Analysis and investigation of the conservativeness condition in the problem of parametric identification of distributed dynamic processes

M G Matveev and E A Sirota

Computer science department, Voronezh State University, Universitetskaya pl. 1, Voronezh, Russia

e-mail: mgmatveev@yandex.ru, atoris@list.ru

Abstract. The problem of bias of OLS estimates arises when solving the problem of parametric identification of distributed dynamic processes. There are various possible solutions to this problem. If the time series is trend-stationary, then these may be "ostationation" methods, which are generally difficult to apply. It is possible to use dimensionality reduction methods, but in this case we will still get biased estimates. In our previous works, it was shown that the problem of biased estimates can be solved using the conservativeness condition. The aim of this work was to investigate the possibility of using the conservativeness condition to improve the quality of estimates of the parametric identification problem, as well as to compare these results with the solution of the problem, in the case of applying a filter to it, as well as ridge regression.

1. Introduction

The problem of bias of OLS estimates in parametric identification of the autoregressive model of nonstationary time series is known [1, 2]. If the time series belongs to the class of TS-series [2], then the possibilities of parametric identification are limited by the methods of "ostationization" of the series: either by the sequential application of the finite difference operator, or by cointegrating methods [2]. The processes of "ostationatization" can provide unbiased estimates, but they are quite complex, cumbersome and require confirmation of a number of statistical hypotheses and, generally speaking, do not ensure the effectiveness of estimates.

Another problem affecting the quality of parametric identification estimates is multicollinearity. When constructing a forecast based on a regression model with multicollinear factors, it is necessary to evaluate the situation by the magnitude of the forecast error. If its value is satisfactory, then the model can be used despite multicollinearity. If the value of the forecast error is large, then the elimination of multicollinear factors from the regression model is one of the methods to improve the quality of the forecast. To date, there are a number of basic ways to eliminate multicollinearity in a multiple regression model.

If the methods did not resolve the multicollinearity of the factors, then go to the use of biased estimation methods of unknown parameters of a regression model, or methods of exclusion of variables from multiple regression models [5], in other words methods of dimension reduction tasks, for example, the method of principal components or ridge regression.
For a certain class of equations, this problem of biased estimates can be solved quite simply by using the conservativeness property [6].

To analyze the quality of the obtained estimates, parametric identification problems usually calculate the bias and standard error of the parameter estimation.

It is obvious that the above problems arise due to the presence of interference. The signal with interference can be filtered out, but this does not exclude distortion of the useful signal. In this case, the useful signal has a characteristic property – the sum of the parameters is equal to one, which is the condition of conservativeness.

Thus, the purpose of this work was to investigate the possibility of using the conservativeness condition to improve the quality of estimates of the parametric identification problem, as well as to compare these results with the solution of the problem, in the case of applying a filter to it, as well as ridge regression.

2. Materials and methods

Let’s consider the research problem in more detail.

If the processes are properly described by linear differential equations, it is convenient to change to difference equations to obtain estimates. For example, for a homogeneous convective diffusion equation with one spatial variable, the reduced difference equation for the \( i \)th node at \( k+1 \) time will be as follows [4]:

\[
y_{i}^{k+1} = a_1 y_{i-1}^k + a_2 y_i^k + a_3 y_{i+1}^k,
\]

with given initial and boundary conditions:

\[
y_i^0 = c_i, \quad y_{i-1}^k = b_{i-1}^k, \quad y_{i+1}^k = b_{i+1}^k, \quad \forall k,
\]

where \( i \) - discrete values of the spatial coordinate, \( k \) - discrete time; in general, the necessary index \( i \) of the parameters \( a = (a_1; a_2; a_3) \) is omitted henceforward for simplicity; condition \( a_1 + a_2 + a_3 = 1 \) ensures conservativeness of the scheme. Here should be recalled, that the scheme is called conservative if it reflects on the grid the same conservation laws that were presented in the original differential problem [5].

The values of the variable \( y_i^k \) are measured at each node \( i \) with an error \( \xi_i^k \) generated by a random white noise process. The measured value will be denoted by \( x_i^k = y_i^k + \xi_i^k \).

Then the expressions (1) and (2) can be rewritten in the form of an autoregressive dependence, describing non-stationary time series:

\[
x_{i}^{k+1} = a_1(x_{i-1}^k - \xi_{i-1}^k) + a_2(x_i^k - \xi_i^k) + a_3(x_{i+1}^k - \xi_{i+1}^k) + \xi_i^{k+1} = a^T \cdot x^k + \omega_i,
\]

Values of the initial and boundary conditions are also determined by the measurement results.

The property of parameters in the sum to give one can be used to prevent distortion of the useful signal. To do this, it is proposed to replace one of the OLS equations in solving the parametric identification problem by an amount equal to one.

Let’s assume that we have high-frequency interference, which we will filter, for example, by a moving average with a certain window parameter.

You need to make sure that filtering generally improves the estimates, but starting at some level of amplitude, the estimates get worse. Our task is to show that the deterioration of the estimates can be avoided or, at least, significantly weakened by the equation the sum of the parameters is equal to one. The conducted experiments clearly show this.

The sample statistics required for the model study were obtained using the analytical solution (1) of the differential equation (2). The analytical solution was discretized in time and spatial coordinate. The
difference equation was obtained with the following parameter values: 

\[ a_1 = 0.2583, a_2 = 0.5, a_3 = 0.2417. \]

In accordance with the research methodology, an additive observation noise was added to the values of the variable in the corresponding grid nodes, obtained using a Gaussian-type independent random number generator with zero expectation and unit variance. The interference intensity was set at different levels indicated in Table 1.

To obtain reliable results of estimates of bias and standard error that allow comparison of numerical values, experiments in all modes were repeated 1000 times and their results were averaged.

We will use the filter-moving average [7], the window size changed from 10 to 70 in increments of 10. Using the filtered signal, we find the OLS estimates and compare them with the unfiltered ones. At the next stage, one of the OLS equations is replaced by the sum of the parameters equal to one.

The table below provides a comparative analysis of estimates of the average bias value in the case of conventional OLS estimates, OLS estimates obtained as a result of filtering, as well as when replacing one equation with an amount equal to one. In the case of filtering, the table shows data for a window equal to 40, it is this window size that showed the best result in terms of offset.

**Table 1. Comparative analysis of estimates of the average value of the bias.**

| \( c \) | Percentage of signal interference (%) | Estimates of the average value of the bias |
|-------|---------------------------------------|------------------------------------------|
|       |                                       | The bias of the OLS – estimators | Bias of OLS estimates in the case of filtering, the window size is 40 | Displacement of OLS estimates when replacing the equation |
|       |                                       | \( \Delta a_1 \) | \( \Delta a_2 \) | \( \Delta a_3 \) | \( \Delta a_1 \) | \( \Delta a_2 \) | \( \Delta a_3 \) | \( \Delta a_1 \) | \( \Delta a_2 \) | \( \Delta a_3 \) |
| 0.6   | 13.75                                 | 0.0444 | 0.056 | 0.017 | 0.04 | 0.05 | 0.013 | 0.036 | 0.04 | 0.01 |
| 0.3   | 7.4                                   | 0.0433 | 0.047 | 0.0134 | 0.0399 | 0.04 | 0.01 | 0.0332 | 0.039 | 0.01 |
| 0.2   | 6.1                                   | 0.0378 | 0.045 | 0.0089 | 0.0356 | 0.038 | 0.0087 | 0.0303 | 0.038 | 0.008 |
| 0.1   | 3.9                                   | 0.0346 | 0.034 | 0.0072 | 0.0331 | 0.031 | 0.0025 | 0.03 | 0.0286 | 0.002 |
| 0.01  | 0.13                                  | 0.0127 | 0.0169 | 0.0034 | 0.0132 | 0.0167 | 0.0012 | 0.011 | 0.016 | 0.00101 |
| 0.0025 | 0.2                                  | 0.0063 | 0.0079 | 0.0017 | 0.006 | 0.0076 | 0.001 | 0.0056 | 0.007 | 0.00093 |
| 0.001 | 0.17                                 | 0.0033 | 0.0052 | 0.001 | 0.00314 | 0.0049 | 0.0009 | 0.00313 | 0.0042 | 0.0007 |

Analysis of the table shows that the best offset result is achieved if the equation is replaced by an amount equal to one.

It is shown in [8] that in some cases, when using ridge regression, a significant reduction in bias and standard error is possible in comparison with the use of standard OLS estimates. As a rule, time series in adjacent nodes of the approximating grid are strongly correlated with each other, so the use of ridge regression, as one of the most effective methods of reducing the dimension, seems reasonable. Therefore, as an additional study, it was decided to compare the best result on the bias in the case of replacing one of the OLS equations by an amount equal to one and the solution of the problem obtained as a result of applying ridge regression.

In [7], the reasons for reducing the dimension in the case of ridge regression are shown. Briefly, the solution of the problem of parametric identification using standard OLS estimates and ridge regression can be described as follows.

Ridge regression is used to reduce the dimension of the feature space, and, consequently, to eliminate such a problem as multicollinearity. To solve this problem, we introduce a restriction on the vector of coefficients in the form of an additional "penalty" term. Thus, there are two evaluation criteria, the first-from the least squares method, and the second-the square of the norm of weights, which should be minimized. These two criteria are combined with a non-negative regularization parameter [7]

\[
Q_b(a) = \| F_a - y \|^2 + \frac{\alpha}{2} \| a \|^2, 
\]
where $\tau$ is the regularization parameter, $\alpha$ is the vector of parameter estimates, $Q_\alpha(a)$ is the error square functional, $y$ - is the target vector, $F$ - is the matrix of feature objects [9]. Differentiating by then equating the derivative to zero, we obtain the expression for $\alpha$

$$\alpha_\tau = (F^TF + \tau I_n)^{-1}F^Ty,$$  \hspace{1cm} (5)

where $I_n$ is a unit matrix of size $n \times n$. The product $\tau I_n$ is called the crest of the regression. To find the vector of coefficients needed to apply the method of singular value decomposition to the matrix $F$ of expression (5)

$$\alpha_\tau = U(D^2 + \tau I_n)^{-1}DV^Ty = \sum_{j=1}^n \frac{\lambda_j}{\lambda_j + \tau} u_j^T(v_j^T y);$$ \hspace{1cm} (6)

Here $VU^T$ are orthogonal matrices, $D$ is the diagonal matrix, $u_j$ and $v_j^T$ are the proper vectors, $\lambda_j$ are the eigenvalues of [9].

It can be seen that the regularizing parameter $\tau$ appears in the solution, it does not allow the denominator to approach zero. Now if the eigenvalues are equal to zero, then the tendency of the terms to infinity is impossible, respectively, the solution becomes more resistant to noise in the data. The square of the norm of the coefficient vectors will also change its form [8]:

$$\|\alpha_\tau\|^2 = \left\| (D^2 + \tau I_n)^{-1}DV^Ty \right\|^2 = \sum_{j=1}^n \frac{\lambda_j}{(\lambda_j + \tau)^2} (v_j^Ty)^2.$$ \hspace{1cm} (7)

It follows from expressions (6) and (7) that for a positive parameter $\tau$, even for singular values close to zero, the phenomenon of multicollinearity can be avoided.

This regression method leads to a reduction in the effective dimension. This effect occurs due to an artificial restriction on the norm of the coefficient vector [8]:

$$\|\alpha\|^2 = \sum_{j=1}^n \frac{\lambda_j}{(\lambda_j + \tau)^2} (v_j^Ty)^2 < \|\alpha_\tau\|^2 = \sum_{j=1}^n \frac{1}{\lambda_j} (v_j^Ty)^2.$$ \hspace{1cm} (8)

to estimate the dimension of the space is invited to consider the trace of the projection matrix:

$$trF(F^TF)^{-1}F^T = tr(F^TF)^{-1}F^T F = trl_n = n.$$ \hspace{1cm} (9)

In the case of ridge regression there is another matrix with the additive in the form of ridge, respectively projected track will change and it will be less than the standard least squares method [9]:

$$tr(F^TF + \tau l_n)^{-1}F^T = trdiag \left( \frac{\lambda_j}{\lambda_j + \tau} \right) = \sum_{j=1}^n \frac{\lambda_j}{\lambda_j + \tau} < n.$$ \hspace{1cm} (10)

Thus, ridge regression really helps in eliminating multicollinearity in the parametric identification problem by reducing the dimension of the feature space and imposing a constraint on the vector of weight coefficients of the regression model.

As it turned out, in the case of ridge regression, with an increase in the regularization parameter, the resistance to noise in the original data increases, but the overall accuracy of the model decreases. At each noise level, the regularization parameter was considered in the range from 0.01 to 20. It is shown in that with the growth of the regularization parameter up to a certain point, the model begins to describe the target values more adequately. This phenomenon is explained by an underestimation of the sum of the vector of weight coefficients, therefore, emissions are given less importance and the model reacts to them more stably. But with an increase in the "crest" parameter, after it reaches a
certain level, the accuracy begins to decrease. Due to the shift of the vector of weight coefficients in the lower direction, the model is also shifted.

3. Results and discussion
Table 2 shows the results of comparing the parameter offsets in the case of replacing the equation by one and applying ridge regression. For ridge regression, the regularization parameter for which the bias estimates are the smallest is selected for inclusion in the table 2.

| c    | Percentage of signal interference (%) | Estimates of the average value of the bias | Displacement of OLS estimates when replacing the equation |
|------|--------------------------------------|------------------------------------------|----------------------------------------------------------|
|      |                                      | Bias of OLS estimates in the case of ridge regression |                                                                 |
|      |                                      | \( \Delta a_1 \) | \( \Delta a_2 \) | \( \Delta a_3 \) | \( \Delta a_1 \) | \( \Delta a_2 \) | \( \Delta a_3 \) |
| 0,6  | 13,75                                | 0.0389 | 0.045 | 0.0125 | 0.036 | 0.04 | 0.01 |
| 0,3  | 7,4                                  | 0.0344 | 0.04 | 0.011 | 0.0332 | 0.039 | 0.01 |
| 0,2  | 6,1                                  | 0.0322 | 0.0383 | 0.0085 | 0.0303 | 0.038 | 0.008 |
| 0,1  | 3,9                                  | 0.0301 | 0.0291 | 0.00243 | 0.03 | 0.0286 | 0.002 |
| 0,01 | 0,13                                 | 0.0123 | 0.0165 | 0.00105 | 0.011 | 0.016 | 0.00101 |
| 0,0025 | 0,2                               | 0.0058 | 0.0073 | 0.001 | 0.0056 | 0.007 | 0.00093 |
| 0,001 | 0,17                                | 0.003136 | 0.0049 | 0.00081 | 0.00313 | 0.0042 | 0.0007 |

The analysis of Table 2 shows that the results of using ridge regression for the initial problem are still slightly worse than the results of using the conservativeness condition in the framework of the OLS application.

The graphs show the dependence of bias on interference for all three parameters in the case of ordinary OLS estimates, OLS estimates obtained as a result of filtering, as well as when replacing one equation with an amount equal to one. In the case of the first parameter, graphs are given for different window sizes in the case of filtering, as well as a regression ridge.

4. Conclusion
Thus, the numerical experiments show that the smallest displacement as a result of solving the parametric identification problem is achieved by replacing one of the equations with an amount equal to one. Even in the case of ridge regression, the estimates of the parameter bias are slightly larger. Signal filtering also improves the bias results compared to conventional OLS estimates, but still these results are worse than ridge regression and, even more so, in the case of replacing the equation by an amount equal to one.

Hence, the property of parameters in the sum to give one can be used to prevent distortion of the useful signal, as well as to reduce the estimates of bias in the case of solving the problem of parametric identification.
Figure 1. Dependence of the offset on the interference for the first parameter.
Figure 2. Dependence of the offset on the interference for the second parameter.

Figure 3. Dependence of the offset on the interference for the third parameter.

References
[1] Guo L Z, Billing S A and Coca D 2010 Identification of partial differential equation models for a class of multiscale spatio-temporal dynamical systems Intr. J. Contr. 83 40-8
[2] Xun X, Cao J, Mallick B, Carrol R and Maity A J 2013 Parameter Estimation of Partial Differential Equation Models J. Am. Stat. Assoc. 108 1009-20
[3] Nosko V P 2011 Econometrics vol 1 ed Yu V Luiso (Moscow: Delo)
[4] Matveev M G, Kopytin A V and Sirota E A 2018 Parameters identification of a distributed dynamical model using combined approach J. Phys. Conf. Ser. 1096 012068
[5] Samarskii A A 2001 The Theory of Difference Schemes (Boca Raton: CRC Press)
[6] Grebenyuk E A 2005 Methods of analyzing the nonstationary time series with implicit changes in their properties *Autom. Remote Control* 66 1871-96

[7] Matveev M G, Sirota E A 2016 Combined models of non-stationary time series with changing states *Bulletin of VSU, series "System analysis and information technologies"* 3 50-61

[8] Matveev M G, Kopytin A V, Sirota E A and Kopytina E A 2017 Modeling of Nonstationary Distributed Processes on the Basis of Multidimensional Time Series *Procedia Engineering* 201 511-6

[9] Kopytin A V, Kopytina E A, Matveev M G 2018 Application of the extended Kalman filter for identification of parameters of the distributed dynamic system *Bulletin of VSU, series "System analysis and information technologies"* 3 44-50

[10] Matveev M G, Sirota E A, Semenov M E and Kopytin A V 2017 Verification of the convective diffusion process based on the analysis of multidimensional time series *CEUR Workshop Proc.* 2022 354-8