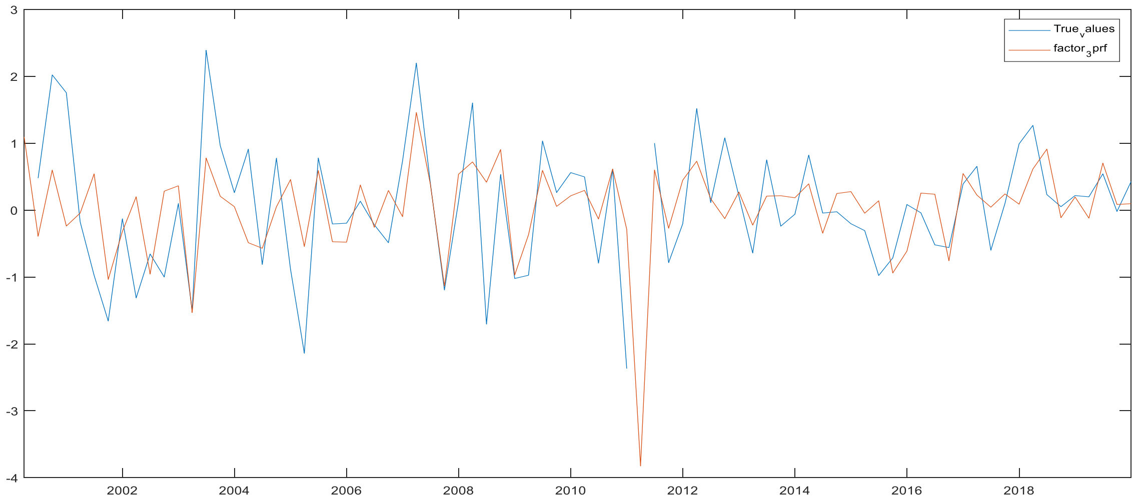
4.1.3 IN-Sample 3PRF:

The first pass consists of estimating parameters at quarterly frequency for right hand side variables. Specifically, the monthly indicators are aggregated to quarterly frequency. Then, we regress each predictor variable on Real GDP and retain slopes estimates.

In the second pass, we extract factors by regressing the cross section of predictor variables on the slope estimates from the first pass. Results are given in the table below:

Variables	*Coefficients*	*st error*	*t-student*
F1	0.042	0.094	0.45
F2	0.212	0.1061	1.998
F3	0.157	0.096	1.58
R-squared :0,6422			

Figure 8: 3PRF factor fitted RGDP



The figure (8) plots the resulting factor over time. The dynamics align well with the Tunisian RGDP growth. The factor declines sharply in 2011 (revolution: recession of the economy) and recovers with the beginning of the expansion in 2018.

4.1.2. In Sample Mixed Dynamic Factor Estimation

The advantage of the mixed dynamic factor models is that they deal with the issue of ragged-edge data, in which series have different publication lags. In this section, we estimate RGDP using a mixed dynamic factor model (MDFM), using the same monthly indicators. The model can be written as:

$$y_t = Hx_t + \varepsilon_t \quad (15)$$

And the transition equation

$$x_t = Ax_{t-1} + e_t \quad (16)$$

Specifically, we estimate a model similar to the mixed frequency dynamic factor model of Mariano and Murasawa (2003), which involves using quarterly RGDP together with a selection of monthly indicators of economic activity to extract a common (monthly) factor that is linked to the business cycle.

In terms of preventing overfitting, the DFM has the additional advantage of dimensions reduction. Moreover, using a DFM allows us to incorporate periods with incomplete observations due to varying publication lags and allows us to assign variables to blocks, such as consumption, production, and financial.

We elect also to use three blocks. In fact, blocks serve two purposes:

- First, blocks allow us to identify the model, so that the observations map to a unique set of factors.
- Second, they allow for more parsimonious estimation by a restricting several parameters to zero.

We estimate our DFM with three lags and three factors using observations excluding outliers to estimate the parameters of the model and allowing for autoregressive error terms, meaning that shocks to variables may be persistent.

Figure 9: Decomposition of RGDP to naming factors

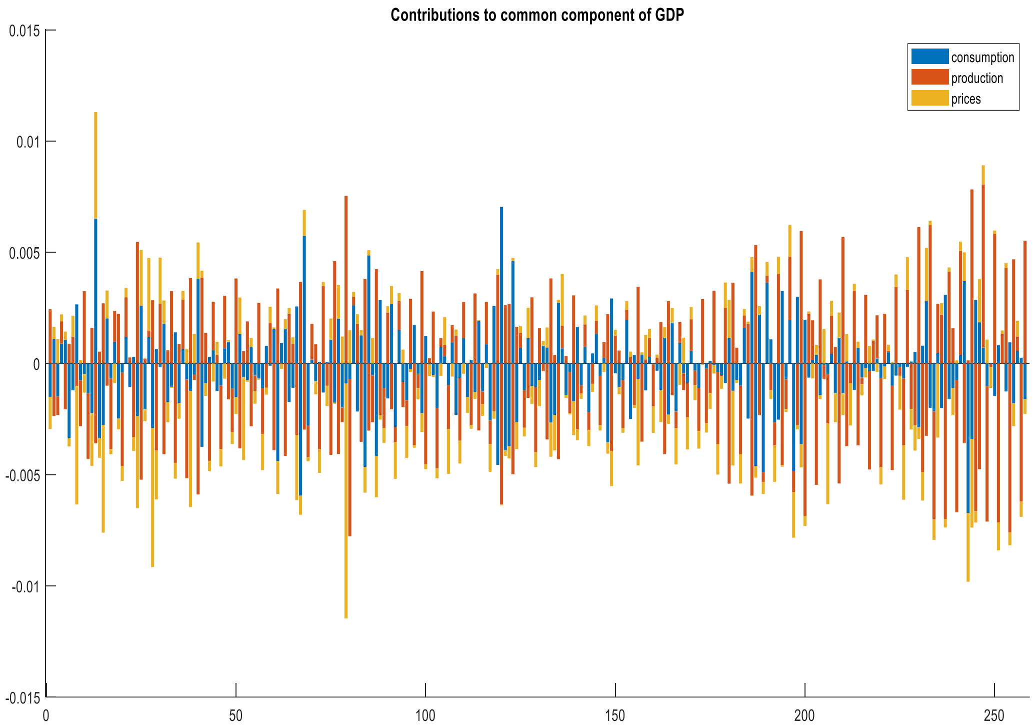


Figure 10: Detrended RGDP Growth and fitted values monthly factor (2000m01-2019m12)

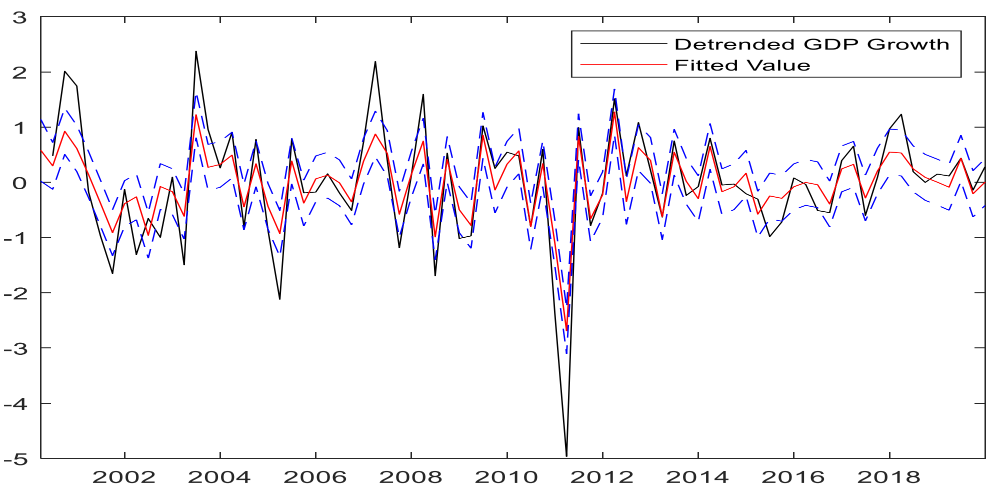


Figure 11: Detrended RGDP Growth VS Fitted values monthly factor (2000m01-2021m04)

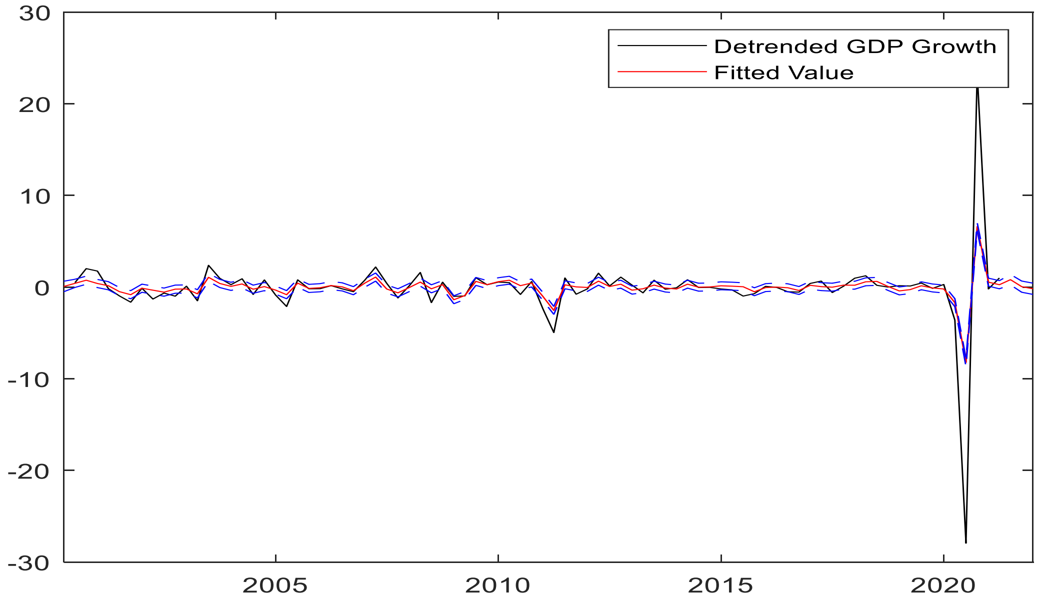


***Figure 11**: DFM is estimation results. This figure displays the time series used in the estimation of the dynamic factor model together with the estimate of the common factor, the estimated factor is reported in standardized units. In the bottom panel, RGDP and the factor are reported in annualized growth units.*

The results look good in sample, but the question is again will the model perform nearly as well out of sample?

4.3 Out of sample (Backtest) models:

In sample performance is not indicative of how the model will perform going forward in time. Indeed, an out of sample nowcast needed to test the performance of the model.

4.3.1 Backtest UMIDAS-AR

We backtest the model over the last 30 periods of the data, that is, over the period (30/06/2012 to30/09/2019), using the data up to the current backtesting date to estimate parameters.

Figure 12: Out of sample Backtest in Multivariate Umidas

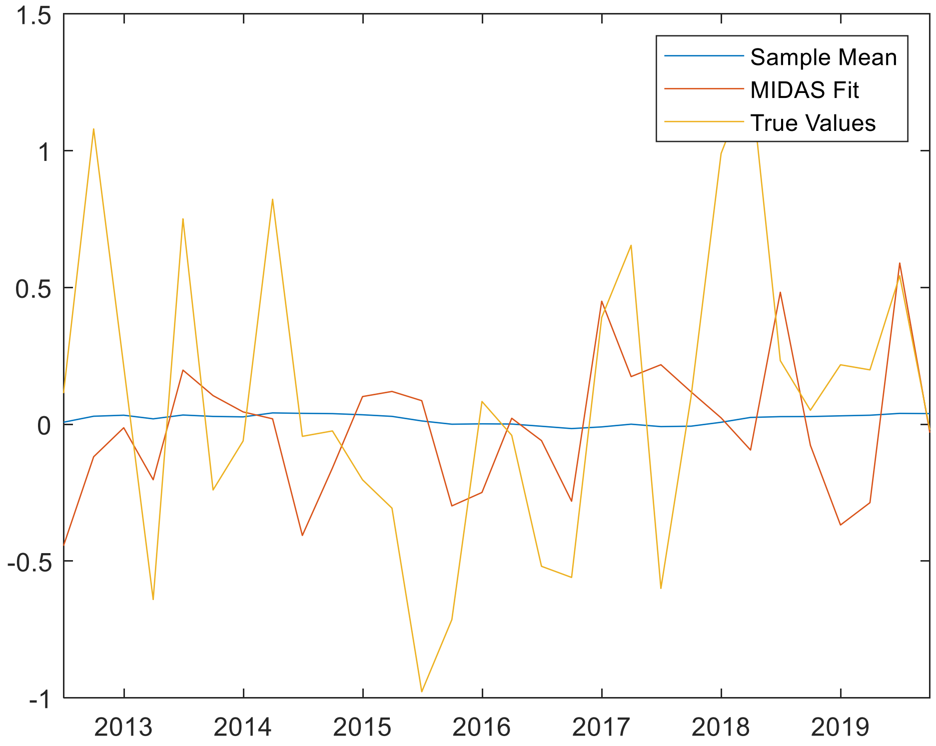
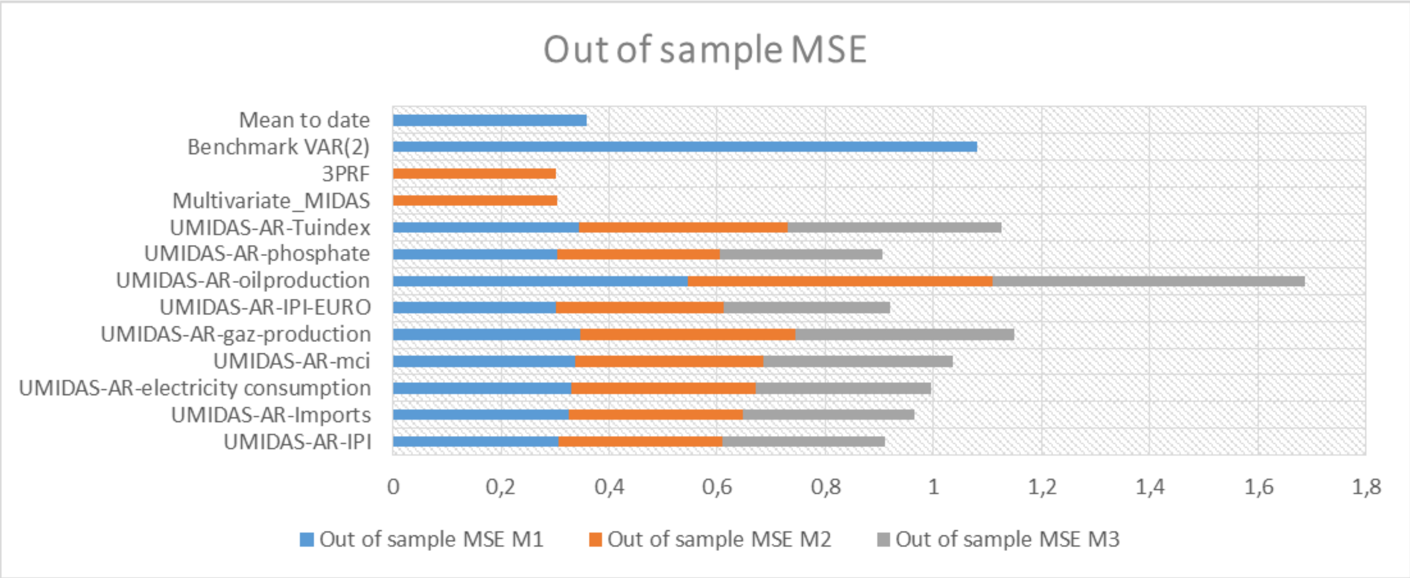


Table (6) *MSE performance of UMIDAS-AR models (out of sample)*

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**	
3PRF		**0,3012**	
Benchmark VAR(2)	1,0808		
Mean to date	0,3579		



Results backtesting show that the MSE of multivariate AR-MIDAS and 3PRF are smaller than the MSE for individual MIDAS models and the VAR(2) benchmark. So these models perform better out of sample.

When *M=3*, imports and consumption electricity become than the industrial production index.

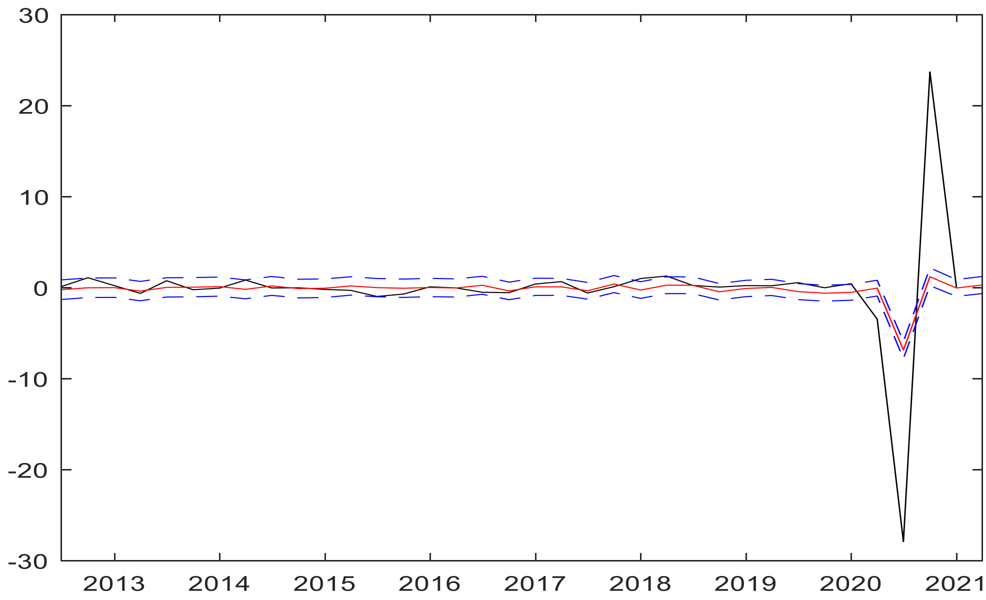
4.3.2 Backtest DFM

We identify the structure of missing observations, so that we can replicate the same structure of missing values in the tail of the data we used in each backtest period. Then, we backtest the model over the last 107 periods of the data over the period (30/06/2012 to 30/04/2021), using the data up to the current backtesting date to estimate parameters.

In the evaluation of the MDFM, we use the out of sample MSE produced in the specified test periods from the third month of 2012 to the end of the quarter 2019. To understand how well the MDFM fit the actual RGDP data, the MSE are then compared against the VAR (2) and the mean of the data of RGDP over the test periods. The best approach to backtesting is to have vintages data, saving real data on each date we want to create a backtest. However, as this not an option in our context, we will simply recreate the pattern of missing observations in the tail of the data.

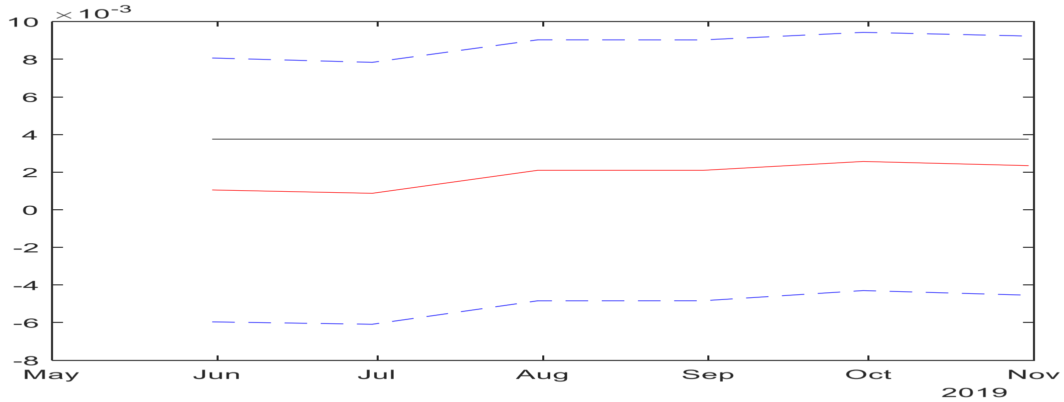
The COVID19 had induced a huge volatility which the model fails to completely capture. Before this period predictions seems good.

Figure 13: Backtest evaluation over the period (2012m06-2021m04)



We are interested in the question of how the Q2 2019 estimates evolves over time. We have six predictions for RGDP in Q2 2019 as shown in the figure below. In the graph below the black line is true RGDP for Q2 2019 and the nowcast of RGDP.

Figure (14): Six predictions for GDP in Q2 2019

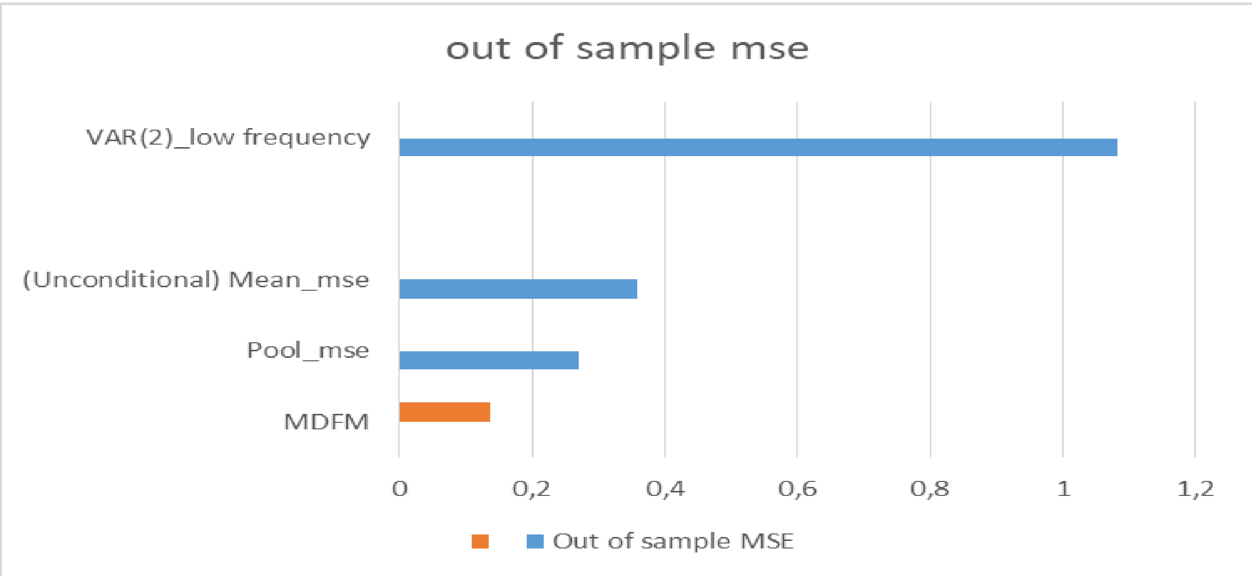


In fact, when if we get more date within the quarter the nowcast of RGDP approaches the true value of RGDP.

The table (7) below summaries the model performance:

Table (6): Model Performance Comparison (MSE)

Models	***Out of sample MSE***
MDFM	0.13654
Pool_mse	**0.2702**
(Unconditional) Mean_mse	0.3579
VAR(2)_low frequency	1.0808



Our results indicate that MFDFM and pooled models (mean of the different models as individual Unrestricted MIDAS, Multivariate MIDAS and MDFM) haved the lowest MSE, showing that these models have potential for nowcasting and forecasting with either higher volatility in RGDP (particularly in period COVID19) and with more limited data availability. Forecasting RGDP for further quarters of the year 2021 using Mixed Dynamic factor models, 3-PRF and pooled forecasts gives the following results:

Table (7): Forecasting further quarters ahead of RGDP growth (Q –Q) in year 2021 based on the releases data in the first and second month of the Q2 2021

Models	**Forecast (Q_Q)**		
	h=2	h=3	h=4
Mixed factor models	19,60%	0,45%	0,13%
Mixed frequency UMIDAS	18,50%	0,28%	0,10%
3 PRF	19,20%	0,31%	0,16%
Pooled forecasts	19,10%	0,34%	0,13%

Conclusion

This paper presents a set of mixed frequency models that improve nowcasting and short-term forecasting of economic activity in Tunisia. We develop and implement a univariate, multivariate MIDAS autoregressive models and mixed frequency dynamic factor models to nowcast quarterly RGDP for Tunisian economy.

We show that MFDFM and combined now-forecasting are more accurate than low frequency predictions using VAR(2); these models have a lowest mean squared errors in both in-sample and out of sample forecasting.

Future research can expand these models by increasing the reliability of nowcasts and forecasts using the same approach, developed by Marcellino and Forroni (2020), ad-hoc adjustments, via adjusting nowcasts by an amount similar to the nowcast and forecast errors made during the revolution period 2011.

References

[1] Carriero, A., Clark, T., and Marcellino, M. (2020). Nowcasting tail risks to economic activity with many indicators. Technical report, Cleveland FED WP 20-13.

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

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

[8] 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.

[9] Marcellino, M. (1998): "Temporal disaggregation, missing observations, outliers, and forecasting: a uniÖed non-model based procedure", Advances in Econometrics, 13, 181-202.

[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

ANNEXES

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 is one possible solution. The chart below decomposes observations in terms of common components (factors), autoregressive and iid errors.