Online Supplementary Materials

Supplementary S1

Table S1. Estimation accuracy metrics for SLMs under varying sample sizes and spatial dependence strength

Bayesian techniques	ρ_{SAR}	Para meters	Observations $n = 100$				Observations $n = 225$				Observations $n = 900$				Observations $n = 2500$			
			Mean	SD	RMSE	MAD	Mean	SD	RMSE	MAD	Mean	SD	RMSE	MAD	Mean	SD	RMSE	MAD
HMC	0.2	ρ_{SEM}	0.196	0.032	0.03	0.024	0.202	0.025	0.025	0.02	0.198	0.012	0.012	0.01	0.2	0.007	0.006	0.005
		σ	1.002	0.074	0.069	0.056	1.009	0.048	0.048	0.036	1.005	0.024	0.025	0.02	1.006	0.014	0.014	0.01
		β_1	1.989	0.056	0.061	0.049	2.004	0.037	0.036	0.029	2.001	0.017	0.018	0.015	1.998	0.01	0.009	0.007
		β_2	-3	0.118	0.119	0.095	-3.003	0.067	0.065	0.054	-3.002	0.032	0.034	0.027	-3.006	0.021	0.018	0.014
	0.5	ρ_{SEM}	0.498	0.027	0.026	0.021	0.496	0.021	0.018	0.014	0.499	0.01	0.01	0.008	0.5	0.006	0.005	0.004
		σ	1.015	0.075	0.083	0.064	1	0.048	0.047	0.038	0.996	0.024	0.024	0.018	1.006	0.014	0.014	0.01
		β_1	1.996	0.057	0.053	0.043	1.999	0.037	0.037	0.029	1.999	0.018	0.017	0.014	1.998	0.01	0.009	0.009
		β_2	-2.973	0.12	0.133	0.105	-3.005	0.067	0.066	0.052	-2.995	0.032	0.032	0.025	-3.007	0.021	0.019	0.014
	0.8	ρ_{SEM}	0.797	0.015	0.014	0.011	0.797	0.012	0.012	0.01	0.8	0.006	0.006	0.005	0.8	0.003	0.004	0.003
		σ	1.009	0.074	0.072	0.059	1.007	0.038	0.036	0.029	2.003	0.018	0.015	0.012	1.998	0.021	0.022	0.017
		β_1	2.007	0.057	0.047	0.039	2.003	0.038	0.036	0.029	2.003	0.018	0.015	0.012	1.998	0.01	0.009	0.007
		β_2	-3.019	0.12	0.115	0.088	-3.004	0.068	0.061	0.048	-3.002	0.033	0.031	0.025	-3.008	0.021	0.022	0.017
VB	0.2	ρ_{SEM}	0.206	0.031	0.033	0.025	0.197	0.024	0.029	0.021	0.201	0.011	0.016	0.014	0.198	0.007	0.012	0.01
		σ	1.03	0.075	0.089	0.064	1.012	0.05	0.052	0.042	1.005	0.025	0.035	0.027	0.999	0.016	0.025	0.019
		β_1	2	0.058	0.071	0.059	2.006	0.039	0.046	0.037	1.999	0.018	0.028	0.023	2.005	0.011	0.023	0.018
		β_2	-3.005	0.118	0.119	0.091	-2.997	0.069	0.079	0.061	-3.005	0.034	0.04	0.03	-2.992	0.021	0.032	0.024
	0.5	ρ_{SEM}	0.497	0.024	0.031	0.025	0.497	0.019	0.024	0.019	0.5	0.009	0.014	0.011	0.499	0.006	0.009	0.008
		σ	1.015	0.075	0.081	0.063	1.006	0.049	0.07	0.056	1.001	0.025	0.03	0.03	0.998	0.016	0.026	0.022
		β_1	2.009	0.055	0.059	0.047	1.994	0.038	0.048	0.039	1.998	0.019	0.027	0.021	2.004	0.011	0.023	0.018
		β_2	-2.982	0.117	0.128	0.104	-2.998	0.07	0.081	0.061	-2.997	0.034	0.039	0.032	-2.998	0.022	0.03	0.025
	0.8	ρ_{SEM}	0.796	0.012	0.016	0.012	0.799	0.009	0.013	0.01	0.799	0.004	0.007	0.005	0.8	0.003	0.005	0.004
		σ	1.026	0.076	0.087	0.069	1.016	0.051	0.057	0.044	1.006	0.025	0.033	0.026	1.001	0.015	0.026	0.021
		β_1	2.012	0.057	0.067	0.054	2.006	0.038	0.044	0.035	2.005	0.019	0.026	0.02	2.002	0.011	0.021	0.017
		β_2	-2.993	0.119	0.132	0.107	-2.994	0.071	0.08	0.061	-3	0.035	0.049	0.035	-3.006	0.022	0.034	0.026
INLA	0.2	ρ_{SEM}	0.199	0.094	0.014	0.012	0.195	0.085	0.024	0.02	0.199	0.045	0.012	0.01	0.201	0.028	0.008	0.007
		σ	1.228	0.211	0.323	0.182	1.178	0.133	0.221	0.103	1.169	0.065	0.18	0.051	1.163	0.038	0.167	0.03
		β_1	2.305	1.156	0.653	0.462	2.403	1.391	0.614	0.369	2.105	0.613	0.236	0.17	2.039	0.318	0.138	0.106
		β_2	-3.171	1.621	1.024	0.766	-3.589	2.044	1.083	0.723	-3.157	0.948	0.417	0.293	-3.013	0.495	0.225	0.18
	0.5	ρ_{SEM}	0.5	0.069	0.017	0.015	0.5	0.054	0.013	0.011	0.502	0.026	0.007	0.006	0.503	0.016	0.007	0.005
		σ	1.181	0.206	0.255	0.142	1.175	0.134	0.216	0.102	1.156	0.065	0.166	0.046	1.159	0.039	0.164	0.032
		β_1	2.031	0.437	0.238	0.199	2.016	0.307	0.145	0.117	1.989	0.143	0.075	0.058	1.98	0.085	0.055	0.042
		β_2	-2.999	0.747	0.436	0.335	-3.023	0.518	0.305	0.251	-3.022	0.235	0.143	0.111	-2.994	0.142	0.099	0.078
	0.8	ρ_{SEM}	0.801	0.024	0.006	0.005	0.799	0.021	0.005	0.004	0.8	0.01	0.002	0.002	0.801	0.006	0.002	0.002
		σ	1.206	0.208	0.288	0.16	1.162	0.131	0.209	0.103	1.162	0.064	0.175	0.052	1.157	0.038	0.163	0.034
		β_1	2.005	0.2	0.144	0.117	2.01	0.137	0.086	0.066	1.995	0.067	0.044	0.035	1.994	0.039	0.027	0.021
		β_2	-3.026	0.352	0.309	0.239	-3.022	0.25	0.18	0.141	-3.013	0.116	0.089	0.072	-2.998	0.07	0.049	0.039

Table S2. Estimation accuracy metrics for SEMs under varying sample sizes and spatial dependence strength

Bayesian techniques	ρ_{SEM}	Param-eters	Observations $n = 100$				Observations $n = 225$				Observations $n = 900$				Observations $n = 2500$			
			Mean	SD	RMSE	MAD	Mean	SD	RMSE	MAD	Mean	SD	RMSE	MAD	Mean	SD	RMSE	MAD
HMC	0.2	ρ_{SEM}	0.212	0.133	0.117	0.096	0.203	0.089	0.087	0.071	0.196	0.045	0.042	0.033	0.198	0.027	0.029	0.024
		σ	1.007	0.074	0.075	0.061	1.006	0.048	0.048	0.038	1.004	0.024	0.025	0.02	1.003	0.014	0.013	0.01
		β_1	2	0.055	0.058	0.045	2.005	0.036	0.037	0.029	2.001	0.017	0.018	0.014	1.997	0.01	0.009	0.007
		β_2	-2.998	0.118	0.118	0.095	-3.001	0.067	0.075	0.062	-3.001	0.032	0.033	0.026	-3.008	0.02	0.021	0.016
	0.5	ρ_{SEM}	0.478	0.114	0.109	0.087	0.499	0.074	0.071	0.058	0.492	0.038	0.039	0.031	0.497	0.023	0.02	0.017
		σ	1.02	0.077	0.073	0.058	1.005	0.049	0.056	0.045	0.997	0.024	0.024	0.019	1.006	0.015	0.014	0.01
		β_1	1.991	0.052	0.052	0.042	1.998	0.035	0.039	0.03	1.999	0.017	0.016	0.013	1.998	0.01	0.008	0.006
		β_2	-3	0.117	0.122	0.094	-3.009	0.067	0.07	0.057	-2.996	0.031	0.031	0.024	-3.007	0.02	0.018	0.013
	0.8	ρ_{SEM}	0.769	0.072	0.082	0.061	0.789	0.046	0.04	0.031	0.794	0.023	0.022	0.017	0.798	0.014	0.011	0.009
		σ	1.013	0.079	0.072	0.058	1.007	0.051	0.055	0.045	1.006	0.025	0.026	0.021	1.006	0.015	0.014	0.01
		β_1	1.999	0.048	0.047	0.037	2.001	0.034	0.034	0.027	2.001	0.016	0.016	0.014	1.998	0.009	0.008	0.006
		β_2	-3.018	0.112	0.108	0.089	-3	0.065	0.063	0.051	-3.001	0.03	0.029	0.022	-3.007	0.019	0.017	0.013
VB	0.2	ρ_{SEM}	0.198	0.13	0.125	0.094	0.192	0.09	0.091	0.069	0.197	0.046	0.05	0.041	0.204	0.028	0.033	0.027
		σ	1.03	0.077	0.084	0.06	1.003	0.05	0.055	0.043	1.001	0.025	0.037	0.029	0.999	0.015	0.026	0.02
		β_1	2.005	0.055	0.058	0.045	1.995	0.038	0.04	0.032	2.001	0.018	0.027	0.022	1.996	0.011	0.023	0.018
		β_2	-2.978	0.118	0.127	0.097	-2.994	0.07	0.076	0.059	-3.006	0.033	0.046	0.037	-2.993	0.021	0.034	0.027
	0.5	ρ_{SEM}	0.47	0.11	0.124	0.097	0.493	0.073	0.079	0.064	0.49	0.039	0.043	0.033	0.501	0.023	0.026	0.022
		σ	1.017	0.074	0.083	0.066	1.015	0.051	0.064	0.048	1.002	0.025	0.035	0.028	1.005	0.015	0.029	0.023
		β_1	2.005	0.052	0.059	0.046	2.004	0.036	0.045	0.035	2.003	0.018	0.027	0.021	1.995	0.01	0.025	0.021
		β_2	-3.001	0.115	0.114	0.09	-3.001	0.069	0.07	0.056	-2.996	0.033	0.039	0.031	-2.996	0.021	0.036	0.029
	0.8	ρ_{SEM}	0.764	0.066	0.087	0.066	0.787	0.044	0.05	0.037	0.796	0.022	0.024	0.018	0.8	0.013	0.015	0.012
		σ	1.023	0.076	0.084	0.064	1.015	0.05	0.059	0.045	1.004	0.025	0.037	0.029	1.001	0.015	0.027	0.021
		β_1	2.006	0.049	0.053	0.041	1.994	0.035	0.042	0.034	1.999	0.017	0.027	0.022	1.996	0.01	0.025	0.019
		β_2	-2.999	0.112	0.118	0.093	-2.998	0.068	0.066	0.053	-2.994	0.032	0.043	0.036	-2.993	0.02	0.031	0.024
INLA	0.2	ρ_{SEM}	0.239	0.099	0.105	0.08	0.217	0.075	0.07	0.053	0.198	0.042	0.039	0.03	0.198	0.027	0.023	0.019
		σ	0.988	0.071	0.081	0.064	0.981	0.047	0.053	0.041	0.999	0.024	0.024	0.019	0.996	0.014	0.013	0.01
		β_1	1.993	0.053	0.051	0.041	2.008	0.035	0.033	0.026	1.999	0.017	0.017	0.014	2.002	0.01	0.011	0.009
		β_2	-2.994	0.115	0.118	0.093	-3.004	0.066	0.068	0.052	-3.003	0.032	0.032	0.025	-3	0.02	0.017	0.012
	0.5	ρ_{SEM}	0.475	0.102	0.116	0.088	0.486	0.072	0.079	0.063	0.497	0.037	0.04	0.031	0.5	0.023	0.014	0.012
		σ	0.997	0.073	0.07	0.053	1.001	0.049	0.052	0.043	1.001	0.024	0.023	0.018	0.999	0.015	0.02	0.015
		β_1	1.996	0.051	0.055	0.045	1.997	0.035	0.029	0.023	1.999	0.017	0.016	0.012	2.002	0.01	0.01	0.009
		β_2	-3.018	0.115	0.124	0.099	-3.004	0.067	0.068	0.055	-2.999	0.031	0.035	0.028	-3.002	0.02	0.021	0.018
	0.8	ρ_{SEM}	0.77	0.066	0.087	0.063	0.783	0.045	0.052	0.04	0.794	0.023	0.025	0.02	0.798	0.014	0.012	0.009
		σ	0.997	0.075	0.086	0.063	0.996	0.05	0.05	0.038	1.001	0.025	0.027	0.021	0.995	0.015	0.015	0.011
		β_1	2.008	0.047	0.045	0.036	1.999	0.033	0.032	0.024	2	0.016	0.015	0.012	1.999	0.009	0.011	0.008
		β_2	-3.008	0.109	0.112	0.091	-3.008	0.065	0.07	0.055	-3.003	0.029	0.03	0.025	-2.996	0.019	0.017	0.013

Supplementary S2. French Municipal Expenditure (2011) Case

Last update: 2025-10-08

Contents

1 Overview **1**

2 Libraries **1**

3 Data and spatial weights **2**

4 Model specification **3**

5 Compile Stan model **4**

6 Fit Model 1: Current expenditure **4**

7 Fit Model 2: Capital expenditure **5**

8 Fit Model 3: Total expenditure **6**

9 Session info **7**

1 Overview

This supplementary provides the R implementation for three SLMs estimated via **mean-field varia-tional Bayes** in **CmdStanR**, employing an inverse-distance spatial weight matrix with a 25 km threshold.

> **Notes** - Ensure the following files exist in your working directory before knitting: - `Case1_Dta_loaded.RData` (contains `COM_IDF2`, `dat2_2011`, etc.) - `Orig_SAR3.stan` (the SAR Stan model).

2 Libraries

```r
suppressPackageStartupMessages({
library(sf)
library(spdep)
library(splm)
library(dplyr)
library(plm)
library(spatialreg)
library(ggplot2)
library(readxl)
library(cmdstanr)
library(posterior)
library(reshape2)
library(kableExtra)
library(htmlTable)
})
```

```
## Warning: package 'splm' was built under R version 4.4.3

## Warning: package 'readxl' was built under R version 4.4.2

## Warning: package 'kableExtra' was built under R version 4.4.3

## Warning: package 'htmlTable' was built under R version 4.4.3
```

3 Data and spatial weights

```r
# Load data objects (adjust the path if needed)
load("Case1_Dta_loaded.RData")

# Build 25 km inverse-distance weights
coords <- st_centroid(st_geometry(COM_IDF2), of_largest_polygon = TRUE)

dmax_25K <- 25000  # meters
nb_d_25K <- dnearneigh(coords, d1 = 0, d2 = dmax_25K, longlat = FALSE)
```

```
## Warning in dnearneigh(coords, d1 = 0, d2 = dmax_25K, longlat = FALSE):
## dnearneigh: longlat argument overrides object
```

```r
# Diagnose isolates
summary(card(nb_d_25K))
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    44.0   137.0   185.0   187.2   243.0   303.0
```

```r
# Inverse-distance lists
dist_list_25K <- nbdists(nb_d_25K, coords, longlat = FALSE)
```

```
## Warning in nbdists(nb_d_25K, coords, longlat = FALSE): dnearneigh: longlat
## argument overrides object
```

```r
glist_inv_25K <- lapply(dist_list_25K, function(d) 1 / d)

# Row-standardized W
listw_idw_25K <- nb2listw(nb_d_25K, glist = glist_inv_25K,
                          style = "W", zero.policy = TRUE)

# Sparse matrix
Mat_idw_25K <- listw2dgCMatrix(listw_idw_25K)

# A simple sparsity diagnostic
Mat_W_inv25K <- reshape2::melt(as.matrix(Mat_idw_25K))
colnames(Mat_W_inv25K) <- c("x", "y", "value")
sparsity_measure <- (1280 - sum(abs(Mat_W_inv25K$value)) /
                       sqrt(sum((Mat_W_inv25K$value)^2))) / 1279
sparsity_measure
```

```
## [1] 0.7205407
```

4 Model specification

```r
# Three empirical models
formlin_CURRENT <- ldepftc ~ ACTOCCUP1564 + SUP + lpop0010 +
  lpop1117 + lpop1824 + lpop65p

formlin_CAPITAL <- ldepinv  ~ ACTOCCUP1564 + SUP + lpop0010 +
  lpop1117 + lpop1824 + lpop65p

formlin_TOTAL   <- ldep     ~ ACTOCCUP1564 + SUP + lpop0010 +
  lpop1117 + lpop1824 + lpop65p

N  <- 1280
K  <- 7
Kc <- 6

# Design matrices (2011)
X_m1 <- model.matrix(formlin_CURRENT, dat2_2011)
X_m2 <- model.matrix(formlin_CAPITAL, dat2_2011)
X_m3 <- model.matrix(formlin_TOTAL,   dat2_2011)
```

```r
# Spatial matrix and eigenvalues
Msar       <- as.matrix(Mat_idw_25K)
eigenMsar  <- as.numeric(eigen(Mat_idw_25K)$values)

# Common Stan parameter names
param_names <- c("Intercept", "sigma", "lagsar", "b[1]",
                 "b[2]", "b[3]", "b[4]", "b[5]", "b[6]")
```

5 Compile Stan model

```r
stan_file <- "Orig_SAR3.stan"
mod <- cmdstan_model(stan_file)
```

```
## Warning in readLines(stan_file): incomplete final line found on
## 'Orig_SAR3.stan'
```

6 Fit Model 1: Current expenditure

```r
data_list1 <- list(
  N = N, Y = as.vector(dat2_2011$ldepftc),
  K = K, X = X_m1, Kc = Kc,
  Msar = Msar, eigenMsar = eigenMsar, prior_only = 0
)

start_time <- Sys.time()
mod1 <- mod$variational(
  data          = data_list1,
  seed          = 10000,
  refresh       = 100,
  algorithm     = "meanfield",
  iter          = 20000,
  grad_samples  = 1,
  adapt_iter    = 125,
  draws         = 2000,
  show_messages = FALSE
)
end_time <- Sys.time()
time1 <- as.numeric(difftime(end_time, start_time,
                             units = "secs"))

draws1 <- mod1$draws(variables = param_names)
sumry1 <- posterior::summarise_draws(
```

```
  draws1,
  mean, median, sd, ~quantile(.x, probs = c(0.025, 0.975))
)

sumry1
```

```
## # A tibble: 9 x 6
##   variable      mean   median       sd    `2.5%`  `97.5%`
##   <chr>        <dbl>    <dbl>    <dbl>     <dbl>    <dbl>
## 1 Intercept  0.489    0.489   0.00856   0.472     0.505
## 2 sigma      0.301    0.301   0.00644   0.288     0.314
## 3 lagsar     0.928    0.928   0.00134   0.925     0.931
## 4 b[1]      -0.0183  -0.0182  0.00276  -0.0236   -0.0128
## 5 b[2]       0.00110  0.00109 0.000833 -0.000517  0.00277
## 6 b[3]       0.138    0.138   0.0580    0.0245    0.254
## 7 b[4]      -0.282   -0.283   0.0556   -0.395    -0.178
## 8 b[5]       0.236    0.235   0.0408    0.156     0.314
## 9 b[6]      -0.0235  -0.0238  0.0307   -0.0829    0.0387
```

```
time1
```

```
## [1] 468.7965
```

7 Fit Model 2: Capital expenditure

```
data_list2 <- list(
  N = N, Y = as.vector(dat2_2011$ldepinv),
  K = K, X = X_m2, Kc = Kc,
  Msar = Msar, eigenMsar = eigenMsar, prior_only = 0
)

start_time <- Sys.time()
mod2 <- mod$variational(
  data          = data_list2,
  seed          = 10000,
  refresh       = 0,
  algorithm     = "meanfield",
  iter          = 20000,
  grad_samples  = 1,
  adapt_iter    = 125,
  draws         = 2000,
  show_messages = FALSE
)
end_time <- Sys.time()
```

```
time2 <- as.numeric(difftime(end_time, start_time,
                             units = "secs"))

draws2 <- mod2$draws(variables = param_names)
sumry2 <- posterior::summarise_draws(
  draws2,
  mean, median, sd, ~quantile(.x, probs = c(0.025, 0.975))
)

sumry2
```

```
## # A tibble: 9 x 6
##   variable      mean   median      sd   `2.5%`  `97.5%`
##   <chr>        <dbl>    <dbl>   <dbl>    <dbl>    <dbl>
## 1 Intercept  1.25     1.25    0.0217   1.21     1.29
## 2 sigma      0.788    0.787   0.0163   0.756    0.820
## 3 lagsar     0.799    0.799   0.00326  0.793    0.805
## 4 b[1]      -0.00462 -0.00449 0.00753 -0.0192   0.0100
## 5 b[2]      -0.00178 -0.00187 0.00228 -0.00614  0.00269
## 6 b[3]       0.0599   0.0638  0.150   -0.238    0.362
## 7 b[4]      -0.285   -0.283   0.128   -0.539   -0.0409
## 8 b[5]       0.373    0.374   0.0965   0.182    0.560
## 9 b[6]      -0.143   -0.141   0.0833  -0.305    0.0156
```

```
time2
```

```
## [1] 178.1609
```

8 Fit Model 3: Total expenditure

```
data_list3 <- list(
  N = N, Y = as.vector(dat2_2011$ldep),
  K = K, X = X_m3, Kc = Kc,
  Msar = Msar, eigenMsar = eigenMsar, prior_only = 0
)

start_time <- Sys.time()
mod3 <- mod$variational(
  data          = data_list3,
  seed          = 10000,
  refresh       = 0,
  algorithm     = "meanfield",
  iter          = 20000,
  grad_samples = 1,
```

```
  adapt_iter   = 125,
  draws        = 2000,
  show_messages = FALSE
)
end_time <- Sys.time()
time3 <- as.numeric(difftime(end_time, start_time,
                             units = "secs"))

draws3 <- mod3$draws(variables = param_names)
sumry3 <- posterior::summarise_draws(
  draws3,
  mean, median, sd, ~quantile(.x, probs = c(0.025, 0.975))
)

sumry3
```

```
## # A tibble: 9 x 6
##   variable      mean   median       sd    `2.5%`  `97.5%`
##   <chr>        <dbl>    <dbl>    <dbl>     <dbl>    <dbl>
## 1 Intercept  0.979    0.979   0.0107    0.958     1.00
## 2 sigma      0.378    0.378   0.00833   0.362     0.395
## 3 lagsar     0.866    0.866   0.00158   0.863     0.869
## 4 b[1]      -0.0124  -0.0123  0.00333  -0.0188   -0.00583
## 5 b[2]       0.00177  0.00176 0.000972 -0.000116  0.00372
## 6 b[3]       0.0630   0.0639  0.0739   -0.0811    0.211
## 7 b[4]      -0.333   -0.335   0.0700   -0.476    -0.202
## 8 b[5]       0.264    0.264   0.0512    0.164     0.362
## 9 b[6]      -0.112   -0.113   0.0382   -0.186    -0.0349
```

```
time3
```

```
## [1] 468.1504
```

9 Session info

```
sessionInfo()
```

```
## R version 4.4.1 (2024-06-14 ucrt)
## Platform: x86_64-w64-mingw32/x64
## Running under: Windows 10 x64 (build 19045)
##
## Matrix products: default
##
##
```

```
## locale:
## [1] LC_COLLATE=Chinese (Simplified)_China.utf8
## [2] LC_CTYPE=Chinese (Simplified)_China.utf8
## [3] LC_MONETARY=Chinese (Simplified)_China.utf8
## [4] LC_NUMERIC=C
## [5] LC_TIME=Chinese (Simplified)_China.utf8
##
## time zone: Asia/Shanghai
## tzcode source: internal
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
##  [1] htmlTable_2.4.3  kableExtra_1.4.0 reshape2_1.4.4   posterior_1.6.0
##  [5] cmdstanr_0.8.1   readxl_1.4.3     ggplot2_3.5.1    spatialreg_1.3-5
##  [9] Matrix_1.7-0     plm_2.6-4        dplyr_1.1.4      splm_1.6-5
## [13] spdep_1.3-6      spData_2.3.3     sf_1.0-17
##
## loaded via a namespace (and not attached):
##  [1] tidyselect_1.2.1     viridisLite_0.4.2    fastmap_1.2.0
##  [4] tensorA_0.36.2.1     TH.data_1.1-2        digest_0.6.37
##  [7] dotCall64_1.1-1      lifecycle_1.0.4      LearnBayes_2.15.1
## [10] processx_3.8.4       survival_3.6-4       magrittr_2.0.3
## [13] compiler_4.4.1       rlang_1.1.4          tools_4.4.1
## [16] utf8_1.2.4           yaml_2.3.10          data.table_1.16.0
## [19] collapse_2.0.16      knitr_1.48           htmlwidgets_1.6.4
## [22] sp_2.1-4             classInt_0.4-10      xml2_1.3.6
## [25] plyr_1.8.9           abind_1.4-8          multcomp_1.4-26
## [28] KernSmooth_2.23-24   withr_3.0.1          grid_4.4.1
## [31] fansi_1.0.6          e1071_1.7-16         colorspace_2.1-1
## [34] scales_1.3.0         MASS_7.3-60.2        cli_3.6.3
## [37] mvtnorm_1.3-1        rmarkdown_2.28       miscTools_0.6-28
## [40] generics_0.1.3       rstudioapi_0.16.0    bdsmatrix_1.3-7
## [43] DBI_1.2.3            proxy_0.4-27         stringr_1.5.1
## [46] splines_4.4.1        parallel_4.4.1       s2_1.1.7
## [49] cellranger_1.1.0     vctrs_0.6.5          boot_1.3-30
## [52] sandwich_3.1-1       jsonlite_1.8.9       Formula_1.2-5
## [55] systemfonts_1.3.1    units_0.8-5          glue_1.8.0
## [58] spam_2.10-0          ps_1.8.0             codetools_0.2-20
## [61] distributional_0.5.0 stringi_1.8.4        gtable_0.3.5
## [64] deldir_2.0-4         lmtest_0.9-40        munsell_0.5.1
## [67] tibble_3.2.1         pillar_1.9.0         htmltools_0.5.8.1
## [70] R6_2.5.1             textshaping_0.4.0    wk_0.9.3
## [73] maxLik_1.5-2.1       Rdpack_2.6.1         evaluate_1.0.0
## [76] lattice_0.22-6       backports_1.5.0      rbibutils_2.2.16
## [79] class_7.3-22         Rcpp_1.0.13          svglite_2.2.1
## [82] coda_0.19-4.1        nlme_3.1-164         checkmate_2.3.2
## [85] xfun_0.47            zoo_1.8-12           pkgconfig_2.0.3
```

Supplementary S3. Ames Housing Case

Last update: 2025-10-08

Contents

1 Overview **1**

2 Libraries **2**

3 Source helper functions **2**

4 Preparing Dataset **2**

5 Define Spatial Connectivity Structure (row-standardized 5-NN) **3**

6 Matrix Sparseness (diagnostic) **3**

7 SEM random effect support **3**

8 INLA SEM fit **4**

9 Hyperparameters: variance and rho marginals **5**

10 Session info **6**

1 Overview

This supplementary presents the R code used to reproduce the SEMs for the **Ames Housing** dataset using **INLA** with an `rgeneric` latent model for the spatial error structure. The workflow follows the original implementation, with minor structural refinements and formatted outputs for clarity.

> **Note:** Make sure the file `Case2_funs.R` (which defines `inla.rgeneric.SEM.model`) is in the same folder as this Rmd before knitting.

2 Libraries

```
suppressPackageStartupMessages({
library(pspatreg)
library(spdep)
library(sf)
library(ggplot2)
library(dplyr)
library(INLA)
library(splm)
library(AmesHousing)
library(reshape2)
  })
```

```
## Warning: package 'splm' was built under R version 4.4.3
```

3 Source helper functions

```
if (!file.exists("Case2_funs.R")) {
  stop("Missing required file: Case2_funs.R.
       Please place it in the same directory as this Rmd.")
}
source("Case2_funs.R")
```

4 Preparing Dataset

```
# Raw Ames Housing Data
ames <- AmesHousing::make_ames()

# Convert to sf using lon/lat
ames_sf <- sf::st_as_sf(ames, coords = c("Longitude", "Latitude"), crs = 4326)

# Keep original lon/lat as attributes for convenience
ames_sf$Longitude <- ames$Longitude
ames_sf$Latitude  <- ames$Latitude

# Log transforms
ames_sf$lnSale_Price    <- log(ames_sf$Sale_Price)
ames_sf$lnLot_Area      <- log(ames_sf$Lot_Area)
ames_sf$lnTotal_Bsmt_SF <- log(ames_sf$Total_Bsmt_SF + 1)
ames_sf$lnGr_Liv_Area   <- log(ames_sf$Gr_Liv_Area)
```

```r
# Drop duplicated coordinates (if any)
ames_sf1 <- ames_sf[duplicated(ames_sf$Longitude) == FALSE,]

coord_sf1 <- cbind(ames_sf1$Longitude, ames_sf1$Latitude)

n <- nrow(ames_sf1)
ames_sf1$idx <- 1:n
```

5 Define Spatial Connectivity Structure (row-standardized 5-NN)

```r
# 5-nearest neighbor with symmetry; coords are lon/lat (longlat = TRUE)
k5nb <- spdep::knn2nb(spdep::knearneigh(coord_sf1, k = 5,
                                        longlat = TRUE, use_kd_tree = FALSE), sym = TRUE)
```

```
## Warning in spdep::knn2nb(spdep::knearneigh(coord_sf1, k = 5, longlat = TRUE, :
## neighbour object has 26 sub-graphs
```

```r
# Row-standardized weights
lw_ames <- spdep::nb2listw(k5nb, style = "W", zero.policy = FALSE)

# Sparse matrix form
W_ames <- splm::listw2dgCMatrix(lw_ames)
```

6 Matrix Sparseness (diagnostic)

```r
Mat_W_ames <- reshape2::melt(as.matrix(W_ames))
colnames(Mat_W_ames) <- c("x", "y", "value")

sparsity_measure <- (2777 - sum(abs(Mat_W_ames$value)) /
                       sqrt(sum((Mat_W_ames$value)^2))) / 2776
sparsity_measure
```

```
## [1] 0.9540157
```

7 SEM random effect support

```r
e.values <- eigen(W_ames)$values
e.values <- as.numeric(e.values)
```

```
## Warning: imaginary parts discarded in coercion
```

```
rho.min <- min(e.values)
rho.max <- max(e.values)
```

8 INLA SEM fit

```
# Define rgeneric SEM model using W_ames (requires Case2_funs.R definition)
SEM.model <- INLA::inla.rgeneric.define(inla.rgeneric.SEM.model, W = W_ames)

# Fix observation noise to (near) zero-variance if desired (as in your example)
zero.variance <- list(prec = list(initial = 15, fixed = TRUE))

# Fit model
sem_mod1 <- INLA::inla(
  lnSale_Price ~ lnLot_Area + lnTotal_Bsmt_SF +
    lnGr_Liv_Area + Garage_Cars + Fireplaces +
    f(idx, model = SEM.model, constr = TRUE),
  data = sf::st_drop_geometry(ames_sf1),
  verbose = F,
  family = "gaussian",
  control.fixed = list(
    mean.intercept = 0,
    prec.intercept = 1/100
  ),
  control.family = list(hyper = zero.variance),
  control.inla = list(strategy = "adaptive", int.strategy = "auto"),
  num.threads = 32,
  control.compute = list(openmp.strategy = "default", smtp = "default",
                         dic = TRUE, waic = TRUE, cpo = TRUE)
)

summary(sem_mod1)
```

```
## Time used:
##     Pre = 0.281, Running = 340, Post = 0.122, Total = 340
## Fixed effects:
##                   mean    sd 0.025quant 0.5quant 0.975quant  mode kld
## (Intercept)      7.550 0.125      7.305    7.550      7.794 7.550   0
## lnLot_Area       0.067 0.010      0.047    0.067      0.087 0.067   0
## lnTotal_Bsmt_SF 0.048 0.003      0.042    0.048      0.054 0.048   0
## lnGr_Liv_Area   0.462 0.015      0.432    0.462      0.491 0.462   0
## Garage_Cars      0.088 0.006      0.076    0.088      0.101 0.088   0
## Fireplaces       0.058 0.006      0.046    0.058      0.071 0.058   0
##
```

```
## Random effects:
##   Name     Model
##     idx RGeneric2
##
## Model hyperparameters:
##                  mean    sd 0.025quant 0.5quant 0.975quant  mode
## Theta1 for idx 3.447 0.028      3.392    3.447      3.502 3.446
## Theta2 for idx 0.769 0.068      0.636    0.769      0.903 0.768
##
## Deviance Information Criterion (DIC) ...............: -30997.24
## Deviance Information Criterion (DIC, saturated) ....: 5553.97
## Effective number of parameters .....................: 2776.97
##
## Watanabe-Akaike information criterion (WAIC) ...: -31849.34
## Effective number of parameters .................: 1388.50
##
## Marginal log-Likelihood:  632.06
## CPO, PIT is computed
## Posterior summaries for the linear predictor and the fitted values are computed
## (Posterior marginals needs also 'control.compute=list(return.marginals.predictor=TRUE)')
```

9 Hyperparameters: variance and rho marginals

```
# Convert marginal of theta1 to sigma via transformation
marg.var.semm1 <- INLA::inla.tmarginal(function(x) { sqrt(exp(-x)) },
                                       sem_mod1$marginals.hyperpar$`Theta1 for idx`)

# Convert theta2 to rho via logistic
marg.rho.semm1 <- INLA::inla.tmarginal(function(x) { 1/(1 + exp(-x)) },
                                       sem_mod1$marginals.hyperpar$`Theta2 for idx`)

INLA::inla.zmarginal(marg.var.semm1, silent = FALSE)
```

```
## Mean            0.178484
## Stdev           0.00245631
## Quantile  0.025 0.173679
## Quantile  0.25  0.176811
## Quantile  0.5   0.178475
## Quantile  0.75  0.180141
## Quantile  0.975 0.183325
```

```
INLA::inla.zmarginal(marg.rho.semm1, silent = FALSE)
```

```
## Mean            0.683161
```

```
## Stdev           0.0145478
## Quantile  0.025 0.654191
## Quantile  0.25  0.673347
## Quantile  0.5   0.683264
## Quantile  0.75  0.693081
## Quantile  0.975 0.711306
```

10 Session info

```
sessionInfo()
```

```
## R version 4.4.1 (2024-06-14 ucrt)
## Platform: x86_64-w64-mingw32/x64
## Running under: Windows 10 x64 (build 19045)
##
## Matrix products: default
##
##
## locale:
## [1] LC_COLLATE=Chinese (Simplified)_China.utf8
## [2] LC_CTYPE=Chinese (Simplified)_China.utf8
## [3] LC_MONETARY=Chinese (Simplified)_China.utf8
## [4] LC_NUMERIC=C
## [5] LC_TIME=Chinese (Simplified)_China.utf8
##
## time zone: Asia/Shanghai
## tzcode source: internal
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
##  [1] reshape2_1.4.4    AmesHousing_0.0.4 splm_1.6-5        INLA_24.10.02
##  [5] sp_2.1-4          Matrix_1.7-0      dplyr_1.1.4       ggplot2_3.5.1
##  [9] spdep_1.3-6       sf_1.0-17         spData_2.3.3      pspatreg_1.1.2
##
## loaded via a namespace (and not attached):
##  [1] tidyselect_1.2.1    viridisLite_0.4.2   fields_16.3
##  [4] fastmap_1.2.0       TH.data_1.1-2       digest_0.6.37
##  [7] dotCall64_1.1-1     lifecycle_1.0.4     Deriv_4.1.6
## [10] LearnBayes_2.15.1   survival_3.6-4      magrittr_2.0.3
## [13] compiler_4.4.1      rlang_1.1.4         tools_4.4.1
## [16] utf8_1.2.4          yaml_2.3.10         collapse_2.0.16
## [19] knitr_1.48          classInt_0.4-10     plyr_1.8.9
## [22] multcomp_1.4-26     KernSmooth_2.23-24  withr_3.0.1
```

```
## [25] numDeriv_2016.8-1.1 grid_4.4.1          fansi_1.0.6
## [28] e1071_1.7-16        colorspace_2.1-1    scales_1.3.0
## [31] MASS_7.3-60.2       cli_3.6.3           mvtnorm_1.3-1
## [34] rmarkdown_2.28      miscTools_0.6-28    generics_0.1.3
## [37] rstudioapi_0.16.0   MBA_0.1-2           bdsmatrix_1.3-7
## [40] minqa_1.2.8         DBI_1.2.3           proxy_0.4-27
## [43] stringr_1.5.1       splines_4.4.1       spatialreg_1.3-5
## [46] maps_3.4.2          parallel_4.4.1      s2_1.1.7
## [49] vctrs_0.6.5         boot_1.3-30         sandwich_3.1-1
## [52] plm_2.6-4           Formula_1.2-5       units_0.8-5
## [55] glue_1.8.0          spam_2.10-0         codetools_0.2-20
## [58] stringi_1.8.4       gtable_0.3.5        deldir_2.0-4
## [61] lmtest_0.9-40       munsell_0.5.1       tibble_3.2.1
## [64] pillar_1.9.0        htmltools_0.5.8.1   R6_2.5.1
## [67] fmesher_0.1.7       wk_0.9.3            maxLik_1.5-2.1
## [70] Rdpack_2.6.1        evaluate_1.0.0      lattice_0.22-6
## [73] rbibutils_2.2.16    MatrixModels_0.5-3  class_7.3-22
## [76] Rcpp_1.0.13         coda_0.19-4.1       nlme_3.1-164
## [79] xfun_0.47           zoo_1.8-12          pkgconfig_2.0.3
```