panel

Joshua Ashkinaze

2023-03-28

Load packages

Load Data

```
##############################################
# LOAD DATA
##############################################
# read in data
df <- read_csv("https://raw.githubusercontent.com/josh-ashkinaze/attention/main/data/trend_merged_data_
```

```
## New names:
## Rows: 16314 Columns: 24
## -- Column specification
## -------------------------------------------------------- Delimiter: "," chr
## (6): search_type, event, kw, index, kwe, period dbl (13): ...1, value,
## rumor_delta, announce_delta, rumor_announce_gap, stu... date (5): date,
## rumor_day, announce_day, max_date, min_date
## i Use `spec()` to retrieve the full column specification for this data. i
## Specify the column types or set `show_col_types = FALSE` to quiet this message.
## * `` -> `...1`
```

```
df$year <- year(df$date)
df$month <- month(df$date)
df$week <- week(df$date)

# kwid is unique kw-id: (event, keyword, search_type)
df$kwid <- paste(paste(df$kw, df$event, "_"), df$search_type, "_")

# kwe is (keyword, event)
df$kwe <- paste(paste(df$kw, df$event, "_"))

df$search_type <- factor(df$search_type)
df$search_type <- relevel(df$search_type, ref = "web")
df$period <- factor(df$period)
df$period <- relevel(df$period, "control")
df$kw <- as.factor(df$kw)
df$event <- as.factor(df$event)
df$log_val <- log(df$value+1)
```

Modeling

```
##############################################
# RANDOM FX VERSION
##############################################
# Make mixed model
model <- lmer(value ~ start_delta + year + month + period*search_type + (1 | event/kw), data = df)

# Look at contrasts:
# For rumors, is attention higher for google news and YT vs web?
# For announcements, is attention higher for web vs google news and YT?
em <- emmeans(model, ~ period*search_type)
```

```
## Note: D.f. calculations have been disabled because the number of observations exceeds 3000.
## To enable adjustments, add the argument 'pbkrtest.limit = 16314' (or larger)
## [or, globally, 'set emm_options(pbkrtest.limit = 16314)' or larger];
## but be warned that this may result in large computation time and memory use.
```

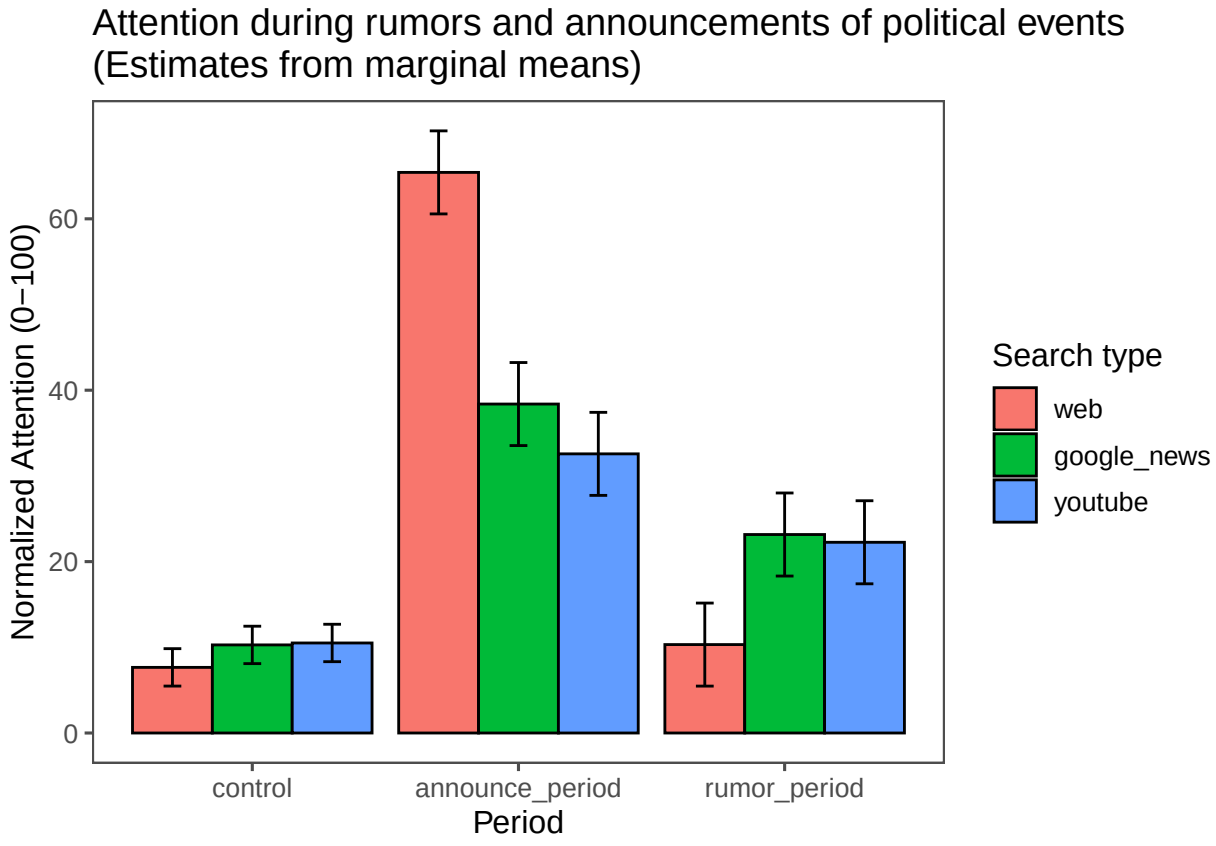
```
## Note: D.f. calculations have been disabled because the number of observations exceeds 3000.
## To enable adjustments, add the argument 'lmerTest.limit = 16314' (or larger)
## [or, globally, 'set emm_options(lmerTest.limit = 16314)' or larger];
## but be warned that this may result in large computation time and memory use.
```

```
em_df <- as.data.frame(em)
pairs <- pairs(em, by = "period", type = "response", rev = TRUE)
print(pairs)
```

```
## period = control:
##  contrast              estimate    SE  df z.ratio p.value
##  google_news - web        2.620 0.345 Inf   7.603  <.0001
##  youtube - web            2.846 0.345 Inf   8.258  <.0001
##  youtube - google_news    0.226 0.345 Inf   0.656  0.7891
##
## period = announce_period:
##  contrast              estimate    SE  df z.ratio p.value
##  google_news - web      -27.031 3.139 Inf  -8.610  <.0001
##  youtube - web          -32.844 3.139 Inf -10.462  <.0001
##  youtube - google_news   -5.812 3.139 Inf  -1.851  0.1531
##
## period = rumor_period:
##  contrast              estimate    SE  df z.ratio p.value
##  google_news - web       12.844 3.139 Inf   4.091  0.0001
##  youtube - web           11.938 3.139 Inf   3.802  0.0004
##  youtube - google_news   -0.906 3.139 Inf  -0.289  0.9551
##
## Degrees-of-freedom method: asymptotic
## P value adjustment: tukey method for comparing a family of 3 estimates
```

```
# Let's graph the Search Type X Period emmeans
em_df$lower <- em_df$asymp.LCL
em_df$upper <- em_df$asymp.UCL
```

```
ggplot(data=data.frame(em_df), aes(x=period, y=emmean, fill=search_type, ymin=lower, ymax=upper)) +
  geom_bar(stat="identity", position=position_dodge(width=0.9), color="black") +
  geom_errorbar(position=position_dodge(width=0.9), width=0.2) +
  labs(x="Period", y="Normalized Attention (0-100)", fill="Search type") + theme_few() + ggtitle("Atten
```




```
################################################
# PANEL MODEL VERSION
################################################
# Fit the fixed effects model and then get newey west standard errors
fem <- plm(value ~ period * search_type, data = df, model = "within", index = c("kwe", "date", "search_
```

```
## Warning in pdata.frame(data, index): duplicate couples (id-time) in resulting pdata.frame
##  to find out which, use, e.g., table(index(your_pdataframe), useNA = "ifany")
```

```
fixed_ses <- summary(fem, vcov = vcovNW)
fem_robust_se <- fixed_ses$coefficients[, 2]
fem_p_values <- fixed_ses$coefficients[, 4]


################################################
# DISLAY MODELS
################################################
stargazer(fem, model, type='text', se=list(fem_robust_se, NULL), p=list(fem_p_values, NULL))
```

```
##
```

```
## ===================================================================================
##                                                        Dependent variable:
##                                                 ---------------------------------------
##                                                               value
##                                                      panel               linear
##                                                      linear          mixed-effects
##                                                       (1)                 (2)
## -----------------------------------------------------------------------------------
## start_delta                                                            0.058***
##                                                                        (0.006)
##
## year                                                                   1.949**
##                                                                        (0.981)
##
## month                                                                  -0.058
##                                                                        (0.099)
##
## periodannounce_period                              58.520***           57.758***
##                                                     (4.610)             (2.234)
##
## periodrumor_period                                   1.895               2.659
##                                                     (2.561)             (2.234)
##
## search_typegoogle_news                              2.620***            2.620***
##                                                     (0.413)             (0.345)
##
## search_typeyoutube                                  2.846***            2.846***
##                                                     (0.414)             (0.345)
##
## periodannounce_period:search_typegoogle_news       -29.652***          -29.652***
##                                                     (6.439)             (3.158)
##
## periodrumor_period:search_typegoogle_news           10.223**            10.223***
##                                                     (4.696)             (3.158)
##
## periodannounce_period:search_typeyoutube           -35.690***          -35.690***
##                                                     (6.232)             (3.158)
##
## periodrumor_period:search_typeyoutube               9.091**             9.091***
##                                                     (4.602)             (3.158)
##
## Constant                                                             -3,928.716**
##                                                                      (1,979.375)
##
## -----------------------------------------------------------------------------------
## Observations                                         16,314              16,314
## R2                                                   0.062
## Adjusted R2                                          0.058
## Log Likelihood                                                      -70,195.290
## Akaike Inf. Crit.                                                   140,420.600
## Bayesian Inf. Crit.                                                 140,536.100
## F Statistic                                 133.648*** (df = 8; 16242)
## ===================================================================================
## Note:                                                 *p<0.1; **p<0.05; ***p<0.01
```

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.