

Country-level analysis: models

Part of the final project for AQMSS II

Polikanov Stepan and Okisheva Vera

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

```

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

```
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	

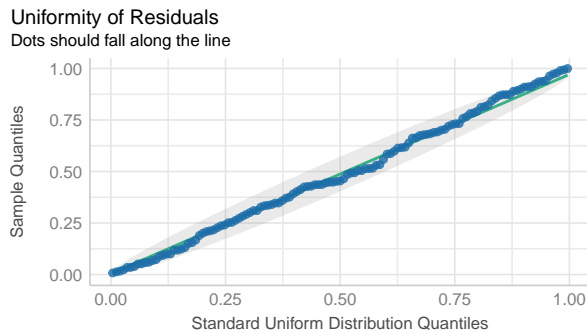
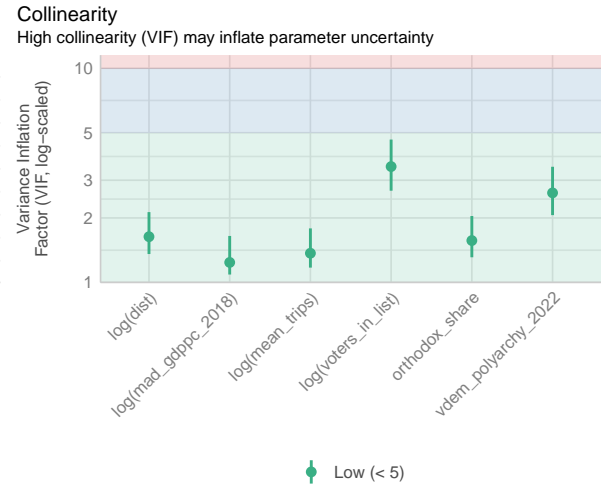
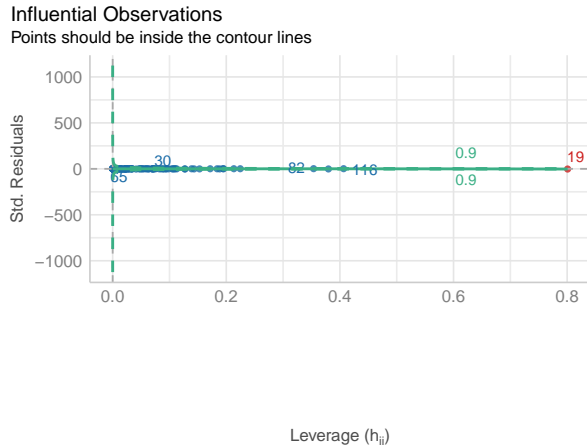
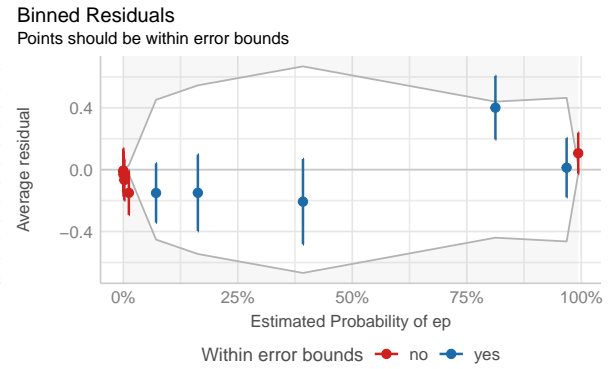
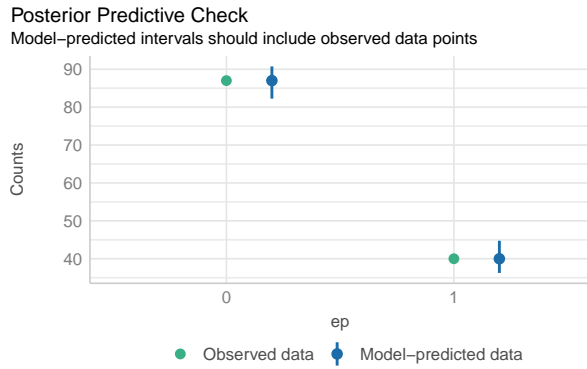
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(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

```
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("scripts", "models", "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. We start with 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)

resizebox.stargazer(m2p, m2d, m2s,
type = "latex",
title = "Linear models for vote shares",
tab.height = "\\textheight", tab.width= "\\textwidth")

```

% Table created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac at gmail.com % Date and time: Thu, May 16, 2024 - 7:25:32 PM

```

check_m2p <- plot(check_model(m2p))

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

check_m2d <- plot(check_model(m2d))

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

check_m2s <- plot(check_model(m2s))

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

check_m2p

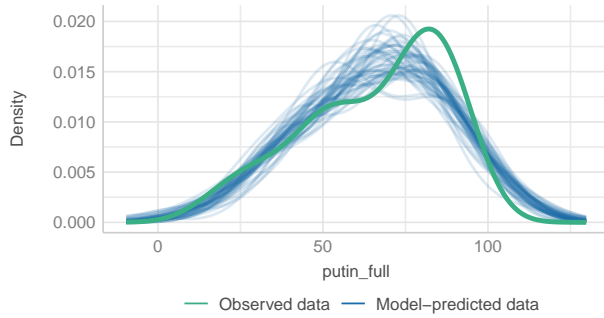
```

Table 1: Linear models for vote shares

	<i>Dependent variable:</i>		
	putin_full	davankov_full	spoiled_full
	(1)	(2)	(3)
orthodox_share	32.482*** (8.433)	−22.382*** (6.535)	−9.951*** (2.281)
vdem_polyarchy_2022	−21.605*** (6.843)	15.022*** (5.303)	7.708*** (1.851)
log(mad_gdppc_2018)	−3.992** (1.612)	3.374*** (1.249)	0.958** (0.436)
obl_type1	−5.746 (3.471)	3.272 (2.689)	1.660* (0.939)
obl_type2	−3.177 (4.463)	1.735 (3.459)	1.509 (1.207)
obl_type3	−8.516 (6.017)	6.848 (4.662)	2.055 (1.627)
obl_type4	2.972 (9.438)	−1.983 (7.313)	0.924 (2.552)
export_share	−3.195** (1.472)	2.251* (1.140)	0.901** (0.398)
import_share	1.614 (1.010)	−1.189 (0.783)	−0.492* (0.273)
friendly_statusUnfriendly	−7.973 (6.411)	4.735 (4.968)	3.463** (1.734)
friendly_statusFriendly	−2.690 (3.133)	2.521 (2.428)	0.685 (0.847)
help	−4.335 (4.605)	2.510 (3.569)	1.991 (1.245)
military_dummy	2.270 (3.671)	−1.869 (2.845)	0.096 (0.993)
log(dist)	−2.606 (2.512)	2.394 (1.946)	−0.573 (0.679)
log(mean_trips)	−0.736 (0.473)	0.552 (0.366)	0.166 (0.128)
Constant	148.891*** (26.415)	−46.176** (20.469)	−4.777 (7.144)
Observations	125	125	125
R ²	0.652	0.593	0.753
Adjusted R ²	0.604	0.537	0.719
Residual Std. Error (df = 109)	13.548	10.498	3.664
F Statistic (df = 15; 109)	13.628***	10.592***	22.116***
<i>Note:</i>		*p<0.1; **p<0.05; ***p<0.01	

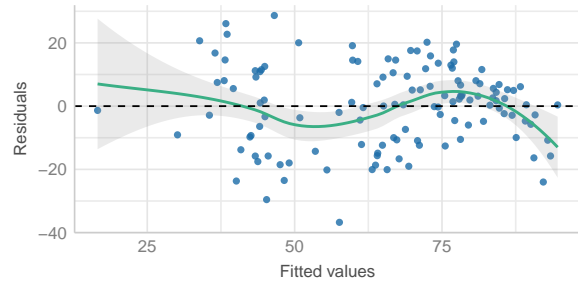
Posterior Predictive Check

Model-predicted lines should resemble observed data line



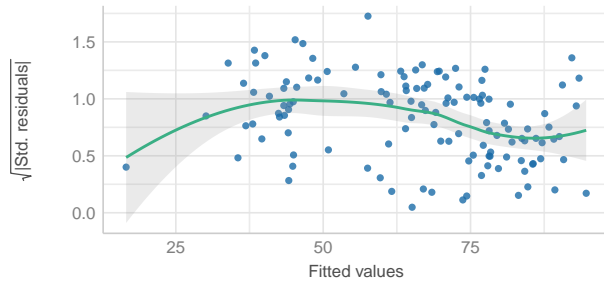
Linearity

Reference line should be flat and horizontal



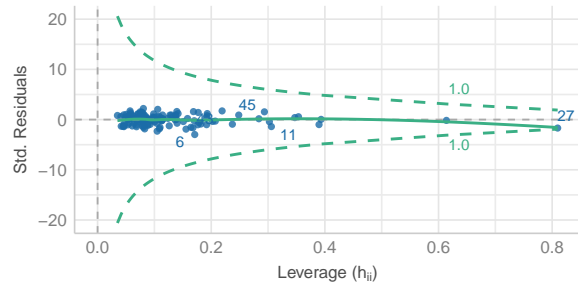
Homogeneity of Variance

Reference line should be flat and horizontal



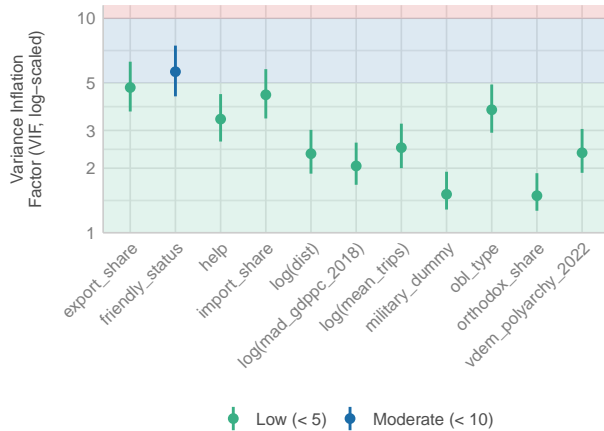
Influential Observations

Points should be inside the contour lines



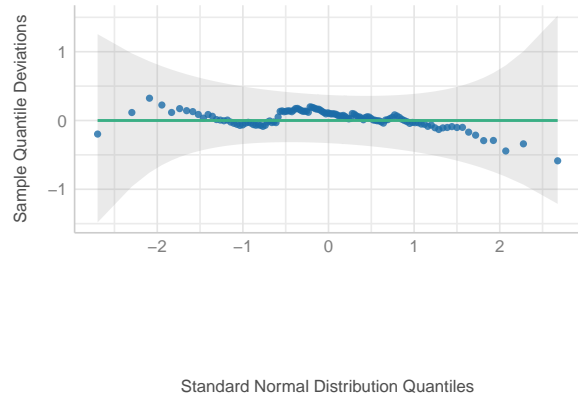
Collinearity

High collinearity (VIF) may inflate parameter uncertainty



Normality of Residuals

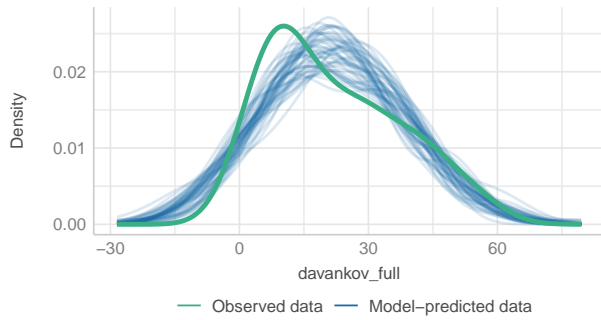
Dots should fall along the line



check_m2d

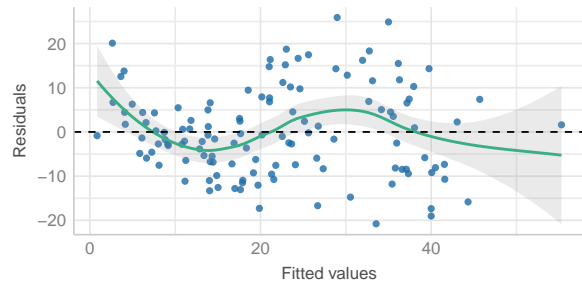
Posterior Predictive Check

Model-predicted lines should resemble observed data line



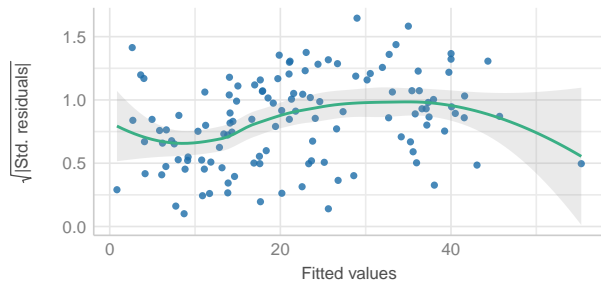
Linearity

Reference line should be flat and horizontal



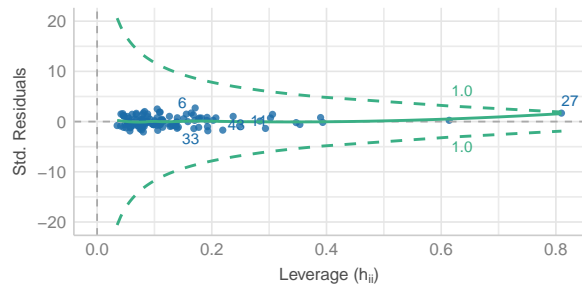
Homogeneity of Variance

Reference line should be flat and horizontal



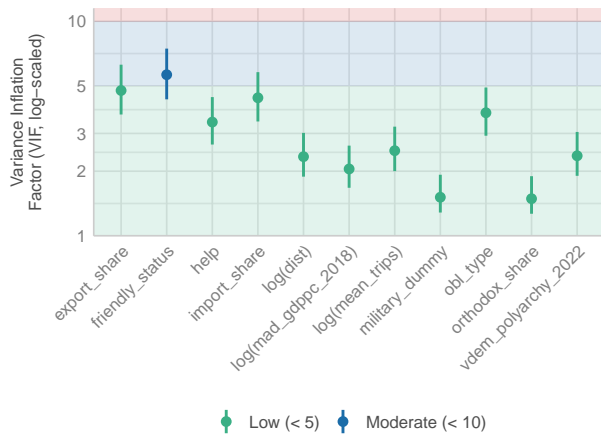
Influential Observations

Points should be inside the contour lines



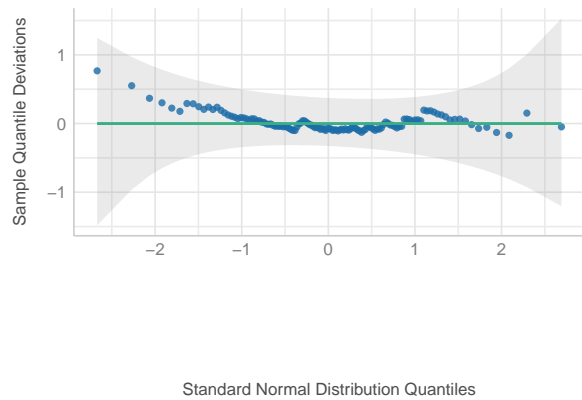
Collinearity

High collinearity (VIF) may inflate parameter uncertainty



Normality of Residuals

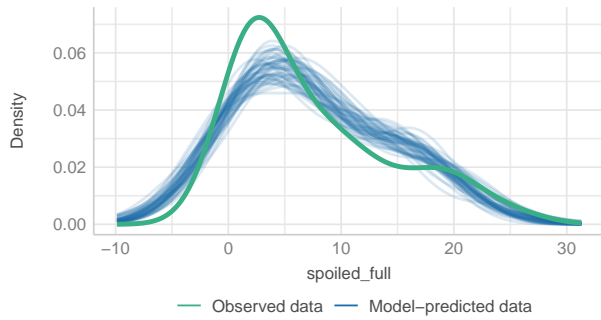
Dots should fall along the line



check_m2s

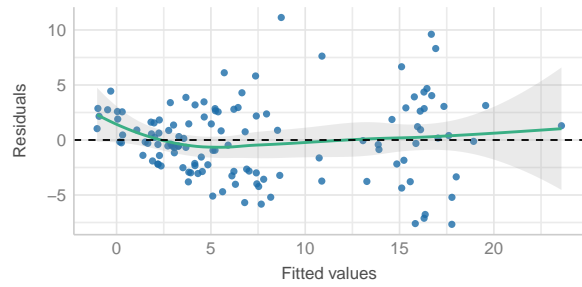
Posterior Predictive Check

Model-predicted lines should resemble observed data line



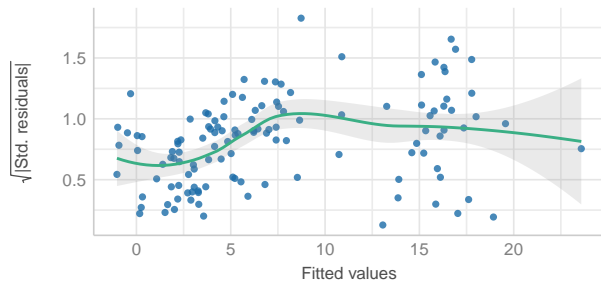
Linearity

Reference line should be flat and horizontal



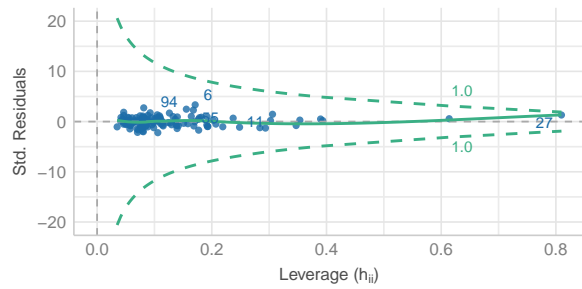
Homogeneity of Variance

Reference line should be flat and horizontal



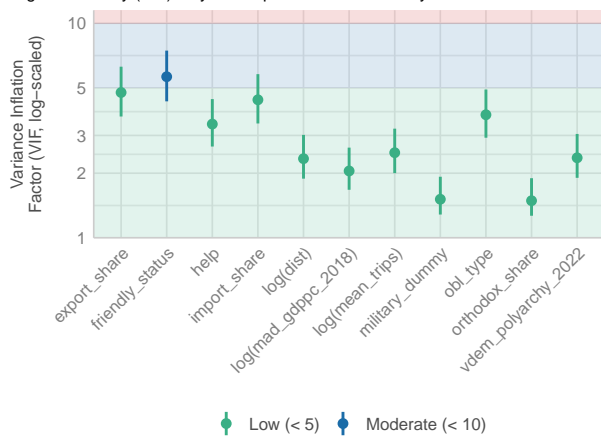
Influential Observations

Points should be inside the contour lines



Collinearity

High collinearity (VIF) may inflate parameter uncertainty



Normality of Residuals

Dots should fall along the line

