# "Plug-in" Statistical Forecasting of Vector Autoregressive Time Series with Missing Values

Yuriy Kharin and Aliaksandr Huryn[1]
Belarusian State University, Minsk

**Abstract:** The problems of statistical forecasting of vector autoregressive time series with missing values are considered. The maximum likelihood forecast is constructed and its mean square risk is evaluated for the case of known parameters. The "plug-in" forecast and statistical estimators are constructed for unknown parameters. Asymptotic properties of constructed estimators are analyzed. Results of numerical experiments are presented.

**Keywords:** Forecasting, Vector Autoregression, Missing Values, Estimators of Parameters.

## 1  Introduction

Missing values are a typical distortion of model assumptions in data analysis (Little and Rubin, 1987; Stockinger and Dutter, 1987). It is fairly common for a time series to have gaps for a variety of reasons (Greene, 2000; Shafer, 1997): 1) the data do not exist at the frequency we wish to observe them; 2) registration errors; 3) deletion of "outliers".

Vector autoregression (VAR) is often used in practice for statistical analysis (statistical estimation of parameters, statistical testing of hypotheses, statistical forecasting) of time series in econometrics (Pantula and Shin, 1993), biometrics (Beran et al., 1998), technometrics (Jones, 1980) and in many other applications.

If the parameters of the VAR-model are a priori known, then the ML-forecast (Kharin2 and Huryn, 2003) can be used. In practice, the parameters are usually unknown, and there are three approaches to forecasting in this prior uncertainty: 1) joint maximum likelihood estimation of the future value of the considered time series and the parameters; 2) application of the EM-algorithm (Little and Rubin, 1987); 3) "plug-in" approach, consisting of two stages: a) estimation of the parameters by some admissible approach; b) calculation of the forecast putting the estimates of the parameters into the ML-forecast.

The first approach is characterized by significant computational complexity; the second approach, and the first approach also, suffice from the multimodality of the objective function (because of the difficult problem of detection of the main maximum for the situation with many local ones). To avoid these difficulties we develop here the third approach.

In Section 2 we define the underlying model of time series and formulate main assumptions on probabilistic characteristics of the model and on the "missing patterns". Section 3 is devoted to the ML-forecasting under missing values for two levels of prior uncertainty: 1) parameters of the VAR model are known; 2) the parameters are unknown. In Section 4 we construct statistical estimators of the parameters $B, G$, "plug-in" forecasting procedure, and also analyze asymptotic properties of the constructed estimators. Section 5 contains some numerical results using real statistical data.

## 2  Mathematical Model

Let the observed $d$-vector time series $Y_t$ be described by the VAR(1) model:

$$Y_t = BY_{t-1} + U_t, \ t \in \mathbb{Z}. \tag{1}$$

where $\mathbb{Z}$ is the set of integers, $Y_t = (Y_{t1}, \ldots, Y_{td})' \in \mathbb{R}^d$, $B = (B_{kl}) \in \mathbb{R}^{d\times d}$ is a matrix of unknown coefficients (all eigenvalues of the matrix $B$ are assumed to be inside the unit circle), $U_t = (U_{t1}, \ldots, U_{td})' \in \mathbb{R}^d$, $\{U_t\}$ are i.i.d. random vectors, $\mathbf{E}\{U_t\} = 0_d$ is the zero $d$-vector, $\mathbf{E}\{U_t U_t'\} = \Sigma = \mathrm{Const}_t$, $|\Sigma| \neq 0$, where by $\mathrm{Const}_{i_1,\ldots,i_n}$ we will denote a mathematical object (variable, vector, matrix, etc.) that is independent of the variables $i_1, \ldots, i_n$, $n \in \mathbb{N}$. There are missing values in observations $\{Y_t\}$. For each vector $Y_t$ the binary vector (called "missing pattern") $O_t = (O_{t1}, \ldots, O_{td})' \in \{0,1\}^d$ is given, where $O_{ti} = \{1,$ if $Y_{ti}$ is observed; $0,$ if $Y_{ti}$ is a missing value$\}$. Define the discrete set $M = \{(t,i), \ t \in \mathbb{Z}, \ i \in \{1,\ldots,d\} : O_{ti} = 1\}$; its elements are assumed to be lexicographically ordered in ascending order; $K = |M|$ is the total number of observed components; $t_- = \min\{t : \sum_{i=1}^d O_{ti} > 0\}$ is the minimal time moment with observed components, $t_+ = \max\{t : \sum_{i=1}^d O_{ti} > 0\}$ is the maximal time moment with observed components. Without loss of generality assume $t_- = 1, t_+ = T$.

Note, that AR(p)-model and VAR(p)-model can be transformed to VAR(1)-model increasing the number of components (Anderson, 1971).

Let us introduce some assumptions on the innovation process $U_t$ and "missing patterns" $\{O_t\}$.

**A1.** The moments of the orders three and four for the innovation process are bounded: $|\mathbf{E}\{U_{ti_1} \cdots U_{ti_3}\}| \leq \mathrm{Const}_{t,i_1,\ldots,i_3}$, $|\mathbf{E}\{U_{ti_1} \cdots U_{ti_4}\}| \leq \mathrm{Const}_{t,i_1,\ldots,i_4}$, $t \in \mathbb{Z}$, $i_1, \ldots, i_4 \in \{1,\ldots,d\}$.

**A2.** The moment of the fourth order for the innovation process is independent of the time moment and the moments of the orders from 5 to 8 for the innovation process are bounded:
$\mathbf{E}\{U_{ti_1} \cdots U_{ti_4}\} = \Sigma^{(4)}_{i_1,\ldots,i_4} = \mathrm{Const}_t$, $|\mathbf{E}\{U_{ti_1} \cdots U_{ti_5}\}| \leq \mathrm{Const}_{t,i_1,\ldots,i_5}$,
$|\mathbf{E}\{U_{ti_1} \cdots U_{ti_6}\}| \leq \mathrm{Const}_{t,i_1,\ldots,i_6}$, $|\mathbf{E}\{U_{ti_1} \cdots U_{ti_7}\}| \leq \mathrm{Const}_{t,i_1,\ldots,i_7}$,
$|\mathbf{E}\{U_{ti_1} \cdots U_{ti_8}\}| \leq \mathrm{Const}_{t,i_1,\ldots,i_8}$, $t \in \mathbb{Z}$, $i_1, \ldots, i_8 \in \{1,\ldots,d\}$.

**A3.** $U_t$ is a Gaussian random vector: $\mathcal{L}\{U_t\} = \mathcal{N}_d(0_d, \Sigma)$.

**A4.** The "missing patterns" $\{O_t\}$ satisfy the asymptotics at $T \to \infty$ $(i,j \in \{1,\ldots,d\})$:

$$T^{-1} \sum_{t=1}^T O_{ti} O_{tj} \to \nu^{(0)}_{ij} \in (0,1], \ (T-1)^{-1} \sum_{t=1}^{T-1} O_{t+1,i} O_{tj} \to \nu^{(1)}_{ij} \in (0,1], \tag{2}$$

$\nu^{(0)}_{ij}$ is the limit frequency for the pair of components $(i,j)$ observed at the same time moment, $\nu^{(1)}_{ij}$ is the limit frequency for the pair of components $(i,j)$ observed at the neighbor time moments.

**A5.** The "missing patterns" $\{O_t\}$ satisfy the asymptotics at $T \to \infty$:

$$(T - |\tau| - 1)^{-1} \sum_{t,t'=1}^{T-1} O_{ti} O_{tj} O_{t'i'} O_{t'j'} \delta_{t-t',\tau} \to \nu^{(2)}_{i,j,i',j'}(\tau) \in [0,1],$$

$\nu_{i,j,i',j'}^{(2)}(\tau)$ is the limit frequency for the pair of components $(i, j)$ observed together with the pair $(i', j')$ at the delay $\tau$, $\nu_{i,j,i,j}^{(2)}(0) \equiv \nu_{ij}^{(0)}$, $\delta_{i,j}$ is the Kronecker symbol;

$$(T - |\tau| - 1)^{-1} \sum_{t,t'=1}^{T-1} O_{ti} O_{tj} O_{t'+1,i'} O_{t'j'} \delta_{t-t',\tau} \to \nu_{i,j,i',j'}^{(3)}(\tau) \in [0, 1].$$

$\nu_{i,j,i',j'}^{(3)}(\tau)$ is the limit frequency for the pair of components $(i, j)$ observed together with the components $(i', j')$ observed at the delay $\tau - 1$ and $\tau$;

$$(T - |\tau| - 1)^{-1} \sum_{t,t'=1}^{T-1} O_{t+1,i} O_{tj} O_{t'+1,i'} O_{t'j'} \delta_{t-t',\tau} \to \nu_{i,j,i',j'}^{(4)}(\tau) \in [0, 1].$$

$\nu_{i,j,i',j'}^{(4)}(\tau)$ is the limit frequency for the pair of components $(i, j)$ observed at the neighbor time moments and the pair $(i', j')$ observed at the neighbor time moments at the delay $\tau$, $\nu_{i,j,i,j}^{(4)}(0) \equiv \nu_{ij}^{(1)}$, where $\tau \in \mathbb{Z}$, $i, j, i', j' \in \{1, \ldots, d\}$.

Let $Y_{T+\tau} \in \mathbb{R}^d$ be a "future vector" to be forecasted for $\tau \geq 1$, $\dot{Y}_{T+\tau} = \hat{Y}_{T+\tau}(X) : \mathbb{R}^K \to \mathbb{R}^d$ be a forecasting statistic (procedure). Introduce the matrix risk $R \in \mathbb{R}^{d \times d}$ and the (scalar) risk $r$ of forecasting:

$$R = \mathbf{E}\left\{ \left( \dot{Y}_{T+\tau} - Y_{T+\tau} \right) \left( \dot{Y}_{T+\tau} - Y_{T+\tau} \right)' \right\}, r = \mathrm{tr}(R) \geq 0.$$

It is known (Greene, 2000), that for the case of complete observations and known parameters $B$, $\Sigma$ the minimal risk $r_0^* = \mathrm{tr}\left( \sum_{i=0}^{\tau-1} B^i \Sigma (B')^i \right) > 0$ is attained for the forecast (Greene, 2000): $\dot{Y}_{T+\tau} = B^\tau Y_T$.

## 3  ML-Forecasting Under Missing Values

Define a bijection $M \leftrightarrow \{1, \ldots, K\} : k = \chi(t, i)$ and the inverse function $(t, i) = \chi(k)$. Compose the $K$-vector of all observed components: $X = (X_1, \ldots, X_K)' \in \mathbb{R}^K$, $X_k = Y_{\chi(k)}$, $k \in \{1, \ldots, K\}$. Note, that if $O_{ti} = 1$, $t \in \{1, \ldots, T\}$, $i \in \{1, \ldots, d\}$, then the process $Y_t$ is observed on $[1, T]$ without any missing value, $K = Td$, $X = (Y_1', \ldots, Y_T')'$, $\chi(t, i) = i + (t - 1)d$; $\chi(k) = ([(k - 1)/d] + 1, (k - 1) \mod d + 1)$, $k \in \{1, \ldots, K\}$. Denote the matrices: $F = (F_{ij}) = \mathrm{cov}\{X, X\} \in \mathbb{R}^{K \times K}$, $H = (H_{ij}) = \mathrm{cov}\{X, Y_{T+\tau}\} \in \mathbb{R}^{K \times d}$, $G = (G_{ij}) = \mathrm{cov}\{Y_{T+\tau}, Y_{T+\tau}\} \in \mathbb{R}^{d \times d}$, $G_1 = \mathrm{cov}\{Y_{T+\tau+1}, Y_{T+\tau}\} \in \mathbb{R}^{d \times d}$, $A_0 = A_0(B, \Sigma) = H'F^{-1} \in \mathbb{R}^{d \times K}$. Note, that according to Anderson (1971) $G = \sum_{i=0}^{\infty} B^i \Sigma (B')^i$, $G_1 = BG$.

**Lemma 1** *Let the model (1) take place. The following expressions for $F$, $H$ hold:*

$$F_{ij} = F_{ji} = \left( B^{\chi_1(i) - \chi_1(j)} G \right)_{\chi_2(i), \chi_2(j)}, \ i, j \in \{1, \ldots, K\}, \ i \geq j; \qquad (3)$$

$$H_{ij} = \left( B^{(T+\tau) - \chi_1(i)} G \right)_{j, \chi_2(i)}, \ i \in \{1, \ldots, K\}, \ j \in \{1, \ldots, d\}.$$

**Proof.** Using the expression for the covariance matrix for the VAR(1) model (Anderson, 1971): $\mathrm{cov}\{Y_i, Y_j\} = B^{i-j}G$, $i \geq j$, we find covariances: $F_{ij} = \mathrm{cov}\{X_i, X_j\} = \mathrm{cov}\{Y_{\chi(i)}, Y_{\chi(j)}\} = (\mathrm{cov}\{Y_{\chi_1(i)}, Y_{\chi_1(j)}\})_{\chi_2(i), \chi_2(j)} = (B^{\chi_1(i)-\chi_1(j)}G)_{\chi_2(i), \chi_2(j)}$. In the same way we find $H$. $\square$

**Theorem 1** *Let the model (1) and the assumption A3 take place. If the true values $B, \Sigma$ are known, and $|F| \neq 0$, then the ML-forecasting statistic and its risk functionals are*

$$\hat{Y}_{T+\tau,\mathrm{ML}} = \mathbf{E}\{Y_{T+\tau}|X\} = A_0 X. \tag{4}$$

$$R_{\mathrm{ML}} = G - H'F^{-1}H \succeq 0, r_{\mathrm{ML}} = \mathrm{tr}(G) - \mathrm{tr}(F^{-1}HH'). \tag{5}$$

**Proof.** Denote $Y_+ = (X', Y_{T+\tau}^T)' \in \mathbb{R}^{K+d}$. By Theorem 1 assumptions, the vector $Y_+$ has the Gaussian distribution. By the Anderson theorem (Anderson, 1971), the likelihood function(w.r.t $Y_{T+\tau}$) is

$$l(Y_{T+\tau}; B, \Sigma) = n_K(X|0_K, F) \, n_d(Y_{T+\tau}|H'F^{-1}X, G - H'F^{-1}H). \tag{6}$$

where $n_K(X|\mu, \Sigma)$ means the $K$-dimensional Gaussian p.d.f. with the parameters $\mu, \Sigma$. The ML-forecast is the solution of the extremum problem: $l(Y_{T+\tau}; B, \Sigma) \to \max_{Y_{T+\tau}}$. Since the first multiplier in (6) does not depend on $Y_{T+\tau}$, we come to the unique solution (4): $\hat{Y}_{T+\tau,\mathrm{ML}} = H'F^{-1}X = A_0 X$. Using the total mathematical expectation formula and (4), we find the risk (5): $R_{\mathrm{ML}} = \mathbf{E}\left\{\left(\hat{Y}_{T+\tau,\mathrm{ML}} - Y_{T+\tau}\right)\left(\hat{Y}_{T+\tau,\mathrm{ML}} - Y_{T+\tau}\right)'\right\} = \mathbf{E}\left\{\mathbf{E}\left\{(\mathbf{E}\{Y_{T+\tau}|X\} - Y_{T+\tau})(\mathbf{E}\{Y_{T+\tau}|X\} - Y_{T+\tau})'|X\right\}\right\} = \mathbf{E}\{\mathrm{cov}\{Y_{T+\tau}, Y_{T+\tau}|X\}\} = \mathbf{E}\{G - H'F^{-1}H\} = G - H'F^{-1}H. \square$

**Theorem 2** *Let the model (1) and the assumption A3 take place. If $B, \Sigma$ are unknown, $|F| \neq 0$, then the ML-forecast of $Y_{T+\tau}$ has the "plug-in" form:*

$$\tilde{Y}_{T+\tau,\mathrm{ML}} = A_0(\tilde{B}, \tilde{\Sigma})X, \tag{7}$$

*where the ML-estimators $\tilde{B}, \tilde{\Sigma}$ of the model parameters are the solution of the minimization problem: $l_1(B, \Sigma) = X'F^{-1}X + \ln|F| + \ln|G - H'F^{-1}H| \to \min_{B, \Sigma}$.*

   **Proof.** According to the equation (6), the joint ML-estimators of $Y_{T+\tau}$, $B, \Sigma$ are the solution of the extremum problem: $l(Y_{T+\tau}; B, \Sigma) \to \max_{Y_{T+\tau}, B, \Sigma}$. From Theorem 1 we get (7), where the ML-estimators $\tilde{B}, \tilde{\Sigma}$ of the model parameters are the solution of the problem: $n_K(X|0_K, F) \, n_d(H'F^{-1}X|H'F^{-1}X, G - H'F^{-1}H) \to \max_{B, \Sigma}$. Taking the logarithm, we come to the statement. $\square$

# 4   "Plug-in" Forecasting in the Case of Unknown $B, \Sigma$

Because of computational complexity of the minimization problem in (7), we propose to construct more suitable estimator for $A_0$ instead of $A_0(\tilde{B}, \tilde{\Sigma})$ in (7).

Let us define the minimal admissible observation time for the observed time series $\{Y_t\}$: $T_0 = \min\left\{T \in \mathbb{N} : \min_{i,j}\sum_{t=1}^{T}O_{ti}O_{tj} > 0, \min_{i,j}\sum_{t=1}^{T-1}O_{t-1,i}O_{tj} > 0\right\}$, and matrices $\hat{G}, \hat{G}_1 \in \mathbb{R}^{d \times d}$ for $T \geq T_0$:

$$\hat{G}_{ij} = \frac{\sum_{t=1}^{T}Y_{ti}Y_{tj}O_{ti}O_{tj}}{\sum_{t=1}^{T}O_{ti}O_{tj}}, \quad \left(\hat{G}_1\right)_{ij} = \frac{\sum_{t=1}^{T-1}Y_{t-1,i}Y_{tj}O_{t-1,i}O_{tj}}{\sum_{t=1}^{T-1}O_{t+1,i}O_{tj}}, \, i,j \in \{1,\dots,d\}. \quad (8)$$

It follows from the underlying model (1), that $|G| \neq 0$ and the matrices $B, G, G_1$ satisfy the matrix equation $G_1 = BG$, consequently $B = G_1 G^{-1}$. Following the "plug-in" principle and using the previous equation let us construct a matrix statistic (if $|\hat{G}| \neq 0$):

$$\hat{B} = \hat{G}_1(\hat{G})^{-1}. \quad (9)$$

Putting then the statistics $\hat{B}, \hat{G}$ (instead of $B, G$) into (3) we get the matrices $\hat{F}, \hat{H}$, the statistic (if $|\hat{F}| \neq 0$)

$$\hat{A}_0 = \hat{H}'\hat{F}^{-1},$$

and the "plug-in" forecasting procedure:

$$\hat{Y}_{T+\tau,\text{plug-in}} = \hat{A}_0 X.$$

According to Lemma 1, theorems 1 and 2, the performance of the "plug-in" forecasting procedure is determined by the performance of the underlying estimators $\hat{B}, \hat{G}, \hat{G}_1$. Let us analyze asymptotic properties ($T \to \infty$) of the proposed estimators (8), (9). Denote functions ($t, t', \tau \in \mathbb{Z}, T \in \mathbb{N}, i, j, i', j' \in \{1,\dots,d\}$) generated by the "missing patterns":

$$c^{(1)}_{\tau,i,j,i',j'}(T) = T\sum_{t,t'=1}^{T-1}O_{ti}O_{tj}O_{t'i'}O_{t'j'}\delta_{t-t',\tau}\left(\sum_{t=1}^{T}O_{ti}O_{tj}\sum_{t=1}^{T}O_{t'i'}O_{t'j'}\right)^{-1},$$

$$c^{(2)}_{\tau,i,j,i',j'}(T) = T\sum_{t,t'=1}^{T-1}O_{ti}O_{tj}O_{t'+1,i'}O_{t'j'}\delta_{t-t',\tau}\left(\sum_{t=1}^{T}O_{ti}O_{tj}\sum_{t=1}^{T-1}O_{t+1,i'}O_{tj'}\right)^{-1},$$

$$c^{(3)}_{\tau,i,j,i',j'}(T) = T\sum_{t,t'=1}^{T-1}O_{t+1,i}O_{tj}O_{t'i'}O_{t'j'}\delta_{t-t',\tau}\left(\sum_{t=1}^{T-1}O_{t+1,i}O_{tj}\sum_{t=1}^{T}O_{t'i'}O_{tj'}\right)^{-1},$$

$$c^{(4)}_{\tau,i,j,i',j'}(T) = T\sum_{t,t'=1}^{T-1}O_{t+1,i}O_{tj}O_{t'+1,i'}O_{t'j'}\delta_{t-t',\tau}\left(\sum_{t=1}^{T-1}O_{t+1,i}O_{tj}\sum_{t=1}^{T-1}O_{t+1,i'}O_{tj'}\right)^{-1},$$

$$C^{(1)}_{\tau,i,j,i',j'} = \nu^{(2)}_{i,j,i',j'}(\tau)\left(\nu^{(0)}_{ij}\nu^{(0)}_{i'j'}\right)^{-1}, \quad C^{(2)}_{\tau,i,j,i',j'} = \nu^{(3)}_{i,j,i',j'}(\tau)\left(\nu^{(0)}_{ij}\nu^{(1)}_{i'j'}\right)^{-1},$$

$$C^{(3)}_{\tau,i,j,i',j'} = \nu^{(3)}_{i',j',i,j}(-\tau)\left(\nu^{(0)}_{i'j'}\nu^{(1)}_{ij}\right)^{-1}, \quad C^{(4)}_{\tau,i,j,i',j'} = \nu^{(4)}_{i,j,i',j'}(\tau)\left(\nu^{(1)}_{ij}\nu^{(1)}_{i'j'}\right)^{-1},$$

and covariances:

$$g^{(1)}_{t-t',i,j,i',j'} = \mathbf{E}\left\{(Y_{ti}Y_{tj} - \mathbf{E}\{Y_{ti}Y_{tj}\})(Y_{t'i'}Y_{t'j'} - \mathbf{E}\{Y_{t'i'}Y_{t'j'}\})\right\},$$

$$g^{(2)}_{t-t',i,j,i',j'} \cdot \mathbf{E}\left\{(Y_{ti}Y_{tj} - \mathbf{E}\{Y_{ti}Y_{tj}\})(Y_{t'+1,i'}Y_{t'j'} - \mathbf{E}\{Y_{t'+1,i'}Y_{t'j'}\})\right\},$$

$$g^{(3)}_{t-t',i,j,i',j'} \cdot \mathbf{E}\left\{(Y_{t+1,i}Y_{tj} - \mathbf{E}\{Y_{t+1,i}Y_{tj}\})(Y_{t'i'}Y_{t'j'} - \mathbf{E}\{Y_{t'i'}Y_{t'j'}\})\right\},$$

$$g^{(4)}_{t-t',i,j,i',j'} \quad \mathbf{E}\left\{(Y_{t+1,i}Y_{tj} - \mathbf{E}\{Y_{t+1,i}Y_{tj}\})(Y_{t'+1,i'}Y_{t'j'} - \mathbf{E}\{Y_{t'+1,i'}Y_{t'j'}\})\right\}.$$

The following lemma is straightforward and the proof is omitted.

**Lemma 2** *Let the model (1) and the assumptions A4, A5 take place. Then at $T \to \infty$ the following asymptotic behavior of the "missing patterns" $\{O_t\}$ takes place:*

$$\frac{1}{T - |\tau| - 1} \sum_{t,t'=1}^{T-1} O_{t+1,i} O_{tj} O_{t'i'} O_{t'j'} \delta_{t-t',\tau} \to \nu^{(3)}_{i',j',i,j}(-\tau) \in [0,1];$$

$$c^{(k)}_{\tau,i,j,i',j'}(T) \longrightarrow C^{(k)}_{\tau,i,j,i',j'}, k \in \{1,2,3,4\};$$

$$C^{(k)}_{\tau,i',j',i,j} \cdot C^{(k)}_{-\tau,i,j,i',j'}, k \in \{1,4\}, C^{(3)}_{\tau,i',j',i,j} \quad C^{(2)}_{-\tau,i,j,i',j'}, \tau \in \mathbb{Z}, i,j,i',j' \in \{1,\ldots,d\}.$$

**Lemma 3** *Let the model (1) and the assumption A2 take place. Then the covariance*

$$\mathbf{E}\left\{(Y_{t+u,i}Y_{tj} - \mathbf{E}\{Y_{t+u,i}Y_{tj}\})(Y_{t'+u',i'}Y_{t'j'} - \mathbf{E}\{Y_{t'+u',i'}Y_{t'j'}\})\right\}$$

*depends functionally on the time moments $t,t'$ through their difference $t - t'$ only,*

$$|\mathbf{E}\left\{(Y_{t+u,i}Y_{tj} - \mathbf{E}\{Y_{t+u,i}Y_{tj}\})(Y_{t'+u',i'}Y_{t'j'} - \mathbf{E}\{Y_{t'+u',i'}Y_{t'j'}\})\right\}| \le \lambda^{|t-t'|}\mathrm{Const}_{i,i'}.$$

*where $t,t',u,u' \in \mathbb{Z}, \lambda \in [0,1)$. and the relations for covariances take place:*

$$g^{(k)}_{\tau,i',j',i,j} \cdots g^{(k)}_{-\tau,i,j,i',j'}, k \in \{1,4\}, g^{(3)}_{\tau,i',j',i,j} - g^{(2)}_{-\tau,i,j,i',j'}, \tau \in \mathbb{Z}, i,j,i',j' \in \{1,\ldots,d\}.$$

**Proof.** The first statement of this lemma follows from the invariance property (w.r.t. time $t$) of the moments $\Sigma, \Sigma^{(4)}$. The second statement – from the inequality for powers of the matrix $B$ (Anderson, 1971): $|(B^\tau)_{ij}| \le \lambda^\tau \mathrm{Const}_{\tau,i,j}$, where $\lambda \in [0,1), \tau \in \mathbb{N}$, $i,j \in \{1,\ldots,d\}$. By replacement of indices one can easy come to the third statement of lemma. $\square$

**Theorem 3** *Let the model (1) and the assumptions A1, A4 take place. Then the estimators (8), (9) are consistent at $T \to \infty$:*

$$\hat{B} \longrightarrow^P B, \hat{G} \longrightarrow^P G, \hat{G}_1 \longrightarrow^P G_1.$$

**Proof.** The statement follows from the expression (9) for matrix $B$ and the properties of estimators for covariances (8). $\square$

Next results on asymptotic normality of $\hat{B}$ are based on the following central limit theorem for $m_T$-dependent random vectors (Shergin, 1976; Maejima, 1978).

**Theorem 4** *Let* $\left\{ Z_t^{(T)}, t = 1, 2, \ldots, k_T \right\}$ *be a sequence of* $m_T$*-dependent random variables with* $\mathbf{E}\left\{ Z_t^{(T)} \right\} = 0$ *for all* $t$ *and* $T$. $k_T \to \infty$ *at* $T \to \infty$. $S_T = \sum_{t=1}^{k_T} Z_t^{(T)}$. $D_T^2 = \mathbf{E}\{S_T^2\}$. $\bar{D}_T^2 = \sum_{t=1}^{k_T} \mathbf{E}\left\{ \left( Z_t^{(T)} \right)^2 \right\}$. $F_T(x) = \mathbf{P}\left( S_T \le D_T x \right)$ *be the distribution function of* $S_T$. $\triangle_T(x) = |F_T(x) - \Phi(x)|$, $\Phi(x)$ *is the standard normal distribution.* $\gamma_\delta = \frac{\delta(\delta+2)}{2(\delta^2+4\delta+2)}$. Const *is a constant,* $0 < \delta \le 1$, $\varepsilon_T = D_T^{-2} m_T^{\frac{3\delta+2}{\delta}}$. *Suppose that* $\mathbf{E}\left\{ \left| Z_t^{(T)} \right|^{2+\delta} \right\} < \infty$ *and the following assumptions are satisfied: 1)* $D_T^2 \longrightarrow \infty$. *2)* $\bar{D}_T^2 = \mathcal{O}(D_T^2)$, *3)* $\sum_{t=1}^{k_T} \mathbf{E}\left\{ \left| Z_t^{(T)} \right|^{2+\delta} \right\} = \mathcal{O}(D_T^2)$, *4)* $k_T = \mathcal{O}(D_T^2)$, *5)* $D_T^8 m_T^{-6} \le k_T^7$ *for large* $T$, *6)* $\varepsilon_T \longrightarrow 0$. *Then* $\triangle_T(x) \le \frac{Const}{(1+|x|)^{2+\delta}} \varepsilon_T^{\gamma_\delta}$ *for all* $x$.

Let a sequence of random vectors $\xi_T = (\xi_{T_i}) \in \mathbb{R}^d$ converges in distribution to random vector $\xi = (\xi_i) \in \mathbb{R}^d$ at $T \to \infty$. If the covariances $\mathrm{cov}\{\xi_i, \xi_j\} < \infty$ exist $\forall i, j \in \{1, \ldots, d\}$, then let us call these covariances as the asymptotic covariances of the sequence $\xi_T$, and let us denote them in the following form:

$$\mathbf{acov}\left\{ (\xi_T)_i, (\xi_T)_j \right\} = \mathbf{cov}\{\xi_i, \xi_j\}, i, j \in \{1, \ldots, d\}.$$

In the same manner define the asymptotic variance and asymptotic mathematical expectation of the sequence $\xi_T$: $\mathbf{aE}\{\xi_T\} = \mathbf{E}\{\xi\}$, $\mathbf{aD}\{\xi_{T_i}\} = \mathbf{D}\{\xi_i\}, i \in \{1, \ldots, d\}$.

**Lemma 4** *Let the model (1) and the assumptions A2, A4, A5 take place. Then at* $T \to \infty$ *the vector, composed of elements of matrices* $\sqrt{T}\left( \hat{G} - G \right)$, $\sqrt{T}\left( \hat{G}_1 - G_1 \right)$, *has the asymptotically normal distribution with zero mean and asymptotic covariances:*

$$\mathbf{acov}\left\{ \sqrt{T}\left( \hat{G} - G \right)_{ij}, \sqrt{T}\left( \hat{G} - G \right)_{i'j'} \right\} = \sum_{\tau=-\infty}^{\infty} g_{\tau,i,j,i',j'}^{(1)} C_{\tau,i,j,i',j'}^{(1)}. \qquad (10)$$

$$\mathbf{acov}\left\{ \sqrt{T}\left( \hat{G}_1 - G_1 \right)_{ij}, \sqrt{T}\left( \hat{G}_1 - G_1 \right)_{i'j'} \right\} = \sum_{\tau=-\infty}^{\infty} g_{\tau,i,j,i',j'}^{(4)} C_{\tau,i,j,i',j'}^{(4)}.$$

$$\mathbf{acov}\left\{ \sqrt{T}\left( \hat{G} - G \right)_{ij}, \sqrt{T}\left( \hat{G}_1 - G_1 \right)_{i'j'} \right\} = \sum_{\tau=-\infty}^{\infty} g_{\tau,i,j,i',j'}^{(2)} C_{\tau,i,j,i',j'}^{(2)},$$

*where* $i, j, i', j' \in \{1, \ldots, d\}$.

**Proof.** According to the theorem of Shiryaev (1995), a vector has an asymptotically normal distribution if and only if any linear combination of its elements has an asymptotically normal distribution. Define the linear combination with arbitrary coefficients $\alpha_{ij}, \beta_{ij} \in \mathbb{R}$:

$$\sqrt{T} \sum_{i,j=1}^{d} \left( \alpha_{ij}\left( \hat{G} - G \right)_{ij} + \beta_{ij}\left( \hat{G}_1 - G_1 \right)_{ij} \right) = \eta_1(T) + \eta_2(T) + \eta_3(T).$$

where

$$
\eta_1(T) \quad \sqrt{T}\sum_{t=1}^{T-1}\sum_{i,j=1}^{d}\left(\frac{\alpha_{ij}\left(Y_{ti}^{(m_T)}Y_{tj}^{(m_T)}-\mathbf{E}\left\{Y_{ti}^{(m_T)}Y_{tj}^{(m_T)}\right\}\right)O_{ti}O_{tj}}{\sum_{t_1=1}^{T}O_{t_1i}O_{t_1j}}\right.
$$

$$
\left.\frac{\beta_{ij}\left(Y_{t+1,i}^{(m_T)}Y_{tj}^{(m_T)}-\mathbf{E}\left\{Y_{t+1,i}^{(m_T)}Y_{tj}^{(m_T)}\right\}\right)O_{t+1,i}O_{tj}}{\sum_{t_1=1}^{T-1}O_{t_1+1,i}O_{t_1j}}\right),
$$

$$
\eta_2(T) \quad \sqrt{T}\sum_{t=1}^{T-1}\sum_{i,j=1}^{d}\left(\frac{\alpha_{ij}\left(Y_{ti}^{(m_T)}U_{tj}^{(m_T)}-\mathbf{E}\left\{Y_{ti}^{(m_T)}U_{tj}^{(m_T)}\right\}\right)O_{ti}O_{tj}}{\sum_{t_1=1}^{T}O_{t_1i}O_{t_1j}}\right.
$$

$$
\frac{\alpha_{ij}\left(U_{ti}^{(m_T)}Y_{tj}^{(m_T)}-\mathbf{E}\left\{U_{ti}^{(m_T)}Y_{tj}^{(m_T)}\right\}\right)O_{ti}O_{tj}}{\sum_{t_1=1}^{T}O_{t_1i}O_{t_1j}}+
$$

$$
\frac{\alpha_{ij}\left(U_{ti}^{(m_T)}U_{tj}^{(m_T)}-\mathbf{E}\left\{U_{ti}^{(m_T)}U_{tj}^{(m_T)}\right\}\right)O_{ti}O_{tj}}{\sum_{t_1=1}^{T}O_{t_1i}O_{t_1j}}+
$$

$$
\frac{\beta_{ij}\left(Y_{t+1,i}^{(m_T)}U_{tj}^{(m_T)}-\mathbf{E}\left\{Y_{t+1,i}^{(m_T)}U_{tj}^{(m_T)}\right\}\right)O_{t+1,i}O_{tj}}{\sum_{t_1=1}^{T-1}O_{t_1+1,i}O_{t_1j}}
$$

$$
\frac{\beta_{ij}\left(U_{t+1,i}^{(m_T)}Y_{tj}^{(m_T)}-\mathbf{E}\left\{U_{t+1,i}^{(m_T)}Y_{tj}^{(m_T)}\right\}\right)O_{t+1,i}O_{tj}}{\sum_{t_1=1}^{T-1}O_{t_1+1,i}O_{t_1j}}+
$$

$$
\left.\frac{\beta_{ij}\left(U_{t+1,i}^{(m_T)}U_{tj}^{(m_T)}-\mathbf{E}\left\{U_{t+1,i}^{(m_T)}U_{tj}^{(m_T)}\right\}\right)O_{t+1,i}O_{tj}}{\sum_{t_1=1}^{T-1}O_{t_1+1,i}O_{t_1j}}\right),
$$

$$
\eta_3(T) \quad \sqrt{T}\sum_{i,j=1}^{d}\frac{\alpha_{ij}\left(Y_{Ti}Y_{Tj}-\mathbf{E}\left\{Y_{Ti}Y_{Tj}\right\}\right)O_{Ti}O_{Tj}}{\sum_{t=1}^{T}O_{ti}O_{tj}},
$$

$Y_t - Y_t^{(m_T)} + U_t^{(m_T)}$, $Y_t^{(m_T)} - \sum_{s=0}^{m_T}B^sU_{t-s}$, $U_t^{(m_T)} := \sum_{s=m_T+1}^{\infty}B^sU_{t-s}$, $m_T \in \mathbb{N}$ is a parameter of decomposition.

By choosing $\delta = 1$, $k_T - T - 1$, $m_T = \left[T^{\frac{7}{36}}\right]$, where $[\cdot]$ means integer part,

$$
S_T - \sum_{t=1}^{T-1}T\sum_{i,j=1}^{d}\left(\frac{\alpha_{ij}\left(Y_{ti}^{(m_T)}Y_{tj}^{(m_T)}-\mathbf{E}\left\{Y_{ti}^{(m_T)}Y_{tj}^{(m_T)}\right\}\right)O_{ti}O_{tj}}{\sum_{t_1=1}^{T}O_{t_1i}O_{t_1j}}+\right.
$$

$$
\left.\frac{\beta_{ij}\left(Y_{t+1,i}^{(m_T)}Y_{tj}^{(m_T)}-\mathbf{E}\left\{Y_{t+1,i}^{(m_T)}Y_{tj}^{(m_T)}\right\}\right)O_{t+1,i}O_{tj}}{\sum_{t_1=1}^{T-1}O_{t_1+1,i}O_{t_1j}}\right),
$$

and using lemmas 2, 3, one can easy verify the assumptions of Theorem 4 and thereby prove that $\eta_1(T)$ has the asymptotically normal distribution with zero mean and asymptotic variance:

$$\sum_{\tau=-\infty}^{\infty} \sum_{i,i',j,j'=1}^{d} \left( \alpha_{ij}\alpha_{i'j'} g^{(1)}_{\tau,i,j,i',j'} C^{(1)}_{\tau,i,j,i',j'} + \alpha_{ij}\beta_{i'j'} g^{(2)}_{\tau,i,j,i',j'} C^{(2)}_{\tau,i,j,i',j'} + \right.$$
$$\left. \beta_{ij}\alpha_{i'j'} g^{(3)}_{\tau,i,j,i',j'} C^{(3)}_{\tau,i,j,i',j'} + \beta_{ij}\beta_{i'j'} g^{(4)}_{\tau,i,j,i',j'} C^{(4)}_{\tau,i,j,i',j'} \right).$$

The convergence in probability of the second and the third terms to zero takes place: $\eta_2(T) \longrightarrow^P 0$, $\eta_3(T) \longrightarrow^P 0$ at $T \to \infty$.

Then according to the theorem of Shiryaev (1995) the vector, which is composed of the elements of the matrices $\sqrt{T}\left(\hat{G} - G\right)$, $\sqrt{T}\left(\hat{G}_1 - G_1\right)$, has the asymptotically normal distribution with zero mean and asymptotic covariances (10). $\square$

**Theorem 5** *Let the model (1) and the assumptions A2, A4, A5 take place. Then at $T \to \infty$ the vector, composed of the elements of the matrix $\sqrt{T}\left(\hat{B} - B\right)$, has the asymptotically normal distribution with zero mean and asymptotic covariances:*

$$\mathbf{acov}\left\{ \sqrt{T}\left(\hat{B} - B\right)_{ij}, \sqrt{T}\left(\hat{B} - B\right)_{i'j'} \right\} = \sum_{l,l'=1}^{d} \sum_{\tau=-\infty}^{\infty} (G^{-1})_{lj} (G^{-1})_{l'j'} \times \quad (11)$$

$$\left( \sum_{k,k'=1}^{d} B_{ik}B_{i'k'} g^{(1)}_{\tau,k,l,k',l'} C^{(1)}_{\tau,k,l,k',l'} + g^{(4)}_{\tau,i,l,i',l'} C^{(4)}_{\tau,i,l,i',l'} - 2\sum_{k=1}^{d} B_{ik} g^{(2)}_{\tau,k,l,i',l'} C^{(2)}_{\tau,k,l,i',l'} \right),$$

*where $i, j, i', j' \in \{1, \ldots, k\}$.*

**Proof.** Transform the normed deviation using the formula (9):

$$\sqrt{T}\left(\hat{B} - B\right) = \sqrt{T}\left( -\hat{G}_1\hat{G}^{-1}\left(\hat{G} - G\right) G^{-1} + \left(\hat{G}_1 - G_1\right) G^{-1} \right).$$

Define the linear combination with arbitrary coefficients $\alpha_{ij} \in \mathbb{R}$:

$$\sum_{i,j=1}^{d} \alpha_{ij}\sqrt{T}\left(\hat{B} - B\right)_{ij} = \sqrt{T}\sum_{i,j=1}^{d} \alpha_{ij} \times$$

$$\left( \sum_{k,l=1}^{d} (-1)\left(\hat{G}_1\hat{G}^{-1}\right)_{ik}\left(\hat{G} - G\right)_{kl}(G^{-1})_{lj} + \sum_{l=1}^{d}\left(\hat{G}_1 - G_1\right)_{il}(G^{-1})_{lj} \right).$$

Using Theorem 3, Lemma 4 and the well known theorem on continuous functional transformations of random vectors (Barndorff-Nielsen and Cox, 1989), we come to the asymptotic normality of the linear combination with zero mean and asymptotic variance:

$$\mathbf{aD}\left\{ \sqrt{T}\sum_{i,j=1}^{d} \alpha_{ij} \left( \sum_{k,l=1}^{d}(-1)(G_1G^{-1})_{ik}\left(\hat{G} - G\right)_{kl}(G^{-1})_{lj} + \right.\right.$$

$$\sum_{l=1}^{d}\left(\dot{G}_1 - G_1\right)_{il}\left(\left(G^{-1}\right)_{lj}\right)\Bigg\} - \sum_{i,j,i',j'=1}^{d} \alpha_{ij}\alpha_{i'j'} \sum_{l,l'=1}^{d} \sum_{\tau=-\infty}^{\infty} \left(G^{-1}\right)_{lj}\left(G^{-1}\right)_{l'j'} \times$$

$$\left(\sum_{k,k'=1}^{d} B_{ik}B_{i'k'}g_{\tau,k,l,k',l'}^{(1)}C_{\tau,k,l,k',l'}^{(1)} + g_{\tau,i,l,i',l'}^{(1)}C_{\tau,i,l,i',l'}^{(1)} - 2\sum_{k=1}^{d} B_{ik}g_{\tau,k,l,i',l'}^{(2)}C_{\tau,k,l,i',l'}^{(2)}\right).$$

Thus the vector, composed of the elements of the matrix $\sqrt{T}\left(\hat{B} - B\right)$, has the asymptotically normal distribution with zero mean and asymptotic covariances (11). □

## 5   Numerical Results

To evaluate the performance of the estimators (9), the experiment on the celebrated "Beveridge price index for wheat 1500 – 1869" (Anderson, 1971) was made. The AR(3) model is $y_{t+1} - 0.7489y_t - 0.3397y_{t-1} + 0.0388y_{t-2} + \xi_t$, where $\{\xi_t\}$ are i.i.d. Gaussian random variables with zero mean and the variance $\sigma^2 - 4$, $t_- \cdot 1$, $t_+ = T - 100$. This model was transformed to the 3-variate VAR(1)-model (1) with $Y_t = (Y_t, Y_{t-1}, Y_{t-2})' \in \mathbb{R}^3$ and was considered for the "missing patterns" $O_t = \left\{0, t = \left[\frac{T-d}{[(T-d)\gamma]+1}\right]i, \, i \in \{1,\ldots,[(T-d)\gamma]\}; \, 1, \text{else}\right\}$, where $\gamma$ is a given portion of missing values, $d - 3$. Dependence of the sample variance $\hat{V} = \frac{1}{400}\sum_{i=1}^{400}\sum_{k,l=1}^{d}(\hat{B}_{kl}^{(i)} - B_{kl})^2$ (for 400 Monte-Carlo replications) for the estimator (9) on the length $T$ of observed time series is presented in Figure 1 for different portions of missing values $\gamma \in \{0, 0.07, 0.1\}$. Also the empirical risk of forecasting $\hat{r} - \frac{1}{400}\sum_{i=1}^{400}(\hat{y}_{T+1}^{(i)} - y_{T+1})^2$ for the forecast $\hat{y}_{T+1} - \left(\hat{Y}_{T+1,\text{plug-in}}\right)_1$ was evaluated, its dependence on the length $T$ is presented in Figure 2.
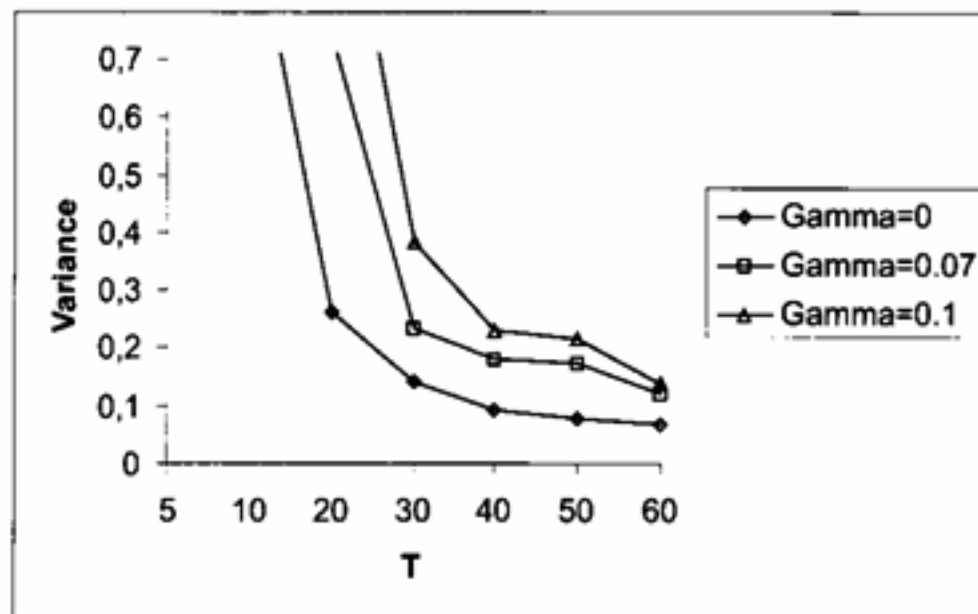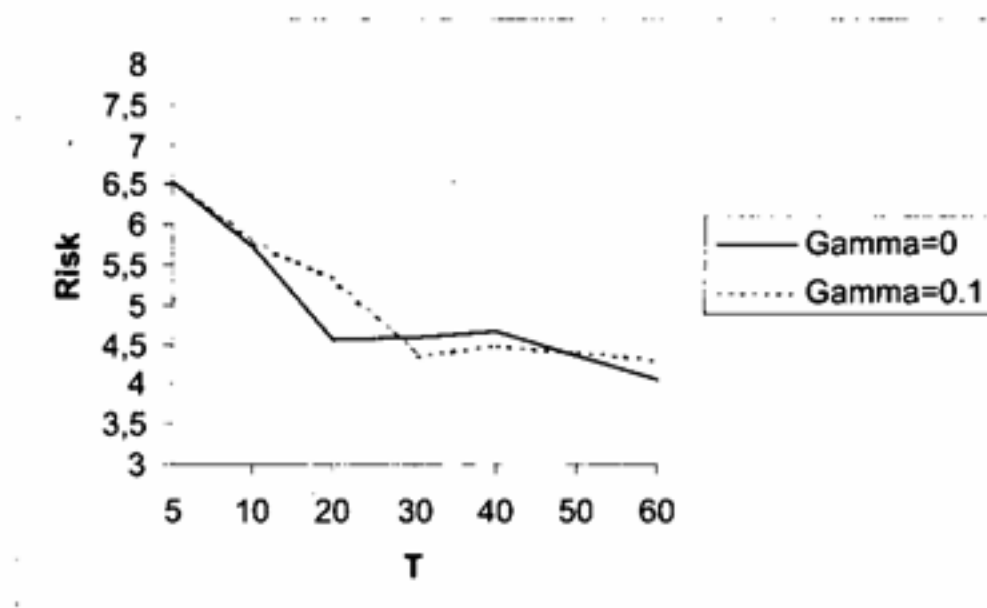


Figure 1: Sample variance of the estimator $\hat{B}$

Figure 2: Empirical risk for the forecast $\hat{Y}_{T+1,\text{plug-in}}$

# References

T.W. Anderson. *The Statistical Analysis of Time Series*. Wiley, New York, 1st edition, 1971.

O.E. Barndorff-Nielsen and D.R. Cox. *Asymptotic Techniques for Use in Statistics*. Chapman and Hall, London, 1st edition, 1989.

J. Beran, R.J. Bhansali, and D. Ocker. On unified model selection for stationary and nonstationary short- and long-memory autoregressive processes. *Biometrica*, 85(4): 921–934, 1998.

W.H. Greene. *Econometric Analysis*. Macmillan, New York, 1st edition, 2000.

R.H. Jones. Maximum likelihood fitting of arma models to time series with missing observations. *Technometrics*, 22(3):389–395, 1980.

Yu.S. Kharin2 and A.S. Huryn. Statistical analysis and forecasting of autoregressive time series under missing values. *Bulletin of the International Statistic Institute*, (1):612–613, 2003.

R. Little and D. Rubin. *Statistical Analysis with Missing Data*. Wiley, New York, 1st edition, 1987.

M. Maejima. A non-uniform estimate in the central limit theorem for m-dependent random variables. *KEIO Engineering Reports*, 31(2):15–20, 1978.

S.G. Pantula and D. Shin. Testing for a unit root in autoregressive processes with systematic but incomplete sampling. *Stat. & Prob. Letters*, 18:183–190, 1993.

J.L. Shafer. *Analysis of Incomplete Data*. Chapman and Hall, London, 1st edition, 1997.

V.V. Shergin. Estimate of the error term in the central limit theorem for m-dependent random variables. *Lietuvos Matematikos Rinkinys*, 16(4):245–250, 1976.

A.N. Shiryaev. *Probability*. Springer-Verlag, New York, 1st edition, 1995.

N. Stockinger and R. Dutter. Robust time series snalysis: A survey. *Kybernetika*, 23: 3–90, 1987.

Corresponding author's address:

Univ.-Prof. Dr. Yuriy Kharin
Department of Mathematical Modelling and Data Analysis
Belarusian State University
Fr. Skoriny av. 4
220050 Minsk
Belarus

Tel. +375 17 2095530
E-mail: Kharin@bsu.by