Bayesian Hierarchical Tweedie Spatio-temporal Modeling Using Hamiltonian Monte-Carlo Algorithm for Semi-continuous Data in the Context of Contraceptive Expenditures

Nazanin Fekri[1*], Mahboubeh Parsaeian[1], Abbas Rahimi Foroushani[1], Ben Swallow[2], Abolghasem Pourreza[3]

[1]Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran.

[2]School of Mathematics and Statistics, University of Glasgow, Glasgow, UK.

[3]Department of Health Management and Economics, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran.

Abstract:

We introduce a model that uses tweedie distribution to analyze nonnegative, skewed to the right and with mass at zero data, taking into account temporal correlation and spatial correlation of areal type. We assess model performance via simulation study and apply the model to investigating modern contraceptive expenditures at the provincial level of Iran.

Keywords: Tweedie distribution, Spatio-temporal model, Hamiltonian Monte-Carlo algorithm, Household Income and Expenditure Survey, contraceptive expenditures.
Mathematics Subject Classification (2010): 60H20**,** 60B34**.**

---

1 Introduction

Space can produce variations in healthcare expenditures, even across relatively small geographic regions and suggests the need for a spatial analysis of areal data at a refined geographic scale (Kibria , 2013). Also spatially sampled studies are often repeated over

---

[*]Speaker: nazanin_fekri@yahoo.com

time. The relationship between observations collected in space and time, is often over-looked, while including a temporal or spatial correlation term is likely to improve the efficiency of the model (Swallow , 2016). Another common feature in healthcare expenditures data is their semi-continuous nature. The topics for dealing with continuous data with a spike at zero is of great practical importance but it has not been given the necessary space in the literature Böhning and Alfò (2016). A popular approach is using two-part mixture models that separately model the probability of occurrence (with a binary component) and the intensity when the response is non-zero (with a continuous component), but this way is not always correct Paradinas (2017). In this paper, we develop a Bayesian hierarchical Tweedie regression model that can accommodate the excess number of zeros by combining information from occurrence and conditional-to-presence abundance, while accounting for spatial and temporal correlation simultaneously. We assess model performance via simulation in two different types of semi-continuous datasets (with high and with low spatial and temporal correlation). Separate studies were performed for a range of values for two parameters of Tweedie distribution, power parameter P which shows the proportion of zeros (1.1, 1.3, 1.7 and 1.9), and dispersion parameter (0.5 and 3). Parameters were estimated using the Hamiltonian Monte-Carlo (HMC) algorithm in the R interface of Stan with package rstan (version 2.14). We run 3 parallel HMC chains of 6000 iterations for each simulated dataset and kept only 3000 iterations per chain for the posterior sample. We compute Median Absolute Bias (MAB) (absolute median of the difference between the posterior means and the true value), Median Relative Bias (MRB) (absolute median of the difference between the posterior means and the true value divided by the true value), the Mean Square Error (MSE) (mean square the difference between the posterior means and the true value divided by the sample size), and the coverage probability (CP) of 95% credible intervals. Simulation results suggest that in all scenarios, CP for regression parameters is more than nominal value of 95%, but in scenarios which P is near to 1 gives biased estimates and low coverage probabilities just for the parameters of Tweedie distribution. It suggests that the proposed model performe strongly in cases where the power parameter is near 2. As the power parameter decrease, the bias of the regression and also power parameter remain low, but the bias of the dispersion parameter increase. We apply the model to data from the Household Income and Expenditure Survey (HIES) conducted annually by the Statistical Center of Iran. We collect information on the factors influencing the average household contraceptive expenditures in the provinces of Iran for the years 2011-2018. Based on the model, divorce rate, unemployment rate, the ratio of income equity (first/last), percentage of urbanization, extended family, human development index, percentage of Muslims and percentage of immigrants were the significant factors on average household contraceptive expenditures in a province.

Conclusion

The use of spatio-temporal tweedie regression model for modeling nonnegative and skewed to the right data with mass at zero and spatial and temporal structures is recommended, but the closer the tweedie power parameter to 2, the more accurate the model performance in estimating the parameters.

Acknowledgement

The present research was a part of research towards a PhD degree in Biostatistics at the School of Public Health, Tehran University of Medical Sciences (TUMS).

References

Böhning, D., Alfò, M. (2016), Special issue on models for continuous data with a spike at zero, *Biometrical Journal*, **58**, 255-258.

Kibria, A., Mancher, M., McCoy, MA., Graham, RP., Garber, AM., Newhouse, JP. (2013). *Health Care Delivery*, National Academies Press.

Paradinas, I., Conesa, D., Lopez-Quilez, A., Bellido, JM. (2017), Spatio-temporal model structures with shared components for semi-continuous species distribution modelling, *Spatial Statistics*, **22**, 434-450.

Swallow, B., Buckland, ST., King, R., Toms, MP. (2016), Bayesian hierarchical modelling of continuous nonnegative longitudinal data with a spike at zero: An application to a study of birds visiting gardens in winter, *Biometrical Journal*, **58**, 357-371.