Plots of China intervention correlation analyses

Kylie Ainslie

06 April, 2020

dat\_all <- readr::read\_csv("dat\_all.csv")

## Parsed with column specification:  
## cols(  
## province = col\_character(),  
## r\_mean = col\_double(),  
## r\_q2.5 = col\_double(),  
## r\_q97.5 = col\_double(),  
## r\_median = col\_double(),  
## date\_lag = col\_date(format = ""),  
## movement = col\_double(),  
## movement\_hubei = col\_double(),  
## rolling.corr.biweekly = col\_double(),  
## rolling.corr.hubei = col\_double()  
## )

dat\_lag\_sa <- readr::read\_csv("dat\_all\_lag\_sens\_analysis.csv")

## Parsed with column specification:  
## cols(  
## province = col\_character(),  
## r\_mean = col\_double(),  
## r\_q2.5 = col\_double(),  
## r\_q97.5 = col\_double(),  
## r\_median = col\_double(),  
## lag = col\_double(),  
## date\_lag = col\_date(format = ""),  
## movement = col\_double(),  
## rolling.corr.biweekly = col\_double()  
## )

dat\_movement <- readRDS("movement\_province\_level\_subset.rds")  
  
dat\_imported <- readr::read\_csv("china\_new\_case\_data\_imported.csv") %>%   
 mutate(date\_confirm = as.Date(date\_confirm)) %>%  
 filter(province %in% c("hubei", "guangdong", "henan", "zhejiang",  
 "hunan", "beijing", "hong\_kong\_sar"))

## Parsed with column specification:  
## cols(  
## date\_confirm = col\_datetime(format = ""),  
## province = col\_character(),  
## new\_cases = col\_double(),  
## imported\_cases = col\_double(),  
## new\_cases\_wo\_imported = col\_double()  
## )

# colours  
col\_cases <- brewer.pal(8, "Reds")[6]   
col\_2019 <- brewer.pal(8, "Blues")[5]   
col\_2020 <- brewer.pal(8, "Blues")[8]   
col\_Rt <- brewer.pal(8, "Greens")[5]   
col\_inter\_corr <- brewer.pal(8, "RdPu")[5]   
col\_intra\_corr <- brewer.pal(8, "RdPu")[8]  
  
col\_text <- "black"   
col\_box <- "#2c3e50" # colour found from running 'calc\_element("axis.text.x", theme\_tq())'  
  
# colours used by KA in earlier versions of plots  
# cols2 <- c("movement" = "#00BFC4", "r\_mean" = "#00BA38", "rolling.corr.biweekly" = "darkblue", "rolling.corr.hubei" = "#FF61CC")  
  
  
# theme for the plots, controlling white space  
 plot\_theme <- theme(legend.position = "none",  
 panel.grid.major = element\_blank(),  
 panel.grid.minor = element\_blank(),  
 panel.background = element\_blank(),  
 axis.text.x = element\_text(angle = 45, hjust = 1),  
 panel.spacing.x = unit(0.2, "lines"),  
 panel.spacing.y = unit(0.2, "lines"),  
 plot.margin = unit(c(0, 0, 0, 0), "cm"),  
 legend.title = element\_blank(),  
 panel.border = element\_rect(fill = NA,   
 size = rel(1/2),   
 color = col\_text))  
  
# theme for facet grid, control label text and box  
facet\_theme <- theme(  
 strip.text.x = element\_text(  
 size = 12, color = col\_text  
 ),  
 strip.text.y = element\_text(  
 size = 12, color = col\_text  
 ),  
 strip.background = element\_blank()  
)  
  
  
  
# facet labels for each province and metric  
  
province\_labels <- c("hubei" = "Hubei",   
 "guangdong" = "Guangdong",   
 "henan" = "Henan",   
 "zhejiang" = "Zhejiang",   
 "hunan" = "Hunan",   
 "beijing" = "Beijing",  
 "cases" = "cases",  
 "movement" = "movement",  
 "hong\_kong\_sar" = "Hong Kong SAR")

# shift the 2019 data to line up on LNY with 2020 data - this is done by '-13'  
dat\_2020 <- dat\_movement %>% filter(year == "2020")  
  
join\_dat <- dat\_movement %>%   
 filter(year == "2020") %>%   
 select(c(date, month\_day, province))  
  
dat\_2019 <- dat\_movement %>%   
 filter(year == "2019") %>%   
 mutate(date = date - 13) %>%  
 filter(date >= "2019-01-01") %>%  
 mutate(month\_day = format(date, "%m-%d")) %>%   
 select(-c(date)) %>%  
 left\_join(join\_dat, by = c("month\_day", "province"))   
  
dat\_movement <- bind\_rows(dat\_2019, dat\_2020)

# Plot of cases and movement for Top 5 and Beijing

# join movement with case data  
dat\_case\_move <- left\_join(dat\_imported, dat\_movement, by = c("date\_confirm" = "date", "province")) %>%  
 select(date\_confirm, province, new\_cases\_wo\_imported, movement, year) %>%  
 rename("cases" = "new\_cases\_wo\_imported") %>%  
 pivot\_wider(names\_from = "year", values\_from = "movement") %>%  
 rename(movement\_2019 = "2019", movement\_2020 = "2020") %>%  
 pivot\_longer(cols = c(cases, movement\_2019, movement\_2020), names\_to = "metric") %>%  
 mutate(metric2 = metric) %>%  
 mutate(metric2 = recode(metric2, movement\_2019 = "movement", movement\_2020 = "movement"))  
  
  
no\_hubei <- dat\_case\_move %>% filter(province %in% c("guangdong", "henan", "zhejiang",  
 "hunan", "beijing"))  
hubei <- dat\_case\_move %>% filter(province == 'hubei')  
  
p1 <- ggplot(data = no\_hubei,aes(x = date\_confirm, y = value, color = metric)) +  
 geom\_line() +  
 xlab("Date") + ylab("") +  
 scale\_color\_manual(values = c(col\_cases, col\_2019, col\_2020),   
 breaks=c("cases","movement\_2019", "movement\_2020"),  
 labels = c("Number of New Cases","Movement 2019", "Movement 2020")) +  
 scale\_x\_date(labels = date\_format("%Y-%m-%d")) +  
 scale\_y\_continuous(limits = c(0, NA)) +  
 guides(color=guide\_legend(nrow=2,byrow=TRUE)) +  
 plot\_theme +  
 theme(axis.title.x=element\_blank()) +  
 facet\_grid(metric2~province, scales = "free\_y", labeller = as\_labeller(province\_labels)) +  
 facet\_theme  
  
legend\_b <- get\_legend(  
 p1 +  
 guides(color = guide\_legend(nrow = 1)) +  
 theme(legend.position = "bottom",  
 legend.key=element\_blank())  
)

## Warning: Removed 36 rows containing missing values (geom\_path).

p2 <- ggplot(data = hubei,aes(x = date\_confirm, y = value, color = metric)) +  
 geom\_line() +  
 # xlab("") +   
 ylab("") +  
 scale\_color\_manual(values = c(col\_cases, col\_2019, col\_2020),   
 breaks=c("cases","movement\_2019", "movement\_2020"),  
 labels = c("Number of New Cases","Movement 2019", "Movement 2020")) +  
 scale\_x\_date(labels = date\_format("%Y-%m-%d")) +  
 scale\_y\_continuous(limits = c(0, NA)) +  
 guides(color=guide\_legend(nrow=2,byrow=TRUE)) +  
 plot\_theme +  
 theme(axis.title.x=element\_blank()) +  
 facet\_grid(metric2~province, scales = "free\_y", labeller = as\_labeller(province\_labels)) +  
 facet\_theme +  
 theme(strip.text.y = element\_blank(),  
 legend.key=element\_blank())  
  
p3 <- plot\_grid(p2, p1, labels = "", nrow = 1, rel\_widths = c(.27, 1))

## Warning: Removed 36 rows containing missing values (geom\_path).  
  
## Warning: Removed 36 rows containing missing values (geom\_path).

p\_fig1 <- plot\_grid(p3,   
 textGrob("Date", gp=gpar(fontsize= 12, col=col\_text, face = "plain")),  
 legend\_b,   
 ncol = 1,   
 rel\_heights = c(1, .05, .1))  
  
  
ggsave("figure1.png", p\_fig1,  
 width = 8, height = 5,  
 dpi = 500)

dat\_all\_long <- dat\_all %>%   
 select(date\_lag, province, r\_mean, movement,  
 rolling.corr.biweekly, rolling.corr.hubei,   
 r\_q2.5, r\_q97.5, r\_median) %>%  
 mutate(date\_lag = as.Date(date\_lag)) %>%  
 gather(metric, value, r\_mean:r\_median) %>%   
 filter(metric %in% c("r\_mean","movement", "rolling.corr.biweekly",   
 "rolling.corr.hubei"))  
  
dat\_lag\_sa\_long <- dat\_lag\_sa %>%  
 select(date\_lag, province, r\_mean, movement,  
 rolling.corr.biweekly, r\_q2.5, r\_q97.5, r\_median) %>%  
 mutate(date\_lag = as.Date(date\_lag)) %>%  
 gather(metric, value, r\_mean:r\_median) %>%   
 filter(metric %in% c("r\_mean","movement", "rolling.corr.biweekly"))

# Plot Top 5 provinces with most confirmed cases and Beijing

# top 6 countries: hubei, guangdong, henan, zhejiang, hunan, beijing  
dat\_all\_no\_hk <- dat\_all %>% filter(province != "hong\_kong\_sar")  
dat\_all\_long\_no\_hk <- dat\_all\_long %>% filter(province != "hong\_kong\_sar")  
  
ymax = 6.6 # maximum y value for plotting  
# truncating the x-axis so that the Fig 2 plot looks nice  
min\_date <- as.Date("2020-01-23") # "2020-01-20"   
max\_date <- as.Date("2020-04-02") # "2020-03-24"  
  
dat\_all\_no\_hk\_trunc <- dat\_all\_no\_hk %>%   
 filter(date\_lag >= min\_date) %>%   
 filter(date\_lag <= max\_date) %>%  
 mutate(r\_q2.5 = ifelse(r\_q2.5 > ymax, ymax, r\_q2.5),  
 r\_q97.5 = ifelse(r\_q97.5 > ymax, ymax, r\_q97.5))  
  
dat\_all\_long\_no\_hk\_trunc <- dat\_all\_long\_no\_hk %>%   
 filter(date\_lag >= min\_date) %>%   
 filter(date\_lag <= max\_date)  
  
  
cols2 <- c("movement" = col\_2020, "r\_mean" = col\_Rt, "rolling.corr.biweekly" = col\_intra\_corr, "rolling.corr.hubei" = col\_inter\_corr)  
  
# plot with: movement, Rt, intra-corr  
p\_fig2 <- ggplot(data = dat\_all\_no\_hk\_trunc, aes(x = date\_lag, y = r\_mean)) +  
 geom\_line(data = filter(dat\_all\_long\_no\_hk\_trunc, !(metric %in% "rolling.corr.hubei")),  
 aes(x = date\_lag, y = value, color = metric)) +   
 geom\_ribbon(aes(ymin=r\_q2.5,ymax=r\_q97.5),   
 fill= col\_Rt ,alpha=0.2)+   
 xlab("Date") + ylab("") +   
 scale\_color\_manual(values = cols2, name = "",   
 breaks=c("movement", "r\_mean", "rolling.corr.biweekly"),  
 labels = c("Movement Index", "Reproduction Number",   
 "Correlation Local")) +  
 scale\_x\_date(labels = date\_format("%Y-%m-%d")) +  
 scale\_y\_continuous(limits = c(-1, ymax)) +  
 geom\_hline(yintercept = 1, linetype="dashed", colour = "black") +  
 geom\_hline(yintercept = -1, linetype="dashed", colour = "black") +  
 geom\_hline(yintercept = 0, colour = "black") +  
 guides(color=guide\_legend(nrow=1,byrow=TRUE)) +  
 plot\_theme +  
 facet\_wrap(~ province, labeller = as\_labeller(province\_labels)) +  
 facet\_theme  
  
legend\_fig2 <- get\_legend(  
 p\_fig2 +  
 guides(color = guide\_legend(nrow = 1)) +  
 theme(legend.position = "bottom",  
 legend.key=element\_blank())  
)

## Warning: Removed 11 rows containing missing values (geom\_path).

p\_fig2 <- plot\_grid(p\_fig2,  
 legend\_fig2,   
 ncol = 1,   
 rel\_heights = c(1, .1))

## Warning: Removed 11 rows containing missing values (geom\_path).

ggsave("figure2.png", p\_fig2,  
 width = 8, height = 7,  
 dpi = 500)

# plot with: movement, Rt, inter-corr, intra-corr  
p\_fig2\_SI <- ggplot(data = dat\_all\_no\_hk, aes(x = date\_lag, y = r\_mean)) +  
 geom\_line(data = dat\_all\_long\_no\_hk,  
 aes(x = date\_lag, y = value, color = metric)) +   
 geom\_ribbon(aes(ymin=r\_q2.5,ymax=r\_q97.5),   
 fill= col\_Rt ,alpha=0.2)+   
 xlab("Date") + ylab("") +   
 scale\_color\_manual(values = cols2, name = "",   
 breaks=c("movement", "r\_mean", "rolling.corr.biweekly",  
 "rolling.corr.hubei"),  
 labels = c("Movement Index", "Reproduction Number",   
 "Correlation Local", "Correlation Hubei")) +  
 scale\_x\_date(labels = date\_format("%Y-%m-%d")) +  
 scale\_y\_continuous(limits = c(-1, 18)) +  
 geom\_hline(yintercept = 1, linetype="dashed", colour = "black") +  
 geom\_hline(yintercept = -1, linetype="dashed", colour = "black") +  
 geom\_hline(yintercept = 0, colour = "black") +  
 #scale\_color\_tq() +  
 guides(color=guide\_legend(nrow=1,byrow=TRUE)) +  
 plot\_theme +  
 facet\_wrap(~ province, labeller = as\_labeller(province\_labels)) +  
 facet\_theme  
  
legend\_fig2\_SI <- get\_legend(  
 p\_fig2\_SI +  
 guides(color = guide\_legend(nrow = 1)) +  
 theme(legend.position = "bottom",  
 legend.key=element\_blank())  
)

## Warning: Removed 54 rows containing missing values (geom\_path).

p\_fig2\_SI <- plot\_grid(p\_fig2\_SI,  
 legend\_fig2\_SI,   
 ncol = 1,   
 rel\_heights = c(1, .1))

## Warning: Removed 54 rows containing missing values (geom\_path).

ggsave("figure2\_SI.png", p\_fig2\_SI,  
 width = 8, height = 7,  
 dpi = 500)

hk\_dat <- dat\_all %>%   
 filter(province == "hong\_kong\_sar")  
hk\_dat\_long <- dat\_all\_long %>%   
 filter(province == "hong\_kong\_sar",  
 metric %in% c("r\_mean","movement","rolling.corr.biweekly"))  
hk\_cases <- dat\_case\_move %>% filter(province == 'hong\_kong\_sar')  
  
  
p\_fig3a <- ggplot(data = hk\_cases, aes(x = date\_confirm, y = value, color = metric)) +  
 geom\_line() +  
 xlab("Date") + ylab("") +  
 scale\_color\_manual(values = c(col\_cases, col\_2019, col\_2020), name = "",   
 breaks=c("cases","movement\_2019", "movement\_2020"),  
 labels = c("Number of New Cases","Movement 2019", "Movement 2020")) +  
 scale\_x\_date(labels = date\_format("%Y-%m-%d")) +  
 guides(color=guide\_legend(nrow=2,byrow=TRUE)) +  
 plot\_theme +  
 facet\_grid(metric2~province, scales = "free\_y", labeller = as\_labeller(province\_labels)) +  
 facet\_theme   
  
  
  
p\_fig3b <- ggplot(data = hk\_dat, aes(x = date\_lag, y = r\_mean)) +  
 geom\_line(data = hk\_dat\_long,  
 aes(x = date\_lag, y = value, color = metric)) +   
 geom\_ribbon(aes(ymin=r\_q2.5,ymax=r\_q97.5),   
 fill= col\_Rt ,alpha=0.2)+   
 xlab("Date") + ylab("") + #ggtitle("Hong Kong") +  
 scale\_color\_manual(values = cols2, name = "",   
 breaks=c("movement", "r\_mean", "rolling.corr.biweekly",  
 "rolling.corr.hubei"),  
 labels = c("Movement Index", "Reproduction Number",   
 "Intra-region Correlation", "Inter-region Correlation")) +  
 scale\_x\_date(labels = date\_format("%Y-%m-%d")) +  
 geom\_hline(yintercept = 1, linetype="dashed", colour = "black") +  
 geom\_hline(yintercept = -1, linetype="dashed", colour = "black") +  
 geom\_hline(yintercept = 0, colour = "black") +  
 plot\_theme  
  
  
  
# create a nonsense plot to get the correct legend for Fig3  
p\_leg <- ggplot(data = hk\_dat, aes(x = date\_lag, y = r\_mean)) +  
 geom\_line(data = hk\_dat\_long,  
 aes(x = date\_lag, y = value, color = metric)) +   
 geom\_line(data = hk\_cases %>% filter(!(metric %in% "movement\_2020")),   
 aes(x = date\_confirm, y = value, color = metric)) +  
 geom\_ribbon(aes(ymin=r\_q2.5,ymax=r\_q97.5),   
 fill= col\_Rt ,alpha=0.2) +  
 scale\_color\_manual(values = c("cases" = col\_cases, "movement\_2019" = col\_2019, cols2), name = "",   
 breaks=c("cases","movement\_2019", "movement", "r\_mean", "rolling.corr.biweekly"),  
 labels = c("Number of New Cases", "Movement 2019",   
 "Movement 2020", "Reproduction Number",   
 "Correlation Local")) +  
 scale\_x\_date(labels = date\_format("%Y-%m-%d")) +  
 theme\_tq() +  
 plot\_theme  
  
  
  
  
leg <- get\_legend(  
 p\_leg +  
 guides(color = guide\_legend(nrow = 2)) +  
 theme(legend.position = "bottom")  
)

## Warning: Removed 34 rows containing missing values (geom\_path).

## Warning: Removed 25 rows containing missing values (geom\_path).

join\_plots <- plot\_grid(p\_fig3a,   
 p\_fig3b,  
 labels = "AUTO",  
 ncol = 2,   
 rel\_widths = c(.5, 1))

## Warning: Removed 34 rows containing missing values (geom\_path).

## Warning: Removed 34 rows containing missing values (geom\_path).

p\_fig3 <- plot\_grid(join\_plots,  
 leg,  
 nrow = 2,   
 rel\_heights = c(1, 0.15))  
  
ggsave("figure3.png", p\_fig3,  
 width = 8, height = 6,  
 dpi = 500)

# Lag sensitivity analysis plots

# top 6 countries: hubei, guangdong, henan, zhejiang, hunan, beijing  
dat\_lag\_no\_hk <- dat\_lag\_sa %>% filter(province != "hong\_kong\_sar")  
dat\_lag\_long\_no\_hk <- dat\_lag\_sa\_long %>% filter(province != "hong\_kong\_sar")  
  
cols2a <- c("movement" = col\_2020, "r\_mean" = col\_Rt, "rolling.corr.biweekly" = col\_intra\_corr)  
  
# plot correlation using province-specific lags  
p\_fig2\_SI2 <- ggplot(data = dat\_lag\_no\_hk, aes(x = date\_lag, y = r\_mean)) +  
 geom\_line(data = dat\_lag\_long\_no\_hk,aes(x = date\_lag, y = value, color = metric)) +   
 geom\_ribbon(aes(ymin=r\_q2.5,ymax=r\_q97.5), fill = cols2a[2] , alpha=0.2)+   
 xlab("Date") + ylab("") +   
 scale\_color\_manual(values = cols2a, name = "",   
 breaks=c("movement", "r\_mean", "rolling.corr.biweekly"),  
 labels = c("Movement Index", "Reproduction Number", "Intra-region Correlation")) +  
 scale\_x\_date(labels = date\_format("%Y-%m-%d")) +  
 scale\_y\_continuous(limits = c(-1, 10)) +  
 geom\_hline(yintercept = 1, linetype="dashed", colour = "black") +  
 geom\_hline(yintercept = -1, linetype="dashed", colour = "black") +  
 geom\_hline(yintercept = 0, colour = "black") +  
 guides(color=guide\_legend(nrow=2,byrow=TRUE)) +  
 #plot\_theme +  
 facet\_wrap(~ province) #+  
 #facet\_theme  
  
# ggsave("figure2\_SI2.png", p\_fig2\_SI2,  
# width = 8, height = 7,  
# dpi = 500)

# movement plots

# p\_mov <- ggplot(data = dat\_movement %>% filter(province %in% c("hubei", "guangdong", "henan", "zhejiang","hunan", "beijing")),  
# aes(x = month\_day, y = movement, group = year,  
# col = as.factor(year))) +  
# geom\_line() +  
# xlab("date") +  
# ylab("movement index") +  
# facet\_wrap(vars(province)) +  
# theme\_tq(base\_size = 12) +   
# theme(panel.grid.major = element\_blank(),  
# panel.grid.minor = element\_blank(),  
# panel.background = element\_blank(),  
# axis.line = element\_line(colour = "black"),  
# axis.text.x = element\_text(angle = 45, hjust = 1)) +  
# labs(col = "year") +  
# scale\_x\_discrete( breaks = c("01-01", "01-16", "02-01", "02-16", "03-01", "03-16")) +  
# scale\_color\_manual(values = c(col\_2019, col\_2020)) +  
# theme(legend.position="bottom") +  
# facet\_wrap(~ province)  
#   
# # LNY 2020 was at "01-23"  
#   
# ggsave("movement.png", p\_mov,  
# width = 10, height = 7,  
# dpi = 300)