Modelling

Peasant unrest and imperial repression

Stepan Polikanov

```
## Building data for the analysis
## Dependencies: data_raw folder

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

conflicts_prefer(sfnetworks::activate)
conflicts_prefer(dplyr::filter)
conflicts_prefer(dplyr::lag)
conflicts_prefer(dplyr::select)
```

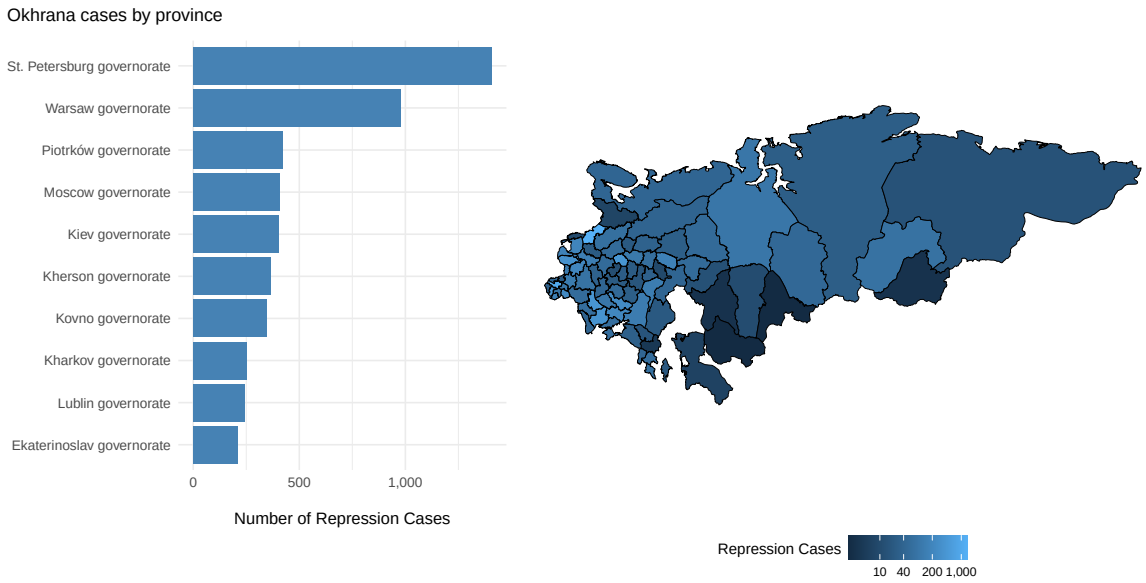
```
okhrana_full <- read_rds(here("data", "data_built", "okhrana_full.rds"))
```

```
bar_plot <- okhrana_full |>
  slice_max(total_cases, n = 10) |>
  ggplot(aes(x = total_cases,
                              y = fct_reorder(prov_ENG, total_cases))) +
  geom_col(fill = "steelblue") +
  labs(
    x = "\nNumber of Repression Cases",
    y = NULL,
  ) +
  scale_x_continuous(labels = scales::comma) +
  theme_minimal(base_size = 12)

map_plot <- ggplot(st_as_sf(okhrana_full)) +
  geom_sf(aes(fill = log1p(total_cases)), color = "black", size = 0.2) +
  scale_fill_viridis_c(option = "plasma", na.value = "grey90") +
  theme_void() +
  scale_fill_continuous(
    name = "Repression Cases\n",
    trans = "identity",
    labels = function(x) scales::comma(round(expm1(x))),
    breaks = log1p(c(0, 10, 40, 200, 1000))  # customize based on your range
  ) +
  theme(legend.position = "bottom")

okhrana_by_province <- bar_plot + map_plot +
  plot_layout(ncol = 2, widths = c(1, 2)) +
  plot_annotation(
    title = "Okhrana cases by province"
  )
```

```
okhrana_by_province
```




```
# write to paper/ as okhrana_by_province.png
ggsave(here("paper", "okhrana_by_province.png"),
       okhrana_by_province, width = 12, height = 6, dpi = 300)
```

```
# Refit models
model1 <- lm(log1p(total_cases) ~ log1p(events), data = okhrana_full)
model2 <- lm(log1p(total_cases) ~ log1p(events) + land_gini, data = okhrana_full)
model3 <- lm(log1p(total_cases) ~ log1p(events) + land_gini + sh_serfs1858 + peasant_share, data = okhrana_full)
model4 <- lm(log1p(total_cases) ~ log1p(events) + land_gini + sh_serfs1858 + peasant_share +
               urbanization_1904 + manufacturing_share_1904, data = okhrana_full)
model5 <- lm(log1p(total_cases) ~ log1p(events) + land_gini + sh_serfs1858 + peasant_share +
               urbanization_1904 * land_gini + manufacturing_share_1904, data = okhrana_full)

# Extract all term names used across models
all_terms <- unique(unlist(lapply(list(model1, model2, model3, model4, model5), function(m) names(coef(m)))))

# Create a default mapping that falls back to original names if not mapped manually
custom_labels <- c(
  "log1p(events)" = "(Log) Pre-1905 unrest",
  "land_gini" = "Land inequality (Gini)",
  "sh_serfs1858" = "Share of serfs (1858)",
  "peasant_share" = "Peasant land share",
  "urbanization_1904" = "Urbanization (1904)",
  "manufacturing_share_1904" = "Manufacturing share (1904)",
  "land_gini:urbanization_1904" = "Urbanization × Inequality",
  "(Intercept)" = "Intercept"
)

# Fill in any unmapped terms to avoid errors
full_coef_map <- setNames(
  custom_labels[all_terms] %||% all_terms,
```

	1. Unrest only	2. + Inequality	3. + Serfdom & Peasants	4. + Urban & Industrial	5. + Urban × Inequality
Intercept	2.888***	3.104**	3.190	2.403	0.939
	(0.794)	(1.113)	(2.026)	(1.596)	(1.896)
(Log) Pre-1905 unrest	0.302	0.485+	0.352	0.534+	0.548*
	(0.194)	(0.255)	(0.339)	(0.268)	(0.265)
Land inequality (Gini)		−1.241	−1.021	−1.561	0.112
		(1.387)	(1.998)	(1.539)	(1.938)
Share of serfs (1858)			0.543	0.060	0.279
			(0.875)	(0.687)	(0.697)
Peasant land share			0.210	−0.234	−0.625
			(1.428)	(1.110)	(1.133)
Urbanization (1904)				6.297**	23.888+
				(1.826)	(12.758)
Manufacturing share (1904)				0.385	−1.325
				(3.804)	(3.957)
Urbanization × Inequality					−19.778
					(14.200)
Num.Obs.	52	49	49	48	48
R2	0.046	0.073	0.081	0.458	0.483
RMSE	1.07	1.08	1.08	0.81	0.79

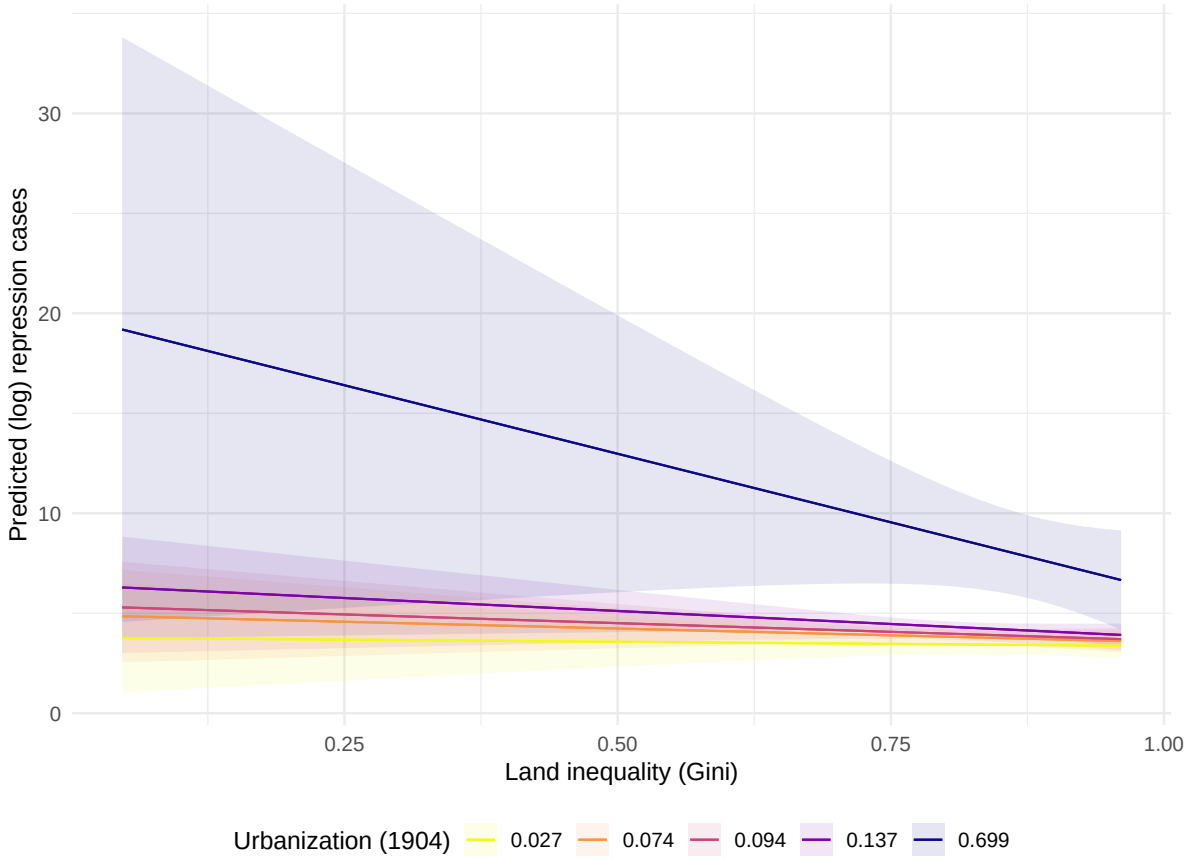
+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

```
  all_terms
)

# Generate clean summary
modelsummary(
  list(
    "1. Unrest only" = model1,
    "2. + Inequality" = model2,
    "3. + Serfdom & Peasants" = model3,
    "4. + Urban & Industrial" = model4,
    "5. + Urban × Inequality" = model5
  ),
  coef_map = full_coef_map,
  gof_omit = "IC|Log|Adj|F|Deviance|AIC|BIC",
  stars = TRUE,
  output = "kableExtra"
) |>
  kable_styling(latex_options = c("scale_down"))
```

```
interaction <- plot_predictions(model5,
                  condition = c("land_gini",
                                "urbanization_1904")) +
  labs(
    x = "Land inequality (Gini)",
    y = "Predicted (log) repression cases",
    color = "Urbanization (1904)",
    fill = "Urbanization (1904)",
  ) +
  scale_color_viridis_d(option = "plasma", direction = -1) +
  scale_fill_viridis_d(option = "plasma", direction = -1) +
  theme_minimal(base_size = 12) +
  theme(legend.position = "bottom")

interaction
```




```
# write to paper/ as interaction.png
ggsave(here("paper", "interaction.png"),
       interaction, width = 8, height = 6, dpi = 300)
```

```
nb_model <- glm.nb(total_cases ~ log1p(events) + land_gini + sh_serfs1858
                   + peasant_share + urbanization_1904
                   + manufacturing_share_1904,
                   data = okhrana_full)

model_rev <- lm(log1p(revolutionary_cases) ~ log1p(events) + land_gini
                + sh_serfs1858 + peasant_share
                + urbanization_1904 + manufacturing_share_1904,
                data = okhrana_full)

model_pop <- lm(log1p(total_cases) ~  log1p(events) + land_gini
                + sh_serfs1858 + peasant_share + urbanization_1904
                + manufacturing_share_1904 +
                  log(population_1904),
                data = okhrana_full)

model_alt_ineq <- lm(log1p(total_cases) ~ log1p(events)
                     + peasant_share + sh_serfs1858
                     + urbanization_1904 + manufacturing_share_1904,
```

	Control for population	Revolutionaries only	Alternative inequality measure	Negative Binomial
Intercept	−9.718*	2.241	1.208	31.326*
	(3.881)	(1.557)	(1.076)	(43.501)
(Log) Pre-1905 unrest	0.154	0.661*	0.465+	1.799*
	(0.265)	(0.262)	(0.260)	(0.420)
Land inequality (Gini)	−1.877	−2.277		0.082+
	(1.379)	(1.502)		(0.109)
Share of serfs (1858)	0.302	−0.289	0.110	1.012
	(0.619)	(0.670)	(0.686)	(0.604)
Peasant land share	0.501	−0.601	0.553	0.475
	(1.017)	(1.084)	(0.794)	(0.459)
Urbanization (1904)	5.653**	4.182*	6.286**	593.362***
	(1.644)	(1.782)	(1.827)	(934.820)
Manufacturing share (1904)	1.221	2.336	0.216	0.709
	(3.411)	(3.712)	(3.802)	(2.331)
Num.Obs.	48	48	48	48
R2	0.577	0.390	0.445	
R2 Adj.	0.503	0.301	0.378	
AIC	520.2	472.8	529.4	531.2
BIC	537.1	487.8	542.4	546.2
Log.Lik.	−51.947	−56.729	−58.502	−257.593
F	7.806	4.377	6.725	7.637
RMSE	0.71	0.79	0.82	183.58

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

```
                  data = okhrana_full)

modelsummary(
  list("Control for population" = model_pop,
       "Revolutionaries only" = model_rev,
       "Alternative inequality measure" = model_alt_ineq,
       "Negative Binomial" = nb_model),
  coef_map = full_coef_map,
  exponentiate = c(F, F, F, T),
  stars = TRUE,
  output = "kableExtra"
) |>
  kable_styling(latex_options = c("scale_down"))
```