EVALUATING FORECAST PERFORMANCE OF GLOBAL VECTOR AUTOREGRESSIVE MODEL

CHUNYEUNG KWOK[*]

Department Westminster Business School, University of Westminster, London NW1 5LS, UK

Abstract: This paper aims to investigate how GVAR (Global Vector Autoregressive) fares against other macro models. For the forecasting exercise, the ability is compared between a generic AR (Autoregressive) model with GVAR ex-ante and GVAR-ex post forecasts. It is easy to see that certain properties are similar among the models such as the long run appears to be unaffected by a monetary shock or that the GDP is negatively affected by it. However, there are also a lot of discrepancies in the short run, particularly in the first 4 quarters. From this, we can conclude that the GVAR model fares best in forecasting that it explicitly allows error correction mechanisms among country models. The paper concludes that the GVAR model is quite adaptable in terms of allowing the data to dictate the short run but also relying on more theory-led identification for the long run.

Keywords: forecasting; modelling; GVAR.

2020 AMS Subject Classification: 62P20, 91G70

INTRODUCTION

The purpose of this paper is to conduct an empirical comparison of the GVAR model with other

models. The strategy for comparing models and testing empirical applications for forecasting is examined. Two GVAR models are estimated using a global dataset, with one model having a restriction on interest rates and the other being unrestricted. The data is forecast within the sample to establish a benchmark for forecasting power. In the second step, results from GVAR model forecasts are compared with forecasts from the standard autoregressive (AR) model. The theoretical foundation of the GVAR model emphasises its ability to use the full information set available, raising the question of whether such extra information enhances forecasting accuracy. However, empirical tests indicate that GVAR ex-ante forecasts do not perform better than simple AR models. Only 15 out of 30 AR estimated models outperformed GVAR forecasts in terms of lower root mean square (RMSE) error. Although the GVAR model's forecasting results are not particularly encouraging, its forecast ability remains inconclusive. For instance, the GVAR model proves to be effective for a few variables, such as exchange rates for specific countries.

The economic forecasting literature, including works by Armstrong [1], Elliott and Timmermann [8], and Granger and Newbold [11], Kwok [12] and Kwok [13], highlights that the value of forecasting can only be comprehended in the context of guiding decisions in areas of economics and finance. Elliott and Timmermann [8] summarised a widely used framework for evaluating economic forecasts that addresses conceptual issues such as how to compare forecasts, how to measure forecast accuracy, and how to manage potential errors. Comparing variables of interest can be challenging as they may not be the same, and this can lead to a risk of comparing "apples and oranges". To mitigate this, variables must be defined clearly to ensure that comparisons are made on the same basis. Model uncertainty is another issue as misspecified models can produce biased forecasts, either predicting too optimistically or pessimistically in certain directions. To overcome this, diagnostic tests can be performed on the estimated model to ensure that it is suitable for forecasting purposes.

Accuracy is a contentious issue in economic forecasts and is highly valued by both users and creators. While the problem of assessing the accuracy of a forecasting model is not particularly difficult as there is typically a "true model" available for comparison, defining the measurement of accuracy remains challenging. Quality is another closely related concept, which becomes relevant when two competing forecasting models produce equally accurate forecasts. In this case, the extra information that a model can convey becomes important and heavily depends on the

user's purpose. If the sole interest is in knowing the forecast, a sophisticated model may not be more valuable than a naïve approach that provides more accurate forecasts. This paper focuses on the accuracy of forecasts, with the discussion of the additional benefits of employing GVAR, which offers a larger information set than simple naïve forecasts, reserved for the end.

Summary Statistics

The aim of this section is to provide a simple and intuitive way to evaluate the accuracy of economic forecasts. In the literature, the most commonly used summary statistics to achieve this goal are the mean error (ME) and mean absolute error (MAE). The ME, which is also known as the bias measure, should be close to zero for a good forecast. It is calculated by summing up the errors (i.e., the differences between forecast values and actual values) and dividing by the forecast horizons. The MAE measures the same thing as the ME, but in absolute values. Therefore, both overestimates and underestimates are treated as positive errors, giving equal penalty to both directions.

Another method of summarising the difference and treating both positive and negative errors equally is by squaring the errors, thus yielding positive values. This is known as the root mean squared error (RMSE) which is equal to $\sqrt{\sum_1^T E_t^2 / T}$. However, it should be noted that the unit of RMSE is directly based on the forecasts it has measured. For example, suppose we have two sequences of forecasts in front of us - one for GDP per capita of a country (forecast of £35,000 and actual result of £36,000) and another one for GDP annual growth rate (forecast of 1.2% and actual result of 1.5%). By definition, the RMSE would be larger for GDP per capita as the base unit of measurement is larger than that of percentage in growth rate. Therefore, RMSEs cannot be used for direct comparison across models. If we wish to compare RMSEs across different models with different values, we need to normalise the calculated RMSEs first. There are several ways to normalise it, but for this chapter, RMSEs were divided by the mean of the sum squared difference from the forecast horizon. Suppose a forecast horizon is 8 periods or two years (8 quarters), then we have n=8 and the difference between forecast and actual would be squared. The RMSE would then be divided by this mean. Similar to other measures, the lower the value, the better the value for the forecasting model and an exact forecast would give a perfect 0 value.

The table below shows the relationship between the two, where there are 8 forecast results for two different sequences. The first forecasts show each period increases by 1% per period, reaching 1.07 after 8 periods. The same increase is also applied to the second forecast sequence which begins a 0.10 and with an increase of 1% per period, reaching 0.11 after 8 periods. As the base unit values they begin with are different, the RMSE would be different for them.

Period	Forecast 1	Actual 1	Forecast 2	Actual 2
1	1.01	1.03	0.10	0.13
2	1.02	1.06	0.11	0.14
3	1.03	1.09	0.12	0.15
4	1.04	1.12	0.13	0.16
5	1.05	1.15	0.14	0.17
6	1.06	1.18	0.15	0.18
7	1.07	1.21	0.16	0.19
8	1.07	1.24	0.17	0.20

As explained in the passage, the RMSE value would be different for the two sequences because the base values they start with are different. In this case, the second sequence starts with a much smaller base value than the first sequence, so even though both sequences have the same percentage increase in forecast errors over each period, the actual value increase for the second sequence is much larger.

To compare the RMSE values across the two sequences, we need to normalize them. In this case, the RMSE mean is calculated as the sum of squared differences between the forecast and actual values divided by the number of periods (n), which is 8 in this case. Then, the RMSE value for each sequence is divided by this mean. The resulting normalized RMSE values are:

Sequence	RMSE	Normalized RMSE
1	0.0265	1.0000
2	0.0026	0.0981

1	2	3	4	5	6	7	8	RMSE	RMSE_Std	RMSE_mean
1.000	1.010	1.020	1.030	1.041	1.051	1.062	1.072	1.018	40.317	0.983
0.100	0.101	0.102	0.103	0.104	0.105	0.106	0.107	0.322	127.495	3.107

Table 1 - RMSE vs RMSE/mean

We can see that the normalised RMSE value for the second sequence is much smaller than that for the first sequence, even though the actual increase in forecast errors over each period was the same for both sequences. This is because the smaller base value of the second sequence led to a larger actual increase in forecast errors, and therefore a larger RMSE value. By normalising the RMSE values, we can compare them across different sequences even when the base values are different.

The rank of RMSEs and Sum of RMSEs

We can then rank the RMSE ranks, with the smallest being the best and the biggest being the worst. The above result shows that, even though the RMSE can be normalised by dividing the mean, it can still favour those with larger base values to begin with. To mitigate this, there are two ways. Instead of comparing RMSE or RMSE/mean across models, we can compare the sum of RMSEs of several models together. For example, suppose we have two GVAR models, GVAR00 and GVAR01, both estimated with the same data but with different specifications. In that case, instead of comparing the individual country model within the GVAR models, we can compare the sum of RMSEs or RMSE/mean of GVAR00 with GVAR01. Since both models would have the same number of country models and the same variables, the sum of RMSEs for both models would not be distorted by the issues mentioned above. In this case, the comparison is much simpler, with the mode that has the smallest RMSE being better.

Theil's U Test

Another measure that is not distorted is Theil's U Test, which is similar to the concepts mentioned above. The formula for calculating it is as follows: first, the sum of squared differences between forecast values (P) and actual values (A) is found, which is then divided by the sum of squared actual values. This indicator is more ideal for assessing the relative quality of a forecast, as it takes into consideration the values of the variables of interest, such as A2. A value of 0 would indicate a perfect forecast in this case.

Table 2:   Directional test

	1	2	3	4	5	6	7	8	Diff 1&2	Diff 2&3	Diff 3&4	Diff 4&5	Diff 5&6	Diff 6&7	Diff 7&8	Pos	Neg	Pos diff	Neg diff
Forecast	5.15446	5.15396	5.15571	5.1576	5.15883	5.15992	5.16041	5.16151	-0.00049	0.00175	0.00188	0.00124	0.00109	0.00048	0.0011	6	1	4	-4
Actual	5.16994	5.19774	5.18922	5.18204	5.17559	5.15863	5.15178	5.16336	0.0278	-0.00852	-0.00718	-0.00645	-0.01696	-0.00685	0.01158	2	5		

$$U_t = \frac{\sqrt{\sum\limits_{t=1}^{n} (P_i - A_i)^2}}{\sqrt{\sum\limits_{t=1}^{n} A_i^2}}$$

Directional tests

Another measure that can be used to evaluate the quality of a forecast is the direction of the forecast. If the goal is simply to determine whether a variable will increase or decrease, then the size of the forecast error may be less important. In this case, the only thing that needs to be measured is the direction of the forecast and the actual outcome.

For example, if a sequence of forecasts is made for 8 periods, any positive value would indicate a positive direction while a negative value would indicate a negative direction. By comparing the sum of directions, we can determine whether the model over or underestimated the actual direction of the results.

To determine the direction of the sequence, we can take the difference between nearby periods, i.e. the difference between the 1st and 2nd period, 2nd and 3rd period, and so on. If the sum of the positive and negative directions of the forecast is equal to the actual results, then the forecast is considered perfect.

However, it should be noted that the difference between positive and negative directions would always be equal, as if one forecast is overestimated, it must be underestimated from the actual result perspective. Summing up positive and negative directions provides a robustness check on the result to ensure that it was calculated correctly.

In the example given, the forecast had 6 positive directions and only 1 negative direction. However, the actual result was 2 positive and 5 negative, indicating an opposite trend.

Comparison with naïve forecasts

Although the user will be able to compare different forecasts using the above tools, they often mean little in isolation and without context as to whether the forecast errors are due to the model

or the nature of the variables being difficult to predict. In this case, a relative comparison can be made with GVAR and other so-called "naïve" models. The purpose of using naïve models is to see if the additional features from GVAR models can add value to forecast accuracy. Several popular naïve models can be used. For example, one can simply generate random numbers given certain parameters that describe the distribution of the variables with Monte Carlo, or a simple model that simply goes up or down by a certain percentage. Therefore, it is expected that GVAR must at least beat randomly generated forecasts, as otherwise, it would prove the model useless. Similarly, a random walk model can also be used to compare whether the GVAR forecasts would be better. In this chapter, autoregressive models were estimated instead, as they tend to be more accurate and would be significantly more meaningful compared with randomly generated models. The forecasts of simple AR models solely rely on their lags; therefore, it should be the most simplistic but also practical alternative to GVAR models, instead of using random walks. The equation below shows the AR(p). Similar to other time series models, the estimation of the models would also be subject to diagnostic checks such as the augmented Dicky-Fuller test for unit roots and AIC/BIC lag selection, etc. Similar to the ordinary linear regression model, it is assumed that the error terms are independently distributed based on a normal distribution with zero mean and a constant variance, and that the error terms are independent of the y values.

$$y_i = \varphi_0 + \varphi_1 y_{i-1} + \varphi_1 y_{i-2} + \dots + \varphi_1 y_{i-p} + \varepsilon_i$$

It is important to note that comparing the RMSE of different models can be a useful way to assess their relative performance, but it should not be the only measure used. Other factors such as the complexity of the models, the stability of the coefficients, and the interpretability of the results should also be taken into account. Additionally, it is important to keep in mind that different models may be better suited to different types of data and forecasting tasks. Therefore, a careful evaluation of multiple models and their respective strengths and weaknesses should be undertaken before making any final conclusions.

Further to the GVAR models, two types of AR forecasts were made, with both ex-post and ex-ante. As ex-post is estimated with the latest available data, it tends to be much better than ex-ante models. The purpose here is not to compare directly with AR (ex-post) to the GVAR model, since

such functionality is not currently available, but this allows us to see the potential room for improvement should we wish to conduct ' nowcasting ' with the latest available data as input. Another purpose of estimating ex-post forecast is that it shows how that particularly depends on the latest available data.

Table 3 - Ranking forecasting models for Saudi Arabia, GDP quarterly growth, the inverse

	Model	2015Q1	2015Q2	2015Q3	2015Q4	2016Q1	2016Q2	2016Q3	2016Q4	Rank	won	
	GVAR1	5.356	5.361	5.364	5.368	5.373	5.377	5.382	5.387	2	57.00%	1
sarbia y	AR (EX Post)	5.365	5.387	5.405	5.391	5.401	5.403	5.402	5.412	1		
	AR (Ex Ante)	5.365	5.383	5.401	5.417	5.432	5.446	5.460	5.474	4		
	GVAR	5.356	5.361	5.365	5.368	5.372	5.376	5.381	5.385	3	58.13%	1
	Actual	5.368	5.386	5.387	5.394	5.394	5.395	5.400	5.415			

of natural log data.

If the difference between ex-post and ex-ante is large and that the ex-post is much more accurate, then it shows that the time series being forecast is much more reliable on its latest data point instead of historical data and the foreign variables (since it is produced from itself hence autoregressive).

Estimating the GVAR model

The GVAR approach involves formulating an individual VARX* model for each country, which relates country-specific variables of interest, such as GDP and inflation. This model represents a large number of variables using linear algebra. The vector of interest, xit, collects macroeconomic variables specific to individual countries over time. The VARX* (2,2) model is represented as xit, x∗it, and uit, where xit is a vector of domestic macroeconomic variables, x∗it is a vector of foreign macroeconomic variables constructed via a weight matrix, and uit is a serially uncorrelated and cross-sectionally weakly dependent process. The weight matrix can reflect trade and/or financial linkages between countries. The VARX* model can also be written in its error-correction form VECMX*, allowing differentiation of short and long-run effects. The GVAR approach involves estimating individual VECMX* models for each country, identifying long-run effects or I(1) relationships across domestic and foreign economies. The second stage of the GVAR approach involves stacking all VARX* models and solving them as a whole. The solution is outlined in di Mauro and Pesaran [7], p.16) and involves the generic VARX* (2,2) model:

$$x_{it} = a_{i0} + a_{i1}t + \varphi_{i1}x_{i,t-1} + \varphi_{i2}x_{i,t-2} + \Lambda_{i1}x_{it}^* + \Lambda_{i1}x_{it-1}^*$$
$$+ \Lambda_{i2}x_{it-2}^* + u_{it}$$

Where the definitions remain the same as defined before, we now introduce a few terms to solve the model as a whole. To form the GVAR model, we first introduce a new term z_{it} define it as:

$$z_{it} = (x'_{it}, x^{*'}_{it})'$$

Therefore we have:

$$A_{i0}W_i x_t = a_{i0} + a_{i1}t + A_{i1}W_i x_{t-1} + A_{i2}W_i x_{t-2} + u_{it}$$

Also recall that for $i = 0, 1, \ldots, N$, which implies the equation above is individual country-specific and require stacking to solve for x t which links all individual models together. We now introduce a few more terms to tidy up the model:

$$G_0 = \begin{pmatrix} A_{00}W_0 \\ A_{10}W_1 \\ \vdots \\ A_{N0}W_N \end{pmatrix}, \quad G_1 = \begin{pmatrix} A_{01}W_0 \\ A_{11}W_1 \\ \vdots \\ A_{N1}W_N \end{pmatrix}, \quad G_2 = \begin{pmatrix} A_{02}W_0 \\ A_{12}W_1 \\ \vdots \\ A_{N2}W_N \end{pmatrix},$$

$$a_0 = \begin{pmatrix} a_{00} \\ a_{10} \\ \vdots \\ a_{N0} \end{pmatrix}, \quad a_1 = \begin{pmatrix} a_{01} \\ a_{11} \\ \vdots \\ a_{N1} \end{pmatrix}, \quad u_1 = \begin{pmatrix} u_{0t} \\ u_{1t} \\ \vdots \\ u_{Nt} \end{pmatrix}$$

thus

$$G_0 x_t = a_0 + a_1 t + G_1 x_{t-1} + G_2 x_{t-2} + u_t$$

As the term G_0 is a known non-singular matrix (invertible matrix). G_0 is called non–singular if there exists an $n \times n$ matrix G_0^{-1} such that $G_0 G_0^{-1} = I_n = G_0^{-1}G_0$. Thus, by multiplying its inverse, the term disappears and we now obtain the GVAR (2) model with 2 lags where:

$$x_t \quad = b_0 + b_1 t + F_1 x_{t-1} + F_2 x_{t-2} + \varepsilon_t$$

Where the new terms collect the inverse of G_0

$$F_1 = G_0^{-1}G_1, F_2 = G_0^{-1}G_2,$$
$$b_0 = G_0^{-1}a_0, b_1 = G_0^{-1}a_1 \; \epsilon_{it} = G_0^{-1}u_{it}$$

The GVAR model above can be solved recursively, see Pesaran [16]. To summarise, as shown above, the GVAR model allows the interactions among the domestic and foreign economies through three diverse channels. The first is the contemporaneous and lagged dependence of domestic variables x_{it} on foreign variables x^*_{it}. In addition, it also allows the effect and

dependence of domestic variables x_{it} on global weakly exogenous variables such as oil and commodity prices. This can also be used as a simulation strategy that can reveal the contemporaneous effects of shocks from country i on j.

Data sources and variables

The model includes 33 countries, with 8 eurozone countries treated as one country in the VARX* model. This list of countries comprises the majority of global output, around 90%. However, due to data quality and availability, some semi-emerging economies such as Russia, Nigeria, Pakistan, and Vietnam are not included. The strict data requirements also mean that most African countries are not included, and the underdeveloped capital markets in emerging markets present a challenge. Other models, such as those developed for soviet economies and developing countries, may be more suitable for accommodating these excluded countries with less stringent data requirements.

GVAR model and Datasets

The datasets contain a large selection of countries and their corresponding economic variables. Currently, the database contains 33 countries, spanning from 1979 to 2016. The model in this study describes the relationships between itself and across 33 countries from 1979q1 – 2016q4. Similarly to Dees et al.(ibid.), the countries in the Eurozone are grouped and considered as 'Euro Area' in the model with its VARX* model. of which 8 eurozone countries are grouped into Euro Area and treated as one country (in the sense of a separate VARX* model). This list of the countries in the model consists of the US, China, Japan, UK, Euro area (Germany, France, Italy, Spain, Netherlands, Belgium, Austria, Finland), Canada, Australia, New Zealand, Sweden, Switzerland, Norway, Korea, Indonesia, Thailand, Philippines, Malaysia, Singapore, India, South Africa, Turkey, Saudi Arabia, Brazil, Mexico, Argentina, Chile and Peru. As it stands, it contains the bulk of the world's output at around 90% ,di Mauro and Pesaran, [7]. Due to data quality and availability, semi-emerging economies such as Russia, Nigeria, Pakistan, and Vietnam are not included.

In terms of variables, there are real output (quarterly in the natural log, seasonally adjusted, with 2015 indexed at 100 for all countries), inflation (constructed from local CPI index, quarterly in the natural log),   real exchange rate (constructed from local currency against USD, where USD is set

as 1, also in the quarter and natural log), real equity price index (from the local largest stock market index, quarterly and in the natural log), short term interest rate (constructed from the local central bank using interest rate, deposit rates, T-bill rates and money market rates, quarterly averages, in natural log, long term interest rate, constructed with interest rates, government securities and bonds, in quarterly averages and natural log. The datasets also include three global variables, namely oil price, raw material and metal price. The oil price is constructed with the Brent crude index, also quarterly and in log. Both raw material and metal prices are taken from primary commodity prices indices and also in the quarterly log.

Real output	33
Inflation	33
Equity price	26
Exchange rate	33
Short interest	32
Long interest	18
Oil Price	1
Material price	1
Metal price	1

Table 3 -GVAR data series

Lag orders of individual VARX* models

Recall that a generic VARX* (p,q) model has lag orders p for both domestic lag orders q for foreign variables. The exact lag orders to be selected are similar to those employed in time series literature with the Akaike information criterion (AIC) or the Schwarz Bayesian criterion (SBC). This is embedded in the GVAR toolbox and the largest values from AIC or SBC are selected for the lag orders.

The table above shows the lag orders selected by either AIC or SBC, whichever value is the highest. It should be noted that it does not matter whether the lag orders of p and q are equal. However, also due to data limitation, an upper limit of two lags is imposed for the test as higher lags would consume too much degree of freedom. This means during the test, the order of (0, 0), (1, 0), (0, 1), and (1, 1) tested for all countries. As the results from the table show, all countries either have the lag order of (2, 1) or (1, 1).

Unit root test

Based on the information provided, it appears that the GVAR approach has an advantage in that it is indifferent to the stationarity or non-stationarity of variables. However, unit root tests, such as the Augmented Dickey-Fuller test, are still useful as they can help identify short-run and long-run relations, such as cointegrating relationships. In this study, the ADF test was conducted on all variables, including real output, inflation, equity price, exchange rate, short-term interest rate, and long-term interest rate. The test was carried out at the 95% confidence level, meaning that if the test statistic for a variable is more negative than the critical value, it will be rejected as having no unit root. The results of the test indicate that most variables have either I(0) or I(1) characteristics, which is ideal for the GVAR approach. These results are displayed in Appendix C of the study, with "N" indicating that the null hypothesis of non-stationarity was not rejected, and "Rej" indicating that it was rejected.

Testing for Cointegrating relationships

The process of identifying cointegrating relationships involves estimating the VECMX* models for each individual country and then using Johansen's trace and maximal eigenvalue statistics to determine the rank of cointegrating relationships for each model. The output from both tests is summarized, and it is noted that the number of cointegrating relationships found differs somewhat from the results reported in Dees et al. [6], which is expected due to newly revised data. Specifically, Japan has the biggest difference between the current estimation and that of Dees et al. [6], with only 2 cointegrating relationships found here but 4 before, while the rest remain similar with a difference of ± 1.

	p	q
ARGENTINA	2	1
AUSTRALIA	1	1
BRAZIL	2	1
CANADA	2	1
CHINA	2	1
CHILE	2	1
EURO	2	1
INDIA	2	1
INDONESIA	2	1
JAPAN	2	1
KOREA	2	1
MALAYSIA	1	1
MEXICO	1	1
NORWAY	2	1
NEW ZEALAND	2	1
PERU	2	1
PHILIPPINES	2	1
SOUTH AFRICA	2	1
SAUDI ARABIA	2	1
SINGAPORE	2	1
SWEDEN	2	1
SWITZERLAND	1	1
THAILAND	2	1
TURKEY	2	1
UNITED KINGDOM	1	1
USA	2	1

Table 4 - VARX order

Testing for weak exogeneity

As mentioned before, the main assumption in the GVAR approach is the weak exogeneity of the foreign variables x_{it}^* concerning the respective VARX* model. As described in Pesaran et al. [6], this assumption is compatible with a certain degree of weak dependence across u_{it} (the residuals). Following the work on weak exogeneity testing by Johansen (1992) and Granger and Lin (1995), the weak exogeneity assumption implies no long-run feedback from x_{it} to x_{it}^*, suggesting that x_{it}^* error correction terms of the individual country VECMX* models do not enter in the marginal model of x_{it}^*, Smith and Galesi, [19]. This implies we can consistently estimate the VARX* models individually and later combine them together to form the GVAR. The proof of weak exogeneity implication on x_{it}^* can be seen in Pesaran [16], ch.23, p.569. The test is as follows:

$$\Delta x^*_{it\ell} = a_{ij} + \sum_{j=1}^{r_i} \delta_{ij\ell} ECM_{ij,t-1} + \sum_{k=1}^{s_i} \Phi'_{ikj} \Delta x_{i,t-k} + \sum_{m=1}^{n_i} \Psi'_{im\ell ikj} \Delta \tilde{x}_{i,\ell-m} + \eta_{it\ell}$$

where ECMij ,t1, j = 1,2,. . . ,*ri* are the estimated error-correction terms corresponding to the cointegrating terms found as shown in previous section. It also should be noted that Δx^*_{itl} is the differenced vector collection of the foreign variables. This is a F-test for the significance of ij, = 0,j = 1,2,. . . ,*ri* above. While the lag orders of p and q were determined earlier via AIC.

The regression was run on the foreign variables in the VARX* models real output (y), inflation (price level, Dp), equity price (eq), short-term interest rate (rs), and long-term interest rate (lr). and also the global variables such as price of metal (pmetal), oil (poil) and raw material (pmat) with 5% significance level. Table on the next page shows the result of 208 regressions run and their F-statistics. Also reported is whether they are rejected or not. The cell of orange indicates it has surpassed the critical value at 5% (defined by its degree of freedom, shown in the second column) which means the assumption of weak exogeneity is not met. Based on all 208 regressions run, only 9 variables (4.3%) are unable to meet the assumption. This result is a slight increase from Dees et al.(2007). Therefore, for example, the foreign long-term interest rate would not enter Australia, Brazil and Turkey VARX* models. Similarly, this applies to other rejected variables.

Testing for structural breaks

Structural breaks are a fundamental problem in econometric modeling that can lead to unreliable models and greater forecast errors. The GVAR literature has extensively discussed the problem of structural breaks, and several test statistics have been developed to assess the structural stability of the estimated coefficients and error variances of the individual VARX*/VECMX* models. These include the maximal OLS cumulative sum (CUSUM) statistic, a test for parameter constancy against non-stationary alternatives, sequential Wald type tests, the QLR statistic, the MW statistic, and the APW statistic.

In this study, the robust versions of these tests were performed on the GVAR model, which included two additional years of data and two global variables. The results show that structural breaks occurred more frequently in the current model compared to those described in the literature.

However, most of the breaks were related to error variances and would not impact the application of the model with impulse responses, as it is based on the bootstrap method for median and confidence boundaries rather than just point estimates.

The tables below show the percentage of variables that were found to have breaks and the estimated dates of the breaks. It is not surprising to find that the dates are mostly related to episodes of financial distresses, as volatility dominates during these periods. Overall, while structural breaks are a significant problem in econometric modeling, the robust versions of the tests used in this study provide a more accurate assessment of the model's stability.

Forecasting

Similar to most econometric models, one of the main outputs of the GVAR model is the forecasts of the economic variables. Recall that the GVAR is constructed by stacking multiple VARX* models. In our case, we have estimated 33 individual VARX* (p,q) models with variable lags and stacked them together and became a GVAR (2) model. We now show that forecasts can be made from the generic GVAR (p) and applied the method to our study. Recall that the individual VARX* (2,2) i.e. two lags for both domestic and foreign variables:

$$x_{it} = a_{i0} + a_{i1}t + \varphi_{i1}x_{i,t-1} + \varphi_{i2}x_{i,t-2} + \Lambda_{i1}x_{it}^* + \Lambda_{i1}x_{it-1}^*$$
$$+ \Lambda_{i2}x_{it-2}^* + u_{it}$$

Where $x_{it}-$ is a vector with a dimension of ki $\times 1$ of domestic macroeconomic variables indexed by individual country i and time as t; x_{it}^*- is a vector with a dimension of ki $\times$ 1 of foreign macroeconomic variables indexed by individual country i and time as and u_it – is a serially uncorrelated and cross-sectionally weakly dependent process. This can be re-written into:

$$A_i(L, P)W_ix_t = \phi_{it}$$

Where ϕ_it equals x_{it}, L as the lag operator; p as the domestic variable lag orders; W as weight matrix and x_t as the domestic variables denoted in t and i denotes the country. In other words, it is simply a re-statement of the VARX* model as a function of domestic variables with lag orders multiplied by their corresponding weights. Also recall that, once the VARX* models have been estimated individually, the next step is to stack the models together to form the GVAR model. Again, using the notations in Dees et al. [6], by stacking the individual VARX* models (written

as ϕ_{it}), we obtain the GVAR (p) model as

$$G(L,\, p)x_t = \phi_t$$

Where

$$G(L,p) = \begin{pmatrix} A_0(L,p)W_0 \\ A_1(L,p)W_1 \\ \vdots \\ A_N(L,p)W_N \end{pmatrix}, \varphi_t = \begin{pmatrix} \varphi_{0t} \\ \varphi_{1t} \\ \vdots \\ \varphi_{Nt} \end{pmatrix}$$

The GVAR ex-ante forecast model has now formed and can be solved via a recursive method at any horizon N.

GVAR ex-ante forecasts

We now turn to the results produced by the estimated GVAR model. As mentioned before, there are 33 countries in total with 8 euro countries which will be estimated as one, therefore there are 26 country models. Each has its combination of lag orders up to a maximum of 2 as determined by AIC/ BIC. It should be noted not all VARX* models have equivalent lag orders nor the same set of domestic and foreign variables due to the specification tests of lag order and weak exogeneity in the last section. In the end, after removing the variables which did not meet the weak exogeneity assumption, we have 271 variables estimated placed in 26 VARX models and one auxiliary model for global variables such as oil price, metal and raw material price for 8 quarters i.e. 2 years. This means 2184 point estimates were created for all variables. For original output see appendix.

GVAR (conditional forecast) and GVAR1 (unconditional forecast)

As mentioned previously, forecasts can either be conditional or unconditional. In this case, we estimate two sets of forecasts from the same estimated GVAR model. Summary statistics like RMSEs were calculated to see which model is more accurate and whether the restrictions imposed improved the forecast accuracy. If there is a strong conviction or that the future values are already known for a variable in advance, then there is a case to impose such restrictions, fixing the values and letting other values be estimated in light of these restrictions. In this case, restrictions were placed on US short and long interest rates setting both at 1% for short and 2% for long. The GVAR forecasts (also denoted in GVAR0 for easy differentiation) with the restriction are simply shown

as GVAR below while the one without restriction is displayed as GVAR1.

Forecasting models comparison

As there are too many forecasts produced and due to space limit, the following shows only a small selection of the forecasts produced. Looking at the forecasts produced in the figure below for the US interest rate, for example, it is easy to see that the GVAR1 forecast was off by a big margin as it was calculated based on previous data, culminating in a negative interest rate lower each quarter. This is not the case in reality, thus the GVAR0 forecasts, with the predetermined restrictions, fared better than the GVAR1 forecasts. Compare the GVAR1 forecast to the AR (ex-ante) and we can see that the AR model is of no use in forecasting the interest rate movement. In this case, a more simplistic approach proved to be more useful than forecasts based on time series alone. In general, AR ex-ante forecasts and also unrestricted GVAR ex-ante forecasts are useless for forecasting interest rates. This is because interest rates are often decided in advance in light of possible future scenarios, therefore, it is a retroactive process. Past influence is likely to be less useful. If we consider AR ex-post forecasts, then we can see that its performance is much better. In this case, we can conclude that if we wish to improve the forecast on the interest rate, we can use the latest figure, thus it would be much closer to a nowcasting exercise.

It seems that the performance of different forecasting models varies greatly depending on the type of data being analysed. In the case of oil price, material and metal price, the fluctuations were too rapid and frequent for the quarterly data to be useful in forecasting. In contrast, for the Argentina equity index, the GVAR forecasts outperformed the AR ex-ante forecast in indicating a downward trend. For both Brazil GDP and UK equity index, none of the models were able to provide useful forecasts, indicating that the time series data alone did not provide much information.

In terms of interest rate forecasts, the GVAR forecasts often performed better than AR models, possibly due to the additional information conveyed by past interrelationships between international central banks. It is interesting to note that in the case of the China inflation rate, the GVAR forecasts provided a middle course and were better than the AR ex-ante forecast, indicating that the interrelationships between different countries may play a role in forecasting accuracy.

Overall, it seems that the choice of forecasting model and approach should be carefully considered depending on the type of data and the specific context in which it is being analyzed.

Since there are too many forecasts to compare with, it is more efficient to compare at a macro level. In this case, RMSEs were calculated for each model of which there are 271 in total. It has been mentioned previously that individual RMSEs should not be used for comparing across models. However, by summing up the totals, we can then use it to compare two GVAR forecasts and decide which is more accurate.

In both cases, we can see that the RMSEs for emerging markets tend to be more accurate than for developed markets. If we rank the RMSEs, where the best has the value of 1 and the worst has 271, then we would have a sum of 9453 ranks in total (1+2+3...+271). The below table shows, for example in GVAR0 model that the combined ranks for the Brazil forecasts are the best of all countries with a rank of 148 only, while USA has a much larger combined rank of 485. We then see that the same is true for Brazil where the combined ranks are now 92 and Switzerland became the worst at 525. This shows that setting restrictions on the interest rates had helped some countries but worse the forecasts for others. If we are to compare two models, we then find that GVAR0 has a sum of RMSE/mean of 43234 and GVAR1 of 32003. Therefore the RMSE/mean is much smaller for the GVAR1 model where the interest rates are left to be estimated by the time series. Now a paradox has appeared. Although setting restrictions on interest rates had increased accuracy for country interest rates, on the whole, it has failed to improve the overall accuracy. The below table summarises the RMSE/mean difference between GVAR and GVAR1 model ΔGvar0; Gvar1 = Gvar0 - Gvar1. Except for the case of the long interest rates (lr), all variables performed better for GVAR0 forecasts as the RMSEs are much smaller. Again, this proves the same conclusion from the previous paragraph.

After comparing the AR (ex-ante) to GVAR models, it is evident that in general, AR models forecast better as their values are smaller. However, this comparison can be distorted by the initial values that were used to start the models. Therefore, Theil's U statistics were also calculated using the equation defined earlier. This measurement provides a more accurate comparison of the

forecasting performance of the models.

From the Theil's U statistics, we can conclude that AR ex-post performed the best, as expected, with a sum of 7.76. However, the GVAR0 model performed much better than the GVAR1 model, with a sum of 11.07 compared to 19.56. This indicates that in this case, the GVAR0 model is better than the GVAR1 model, and even better than the AR ex-ante model. This proves that GVAR forecasts are better than a simplistic AR model if restrictions are not set.

Directional test

The directional test was applied to evaluate whether the GVAR0 model could anticipate the direction of change and whether the forecasts were moving up or down. The GVAR0 model produced 931 forecast points, with 48% indicating a positive change and 52% indicating a negative change, with no variable staying on the same course. In comparison, the actual results showed 56% moving up and 44% moving down, with no variable staying on the same course over eight quarters (two years). This indicates that the GVAR model overestimated 8% of its forecasts and underestimated 8%, resulting in 77 incorrect calls out of 931, or a 92% accuracy rate.

SUMMARY AND CONCLUSION

Indeed, the accuracy of a forecast is not the only factor that determines its usefulness. Even if a forecast is not perfect, it can still provide valuable information to decision-makers if it captures important relationships between economic variables or identifies emerging trends. GVAR models are particularly useful in this regard, as they allow for the analysis of complex interdependencies among countries and regions. They also provide a framework for exploring the effects of different economic policies and shocks on global economic outcomes.

In addition to their analytical capabilities, GVAR models also have practical applications. For example, they can be used to generate scenarios for stress testing financial systems, or to assess the economic impact of geopolitical events such as trade wars or political instability. GVAR models can also be used by policymakers to evaluate the effectiveness of different policy interventions in different countries or regions.

In summary, while forecast accuracy is an important consideration, it is not the only factor that determines the usefulness of a forecasting model. GVAR models provide a valuable tool for understanding the complex interdependencies of the global economy and for generating scenarios for policy analysis and stress testing.

Figure 1 Forecasts comparison
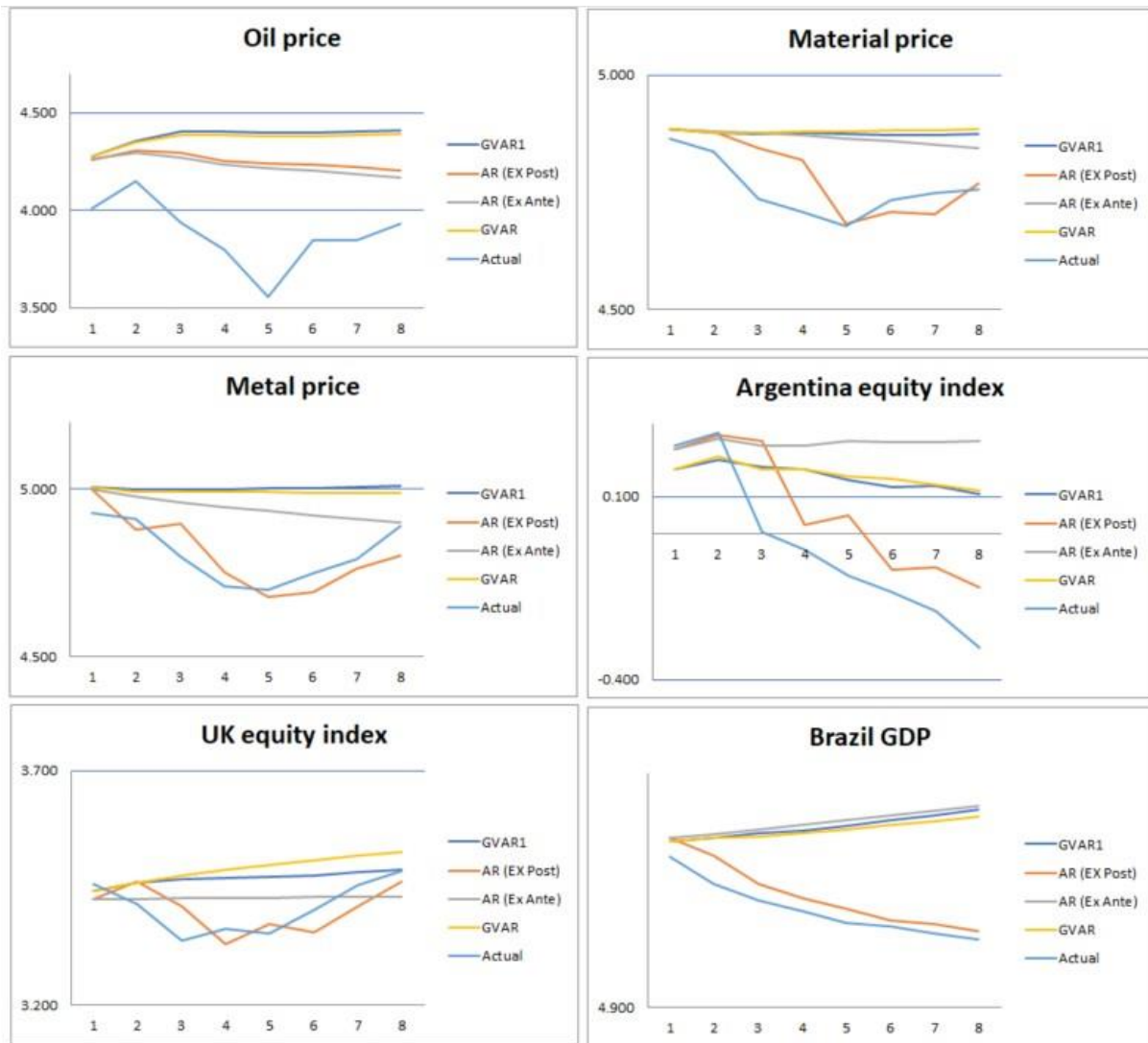
Figure 2 - Forecast for various variables
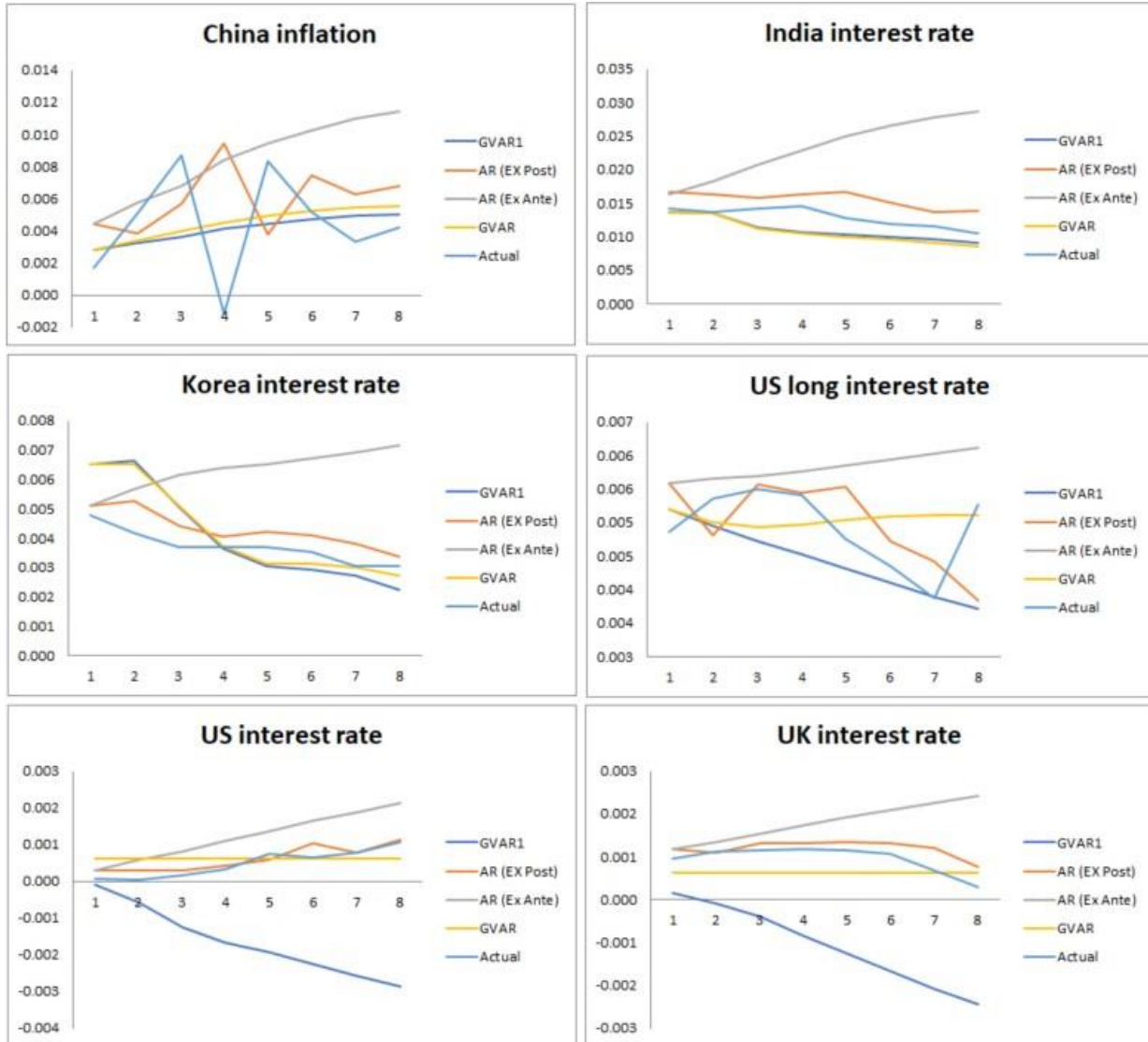
Figure 3 – GVAR country ranks

Country	Data Sum of Rank	Sum of RMSE_mean	Average of RMSE_mean
du_model	29	14.3	4.77
bra	148	75.1	18.77
per	165	90.5	22.62
arg	199	108.9	21.77
sarbia	204	230.5	76.83
chl	248	351.1	70.22
turk	219	446.6	111.65
mex	253	492.2	123.05
phlp	326	572.6	114.51
indns	323	785.1	196.28
thai	358	808.4	161.69
india	432	1037.2	172.87
nzld	429	1147.4	191.24
switz	485	1508.4	251.40
nor	402	1667.3	277.88
mal	313	1792.4	358.49
swe	432	1842.0	307.01
austlia	454	1996.5	332.75
china	340	2011.7	502.93
sing	349	2294.6	458.92
safrc	459	2331.2	388.53
euro	473	2577.7	429.62
kor	448	2825.4	470.89
can	481	3587.5	597.92
japan	496	3841.6	640.26
uk	503	4081.3	680.22
usa	485	4716.4	943.28
Grand Total	**9453**	**43234.0**	**315.58**

Figure 4 -GVAR variable ranks

Variable	Data Sum of Rank	Sum of RMSE_mean	Average of RMSE_mean
poil	1	1.9	1.95
pmetal	8	5.0	5.01
pmat	20	7.4	7.36
eq	434	169.0	8.90
ep	730	359.1	14.36
lr	1594	14415.3	1108.87
y	1764	2293.1	88.20
Dp	2245	5714.2	219.78
r	2657	20268.8	810.75
Grand Total	**9453**	**43234.0**	**315.58**

Figure 5- GVAR   1 country ranks

Country	Data Sum of Rank	Sum of RMSE_mean	Average of RMSE_mean
du_model	31	12.76166921	4.253889737
bra	92	29.16544716	7.291361791
arg	178	69.23511708	13.84702342
per	186	80.98285726	20.24571432
sarbia	215	201.1677103	67.05590344
turk	234	457.1112637	114.2778159
chl	247	258.36996	51.67399201
mex	271	515.7965162	128.949129
phlp	293	277.9633973	55.59267946
thai	320	551.310971	110.2621942
mal	324	1698.011169	339.6022338
indns	337	760.8596665	190.2149166
sing	341	1042.03751	208.4075019
china	357	1788.455377	447.1138441
nor	421	1709.587074	284.931179
india	423	634.1435858	105.6905976
usa	446	1439.597501	287.9195003
nzld	447	997.8389228	166.3064871
kor	448	1556.865097	259.4775161
swe	450	3015.005359	502.5008931
safrc	455	1543.850439	257.3084065
austlia	455	1633.845196	272.3075326
can	466	1944.267464	324.0445773
uk	496	1776.007803	296.0013006
euro	497	2559.35763	426.5596051
japan	498	3309.130892	551.5218153
switz	525	2141.131881	356.8553135
Grand Total	9453	32003.85748	233.6047991

Figure 6- GVAR   1- Variable ranks

Variable	Data Sum of Rank	Sum of RMSE_mean	Average of RMSE_mean
poil	1	1.65	1.65
pmetal	9	4.32	4.32
pmat	21	6.79	6.79
eq	509	151.72	7.99
ep	714	277.59	11.10
lr	1627	12997.62	999.82
y	1810	2005.14	77.12
Dp	2210	4255.78	163.68
r	2552	12303.25	492.13
Grand Total	9453	32003.86	233.60

Figure 7 - Difference between GVAR and GVAR1

Variable	Count of RMSE_mean	Sum of RMSE_mean	Average of RMSE_mean	Max of RMSE_mean	Min of RMSE_mean	StdDevp of RMSE_mean
Dp	0.0	-1458.5	-56.1	-37.1	-6.5	-22.4
ep	0.0	-81.5	-3.3	-21.8	-0.5	-4.5
eq	0.0	-17.3	-0.9	-6.1	-1.2	-0.9
lr	0.0	-1417.7	-109.1	872.3	-62.5	145.1
pmat	0.0	-0.6	-0.6	-0.6	-0.6	0.0
pmetal	0.0	-0.7	-0.7	-0.7	-0.7	0.0
poil	0.0	-0.3	-0.3	-0.3	-0.3	0.0
r	0.0	-7965.6	-318.6	-819.9	-6.1	-288.4
y	0.0	-288.0	-11.1	61.9	-4.2	16.3
Grand Total	0.0	-11230.1	-500.6	47.7	-82.5	-154.8
Average	0.0	-1247.8	-55.6	5.3	-9.2	-17.2

Figure 8 - Forecast evaluation with RMSE

Model	Count of RMSE	Count of Rank	Average of RMSE	Average of Rank	StdDevp of RMSE	Max of RMSE	Min of RMSE
AR (Ex Ante)	30	30	0.057	2.733	0.095	0.373	0.001
AR (EX Post)	30	30	0.043	1.600	0.096	0.394	0.000
GVAR	30	30	0.080	2.867	0.133	0.513	0.000
GVAR1	30	30	0.082	2.800	0.135	0.529	0.000
Grand Total	120	120	0.066	2.500	0.118	0.529	0.000

Figure 9 -Theil's U statistics

Model		Count of Theil's U	Sum of Theil's U2	Max of Theil's U2	Average of Theil's U2	Min of Theil's U2
	Data					
AR (EX Post)		30	7.76	1.797	0.259	0.000369
GVAR		30	11.07	1.857	0.369	0.000799
AR (Ex Ante)		30	14.34	1.752	0.478	0.000539
GVAR1		30	19.56	4.407	0.652	0.000473
Grand Total		120	52.73	4.407	0.439	0.000369

Figure 10 - Directional test

Forecast	Positive	446	931		48%		-8%
	Negative	485			52%		8%
Actual	Positive	523	931		56%		
	Negative	408			44%		

CONFLICT OF INTERESTS

The author declares that there is no conflict of interests.

REFERENCES

[1] J.S. Armstrong, Principles of forecasting: a handbook for researchers and practitioners, Springer New York, 2001.

[2] B. Bernanke, J. Boivin, P. Eliasz, Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach, National Bureau of Economic Research, Cambridge, MA, 2004.

[3] B.S. Bernanke, J. Boivin, P. Eliasz, Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach, Quart. J. Econ. 120 (2005), 387-422.

[4] I. Carabenciov, I. Ermolaev, C. Freedman, et al.   A small quarterly projection model of the US economy. IMF Working Paper 08/278, 2013.

[5] N. Carnot, V. Koen, B. Tissot, Economic forecasting and policy, Palgrave Macmillan, 2nd edition, 2011.

[6] S. Dees, F. di Mauro, M.H. Pesaran, L.V. Smith, Exploring the international linkages of the euro area: a global VAR analysis, J. Appl. Econ. 22 (2007), 1–38.

[7] F. di Mauro, M.H. Pesaran, The GVAR handbook: structure and applications of a macro model of the global economy for policy analysis, Oxford University Press, (2013).

[8] G. Elliott, A. Timmermann, Handbook of economic forecasting, Elsevier, (2013).

[9] C.W.J. Granger, Y. Jeon, Thick modeling, Econ. Model. 21 (2004), 323-343.

[10] C.W.J. Granger, Y. Jeon, Evaluation of global models, Econ. Model. 24 (2007), 980-989.

[11] C.W.J. Granger, P. Newbold, Forecasting economic time series, Academic Press, 2014.

[12] C. Kwok, Estimating structural shocks with the GVAR-DSGE model: Pre-and post-pandemic, Mathematics, 10 (2022), 1773.

[13] J. Kwok, Macroeconometric models for portfolio management, Vernon Press, 2021.

[14] J.C. Williams, A. Levin, V. Wieland, The performance of forecast-based monetary policy rules under model uncertainty, SSRN. (2001). https://doi.org/10.2139/ssrn.288556.

[15] A. Levin, V. Wieland, J.C. Williams, The performance of forecast-based monetary policy rules under model uncertainty, Amer. Econ. Rev. 93 (2003), 622-645.

[16] M.H. Pesaran. Theory and practice of GVAR modelling, in: Time Series and Panel Data Econometrics, chapter 33, pages 900–935. Oxford University, 2015.

[17] CC.A. Sims, Interpreting the macroeconomic time series facts, Eur. Econ. Rev. 36 (1992), 975–1000.

[18] J.B. Taylor, Macroeconomic policy in a world economy: from econometric design to practical operation, W.W. Norton, New York, 1993.

[19] L.V. Smith, A. Galesi, GVAR Toolbox 2.0 User Guide, Technical Report, 2017.