

2.1. Autoregressions in economic forecasting

Historically, the logic behind the formation of autoregressive models was as follows.

The researchers faced two different tasks of economic forecasting:

- 1) forecasting the indicator for the future as a continuation of some trend over time or as some factor dependence;
- 2) forecasting of an indicator that does not tend to grow, but fluctuates under the influence of random factors in a certain range.

These two tasks differed in different methods and different forecasting time periods. The first problem was solved with the help of regression-correlation analysis and made it possible to carry out medium and long-term economic forecasts. The second task was solved by calculating different variants of the average value indicator and made it possible to carry out short-term economic forecasts. And today these two tasks represent two independent directions in economic forecasting.

In fact, the task of medium- and long-term economic forecasting is to identify, describe and predict trends and relationships. The task of short-term forecasting is to predict deviations from this general trend or relationships.

The object of our scientific research is precisely the second direction – short-term forecasting.

Since the processes initially studied and predicted in this case had no tendencies to increase or decrease, they began to be called "stationary".

The dominant point of view on how the "stationary process" is understood today is most clearly reflected by such a definition: "In practice, there are often random processes that occur approximately uniformly in time and have the form of continuous random fluctuations around a certain average value, and neither the average amplitude nor the nature of these fluctuations shows significant changes over time. Such random processes are called stationary" (Wentzel, 2010, p. 479). Most often, such processes have a normal probability distribution. And since fluctuations in the stationary process occur relative to a "certain average value", then the natural model of short-term forecasting is this very average value.

The simplest is the arithmetic mean of the last p observations:

$$\bar{y} = \frac{1}{p} \sum_{\tau=0}^{p-1} y_{t-\tau} \quad (2.1.1)$$

Here y_t are the values of a series varying in time t . These future values of the series are the object of forecasting. It is logical to assume that the next observed value of the considered series y_{t+1} will be close to this arithmetic mean, that is, the arithmetic mean acts as a forecast for the next step of observation:

$$\hat{y}_{t+1} = \bar{y} \quad (2.1.2)$$

The main problem faced by forecasters when using the arithmetic mean as a forecasting model was the choice of the averaging period p . This turned out to be especially important for small samples. When choosing different averaging periods, different arithmetic means are obtained, which predicted the series in different ways. Therefore, for different series, the researchers experimentally selected different periods of averaging p . And after choosing this averaging period and obtaining new values, the average (2.1.1) was recalculated taking into account the new values, but keeping the averaging period p unchanged. Since such an average "moved" along the series, it became known as the "moving average".

The next stage in the development of short-term forecasting models was the understanding that the weighted average can be better than the arithmetic mean:

$$\hat{y}_{t+1} = \sum_{\tau=0}^{p-1} b_{\tau} y_{t-\tau}, \quad \sum_{\tau=0}^{p-1} b_{\tau} = 1 \quad (2.1.3)$$

Here, the b_{τ} weights reflect the different contribution of observations to the predicted value of the indicator. An important requirement is that the sum of these weights is equal to one. Only then can we talk about the average. If this rule is violated, the amount will not be the average. Since it is a weighted moving average, it is denoted by $MA(p)$.

Weights b_{τ} can be set in a variety of ways. Scientists used different procedures for this, including weighing forward, backward and in the middle of the averaging segment. By changing these averaging centers, one can get different options for describing the series and change the accuracy of its prediction. To smooth the graph and eliminate the influence of strong random deviations on the overall view of the series, averaging over the sample center can be used. And for economic forecasting, the models whose weights decreased with decreasing indicators in the past turned out to be more accurate, that is, when, in addition to (2.1.3), the following condition was also fulfilled:

$$b_0 > b_1 > \dots > b_{p-1} \quad (2.1.4)$$

In the economic short-term forecasting moving weighted average models began to be successfully used. But in this case, there was one problem that the forecaster faced every time he faced the need to predict a new series – how to choose weights in the moving average (2.1.3) so as to get the best forecast in terms of accuracy? After all, one can set weights as he likes! It was important for scientists to set the weights in such a way that it would become universal. This was the method for calculating the weights of observations as members of a series of an infinite geometric progression (Brown, 1956; Holt, 1957):

$$b_{\tau} = \alpha(1 - \alpha)^{\tau} \quad (2.1.5)$$

There is no need to think about what values each weight takes – they are calculated using this formula. Therefore, it is enough just to set the value α and all the weights b_{τ} are easily calculated.

It is important to set the smoothing constant α , so that the sum of the terms of the series (2.1.4) calculated with the help of (2.1.5) is equal to one.

For an infinite geometric progression, this means that the condition must be met:

$$|1 - \alpha| < 1 \quad (2.1.6)$$

Hence, the boundaries in which the values of the smoothing constant should lie are clear (Svetunkov, 1997):

$$0 < \alpha < 2 \quad (2.1.7)$$

If we substitute in (2.1.3) the weights of b_{τ} , which are calculated using rule (2.1.5) under condition (2.1.7), and assume that the series of observations is infinite, then the mean will take the following form:

$$\hat{y}_{t+1} = \alpha y_t + (1 - \alpha) \hat{y}_t \quad (2.1.8)$$

This forecasting model, called exponential smoothing, turned out to be convenient for calculations due to its simplicity and due to the fact, that by changing the values of the smoothing constant in the range (2.1.7), one can find such a value α , at which the model will give the smallest error in short-term economic forecasting.

Historically, forecasters began to use its truncated band instead of the band of values (2.1.7):

$$0 < \alpha \leq 1 \quad (2.1.9)$$

It was not clear to them what meaning the model acquires beyond these limits. But if the model works within the limits (2.1.9), then its meaning is clear - the model is adaptive to the current information to varying degrees, and the smoothing constant α characterizes the degree of this adaptation - at $\alpha=0$ the model is absolutely not adaptive to the current information, and at $\alpha=1$ - it fully considers the current information, ignoring the past information.

Quite often, when finding the optimal value of the smoothing constant α , scientists and practicing economists were convinced that this optimal value is equal to one. This means that

$$\hat{y}_{t+1} = 1y_t + (1-1)\hat{y}_t = y_t \quad (2.1.10)$$

That is – the current value of the indicator turned out to be the best predictive value. Forecasters began to call this situation the *NAÏVE* model.

In fact, there are almost no situations when the optimal value of the smoothing constant is equal to one ($\alpha=1$). It is just that in such cases the smoothing constant goes beyond (2.1.9) and its optimal value lies in the range

$$1 < \alpha < 2 \quad (2.1.11)$$

In this range, the exponential smoothing model acquires special properties. In order to understand them, we will present under the conditions (2.1.9) the model (2.1.8) in this form:

$$\hat{y}_{t+1} = \hat{y}_t + \alpha(y_t - \hat{y}_t) = \hat{y}_t + \alpha\varepsilon_t \quad (2.1.12)$$

It can be seen from this form of the model representation that it works like this: the forecast value is calculated as the previous calculated value of the indicator, adjusted for the current error ε_t . The degree of this correction is determined by the value of the smoothing constant α . If it is equal to one, the *NAÏVE* model is obtained

In the infinite set of the smoothing constants (2.1.11), when the smoothing constant is greater than one, the model behaves differently:

$$\hat{y}_{t+1} = y_t + (\alpha - 1)(y_t - \hat{y}_t) = y_t + \alpha\varepsilon_t \quad (2.1.13)$$

That is – not the calculated, but the actual value of the predicted indicator is adjusted for the current forecast error ε_t . The model acquires the property of self-learning.

The success of the exponential smoothing model in the short-term economic forecasting was somewhat obscured by the fact that the sum of the weights (2.1.5) will be equal to one only if the number of observations, and, therefore, the number of weighting coefficients is equal to infinity. That is, strictly speaking, the model (2.1.8) is not an average for small and medium samples. But

with a large number of observations, the sum of the series (2.1.5) differs from one by such small values that this disadvantage can be neglected.

The fact that a stationary series can be successfully predicted using a model in which the sum of the weighting coefficients is not equal to one and this model is not some form of average has been known for quite a long time. As far back as 1907, A.A. Markov initiated the formation of an extensive class of stochastic processes with a discrete time component, which were named after him. Markov's processes describe the following probabilistic state of the process depending on the current state. These were the first autoregressive models that have undergone significant development since then. In order to avoid difficulties with indices, the calculated and predicted value of the indicator is referred to the current time t and written as follows:

$$\hat{y}_t = a_1 y_{t-1} + a_2 y_{t-2} + a_3 y_{t-3} + \dots + a_p y_{t-p} = \sum_{\tau=1}^p a_\tau y_{t-\tau} \quad (2.1.14)$$

This model is usually denoted as $AR(p)$, where p is the autoregression order.

It is easy to notice that in the case when $p \rightarrow \infty$, and the coefficients a_τ take the values of the terms of the geometric progression series (2.1.5), the $AR(p)$ model turns into an exponential smoothing model. Consequently, the autoregressive model (2.1.14) is a general model of short-term forecasting, the frequent cases of which are both the exponential smoothing model and the *NAÏVE* model.

The actual values of the predicted indicator differ from the calculated values by some error:

$$y_t - \hat{y}_t = \varepsilon_t \quad (2.1.15)$$

It follows from this that this error, in turn, can be considered as a certain time series that does not tend to increase or decrease its indicators, but is related to the predicted indicator. Therefore, a series of y_t values can be predicted, for example, using the moving average of this error:

$$\hat{y}_t = \sum_{\tau=1}^q b_\tau \varepsilon_{t-\tau}, \quad \sum_{\tau=1}^q b_\tau = 1 \quad (2.1.16)$$

The order of averaging q of this moving average does not necessarily have to coincide with the order of autoregression p , since approximation errors behave differently than the simulated series. This model of the moving average approximation error became known as $MA(q)$. If we remove the restrictions on the equality of the sum of coefficients in the $MA(q)$ model to one, then the moving average will cease to be an average, but will become an autoregression. And autoregression is a more general and more accurate forecasting model than a moving average

model. Today, a non-sliding weighted average (2.1.16) is used, namely, the autoregression of the error, when the sum of the coefficients b_τ is not equal to one. But until now, autoregression, in which the errors ε_t , are the factors, continue to be called the $MA(q)$ model.

The $MA(q)$ model will give different predictive values of the y_t , indicator than the model (2.1.14). In order to take advantage of each model – $AR(p)$ and $MA(q)$ – they were combined into one $ARMA(p, q)$ model:

$$\hat{y}_t + \hat{\varepsilon}_t = \sum_{\tau=1}^p a_\tau y_{t-\tau} + \sum_{\tau=1}^q b_\tau \varepsilon_{t-\tau} \quad (2.1.17)$$

A lot of educational and scientific literature is devoted to the study of the properties of these models. It is important for us that this literature shows the relationship between the coefficients a_τ and b_τ , as well as the interdependence of the order p from the order q . Using this property, a procedure for detecting lags and evaluating the coefficients of the model (2.1.17) was developed, which is called the "Box-Jenkins` methodology" (*Box-Jenkins*).

The main difficulty in practical application of the $ARMA(p, q)$ model is to determine the autoregression order p and the order q . In simple cases, studies of autocorrelation and partial autocorrelation functions can be useful. But in cases where the orders of p and q in the original series are greater than 3, such studies are ineffective.

Today there are several generally accepted methods of constructing models (2.1.17), which can be divided into two groups:

1) simultaneous evaluation of the coefficients a_τ and b_τ of the model (2.1.17), when, gradually increasing the orders of the model p and q and calculating one of the information criteria, the best model is selected according to this criterion;

2) the lag of the $AR(p)$ model is estimated and the coefficients a_τ are calculated, after which the errors ε_t are calculated and the $MA(q)$ model is constructed on their basis. Then, based on the $MA(q)$ model, the coefficients of the $AR(p)$ model are adjusted. Since the $AR(p)$ model obtained at the second step will not always correspond to the original model, a multi-iterative procedure for "debugging" the model is carried out – this is the practical implementation of the Box-Jenkins methodology.

In practice, not every economic process can be classified as stationary.

Therefore, the $ARMA(p, q)$ model cannot always be used directly for economic forecasting. In such non-stationary cases, researchers transform the original series of values so that the resulting

series does not increase or decrease. Most often, this is achieved by calculating the finite differences of the original series. The number of this finite difference is denoted as d and it is included in the name of the model, which, taking into account such transformations, is called *ARIMA* (p, d, q).

The basic *ARIMA* (p, d, q). model today is the basis for the further development of short-term forecasting models that are used to solve various particular problems. Exogenous variables are added to this model and it is denoted as *ARIMAX* (p, d, q, b), autoregression is made nonlinear and denoted as *NARMA* (p, q), the seasonality factor is included in the model and this model is called *SARMA* (p, q) etc.

2.2. Vector autoregressions

No economic indicator develops autonomously. Its dynamics is influenced by many different factors. There are also cases in economy when some indicators change, influencing each other. Such a joint interdependent dynamics was described for some time by a system of simultaneous equations, and later it began to be replaced by vector regressions.

In the case when the vector of indicators at time t is determined by the values of the same vector at previous points in time, it is appropriate to say that such processes are described by vector autoregressions.

Vector autoregression of order p , denoted as *VAR*(p), can be presented in this form (Lütkepohl, 2005, p.13):

$$\hat{Y}_t = A_0 + A_1 Y_{t-1} + A_2 Y_{t-2} + \dots + A_p Y_{t-p}. \quad (2.2.1)$$

Here Y_t is a k -dimensional vector of variables;

A_0 - k -dimensional vector of coefficients;

A_t - $k \times k$ - dimensional constant real matrices.

In the two-dimensional case, *VAR* (I) will take the form:

$$\begin{pmatrix} \hat{y}_{1t} \\ \hat{y}_{2t} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix}. \quad (2.2.2)$$

VAR (I) in the three-dimensional case will look like this:

$$\begin{pmatrix} \hat{y}_{1t} \\ \hat{y}_{2t} \\ \hat{y}_{3t} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \\ y_{3t-1} \end{pmatrix}. \quad (2.2.3)$$

$VAR(1)$ for the four-dimensional case will be:

$$\begin{pmatrix} \hat{y}_{1t} \\ \hat{y}_{2t} \\ \hat{y}_{3t} \\ \hat{y}_{4t} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \\ y_{3t-1} \\ y_{4t-1} \end{pmatrix}. \quad (2.2.4)$$

It is easy to notice that the number of coefficients to be estimated for first-order vector autoregression is equal to k^2 , where k is the size of the vector used. And in general, the number of coefficients to be estimated for $VAR(p)$ will be equal to $(k^2 \cdot p)$. For example, for $VAR(3)$ at $k=5$, it is necessary to estimate $(5^2 \cdot 3=75)$ unknown coefficients. That is, it is necessary to solve a system of 75 linear equations with 75 unknown quantities. It is obvious that the solution of such problems in economic practice is exotic today.

If we compare the one-dimensional autoregression model and the vector autoregression model with each other, we can see that the complexity of the autoregression models $AR(p)$ is determined by the lag order p , and the complexity of the vector autoregression is determined primarily by the size of the vector k , and only then by its lag p . The theory of vector autoregressions has been developed in general, but these autoregressions have not found wide application in practice, since it is necessary to evaluate a large number of coefficients and in publications devoted to the construction of $VAR(p)$ in practice, vectors with a dimension higher than four and a lag greater than two are rarely occurred. Therefore, the dimension of the vector is not indicated anywhere in the model notation, and it is the dimension of the vector, as can be seen from the previous arguments, that has a decisive influence on the model complexity. That is why, when using one or another model of vector autoregression, it is mandatory to indicate the dimension of its vector. Based on these considerations, hereinafter, referring to vector autoregressions, the dimension of the vector in the model designation will be indicated with the symbol k . Consequently, by the abbreviation $VAR^k(p)$ we will denote the autoregression of the order p of a k -dimensional vector, immediately imagining that when using this model in practice it will be necessary to estimate $(k^2 \cdot p)$ coefficients.

2.3. Complex-valued autoregression as a competitor to the $VAR^k(p)$ model

When modeling many economic processes, the use of models and methods of the theory of functions of a complex variable turns out to be no worse, and in some cases better, than models of real variables. For example, when modeling production processes, the production functions of complex variables describe these processes in more detail, and in some cases demonstrate greater accuracy in forecasting than the production functions of real variables (Svetunkov, 2012).

This gave rise to a desire to test the possibility of using complex-valued economics with regard to the economic forecasting problems, in particular, in relation to the tasks of short-term economic forecasting using autoregression models

In general, the complex autoregression model can be written as follows:

$$y_{1t} + iy_{2t} = \sum_{\tau=1}^p F(y_{1(t-\tau)} + iy_{2(t-\tau)}) + (\varepsilon_{1t} + i\varepsilon_{2t}) \quad (2.3.1)$$

Here y_{1t} and y_{2t} are real variables predicted at point of time t ;

i is an imaginary unit, $i = \sqrt{-1}$;

F is some complex-valued function;

τ is the autoregression lag;

p is the autoregression order;

ε_{1t} and ε_{2t} are approximation errors of the first and second variables at point of time t .

Depending on the type of complex-valued function F , complex autoregressions (2.3.1) can be either linear or nonlinear. Nonlinear autoregressive models of real variables are not often found both in practical application and in theoretical research. Therefore, in our study, we will focus our attention on linear autoregressions, and from now on we will understand complex autoregressions to be linear forms of the model (2.3.1), and will denote these models as $CAR(p)$.

Thus, the complex-valued $CAR(p)$ autoregressive models under consideration will be generally represented in this form:

$$y_{1t} + iy_{2t} = (b_0 + ib_1) + \sum_{\tau=1}^p (a_{0\tau} + ia_{1\tau})(y_{1(t-\tau)} + iy_{2(t-\tau)}) + (\varepsilon_{1t} + i\varepsilon_{2t}) \quad (2.3.2)$$

where b_0 and b_1 are coefficients (free terms) reflecting the initial value of the complex series;

a_{0t} and a_{1t} are coefficients of proportionality

Usually, when presenting autoregression models, free terms are omitted, since they can be eliminated by centering the original variables with respect to their arithmetic means. Therefore, we will further assume that the coefficients b_0 and b_1 of the $CAR(p)$ model are equal to zero.

Then the complex autoregression of the first order $CAR(1)$ can be represented either like this:

$$\hat{y}_{1t} + i\hat{y}_{2t} = (a_{01} + ia_{11})(y_{1(t-1)} + iy_{2(t-1)})$$

or in a vector form:

$$\begin{pmatrix} \hat{y}_{1t} \\ \hat{y}_{2t} \end{pmatrix} = \begin{pmatrix} a_{01} & -a_{11} \\ a_{11} & a_{01} \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \end{pmatrix}. \quad (2.3.3)$$

While comparing (2.3.3) with (2.3.2), we can make sure that we are dealing with a special case of 2-dimensional vector autoregression

$VAR^2(1)$. At the same time, for the practical application of the $VAR^2(1)$ model, it is necessary to estimate four unknown coefficients, and for the practical application of the $CAR(1)$ model, only two coefficients are needed. This means that the $VAR^2(1)$ model, with the help of two additional coefficients, will take into account some nuances, and therefore it will describe the initial process more accurately than complex autoregression.

Perhaps this will result in a greater prognostic accuracy of this model compared to the $CAR(1)$ model. But the undeniable advantage of the $CAR(1)$ model is that in order to apply it in practice, it is necessary to estimate only two coefficients, not four.

Can $CAR(p)$ be applied in case when k-dimensional vectors of order greater than three are used? There is such a possibility. For example, for $k=3$ and $p=1$, the following $CAR^3(1)$ model should be used:

$$\begin{cases} \hat{y}_{1t} + i\hat{y}_{2t} = (a_{11} + ia_{12})(y_{1t-1} + iy_{2t-1}) + (b_{11} + ib_{12})y_{3t-1} \\ \hat{y}_{3t} = a_{31}y_{1t-1} + a_{32}y_{2t-1} + a_{33}y_{3t-1} \end{cases} \quad (2.3.4)$$

Or in a vector form:

$$\begin{pmatrix} \hat{y}_{1t} \\ \hat{y}_{2t} \\ \hat{y}_{3t} \end{pmatrix} = \begin{pmatrix} a_{11} & -a_{12} & b_{11} \\ a_{12} & a_{11} & b_{12} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \\ y_{3,t-1} \end{pmatrix}. \quad (2.3.5)$$

In this case, it is necessary to estimate seven unknown coefficients, and in the 3-dimensional model $VAR^3(1)$ (2.3.3) it is necessary to estimate nine unknown coefficients. The efforts saving in estimating the coefficients is obvious.

This saving is even more obvious if we compare the 4-dimensional vector $VAR^4(1)$ (2.3.4) and the $CAR^4(1)$ model. In the classical vector autoregression $VAR^4(1)$ it is necessary to estimate 16 unknown coefficients. And the $CAR^4(1)$ model will be written like this:

$$\begin{pmatrix} y_{1t} + iy_{2t} \\ y_{3t} + iy_{4t} \end{pmatrix} = \begin{pmatrix} a_{11} + ia_{12} & b_{11} + ib_{12} \\ a_{21} + ia_{22} & b_{21} + ib_{22} \end{pmatrix} \begin{pmatrix} y_{1t-1} + iy_{2t-1} \\ y_{3t-1} + iy_{4t-1} \end{pmatrix} \quad (2.3.6)$$

Whence it can be seen that in order to use this model in practice, it is necessary to estimate 8 coefficients, and not 16 coefficients. This is two times less than that of the model $VAR^4(1)$.

If we reduce the $CAR^k(p)$ model to a vector form and compare it with the vector autoregression model, then we can determine the number of coefficients of the $VAR^k(1)$ and $CAR^k(1)$ models of different dimensions k and understand their difference. This has been done in Table 2.1.

Table 2.1.

Coefficients of $VAR^k(1)$ and $CAR^k(1)$ models for different dimensions of vector k

k	$VAR^k(p)$	$CAR^k(p)$	Number of coefficients	
			$VAR^k(p)$	$CAR^k(p)$
2	$a_{11} \quad a_{12}$ $a_{21} \quad a_{22}$	$a_{11} \quad -a_{12}$ $a_{12} \quad a_{11}$	4	2
3	$a_{11} \quad a_{12} \quad a_{13}$ $a_{21} \quad a_{22} \quad a_{23}$ $a_{31} \quad a_{32} \quad a_{33}$	$a_{11} \quad -a_{12} \quad a_{13}$ $a_{12} \quad a_{11} \quad a_{22}$ $a_{31} \quad a_{32} \quad a_{33}$	9	7

4	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}	a_{11} $-a_{12}$ a_{13} $-a_{14}$ a_{12} a_{11} a_{14} a_{13} a_{31} $-a_{32}$ a_{33} $-a_{34}$ a_{32} a_{31} a_{34} a_{33}	16	8
5	a_{11} a_{12} a_{13} a_{14} a_{15} a_{21} a_{22} a_{23} a_{24} a_{25} a_{31} a_{32} a_{33} a_{34} a_{35} a_{41} a_{42} a_{43} a_{44} a_{45} a_{51} a_{52} a_{53} a_{54} a_{55}	a_{11} $-a_{12}$ a_{13} $-a_{14}$ a_{15} a_{12} a_{11} a_{14} a_{13} a_{25} a_{31} $-a_{32}$ a_{33} $-a_{34}$ a_{35} a_{32} a_{31} a_{34} a_{33} a_{45} a_{51} a_{52} a_{53} a_{54} a_{55}	25	17
6	a_{11} a_{12} a_{13} a_{14} a_{15} a_{16} a_{21} a_{22} a_{23} a_{24} a_{25} a_{26} a_{31} a_{32} a_{33} a_{34} a_{35} a_{36} a_{41} a_{42} a_{43} a_{44} a_{45} a_{46} a_{51} a_{52} a_{53} a_{54} a_{55} a_{56} a_{61} a_{62} a_{63} a_{64} a_{65} a_{66}	a_{11} $-a_{12}$ a_{13} $-a_{14}$ a_{15} $-a_{16}$ a_{12} a_{11} a_{14} a_{13} a_{16} a_{15} a_{31} $-a_{32}$ a_{33} $-a_{34}$ a_{35} $-a_{36}$ a_{32} a_{31} a_{34} a_{33} a_{36} a_{35} a_{51} $-a_{52}$ a_{53} $-a_{54}$ a_{55} $-a_{56}$ a_{52} a_{51} a_{54} a_{53} a_{56} a_{55}	36	18
7	a_{11} a_{12} a_{13} a_{14} a_{15} a_{16} a_{21} a_{22} a_{23} a_{24} a_{25} a_{26} a_{31} a_{32} a_{33} a_{34} a_{35} a_{36} a_{41} a_{42} a_{43} a_{44} a_{45} a_{46} a_{51} a_{52} a_{53} a_{54} a_{55} a_{56} a_{61} a_{62} a_{63} a_{64} a_{65} a_{66} a_{71} a_{72} a_{73} a_{74} a_{75} a_{76}	a_{11} $-a_{12}$ a_{13} $-a_{14}$ a_{15} $-a_{16}$ a_{12} a_{11} a_{14} a_{13} a_{16} a_{15} a_{31} $-a_{32}$ a_{33} $-a_{34}$ a_{35} $-a_{36}$ a_{32} a_{31} a_{34} a_{33} a_{36} a_{35} a_{51} $-a_{52}$ a_{53} $-a_{54}$ a_{55} $-a_{56}$ a_{52} a_{51} a_{54} a_{53} a_{56} a_{55} a_{71} $-a_{72}$ a_{73} $-a_{74}$ a_{75} $-a_{76}$ a_{72} a_{71} a_{74} a_{73} a_{76} a_{75}	49	31

The table shows the coefficients of the two models for vectors with dimension up to $k=7$. It makes no sense to continue this table, since the pattern is already obvious – to build a model of complex vector autoregression, it is always necessary to estimate fewer coefficients than for the classical model of vector autoregression. And it is already clear that, for example, to use the $VAR^8(I)$ model, at $k=8$, it will be necessary to estimate 64 unknown coefficients, while for the $CAR^8(I)$ model, only 32 unknown coefficients. For the vector autoregressions with an even vector dimension, to construct a complex autoregression, it is necessary to estimate two times less coefficients than for classical vector autoregression. As far as information criteria are used to select complex models, this circumstance can be decisive when choosing a model, since in these criteria the number of model coefficients significantly affects the value of the criterion. Therefore, it can be concluded from Table 2.1 that the two models under consideration - $VAR^k(p)$ and $CAR^k(p)$ - can really compete with each other only if two- or three-dimensional vector

autoregressions are used. In these cases, an increase in the number of coefficients of the $VAR^k(p)$ model will be compensated by a decrease in the variance of the approximation error or forecasting error, and the information criterion can make a choice in favor of the $VAR^k(p)$ model. But in all other cases, one should expect that the choice will fall on the $CAR^k(p)$ model.

Let us demonstrate this statement with a concrete example.

We have at our disposal 118 data on exchange rates on the Moscow Interbank Currency Exchange from 20.09.2018 to 16.03.2019. This period of time has been chosen deliberately, because during this period the behavior of currencies can well be attributed to a stationary process. We will use the following exchange rates in Russian rubles:

- 1) US dollar, y_{1t} ;
- 2) euro, y_{2t} ;
- 3) pound sterling, y_{3t} ;
- 4) Swiss franc, y_{4t} .

We will sequentially build models $VAR^2(1)$ and $CAR(1)$ for the first two variables (y_{1t}, y_{2t}), $VAR^3(1)$ and $CAR^3(1)$ for the first three variables (y_{1t}, y_{2t}, y_{3t}) and $VAR^4(1)$ and $CAR^4(1)$ for the four-dimensional vector ($y_{1t}, y_{2t}, y_{3t}, y_{4t}$).

Let us start with a simple two-dimensional case. By estimating the coefficients of the $VAR^2(1)$ and $CAR^2(1)$ models on these data using the least squares method, it is possible to obtain the models that, as expected, describe the original data differently. The results of the available data approximation by these models are summarized in Table 2.2.

In this table and in all subsequent tables of this paragraph, the indices i and j indicate the currency numbers in this order: 1 - US dollar, 2 - euro, 3 - pound sterling, 4 - Swiss franc. Therefore, for example, at the intersection of the line denoted by the number 1 and the column denoted by the number 2, the value of the coefficient $a_{12} = 0.026$ of two-dimensional vector autoregression is given. This coefficient shows the influence of the euro exchange rate at the preceding moment on the dollar exchange rate at the present moment.

The last column of the table shows the variances of approximation error for each model as a whole and for each currency type individually. So, the approximation error of the $VAR^2(1)$ model was 0.3695 - this is shown in the first line of the table, and the approximation error of the dollar rate by this model turned out to be 0.1661.

The first part of the table shows data on the $VAR^2(I)$ model, and the second part shows data on the $CAR(I)$ model.

Table 2.2.

Comparative results of currency exchange rate modeling by $VAR^2(I)$ and $CAR(I)$ models

j \ i	Coefficients		σ^2
	1	2	
Model $VAR^2(I)$			0,3695
1	0,9968	0,0026	0,1661
2	0,1409	0,8761	0,2034
Model $CAR(I)$			0,3774
1	0,9996	-0,0001	0,1660
2	0,0001	0,9996	0,2114

A simple comparison of the results shows that, as expected, the $VAR^2(I)$ model turned out to be generally more accurate than the $CAR(I)$ model, but insignificantly: the total variance of vector autoregression was 0.3695, while the total variance of complex autoregression turned out to be higher - 0.3774, which is only 2.15% worse.

The variance of approximation error of the dollar rate using the $VAR^2(I)$ model is almost equal to the variance of approximation error of the dollar rate when using $CAR(I)$. The difference between them was only one ten thousandth: 0.1661-0.1660. And the complex autoregression model turned out to be more accurate.

The euro exchange rate model $VAR^2(I)$ described more accurately: the variance of the approximation error is 0.2034, while for the $CAR(I)$ model, the variance of approximation error of euro rate turned out to be 0.2114, which is 5% more.

Since information criteria are used to select the best model as a compromise between the desire to choose the most accurate model in the approximation and, at the same time, to choose a simpler model, Bayesian information criteria were calculated:

$$BIC = \ln \sigma^2 + \frac{k}{N} \ln(N) \quad (2.3.7)$$

Here N is the number of observations.

The calculated values of the criteria differ from each other

$$-BIC_{VAR^2(I)} = -0,834;$$

$$-BIC_{CAR(I)} = -0,894.$$

The information criteria turned out to be negative, since the variances of approximation errors are less than one and the logarithms of these variances become negative: for the first model, this logarithm is equal to

(-0.9957), and for the second model - (-0.9744). Since four coefficients are used in the $VAR^2(I)$ model, and two coefficients are used in the $CAR(I)$ model, the information criterion for complex autoregression turned out to be less and the preference should be given to it, since the increase in the accuracy of the model with its simultaneous complication turned out to be not so significant as to sacrifice the simplicity of the model.

The results for a three-dimensional vector, $k=3$., turned out to be even more significant for the complex autoregression. They are presented in table 2.3.

Table 2.3.

Comparative results of currency exchange rate modeling by $VAR^3(I)$ and $CAR^3(I)$ models

i j	Coefficients			σ^2
	1	2	3	
Model $VAR^3(I)$				0,7456
1	1,0331	-0,0091	-0,0177	0,1654
2	0,1521	0,8968	-0,0269	0,2023
3	0,2111	-0,1065	0,9307	0,3779
Model $CAR^3(I)$				0,7545
1	1,0149	0,0083	-0,0044	0,1657
2	-0,0083	1,0149	-0,0199	0,2110
3	0,2098	-0,1020	0,9277	0,3779

For the three-dimensional case, the first two variables y_{1t} and y_{2t} are better described by the $VAR^3(I)$ model, and the third variable is described equally well by both the $VAR^3(I)$ model and the $CAR^3(I)$ model. And in general, the variance of vector autoregression 0.7456 is less than the variances of complex-valued autoregression 0.7545. This difference in approximation accuracy is 1.2%.

But for $VAR^3(I)$, the information criterion turned out to be equal to $BIC=0,0703$, and for $CAR^3(I)$ it was significantly less, namely, $BIC=0,0014$

That is, according to the information criterion, the model $CAR^3(I)$ is more preferable than the $VAR^3(I)$ model and it is the model of complex vector autoregression that the researcher will choose the without hesitation.

In the four-dimensional vector autoregression $VAR^4(I)$, the estimates require 16 unknown coefficients, and in the complex-valued autoregression $CAR^4(I)$ - only 8 unknown coefficients. All necessary calculations were performed and the coefficients of these two models were found. The values of these coefficients and the variances of approximation errors by these two models of the initial series are given in Table. 2.4.

Table 2.4.

Comparative results of currency exchange rate modeling by $VAR^4(I)$ and $CAR^4(I)$ models

j \ i	Coefficients				σ^2
	1	2	3	4	
Model $VAR^4(I)$					0,9023
1	1,0717	0,2750	-0,0627	-0,3030	0,1574
2	0,1701	0,9501	-0,0365	-0,0660	0,2010
3	0,2531	0,1768	0,8820	-0,3007	0,3713
4	0,1990	0,2474	-0,0842	0,6291	0,1727
Model $CAR^4(I)$					0,9497
1	1,0114	0,0254	-0,0063	-0,0253	0,1669
2	-0,0254	1,0114	0,0253	-0,0063	0,2092

3	0,1009	0,0302	0,9132	-0,0453	0,3869
4	-0,0302	0,1009	0,0453	0,9132	0,1866

And again, we can make sure that if we use the criterion of the minimum variance of approximation error for selecting the best forecasting model, then the model $VAR^4(I)$ is slightly more accurate over all positions than the $CAR^4(I)$ model. And, in general, the variance of approximation error of the $VAR^4(I)$ model is less than the approximation error of the $CAR^4(I)$ model by 5.3%, which seems to be significant. But the information criterion for vector autoregression is $BIC=0,5440$, and for the complex-valued autoregression it is two times less - $BIC=0,2718$. So, if we use the information criterion to choose the best model, then it will unambiguously indicate that the complication of the model with a simultaneous increase in the number of estimated coefficients does not make sense. Complex-valued autoregression should be used, which describes the vector of changing variables maybe a little worse, but it is much simpler and therefore it should be expected to be more stable in the results of short-term forecasting and more accurate.

Interesting results were obtained by Yevgeny Goltsev, who, at my request, built two first-order vector autoregressions with a dimension $k=8$, namely, $VAR^8(I)$ and $CAR^8(I)$ using the example of the Moscow Exchange economic conditions indices. 8 industry indices are distinguished on it¹:

1) consumer sector y_{1t} ;

2) chemistry and petrochemistry y_{2t} ;

3) finance y_{3t} ;

4) power industry y_{4t} ;

5) metals and mining y_{5t} ;

6) oil and gas y_{6t} ;

7) telecommunications y_{7t} ;

8) transport y_{8t} .

¹ <https://www.moex.com/>

According to the series of changes in these indices from 01.05.2016 to 20.09.2020, the coefficients of these two models were evaluated - $VAR^8(I)$ (64 unknown coefficients) and $CAR^8(I)$ (32 unknown coefficients)

The model $VAR^8(I)$ looks like this:

$$\begin{pmatrix} \hat{y}_{1t} \\ \hat{y}_{2t} \\ \hat{y}_{3t} \\ \hat{y}_{4t} \\ \hat{y}_{5t} \\ \hat{y}_{6t} \\ \hat{y}_{7t} \\ \hat{y}_{8t} \end{pmatrix} = \begin{pmatrix} 0,996 & 0,013 & 0,047 & -0,244 & -0,001 & 0,055 & 0,052 & -0,433 \\ 0,007 & 0,834 & -0,053 & 0,158 & 0,189 & 0,024 & -0,082 & -0,143 \\ 0,007 & -0,098 & 0,941 & 0,167 & 0,103 & 0,030 & -0,075 & -0,139 \\ 0,003 & 0,011 & 0,019 & 0,884 & 0,005 & 0,010 & 0,008 & -0,096 \\ -0,023 & -0,010 & 0,026 & -0,066 & 0,992 & 0,029 & -0,012 & -0,097 \\ 0,010 & -0,009 & 0,088 & -0,376 & 0,067 & 0,989 & 0,079 & -0,115 \\ 0,001 & 0,003 & -0,012 & 0,028 & 0,007 & 0,014 & 0,956 & -0,098 \\ 0,003 & -0,005 & 0,008 & 0,039 & 0,011 & 0,012 & -0,024 & 0,832 \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \\ y_{3,t-1} \\ y_{4,t-1} \\ y_{5,t-1} \\ y_{6,t-1} \\ y_{7,t-1} \\ y_{8,t-1} \end{pmatrix} \quad (2.3.8)$$

It describes the basic data with a standard error $\sigma=430,27$. The information criterion for it is $BIC=13,65$.

The model of complex vector autoregression $CAR^8(I)$ has other coefficients:

$$\begin{pmatrix} \hat{y}_{1t} \\ \hat{y}_{2t} \\ \hat{y}_{3t} \\ \hat{y}_{4t} \\ \hat{y}_{5t} \\ \hat{y}_{6t} \\ \hat{y}_{7t} \\ \hat{y}_{8t} \end{pmatrix} = \begin{pmatrix} 0,992 & 0,003 & -0,023 & -0,003 & 0,017 & -0,012 & -0,042 & 0,005 \\ -0,003 & 0,992 & 0,003 & -0,023 & 0,012 & 0,017 & -0,005 & -0,012 \\ 0,003 & 0,008 & 0,974 & -0,010 & 0,014 & -0,015 & -0,053 & 0,017 \\ -0,008 & 0,003 & 0,010 & 0,974 & 0,015 & 0,014 & -0,017 & -0,053 \\ -0,009 & 0,001 & 0,013 & 0,008 & 1,007 & -0,002 & -0,049 & -0,005 \\ -0,001 & -0,009 & -0,008 & 0,013 & 0,002 & 1,007 & 0,005 & -0,049 \\ -0,001 & -0,002 & -0,013 & 0,006 & 0,007 & -0,008 & 0,963 & 0,012 \\ 0,002 & -0,001 & -0,006 & -0,013 & 0,008 & 0,007 & -0,012 & 0,963 \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \\ y_{3,t-1} \\ y_{4,t-1} \\ y_{5,t-1} \\ y_{6,t-1} \\ y_{7,t-1} \\ y_{8,t-1} \end{pmatrix} \quad (2.3.9)$$

And it describes the basic data somewhat worse. It has a standard error equal to $\sigma=441,94$. But due to the fact that the latter model estimates twice as few coefficients, its information criterion is less than that of the $VAR^8(I)$ model, and is $BIC=12,94$.

Therefore, for short-term forecasting purposes, a preference should be given to the complex autoregression model $CAR^8(I)$.

Based on the data from 09/27/2020 to 04/25/2021, Yevgeny Goltsev with the help of these models made retro forecasts for one-step of observations and compared the prognostic values with the actual ones. The results of this comparison are shown in Table 2.5.

Table 2.5.

Results of the retro forecast of the economic conditions indices of the models (2.3.8) and (2.3.9)

№	Industry Index	Standard error of retro forecast, σ		Improved accuracy %
		$VAR^8(I)$ (2.3.8)	$CAR^8(I)$ (2.3.9)	
1.	Oil and gas	331,26	267,52	21,29
2.	Finance	344,25	307,14	11,39
3.	Metal and mining	321,45	295,21	8,51
4.	Power industry	89,63	54,09	49,46
5.	Consumer	177,91	164,85	7,62
6.	Chemistry and petrochemistry	331,38	372,46	-11,67
7.	Transport	45,16	36,96	19,98
8.	Telecom	64,20	36,42	55,23

Only for one index, namely, the Chemistry and Petrochemistry index of Russia, the retro forecast, with the help of complex autoregression, turned out to be worse than the forecast performed by using vector autoregression. All other indices of the Moscow Exchange are predicted better using the $CAR^8(I)$ (2.3.9) model, and in the case of Power Industry and Telecom indices, the forecasting accuracy is even twice as high.

We took the simplest and most accessible data as an example. And we found that to predict this series of values, BIC always recommends to use complex-valued autoregression, rather than vector autoregression. This means that in most practical cases complex-valued autoregression will be preferable to vector autoregression. And the greater the dimension of the vectors and the higher the order of autoregression, the lower the value will be taken by the BIC criterion for this model in comparison with vector autoregression.

This means that the complex-valued autoregression model is a competing vector autoregression model, and a significant reduction in the complexity of new models compared to vector autoregression models allows them to be widely used in economic practice.

2.4. $CARMA^k(p,q)$ in economic forecasting

We have seen that the $CAR^k(p)$ model is a good alternative to the vector autoregression models – they are simpler, contain fewer coefficients that need to be estimated, and they do not

significantly lose in approximation accuracy to the $VAR^k(p)$ models. That is why the information criterion for the models considered in the previous paragraph recommends using a complex autoregression model, rather than a vector one.

Earlier it was shown how the $AR(p)$ autoregression model was transformed into the ARMA (p, q) model. The same logic was used by scientists to form a more complex $VARMA^k(p, q)$ model based on the vector autoregression model:

$$\hat{Y}_t = A_0 + A_1 Y_{t-1} + \dots + A_p Y_{t-p} + M_1 U_{t-1} + \dots + M_q U_{t-q}. \quad (2.4.1)$$

Here U is the vector of approximation errors which has the same dimension k as the vector of indicators Y . In the general case, the number q of preceding vectors of approximation error U is not equal to the number p of preceding vectors of predicted indicators Y .

At the same time, as in the case of $ARMA(p, q)$, a relationship is assumed between $VAR^k(p)$ and $MA^k(q)$ (Lütkepohl, 2005, p. 436).

In the two-dimensional case, the model $VARMA^2(1, 1)$, taking into account the notations we have adopted, will be written as follows (Lütkepohl, 2005, p. 443):

$$\begin{pmatrix} \hat{y}_{1t} \\ \hat{y}_{2t} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} \varepsilon_{1,t-1} \\ \varepsilon_{2,t-1} \end{pmatrix} \quad (2.4.2)$$

Concerning the errors of the ε_{it} vector U , they are considered to be "white noise" with zero mathematical expectation. In order to use this model in practice, it is necessary to estimate its eight coefficients: 4 coefficients of the component $VAR^2(1)$ and 4 coefficients of the second component $MA^2(1)$.

Considering the vector complex autoregression model as a special case of vector autoregression, one can obtain a similar (2.4.1) $CARMA(p, q)$ model. We will not consider the option when the component $MA(q)$ is represented as a simple k -dimensional vector. We will also consider this component in a complex form. Then, for the two-dimensional case, $CARMA(1, 1)$ will be written as:

$$\begin{pmatrix} \hat{y}_{1t} \\ \hat{y}_{2t} \end{pmatrix} = \begin{pmatrix} a_0 & -a_1 \\ a_1 & a_0 \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} + \begin{pmatrix} b_0 & -b_1 \\ b_1 & b_0 \end{pmatrix} \begin{pmatrix} \varepsilon_{1,t-1} \\ \varepsilon_{2,t-1} \end{pmatrix} \quad (2.4.3)$$

Or, in a complex form:

$$\hat{y}_{1t} + i\hat{y}_{2t} = (a_0 + ia_1)(y_{1,t-1} + iy_{2,t-1}) + (\varepsilon_{1t} + i\varepsilon_{2t}) + (b_0 + ib_1)(\varepsilon_{1,t-1} + i\varepsilon_{2,t-1}) \quad (2.4.4)$$

Here, as can be seen, it is necessary to estimate the values of only 4 unknown coefficients, and not 8 coefficients, as is required for the practical application of the vector autoregression model (2.4.2).

In order to estimate the practical acceptability of the $CARMA(p,q)$ model, let us consider a simple case of constructing the $VARMA^4(2,1)$ model, which will be written as follows:

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

It is easy to see that for the practical use of this model, it is necessary to estimate 48 unknown coefficients from statistical data. This is not an easy task for practicing economists, although the four-dimensional case of the economic vector seems to be a fairly simple reflection of real economic situations. Therefore, despite the simplicity of the model, few practicing forecasters will dare to build such a model and use it in practice.

But a similar to it model $CARMA^4(2,1)$, in which not only the variables are presented in a complex form, but the error vector is also presented in a complex form, will be written as follows:

$$\begin{pmatrix} \hat{y}_{1t} \\ \hat{y}_{2t} \\ \hat{y}_{3t} \\ \hat{y}_{4t} \end{pmatrix} = \begin{pmatrix} a_{11} & -a_{12} & a_{13} & -a_{14} \\ a_{12} & a_{11} & a_{14} & a_{13} \\ a_{31} & -a_{32} & a_{33} & -a_{34} \\ a_{32} & a_{31} & a_{34} & a_{33} \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \\ y_{3,t-1} \\ y_{4,t-1} \end{pmatrix} + \begin{pmatrix} c_{11} & -c_{12} & c_{13} & -c_{14} \\ c_{12} & c_{11} & c_{14} & c_{13} \\ c_{31} & -c_{32} & c_{33} & -c_{34} \\ c_{32} & c_{31} & c_{34} & c_{33} \end{pmatrix} \begin{pmatrix} y_{1,t-2} \\ y_{2,t-2} \\ y_{3,t-2} \\ y_{4,t-2} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \\ \varepsilon_{4t} \end{pmatrix} + \begin{pmatrix} b_{11} & -b_{12} & b_{13} & -b_{14} \\ b_{12} & b_{11} & b_{14} & b_{13} \\ b_{31} & -b_{32} & b_{33} & -b_{34} \\ b_{32} & b_{31} & b_{34} & b_{33} \end{pmatrix} \begin{pmatrix} \varepsilon_{1,t-1} \\ \varepsilon_{2,t-1} \\ \varepsilon_{3,t-1} \\ \varepsilon_{4,t-1} \end{pmatrix}$$

In order to use this model in the practice of short-term economic forecasting, it is necessary to estimate only 24 coefficients - two times less.

And this task can be solved by almost any economic forecaster.

The number of coefficients of the $VARMA^k(p,q)$ model, which should be estimated on the available set of values, depends both on the orders p and q of the model (2.4.1) and on the dimension k of the vector of variables, and it is the dimension of the vector k that plays a decisive role in increasing the complexity of the model. The total number of coefficients of such a model can be described by the formula: $k^2 \cdot (p+q)$. The number of coefficients of the $CARMA^k(p,q)$ model will always be less than that of the $VARMA^k(p,q)$ model. And for an even number of vector dimension k , the number of these coefficients will always be two times less and equal to $k^2 \cdot (p+q)/2$. For example, 192 unknown coefficients should be estimated for the $VARMA^4(3,4)$ model, and 86 coefficients for the $CARMA^4(3,4)$ model.

This means that the higher the dimension of the vector in a vector autoregression, the less chance the vector autoregression has of being better compared to a complex autoregression of the same order, since the information criteria will always recommend models with fewer coefficients number.

Let us define, for example, the condition under which the forecaster is hesitant about which model to prefer - $VARMA^4(2,1)$ or $CARMA^4(2,1)$?

These models have been given above in full form. The situation of choice is possible only if the information criteria calculated for each model turn out to be equal to each other, that is, it is fulfilled:

$$\ln \sigma_{VAR}^2 + \frac{48}{N} \ln N = \ln \sigma_{CAR}^2 + \frac{24}{N} \ln N \quad (2.4.5)$$

Hence, we have for the variances:

$$\ln \frac{\sigma_{CAR}^2}{\sigma_{VAR}^2} = \frac{24}{N} \ln N \rightarrow \frac{\sigma_{CAR}^2}{\sigma_{VAR}^2} = N^{\frac{24}{N}} \quad (2.4.6)$$

In the previous paragraph, 118 observations were used for calculations. Let us substitute this number $N=118$ в (2.4.6). Then we get that the variance of $CARMA^4(2,1)$ should be 2.64 times greater than the variance of $VARMA^4(2,1)$, that is, the variance of the approximation error of the complex autoregression $CARMA^4(2,1)$ should exceed the variance of the approximation error of the $VARMA^4(2,1)$ model by 264%! Obviously, such a situation is unlikely to be encountered in practice.

We draw an unambiguous conclusion from this: $CARMA^k(p,q)$ models will almost always be preferable to $VARMA^k(p,q)$ models at $k>3$. But even in the case when $k=2$, the complex autoregression has a good chance of becoming a better model than the vector autoregression model, which is demonstrated by the example in Table 2.2, when simple $VAR^2(1)$ and $CAR(1)$ models were considered without the $MA(q)$. component.