

Using robust regression methods for improve the accuracy of the estimating of observational models parameters

Kazan Federal University, 420008, Kremlevskaya 18, Kazan, Russia

Abstract

© SGEM2018. Development of information and computer technology is a factor in increased use of statistical methods in all sphere of action (geology, geophysics, ecology, astronomy and many others). Regression analysis (RA) is widely used in the prediction and assessment of model parameters. For many years the ordinary least squares estimation method (LSE, OLS), a computational scheme of RA, and its generalizations are widely used in practice. The disadvantage of the classical scheme of RA, is that with the compliance LSE assumptions, are not verified, in consequence condition which the model can fail to match observations. Most authors of books on regression analysis are inclined to believe that it is necessary to check execution the conditions of RA and in case of breach of the use of alternative computational schemes. In the present paper an approach of the additive regression model, based on special software and provided for the automatic operation model, verification of compliance with the assumptions of RA – LSE and adaptation in the case of violations, is presented. The method of joint solution of the LSE estimation in violation of the assumptions about the normality of error distribution and the independence of the covariates (the absence of multicollinearity) are proposed. Estimation methods that take into account the existence of «grave error» and allowing at the same time accurately enough to determine the parameter estimates are called robust or stable. Develop sustainable methods of estimating was motivated to improve existing schemes LSE so that emissions have the least possible impact on the final results of the evaluation. A large loss of efficiency LSE – count occurs in the presence of even a small fraction of large emissions. In such situations it is necessary to use robust estimation methods that significantly reduce the harmful effects of high emissions estimate and to obtain an acceptable final assessment required parameters. If the condition is violated the independence of the covariates, there is a serious problem common to many methods of estimation. To solve this problem is currently used methods of «Ridge regression» and «Regularization». The study our aims to improve the accuracy of estimating the model parameters and prognosis in the presence of multicollinearity and outliers on the basis of sustainable ridge estimation. Based on the goals, accomplish three tasks: a) the synthesis algorithm of the method of sustainable ridge estimation, adapted together to marked disturbances; b) the software implementation of the new method; c) the testing the program on empirical data.

<http://dx.doi.org/10.5593/sgem2018/2.1/S07.105>

Keywords

Multiple analysis, Regression analysis, Robust methods, Software

References

- [1] Valeev S.G., Regression modeling at data processing, Kazan: FAN, 2001, pp 1-296 [in Russian].
- [2] Friedman J.H., Stuetzle W., Projection Pursuit Regression, Journal of the American Statistical Association, vol. 76/no. 376, pp 817-823, 1981.
- [3] Cheng M.-Y., Fan J., Peter Hall's contributions to nonparametric function estimation and modeling, The Annals of Statistics, vol. 44/ no. 5, pp 1837-1853, 2016.
- [4] Yuan M., On the identifiability of additive index models, Statistica Sinica, vol. 21/ no. 4, pp 1901-1911, 2011.
- [5] Nefedyev Y.A., Bezmenov V.M., Demin S.A., Andreev A.O., Demina N.Y., Application of antijamming robust analysis method for selenocentric reference net building, Nonlinear Phenomena in Complex Systems, vol. 19/no. 1, pp 102-106, 2016.
- [6] Nefedjev Yu.A., Rizvanov N.G., The results of an accurate analysis of EAO charts of the Moon marginal zone constructed on the basis of lunar occultations, Astronomische Nachrichten, vol. 323/no. 2, pp 135-138, 2002.
- [7] Valeev S.G., Regressionnoe modelirovanie pri obrabotke nablyudenyi, Moscow: Nauka, 1991, pp 1-273 [in Russian].
- [8] Khyuber P., Robastnost' v statistike, Moscow: Mir, 1984, pp 1-304 [in Russian].
- [9] Varaksina N.Y., Nefedyev Y.A., Churkin K.O., Zabbarova R.R., Demin S.A., Selenocentric reference coordinates net in the dynamic system, Journal of Physics: Conference Series, vol. 661, art. no. 012014, 2015.
- [10] Demina N.Y., Andreev A.O., Demin S.A., Nefedyev Y.A., The method for celestial bodies' center of mass position relative to their figures determination on the basis of harmonic analysis of the expansion in spherical functions in order to refine the physical libration parameters, Journal of Physics: Conference Series, vol. 929, art. no. 012013, 2017.
- [11] Varaksina N.Y., Nefedyev Y.A., Churkin K.O., Zabbarova R.R., Demin S.A., Lorentzian' analysis of the accuracy of modern catalogues of stellar positions, Journal of Physics: Conference Series, vol. 661, art. no. 012015, 2015.