SJSI

Nowcasting GDP in Parsimony Data Environment Using Bayesian Mixed Frequency VAR (Empirical Evidence from Syria)

Authors:
Khder Alakkari

Affiliation:
Department of Statistics and Programming - Faculty of Economics - Tishreen University, Tartous - Syria

ABSTRACT

Monitoring economic conditions in real-time or Nowcasting is among the most important tasks routinely performed by economists as it is important in describing the investment environment in any country. Nowcasting brings some key challenges that characterize modern frugal data analyses in developing countries, often referred to as the three (V)s. These include: the small number of continuously published time series (volume), the complexity of the data covering different sectors of the economy and being asynchronous with different frequency and accuracy to be published (variety), and the need to incorporate new information within months of its publication (velocity). In this article, we explored alternative ways to use Bayesian Mixed Frequency Vector Autoregressive (BMFVAR) models to address these challenges. The research found that BMFVAR can effectively handle the three (V)s and create real-time accurate probabilistic forecasts of the Syrian economic activity and, beyond that, a powerful narrative via scenario analysis.

Keywords: Bayesian Analysis – Mixed Frequency – Parsimony Data – VAR Model – Nowcasting.

INTRODUCTION

Nowcasting refers to the projection of information about the present, the near future, and even the recent past. The importance of nowcasting has shifted from weather forecasting to economics, where economists use it to track the economy status through real-time GDP forecasts, as it is the main low-frequency (quarterly - annually) indicator reflecting the state of the country's economy. This is like satellites that reflect the weather on earth. It does this by relying on high-frequency measured variables (daily - monthly) that are reported in real-time. The importance of using nowcasting in the economy stems from the fact that data issuers, statistical offices and central banks release the main variables of the economy, such as gross domestic product and its components, with a long lag. In some countries, it may take up to five months. In other countries, it may take two years depending on the capabilities each country has, leading to a state of uncertainty about the economic situation among economic policy makers and state

followers business. Real-time indicators related to the economy (e.g consumer prices and exchange rates) are used here in order to obtain timely information about variables published with a delay. The first use of nowcasting technology in economics dates back to Giannone et al 2008, by developing a formal forecasting model that addresses some of the key issues that arise when using a large number of data series released at varying times and with varying delays [1]. They combine the idea of "bridging" the monthly information with the nowcast of quarterly GDP and the idea of using a large number of data releases in a single statistical framework. Banbura et al proposed a statistical model that produces a series of Nowcasting forecasts on real-time releases of various economic data [2,3]. The methodology enables the processing of a large amount of information from nowcasting's Eurozone GDP Q4 2008 study. Since that time, the models that can be used to create nowcasting have expanded. Kuzin et al compared the mixed-frequency data sampling (MIDAS) approach proposed by Ghysels et al [4,5] with the mixed frequency VAR (MF-VAR) proposed by Zadrozny and Mittnik et al [6,7], with model specification in the presence of mixed-frequency data in a policy making situation, i.e. nowcasting and forecasting quarterly GDP growth in Eurozone on a monthly basis. After that time, many econometric models were developed to allow the use of nowcasting and solve many data problems. Ferrara et al [8] proposed an innovative approach using nonparametric methods, based on nearest neighbor's approaches and on radial basis function, to forecast the monthly variables involved in parametric modeling of GDP using bridge equations. Schumacher et al [9] compare two approaches from nowcasting GDP: Mixed Data Sampling (MIDAS) regressions and bridge equations. Macroeconomics relies on increasingly non-standard data extracted using machine learning (text analysis) methods, with the analysis covering hundreds of time series.

Some studies examined US GDP growth forecast using standard high-frequency time series and non-standard data generated by text analysis of financial press articles and proposed a systematic approach to high-dimensional time regression problems [10-11]. Another team of researchers worked on dynamic factor analysis models for nowcasting GDP [12], using a Dynamic Factor Model (DFM) to forecast Canadian GDP in real time. The model was estimated using a mix of soft and hard indices and the authors showed that the dynamic factor model outperformed univariate criteria as well as other commonly used nowcasting models such as MIDAS and bridge regressions. Anesti et al [13] proposed a release-enhanced dynamic factor model (RA-DFM) that allowed quantifying the role of a country's data flow in the nowcasting of both early (GDP) releases and later revisions of official estimates. A new mixed-frequency dynamic factor model with time-varying parameters and random fluctuations was also designed for macroeconomic nowcasting, and a fast estimation algorithm was developed [14]. Deep learning models also entered the field of GDP nowcasting, as in many previous reports [15-17]. In Syria, there are very few attempts to nowcasting GDP, among which we mention a recent report that uses the MIDAS Almon Polynomial Weighting model to nowcasting Syria's annual GDP based on the monthly inflation rate data [18]. Our research aims to solve a problem that exists in the Arab and developing countries in general, and Syria in particular, namely the inability to collect data in real time on the main variable in the economy due to the weakness of material and technical skills. Therefore, this study uses Bayesian mixed-frequency VAR models to nowcasting GDP in Syria based on a set of high-frequency data. The rationale for choosing these models is that they enable the research goal to be achieved within a structural economic framework that reduces statistical uncertainty in the domain of

high-dimensional data in a way that distinguishes them from the nowcasting family of models, according to a work by Cimadomo et al and Crump et al [19-20]. The first section of this research includes the introduction and an overview of the previous literature. The second section contains the research econometric framework, through which the architecture of the research model is developed within a mathematical framework. The third section is represented by the data used in the research including the exploratory phase. The fourth section contains the discussion and interpretation of the results of the model after the evaluation. The fifth section presents the results of the research and proposals that can consider realistic application by the authorities concerned.

MATERIALS AND METHODS

The working methodology in this research is divided into two main parts. The *first* in which the low-frequency variable (Gross Domestic Product) is converted from annual to quarterly with the aim of reducing the forecast gap and tracking the changes in the GDP in Syria in more real-time., by reducing the gap with high-frequency data that we want to predict their usage. To achieve this, we used Chow-Lin's Litterman: random walk variant method.

Chow-Lin's Litterman Method:

This method is a mix and optimization of two methods. The Chow-Lin method is a regression-based interpolation technique that finds values of a series by relating one or more higher-frequency indicator series to a lower-frequency benchmark series via the equation:

$$x(t) = \beta Z(t) + a(t) \tag{1}$$

Where β is a vector of coefficients and a random variable with mean zero and covariance matrix V. Chow and Lin [21] used generalized least squares to estimate the covariance matrix, assuming that the errors follow an AR (1)

process, from a state space model solver with the following time series model:

$$a(t) = \rho a(t-1) + \epsilon(t) \tag{2}$$

Where $\epsilon(t) \sim N(0, \sigma^2)$ and $|\rho| < 1$. The parameters ρ and β are estimated using maximum likelihood and Kalman filters, and the interpolated series is then calculated using Kalman smoothing. In the Chow-Lin method, the calculation of the interpolated series requires knowledge of the covariance matrix, which is usually not known. Different techniques use different assumptions about the structure beyond the simplest (and most unrealistic) case of homoscedastic uncorrelated residuals. A common variant of Chow Lin is Litterman interpolation [22], in which the covariance matrix is computed from the following residuals:

$$a(t) = a(t-1) + \epsilon(t) \tag{3}$$

Where $\epsilon(t) \sim N(0, V)$

$$\epsilon(t) = \rho\epsilon(t-1) + e(t) \tag{4}$$

and the initial state $a(0) = 0$. This is essentially an ARIMA (1,1,0) model.

In the *second* part, an econometric model suitable for this study is constructed by imposing constraints on the theoretical VAR model to address a number of statistical issues related to the model's estimation, represented by the curse of dimensions.

Curse of Dimensions

The dimensional curse basically means that the error increases with the number of features (variables). A higher number of dimensions theoretically allows more information to be stored, but rarely helps as there is greater potential for noise and redundancy in real data. Collecting a large number of data can lead to a dimensioning problem where very noisy

dimensions can be obtained with less information and without much benefit due to the large data [23]. The explosive nature of spatial scale is at the forefront of the Curse of Dimensions cause. The difficulty of analyzing high-dimensional data arises from the combination of two effects: **1-** Data analysis tools are often designed with known properties and examples in low-dimensional spaces in mind, and data analysis tools are usually best represented in two- or three-dimensional spaces. The difficulty is that these tools are also used when the data is larger and more complex, and therefore there is a possibility of losing the intuition of the tool's behavior and thus making wrong conclusions. **2-** The curse of dimensionality occurs when complexity increases rapidly and is caused by the increasing number of possible combinations of inputs. That is, the number of unknowns (parameters) is higher than the number of observations. Assuming that **m** denotes dimension, the corresponding covariance matrix has **m (m+1)/2** degrees of freedom, which is a quadratic term in **m** that leads to a high dimensionality problem. Accordingly, by imposing a skeletal structure through the initial information of the Bayesian analysis, we aimed to reduce the dimensions and transform the high-dimensional variables into variables with lower dimensions and without changing the specific information of the variables. With this, the dimensions were reduced in order to reduce the feature space considering a number of key features.

Over Parameterizations

This problem, which is an integral part of a high dimensionality problem, is statistically defined as adding redundant parameters and the effect is an estimate of a single, irreversible singular matrix [24]. This problem is critical for statistical estimation and calibration methods that require matrix inversion. When the model starts fitting the noise to the data and the estimation parameters, i.e. H. a high degree of correlation existed in the co-correlation matrix of the residuals, thus producing predictions with large out-of-sample errors. In other words, the uncertainty in the estimates of the parameters and errors increases and becomes uninterpretable or far from removed from the realistic estimate. This problem is addressed by imposing a skeletal structure on the model, thereby transforming it into a thrift. Hence, it imposes constraints that allow a correct economic interpretation of the variables, and reducing the number of unknown parameters of the structural model, and causing a modification of the co - correlation matrix of the residuals so that they become uncorrelated with some of them, in other words, become a diagonal matrix.

Overfitting and Underfitting

Overfitting and Underfitting are major contributors to poor performance in models. When Overfitting the model (which works perfectly on the training set while ineffectively fitting on the test set), it begins with matching the noise to the estimation data and parameters, producing predictions with large out-of-sample errors that adversely affect the model's ability to generalize. An overfit model shows low bias and high variance [25]. Underfitting refers to the model's inability to capture all data features and characteristics, resulting in poor performance on the training data and inability to generalize the model results [26]. To avoid and detect overfitting and underfitting, we tested the validity of the data by training the model on 80% of the data subset and testing the other 20% on the set of performance indicators.

Theoretical VAR model:

VAR model developed by Sims [27] have become an essential element in empirical macroeconomic research. Autoregressive models are used as tools to study economic shocks because they are based on the concept of dynamic behavior between different lag values for all variables in the model, and these models are considered to be generalizations of autoregressive (AR) models. The p-rank VAR is called the VAR$_p$ model and can be expressed as:

$$y_t = C + \beta_1 y_{t-1} + \cdots + \beta_P y_{t-P} + \epsilon_t \quad \epsilon_t \sim (0, \Sigma) \qquad (5)$$

Where y_t is a $K \times 1$ vector of endogenous variables. β_P Is matrix of coefficients corresponding to a finite lag in y_t, ϵ_t: random error term with mean 0 representing external shocks, Σ: matrix (variance – covariance). The number of parameters to be estimated is $K + K^2 p$, which increases quadratic with the number of variables to be included and linearly in order of lag. These dense parameters often lead to inaccuracies regarding out-of-sample prediction, and structural inference especially for high-dimensional models.

Bayesian Inference for VAR model

The Bayesian approach to estimate VAR model addresses these issues by imposing additional structure on the model, the associated prior information, enabling Bayesian inference to solve these issues [28-29], and enabling the estimation of large models [2]. It moves the model parameters towards the parsimony criterion and improves the out-of-sample prediction accuracy [30]. This type of contraction is associated with frequentist regularization approaches [31]. Bayesian analysis allows us to understand a wide range of economic problems by adding prior information in a normal way, with layers of uncertainty can be explained through hierarchical modelling [32].

Prior Information:

The basic premise for starting a Bayesian analysis process must have prior information, and identifying it correctly is very important. Studies that try not to impose prior information result in unacceptable estimates and weak conclusions. Economic theory is a detailed source of prior information, but it lacks many settings, especially in high-dimensional models. For this reason, Villani [33] reformulates the model and places the information at a steady state, which is often the focus of economic theory and which economists better understand. It has been proposed to determine the initial parameters of the model in a data-driven manner, by treating them as additional parameters to be estimated. According to the hierarchical approach, the prior parameters are assigned to hyper priors. This can be expressed by the following Bayes law:

$$p(\gamma|y) = p(y|\theta, \gamma) p(\theta|\gamma) p(\gamma) \qquad (6)$$

$$p(y|\gamma) = \int p(y|\theta, \gamma) p(\theta|\gamma) d\theta \qquad (7)$$

Where $y = (y_{p+1} + \cdots + y_T)^T$, θ: is coefficients AR and variance for VAR model, γ: is hyper parameters. Due to the coupling of the two equations above, the ML of the model can be efficiently calculated as a function of γ. Giannone et al [34] introduced three primary information designs, called the Minnesota Litterman Prior, which serves as the basis, the sum of coefficients [35] and single unit root prior [36].

Minnesota Litterman Prior:

Working on Bayesian VAR priors was conducted by researchers at the University of Minnesota and the Federal Reserve Bank of Minneapolis [37] and these early priors are often referred to as Litterman prior or Minnesota Prior. This family of priors is based on assumption that Σ is

known; replacing Σ with an estimate Σ. This assumption leads to simplifications in the prior survey and calculation of the posterior.

The prior information basically assumes that the economic variables all follow the random

$$[((A_s)_{ij}|\Sigma)] = \begin{cases} 1 & if\ i = j\ and\ s = 1 \\ 0 & if\ otherwise \end{cases}$$

$$cov((A_S)_{IJ},\ (A_r)_{kl}|\Sigma) = \begin{cases} \lambda^2\ \dfrac{1}{s^\propto}\ \dfrac{\Sigma_{ik}}{\psi_j\ /(d-M-1)} & if\ l = j\ and\ r = s \\ 0 & otherwise \end{cases} \qquad (8)$$

The key parameter is λ which controls the magnitude influence of the prior distribution, i.e. it weighs the relative importance of the primary data. When $\lambda \longrightarrow 0$ the prior distribution is completely superimposed, and in the case of $\lambda \longrightarrow \infty$ the estimate of the posterior distributions will approximate the OLS estimates. $\propto$ is used to control the break attenuation, and ψ is used to control the prior standard deviation when lags are used. Prior Minnesota distributions are implemented with the goal of de-emphasizing the deterministic component implied by the estimated VAR models in order to fit them to previous observations. It is a random analysis-based methodology for evaluating systems under non-deterministic (stochastic) scenarios when the analytical solutions are complex, since this method is based on sampling independent variables by generating phantom random numbers.

walk process. These specifications lead to good performance in forecasting the economic time series. Often used as a measure of accuracy, it follows the following torques:

Dummy Observations

Starting from the idea of Sims and Zha [38] to complement the prior by adding dummy observations to the data matrices to improve the predictive power of Bayesian VAR. These dummy observations consist of two components; the sum-of-coefficient component and the dummy-initial-observation component. The sum-of-coefficients component of a prior was introduced by Doan et al [37] and demonstrates the belief that the average of lagged values of a variable can serve as a good predictor of that variable. It also expresses that knowing the average of lagged values of a second variable does not improve the prediction of the first variable. The prior is constructed by adding additional observations to the top (pre-sample) of the data matrices. Specifically, the following observations are constructed:

$$y^+_{M\times M} = diag\left(\dfrac{\bar{y}}{\mu}\right)$$

$$x^+_{M\times(1+Mp)} = [0, y^+, \dots, y^+] \qquad (9)$$

Where y is the $M \times 1$ vector of the means over the first p observed by each variable, the key parameter μ is used to control for variance and hence the impact of prior information. When $\mu \longrightarrow \infty$ the prior information becomes

uninformative and when $\mu \longrightarrow 0$ the model is formed into a formula consisting of a large number of unit roots as variables and without co-integration. The dummy initial observation component [36] creates a single dummy

observation that corresponds to the scaled average of the initial conditions, reflecting the belief that the average of the initial values of a variable is likely to be a good prediction for that variable. The observation is formed as:

$$y^{++}_{1 \times M} = \frac{\bar{y}}{\delta}$$

$$x^{++}_{1 \times (1+MP)} = \left[\frac{1}{\delta}, y^{++}, \dots, y^{++}\right] \tag{10}$$

Where y is the $M \times 1$ vector of the means over the first p observed by each variable, the key parameter δ is used to control on variance and hence the impact of prior information. As $\delta \to 0$, all endogenous variables in VAR are set to their unconditional mean, the VAR exhibits unit roots without drift, and the observation agrees with co-integration.

Structural Analysis

The nowcasting technique within the VAR model differs from all nowcasting models in the possibility of an economic interpretation of the effect of high-frequency variables on the low-frequency variable by measuring the reflection of nonlinear changes in it, which are defined as (impulse and response). A shock to the y_t variable not only affects the y_t variable directly, but it is also transmitted to all other endogenous variables through the dynamic (lag) structure of VAR. An impulse response function tracks the impact of a one-off shock to one of the innovations on current and future values of the endogenous variable.

BVAR is estimated reductively, i.e., without a contemporary relationship between endogenous variables in the system. While the model summarizes the data, we cannot determine how the variables affect each other because the reduced residuals are not orthogonal. Recovering the structural parameters and determining the precise propagation of the shocks requires constraints that allow a correct economic interpretation of the model, identification constraints that reduce the number of unknown parameters of the structural model,

and lead to a modification of the co-correlation matrix of the residuals so that they become uncorrelated, i.e., they become a diagonal matrix. This problem is solved by recursive identification and achieved by Cholesky's analysis [39] of the variance and covariance matrix of the residuals Σ, Where Cholesky uses the inverse of the Cholesky factor of the residual covariance matrix to orthogonalize the impulses. This option enforces an ordering of the variables in the VAR and attributes the entire effect of any common component to the variable that comes first in the VAR system. For Bayesian frame, Bayesian sampling will use a Gibbs or Metropolis-Hasting sampling algorithm to generate draws from the posterior distributions for the impulse. The solution to this problem can be explained mathematically from the VAR model:

$$y_t = \beta_0 y_t + \beta_1 y_{t-1} + \epsilon_t \tag{11}$$

Where β_0 show us contemporaneous correlation. β_1 Coefficient matrix at lag1, ϵ_t error term where:

$$\epsilon \sim N(0,1)$$

$$\Sigma_\epsilon = E(\epsilon_t \epsilon'_t) = I \tag{12}$$

In order to estimate the previous equation, we must abandon the contemporaneous relations between the endogenous variables y_t since we transfer them to the other side:

$$(I - \beta_0)y_t = A_1 y_{t-1} + \epsilon_t$$

$$B_0^* y_t = A_1 y_{t-1} + \epsilon_t \text{ , where } B_0^* = I - \beta_0 \tag{13}$$

$$y_t = B_0^{*-1} A_0 y_{t-1} + B_0^{*-1} \epsilon_t$$

Now parsimony VAR model can be estimated:

$$y_t = A_1 y_{t-1} + u_t \tag{14}$$

Where A_1 are reduced coefficient. u_t Represent the weighted averages of the structural coefficients β_1, ϵ_t. Where:

$$A_1 = B_0^{*-1} A_t$$

$$u_t = B_0^{*-1} \epsilon_t \tag{15}$$

u_t is an uncorrelated and exogenous series and $\sum_u$ become a diagonal matrix.

Mixed Frequency VAR

Complementing the constraints imposed on the theoretical VAR model, as a prelude to achieving the appropriate form for the research model and developing the model presented by AL-Akkari and Ali [40] to predict macroeconomic data in Syria, we formulate the appropriate mathematical framework for the possibility to benefit from high-frequency data emitted in real-time in nowcasting of GDP in Syria.

We estimate a mixed frequency VAR model with no constant or exogenous variables with only two data frequencies, low and high (quarterly-monthly), and that there are m high frequency periods per low frequency period. Our model consists of k_L variable y_L^i observed at low frequency and k_H variables y_H^i observed at high frequency. Where:

$y_{L,t}^i$: represent the i low-frequency variable observed during the low-frequency period t_L.

$y_{H,t}^i$: represent the i high-frequency variable observed during the low-frequency period t_H.

By stacking the k_L and k_H variables into the matrices Y_L and Y_H ignoring the intersections and exogenous variables for brevity, respectively, we can write the VAR:

$$\begin{vmatrix} Y_{H,t_L,1} \\ Y_{H,t_L,2} \\ \cdots \\ Y_{H,t_L,m} \\ Y_{L,t_L} \end{vmatrix} = \sum_{j=1}^{p} \begin{bmatrix} \Pi_j^{1,1} & \Pi_j^{1,2} & \cdots & \Pi_j^{1,m} & \Pi_j^{1,m+1} \\ \Pi_j^{2,1} & \cdots & \cdots & \cdots & \Pi_j^{2,m+1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \Pi_j^{m,1} & \cdots & \cdots & \Pi_j^{m,m} & \Pi_j^{m,m+1} \\ \Pi_j^{m+1,1} & \cdots & \cdots & \Pi_j^{m+1,m} & \Pi_j^{m+1,m+1} \end{bmatrix} \begin{vmatrix} Y_{H,t_L-j,1} \\ Y_{H,t_L-j,2} \\ \cdots \\ Y_{H,t_L-j,m} \\ Y_{L,t_L-j} \end{vmatrix} + \begin{vmatrix} E_{H,t_{L,1}} \\ E_{H,t_{L,2}} \\ \cdots \\ E_{H,t_{L,m}} \\ E_L, t_L \end{vmatrix}$$

$$(16)$$

Where $\Pi_j^{a,b}$ is $k_H \times k_H$, $\Pi_j^{m+1,b}$ is $k_L \times k_H$, and $\Pi_j^{a,m+1}$ is $k_H \times k_L$, for all $j, a, b = 1, \ldots, m$

and $\Pi_j^{m+1,m+1}$ is $k_L \times k_L$.

Performance indicators

We use indicators to assess the performance of the models which are used to determine their ability to explain the characteristics and information contained in the data. It does this by examining how closely the values estimated from the model correspond to the actual values, taking into account the avoidance of an underfitting problem that may arise from the training data and an overfitting problem that arises from the test data. The following performance indicators include:

Theil coefficient (U):

$$\frac{\sqrt{\frac{1}{n}\sum_t^n (y_t - \hat{y}_t)^2}}{\sqrt{\frac{1}{n}\sum_t^n y^2} + \sqrt{\frac{1}{n}\sum_t^n \hat{y}^2}} \tag{17}$$

Mean Absolute Error (MAE):

$$\frac{1}{n}\sum_{t=1}^{n}\left|y_t - \hat{y}_t\right| \quad\quad\quad (18)$$

Root Mean Square Error (RMSE):

$$\sqrt{\frac{\sum_{t=1}^{n}(\hat{y}_t - y_t)^2}{n}} \quad\quad\quad (19)$$

Where $\hat{y}_t$ the forecast is value; y_t is the actual value; and n is the number of fitted observed. The smaller the values of these indicators, the better the performance of the model.

RESULTS & DISCUSSION

As mentioned, the data consists of high and low frequency variables collected from the websites of official Syrian organizations and the World Bank [41-46], summarized in Table 1, which shows basic information about data. The data of the variable GDP were collected annually and converted using the combination of Chow-Lin's Litterman methods, (Fig 1), which shows that the volatility in annual GDP was explained by the quarter according to the hybrid method used, and gives us high reliability for the possibility of using this indicator in the model. For high-frequency data, the monthly closing price was collected from the sources listed in Table (1) and its most recent version was included in the model to provide real-time forecasts.

| Table 1. basic information about data				
Variable	Code	Frequency	Time period	Source
Gross Domestic Product	GDP	Quarterly	2010Q1-2020Q1	(Central Bureau Of Statistics , 2021)
Consumer Prices Index	CPI	Monthly	2010M1-2022M6	(Central Bureau Of Statistics, 2022)
Long Term Interest Rate	INT	Monthly	2010M1-2022M3	(Central Bank of Syria, 2022)
Financial Market Price Index	DWX	Monthly	2010M1-2023M3	(Damascus Stock Exchange, 2023)
Parallel Exchange Rate	EXR	Monthly	2010M1-2023M3	(Central Bank of Syria, 2023)
Gold Prices	GOP	Monthly	2010M1-2023M3	(Craftsmanship Association, 2023)
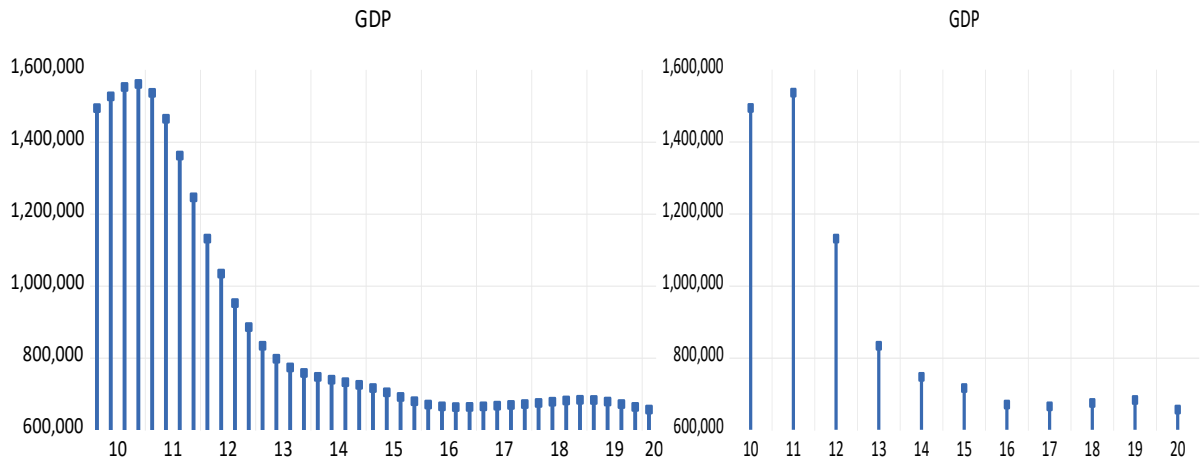
Figure 1. Annual and Quarterly GDP in Syria.

Figure (2) shows the evolution of these variables. We presented in Figure (2), the data in its raw form has a general stochastic trend and differs in its units of measurement. To unify the characteristics of this data, growth rates were used instead of prices in this case, or called log, because of their statistical properties. These features are called stylized facts; first, there is no normal distribution. In most cases, the distribution deviates to the left and has a high kurtosis. It has a high peak and heavy tails [47]. Second, it has the property of stationary and there is almost no correlation between the different observations [48]. The log of this series is calculated.
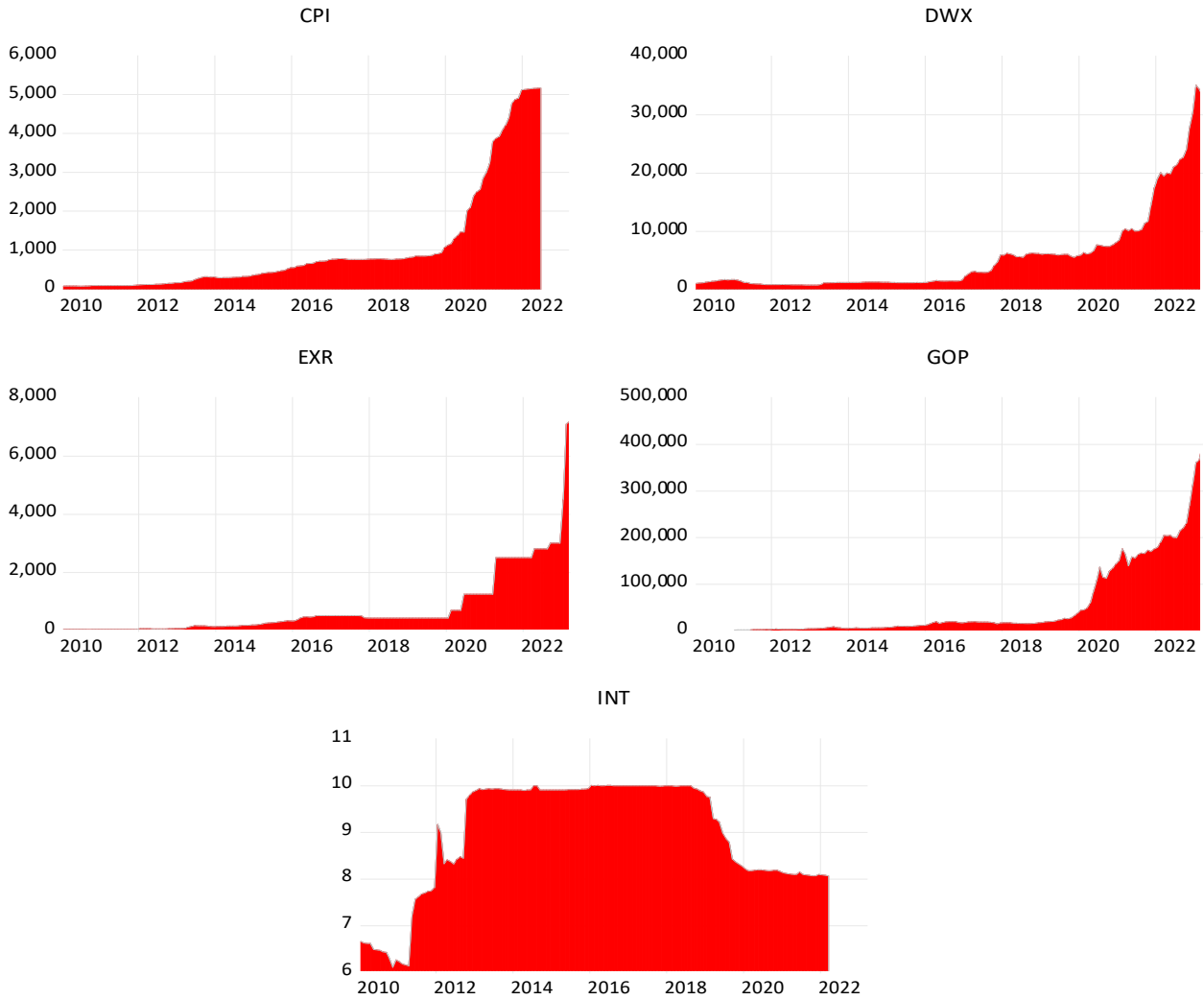
Figure 2. Evolution of high frequency study variables

Figure (3) shows the magnitude growth of the macroeconomic and financial variables in Syria during the indicated period. We note that the data is characterized by the lack of a general trend and volatility around a constant. We found that the fluctuations change over time and follow the characteristic of stochastic volatility that decreases, increases and then decreases. We found that the most volatile variable is the exchange rate (EXR) and the period between 2019 and 2020 is the one when the variables fluctuated the most due to uncertainty and US economic sanctions. We also noted that the periods of high volatility are offset by negative growth in Syria's GDP. Using the following Table (2), we show the most important descriptive statistics for the study variables.
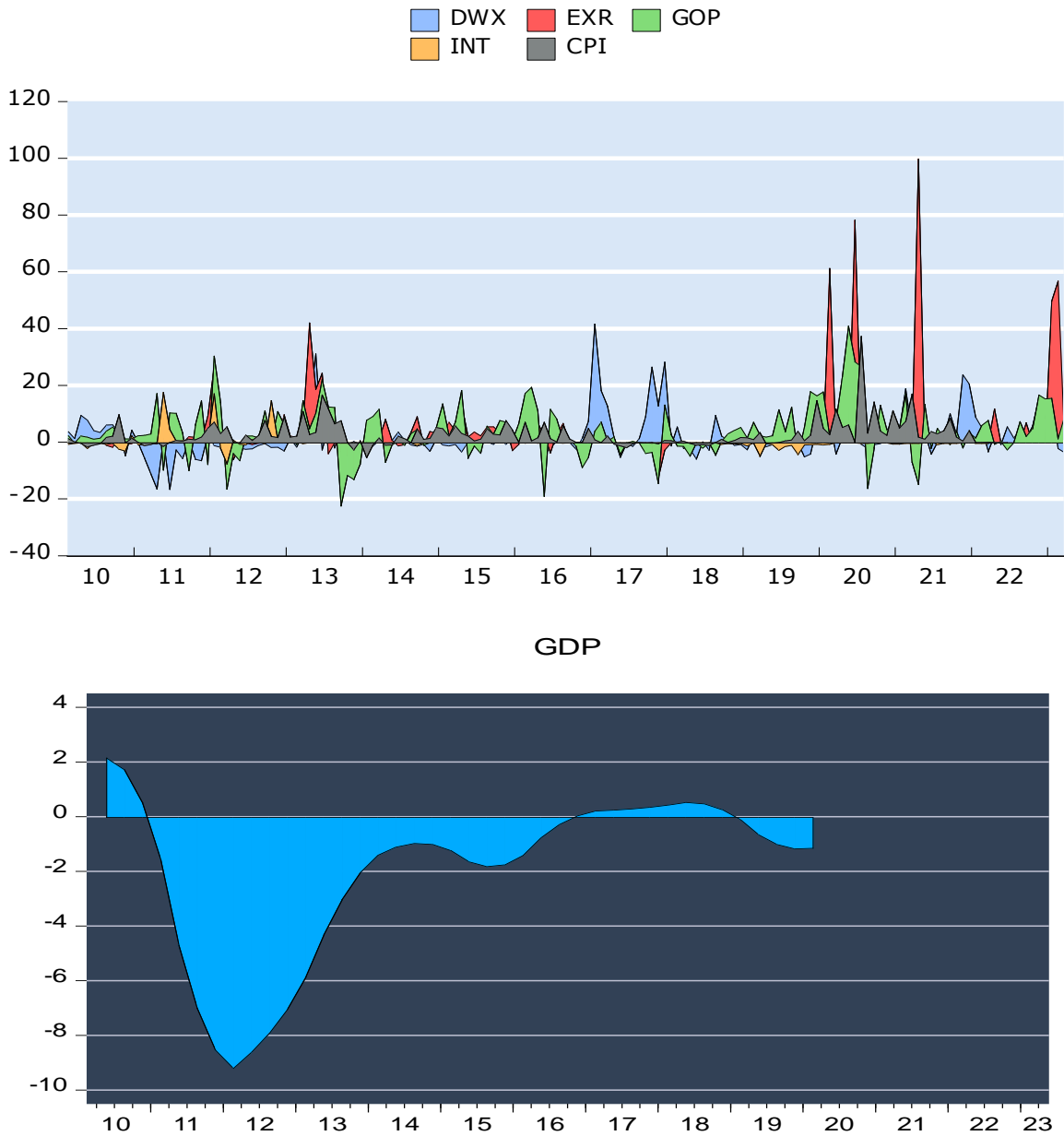
GDP

Figure 3. Evolution of log high - low frequency study variables.

Table 2. Descriptive statistics for variables							
Variables	Normality J-B	Mean	Standard Deviation	Maximum	Minimum	Skewness	Kurtosis
GDP	34.15***	-1.98	3.01	2.166	-9.181	-1.15	3.189
DWX	404.52***	2.25	7.62	41.84	-16.38	2.05	10.03
EXR	6915.8***	12.7	12.7	100.1	-12.31	5.33	34.9**8**
GOP	28.061***	3.88	9.36	**41.17**	**-22.22**	**0.51**	**4.893**
INT	6460.3***	0.16	2.62	**17.75**	**-7.392**	**4.97**	**34.02**
CPI	2569.8***	2.91	4.71	**37.05**	-5.174	3.45	22.35

*** denotes significance of the statistical value at 1%. ** at 5%. * at 10%.

Table (2) shows that the probability value of the normal distribution test statistic is significant at 1%, and we conclude that the data on the growth rates of the Syrian economy is not distributed according to the normal distribution, and both the mean and the standard deviation were not useful for the prediction in this case since they have become a breakdown. We also note from Table (2) the positive value of the skewness coefficient for all variables, and therefore the growth rates of economic variables are affected by shocks in a way that leads to an increase in the same direction, except for the economic growth rate, in which the shocks are negative and the distribution is skewed to the left. We also found that the value of the kurtosis coefficient is high (greater than 3 for all variables), indicating a

tapered leptokurtic-type peak. Additionally, we noted that the largest difference between the maximum and minimum value of the variable is the exchange rate, which indicates the high volatility due to the devaluation of the currency with the high state of uncertainty after the start of the war in Syria. Also, one of the key characteristics of growth rates is that they are stable (i.e. they do not contain a unit root). Since structural changes affect expectations, as we have seen in Figure (1), there is a shift in the path of the variable due to political and economic events. We hence used the breakpoint unit root test proposed by Perron et al [49-50], where we assume that the structural change follows the course of innovation events and we test the largest structural breakpoint for each variable and get the following results:

$$y_t = c + \alpha t + \theta DU_t(T_b) + \gamma DT_t(T_b) + \omega D_t(T_b) + \delta y_{t-1} + \sum_{i=1}^{p} \varphi_i \Delta y_{t-i} + \varepsilon_t \qquad (20)$$

Where $\alpha, \theta, \gamma, \omega$ is break structural coefficients, DU_t, DT_t, D_t is dummy variable indicated on structural change, c: intercept, p: lag degree for AR model. Figure (4) shows the structural break estimate for the study variables. Structural break refers to a change in the behavior of a variable over time, shifting relationships between variables. Figure (4) also demonstrates that all macroeconomic variables have suffered from a structural break, but in different forms and at different times, although they show the impact of the same event, namely the war in Syria and the resulting economic blockade by Arab and Western countries, as all structural breaks that have occurred after 2011. We found that in terms of the rate of economic growth, it has been quickly influenced by many components and patterns. For EXR, CPI, GOP, the point of structural change came after 2019, the imposition of the Caesar Act and the Corona Pandemic in early 2020, which resulted in significant increases in these variables. We noted that the structural break in the Damascus Stock Exchange Index has occurred in late 2016 as the security situation in Syria improved,

resulting in restored investor confidence and realizing returns and gains for the market.

The step involves imposing the initial information on the structure of the model according to the properties of the data. Based on the results of the exploratory analysis, the prior Minnesota distributions were considered a baseline, and the main parameter is included λ in the hierarchical modeling:

Rho H = 0.2 is high frequency AR(1), Rho L = 0 is low frequency AR(1), Lambda = 5 is overall tightness, Upsilon HL = 0.5 is high-low frequency scale, Upsilon LH = 0.5 is low-high frequency scale, Kappa = 1 is exogenous tightness. C1 = 1 is residual covariance scale. The number of observations in the frequency conversion specifies the number of high frequency observations to use for each low frequency period. When dealing with monthly and quarterly data, we can specify that only two months from each quarter should be used. Last observations indicate that the last set of high-frequency observations from each low-frequency period should be used. Initial covariance gives the initial estimate of the

residual covariance matrix used to formulate the prior covariance matrix from an identity matrix, specifying the number of Gibbs sampler draws, the percentage of draws to discard as

burn-in, and the random number seed. The sample is divided into 90% training (in–of – sample) and 10% testing data (out-of–sample).
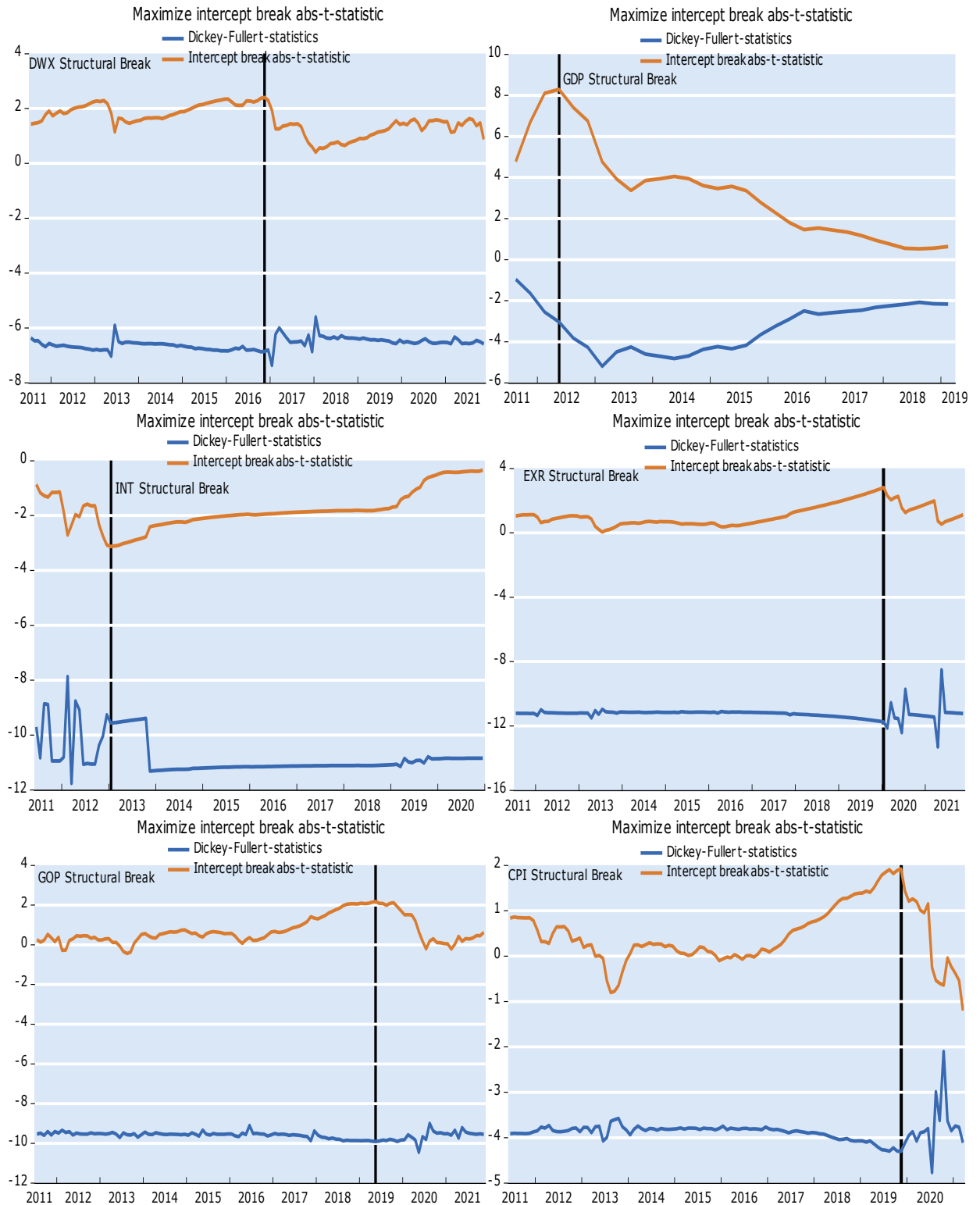
Figure 4. Structural Breaks of log high - low frequency study variables.

Table (3) provides us with the results of estimating the model to predict quarterly GDP in Syria. The results show the basic information to estimate the prior and posterior parameters, and the last section shows the statistics of the models such as the coefficient of determination and the F-statistic. We note that every two months were determined from a monthly variable to forecast each quarter of GDP in the BMFVAR model. Although there is no explanation due to the imposition of constraints, the model shows good prediction results with a standard error of less than 1 for each parameter and a high coefficient of determination of the GDP prediction equation explaining 93.6% of the variance in GDP.

Table 3. BMFVAR model estimation results

Mixed Frequency VAR Estimates
Sample: 2010Q3 2019Q1
Included observations: 35
Est type: Bayesian
Frequency conversion: Last
Initial covariance: Identity
Hyper-parameters: Lambda: 5, UpsilonHL: 0.5, UpsilonLH: 0.5, RhoH: 0.2, RhoL: 0,
Kappa: 1, C1: 1
10000 draws, 10% burnin, seed=1
Standard errors in ( )

	INT_2	INT_1	EXR_2	EXR_1	CPI_2	CPI_1	DWX_2	DWX_1	GOP_2	GOP_1	GDP
INT_2(-1)	-0.210395	-0.013337	0.005187	0.376769	0.022245	-0.123205	-0.035831	0.454396	-0.074473	-0.520494	-0.143971
	(0.16361)	(0.12096)	(0.18951)	(0.24014)	(0.08969)	(0.15894)	(0.27773)	(0.32058)	(0.29253)	(0.37566)	(0.05967)
INT_1(-1)	0.141927	-0.034539	-0.177912	0.131036	-0.055789	0.032988	0.003324	-0.156873	0.081991	-0.010867	0.016145
	(0.16636)	(0.17589)	(0.16582)	(0.27192)	(0.11273)	(0.22632)	(0.19133)	(0.32574)	(0.18992)	(0.35387)	(0.08674)
EXR_2(-1)	-0.119508	0.031038	0.140408	0.145814	0.118771	-0.149111	0.193608	0.557508	-0.147209	-0.531904	-0.023000
	(0.16137)	(0.12262)	(0.18085)	(0.23334)	(0.08952)	(0.16123)	(0.25818)	(0.30843)	(0.27007)	(0.35744)	(0.06027)
EXR_1(-1)	-0.017897	-0.103124	0.153804	-0.104820	0.239178	0.169465	0.001997	0.037449	0.039490	-0.071046	-0.011120
	(0.11119)	(0.08676)	(0.12076)	(0.16429)	(0.06153)	(0.11523)	(0.16255)	(0.22352)	(0.16542)	(0.26134)	(0.04232)
CPI_2(-1)	-0.087834	-0.015821	0.260409	0.030926	0.531593	-0.001680	0.020448	-0.074235	-0.088794	0.347960	0.105574
	(0.18388)	(0.13512)	(0.20441)	(0.26116)	(0.10000)	(0.17826)	(0.29192)	(0.34595)	(0.30341)	(0.39860)	(0.06625)
CPI_1(-1)	-0.004568	0.103836	-0.060392	0.374433	-0.198834	0.455771	0.173783	-0.129654	0.272139	0.150932	0.048595
	(0.11287)	(0.09163)	(0.12070)	(0.16406)	(0.06398)	(0.12059)	(0.16254)	(0.22094)	(0.16588)	(0.25869)	(0.04472)
DWX_2(-1)	0.020510	0.047563	-0.229432	-0.313470	-0.041321	-0.127081	-0.201227	0.044119	0.013937	-0.248122	0.054399
	(0.08338)	(0.05636)	(0.10835)	(0.13247)	(0.04546)	(0.07538)	(0.16647)	(0.17485)	(0.18351)	(0.21756)	(0.02775)
DWX_1(-1)	-0.214794	-0.062982	0.052018	0.207572	0.044297	-0.025643	0.307178	0.373465	-0.119579	-0.062903	-0.005737
	(0.08752)	(0.06372)	(0.09997)	(0.12989)	(0.04721)	(0.08486)	(0.14107)	(0.17607)	(0.14670)	(0.20978)	(0.03124)
GOP_2(-1)	0.025193	-0.030505	-0.120506	0.038249	-0.092885	-0.034359	-0.195787	-0.150071	-0.120841	0.092407	-0.030205
	(0.07746)	(0.05047)	(0.10220)	(0.12397)	(0.04175)	(0.06791)	(0.15809)	(0.16428)	(0.17467)	(0.20515)	(0.02485)
GOP_1(-1)	0.033009	0.014886	0.227830	0.099717	0.073085	0.127310	0.044275	-0.199702	0.092435	0.308335	0.002931
	(0.06381)	(0.04268)	(0.08020)	(0.09962)	(0.03370)	(0.05737)	(0.12035)	(0.13634)	(0.12865)	(0.16853)	(0.02100)
GDP(-1)	-0.020380	0.083859	0.015520	-0.060208	-0.073858	-0.310689	0.003122	0.002359	-0.003236	-0.003752	0.833066
	(0.07986)	(0.09863)	(0.04339)	(0.08302)	(0.05896)	(0.12188)	(0.03914)	(0.07616)	(0.03272)	(0.06490)	(0.04874)
C	0.883269	-0.172775	0.056207	-0.070877	0.174346	0.595522	-0.126913	0.165234	0.101768	0.140199	-0.698764
	(0.37950)	(0.29519)	(0.30641)	(0.34127)	(0.21336)	(0.36462)	(0.35382)	(0.35045)	(0.32483)	(0.33043)	(0.15035)
R-squared	0.230921	0.158642	0.325223	0.276615	0.669043	0.647527	0.196907	0.215540	0.111423	0.194525	0.936309
Adj. R-squared	-0.136899	-0.243746	0.002504	-0.069352	0.510759	0.478953	-0.187181	-0.159637	-0.313549	-0.190702	0.905847
Sum sq. resids	249.8841	91.01440	601.0143	839.7380	82.44195	170.9097	1549.055	1252.523	2120.300	2273.684	21.10105
S.E. equation	2.671993	1.612579	4.143892	4.898215	1.534759	2.209781	6.652721	5.982172	7.783316	8.059926	0.776458
F-statistic	0.955361	0.599948	1.533547	1.216692	6.432167	5.845305	0.780136	0.874241	0.398983	0.768420	46.77495
Mean dependent	0.394730	-0.173100	1.496440	2.461806	1.732614	3.382708	1.539342	1.582264	1.601329	2.609147	-2.219212
S.D. dependent	3.091323	1.783714	5.118267	5.843157	2.706754	3.776425	7.532011	6.852798	8.377445	9.111703	3.121563

Figure (5) also shows us the reliability of the prediction results. It is clear that the roots of the estimation of the parameters follow the inverse AR polynomial process. The inverted AR roots have modules very close to one which is typical of many macro time series models.

Table (4) shows the evolution of GDP Forecasting in Syria in and out the sample based on a number of indicators. When the value of these indicators is close to 0, the estimated values are close to the actual values, which is shown by Table 4, since the values of these indicators are all less than 1. The out-of-sample predicted values show a better performance of the model, and hence it can be adopted to track changes in quarterly GDP over time. Up-to-date with current versions.
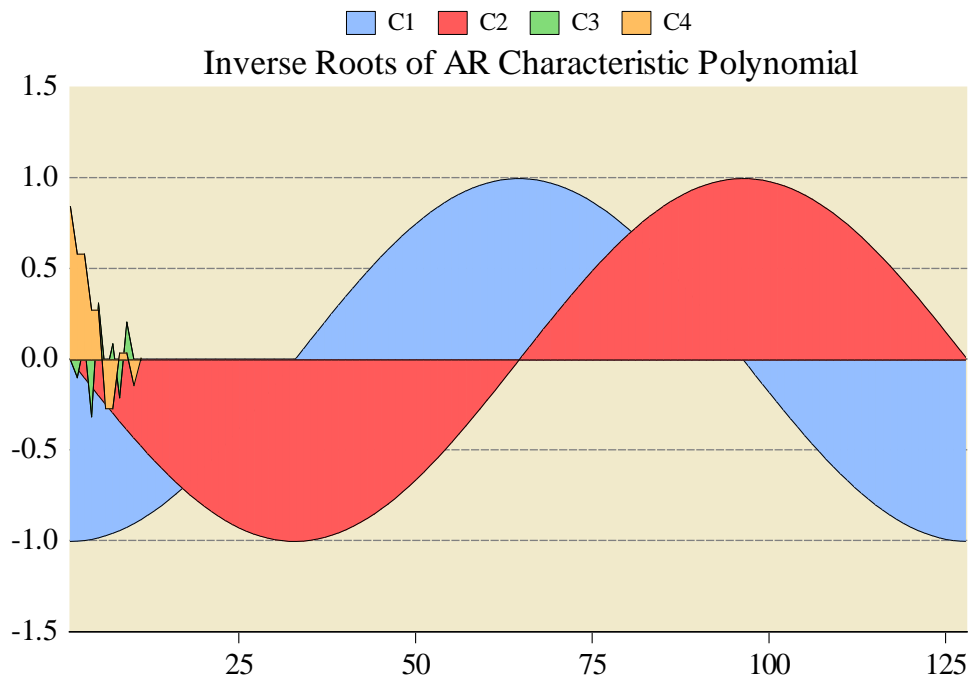
Figure 5. Inverse Roots of AR Characteristic polynomial

Table 4. Accuracy performance forecasting			
Indicators Performance	RMSE	MAE	Theil
In – of – Sample	0.763	0.575	0.104
Out – of - Sample	0.171	0.171	0.070

Through the data visualization technique, Figure (6) shows us the closeness of the expected values of quarterly GDP in Syria in of sample, which leads to the exclusion of the presence of an undue problem in the estimate (overfitting - underfitting). The median is used in the prediction because it is immune to the values of structural breaks and the data are not distributed according to the normal distribution. The out-of-sample forecast results (Fig 7) also indicate negative rates of quarterly GDP growth in Syria with the negative impact of internal and external shocks on the Syrian economy with the accumulation of the impact of sanctions, the lack of long-term production plans and the ineffectiveness of internal monetary policy tools, as the latest forecasts for GDP growth this quarter are (-3.69%).

The uncertainty at each point in time is included in the Syrian GDP projections, thereby achieving two objectives: incorporating the range of GDP change at each point in time and knowing the amount of error in the forecasts. We found that uncertainty increases as the forecast period lengthens (Figures 8-10).

The model also provides us with important results through scenario analysis. Figure 11 shows that the quarterly GDP in Syria is affected by the shocks of high-frequency variables, since this Figure shows that the shocks have negative impacts. This recommends a better activation of the instruments of The Central Bank and those responsible for monetary policy in Syria. These results are considered important tools for them to know and evaluate the effectiveness of their tools.

GDP
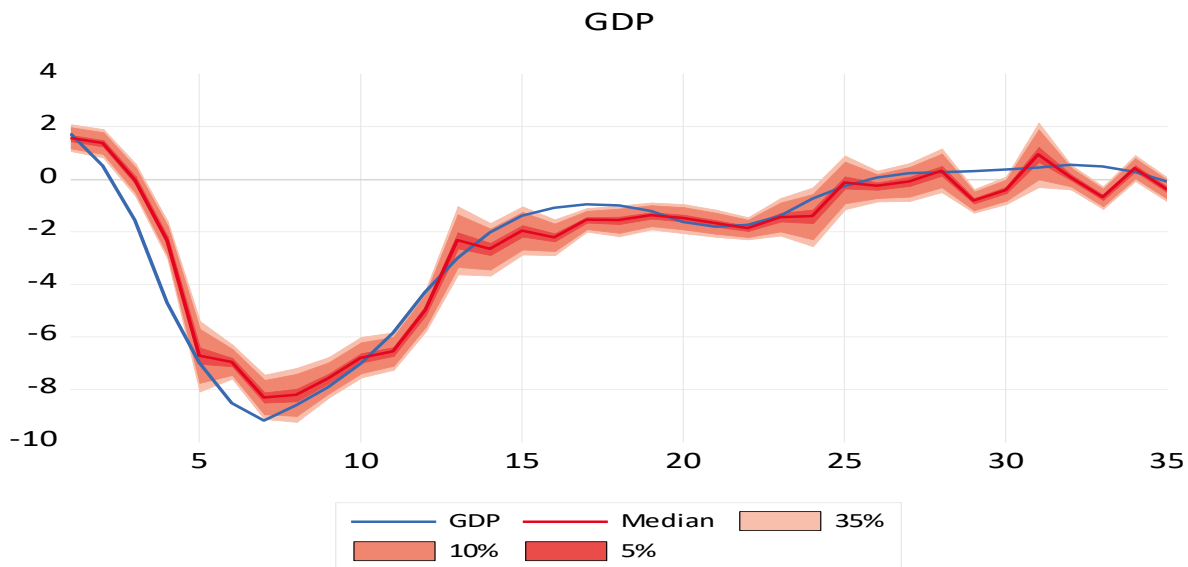
Figure 6. Forecasting quarterly GDP in Syria in of sample with 35%, 10%, 5% distribution quantities
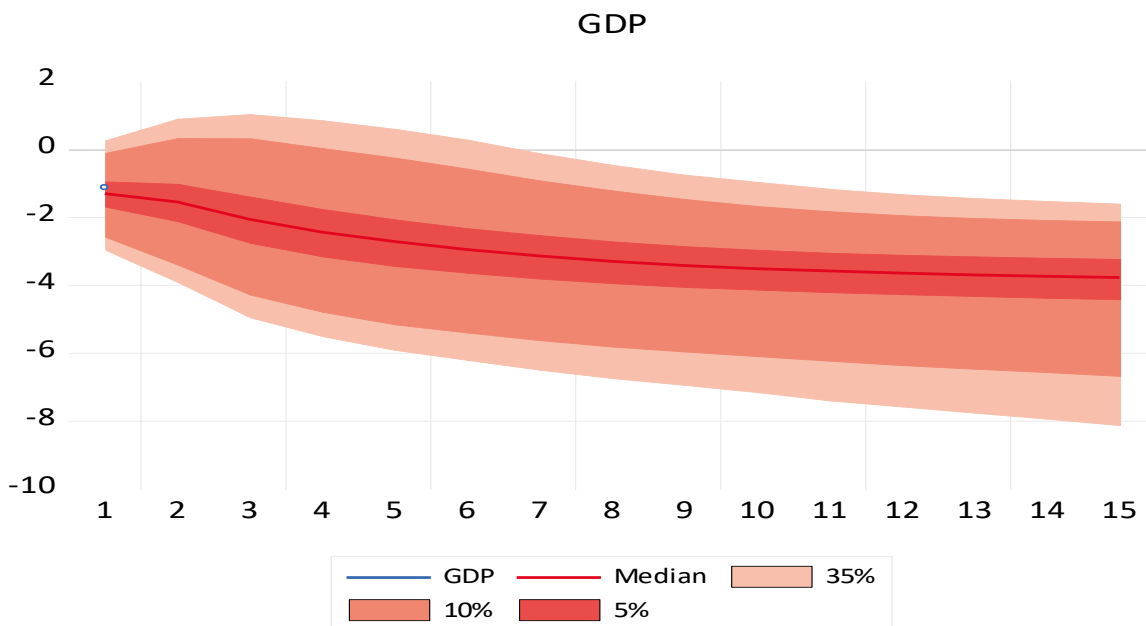
GDP

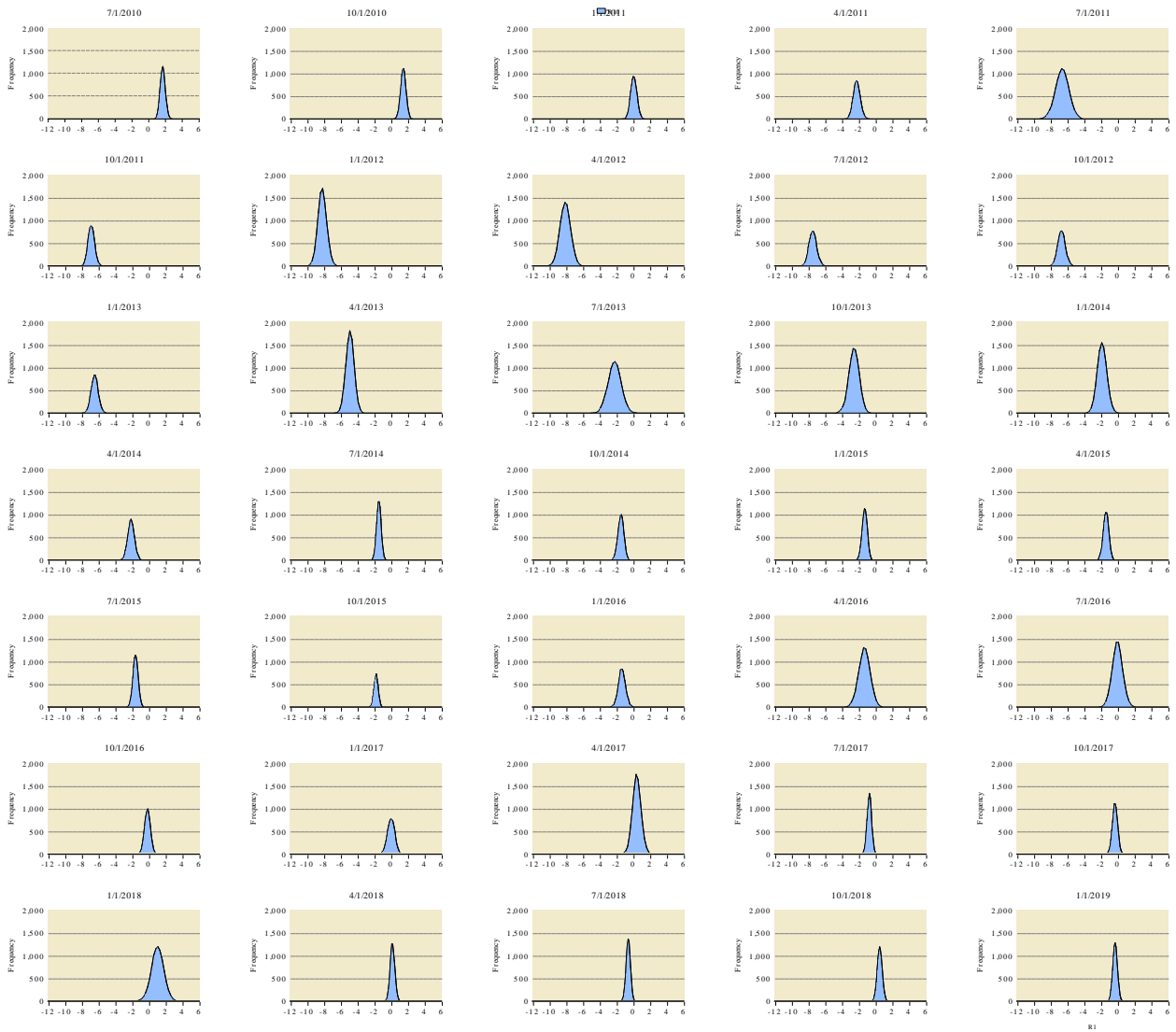**Figure 7. Forecasting quarterly GDP in Syria out of sample with 35%, 10%, 5% distribution quantities**

Figure 8. Uncertainty for forecasting quarterly GDP in Syria in-of –sample
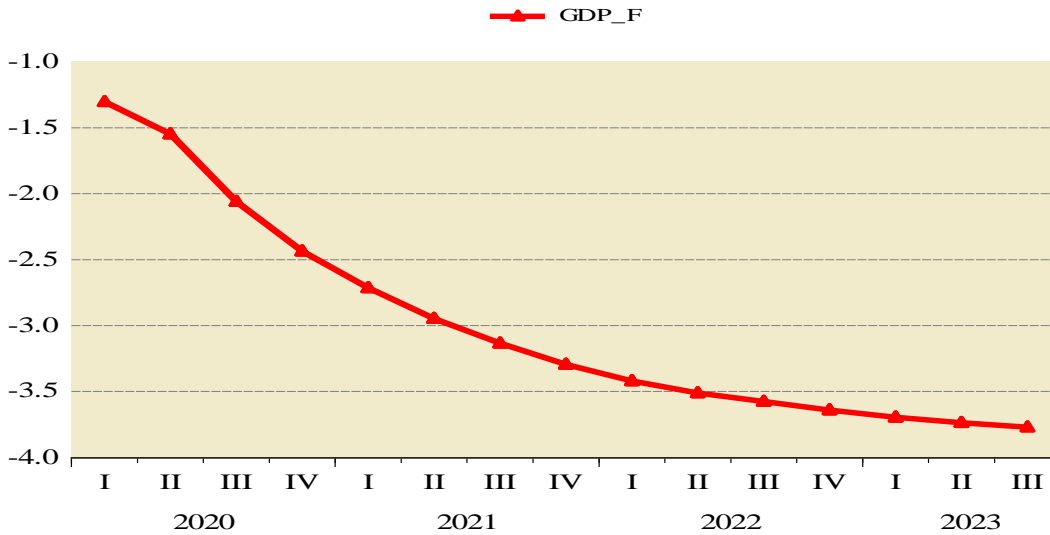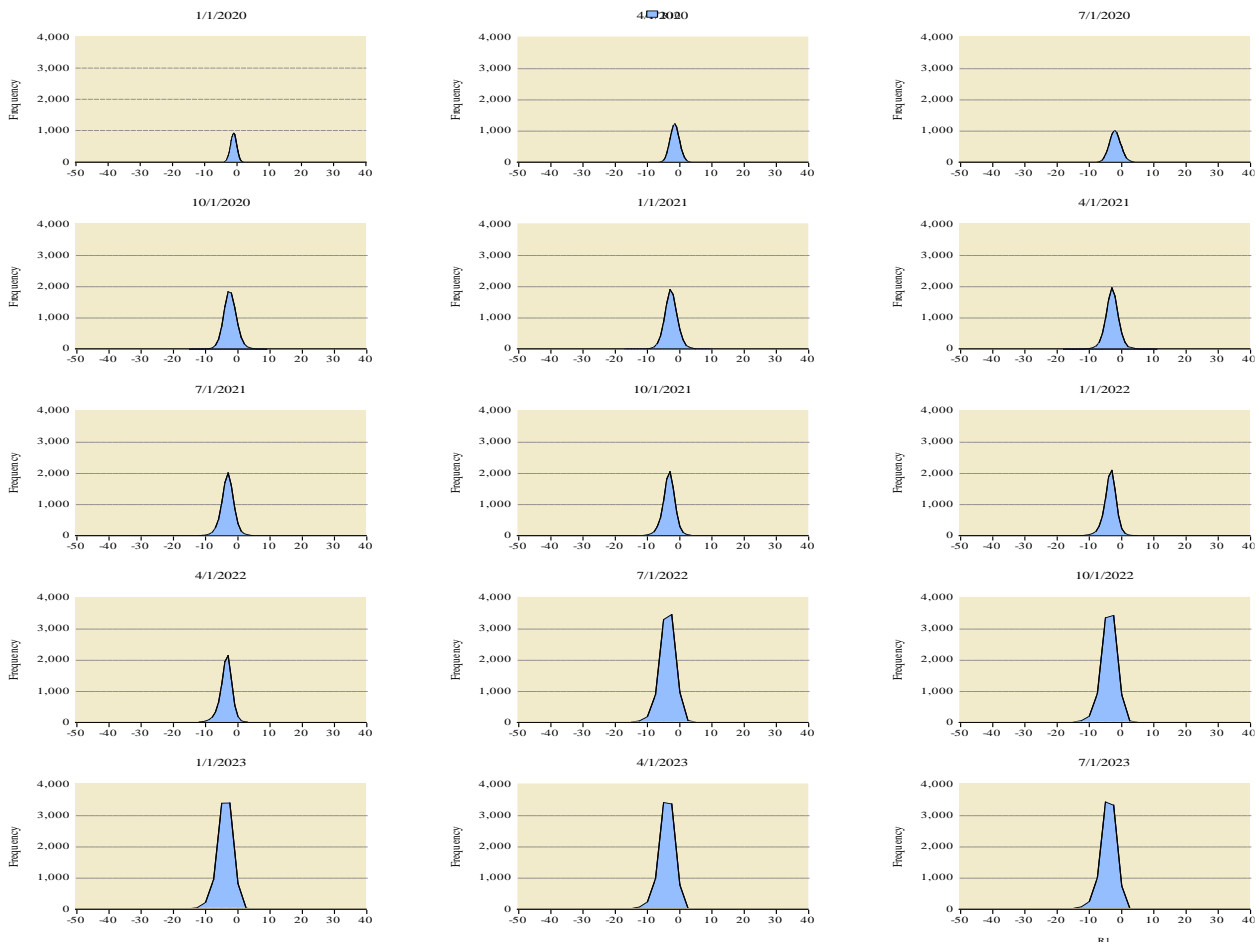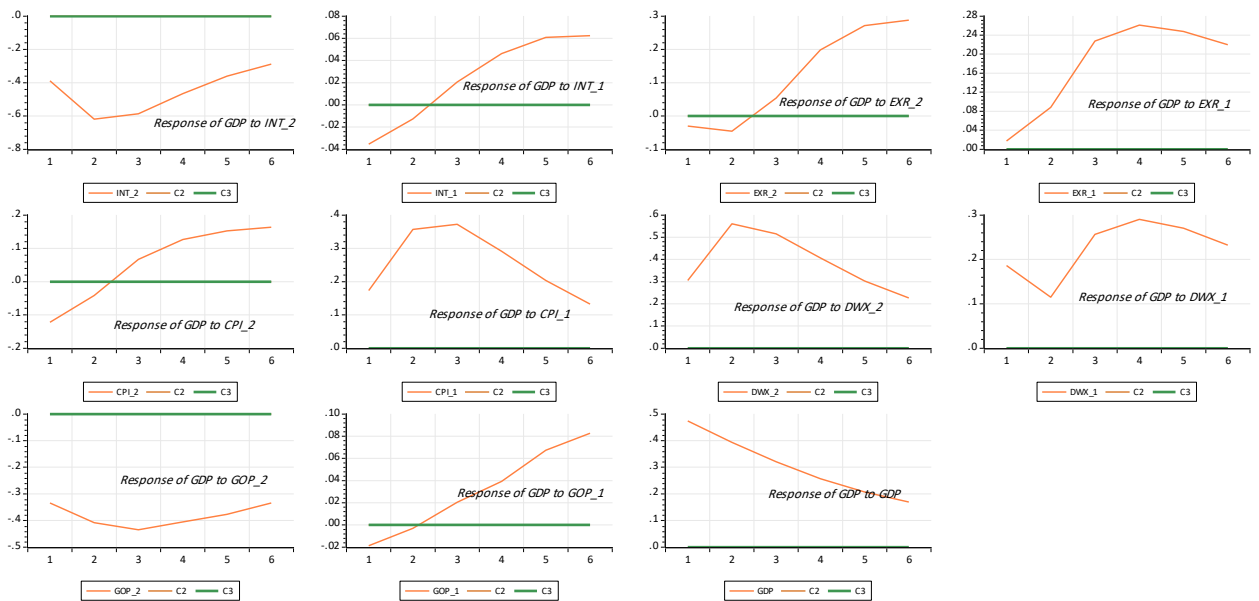

Figure 9. Centered Nowcasting quarterly GDP in Syria 2020q2-2023q1

Figure 10. Uncertainty for forecasting quarterly GDP in Syria out-of –sample

Figure 11. Shocks of high-frequency variables in the quarterly GDP of Syria

CONCLUSIONS AND RECOMMENDATIONS

This paper showed that BMFVAR could be successfully used to handle Parsimony Environment Data, i.e., a small set of macroeconomic time series with different frequencies, staggered release dates, and various other irregularities – for real-time nowcasting. BMFVAR are more tractable and have several other advantages compared to competing nowcasting methods, most notably Dynamic Factor Models. For example, they have general structures and do not assume that shocks affect all variables in the model at the same time. They require less modeling choices (e.g., related to the number of lags, the block-structure, etc.), and they do not require data to be made stationary. The research main finding was presenting three strategies for dealing with mixed-frequency in the context of VAR; First, a model – labelled "Chow-Lin's Litterman Method" – in which the low-frequency variable (Gross Domestic Product) is converted from annual to quarterly with the aim of reducing the forecast gap and tracking the changes in the GDP in Syria in more real-time and reducing the gap of high-frequency data that we want to predict usage. Second, the research adopts a methodology known as "blocking", which allows to treat higher frequency data as multiple lower-frequency variables. Third, the research uses the estimates of a standard low-frequency VAR to update a higher-frequency model. Our report refers to this latter approach as "Polynomial-Root BVAR". Based on a sample of real-time data from the beginning of 2010 to the end of the first quarter of 2023, the research shows how these models will have nowcasted Syria GDP growth. Our results suggests that these models have a good nowcasting performance. Finally, the research shows that mixed-frequency BVARs are also powerful tools for policy analysis, and can be used to evaluate the dynamic impact of shocks and construct scenarios. This increases the attractiveness of using them as tools to track economic activity for both the Central Bank of Syria and the Central Bureau of Statistics.

REFERENCES:

1. Giannone D, Reichlin L, Small D. Nowcasting: The real-time informational content of macroeconomic data. Journal of monetary economics. 2008; 55(4): 665-676.

2. Banbura M, Giannone D, Reichlin L. Nowcasting. ECB Working Paper No. 1275, Available at SSRN: https://ssrn.com/abstract=1717887. 2010;: 40.

3. Kuzin V, Marcellino M, Schumacher C. MIDAS vs. mixed-frequency VAR: Nowcasting GDP in the euro area. International Journal of Forecasting. 2011; 27(2): 529-542.

4. Ghysels E, Santa-Clara P, Valkanov R. The MIDAS touch: mixed data sampling regression models. University of North Carolina. Mimeo. 2004.

5. Ghysels E, Sinko A, Valkanov R. MIDAS regressions: further results and new directions. Econometric Reviews. 2007; 26: 53-90.

6. Zadrozny AP. Gaussian-likelihood of continuous-time ARMAX models when data are stocks and flows at different frequencies. Econometric Theory. 1988; 4: 108-124.

7. Mittnik S, Zadrozny AP. Forecasting German GDP at monthly frequency using monthly IFO business conditions data. In J.-E. Sturm, & T. Wollmershauser (Eds.). Ifo survey data in business cycle and monetary policy analysis. 2005;: 19-48.

8. Ferrara L, Guégan D, Rakotomarolahy P. GDP nowcasting with ragged-edge data: a semi-parametric modeling. Journal of Forecasting. 2010; 29(1-2): 186-199.

9. Schumacher C. A comparison of MIDAS and bridge equations. International Journal of Forecasting. 2016; 32(2): 257-270.

10. Babii A. High-dimensional mixed-frequency IV regression. ournal of Business & Economic Statistics 40. 2022; 4: 1470-1483.

11. Babii A, Ghysels E, Striaukas G. Machine learning time series regressions with an application to nowcasting. Journal of Business & Economic Statistics. 2021; 40(3): 1094-1106.

12. Chernis T, Sekkel R. A dynamic factor model for nowcasting Canadian GDP growth. Empirical Economics. 2017; 53: 217-234.

13. Anesti N, Galvão AB, Miranda-Agrippino S. Uncertain Kingdom: Nowcasting gross domestic product and its revisions. Journal of Applied Econometrics. 2022; 37(1): 42-62.

14. Eraslan S, Schröder M. Nowcasting GDP with a pool of factor models and a fast estimation algorithm. International Journal of Forecasting. 2022.

15. Hopp D. Economic nowcasting with long short-term memory artificial neural networks (LSTM). Journal of Official Statistics. 2022; 38(3): 847-873.

16. Longo L, Riccaboni M, Rungi A. A neural network ensemble approach for GDP forecasting. Journal of Economic Dynamics and Control. 2022; 134: 104278.

17. Dauphin MJF, Dybczak MK, Maneely M, Sanjani MT, Suphaphiphat MN, Wang Y, et al. Nowcasting GDP-A Scalable Approach Using DFM, Machine Learning and Novel Data, Applied to European Economies: International Monetary Fund; 2022.

18. Mansour Y, Ismail R, Alakkari K. Nowcasting As An Approach To Mitigate Economic Uncertainty In Syria. Tishreen University Journal- Economic and Legal Sciences Series. 2021; 42(6): 1-20.

19. Cimadomo J, Giannone D, Lenza M, Monti F, Sokol A. Nowcasting with large Bayesian vector autoregressions. Journal of Econometrics. 2022; 231(2): 500-519.

20. Crump RK, Eusepi S, Giannone D, Qian E, Sbordone AM. A large Bayesian VAR of the United States economy. FRB of New York Staff Report. 2021; 976.

21. Chow GC, Lin AL. Best linear unbiased interpolation, distribution, and extrapolation of time series by related series. The review of Economics and Statistics. 1971; 372-375.

22. Litterman RB. A random walk, Markov model for the distribution of time series. Journal of Business & Economic Statistics. 1983; 1(2): 169-173.

23. Verleysen M, François D. The curse of dimensionality in data mining and time series prediction. In In Computational Intelligence and Bioinspired Systems: 8th International Work-Conference on Artificial Neural Networks; 2005; Barcelona, Spain: Springer Berlin Heidelberg. p. 758-770.

24. Eisenstat E, Chan JC, Strachan RW. Stochastic model specification search for time-varying parameter VARs. Econometric Reviews. 2016; 35(8-10): 1638-1665.

25. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In " in IEEE Conf. Comput. Vis Pattern Recognit; 2016: IEEE. p. 770–778.

26. Zhang H, Zhang L, Jiang Y. Overfitting and underfitting analysis for deep learning based end-to-end communication systems. In In 2019 11th International Conference on Wireless Communications and Signal Processing (WCSP); 2019: IEEE. p. 1-6.

27. Sims CA. Macroeconomics and reality. Econometrica: journal of the Econometric Society. 1980;: 1-48.

28. Al-Akkari K, Ali B. Using Bayesian Inference for Stochastic Volatility Model (Empirical Evidence from Damascus Securities Exchange). Tishreen University Journal- Economic and Legal Sciences Series. 2022; 44(3): 11-32.

29. Alakkari K. Measuring Economic Policy Uncertainty in Syria Using Bayesian Stochastic Volatility Model. Magallat al-Tanmiyat wa-al-Siyasat al-Iqtisadiyyat. 2022; 24(2): 63-93.

30. Koop GM. Forecasting with medium and large Bayesian VARs. Journal of Applied Econometrics. 2013; 28(2): 177-203.

31. De Mol, C, Giannone D, Reichlin L. Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components? Journal of Econometrics. 2008; 146(2): 318-328.

32. Gelman A, Carlin JB, Stern HS, Rubin DB. Bayesian Data Analysis. Chapman & Hall/CRC Texts in Statistical Science. 2013; 3.

33. Villani M. Steady-state priors for vector autoregressions. Journal of Applied Econometrics. 2009; 24(4): 630-650.

34. Giannone D, Lenza M, Primiceri GE. Prior selection for vector autoregressions. Review of Economics and Statistics. 2015; 97(2): 436-451.

35. Litterman RB. Bayesian procedure for forecasting with vector autoregressions. Massachusetts Institute of Technology. 1980.

36. Sims CA. A nine-variable probabilistic macroeconomic forecasting model. In Business cycles, indicators, and forecasting. 1993; 179-212.

37. Doan T, Litterman R, Sims C. Forecasting and conditional projection using realistic prior distributions. Econometric reviews. 1984; 3(1): 1-100.

38. **Sims CA, Zha T. Bayesian methods for dynamic multivariate models. International Economic Review. 1998; 949-968.**

39. **Higham NJ. Analysis of the Cholesky decomposition of a semi-definite matrix. 1990; 161-185.**

40. **AL-Akkari K, Ali B. Forecasting Macroeconomic Data in Syria with Hierarchical Bayesian Structural VAR Model. Tartous University Journal for Research and Scientific Studies - Economic and Legal Sciences Series. 2022; 6(7): 1-25.**

41. **Central Bureau Of Statistics. Statistical Abstract. Damascus:; 2021.**

42. **Central Bureau Of Statistics. Consumer price index CPI. Damascus; 2022.**

43. **Central Bank of Syria. market interest rates. Damascus; 2022.**

44. **Damascus Stock Exchange. DSE Weighted Index (DWX). Damascus; 2023.**

45. **Central Bank of Syria. Remittance and exchange bulletin. Damascus; 2023.**

46. **Craftsmanship Association. The price of a gram of 21 karat gold. Damascus; 2023.**

47. **Mandelbrot BB. The variation of certain speculative prices. Springer New York. 1997;: 371-418.**

48. **Fama EF. The behavior of stock-market prices. The journal of Business. 1965; 38(1): 34-105.**

49. **Perron P, Vogelsang TJ. Testing for a unit root in a time series with a changing mean: corrections and extensions. Journal of business & economic statistics. 1992; 10(4): 467-470.**

50. **Perron P. Dealing with structural breaks. Palgrave handbook of econometrics. 2006; 1(2): 278-352.**

Competing interests: The Author declare that they have no competing interests.

Data and materials availability: All data are available in the cited official websites, and attached to an excel file with article.