Country-level analysis: models

Part of the final project for AQMSS II

Polikanov Stepan and Okisheva Vera

```r
source(here::here("utilities", "check_packages.R"))
source(here::here("utilities", "functions.R"))

conflicts_prefer(dplyr::filter)
```

```r
data_country <- read_rds(here("data", "data_built", "data_country.rds"))
```

Aggregated models

We first run models aggregated to the country-level. This approach has an attractive feature of simplicity and the inability to run out of degrees of freedom. We aggregate to the country- rather than city- or voting station- level because the indicators and variables available to us were either available only at the country-level, are possible to measure only at the country-level (export and import for example) or are comparable only between countries. This means that we may miss some of the between-city variation in the data, however due to selection this is better addressed in the exit-poll sample choice models.

Selection into conducting exit polls

As mentioned elsewhere, exit polls are not a representative sample of all voting stations abroad. As initiative that conducted them self-described as founded by "free people and independent activists from Russia living abroad". This means that this endeavor is easily labelled within the "non-systemic opposition" in Russia by both descriptive signals (civic engagement, control of elections to avoid electoral fraud and simply "activism") and scope of operation (WEIRD countries, mostly within OECD and popular Russian tourist or immigration spots such as Vietnam or Kazakhstan).

To empirically confirm this we run regressions with an outcome denoting whether an exit poll was conducted at a voting station abroad. We relate this to variables that might affect migration choice as well as the baseline number of migrants in the country.

```r
m1 <- lm(ep ~ vdem_polyarchy_2022 + log(mad_gdppc_2018) + orthodox_share
         + log(dist) + log(voters_in_list) + log(mean_trips), data = data_country)

m1.log <- glm(ep ~ vdem_polyarchy_2022 + log(mad_gdppc_2018) + orthodox_share
              + log(dist) + log(voters_in_list)+ log(mean_trips),
              data = data_country, family = "binomial")

summary(m1)
```

```
Call:
lm(formula = ep ~ vdem_polyarchy_2022 + log(mad_gdppc_2018) +
    orthodox_share + log(dist) + log(voters_in_list) + log(mean_trips),
    data = data_country)

Residuals:
    Min      1Q  Median      3Q     Max
-0.7548 -0.2099  0.0293  0.1848  0.6735

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)         -0.292865   0.537066  -0.545   0.5866
vdem_polyarchy_2022  0.746300   0.113408   6.581 1.29e-09 ***
log(mad_gdppc_2018)  0.022125   0.033887   0.653   0.5151
orthodox_share      -0.178875   0.177499  -1.008   0.3156
log(dist)           -0.090970   0.047854  -1.901   0.0597 .
log(voters_in_list)  0.124133   0.027041   4.590 1.10e-05 ***
log(mean_trips)      0.001009   0.011522   0.088   0.9304
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3014 on 120 degrees of freedom
  (15 observations deleted due to missingness)
Multiple R-squared:  0.6022,    Adjusted R-squared:  0.5823
F-statistic: 30.27 on 6 and 120 DF,  p-value: < 2.2e-16
```

```r
summary(m1.log)
```

```
Call:
glm(formula = ep ~ vdem_polyarchy_2022 + log(mad_gdppc_2018) +
    orthodox_share + log(dist) + log(voters_in_list) + log(mean_trips),
    family = "binomial", data = data_country)
```

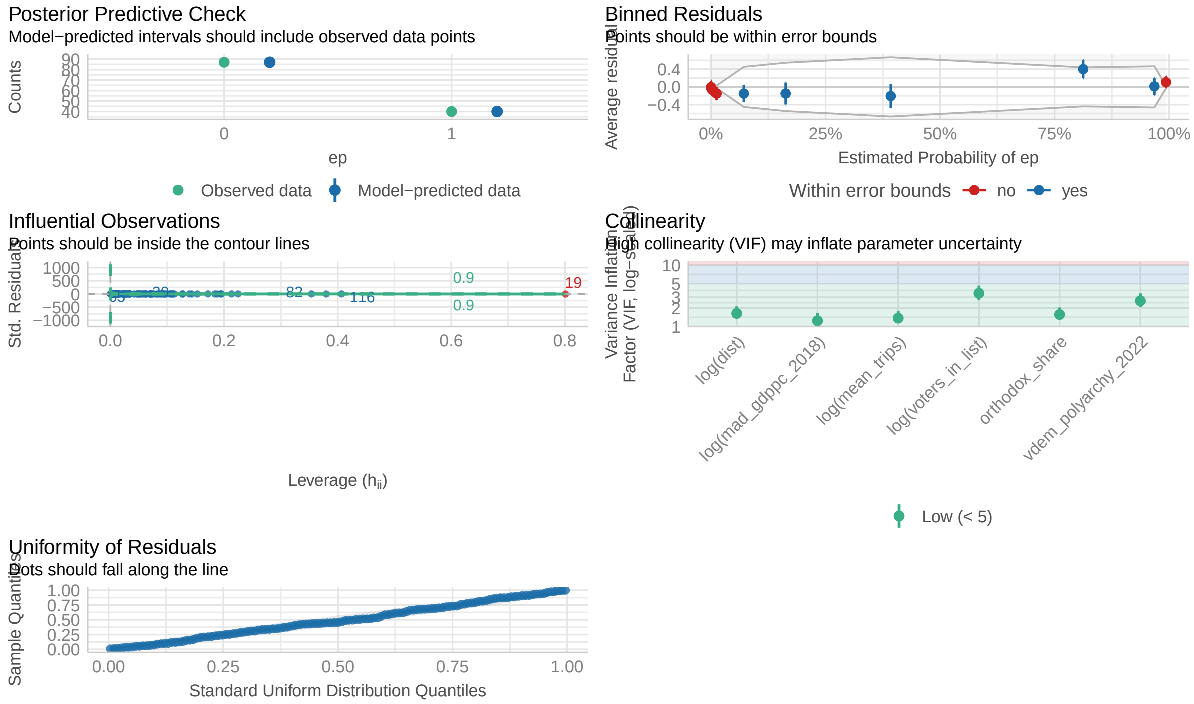
```
Coefficients:
                     Estimate Std. Error z value Pr(>|z|)    
(Intercept)         -29.85513   10.39357  -2.872 0.004073 ** 
vdem_polyarchy_2022   8.39480    2.25693   3.720 0.000200 ***
log(mad_gdppc_2018)   1.04301    0.58517   1.782 0.074682 .  
orthodox_share       -2.21453    1.90451  -1.163 0.244918    
log(dist)             0.02696    0.54767   0.049 0.960737    
log(voters_in_list)   1.95444    0.54075   3.614 0.000301 ***
log(mean_trips)       0.05063    0.12307   0.411 0.680811    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 158.245  on 126  degrees of freedom
Residual deviance:  47.089  on 120  degrees of freedom
  (15 observations deleted due to missingness)
AIC: 61.089

Number of Fisher Scoring iterations: 8
```

```r
check_m1.log <- plot(check_model(m1.log))
check_m1.log[[4]] <- check_m1.log[[4]] + theme(axis.text.x = element_text(
  angle = 45,hjust = 1)
  )
check_m1.log
```




```
save(m1.log, file = here("data", "data_built", "m1_log.rds"))
```

Vote shares

As the dependent variables here are vote shares (percent, 0-100 for use of interpretation), we use a linear regression model. Here is the simplest model using our preferred variables.

```
m2p <- lm(putin_full ~ orthodox_share + vdem_polyarchy_2022
          + log(mad_gdppc_2018) + obl_type + export_share + import_share
          + friendly_status + help + military_dummy + log(dist) + log(mean_trips),
          data = data_country)

m2d <- lm(davankov_full ~ orthodox_share + vdem_polyarchy_2022
          + log(mad_gdppc_2018) + obl_type + export_share + import_share
          + friendly_status + help + military_dummy + log(dist) + log(mean_trips),
          data = data_country)

m2s <- lm(spoiled_full ~ orthodox_share + vdem_polyarchy_2022
          + log(mad_gdppc_2018) + obl_type + export_share + import_share
          + friendly_status + help + military_dummy + log(dist) + log(mean_trips),
          data = data_country)
```

```
save(list = c("m2p", "m2d", "m2s"),
     file = here("data", "data_built", "lin.RData"))
```

```r
coef_map_lm <- c("(Intercept)",
                 "orthodox_share" = "Share of Orthodox Christians",
                 "vdem_polyarchy_2022" = "Polyarchy index",
                 "log(mad_gdppc_2018)" = "GDP per capita (log)",
                 "obl_type1" = "Military agreements: 1 (ref 0)",
                 "obl_type2" = "Military agreements: 2 (ref 0)",
                 "obl_type3" = "Military agreements: 3 (ref 0)",
                 "obl_type4" = "Military agreements: 4 (ref 0)",
                 "export_share" = "Export share",
                 "import_share" = "Import share",
                 "friendly_statusUnfriendly" =
                   "Unfriendly status (ref Neutral)",
                 "friendly_statusFriendly" =
                   "Friendly status (ref Neutral)",
                 "help" = "Help to Ukraine",
                 "military_dummy" = "Russian military presence",
                 "log(dist)" = "Geodesic distance (log)")
```

```r
modelsummary(list("Putin" = m2p, "Davankov" = m2d, "Spoiled" = m2s),
             output = "kableExtra",
             stars = T, vcov = "robust",
             coef_map = coef_map_lm) |>
  kable_styling(latex_options = c("scale_down"))
```

```r
check_m2p <- plot(check_model(m2p))
```

```r
check_m2p[[5]] <- check_m2p[[5]] + theme(axis.text.x = element_text(angle = 45,
                                                                      hjust = 1))


check_m2d <- plot(check_model(m2d))
```

```r
check_m2d[[5]] <- check_m2d[[5]] + theme(axis.text.x = element_text(angle = 45,
                                                                      hjust = 1))

check_m2s <- plot(check_model(m2s))
```

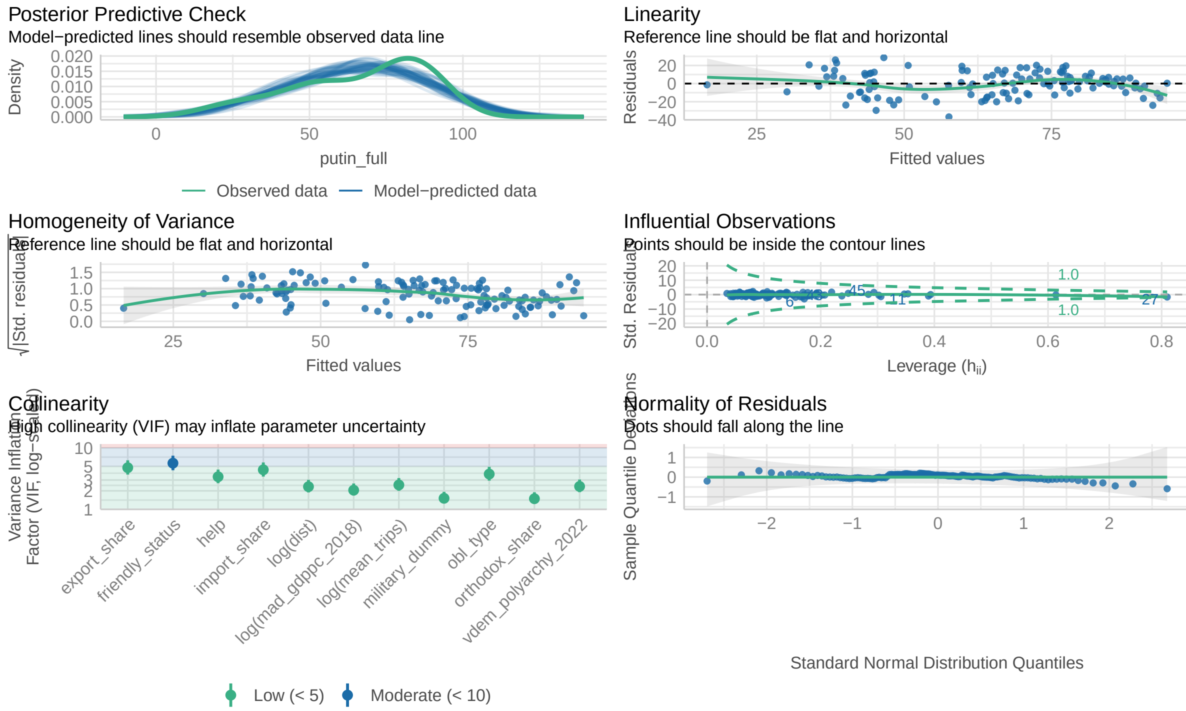
```r
check_m2s[[5]] <- check_m2s[[5]] + theme(axis.text.x = element_text(angle = 45,
                                                                      hjust = 1))
```

```r
check_m2p
```

Table 1: Linear models for vote shares

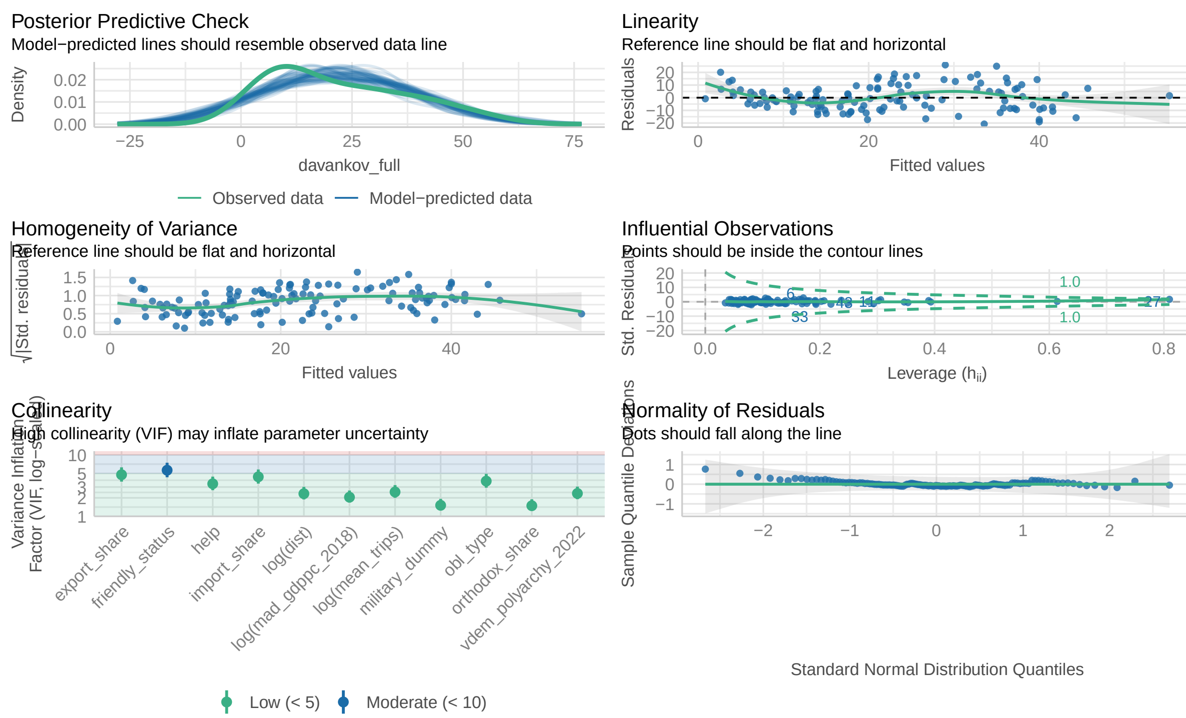
	Putin	Davankov	Spoiled
Share of Orthodox Christians	32.482**	−22.382*	−9.951***
	(11.218)	(9.315)	(2.830)
Polyarchy index	−21.605**	15.022**	7.708***
	(7.132)	(5.570)	(1.973)
GDP per capita (log)	−3.992*	3.374**	0.958*
	(1.587)	(1.212)	(0.376)
Military agreements: 1 (ref 0)	−5.746	3.272	1.660+
	(3.647)	(2.779)	(0.982)
Military agreements: 2 (ref 0)	−3.177	1.735	1.509
	(5.102)	(3.781)	(1.284)
Military agreements: 3 (ref 0)	−8.516	6.848	2.055
	(7.284)	(5.700)	(2.048)
Military agreements: 4 (ref 0)	2.972	−1.983	0.924
	(8.597)	(6.117)	(2.105)
Export share	−3.195**	2.251**	0.901*
	(1.055)	(0.842)	(0.421)
Import share	1.614	−1.189	−0.492
	(2.037)	(1.578)	(0.474)
Unfriendly status (ref Neutral)	−7.973	4.735	3.463
	(7.458)	(5.474)	(2.272)
Friendly status (ref Neutral)	−2.690	2.521	0.685
	(3.009)	(2.270)	(0.823)
Help to Ukraine	−4.335	2.510	1.991
	(6.758)	(4.666)	(2.005)
Russian military presence	2.270	−1.869	0.096
	(3.148)	(2.452)	(0.894)
Geodesic distance (log)	−2.606	2.394	−0.573
	(2.843)	(2.156)	(0.822)
Num.Obs.	125	125	125
R2	0.652	0.593	0.753
R2 Adj.	0.604	0.537	0.719
AIC	1023.2	959.4	696.2
BIC	1071.3	1007.5	744.3
Log.Lik.	−494.588	−462.708	−331.124
F	15.710	12.588	21.638
RMSE	12.65	9.80	3.42
Std.Errors	HC3	HC3	HC3

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$



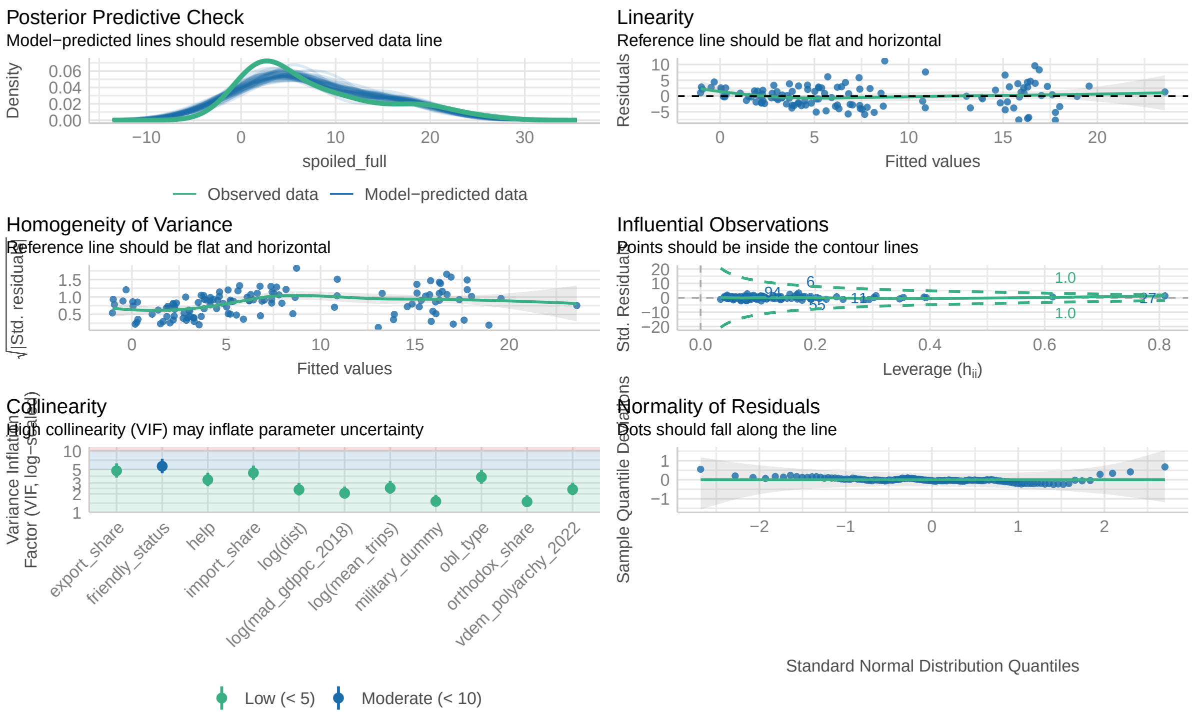
Posterior Predictive Check
Model−predicted lines should resemble observed data line
Density
putin_full
Observed data
Model−predicted data
Linearity
Reference line should be flat and horizontal
Residuals
Fitted values
Homogeneity of Variance
Reference line should be flat and horizontal
√|Std. residuals|
Fitted values
Influential Observations
Points should be inside the contour lines
Std. Residuals
Leverage (h_{ii})
Collinearity
High collinearity (VIF) may inflate parameter uncertainty
Variance Inflation Factor (VIF, log−scaled)
export_share
friendly_status
help
import_share
log(dist)
log(mad_gdppc_2018)
log(mean_trips)
military_dummy
obl_type
orthodox_share
vdem_polyarchy_2022
Low (< 5)
Moderate (< 10)
Normality of Residuals
Dots should fall along the line
Sample Quantile Deviations
Standard Normal Distribution Quantiles


```
check_m2d
```



Posterior Predictive Check
Model−predicted lines should resemble observed data line
Density
davankov_full
Observed data
Model−predicted data
Linearity
Reference line should be flat and horizontal
Residuals
Fitted values
Homogeneity of Variance
Reference line should be flat and horizontal
√|Std. residuals|
Fitted values
Influential Observations
Points should be inside the contour lines
Std. Residuals
Leverage (h_{ii})
Collinearity
High collinearity (VIF) may inflate parameter uncertainty
Variance Inflation Factor (VIF, log−scaled)
export_share
friendly_status
help
import_share
log(dist)
log(mad_gdppc_2018)
log(mean_trips)
military_dummy
obl_type
orthodox_share
vdem_polyarchy_2022
Low (< 5)
Moderate (< 10)
Normality of Residuals
Dots should fall along the line
Sample Quantile Deviations
Standard Normal Distribution Quantiles

```
check_m2s
```



Posterior Predictive Check
Model−predicted lines should resemble observed data line
Density
spoiled_full
Observed data
Model−predicted data
Linearity
Reference line should be flat and horizontal
Residuals
Fitted values
Homogeneity of Variance
Reference line should be flat and horizontal
√|Std. residuals|
Fitted values
Influential Observations
Points should be inside the contour lines
Std. Residuals
Leverage (h_{ii})
Collinearity
High collinearity (VIF) may inflate parameter uncertainty
Variance Inflation Factor (VIF, log−scaled)
export_share
friendly_status
help
import_share
log(dist)
log(mad_gdppc_2018)
log(mean_trips)
military_dummy
obl_type
orthodox_share
vdem_polyarchy_2022
Low (< 5)
Moderate (< 10)
Normality of Residuals
Dots should fall along the line
Sample Quantile Deviations
Standard Normal Distribution Quantiles