Vietnamese Export Growth Prediction Applying MIDAS and MF-VAR on Mixed-Frequency Data

Nguyen Thi Hien[1], Hoang Anh Tuan[2], Dinh Thi Ha[3](✉), Le Mai Trang[2], Tran Kim Anh[2], and Dao The Son[2]

[1] Mathematics Department, Thuongmai University, Hanoi, Vietnam
hiennguyen@tmu.edu.vn
[2] Economics Department, Thuongmai University, Hanoi, Vietnam
{hoanganhtuan, lmtrang2000, trankimanh, daotheson}@tmu.edu.vn
[3] Informatics Department, Thuongmai University, Hanoi, Vietnam
dinhha@tmu.edu.vn

Abstract. Import and export growth forecasting is always a concern for researchers as well as policymakers in any country. However, forecasting this index through methods that employ data sets with the same frequency does not reflect reality due to the fact that the economic and financial indicators have different published frequencies. Therefore, between two data periods, some important information affecting the target variable maybe not be included in the model, leading to the limited accuracy of the forecast. The approach of data analysis with multiple frequencies has gained a lot of interest recently in order to overcome data restrictions and increase forecasting performance. In this article, we study and apply a number of models for mixed-frequency data such as MF-VAR and MIDAS to predict Vietnam's exports based on collected data sets in the period from 2006 to 2020. The findings show that the MF-VAR model has high forecasting error and is not suitable for predicting the export growth of Vietnam, while the MIDAS model gives good prediction results on the same data set. In addition, the prediction results also reveal that the MIDAS model is effective for short-term forecasting. This result is quite similar to that of several previously published studies. The findings of this research also open a promising direction in studying and applying the mixed frequency data models to forecast export growth as well as other economic indicators .

Keywords: Export growth · Forecasting · Mixed-frequency data · MF-VAR · MIDAS

1 Introduction

Trade is a key component of GDP for almost any country. That is particularly true for an export-oriented economy like Vietnam. Trade as a percentage of GDP for Vietnam has increased steadily for the last four decades, starting with 23.3% in 1986 to 209.3% in 2020. Last year, the export of goods and services from Vietnam reached 287.8

P. Cong Vinh and N. Huu Nhan (Eds.): ICTCC 2021, LNICST 408, pp. 1–19, 2021.
https://doi.org/10.1007/978-3-030-92942-8_1

billion USD or the equivalent of 106.1% GDP, while that number for import is 279.8 billion USD or 103.2% GDP (Nguyen et al. 2020). This makes trade an essential part of GDP forecasting for the purpose of macroeconomic policies.

There are numerous swings that accompany the growing level of economic openness and integration. According to trade reports, Vietnam's situation of exports and imports can face uncertainties from many domestic and international factors. Therefore, an accurate forecast of export growth will help policymakers build national economic growth scenarios, as well as help countries be more proactive in their foreign exchange reserves. Besides, the correct and timely forecast of export growth also helps researchers and policymakers to forecast shocks to the economy, because exports have a great influence on the business cycle.

For any macroeconomic forecast, there is a need to have the most precise and most timely estimate possible. This goal does have a common challenge from macroeconomic data when reports contain data with different frequencies. They are not officially made available at the same time, or in many cases, there can be lags in reporting.

This study aims to investigate the performance of Mixed-Frequency models on forecasting export growth in Vietnam. These models create possibilities to utilize both low-frequency and high-frequency data, to combine lag reports with real-time data to give accuracy and timely forecast for Vietnamese export MIDAS and MF-VAR stay in the focusing point, and its parameters are modified to capture the fluctuation of data and to improve forecasting accuracy. The study collects data from different public sources on the supply and demand sides of exports. Data is then used to test the accuracy of different mixed-frequency forecasting models by comparing predicted values with real export growths. This experimental work on export data can serve as a preliminary study for further applications on other essential macroeconomic indicators for the purpose of monitoring and making policies.

The remainder of this paper is structured as follows: Sect. 2 gives an overview of the literature on export forecasting techniques with both types of models, using the same frequency and mixed-frequency data. Section 3 describes the data collection and preprocessing processes. Section 4 presents experiment results with mixed-frequency forecasting models on export data. Section 5 closes with a discussion on the results and future works.

2 Related Works

The forecasting microeconomic indicators in general and export, in particular, are very interested and studied by many researchers. In which most studies use data with the same frequency with many different methods such as the gravity model, export demand model, Bayesian regression model, etc. and machine learning methods such as Neural networks, fuzzy set theory, genetic algorithms.

From the perspective of international trade, many export forecasting models approach from classical to modern theory. The classical theoretical models for export forecasting include the Ricardian model, the technology-based model, and the Heckscher-Ohlin resource factor model (Bussière et al. 2005). The endogenous growth and gravity models have also contributed to theoretical advances along with the theory of competitive advantage. In these models, factors such as foreign trade return to scale, monopolistic competition, diversity or market failure are considered to be included in the forecasting model (Feenstra 2015). Mehta and Mathur (2003) reviewed existing models for short-term forecasting of Indian exports in which export turnover depends on the import demand for domestically produced products of the partner countries, on the exchange rate and export price index and their lagged variables. Bussière et al. (2005) analysed the rapid trade integration of the Central and Eastern European countries (CEECs) with the euro area since 1980 and drew implications for further integration by using as a benchmark an enhanced gravity model. In this study, the results of the fixed-effects estimator and the dynamic OLS specification indicated that the economic size of neighbouring countries has a highly significant and less than the proportional impact on bilateral trade, the real exchange rate variables also enter the regression significantly. Similarly, having a common border and speaking the same language increase trade between the two countries, while the common territory dummy is not related to total trade. The authors also showed that the predicted trade values derived from the gravity model may be biased if not taking into account adjustments to standard trade conditions after the opening-up of Eastern Europe, which may translate into distorted estimates for the fixed effects.

Recently, according to research by Siggel (2006), one of the most commonly used models to forecast export turnover is the export demand model. This model assumes that supply is infinitely elastic, i.e. when demand exists, any supply can be produced. In the export demand model, most variables such as exchange rate, price index and the relative price of exports are used, in which relative price is one of the very important factors determining the competitiveness of export activities and comparative advantage. Stoevsky (2009) used the export demand model to forecast the total exports of Bulgaria, where the predictors in the model include the export price index of goods and services (nominal or real), the exchange rate (nominal or real), a composite index that measures external demand for domestically produced products, the price vectors of the economy's key exports, and their lagging variables. Lehmann (2015) forecasted the import-export turnover index of most European countries (including 18 Eastern and Central European countries, 28 EU countries) to study the economies of these countries. Survey-based variables (called soft variables) such as business confidence index, consumer confidence index, production expectation index, etc., are included in the import and export demand model to see if the forecast error is lower when compared with the model that includes only hard variables such as price variables and price indexes, aggregate demand variables, exchange rates, etc. However, the results in most countries show that adding a soft variable to the import and export forecasting model using the demand model method gives a much lower forecast. In fact, there are many other factors (or variables) that also affect the economy's exports, such as national production capacity, the skills of employees, inventories and credit balances of manufacturing industries, domestic and international political-economic situation, etc.

With the Bayesian regression model, Eckert et al. (2019) used a comprehensive data set containing Swiss exports to forecast exports. The all-time series covered the period from 1988 to 2018 with monthly frequency and was calculated in Swiss francs, they were not seasonally adjusted. In order to check the accuracy of all forecasts, the authors computed the log of the RMS forecasting error of the base forecasts relative to the MSE of the coherent forecasts from each method. The results showed values above zero indicate a better forecast performance. This also provided strong evidence that in addition to making coherent predictions, the Bayesian estimate helped to improve forecast accuracy.

In the approach using machine learning methods, Urrutia et al. (2019) forecasted Philippine imports and exports using Bayesian Artificial Neural Network and ARIMA model regression. In this study, data were collected from the Philippine Statistics Authority with a total of 100 observations in order to forecast values of imports and exports from the first quarter of the year 2018 to the fourth quarter of the year 2022 using the most fitted model. By computing and comparing the MSE, RMSE, NMSE, MAE, and MAPE of each model, the researcher deduced the fitted model that can use in forecasting the imports and exports of the Philippines is the Bayesian Artificial Neural Network. Upon using the Paired T-test, the p-value for the imports is 0.055 and for the exports is 0.054 which means that there is no significant difference between actual and predicted values for both imports and exports of the Philippines. This study could help the economy of the Philippines by considering the forecasted Imports and Exports which can be used in analyzing the economy's trade deficit.

Qu et al. (2019) forecasted import and export trends of Shandong province, China using the LSTM model (Long short term memory neural network). The study used monthly import and export data of Shandong province during the period from January 2001 to June 2018. Monthly import and export data often have large-scale, nonlinear and difficult to fit, so it is difficult to accurately forecast their trends. To solve this problem, the researchers proposed to use the LSTM model in combination with the cubic exponential smoothing method of the time series model. The results showed that the LSTM method gives a better result than the traditional time series regression model. Meanwhile, in order to forecast the demand of Chinese exports, Bin and Tianli (2020) based on artificial neural network theory, the theory of fuzzy systems and established forecasting models. Based on the system of export forecasting criteria determined on the basis of analyzing the research results of foreign trade export, the authors chose the evolutionary morphological neural network model to apply the export prediction and considered the influence of different factors, the obtained prediction results were ideal. On the other hand, Xie and Xie (2009) forecasted the total volume of import and export trade based on Gray forecast modelling (GM) which are optimized by a genetic algorithm. This paper introduced a combination prediction method based on fully utilize the advantages of the GM model and the characteristics of the genetic algorithm to find the global optimal combination, so the predictive model is more accurate. The results indicated that the model can be used as an effective tool to forecast the total volume of import and export trade.

For the above 3 studies, to forecast the total amount of a province's import-export trade, the authors performed on the data set with the same frequency, namely by month (in the study of Qu et al. (2019) or by year (as in the study of Xie and Xie (2009) and Bin and Tianli (2020).

Similarly, also with the same frequency of data by month, in 2020, Hai et al. built a model to forecast export turnover using the dimensional reduction method based on the Kernel trick. Different from traditional forecasting models, the study developed the model based on a large data set of 144 variables collected from various sources during the period from January 2013 to June 2019. By using a factor bridging equation model, factors were extracted from a large data set using a combination of attribute selection and attribute learning methods. The attribute selection method was used to remove the redundant and noisy information in the original data set while the attribute learning method was used in combination to extract the factors in the original data set after removing the noise or redundancy information. The test forecast results of the model built on the real data set of Vietnam showed that the out-of-sample prediction accuracy is about less than 3% compared with reality.

Trending of using different frequency data on the prediction of macroeconomic indicators.

The characteristics of macroeconomic indicators are that they are often published with different frequencies, some are published annually, but others are published quarterly, monthly, or daily. In addition to that, macroeconomic indicators often have different published times, making it difficult to exploit data for forecasting. To deal with those issues, one of the macroeconomic forecasting methods in recent years, which has been studied and applied by many scientists around the world, is models with variables collected and published with different frequencies. The research results show that this analytical method has the great advantage of optimally exploiting published data at different frequencies, possibly by day, month, quarter, or year. Some of the most popular models for analyzing data with mixed frequencies including the mixed data sampling (MIDAS) model, the bridge equations, and the mixed frequency VAR (MF-VAR) model. However, published studies have only focused mainly on applying mixed frequency data models for forecasting GDP, the indexes reflecting the growth of economies, and forecasting inflation without many forecasts for other important macro indicators such as exports. Meanwhile, studies in Vietnam have hardly mentioned the model of using datasets with mixed frequency in forecasting macroeconomic variables in general and exports in particular. Using mixed-frequency models for export forecasting is a large research gap. Therefore, this study investigates the performance of models that work on mixed frequencies data for the export forecasting problem.

3 Research Models and Data

3.1 Mixed-Frequency Models

Mixed frequency models are proposed with the aim of exploiting the availability of data at different frequencies and overcoming the disadvantages of the traditional VAR model when building a predictive model on a mixed frequency data set based on the VAR structure. One of the highest accuracy forecasting methods on datasets with different frequencies is the Kalman filter thanks to its ability to estimate the past, present, and future. However, it requires a lot of calculations, this method is not widely applied in the field of financial economics but instead in the field of automatic control. Besides, there are other popular mixed frequency methods such as the bridge equation model, the group of mixed data sampling model (MIDAS), and the mixed frequency VAR model (MF-VAR). These methods belong to the group of dynamic factor models and are applied in many fields of finance and economics (Kuzin et al. 2011; Foroni and

Marcellino 2013). In which, the bridge equation model and the MIDAS model are methods that approach single equations, i.e. have only one dependent variable, and MF-VAR is the model that approaches the system of equations in which all variables are dependent variables. Therefore, MF-VAR is often limited by the dimensional curse/ the curse of dimensionality, and determining the optimal delay of MF-VAR is often not as accurate as other methods. In this article, we study and apply two models MF-VAR and MIDAS to forecast Vietnam's exports based on collected data sets in the period from 2006 to 2020.

- *Mixed-frequency VAR model (MF-VAR)*

The MF-VAR model is a VAR model that analyzes time series datasets with mixed frequencies. This model is based on the state space approach and was proposed by Mariani and Murasawa in 2010 (Foroni and Marcellino 2013). In this model, to process data with different frequencies, the variables with low frequency is considered as a high-frequency variable with missing observations. Then, the low-frequency variables will be applied techniques to transform to the same frequency of the highest frequency variables in the model (for example, the Kalman filter is applied to estimates missing observations). MF-VAR usually has many parameters that are estimated by two following methods. The first method basing on the classical approach to estimate by using the maximum-likelihood technique while the second method is to use Bayesian inference with the expectation-maximization algorithm.

According to Mariano and Murasawa (2010), with the classical approach, when considering the VAR model in the state space, the disaggregation of the variable with the quarterly frequency y_{t_m} observed at time $t_m = 3, 6, 9, \ldots .T^m$ into the unobserved monthly variable $y^*_{t_m}$ is described by the following aggregation equation:

$$y_{t_m} = \frac{1}{3} y^*_{t_m} + \frac{2}{3} y^*_{t_m - 1} + y^*_{t_m - 2} + \frac{2}{3} y^*_{t_m - 3} + \frac{1}{3} y^*_{t_m - 4}$$

This equation comes from the assumption that the series y_{t_m} is the average of the series $y^*_{t_m}$ with its lagged variables. Then, the process VAR(p) is represented as the following equation:

$$\phi(L_m) \begin{pmatrix} y^*_{t_m} & \mu^*_y \\ x_{t_m} & \mu_y \end{pmatrix} = u_{t_m}$$

Where

x_{t_m} and $y^*_{t_m}$ are variables with high frequency.
u_{t_m} is a random error with zero expectation and constant variance.

The parameters in the model are interpreted and estimated based on the state-space model with the maximum-likelihood estimation. According to Vladimir Kuzin et al. (2011), MF-VAR is more effective in long-term forecasting than nowcasting.

- *Mixed-data Sampling Model MIDAS*

MIDAS model is an analysis model with mixed frequency data, proposed by Eric Ghysels, Arthur Sinko & Rossen Valkanov in 2002, in which the explanatory variables have different frequencies that are equal or higher than the frequency of the dependent variable. In this model, to the explanatory variables with higher frequency, lagging polynomials are used to prevent an increase in the number of parameters as well as problems related to lag selection of the independent variables in the model.

The basic MIDAS model for a single explanatory variable and h_q- step- ahead forecasting, with $h_q = h_m/m$ is defined by:

$$y_{t_q+h_q} = y_{t_m+h_m} = \beta_0 + \beta_1 b(L_m;\theta)x^{(m)}_{t_m+w} + \varepsilon_{t_m+h_m}$$

In which:

y_{t_q} is a dependent variable with low frequency and x_{t_m} is an explanatory variable with high frequency.
t_q is the time index, t_q = 1, 2, …, T^y_q, in which T^y_q is the final time that y_{t_q} have available data and $t_q + h_q$ is the time to be forecast.
t_m is the time index with higher frequency, t_m = 1, 2, …, T^x_m, in which T^x_m is the final time that x_{t_m} have available data.
m is the index to defines a higher degree of frequency of the independent variable than the dependent variable. For example, if y has a quarterly frequency and x has a monthly frequency, then m = 3, and if y has a quarterly frequency and x has a weekly frequency, then m = 12.
w is the number of time periods based on high frequency for which the explanatory variable has data available earlier than the dependent variable.
$b(L_m;\theta) = \sum_{k=0}^{K} c(k;\theta)L^k_m$ is lag polynomials with L_m is the lag operator that is defined by:
$L^k_m x^{(m)}_{t_m} = x^{(m)}_{t_m-x}.x^{(m)}_{t_m+w}$ is skip-sampled from the high-frequency indicator x_{t_m}.
$c(k;\theta)$ are the parameters with lagged variables of the independent variable to be estimated, where k is the lag index.
β_0, β_1 are the coefficients of the regression model.
ε is the error of the model.

One of the main issues of the MIDAS method is finding a suitable parameterization for the lagging coefficients $c(k;\theta)$ to prevent the parameter increase in the model, avoid the situation that too many parameters cause difficulty for model estimation. This problem is due to x_{t_m} has a higher frequency than y_{t_q}. Several weight schemes have been developed to overcome this problem, the most popular of which is the Almon polynomial, also known as the "Exponential Almon Lag". Specifically, the Almon scheme is expressed as follows:

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

where Q is the number of parameters of θ, or $\theta = (\theta_1, \theta_2, \ldots, \theta_Q)$ are the parameters to be estimated. This function is quite flexible and can take various shapes with only a few parameters. Ghysel et al. (2004) applied this functional form with two parameters, that allows great flexibility and defines how many lags are included in the regression.

3.2 Research Data

- *Variable selection*

Based on international experience, the characteristics of export growth in Vietnam and the availability of the database, this research selects the variables for the model as shown in Tables 1, 2 and 3. For the low-frequency dependent variable, the research uses the quarterly export growth.

Variables representing the supply side include GDP growth (denoted by GDP_G), Net Foreign Direct Investment (NFDI), Industrial Production Growth Rate (IP_G); Variables representing the demand side include Retail sales growth (denoted by RS_G), Commodity import growth (denoted by CI_G), inflation rate (denoted by CPI); variables representing the export competitiveness include export commodity prices including exchange rate (current exchange rate and denoted by XR), weekly exchange rate (XR_W); The remaining variables represent the monetary sector, banking, stock market, futures contracts including Balance of payments (denoted by BOP), Capital account (denoted by KA), Money supply growth (denoted by MS_G), Reserve Ratio (denoted by RR), Composite Stock Price Index (denoted by CSPI), Interest Rate (denoted by IR), Total International Reserves (denoted by TIR), Balance of Trade (denoted by BOT). The inclusion of monetary variables also has important explanatory implications because Vietnam's import-export enterprises in general are highly dependent on bank capital. Large fluctuations in the stock market and futures market have a significant impact on export activities.

To forecast the quarterly export growth rate, the study uses a dataset of 22 variables with 19 economic indicators, of which: 5 quarterly frequency variables, 14 monthly frequency variables and 3 weekly frequency variables. They are described in detail in the following tables:

Table 1. The quarterly-frequency economics indicators

Variable	Unit	Indicator
EX_G	(%)	Export growth
GDP_G	(%)	GDP growth
BOP	(%)	Balance of payments
NFDI	(%)	Net foreign direct investment
KA	(%)	Capital account

Table 2. The monthly-frequency economics indicators

Variable	Unit	Economics indicator	Variable	Unit	Economics indicator
RS_G	(%)	Retail sales growth	**TIR**	(%)	Total international reserves (excluding gold)
IP_G	(%)	Industrial production growth	**CI_G**	(%)	Commodity import growth
MS_G	(%)	Money supply growth	**BOT**	(%)	Balance of trade
RR	(%)	Reserve ratio	**CPI**	(%)	Inflation rate
CSPI	Index	Composite stock price index	**XR_M**	USD/VND	Monthly exchange rate (average)
IR	(%)	Interest rate	**G_M**	USD/OUNCE	The future contract of gold
XR	Index	Exchange rate	**O_M**	USD/Barrel	The future contract of raw oil

Table 3. The weekly-frequency economics indicators

Variable	Unit	Economics indicator
XR_W	USD/VND	Weekly exchage rate
G_W	USD/OUNCE	The future contract of gold
O_W	USD/ Barrel	The future contract of raw oil

- *Data collection*

The research is carried out on a dataset of macroeconomic indicators, collected with different frequencies (quarterly, monthly, weekly) from the websites of the General Statistics Office, the IMF, the World Bank, ADB, Bloomberg. The data preprocessing and result representation is performed on Excel, while the analysis and forecasting with models are done on Eviews 11.

- *Data characteristics and preprocessing*

Due to the lag in data publication, there are differences in the number of observations obtained of the indicators in the last year of the research period, 3 data series are missing information (Money supply growth, Reserve ratio, total international reserves excluding gold).

Some variables do not have published data by month, by quarter but have published data at a higher frequency such as day, week, month. Therefore, in this study, high-frequency data were used to generate lower frequency data for those variables. The low-frequency variables are calculated as the average of higher frequency variables.

For example, Vietnam's export growth in the first quarter of 2020 will be taken as the average of Vietnam's export growth data for January, February, and March of 2020.

After collecting and pre-processing, the dataset was divided into 2 datasets of 2 different stages. In which, the data set in the period 2006 to 2018 is used to estimate the parameters in the regression models; the data set from 2019 to 2020 is used to forecasts.

- *Missing data issue*

In the 22 mentioned variables, 3 variables have missing data at the end of the research period, including Money supply growth, Reserve ratio and Total international reserves excluding gold. Typically, Money supply growth, Reserve ratio misses data of 12 months in 2020 and Total international reserves excluding gold misses data of 3 months of the fourth quarter of 2020. To overcome the issue of missing data, this research makes predictions for the missing values. Before forecasting for missing values and forecasting monthly export growth for Vietnam, the authors considered the correlation between independent variables of the same frequency in the three groups by the Pearson correlation coefficient and checked the stationarity of the data series by the unit root test method. The results showed that the data series are stationary, there are some series that are not stationary but overcome by taking the first-order difference.

- *Forecast for missing data (monthly frequency)*

The ARIMA model is one of the most effective short-term forecasting models for macroeconomic indicators. Therefore, this study applied the ARIMA (p, d, q) model to predict for 3 series with missing data. After considering the correlation schema of 3 series MS_G, RR, TIR, combined with the results of testing the stationarity of the series, the orders of AR (p), MA (q) and the order of difference (d) for stationary series have determined. Then the study regressed the different ARIMA models based on p, d, q levels and compared the models according to AIC and SIC standards to find the most suitable model. The results obtained are as follows (Table 4):

Table 4. The selection result of the ARIMA model

	MS_G	RR	TIR
P	2	1	1
D	1	1	1
Q	4	2	2
ARIMA (selected)	**(2, 1, 0)**	**(1, 1, 0)**	**(1, 1, 0)**
MAE	4.223	2.109	8.899
RMSE	5.835	2.577	11.183

(Source: The analysis result of authors)

4 Forecast Results

4.1 Overview of Real Quarterly Export Growth of Vietnam in the Period 2006–2020

From 2006 to now is the period of extensive economic integration with the important event that Vietnam became the 150th member of the WTO. As a result, in line with prior foreign trade policies and strategies, Vietnam's orientation is to continue leveraging its relative advantages to exploit present export markets while also developing new markets to support economic restructuring[1]. Shifting the export structure in favour of encouraging the export of high-value-added products, processed and manufactured goods, items with advanced technologies and brain content, while gradually lowering raw exports.

Figure 1 depicts the outcomes of Vietnam's quarterly export growth over this time period (Table 5).

Table 5. Descriptive statistics on quarterly export growth of Vietnam in the period 2006–2020

Observations	Mean	Median	Maximum	Minimum	Std. Dev
60	16.77	15.39	40.85	−19.6	12.08

(Source: data analysis results of the authors)

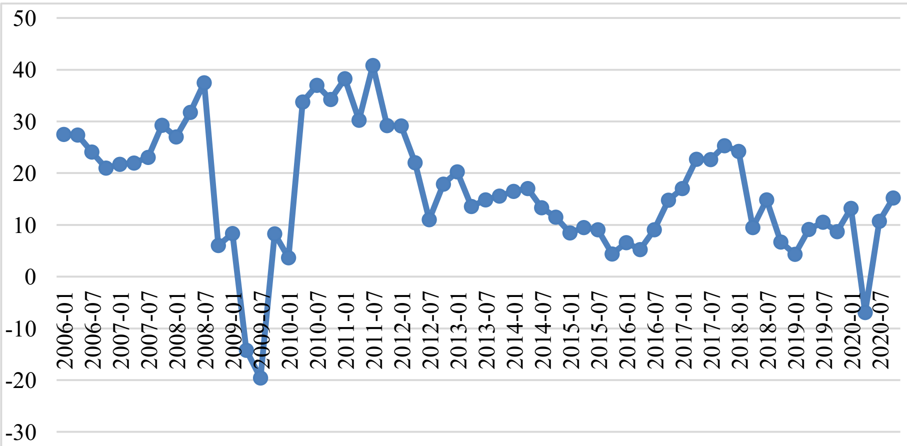

Fig. 1. Vietnam's quarterly export growth for the period 2006–2020 (Source: data analysis results of the authors)

[1] https://moit.gov.vn/tin-tuc/hoat-dong/xuat-nhap-khau-dong-luc-quan-trong-cho-tang-truong-kinh-te-d.html accessed on 01/08/2021.

Figure 1 shows that quarterly export growth has fluctuated throughout time. In particular, from 2008 to 2010, Vietnam's exports underwent significant fluctuations, dropping in the first quarter of 2008, hitting a low point in the third quarter of 2009 with a negative growth rate of almost −20%, and then beginning to recover; in the third quarter of 2011, Vietnam's export growth reached the highest level of about 41%. However, there were signals of a downturn starting in the fourth quarter of 2011, and this trend maintained from 2011 to the first half of 2016. Exports showed signs of increasing again from the end of 2016, reaching a new high in the fourth quarter of 2017 with a staggering 26.3% increase, but by 2018 they had reverted to the level of 2015–2016. Due to the impact of Covid-19 in the second quarter of 2020, Vietnam's export growth marked a serious decrease with a growth rate of −7%. However, the average export per quarter of the whole period 2006 to 2020 increased by around 16.8%.

When we examine the trend closely, we can see that the two periods with the lowest export value are the third quarter of 2009 (−19%) – the period when Vietnam's economy was severely impacted by the global financial crisis – and the second quarter of 2020 (−7%) – the period when Vietnam is facing the second wave of Covid-19, which means that shocks can have a significant impact on the country's exports. Furthermore, Vietnam's export growth trend is not stable, which further confirms the importance of forecasting. These actual results will be compared to the results obtained by the research team when conducting regression of mixed frequency data in Vietnam, which will be analyzed and presented in the following section.

4.2 Forecast Results of Vietnam's Export Growth Using the MF-VAR Model

The MF-VAR model is used in this study to predict Vietnam's export growth. There are some economic indicators at varying frequencies are included in the data collection. As a result, in order to develop the optimal model for predicting export growth, we examine the MF-VAR model using three scenarios in which the data of some economic indicators change with increasing frequency. This analysis also aims to see if the projected outcomes from utilizing the MF-VAR model with data containing multiple high-frequency variables are better or not. All three models include 16 variables mentioned in the research data, in which:

- Model 1: includes 5 quarterly frequency variables and 11 monthly frequency variables, with a lag after analysis and selection at level 2.
- Model 2: includes 2 quarterly frequency variables and 14 monthly frequency variables. The difference from model 1 is that the three quarterly frequency variables have been replaced by three monthly frequency variables including gold price, USD price and crude oil price.
- Model 3: includes 2 quarterly frequency variables, 11 monthly frequency variables and 3 weekly frequency variables. The difference between model 1 and model 2 is that model 3 includes datasets with all quarterly, monthly and weekly frequencies.

To compare the prediction results of the three models, the accuracy (or error) of the results must be measured. There is no one-size-fits-all indicator, therefore this is a difficult assignment. Currently, MAE (Mean absolute) and RMSE (Root mean squared) are regarded as two excellent predictive accuracy measurement criteria that are commonly applied in many studies. As a result, we examine the predicted accuracy of three models using those two criteria in this study (Table 6).

Table 6. Evaluation of prediction error of MF-VAR models

Time	Export growth (Quarterly)	MF_VAR1	Absolute error	MF_VAR2	Absolute error	MF_VAR3	Absolute error
2019Q1	4.34	24.17	19.83	20.56	16.22	43.28	38.94
2019Q2	9.15	24.69	15.54	20.67	11.52	47.62	38.47
2019Q3	10.52	25.23	14.71	20.76	10.24	50.71	40.19
2019Q4	8.70	25.72	17.02	20.81	12.11	40.47	31.77
2020Q1	13.21	23.14	9.92	20.82	7.61	37.88	24.66
2020Q2	−6.94	26.47	33.41	20.81	27.75	24.78	31.72
2020Q3	10.73	26.72	15.99	20.78	10.05	−2.71	13.44
2020Q4	15.22	26.93	11.70	20.76	5.53	−41.04	56.26
R2		**0.82**		**0.85**		**0.85**	
RMSE		**11.05**		**11.00**		**21.05**	
MAE		**7.20**		**7.79**		**16.88**	

(Source: data analysis results of the authors)

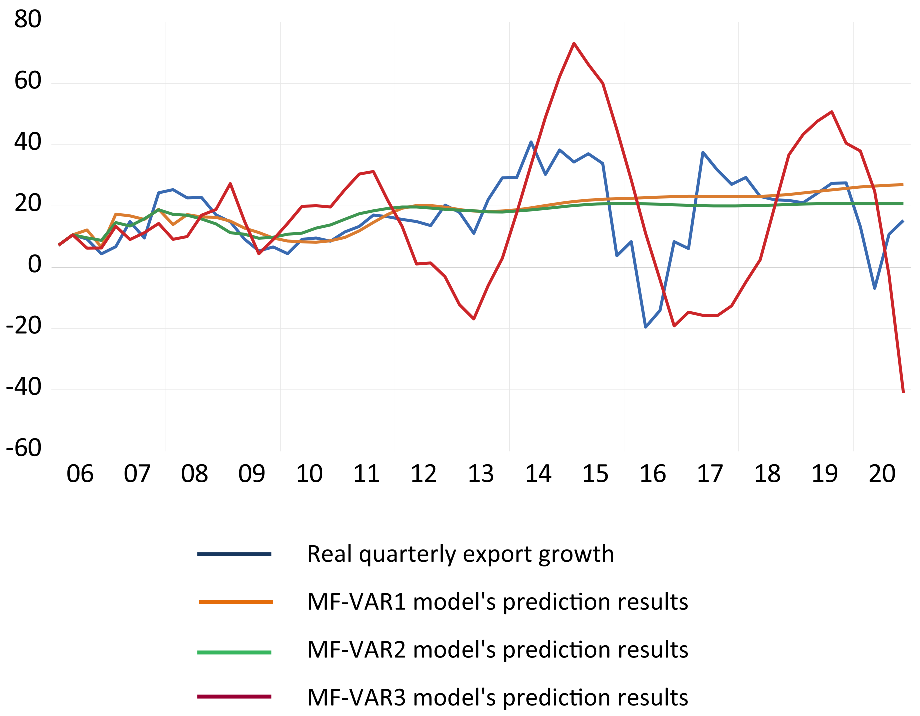

Fig. 2. Vietnam's real quarterly export growth and 3 forecast results *(Source: data analysis results of the authors)* (Color figure online)

Figure 2 shows that model 1 (orange line) delivers the closest predicted results to the true data series (blue line) across the whole period, with the smallest RMSE and MAE error values among the three models. Model 3 (red line) simulates the trend better, but the discrepancy between the model 3 forecast points and the original data is fairly large; there are variances of up to 30% in the fourth quarter of 2015, nearly 40% in Q4 2013, and more than 40% in Q4 2017.

The average error of the three models, according to MAE standards, is greater than 7%. Furthermore, because models 1 and 2 do not predict the peaks and troughs of the export growth cycle, determining the direction of recovery, growth, or decrease is difficult. Models 1 and 2's forecast values are almost ineffectual from 2012 onwards because export growth changes regularly with a significant margin, and model 3's amplitude fluctuates far beyond the actual value. As can be observed, the MF-VAR model is not suitable for forecasting Vietnam's export growth over the research period using the given collection of macroeconomic data.

4.3 Forecast Results of Vietnam's Export Growth by MIDAS Model

The MIDAS model has many different variations, this study applied the basic MIDAS model using the variables presented in Sect. 3.2. The research takes into account three variant models of MIDAS that are compared to the MF-VAR model, each consisting of 16 variables, of which 15 are independent.

- Model 1: 15 independent variables including 4 quarterly frequency variables: GDP_G, BOP, NFDI, KA and 11 monthly frequency variables.
- Model 2: 15 independent variables including 1 quarterly frequency of GDP; 14 monthly frequency variables (11 old variables and 3 new variables: XR_M, G_M, O_M).
- Model 3: 15 independent variables including 1 quarterly frequency variable, 11 monthly frequency variables and 3 weekly frequency variables (XR_W, G_W, O_W).

According to published researches, such as Kuzin (2011), Foroni and Marcellino (2013), the MIDAS model is useful in short-term forecasting since it is based on data that is closest to the forecast time. As a result, using high-frequency data collected through December 2020, the research team applies the MIDAS models to forecast Vietnam's export growth from the first quarter of 2019 to the fourth quarter of 2020.

The forecasting process is carried out as follows: Based on the data from the previous quarter, the author team will generate a forecast for the next quarter using a built model. The forecasted value is compared with the actual data of that quarter when it is published to estimate the error of the model. (Example: use the results of the fourth quarter of 2018 to forecast the results of the first quarter of 2019 and compare the results with the actual published data of the first quarter of 2019). The forecast results of Vietnam's export growth by quarter are presented in Fig. 3 and Table 7 below:

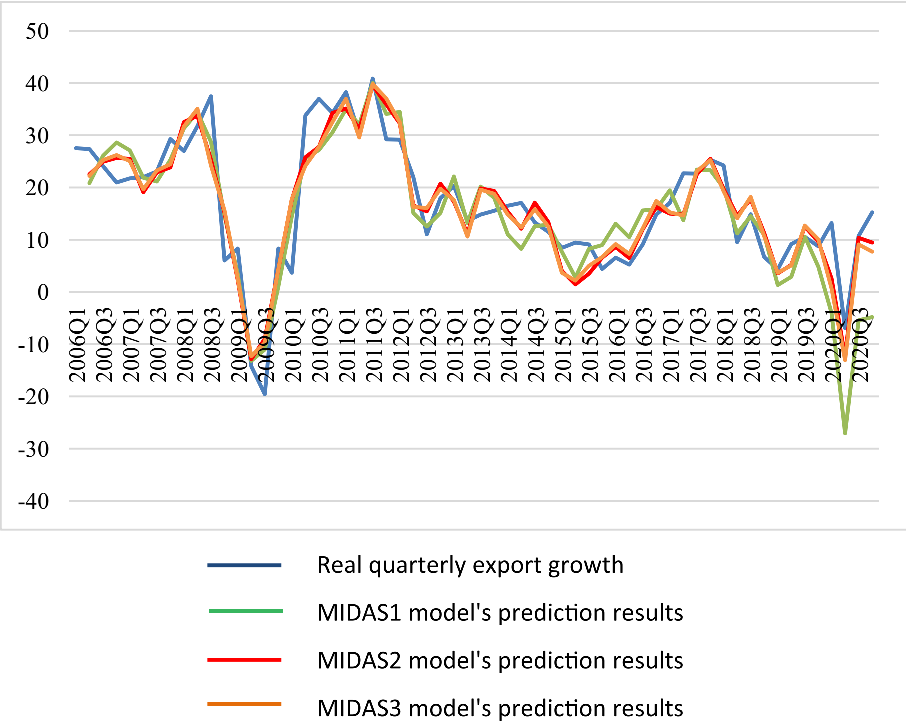

Fig. 3. Quarterly export growth forecast results of the model MIDAS1, MIDAS2, MIDAS3 *(Source: data analysis results of the authors)*

Figure 3 indicates that, when compared to the actual data series, the three MIDAS models accurately anticipate the variation trend of Vietnam's export growth chain. Most of the time, anticipated export growth follows real export growth, with only a few exceptions, such as the period early 2006 to mid-2007, the first two quarters of 2014, and 2015. The three forecast lines' range of fluctuations is fairly tiny in comparison to the value of the actual line, with the average absolute error of the three lines ranging from 4% to 5%. Furthermore, the forecasted outcomes closely track the export cycle's peaks and troughs. When comparing the prediction lines of the MIDAS2 and MIDAS3 models, it is clear that the two models provide very comparable forecast results; in fact, the two lines are practically identical.

The forecast results of the MIDAS2 and MIDAS3 models in the period 2019–2020 are pretty similar to the actual data when compared to model 1. Because of the advantage of using various and high-frequency data in the model, the mixed-frequency model outperforms classic models (VAR, SVAR, BVAR, VECM) in capturing rapid and unexpected changes in export activity. Traditional models, on the other hand, are frequently constrained by the use of data with the same frequency (for example, monthly models only use data with a monthly frequency, quarterly models only use data with a quarterly frequency), so their flexibility will be limited compared to the MIDAS model.

- *Evaluation of forecast errors*

To compare the forecast results of 3 MIDAS models, we also use two common predictive accuracy measurement criteria in many studies, namely MAE (Mean absolute) and RMSE (Root mean squared).

Table 7. Evaluation of prediction error of 3 MIDAS models

Time	Export growth (Quarterly)	MIDAS1	Absolute error	MIDAS2	Absolute error	MIDAS3	Absolute error
2019Q1	4.34	1.33	**3.00**	3.57	**0.77**	3.69	**0.65**
2019Q2	9.15	2.89	**6.26**	5.19	**3.96**	5.14	**4.01**
2019Q3	10.52	10.59	**0.06**	12.52	**1.99**	12.72	**2.20**
2019Q4	8.70	4.81	**3.89**	9.64	**0.94**	10.05	**1.35**
2020Q1	13.22	−4.056	17.27	2.57	10.65	0.65	12.56
2020Q2	−6.94	−27.09	20.15	−12.38	5.44	−13.08	6.14
2020Q3	10.73	−5.46	16.19	10.34	0.39	9.06	1.67
2020Q4	15.22	−4.84	20.06	9.46	5.76	7.69	7.53
R2		**0.82**		**0.82**		**0.81**	
RMSE		**6.87**		**5.17**		**5.141**	
MAE		**5.14**		**4.13**		**4.07**	

(Source: data analysis results of authors)

Table 7 shows that the three proposed MIDAS models with Almon weights provide a good forecast of Vietnam's quarterly export growth, particularly in the four quarters of 2019, the MAE of all three models are mostly smaller than the four quarters of 2020. In 2019, model 1 has 3 quarters, models 2 and 3 have all 4 quarters with forecast errors smaller than the average absolute error MAE, while the 4 quarters of 2020 mostly have errors bigger than MAE. If only considering model 1, the forecast for all quarters of 2019 is quite good with a low prediction error; the best forecast result belongs to the 3rd quarter with an error of only 0.06%, followed by the 1st quarter (3%) and the 4th quarter (3.89%); however, when entering 2020, the forecast error has a rather large difference, with all quarters having an error level of over 16%. For models 2 and 3, the scenario is similar, but the forecast error has improved significantly. If we only consider the year 2019, both models, in particular, have a low prediction error of roughly 4%; model 2 has up to three-quarters of anticipated results under 2%, while model 3 has three-quarters of forecast results under 2.2. Models 2 and 3 also have significantly smaller forecast errors than model 1 when calculated solely for 2020. This finding demonstrates that the MIDAS model is superior in short-term forecasting. The forecast error according to two criteria MAE and RMSE once again show that MIDAS2 and MIDAS3 models predict export growth of Vietnam better than MIDAS1 models, and the forecast results of two models (MIDAS2 and MIDAS3) is almost equally good.

This is due to the fact that the MIDAS2 and MIDAS3 models used more variables with higher frequency data than the original model. Specifically, three quarterly financial and monetary indices (capital account, balance of payments, and net direct investment) are replaced by three monthly monetary indices (gold price, USD/VND exchange rate, and crude oil price) in the MIDAS2 model, and three weekly monetary indices (gold price, USD/VND exchange rate, and crude oil price) in the MIDAS3 model. However, the sets of economic indicators included in the export growth forecast model are the same in models 2 and 3, with the exception of the monthly and weekly frequency, thus the difference in forecast results is not considerable, but the MIDAS3 model still outperforms the MIDAS2 model. At the same time, this forecast result demonstrates that the Midas model outperforms the MF-Var model in terms of predictive efficiency.

This result proves that the development of these forecasting models is on track, and the input data is suitably chosen. Furthermore, adding high-frequency financial data to the MIDAS regression model also enhanced prediction accuracy, demonstrating that financial data has an important role to play in predicting export growth. This finding implies that in order to fully utilize the forecasting ability of financial indicators, they must be combined with macroeconomic data.

5 Conclusion and Future Work

The study used the MF-VAR and MIDAS models to anticipate Vietnam's export growth from 2006 to 2020 using a data collection of varied frequency economic indicators. The analysis results show that the MF-VAR model has high forecasting error and is not suitable for forecasting the export growth of Vietnam, whereas the MIDAS model produces good forecast results using the same data set. Furthermore, as with several previously published study findings, the forecast results suggest that the MIDAS model is effective for short-term forecasting. The findings also suggest that high-frequency financial variables can be used to forecast Vietnam's export growth. This could be linked to the recent development of Vietnam's financial market.

From a policy standpoint, our findings imply that financial variables need to be closely monitored to predict fluctuations in the export cycle. In terms of modelling, our results point to the importance of linking financial sectors and economic realities in macroeconomic models. The role of financial variables in predicting export growth is not only due to their forward-looking nature but also to the close association between financial markets and a country's import and export activities.

In recent years, mixed frequency data analysis models have gotten a lot of attention and use in the economic and financial disciplines, particularly in forecasting national macroeconomic indicators. However, in Vietnam, there is almost no research in this field. Therefore, the application of models with mixed frequency data such as MF-VAR and MIDAS to forecast Vietnam's macroeconomic indicators is a promising new research direction, but due to the lag in publication of data on economic indicators, lead to gathering data for the study is still difficult, the research team has limited itself to estimating Vietnam's export growth at quarterly level, rather than at a higher frequency such as monthly or weekly.

Short-term forecasts of macroeconomic indicators will be critical in formulating policy and presenting each country's economic development strategies. As a result, with the goal of making timely and relevant forecasts in the future, the research team will continue to build and deploy models with varying frequency data to anticipate macroeconomic indicators.

Acknowledgments. This research is a part of the project funded by the Ministry of Education and Training and Thuongmai University under grant number B2020-TMA-04.

References

Başer, U., Bozoğlu, M., Eroğlu, N.A., Topuz, B.K.: Forecasting chestnut production and export of Turkey using ARIMA model. Turk. J. Forecast. **2**(2), 27–33 (2018)

Bin, J., Tianli, X.: Forecast of export demand based on artificial neural network and fuzzy system theory. J. Intell. Fuzzy Syst. (Preprint), 1–9 (2020)

Bui, T.M.N., Nguyen, T.Q.N., Nguyen, T.Q.C.: Using Arima model in forecasting the value of Vietnam's exports. J. Financ. Account. Res. **186**(1), 58 (2019)

Bussière, M., Fidrmuc, J., Schnatz, B.: Trade integration of Central and Eastern European countries:lLessons from a gravity model. ECB Working Article, No. 545 (2005)

Dung, T.T.M., Tuan, V.T., Huong, D.T.M., Luan, N.M.: Applying quantitative methods in forecasting Pangasius exports. Sci. J. Can Tho Univ., 123–132 (2014)

Eckert, F., Hyndman, R.J., Panagiotelis, A.: Forecasting Swiss exports using Bayesian forecast reconciliation (No. 457). KOF Working Papers (2019)

Feenstra, R.C.: Advanced International Trade: Theory and Evidence. Princeton University Press, Princeton (2015)

Foroni, C., Marcellino, M.G.: A survey of econometric methods for mixed-frequency data. Available at SSRN 2268912 (2013)

Ghysels, E., Santa-Clara, P., Valkanov, R.: The MIDAS touch: mixed data sampling regression models. Powered by the California Digital Library University of California (2004). 34p. https://escholarship.org/uc/item/9mf223rs

Hai, N.M., Dung, N.D., Thanh, D.V., Lam, P.N.: Building export value forecast model using dimensional reduction method based on kernel technique. In: The XXIII National Conference: Some Selected Issues of Information and Communication Technology (2020)

IMF: IMF Country Report No. 21/42 (2021)

Kuzin, V., Marcellino, M., Schumacher, C.: MIDAS vs. mixed-frequency VAR: Nowcasting GDP in the euro area. Int. J. Forecast. **27**(2), 529–542 (2011)

Le, N.B.N., T. A., Le, L.Q.T.: Forecast model for shrimp export prices of Vietnam. Sci. J. Can Tho Univ., 188–195 (2018)

Lehmann, R.: Survey-based indicators vs. hard data: what improves export forecasts in Europe? (No. 196). ifo Working Paper (2015)

Mariano, R., Murasawa, Y.: A coincident index, common factors, and monthly real GDP. Oxford Bull. Econ. Stat. **72**(1), 27–46 (2010)

Pyo, H.K., Oh, S.H.: A Short-term Export Forecasting Model using Input-Output Tables (2016)

Qu, Q., Li, Z., Tang, J., Wu, S., Wang, R.: A trend forecast of import and export trade total volume based on LSTM. In: IOP Conference Series: Materials Science and Engineering, vol. 646, no. 1, p. 012002. IOP Publishing (2019)

Schorfheide, F., Song, D.: Real-time forecasting with a mixed frequency VAR. Mimeo (2011)

Siggel, E.: International competitiveness and comparative advantage: a survey and a proposal for measurement. J. Ind. Compet. Trade **6**(2), 137–159 (2006)

Stoevsky, G.: Econometric Forecasting of Bulgaria's Export and Import Flows. Bulgarian National Bank Discussion Articles DP/77/2009 (2009)

Nguyen, T.N.A., Pham, T.H.H., Vallée, T.: Trade volatility in the association of Southeast Asian nations plus three: impacts and determinants. Asian Dev. Rev. **37**(2), 167–200 (2020)

Urrutia, J.D., Abdul, A.M., Atienza, J.B.E.: Forecasting Philippines imports and exports using Bayesian artificial neural network and autoregressive integrated moving average. In: AIP Conference Proceedings, vol. 2192, no. 1, p. 090015. AIP Publishing LLC (2019)

World Bank: World Development Indicator (2020)

Xie, Q., Xie, Y.: Forecast of the total volume of import-export trade based on grey modelling optimized by genetic algorithm. In: 2009 Third International Symposium on Intelligent Information Technology Application, vol. 1, pp. 545–547. IEEE (2009)