Code for Empirical Bayes Small Area Estimation Under a Zero-inflated Lognormal Model with correlated random area effects

Xiaodan Lyu, Emily Berg, Heike Hofmann

R package "saezero"

An R package called "saezero" has been developed for implementing the methodologies introduced in the paper, which includes the following functions

- `as.2pdata`: convert a data frame to a list made for fitting the two-part model proposed in the paper;
- `mleEBH`: obtain maximum likelihood estimates under our model assumption;
- `ebLBH`: obtain Empirical Bayes small area predictor and associated one-step MSE estimator;
- `simLBH`: simulate responses under our model assumption.

The packages is available and maintained at https://github.com/XiaodanLyu/saezero.

How to execute the code

This folder contains R code for running both the simulation and the case study. Below is the file structure.

```
|-- README.pdf
|-- case_study
|   |-- data
|   |-- case_study.R
|   |-- case_study.pdf
|   `-- case_study.rmd
|-- simulation
|   |-- intermediate_results
|   |-- simulation.R
|   |-- simulation_results.Rmd
|   |-- simulation_results.pdf
|   `-- utility-functions.R
```

Simulation

The file `utility-functions.R` contains helper functions for the simulation. The file `simulation.R` is the main function to be executed and results in the output data files in the folder `intermediate_results`. The file `simluation_results.Rmd` contains the output and the R code for reproducing the tables and images related to the simulation section (Section 4 in the paper) and results in `simulation_results.pdf`.

Case study

Our R package "saezero" provides two data sets:

- `Xaux`: the auxiliary information for predicting cropland RUSLE2;
- `erosion`: the simulated data that mimics the real data and produces similar results.

For detailed description of the two data sets, please refer to the package reference manual or run `?saezero::erosion` and `?saezero::Xaux` in R. The way we produced the data is given in the Section 4.1 of our paper. The file `case_study.rmd` combines the output and the R code (`case_study.R`) for reproducing the tables and images related to the data analysis section (Section 4 in the paper) and results in `case_study.pdf`. The folder `data` contains an intermediate data file `eb_mmse_boot.RData` that contains the parametric bootstrap results of the soil erosion data.

Author information

Xiaodan Lyu is mainly responsible for writing the R code of this paper. Readers can email the author at annielyu8@gmail.com for any questions, comments and remarks on the code. Readers can also report bugs at https://github.com/XiaodanLyu/zero-sae-bj-2020.

Computing Platform

For the parametric bootstrap part, 25 cores were used to speed up the computing process. The results shall be reproducible on a high performance computer using the same random seed and 25 cores. All the simulation code was run under the following configuration:

```
R version 3.5.0 (2018-04-23)
Platform: x86_64-pc-linux-gnu (64-bit)
Running under: Red Hat Enterprise Linux

Matrix products: default
BLAS/LAPACK: /opt/rit/spack-app/linux-rhel7-x86_64/gcc-4.8.5/openblas-0.3.3-vbnrrlvu244vbbzbjg45z4zudqv

locale:
 [1] LC_CTYPE=en_US.UTF-8       LC_NUMERIC=C
 [3] LC_TIME=en_US.UTF-8        LC_COLLATE=en_US.UTF-8
 [5] LC_MONETARY=en_US.UTF-8    LC_MESSAGES=en_US.UTF-8
 [7] LC_PAPER=en_US.UTF-8       LC_NAME=C
 [9] LC_ADDRESS=C               LC_TELEPHONE=C
[11] LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C

attached base packages:
[1] stats     graphics  grDevices utils     datasets  methods   base

other attached packages:
[1] lme4_1.1-21   Matrix_1.2-17 dplyr_0.7.5   saezero_0.1.0

loaded via a namespace (and not attached):
 [1] Rcpp_0.12.16     lattice_0.20-38  assertthat_0.2.0 MASS_7.3-49
 [5] grid_3.5.0       R6_2.2.2         nlme_3.1-137     magrittr_1.5
 [9] rlang_0.2.2      minqa_1.2.4      nloptr_1.2.2     bindrcpp_0.2.2
[13] boot_1.3-20      splines_3.5.0    glue_1.2.0       purrr_0.2.4
[17] compiler_3.5.0   pkgconfig_2.0.1  bindr_0.1.1      tidyselect_0.2.3
[21] tibble_1.3.4
```