

Poverty Mapping in Sri Lanka

Zewei (Whiskey) Liao

2022-08-31

Contents

Preprocess the Data	1
Frequentist Models	1
Bayesian Models	4
Fitting the INLA-Besag Model	4
Validating the Model against DSD-level Poverty Headcount Index	5

Preprocess the Data

The following codes are largely adapted from `pre-process.R`.

```
## Reading layer `sri_lanka_gnd' from data source
##   `/Users/wliao0504/OneDrive - The University of Chicago/Repo/LIRNEasia/poverty-mapping/data/sri_lanka_gnd.shp'
##   using driver `ESRI Shapefile'
## Simple feature collection with 14022 features and 5 fields
## Geometry type: MULTIPOLYGON
## Dimension:      XY
## Bounding box:   xmin: 351691.4 ymin: 654223.4 xmax: 596874.6 ymax: 1087294
## Projected CRS: WGS 84 / UTM zone 44N
```

Frequentist Models

```
##
## Lagrange multiplier diagnostics for spatial dependence
##
## data:
## model: lm(formula = pc_1 ~ ., data = data_train)
## weights: neighbour_train
##
## LMerr = 626.95, df = 1, p-value < 2.2e-16
##
## Lagrange multiplier diagnostics for spatial dependence
```

```

##
## data:
## model: lm(formula = pc_1 ~ ., data = data_train)
## weights: neighbour_train
##
## LMLag = 323.55, df = 1, p-value < 2.2e-16
##
##
## Lagrange multiplier diagnostics for spatial dependence
##
## data:
## model: lm(formula = pc_1 ~ ., data = data_train)
## weights: neighbour_train
##
## RLMerr = 309.96, df = 1, p-value < 2.2e-16
##
##
## Lagrange multiplier diagnostics for spatial dependence
##
## data:
## model: lm(formula = pc_1 ~ ., data = data_train)
## weights: neighbour_train
##
## RLMlag = 6.5599, df = 1, p-value = 0.01043
##
##
## Lagrange multiplier diagnostics for spatial dependence
##
## data:
## model: lm(formula = pc_1 ~ ., data = data_train)
## weights: neighbour_train
##
## SARMA = 633.51, df = 2, p-value < 2.2e-16

##
## Likelihood ratio for spatial linear models
##
## data:
## Likelihood ratio = 155.88, df = 27, p-value < 2.2e-16
## sample estimates:
## Log likelihood of model_sdem   Log likelihood of model_sem
##                -20965.58                -21043.52

##
## Likelihood ratio for spatial linear models
##
## data:
## Likelihood ratio = 506.19, df = 1, p-value < 2.2e-16
## sample estimates:
## Log likelihood of model_sdem   Log likelihood of model_slx
##                -20965.58                -21218.67

##

```

```

## Likelihood ratio for spatial linear models
##
## data:
## Likelihood ratio = 720.41, df = 28, p-value < 2.2e-16
## sample estimates:
## Log likelihood of model_sdem   Log likelihood of model_ols
##                               -20965.58                     -21325.78

##
## studentized Breusch-Pagan test
##
## data:
## BP = 848.95, df = 54, p-value < 2.2e-16

##
## Call:
## errorsarlm(formula = pc_1 ~ ., data = data_train, listw = neighbour_train,
##             Durbin = TRUE, zero.policy = TRUE)
##
## Residuals:
##           Min           1Q       Median           3Q          Max
## -7.9661182 -0.9704006 -0.0089015  0.9541277 15.6131207
##
## Type: error
## Regions with no neighbours included:
## 2139470 2103120 2118085 5136060 1203165 1321135 1203120 1224450 3327070 4130265 4103015 1333130 131
## Coefficients: (asymptotic standard errors)
##
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -0.0108473   0.0207727  -0.5222 0.6015396
## avg_nighttime_call_duration      0.0618961   0.0318869   1.9411 0.0522444
## radius_of_gyration_log      -0.1554168   0.0352048  -4.4146 1.012e-05
## unique_tower_count      0.7886468   0.0518688  15.2047 < 2.2e-16
## spatial_entropy      0.0651196   0.0206096   3.1597 0.0015794
## avg_call_count_per_contact      -0.2110109   0.0365481  -5.7735 7.764e-09
## avg_call_duration_per_contact      0.1178247   0.0339156   3.4741 0.0005127
## contact_count      -0.1000210   0.0483538  -2.0685 0.0385906
## social_entropy      -0.0464305   0.0283035  -1.6405 0.1009115
## travel_time_major_cities_log      -0.2534029   0.0315154  -8.0406 8.882e-16
## population_count_ciesin_log      -0.2104062   0.0252857  -8.3211 < 2.2e-16
## population_density_log      1.3793794   0.0417804  33.0150 < 2.2e-16
## evapotranspiration      -0.1473175   0.0444864  -3.3115 0.0009279
## elevation_log      -0.0497755   0.0443710  -1.1218 0.2619472
## vegetation      0.0384752   0.0283046   1.3593 0.1740418
## distance_roadways_motorway      -0.0123197   0.0342025  -0.3602 0.7186990
## distance_roadways_primary_log      -0.2080477   0.0191633 -10.8566 < 2.2e-16
## distance_roadways_tertiary_log      -0.2558403   0.0181921 -14.0633 < 2.2e-16
## distance_waterways_log      -0.0657847   0.0172099  -3.8225 0.0001321
## urban_rural_fb_log      0.5433002   0.0335116  16.2123 < 2.2e-16
## urban_rural_ciesin      0.2174607   0.0237596   9.1526 < 2.2e-16
## protected_areas_log      -0.0071931   0.0184147  -0.3906 0.6960798
## land_cover_woodland      -0.2743202   0.0491636  -5.5797 2.409e-08
## land_cover_grassland_log      -0.0667051   0.0223223  -2.9883 0.0028056
## land_cover_cropland      -0.2488964   0.0380301  -6.5447 5.961e-11
## pregnancies_log      0.2500130   0.0237921  10.5082 < 2.2e-16

```

```

## precipitation          0.2249137  0.0390617   5.7579 8.516e-09
## temperature           -0.0061307  0.0223063  -0.2748 0.7834390
## lag.avg_nighttime_call_duration -0.0279169  0.0560175  -0.4984 0.6182294
## lag.radius_of_gyration_log      -0.0268633  0.0564575  -0.4758 0.6342068
## lag.unique_tower_count          0.2926737  0.0801805   3.6502 0.0002621
## lag.spatial_entropy            -0.0733832  0.0356233  -2.0600 0.0394005
## lag.avg_call_count_per_contact   -0.2732304  0.0589354  -4.6361 3.550e-06
## lag.avg_call_duration_per_contact 0.1042731  0.0588795   1.7710 0.0765676
## lag.contact_count              -0.1186167  0.0793460  -1.4949 0.1349328
## lag.social_entropy              0.0219028  0.0492614   0.4446 0.6565915
## lag.travel_time_major_cities_log -0.0449946  0.0489266  -0.9196 0.3577636
## lag.population_count_ciesin_log  -0.1217125  0.0463556  -2.6256 0.0086490
## lag.population_density_log       0.0612430  0.0718201   0.8527 0.3938102
## lag.evapotranspiration           0.1877668  0.0658382   2.8519 0.0043453
## lag.elevation_log               0.3010859  0.0656787   4.5842 4.557e-06
## lag.vegetation                 0.0648550  0.0481339   1.3474 0.1778559
## lag.distance_roadways_motorway   -0.1301763  0.0469350  -2.7735 0.0055450
## lag.distance_roadways_primary_log -0.0312287  0.0346807  -0.9005 0.3678732
## lag.distance_roadways_tertiary_log 0.0610863  0.0347092   1.7599 0.0784175
## lag.distance_waterways_log       -0.0052836  0.0316974  -0.1667 0.8676161
## lag.urban_rural_fb_log           0.0363273  0.0627610   0.5788 0.5627107
## lag.urban_rural_ciesin           0.1281148  0.0393031   3.2597 0.0011155
## lag.protected_areas_log          -0.0284629  0.0350893  -0.8112 0.4172773
## lag.land_cover_woodland          0.0358925  0.0883304   0.4063 0.6844903
## lag.land_cover_grassland_log      0.0332351  0.0415537   0.7998 0.4238202
## lag.land_cover_cropland           0.0317661  0.0680328   0.4669 0.6405551
## lag.pregnancies_log              0.0151269  0.0400172   0.3780 0.7054229
## lag.precipitation               -0.0848478  0.0536413  -1.5818 0.1137041
## lag.temperature                 0.1169126  0.0413291   2.8288 0.0046720
##
## Lambda: 0.28139, LR test value: 506.19, p-value: < 2.22e-16
## Asymptotic standard error: 0.012224
##      z-value: 23.019, p-value: < 2.22e-16
## Wald statistic: 529.88, p-value: < 2.22e-16
##
## Log likelihood: -20965.58 for error model
## ML residual variance (sigma squared): 2.4797, (sigma: 1.5747)
## Nagelkerke pseudo-R-squared: 0.76632
## Number of observations: 11138
## Number of parameters estimated: 57
## AIC: 42045, (AIC for lm: 42549)

## [1] 0.7393339

## [1] 0.7434229

```

Bayesian Models

Fitting the INLA-Besag Model

```
## [1] 0.8448995
```

```
## [1] 0.7638455
## [1] 0.8531872
## [1] 0.7494516
## [1] 0.8309729
## [1] 0.5937772
## [1] 1.296221
## [1] 1.612706
```

Validating the Model against DSD-level Poverty Headcount Index

```
## [1] 0.2648532
## [1] 0.6343936

##
## Spearman's rank correlation rho
##
## data: final.df$hci_std and final.df$fit_wted_std
## S = 2209747, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##      rho
## 0.6343936

##
## Call:
## lm(formula = final.df$hci_std ~ final.df$fit_wted_std)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.3599 -0.2744  0.0975  0.5072  1.4436
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -4.969e-16  4.264e-02   0.00      1
## final.df$fit_wted_std  6.324e-01  4.271e-02  14.81 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7758 on 329 degrees of freedom
## Multiple R-squared:  0.4, Adjusted R-squared:  0.3981
## F-statistic: 219.3 on 1 and 329 DF, p-value: < 2.2e-16
```

```
##
## Call:
## lm(formula = final.df$headcount_index ~ final.df$fit_wted)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -9.0287 -3.1724 -0.6097  1.7160 27.2681
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      9.10403    0.26890   33.86  <2e-16 ***
## final.df$fit_wted -1.42658    0.09633  -14.81  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.852 on 329 degrees of freedom
## Multiple R-squared:  0.4, Adjusted R-squared:  0.3981
## F-statistic: 219.3 on 1 and 329 DF, p-value: < 2.2e-16
```

	All	RS Only	CDR Only	DSD Level
OLS: Train		0.7148288	0.4838027	
OLS: Validate	0.7393339			
SDEM: Train				
SDEM: Validate	0.7434229			
INLA-Besag: Train	0.8448995	0.8531872	0.8309729	40%
INLA-Besag: Validate	0.7638455	0.7494516	0.5937772	