



Big Data Inferences and Modelling for Official Statistics with Application to Satellite Imagery Analysis

Siu-Ming Tam*

Australian Bureau of Statistics, Australia – Siu-Ming.Tam@abs.gov.au

Paul Schubert

Australian Bureau of Statistics, Australia – Paul.Schubert@abs.gov.au

Most, if not all, Big Data sources suffer from coverage biases, self-selection biases, representation biases, response biases or measurement errors.

To address these issues, this paper will present a Bayesian framework for making Big Data inferences, based on conceptualised transformation, sampling and censoring processes applied to the Big Data measurements.

Proper inference will require modelling of all three processes, which can be very complex, if at all possible. However, where certain sampling and censoring ignorability conditions are fulfilled, inference can be made on the Big Data measurements as if they are acquired from a random sample.

Finally, the paper will apply the Bayesian framework to the analysis of satellite imagery data for predicting crop areas and yields, by employing a binomial logistic regression model for the selected crop type, a log linear model for the crop quantities, and an autoregressive model for the regression coefficients of the regression coefficients of both models.

Keywords: Big Data; dynamic logistic regression models; Kalman Filters; satellite imagery analysis.