

Analysis of the use of vector autoregressions in economic forecasting

S.G. Svetunkov*

Peter the Great St.Petersburg Polytechnic University (SPbPU), Saint Petersburg, Russia

Abstract. Using vector autoregressions is a promising direction in short-term economic forecasting. They do not simply model the relationship between different factors, but also model the time-distributed relationship of these factors. Vector autoregressions are suitable for modeling complex dynamic economic multifactor processes. The complexity of the problem of estimating coefficients, which increases with the dimensionality of vectors, prevents the widespread use of autoregressions in practice. Vector autoregressions in complex-valued form with the same dimensionality as the modeled vector contain a much smaller number of coefficients. This facilitates the estimation of the coefficients of vector autoregressions. Some problems requiring further investigation arise when using vector autoregressions in complex form. Among them is the problem of selecting the best model. The information criteria used for this purpose limit the variety of vector autoregressions, reducing them to elementary models. The study was supported by the Russian Science Foundation grant No. 23-28-01213, <https://rscf.ru/project/23-28-01213>.

1 Introduction

A scientific paper [1] gave the first insight into vector autoregressive (VAR) models as early as 1980, but vector autoregressive models only became the subject of intense scientific attention in the early 21st century. Almost every scientist could build these models because of the expansion of the computational capabilities of personal computers. Today, VAR theory is a complete doctrine describing the application of VARs and explaining their most important properties [2, 3, 4].

The practical application of vector autoregressions is still at a very low level. If at the end of January 2023, the GOOGLE search engine returned 222,000 results for articles containing the term "vector autoregressions", only 2,000 of them contained results of applying vector autoregressions in practice. If one examines these 2,000 cases of the practical application of VAR closely, one can see that most of these vector autoregressions were two-dimensional. Three- or four-dimensional autoregressions are much rarer, and autoregressions of vectors greater than ten dimensions are the rarest cases in modeling and forecasting practice [5,6]. What causes such a discrepancy between the almost perfect theory and the vanishingly small practice of application of this theory?

Let us give our answer to this question. Let us present the vector autoregression model:

* Corresponding author: sergey@svetunkov.ru

$$\hat{Y}_t = A_0 + A_1 Y_{t-1} + A_2 Y_{t-2} + \dots + A_p Y_t \tag{1}$$

Here Y_t – is k -dimensional vector of variables;
 A_0 – k -dimensional vector of free coefficients;
 A_t – $k \times k$ -dimensional constant real matrix of coefficients.

The vector autoregression model is commonly referred to as VAR(p), where p is the order of autoregression.

From (1) it is easy to see that the number of vector autoregressive coefficients will be equal to $(k^2 \cdot (p+1))$. The original series of variables are usually centered on its arithmetic mean, thus eliminating the k -dimensional vector of free coefficients A_0 . The dimensionality of the problem is reduced and we only have to estimate $k^2 \cdot p$ unknown coefficients. So what is this number?

Suppose a forecaster is going to use VAR(p) for the 10-dimensional ($k=10$) vector of some interrelated economic indicators and assuming the forecast process has three lags ($p=3$). Then he has to solve the problem of estimating from the statistics $(10^2 \cdot 3) = 300$ unknown coefficients. If he wants to increase the dimensionality of the vector by at least one and increase the number of modelled indicators to $k=11$, he will have to estimate $(11^2 \cdot 3) = 363$ unknown coefficients.

The non-linear growth in the number of estimated VAR(p) coefficients with increasing vector dimensionality k limits the practical application of vector autoregressions because of the increasing computational complexity of the problem. Not every researcher has the skills to solve problems of this dimension, so vector autoregressive models do not find the wide practical application they deserve. This explains the fact that high-dimensional autoregressions apply to solving natural science or engineering problems. Only specialists strong in mathematics and mathematical statistics can estimate such many coefficients. A review of scientific publications on the application of vector autoregressions shows their infrequent application in economics.

The order k of a vector autoregression is an important characteristic of a vector autoregression, so when denoting a vector autoregression we will show its dimensionality. We denote by VAR ^{k} (p) the autoregression of a k -dimensional vector of order p .

2 Research methodology

Physics and engineering extensively use the theory of functions of a complex variable. This tool allows us to describe simple complex processes that are poorly formalizable in the real variable domain. The application of the tool in economics has opened a new direction of economic and mathematical modeling - "complex-valued economics". [7]. Many areas of economic and mathematical modeling, including the example of univariate autoregressions, have shown the success of the tools of this scientific section. A simple first-order complex-valued autoregression has the following form [8]:

$$\hat{y}_{1t} + i\hat{y}_{2t} = (a_0 + ia_1)(y_{1(t-1)} + iy_{2(t-1)}) \tag{2}$$

Here y_{1t} and y_{2t} – modeled variables,
 a_0 and a_1 – actual autoregressive coefficients,
 i – is an imaginary unit, the square of which is $i^2=-1$.

Since the equality sign in (2) means that the real and imaginary parts of this equality are equal to each other, after opening the brackets in the right part, multiplying the numbers by each other, and grouping the real and imaginary parts, we can represent it in vector form:

$$\begin{pmatrix} \hat{y}_{1t} \\ \hat{y}_{2t} \end{pmatrix} = \begin{pmatrix} a_0 & -a_1 \\ a_1 & a_0 \end{pmatrix} \begin{pmatrix} y_{1(t-1)} \\ y_{2(t-1)} \end{pmatrix} \tag{3}$$

A simple complex-valued autoregression is a kind of bivariate vector autoregression in which some coefficients are equal to each other. To be sure of this, we present below $VAR^2(I)$:

$$\begin{pmatrix} \hat{y}_{1t} \\ \hat{y}_{2t} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} y_{1(t-1)} \\ y_{2(t-1)} \end{pmatrix} \tag{4}$$

Comparing (3) and (4) with each other, we can see that (3) is a particular case of model (4). To construct model (3) it is necessary to estimate only two unknown coefficients, while to use model (4) it is necessary to estimate four coefficients, that is twice as many.

Let us denote vector autoregression of complex variables as $CVAR^k(p)$ to distinguish it from vector autoregression of real variables.

For the case of a four-dimensional vector, write the $CVAR^4(I)$ model in the complex-valued form:

$$\begin{aligned} \hat{y}_{1t} + i\hat{y}_{2t} &= (a_{11} + ia_{12})(y_{1(t-1)} + iy_{2(t-1)}) + (a_{13} + ia_{14})(y_{3(t-1)} + iy_{4(t-1)}) \\ \hat{y}_{3t} + i\hat{y}_{4t} &= (a_{21} + ia_{22})(y_{1(t-1)} + iy_{2(t-1)}) + (a_{23} + ia_{24})(y_{3(t-1)} + iy_{4(t-1)}) \end{aligned}$$

We can also represent it in matrix form:

$$\begin{pmatrix} \hat{y}_{1t} \\ \hat{y}_{2t} \\ \hat{y}_{3t} \\ \hat{y}_{4t} \end{pmatrix} = \begin{pmatrix} a_{11} & -a_{12} & a_{13} & -a_{14} \\ a_{12} & a_{11} & a_{14} & a_{13} \\ a_{21} & -a_{22} & a_{23} & -a_{24} \\ a_{22} & a_{21} & a_{24} & a_{23} \end{pmatrix} \begin{pmatrix} y_{1(t-1)} \\ y_{2(t-1)} \\ y_{3(t-1)} \\ y_{4(t-1)} \end{pmatrix} \tag{5}$$

Its analogous model $VAR^4(I)$ would be:

$$\begin{pmatrix} \hat{y}_{1t} \\ \hat{y}_{2t} \\ \hat{y}_{3t} \\ \hat{y}_{4t} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix} \begin{pmatrix} y_{1(t-1)} \\ y_{2(t-1)} \\ y_{3(t-1)} \\ y_{4(t-1)} \end{pmatrix} \tag{6}$$

Comparing these two models with each other, we can see that model (5) requires estimating fewer coefficients than model (6). If for the model in complex-valued form (5) it is necessary to estimate values of 8 unknown coefficients, then for vector autoregression of similar dimension (6) it is necessary to find 16 unknown coefficients.

For any even number of the dimensionality of vector k the number of coefficients of model $CVAR^k(p)$ will always be two times less than the number of coefficients of model $VAR^k(p)$. And this is a significant reduction of the dimensionality of the problem to be solved. For example, for $VAR^{10}(1)$ model we need to estimate 100 unknown coefficients, and for $CVAR^{10}(1)$ model we need 50 coefficients.

When a vector of odd dimension k is modeled, the situation is not better. The $CVAR^3(1)$ model for a three-dimensional vector of modeled variables in basic form would look like this:

$$\begin{aligned} \hat{y}_{1t} + i\hat{y}_{2t} &= (a_{11} + ia_{12})(y_{1(t-1)} + iy_{2(t-1)}) + (a_{13} + ia_{14})y_{3(t-1)} \\ \hat{y}_{3t} &= (a_{21} + ia_{22})(y_{1(t-1)} + iy_{2(t-1)}) + a_{23}y_{3(t-1)} \end{aligned}$$

Or, representing it in matrix form:

$$\begin{pmatrix} \hat{y}_{1t} \\ \hat{y}_{2t} \\ \hat{y}_{3t} \end{pmatrix} = \begin{pmatrix} a_{11} & -a_{12} & a_{13} \\ a_{12} & a_{11} & a_{14} \\ a_{21} & -a_{22} & a_{23} \end{pmatrix} \begin{pmatrix} y_{1(t-1)} \\ y_{2(t-1)} \\ y_{3(t-1)} \end{pmatrix} \quad (7)$$

Here we have to estimate 7 unknown coefficients. The three-dimensional vector autoregression $VAR^3(1)$ will look like this:

$$\begin{pmatrix} \hat{y}_{1t} \\ \hat{y}_{2t} \\ \hat{y}_{3t} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} y_{1(t-1)} \\ y_{2(t-1)} \\ y_{3(t-1)} \end{pmatrix} \quad (8)$$

To use this model in practice it is necessary to find values of 9 unknown coefficients - two more coefficients than in the $CVAR^3(1)$ model.

3 Research results

The greater the number of coefficients a model contains, the more accurately it can describe the process being modeled. This rule is not always satisfied, but it is true. Therefore, perhaps in most cases in practice, the $VAR^k(p)$ model will be more accurate than the analogous $CVAR^k(p)$ model. There may be an impression that by clipping the problem into a complex form and making it easier to solve, we degrade the approximation and prediction properties of a vector autoregression. Complex-valued model in this case will always be inferior to $VAR^k(p)$ models, and it will be of theoretical interest only.

Let's check whether these fears have a right to exist.

Hirotsugu Akaike presented a paper "Information Theory and an Extension of the Maximum Likelihood Principle" at the Second International Symposium on Informatization in Budapest (1973) [9]. He proposed a universal method for choosing the best i -th autoregressive model with coefficients θ from a set of models $i=1, 2, \dots, k, \dots, L$ using the maximum likelihood principle. We reduced his proposal to the recommendation to choose a model by the criterion:

$${}_k\lambda_L = -2 \sum_{i=1}^N \log f(x_i | {}_k\hat{\theta}) + 2k \quad (9)$$

Here $f(x_i | {}_k\hat{\theta})$ is the likelihood function of the initial variables x for a set of parameters ${}_k\hat{\theta}$, k - «some ... equivalents, is often called the order of the model» [9].

The following year Akaike published an article "A New Look at Identification of a Statistical Model", where he called the criterion of choosing the best model AIC (Akaike's informational criterion) and presented the formula of this criterion in a convenient form for practical use in the problem of choosing the best autoregressive model ARIMA(p, d, q) [10, p. 720]:

$$AIC = N \log \sigma^2 + 2(p + q) \tag{10}$$

Here N is the number of observations, σ^2 is the variance of the model, p is the autoregressive order AR(p), and q is the model order MA(q).

Akaike's idea - when choosing the best regression model to consider not only the accuracy of approximation but also the complexity of the model, other mathematicians have continued and proposed various modifications of the criterion for choosing the best AIC model. One popular version of this idea was the "Bayesian information criterion" - best model selection criterion proposed by Gideon Schwartz in 1978.

Schwartz did not use the maximum likelihood function: "We therefore assume that observations come from a Koopman-Darmois family, i.e., relative to some fixed measure on the sample space they possess a density of the form

$$f(x, \theta) = \exp(\theta \cdot y(x) - b(\theta)) \tag{11}$$

where θ ranges over the natural parameter space Θ , a convex subset of the K -dimensional Euclidean space, and y is the sufficient K -dimensional statistic...

Via Bayes' formula that is equivalent to choosing the j maximizes

$$S(Y, n, j) = \log \int \alpha_j \exp((Y \circ \theta - b(\theta))n) d\mu_j(\theta) \tag{12}$$

where the integral extend over $m_j \cap \Theta$, and Y is the averaged y -statistic $(1/n) \sum y(X_i)$. [11, p.462].

Schwartz, solving this problem, got another criterion for choosing the best model, which he called «Bayesian information criterion» (BIC):

$$BIC = \ln \sigma^2 + (p + q) \frac{\ln N}{N} \tag{13}$$

The Schwartz BIC criterion is today the primary criterion when using autoregressions in economics and when solving the problem of choosing the best autoregressive model.

Both the Akaike criterion (10) and the Schwartz criterion (13) represent a compromise between the accurate in the past and the least complex model.

The more complex the model, the more accurately it describes the past, but the more coefficients it has. There is a danger that more complex models describe not only essential elements of past economic dynamics but also non-essential elements that acted in the past but will not act in the future. Making a model more complex slightly improves its accuracy, but significantly impairs its predictive properties. An insignificant decrease in the logarithm of variance is compensated by a more noticeable increase in the number of coefficients and leads to an increase in the BIC as a whole. Then BIC of this model in comparison with BIC of the previous simple model turns out to be bigger and the information criterion recommends choosing a simple model.

The procedure for selecting the best model using information criteria is formalized and reduced to selecting the model whose BIC criterion is minimal.

Earlier we expressed a concern that the $VAR^k(p)$ model would be more accurate than the analogous $CVAR^k(p)$ model. And while the information criteria help to make the model choice, the smaller number of coefficients of the complex-valued model may compensate for the lower accuracy of the $CVAR^k(p)$ model's approximation error compared to the $VAR^k(p)$ model.

Let us compare with each other the BIC for each of the vector autoregressive models - real variables and complex ones. Consider the case where the modeled vector comprises an even number of elements.

Then for $VARMA^k(p,q)$ we get the following formula for calculating the information criterion:

$$BIC = \ln \sigma_{VAR}^2 + k^2(p+q) \frac{\ln N}{N} \tag{14}$$

and for $CVARMA^k(p,q)$ (at even k) – such:

$$BIC = \ln \sigma_{CVAR}^2 + \frac{k^2(p+q)}{2} \frac{\ln N}{N} \tag{15}$$

Now we can get an answer to the question of how high should the variance of the complex autoregression be compared to the variance of the vector autoregression of the real variables to concede to it according to the BIC criterion?

Let us first find the condition under which the two criteria are equal to each other. By equating the right-hand sides of (14) and (15), we get:

$$\ln \frac{\sigma_{CVAR}^2}{\sigma_{VAR}^2} = \frac{k^2(p+q)}{2N} \ln N \rightarrow \frac{\sigma_{CVAR}^2}{\sigma_{VAR}^2} = N^{\frac{k^2(p+q)}{2N}} \tag{16}$$

Both models will be equal in terms of BIC criterion when the variance of the complex autoregression $CVAR$ is exactly $N^{\frac{k^2(p+q)}{2N}}$ times larger than the variance of the vector autoregression VAR .

Let us show what this means by a conditional example.

Let's consider a four-dimensional vector ($k=4$) for autoregression with lags $p=2$ and $q=1$ at $N=40$ observations. Then, according to the BIC criterion, a complex autoregressive model will have the same chances to be chosen for modeling if its variance is larger than the variance of VAR in

$$\frac{\sigma_{CVAR}^2}{\sigma_{VAR}^2} = 40^{\frac{4^2(2+1)}{80}} = 40^{0.6} = 9,15 \text{ times.}$$

If the $VARMA^4(2,1)$ model describes past data and has a variance equal to 10%, it will be selected under the BIC information criterion only when the variance of the approximation error of the $CARMA^4(2,1)$ model must be greater than 91.5%. For a VAR to have a chance of being selected under the BIC criterion, the $CVAR$ model must be simply terrible.

Research on many examples has shown that the VAR model more often approximates the data slightly more accurately than the $CVAR$ model - 2 or 3%. We're not talking about 90%. The VAR model has no chance of beating the $CVAR$ model in the information criterion competition.

When the performance vector being modeled contains an odd number of elements, the VAR model has a slightly better chance for BIC selection than with an even number of elements. But here we should talk about VAR variance being several times smaller than CVAR variance, and this is not practically possible.

The situation is not at all as expected: CVAR models are not at all some alternative to VAR models. CVAR models have no alternative in practical application because they will always be better than VAR models by BIC or AIC criteria.

4 Discussion of results

CVAR models, having less of coefficients requiring estimation, will open new perspectives of using a remarkable modeling tool - vector autoregression in practice. Their construction of an example of proper objects of modeling is possible for the researcher who does not possess knowledge of special sections of mathematics.

For these prospects to open for researchers, it is necessary to solve some problems.

The BIC used to select the best autoregressive model will almost always give preference to CVAR. But the family of vector autoregressions includes not only CVAR(p) autoregressions themselves but also autoregressions considering errors, which we will denote as CVARMA^k(p,d,q), where p is the order of autoregressive factors, d is the order of finite differences, with which the original data are reduced to a stationary form, q is the lag order for autoregressions of errors. The order of finite differences does not play any role in our study, and we will not specify it.

Let us write the VARMA^k(p,q) model form [3]:

$$\hat{Y}_t = A_0 + A_1 Y_{t-1} + \dots + A_p Y_{t-p} + M_1 U_{t-1} + \dots + M_q U_{t-q} \tag{17}$$

Here U is the vector of approximation errors, which has the same dimension k as the vector of indices Y, and M_j is a square matrix of coefficients. Regarding the errors ε_t of vector U considered, they represent "white noise" with zero mathematical expectation.

The number q of antecedent vectors of approximation errors U is not equal to the number p of antecedent vectors of predicted indicators Y.

For the two-dimensional case, we write the model (17) in the complex form CVARMA²(1,1):

$$\begin{pmatrix} \hat{y}_{1t} \\ \hat{y}_{2t} \end{pmatrix} = \begin{pmatrix} a_0 & -a_1 \\ a_1 & a_0 \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} + \begin{pmatrix} m_0 & -m_1 \\ m_1 & m_0 \end{pmatrix} \begin{pmatrix} \varepsilon_{1,t-1} \\ \varepsilon_{2,t-1} \end{pmatrix} \tag{18}$$

This model, as compared to the CVAR²(1) model, contains twice as few coefficients, since the coefficients of the matrix M_j appear. The transition from a simple model CVAR^k(p) to a more complex model CVARMA^k(p,q) increases the dimensionality of the problem to be solved (twice as much in the case under consideration). Applying the BIC criterion or some other informational criterion to select the best model, we again encounter the fact that these criteria act as prohibitive barriers to making the model more complex.

Let us check under what conditions the transition from the complex CVAR^k(p) autoregression to its more precise modification CVARMA^k(p,q) is possible.

The BIC for the first model will be calculated:

$$BIC = \ln \sigma_{CVAR}^2 + \frac{k^2 p \ln N}{2 N} \tag{19}$$

For the second model, it is:

$$BIC = \ln \sigma_{CVARMA}^2 + \frac{k^2(p+q)}{2} \frac{\ln N}{N} \quad (20)$$

These two criteria will coincide when the equality is fulfilled:

$$\ln \sigma_{CVAR}^2 + \frac{k^2 p}{2} \frac{\ln N}{N} = \ln \sigma_{CVARMA}^2 + \frac{k^2(p+q)}{2} \frac{\ln N}{N} \quad (21)$$

From

$$\ln \frac{\sigma_{CVAR}^2}{\sigma_{CVARMA}^2} = \frac{k^2}{2} \frac{q}{N} \ln N \rightarrow \frac{\sigma_{CVAR}^2}{\sigma_{CVARMA}^2} = N^{\frac{k^2 q}{2 N}} \quad (22)$$

We use the $CVAR^4(I)$ model. Alternatively, we use $CVARMA^4(I, I)$. How much more accurate should the second model be than the first one at $N=40$? Let us substitute this data into (22) and get that the variance of the $CVARMA^4(I, I)$ model should be 2.09 times smaller than that of $CVAR^4(I)$. Such an increase in accuracy is extremely rare in the domain of simple one-dimensional $ARIMA(p, d, q)$ autoregressions and will also be very rare with vector autoregressions.

The information criterion does not allow us to use models other than $CVAR$ to model complex economic processes.

We will get similar results if we compare a simple vector autoregression model with one step $CVAR^k(I)$ with the same models, but with larger lags $CVAR^k(p)$, where $p > 1$.

5 Conclusions

For wide use in the practice of vector autoregressions it is necessary to use their complex-valued form. This facilitates estimating multiple coefficients of vector autoregressions and allows us to estimate coefficients of these models of different dimensions without difficulty.

When using any information criterion to select the best vector autoregressive model of dimension k , we face the fact that this criterion shows as the best model - its simplest form $CVAR^k(I)$. $CVAR^k(p)$ models and, even more so, $CVARMA^k(p, q)$ models lose the competition with this simple model by information criteria.

Actual objects modeled with vector autoregressions are systems with time-distributed lags $p > 1$ and a complex structure of relationships with the influence of previous errors on the current result. By the meaning of the simulated processes, $CVARMA^k(p, q)$ models should often be the best, but they do not pass the information criterion.

To incorporate the rich toolbox of vector autoregressions into modeling practice, another criterion for selecting the best autoregressive model must be substantiated than the informational criteria that exist today.

References

1. C.A. Sims, Macroeconomics and Reality. *Econometrica* **48**, 1- 48 (1980).
2. H. Lütkepohl, *New Introduction to Multiple Time Series Analysis* (Springer, Berlin, 2005).

3. L. Kilian, H. Lütkepohl, *Structural Vector Autoregressive Analysis* (Cambridge University Press, 2017).
4. D. F. Ahelegbey, *Sparse Graphical Vector Autoregression: A Bayesian Approach: SSRN* (2016).
5. A. Rusyana, N. Tatsara, R. Balqis, S. Rahmi, "Application of Clustering and VARIMA for Rainfall Prediction", in *IOP Conference Series: Materials Science and Engineering*, 428–438 (2020).
6. A. Meimela, S.S. Lestari, I.F. Mahdy, T. Toharudin, B.N. Ruchjana, *J. of Physics: Conference Series*, 55–79 (2021).
7. S. Svetunkov, *Complex-Valued Modeling in Economics and Finance* (Springer Science+Business Media, New York, 2012).
8. S.G. Svetunkov, *Economic forecasting using complex autoregressions* (Polytech-PRESS, St. Petersburg, 2021).
9. H. Akaike, "Information theory as an extension of the maximum likelihood principle", in *Second International Symposium on Information Theory*, 267–281 (Akademiai Kiado, Budapest, 1973).
10. H. Akaike, *IEEE transactions on automatic control* **19(6)**, 716–723 (1974).
11. G. Schwarz, et al., *The annals of statistics* **6(2)**, 461–464 (1978).