Code Appendix

Thomas Hsiao

3/14/2019

```r
library(data.table)
library(ggplot2)
library(INLA)
library(RColorBrewer)
library(gridExtra)
library(spdep)
library(Hmisc)

#' Set working directory
home_dir <- "C:/Users/twh42/Documents/UW_Class/CSSS_554/final_project"
setwd(home_dir)
shp_path <- "data/NGA/shape_files/2013/shps/"

#' Set parameters to pull data
recall   <- 15
length   <- 3
age_bins <- 5

#' Read in data - only admin 2 subnational estimates
index_map <- fread(paste0(shp_path, "/index_map.csv"))
asfr_path <- sprintf("data/prepped/asfr_recall_%d_length_%d_age_%d.csv", recall, length, age_bins)
tfr_path  <- sprintf("data/prepped/tfr_recall_%d_length_%d_age_%d.csv", recall, length, age_bins)
asfr      <- fread(asfr_path)[STATE != "national" & SurveyId != "NG2015MIS"]#[SurveyId == sid & STATE !
tfr       <- fread(tfr_path)

#' Create time interval periods (1995-1999, 2000-2004, 2005-2009, 2010-2014) for temporal RW smoothing
asfr <- asfr[year >= 1995]
asfr[, time_period := Hmisc::cut2(year, seq(1995, 2013, 3))]
asfr <- asfr[order(time_period)]
asfr[, time_step := .GRP, time_period]
asfr[time_period == "[1995,1998)", time_period_short := "'95-'97"]
asfr[time_period == "[1998,2001)", time_period_short := "'98-'00"]
asfr[time_period == "[2001,2004)", time_period_short := "'01-'03"]
asfr[time_period == "[2004,2007)", time_period_short := "'04-'06"]
asfr[time_period == "[2007,2010)", time_period_short := "'07-'09"]
asfr[time_period == "[2010,2013]", time_period_short := "'10-'13"]

#' interpretable time steps for INLA
asfr[, time.struct := time_step]
asfr[, time.unstruct := time_step]

#' Match indices up with the indices in graph/shapefile - VERY IMPORTANT
asfr <- merge(asfr, index_map, by = "STATE")
asfr[, state.struct := graph_index]
asfr[, state.unstruct := graph_index]

#' Space-time interaction
```

```r
asfr[, time.space := paste0(time.struct,"_", state.struct)]

# 1. Fit Separate Models by age: 15-19, 20-24, 25-29, 30-34, and 35-39
# 2. INLA Poisson Lognormal Spatial ICAR model - Besag and IID
# 3. INLA Poisson Lognormal Spatial ICAR model - Besage and IID and RW2
# 4. INLA Poisson Lognormal Spatial ICAR model with

fit_age_specific_inla <- function(asfr, age){
  #' @description Fits INLA model for given age group
  df <- copy(asfr[age_start == age])

  #### PRIORS #######
  # IID parameters
  a.iid <- 0.5
  b.iid <- 0.001488
  # ICAR parameters
  a.icar <- 0.5
  b.icar <- 0.003602143
  # RW2 parameters
  a.rw <- 1
  b.rw <- 0.01
  ######## DIFFERENT MODEL FORMULAS ################
  #' Besag Spatial Model: Poisson-LogNormal-Spatial Model
  formula <- nbirths ~ 1 +
    f(state.struct, model = "besag", hyper = list(prec=list(prior="loggamma", param=c(a.icar, b.icar)))
    f(state.unstruct, model = "iid", hyper = list(prec=list(prior="loggamma",param=c(a.iid,b.iid),initia

  #' BYM2 Spatial Model
  formula <- nbirths ~ 1 +
    f(state.struct, model = "bym2", hyper = list(phi=list(prior="pc", param=c(0.5, 0.5), initial = 1),pr
      graph = paste0(shp_path, "nga.graph"),
      scale.model = T, constr = T)

  #' Besag Spatial and RW Temporal Model - Structured and unstructured effects:
  formula <- nbirths ~ 1 +
    f(state.struct, model = "besag", hyper = list(prec=list(prior="loggamma", param=c(a.icar, b.icar)))
    f(state.unstruct, model = "iid", hyper = list(prec=list(prior="loggamma",param=c(a.iid,b.iid),initia
    f(time.struct, model = "rw2", hyper = list(prec=list(prior="pc.prec",param=c(a.rw, b.rw)))) +
    f(time.unstruct, model = "iid", hyper = list(prec=list(prior="loggamma",param=c(a.iid,b.iid),initial
    f(time.space, model = "iid", hyper = list(prec=list(prior="loggamma",param=c(a.iid,b.iid),initial=1)

  ######### FITTING THE MODEL IN INLA#################
  #' Fit Spatial INLA model
  inla.fit <- inla(formula, family = "poisson",
                   control.compute=list(dic=T,mlik=T,cpo=T), # What fit statistics to output
                   control.predictor=list(compute=TRUE),     # Output marginal of linear predictor - ba
                   data = df,
                   E = py)
  df[, inlaest := inla.fit$summary.fitted.values$`0.5quant`]
  df[, inlalower := inla.fit$summary.fitted.values$`0.025quant`]
  df[, inlaupper := inla.fit$summary.fitted.values$`0.975quant`]
  list(fit = inla.fit, df = df)
}
```

```r
#' Save Fit and DF objects per age group
all_ages <- seq(15, 45, 5)
fits <- final_df <- fixed_effects <- vector(mode = "list", length = length(all_ages))
names(fits) <- names(final_df) <- names(fixed_effects) <- all_ages
for (i in 1:length(all_ages)){
  a <- all_ages[i]
  message(a)
  obj <- fit_age_specific_inla(asfr, a)
  fits[[i]] <- obj$fit
  final_df[[i]] <- obj$df
}

final_df <- rbindlist(final_df)
final_df[, c("inlaest_thous", "inlalower_thous", "inlaupper_thous") := lapply(.SD, function(x) round(10
         .SDcols = c("inlaest", "inlalower", "inlaupper")]

########### PROPORTION OF VARIANCE AND VARIANCE SUMMARIES FOR EACH COMPONENT ##########################
nareas <- 37
nperiods <- 6
index <- 1

create.prop.var.table <- function(index, fits, nareas = 37, nperiods = 6){
  message(index)
  all_ages <- seq(15, 45, 5)
  mat.marg.icar <- matrix(NA, nrow = nareas, ncol = 1000)
  mat.marg.rw <- matrix(NA, nrow = nperiods, ncol = 1000)
  m <- fits[[index]]$marginals.random$state.struct
  t <- fits[[index]]$marginals.random$time.struct
  for (i in 1:nareas) {
    Sre <- m[[i]]
    mat.marg.icar[i, ] <- inla.rmarginal(1000, Sre)
  }
  for (i in 1:nperiods){
    Tre <- t[[i]]
    mat.marg.rw[i, ] <- inla.rmarginal(1000, Tre)
  }
  var.state.struct <- apply(mat.marg.icar, 2, var)
  var.time.struct <- apply(mat.marg.rw, 2, var)
  var.state.unstruct <- inla.rmarginal(1000, inla.tmarginal(function(x) 1/x, fits[[index]]$marginals.hyp
  var.time.unstruct <- inla.rmarginal(1000, inla.tmarginal(function(x) 1/x, fits[[index]]$marginals.hype
  var.timespace <- inla.rmarginal(1000, inla.tmarginal(function(x) 1/x, fits[[index]]$marginals.hyper$"

  data.table(`Age Group` = all_ages[index],
             `Spatial Structure` = median(var.state.struct),
             `Spatial Unstructure` = median(var.state.unstruct),
             `Time Structure` = median(var.time.struct),
             `Time Unstructure` = median(var.time.unstruct),
             `Space-Time` = median(var.time.space))

  total.var <- var.state.struct + var.time.struct + var.state.unstruct + var.time.unstruct + var.timespa
  perc.var.state.struct <- mean(var.state.struct/total.var)
  perc.var.state.unstruct <- mean(var.state.unstruct/total.var)
  perc.var.time.struct <- mean(var.time.struct/total.var)
```

```r
  perc.var.time.unstruct <- mean(var.time.unstruct/total.var)
  perc.var.time.space <- mean(var.timespace/total.var)

  data.table(`Age Group` = all_ages[index],
             `Spatial Structure` = perc.var.state.struct,
             `Spatial Unstructure` = perc.var.state.unstruct,
             `Time Structure` = perc.var.time.struct,
             `Time Unstructure` = perc.var.time.unstruct,
             `Space-Time` = perc.var.time.space)
}

create.var.sum.table <- function(index, fits, nareas = 37, nperiods = 6){
  message(index)
  all_ages <- seq(15, 45, 5)
  mat.marg.icar <- matrix(NA, nrow = nareas, ncol = 1000)
  mat.marg.rw <- matrix(NA, nrow = nperiods, ncol = 1000)
  m <- fits[[index]]$marginals.random$state.struct
  t <- fits[[index]]$marginals.random$time.struct
  for (i in 1:nareas) {
    Sre <- m[[i]]
    mat.marg.icar[i, ] <- inla.rmarginal(1000, Sre)
  }
  for (i in 1:nperiods){
    Tre <- t[[i]]
    mat.marg.rw[i, ] <- inla.rmarginal(1000, Tre)
  }
  var.state.struct <- apply(mat.marg.icar, 2, var)
  var.time.struct <- apply(mat.marg.rw, 2, var)
  var.state.unstruct <- inla.rmarginal(1000, inla.tmarginal(function(x) 1/x, fits[[index]]$marginals.hyp
  var.time.unstruct <- inla.rmarginal(1000, inla.tmarginal(function(x) 1/x, fits[[index]]$marginals.hype
  var.timespace <- inla.rmarginal(1000, inla.tmarginal(function(x) 1/x, fits[[index]]$marginals.hyper$"

  data.table(`Age Group` = all_ages[index],
             `Spatial Structure` = paste0(round(median(var.state.struct), 3), " (",round(quantile(var.st
             `Spatial Unstructure` = paste0(round(median(var.state.unstruct), 3), " (",round(quantile(va
             `Time Structure` = paste0(round(median(var.time.struct), 3), " (",round(quantile(var.time.
             `Time Unstructure` = paste0(round(median(var.time.unstruct), 3), " (",round(quantile(var.ti
             `Space-Time` = paste0(round(median(var.timespace), 3), " (",round(quantile(var.timespace,
}

# Just create table for 15-44 - Proportion of Total Variance
prop.var.table <- rbindlist(lapply(1:6, create.prop.var.table, fits = fits))
prop.var.table %>% View
prop.var.table[, `Age Group` := paste0(`Age Group`, "-", `Age Group` + 4)]
write_csv(prop.var.table, paste0(plot_path, "/prop_var_table.csv"))

# Just create table for 15-44 - Variance Summaries
var.sum.table <- rbindlist(lapply(1:6, create.var.sum.table, fits = fits))
var.sum.table[, `Age Group` := paste0(`Age Group`, "-", `Age Group` + 4)]
write_csv(var.sum.table, paste0(plot_path, "/var_sum_table.csv"))

############# FIGURES AND TABLES FOR REPORT ###################################
plot_path <- "report/plots/"
```

```
#' Subnational State Specific Time Series plots for each age group - Data and smoothed estimate
pdf(paste0(plot_path, "inlaest_and_data_loc_specific_yfixed.pdf"), width = 12, height = 8)
for(state in unique(final_df$STATE)){
  message(state)
  gg <- ggplot(final_df[STATE == state & age_start != 45], aes(as.factor(time_period), (asfr))) +
    geom_pointrange(aes(color = SurveyId, ymin=lower, ymax=upper), position = position_jitter(), alpha =
    #geom_errorbar(aes(ymin = lower, ymax = upper, color = SurveyId), width = 0, position = position_ji
    #geom_line(aes(color = SurveyId, group = SurveyId), linetype = "dashed") +
    geom_point(aes(as.factor(time_period), (inlaest), group = age_start), color = "black", size=2) +
    geom_errorbar(aes(ymin=inlalower, ymax = inlaupper, group = age_start), color = "black", width = 0,
    geom_line(aes(as.factor(time_period), inlaest, group = age_start), linetype = "dashed") +
    #geom_ribbon(aes(ymin = (inlalower), ymax = (inlaupper), group = age_start), fill = "blue", alpha =
    theme_classic() +
    facet_wrap(~age_start) +
    scale_x_discrete(labels = unique(asfr$time_period_short)) +
    xlab("Time") +
    ylab("ASFR") +
    ggtitle(toupper(state)) +
    coord_cartesian(ylim = c(0, 0.3)) +
    labs(x=NULL, y=NULL)
  print(gg)
}
dev.off()

# Maps of each age groups estimate
map    <- rgdal::readOGR(dsn = paste0(shp_path, "sdr_subnational_boundaries2.shp"))
map_df <- data.table(broom::tidy(map, region = "REGNAME"))
map_df <- merge(map_df, final_df[,.(STATE, age_start, time_period, time_period_short, inlaest_thous)],
                by.x = "id", by.y = "STATE", allow.cartesian = T)

for(a in unique(final_df$age_start)){

  message(a)
  pdf(paste0(plot_path, "inlaest_maps_asfr_",a,".pdf"), width = 11.5, height = 8)
  gg <- ggplot() +
    geom_polygon(data = map_df[age_start == a], aes(x = long, y = lat, group = group, fill = inlaest_th
    #theme_void() +
    ggtitle(paste0("Nigeria Age Group ", a, "-", a+4)) +
    coord_quickmap() +
    facet_wrap(~time_period) +
    #scale_fill_gradient(low = "blue", high = "yellow") +
    scale_fill_distiller(type = "div", palette = "RdYlBu", name = "ASFR (per 1000)") +
    theme_classic()
  print(gg)
  dev.off()
}


########## TFR MAP #############
tfr <- final_df[, .(asfr = mean(inlaest)), by = .(STATE, time_period, age_start)]
tfr <- tfr[, .(tfr = 5 * sum(asfr), num_age_groups = .N), by = .(time_period, STATE)]
#tfr <- tfr[num_age_groups >= 6]
```

```r
map    <- rgdal::readOGR(dsn = paste0(shp_path, "sdr_subnational_boundaries2.shp"))
map_df <- data.table(broom::tidy(map, region = "REGNAME"))
map_df <- merge(map_df, tfr[,.(STATE, time_period, tfr)],
                by.x = "id", by.y = "STATE", allow.cartesian = T)

pdf(paste0(plot_path, "inlaest_maps_tfr.pdf"), width = 11.5, height = 8)
  gg <- ggplot() +
    geom_polygon(data = map_df, aes(x = long, y = lat, group = group, fill = tfr), color = 'black', siz
    #theme_void() +
    ggtitle(paste0("Nigeria TFR")) +
    coord_quickmap() +
    facet_wrap(~time_period) +
    #scale_fill_gradient(low = "blue", high = "yellow") +
    scale_fill_distiller(type = "div", palette = "RdYlBu", name = "TFR") +
    theme_classic()
  print(gg)
dev.off()

############ SCATTER OF SHRINKAGE #############
ggplot(final_df[age_start == 15], aes(asfr, inlaest)) +
  geom_point(alpha=0.5) +
  facet_wrap(~age_start) +
  geom_abline(color = "red") +
  coord_cartesian(xlim = c(0, 0.5), ylim = c(0, 0.5))
```