Stratified Analysis

Alvin Sheng

```
library(here)
```

```
## Warning in readLines(f, n): line 1 appears to contain an embedded nul
```

```
## here() starts at /Volumes/ALVINDRIVE/usb_NIEHS_work/flood-risk-health-effects
```

```
library(coda)
library(CARBayes)
```

```
## Loading required package: MASS
```

```
## Loading required package: Rcpp
```

```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
```

```
library(ggplot2)
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v tibble  3.1.6     v dplyr   1.0.7
## v tidyr   1.1.4     v stringr 1.4.0
## v readr   2.1.1     v forcats 0.5.1
## v purrr   0.3.4
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## x dplyr::select() masks MASS::select()
```

```
fhs_model_df <- readRDS("intermediary_data/fhs_model_df_all_census_tract_pc.rds")
```

Non-Spatial Modeling

CAR model results, Coronary Heart Disease Stratified on Poverty

Inference is based on 3 markov chains, each of which has been run for 110000 samples, the first 10000 of which has been removed for burn-in. The remaining 100000 samples are thinned by 2, resulting in 150000 samples for inference across the 3 Markov chains.

```
load(here("modeling_files/model_stratif_poverty.RData"))
```

Beta samples

```
beta_samples <- mcmc.list(chain1$samples$beta, chain2$samples$beta,
                          chain3$samples$beta)
```

```
effectiveSize(beta_samples)
```

```
##       var1       var2       var3       var4       var5       var6       var7
## 68447.2046 36382.4841 37761.1319 53281.5366 68914.2531 78405.5864 87890.6361
##       var8       var9      var10      var11      var12      var13      var14
## 41973.9980 61353.3136 50147.4980 60253.3373 74957.3127 89787.3447 35155.2520
##      var15      var16      var17      var18      var19      var20      var21
## 63037.1789 62136.1468 70022.6252 73724.5058 32342.7837 88478.5492 57534.7711
##      var22      var23      var24      var25      var26      var27      var28
## 10909.7609  6229.8757   323.8625  4500.4463  2003.7053  2883.0879   996.6574
##      var29      var30      var31      var32      var33      var34      var35
##   161.4431   672.4907  1167.2441 28855.9245 79627.2383 39030.9892 34364.0083
##      var36      var37      var38      var39      var40      var41      var42
## 52293.7785 66509.7765 69269.6952 76671.2019 56050.2334 50185.4620 50495.6724
##      var43      var44      var45      var46      var47      var48      var49
## 60423.8387 60221.4781 80050.1425 28753.4787 41256.5990 55529.6434 37611.9176
##      var50      var51      var52      var53      var54      var55      var56
## 63008.8724 35946.8002 79938.2627 47002.3867  8686.7542  6769.6085   325.5017
##      var57      var58      var59      var60      var61      var62      var63
##  3937.8319  2135.7717  2707.4575  1039.3790   163.8511   680.0737  1232.0642
##      var64
## 28006.0742
```

Examining sigma2, nu2, rho

```
sigma2_samples <- mcmc.list(chain1$samples$sigma2, chain2$samples$sigma2,
                            chain3$samples$sigma2)

nu2_samples <- mcmc.list(chain1$samples$nu2, chain2$samples$nu2,
                         chain3$samples$nu2)
```

```
effectiveSize(sigma2_samples)
```

```
##     var1
## 6808.768
```

```
effectiveSize(nu2_samples)
```

```
##     var1
## 9364.658
```

Examining a sample of the 3108 phi parameters

```
phi_samples <- mcmc.list(chain1$samples$phi, chain2$samples$phi, chain3$samples$phi)
```

```
set.seed(1157, kind = "Mersenne-Twister", normal.kind = "Inversion", sample.kind = "Rejection")

phi_subset_idx <- sample(1:ncol(phi_samples[[1]]), size = 10)
```

```
phi_samples_subset <- phi_samples[, phi_subset_idx]
```

```
effectiveSize(phi_samples_subset)
```

```
##      var1      var2      var3      var4      var5      var6      var7      var8
## 37202.660  4572.161  6806.516 20135.793 93307.470 39089.413 33554.743 40535.330
##      var9     var10
## 70212.189  4821.368
```

Inference

```
beta_samples_matrix <- rbind(chain1$samples$beta, chain2$samples$beta, chain3$samples$beta)

# TODO: change to var_names later
colnames(beta_samples_matrix) <- c("strat0",
"strat0:flood_risk_pc1",
"strat0:flood_risk_pc2",
"strat0:flood_risk_pc3",
"strat0:flood_risk_pc4",
"strat0:flood_risk_pc5",
"strat0:EP_UNEMP",
"strat0:EP_PCI",
"strat0:EP_NOHSDP",
"strat0:EP_AGE65",
"strat0:EP_AGE17",
"strat0:EP_DISABL",
"strat0:EP_SNGPNT",
"strat0:EP_MINRTY",
"strat0:EP_LIMENG",
"strat0:EP_MUNIT",
"strat0:EP_MOBILE",
"strat0:EP_CROWD",
"strat0:EP_NOVEH",
"strat0:EP_GROUPQ",
"strat0:EP_UNINSUR",
"strat0:co",
"strat0:no2",
"strat0:o3",
"strat0:pm10",
"strat0:pm25",
"strat0:so2",
"strat0:summer_tmmx",
"strat0:winter_tmmx",
"strat0:summer_rmax",
"strat0:winter_rmax",
"strat0:Data_Value_CSMOKING",
"strat1",
"strat1:flood_risk_pc1",
"strat1:flood_risk_pc2",
"strat1:flood_risk_pc3",
"strat1:flood_risk_pc4",
"strat1:flood_risk_pc5",
```

```
"strat1:EP_UNEMP",
"strat1:EP_PCI",
"strat1:EP_NOHSDP",
"strat1:EP_AGE65",
"strat1:EP_AGE17",
"strat1:EP_DISABL",
"strat1:EP_SNGPNT",
"strat1:EP_MINRTY",
"strat1:EP_LIMENG",
"strat1:EP_MUNIT",
"strat1:EP_MOBILE",
"strat1:EP_CROWD",
"strat1:EP_NOVEH",
"strat1:EP_GROUPQ",
"strat1:EP_UNINSUR",
"strat1:co",
"strat1:no2",
"strat1:o3",
"strat1:pm10",
"strat1:pm25",
"strat1:so2",
"strat1:summer_tmmx",
"strat1:winter_tmmx",
"strat1:summer_rmax",
"strat1:winter_rmax",
"strat1:Data_Value_CSMOKING") # var_names
```

```
(beta_inference <- round(t(apply(beta_samples_matrix, 2, quantile, c(0.5, 0.025, 0.975))),5))
```

```
##                               50%     2.5%    97.5%
## strat0                    6.43057  6.41684  6.44428
## strat0:flood_risk_pc1    -0.01479 -0.02603 -0.00362
## strat0:flood_risk_pc2     0.00961 -0.00341  0.02245
## strat0:flood_risk_pc3    -0.00201 -0.01169  0.00766
## strat0:flood_risk_pc4     0.00792 -0.00135  0.01719
## strat0:flood_risk_pc5     0.00190 -0.00692  0.01075
## strat0:EP_UNEMP           0.05020  0.03500  0.06527
## strat0:EP_PCI            -0.04941 -0.06333 -0.03545
## strat0:EP_NOHSDP          0.23345  0.20579  0.26097
## strat0:EP_AGE65           1.23105  1.21750  1.24458
## strat0:EP_AGE17           0.15970  0.14417  0.17523
## strat0:EP_DISABL          0.22445  0.20776  0.24128
## strat0:EP_SNGPNT          0.01504 -0.00063  0.03067
## strat0:EP_MINRTY         -0.17432 -0.19679 -0.15201
## strat0:EP_LIMENG         -0.02997 -0.05533 -0.00480
## strat0:EP_MUNIT          -0.05975 -0.07265 -0.04682
## strat0:EP_MOBILE          0.07762  0.06456  0.09065
## strat0:EP_CROWD           0.01336 -0.00717  0.03394
## strat0:EP_NOVEH           0.08992  0.06770  0.11209
## strat0:EP_GROUPQ         -0.09386 -0.10680 -0.08081
## strat0:EP_UNINSUR         0.13585  0.11884  0.15277
## strat0:co                -0.11942 -0.15848 -0.07998
## strat0:no2                0.01534 -0.03633  0.06636
## strat0:o3                -0.14595 -0.21856 -0.07619
```

```
## strat0:pm10                -0.19961 -0.23314 -0.16712
## strat0:pm25                 0.43798  0.39060  0.48763
## strat0:so2                  0.05621  0.02290  0.09022
## strat0:summer_tmmx          0.13280  0.08339  0.18615
## strat0:winter_tmmx         -0.31593 -0.47846 -0.18847
## strat0:summer_rmax          0.00254 -0.06569  0.07049
## strat0:winter_rmax          0.05320  0.00389  0.10322
## strat0:Data_Value_CSMOKING  0.70817  0.67899  0.73708
## strat1                      0.31696  0.29857  0.33528
## strat1:flood_risk_pc1      -0.01268 -0.02315 -0.00227
## strat1:flood_risk_pc2       0.00658 -0.00523  0.01846
## strat1:flood_risk_pc3       0.00569 -0.00365  0.01504
## strat1:flood_risk_pc4      -0.00259 -0.01103  0.00583
## strat1:flood_risk_pc5       0.00154 -0.00691  0.00993
## strat1:EP_UNEMP             0.05283  0.04390  0.06177
## strat1:EP_PCI              -0.08201 -0.10645 -0.05750
## strat1:EP_NOHSDP            0.14347  0.12630  0.16066
## strat1:EP_AGE65             1.63400  1.61986  1.64827
## strat1:EP_AGE17             0.30099  0.28704  0.31510
## strat1:EP_DISABL            0.22420  0.21232  0.23611
## strat1:EP_SNGPNT           -0.05794 -0.06914 -0.04666
## strat1:EP_MINRTY            0.01348 -0.00424  0.03122
## strat1:EP_LIMENG           -0.04214 -0.05721 -0.02698
## strat1:EP_MUNIT            -0.00647 -0.01763  0.00469
## strat1:EP_MOBILE            0.05158  0.04197  0.06119
## strat1:EP_CROWD            -0.02294 -0.03467 -0.01119
## strat1:EP_NOVEH             0.19940  0.18479  0.21408
## strat1:EP_GROUPQ           -0.05314 -0.06171 -0.04457
## strat1:EP_UNINSUR           0.08819  0.07621  0.10005
## strat1:co                  -0.14840 -0.19083 -0.10638
## strat1:no2                 -0.02954 -0.08115  0.02160
## strat1:o3                  -0.15649 -0.22953 -0.08649
## strat1:pm10                -0.14592 -0.18058 -0.11255
## strat1:pm25                 0.45173  0.40449  0.50151
## strat1:so2                  0.01675 -0.01555  0.04973
## strat1:summer_tmmx          0.04595 -0.00426  0.09936
## strat1:winter_tmmx         -0.16618 -0.32870 -0.03905
## strat1:summer_rmax         -0.07471 -0.14382 -0.00705
## strat1:winter_rmax          0.04073 -0.00820  0.09070
## strat1:Data_Value_CSMOKING  1.03745  1.01707  1.05807
```

List of significant beta coefficients:

```
colnames(beta_samples_matrix)[sign(beta_inference[, 2]) == sign(beta_inference[, 3])]
```

```
##  [1] "strat0"                    "strat0:flood_risk_pc1"
##  [3] "strat0:EP_UNEMP"           "strat0:EP_PCI"
##  [5] "strat0:EP_NOHSDP"          "strat0:EP_AGE65"
##  [7] "strat0:EP_AGE17"           "strat0:EP_DISABL"
##  [9] "strat0:EP_MINRTY"          "strat0:EP_LIMENG"
## [11] "strat0:EP_MUNIT"           "strat0:EP_MOBILE"
## [13] "strat0:EP_NOVEH"           "strat0:EP_GROUPQ"
## [15] "strat0:EP_UNINSUR"         "strat0:co"
## [17] "strat0:o3"                 "strat0:pm10"
## [19] "strat0:pm25"               "strat0:so2"
```

```
## [21] "strat0:summer_tmmx"         "strat0:winter_tmmx"
## [23] "strat0:winter_rmax"         "strat0:Data_Value_CSMOKING"
## [25] "strat1"                     "strat1:flood_risk_pc1"
## [27] "strat1:EP_UNEMP"            "strat1:EP_PCI"
## [29] "strat1:EP_NOHSDP"           "strat1:EP_AGE65"
## [31] "strat1:EP_AGE17"            "strat1:EP_DISABL"
## [33] "strat1:EP_SNGPNT"           "strat1:EP_LIMENG"
## [35] "strat1:EP_MOBILE"           "strat1:EP_CROWD"
## [37] "strat1:EP_NOVEH"            "strat1:EP_GROUPQ"
## [39] "strat1:EP_UNINSUR"          "strat1:co"
## [41] "strat1:o3"                  "strat1:pm10"
## [43] "strat1:pm25"                "strat1:winter_tmmx"
## [45] "strat1:summer_rmax"         "strat1:Data_Value_CSMOKING"
```

Credible Interval plots for the coefficients, in ggplot

```
# first, process the beta_inference matrix in a form ggplot can understand

beta_inference_df <- as.data.frame(beta_inference)

beta_inference_df <- mutate(beta_inference_df, var_name = row.names(beta_inference_df))

beta_inference_df <- rename(beta_inference_df,
                            post_median = `50%`,
                            post_2.5 = `2.5%`,
                            post_97.5 = `97.5%`)

beta_inference_df$var_name <- factor(beta_inference_df$var_name, levels = beta_inference_df$var_name)
```

Splitting up the beta coefficients for each strata

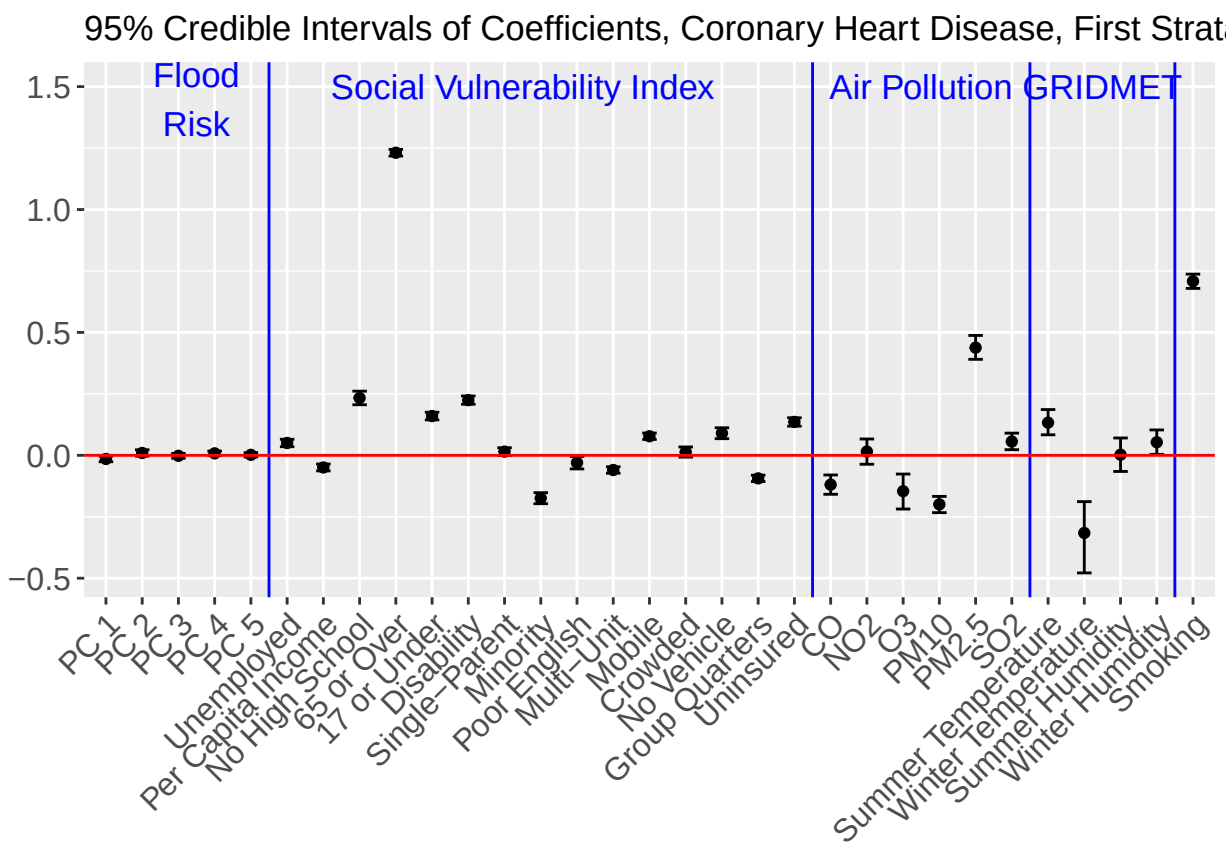
```
beta_inference_df_strat0 <- beta_inference_df[1:(nrow(beta_inference_df)/2),]

beta_inference_df_strat1 <- beta_inference_df[(nrow(beta_inference_df)/2 + 1):nrow(beta_inference_df),]
```

Note: The intercept for both strata (corresponding to poverty) is not included.

```
p <- ggplot(beta_inference_df_strat0[-1, ], aes(x = var_name, y = post_median)) +
  geom_point() +
  theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust=1), axis.title.x = element_blank(), axis
        axis.text=element_text(size=12),
        plot.margin = margin(5.5, 5.5, 5.5, 10)) +
  geom_errorbar(aes(ymin = post_2.5, ymax = post_97.5, width = 0.4)) +
  geom_vline(xintercept = c(5.5, 20.5, 26.5, 30.5), col = "blue") +
  geom_hline(yintercept = 0, col = "red") +
  annotate(geom = "text", x = 3.5, y = 1.45, label = "Flood\nRisk",
           col = "blue", size = 4.5) +
  annotate(geom = "text", x = 12.5, y = 1.5, label = "Social Vulnerability Index",
           col = "blue", size = 4.5) +
  annotate(geom = "text", x = 23.5, y = 1.5, label = "Air Pollution",
           col = "blue", size = 4.5) +
  annotate(geom = "text", x = 28.5, y = 1.5, label = "GRIDMET",
           col = "blue", size = 4.5) +
  scale_x_discrete(labels = c("PC 1", "PC 2", "PC 3", "PC 4", "PC 5",
```

```
                                "Unemployed", "Per Capita Income", "No High School",
                                "65 or Over", "17 or Under", "Disability",
                                "Single-Parent", "Minority", "Poor English",
                                "Multi-Unit", "Mobile", "Crowded",
                                "No Vehicle", "Group Quarters", "Uninsured",
                                "CO", "NO2", "O3", "PM10", "PM2.5", "SO2",
                                "Summer Temperature", "Winter Temperature", "Summer Humidity", "Winter Hu
                                "Smoking")) + ggtitle("95% Credible Intervals of Coefficients, Coronary H
```

```
p
```




```
p <- ggplot(beta_inference_df_strat1[-1, ], aes(x = var_name, y = post_median)) +
  geom_point() +
  theme(axis.text.x = element_text(angle = 45, vjust = 1, hjust=1), axis.title.x = element_blank(), axis
        axis.text=element_text(size=12),
        plot.margin = margin(5.5, 5.5, 5.5, 10)) +
  geom_errorbar(aes(ymin = post_2.5, ymax = post_97.5, width = 0.4)) +
  geom_vline(xintercept = c(5.5, 20.5, 26.5, 30.5), col = "blue") +
  geom_hline(yintercept = 0, col = "red") +
  annotate(geom = "text", x = 3.5, y = 1.45, label = "Flood\nRisk",
           col = "blue", size = 4.5) +
  annotate(geom = "text", x = 12.5, y = 1.5, label = "Social Vulnerability Index",
           col = "blue", size = 4.5) +
  annotate(geom = "text", x = 23.5, y = 1.5, label = "Air Pollution",
           col = "blue", size = 4.5) +
  annotate(geom = "text", x = 28.5, y = 1.5, label = "GRIDMET",
           col = "blue", size = 4.5) +
```

```
  scale_x_discrete(labels = c("PC 1", "PC 2", "PC 3", "PC 4", "PC 5",
                              "Unemployed", "Per Capita Income", "No High School",
                              "65 or Over", "17 or Under", "Disability",
                              "Single-Parent", "Minority", "Poor English",
                              "Multi-Unit", "Mobile", "Crowded",
                              "No Vehicle", "Group Quarters", "Uninsured",
                              "CO", "NO2", "O3", "PM10", "PM2.5", "SO2",
                              "Summer Temperature", "Winter Temperature", "Summer Humidity", "Winter Hu
                              "Smoking")) + ggtitle("95% Credible Intervals of Coefficients, Coronary H

p
```