Analysing Vote Choice Data

Assignment 3

Jacob Edenhofer*

14 May 2023

Preliminaries

Let us import the necessary packages and the data:

```
# packages
library(tidyverse)
library(here)
library(modelsummary)
library(haven)
library(ggpubr)
library(knitr)
library(kableExtra)
library(ggeffects)
library(fixest)
library(lme4)
library(margins)
library(bife)

# data
ess789 <- read_dta(paste0(here(), "/Data/ESS789.dta"))
```

Exercise 1

1.1

Make sure that variable gndr is a dummy taking values 0/1, Rescale variables ipequopt and impfree so that higher values measure higher importance, Create variable year for each wave of the survey, Create a categorical variable cohort that measure in which decade the respondent was born. Make thevariable have only 4 levels, one for each quartile of the year of birth distribution

To prepare the data in the desired way, I run:

*jacob.edenhofer@some.ox.ac.uk

```
ess789_mod <- ess789 %>%
  # 1 for males, 0 for females
  mutate(gndr_dummy = ifelse(gndr == 1, 1, 0),
         gndr_dummy = factor(gndr_dummy),
         ipeqopt_recoded = recode(as.numeric(ipeqopt),
                                  "1" = 6,
                                  "2" = 5,
                                  "3" = 4,
                                  "4" = 3,
                                  "5" = 2,
                                  "6" = 1),
         impfree_recoded = recode(as.numeric(impfree),
                                  "1" = 6,
                                  "2" = 5,
                                  "3" = 4,
                                  "4" = 3,
                                  "5" = 2,
                                  "6" = 1),
         # from ess website
         year = case_when(essround == 7 ~ 2014,
                          essround == 8 ~ 2016,
                          TRUE ~ 2018),
         # quartiles obtained by running quantile(ess789_mod$agea, na.rm = T)
         cohort = case_when((agea >=14 & agea < 35) ~ "[14, 35)",
                            (agea >= 35 & agea < 50) ~ "[35, 50)",
                            (agea >= 50 & agea < 64) ~ "[50, 64)",
                            TRUE ~ "[64, 114)"),
         mnrchy_factor = factor(mnrchy),
         eummbr_factor = factor(eummbr))
```

1.2

Look at the variables in the dataset: which ones vary at the individual level? Which at the country level? And which at the country-year level?

I summarise the levels of variation for the different variables in table 1:

```
# dataframe
df_var <- tribble(~"Variable", ~"Description",
                  "env", "level of green attitudes in a given country in a given year",
                  "cons", "level of social conservativism in a given country in a given year",
                  "eummbr", "EU membership dummy",
                  "mnrchy", "Consitutional monarchy dummy",
                  "ipequopt", "whether respondent believes that it is important that people are treated
```

```
                  "impfree", "whether the respondent believes that it is important to make own decision
                  "uemp5yr", " periods of unemployment experienced by the respondent in the five previo
                  "gndr", "respondent's gender",
                  "agea", "respondent's age")

# table
df_var %>%
  kbl(booktabs = T, caption = "Summary table of levels of variation") %>%
  kable_styling(latex_options = "hold_position") %>%
  pack_rows("varies at country-year level", 1, 2) %>%
  pack_rows("varies at country level", 3, 4) %>%
  pack_rows("varies at individual level", 5, 9)
```

Table 1: Summary table of levels of variation

Variable	Description
varies at country-year level	
env	level of green attitudes in a given country in a given year
cons	level of social conservativism in a given country in a given year
varies at country level	
eummbr	EU membership dummy
mnrchy	Consitutional monarchy dummy
varies at individual level	
ipequopt	whether respondent believes that it is important that people are treated equally and have
impfree	whether the respondent believes that it is important to make own decisions and be free
uemp5yr	periods of unemployment experienced by the respondent in the five previous years
gndr	respondent's gender
agea	respondent's age

What's the mean value of variables capturing the importance of freedom and equality for respondents?, Do they differ between countries with a Constitutional Monarchy and those without? And between EU members and non-members? Report your results in a nice, tidy table.

To compare the mean values of `impfree_recoded` and `ipeqopt_recoded` between respondents living in constitutional monarchies, as opposed to those who do not, I run:

```
ess789_mod %>%
 dplyr::select(ipeqopt_recoded, impfree_recoded, mnrchy_factor) %>%
 datasummary_balance(~mnrchy_factor, fmt = 3,
                     dinm_statistic = "p.value",
                     title = "Comparing mean values between constitutional monarchies and republics",
                     output = "kableExtra",
                     data = .) %>%
  kable_styling(latex_options = "hold_position")
```

Table 2 shows that respondents in constitutional monarchies, on average, accord greater importance to equality

Table 2: Comparing mean values between constitutional monarchies and republics

	0 (N=74502)		1 (N=25394)			
	Mean	Std. Dev.	Mean	Std. Dev.	Diff. in Means	p
ipeqopt_recoded	4.825	1.086	5.035	0.942	0.211	<0.001
impfree_recoded	4.821	1.106	4.856	1.060	0.035	<0.001

than their counterparts in republics, with the difference being significant at the 1% level. The same holds for `impfree`, though difference in means is small.

To compare the mean values of `impfree_recoded` and `ipeqopt_recoded` between respondents living in EU member states, as opposed to those who do not, I run:

```
ess789_mod %>%
  dplyr::select(ipeqopt_recoded, impfree_recoded, eummbr_factor) %>%
  datasummary_balance(~eummbr_factor, fmt = 3,
                      dinm_statistic = "p.value",
                      title = "Comparing mean values between EU members and non-members",
                      output = "kableExtra",
                      data = .) %>%
  kable_styling(latex_options = "hold_position")
```

Table 3: Comparing mean values between EU members and non-members

	0 (N=8986)		1 (N=90910)			
	Mean	Std. Dev.	Mean	Std. Dev.	Diff. in Means	p
ipeqopt_recoded	4.901	1.044	4.876	1.056	-0.025	0.031
impfree_recoded	4.941	1.074	4.819	1.096	-0.123	<0.001

Table 3 shows that respondents in EU member states, on average, accord less importance to equality than their counterparts in non-EU member states, with the difference being significant at the 1% level. The same holds for `impfree`, though difference in means is small.

Finally, for each observation, create a variable indicating how much more (or less) the respondent value freedom over equality

To create this variable, I subtract `ipeqopt_recoded` from `impfree_recoded`. This variable is zero for respondents who agree to the same extent with both items, negative for those who agree more strongly with `ipeqopt` than with `impfree`, and positive for those who agree more strongly with `impfree` and `ipeqopt`.

```
ess789_mod <- ess789_mod %>%
  mutate(free_equal_diff = impfree_recoded - ipeqopt_recoded)
```

1.3

Which are the factors that better predict whether a respondent prefers freedom over equality? (Hint: build your dependent variable first). Plot the coefficients and comment their significance. Plot how the predicted probabilities of preferring freedom over equality change for male and female respondents conditionally on their experience of unemployment.

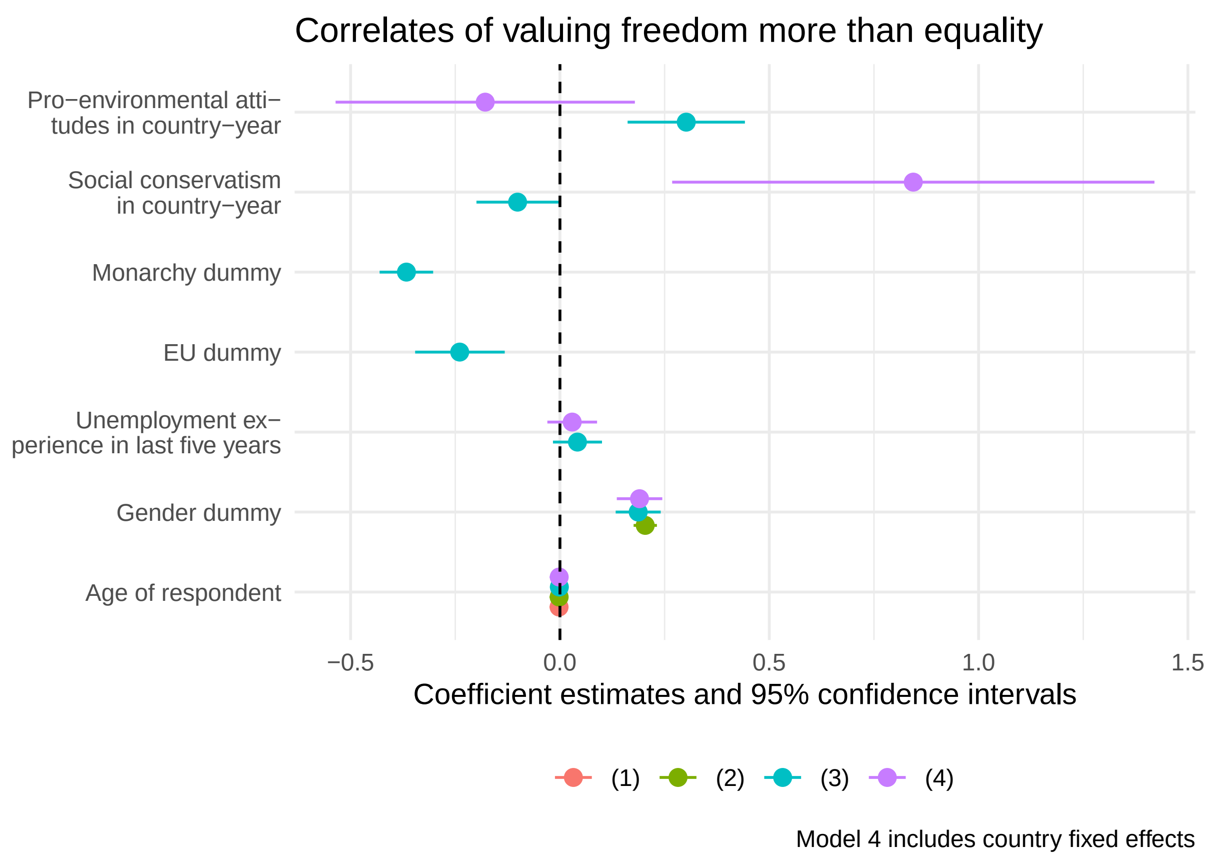
My dependent variable, `free_better_dummy`, is a binary variable that takes the value of one if `free_equal_diff` is positive, i.e. if a respondent agrees more strongly with `impfree` than with `ipeqopt`, and is zero otherwise. I then estimate four logit specifications, with that variable as my dependent variable:

- I start by regressing `free_better_dummy` on respondents' age following the literature on long-term value changes (e.g. Inglehart and Welzel 2010).
- Then, I add dummy for respondents' gender, reflecting recent arguments that men and women have systematically different social attitudes (e.g. Anduiza and Rico 2022; Oshri et al. 2022).
- Next, I add a dummy for unemployment experience, given that adverse economic shocks may affect beliefs about equality and freedom. Since country's EU (non-)membership and its status as a constitutional monarchy might also influence respondents' social attitudes I include dummies for these as well. Finally, I include a country's overall level of social conservatism and environmental concern in a given year since these can be construed as proxies for the broader societal context within which individuals form their own attitudes.
- The final model is almost identical to model three, except for `eummbr_factor` and `mnrchy_factor` being excluded. This is because model four includes country fixed effects, which control for all (un)observed factors that vary across countries, but are constant over time. Since `eummbr_factor` and `mnrchy_factor` are constant over time, their inclusion is rendered superfluous by the country fixed effects.

```r
# dependent variable
ess789_mod <- ess789_mod %>%
  mutate(free_better_dummy = ifelse(free_equal_diff > 0, 1, 0),
         uemp5yr_factor = factor(uemp5yr))

# model
free_better_model1 <- glm(free_better_dummy ~ agea,
                          family = binomial(link = "logit"),
                          data = ess789_mod)
free_better_model2 <- glm(free_better_dummy ~ agea + gndr_dummy,
                          family = binomial(link = "logit"),
                          data = ess789_mod)
free_better_model3 <- glm(free_better_dummy ~ agea + gndr_dummy + uemp5yr
                           + eummbr_factor + mnrchy_factor + cons + env,
                          family = binomial(link = "logit"),
                          data = ess789_mod)
free_better_model4 <- bife(free_better_dummy ~ agea + gndr_dummy + uemp5yr + cons + env | cntry,
                                  model = "logit", data = ess789_mod)
```

```r
# coefficient plot
modelplot(list(free_better_model1, free_better_model2,
               free_better_model3, free_better_model4),
               coef_map = c("agea" = "Age of respondent",
                            "gndr_dummy1" = "Gender dummy",
                            "uemp5yr" = "Unemployment ex-\nperience in last five years",
                            "eummbr_factor1" = "EU dummy",
                            "mnrchy_factor1" = "Monarchy dummy",
                            "cons" = "Social conservatism\nin country-year",
                            "env" = "Pro-environmental atti-\ntudes in country-year")) +
  geom_vline(xintercept = 0, linetype = "dashed") +
  expand_limits(x = -0.5) +
  labs(title = "Correlates of valuing freedom more than equality",
       caption = "Model 4 includes country fixed effects") +
  theme(legend.position = "bottom")
```



The coefficient plot implies four lessons:

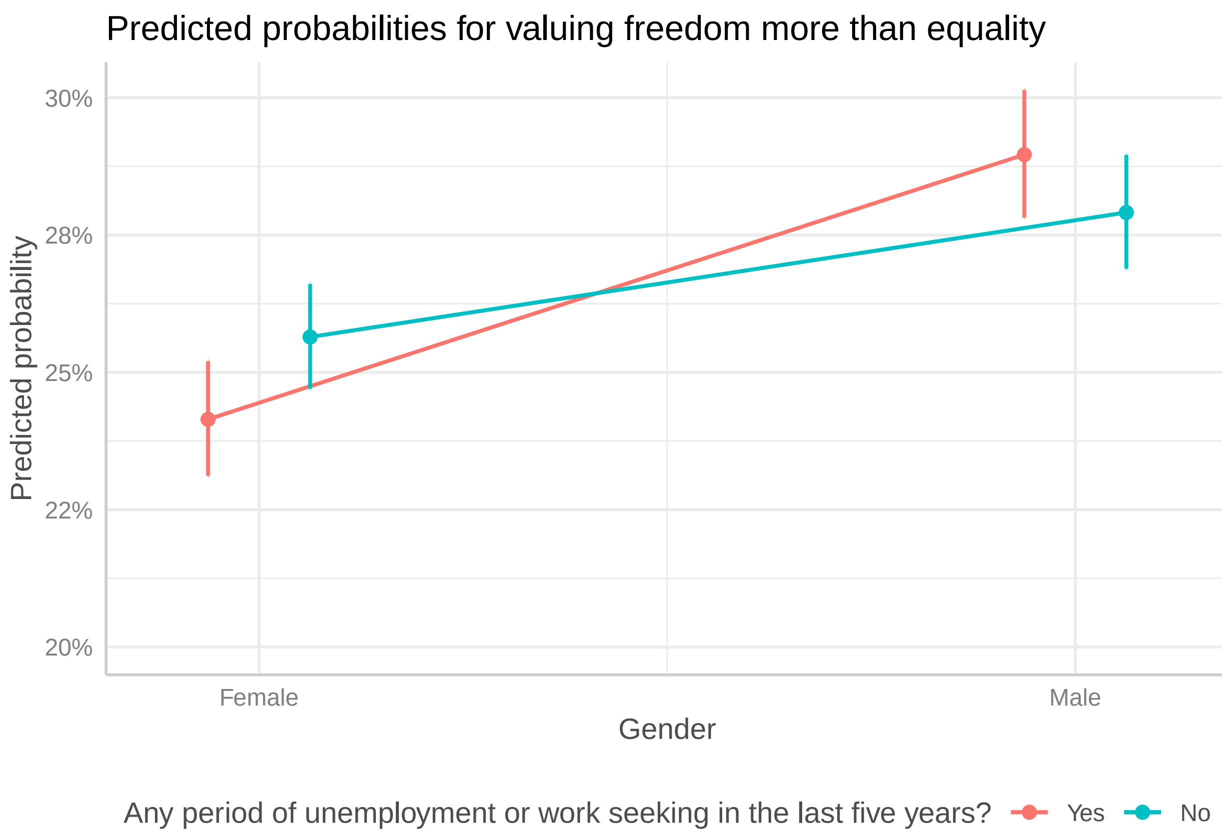
- Gender is consistently a statistically significant (at the 5% level) predictor of valuing freedom more than equality, with men being, on average, more likely to do so than females, holding all other included covariates constant. Substantively, the log odds are roughly 20% ($100 * (exp(0.18) - 1)$) for men than for females.

- Age and unemployment experience do not significantly predict preferring freedom over equality.
- On average, respondents residing in EU countries are, compared to their non-EU counterparts, significantly less likely to prefer freedom over equality, holding all other included covariates constant. Similarly, respondents residing in constitutional monarchies are less likely to express such a preferences, relative to those living in republics. Substantively, the log odds are approximately 20% lower $(100 * (exp(-0.239) - 1))$ for EU, as opposed to non-EU, respondents, and 30% $(100 * (exp(-0.367) - 1))$ lower for respondents in constitutional monarchies.
- The inclusion of country fixed effects leads to a strongly positive association between overall social conservatism in a given year and a preference for freedom over equality, with the log odds increasing by roughly 130% for a unit increase in social conservatism $(100 * (exp(0.84) - 1))$. By contrast, the coefficient estimate for pro-environmental attitudes becomes insignificant when including country fixed effects, suggesting that the original positive association is driven by (un)observed confounders.

Plot how the predicted probabilities of preferring freedom over equality change for male and female respondents conditionally on their experience of unemployment.

To plot the predicted probabilities, I use the `ggpredict()` function applied to a simple regression of `free_better_dummy` on the interaction between `gndr_dummy1` and `uemp5yr_factor`.

```
# data
ess789_mod <- ess789_mod %>%
  mutate(gndr_dummy1 = factor(gndr_dummy, levels = c("0", "1"),
                              labels = c("Female", "Male")))
# model
free_better_model5 <- glm(free_better_dummy ~ gndr_dummy1*uemp5yr_factor,
                         family = binomial(link = "logit"),
                         data = ess789_mod)
# plot
plot(ggpredict(free_better_model5, terms = c("gndr_dummy1", "uemp5yr_factor")),
     connect.lines = T) +
  scale_colour_discrete("Any period of unemployment or work seeking in the last five years?",
                        labels = c("1" = "Yes",
                                   "2" = "No")) +
  labs(x = "Gender", y = "Predicted probability",
      title = "Predicted probabilities for valuing freedom more than equality") +
  expand_limits(y = c(0.2, 0.3)) +
  theme(legend.position = "bottom")
```



The vertical differences between the point estimates represent the marginal effect of unemployment experience for men and women respectively. For women, the marginal effect of having experience unemployment within the last five years is negative, while it is positive for men. That is, men become more likely to prefer freedom over equality after having experienced unemployment (the difference is significant at the 5% level, which can be seen by running `summary(free_better_model5)`), with the reverse holding for women.

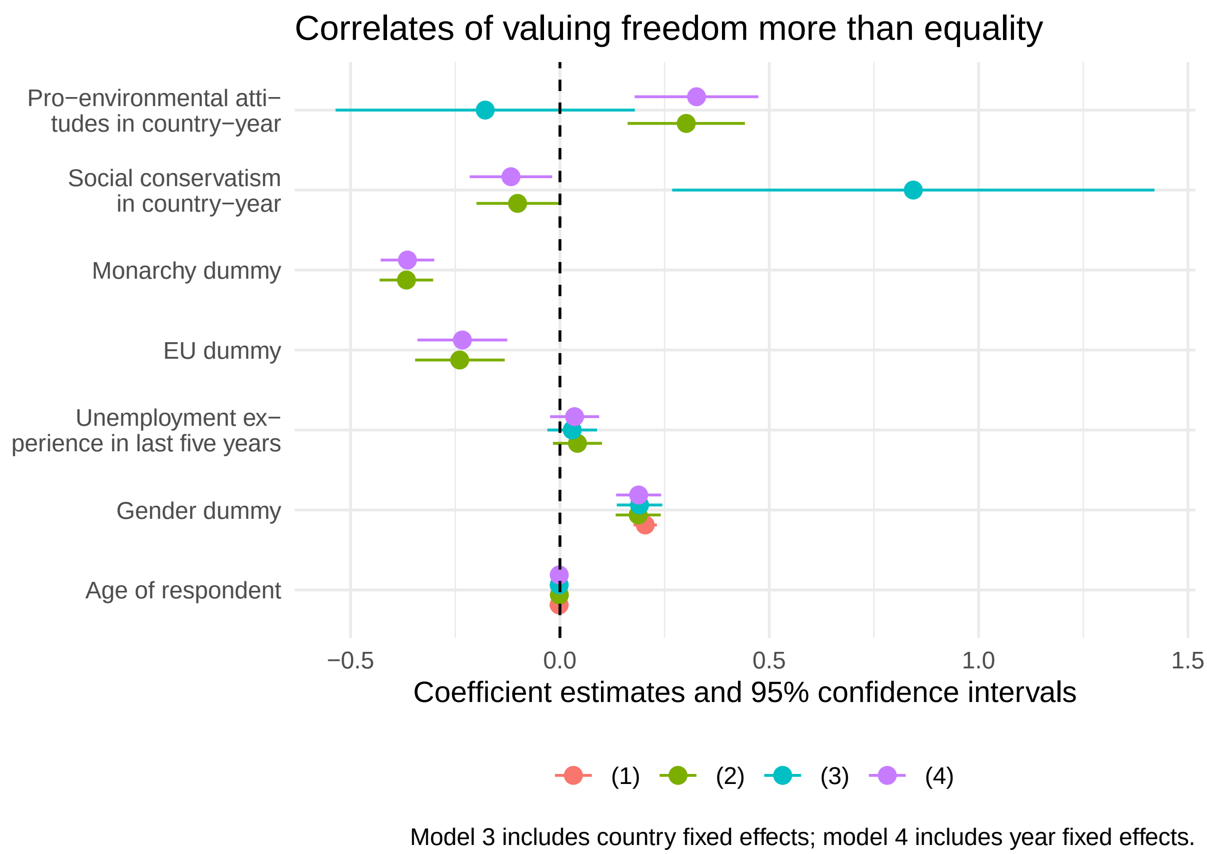
1.4

Estimate the model above using year-level fixed effects: What do the year-level fixed effects exactly do?, What are the variables that change? How? And why those in particular?

By including year fixed effects, we restrict our attention to cross-country variation within each ESS wave. Doing so allows us to account for (confounding) factors, both observable and unobservable, that vary over time and are constant across countries, such as common economic shocks. Hence, I run:

```
free_better_model6 <- bife(free_better_dummy ~ agea + gndr_dummy + uemp5yr + eummbr_factor + mnrchy_fac
           model = "logit", data = ess789_mod)
# plot
modelplot(list(free_better_model2, free_better_model3,
               free_better_model4, free_better_model6),
          coef_map = c("agea" = "Age of respondent",
                       "gndr_dummy1" = "Gender dummy",
```

```
                         "uemp5yr" = "Unemployment ex-\nperience in last five years",
                         "eummbr_factor1" = "EU dummy",
                         "mnrchy_factor1" = "Monarchy dummy",
                         "cons" = "Social conservatism\nin country-year",
                         "env" = "Pro-environmental atti-\ntudes in country-year")) +
  geom_vline(xintercept = 0, linetype = "dashed") +
  labs(title = "Correlates of valuing freedom more than equality",
       caption = "Model 3 includes country fixed effects; model 4 includes year fixed effects.") +
  theme(legend.position = "bottom")
```



change in variables

1.5

If you were asked at which other level you would add fixed effects, what would you answer?

Ideally, I would want to probe the robustness of the above results by including year-wave fixed effects. In this way, we could control both for country-specific, time-invariant (un)observable confounders (country fixed effects), and for wave-specific, country-invariant (un)observable confounders (wave fixed effects).

Exercise 2

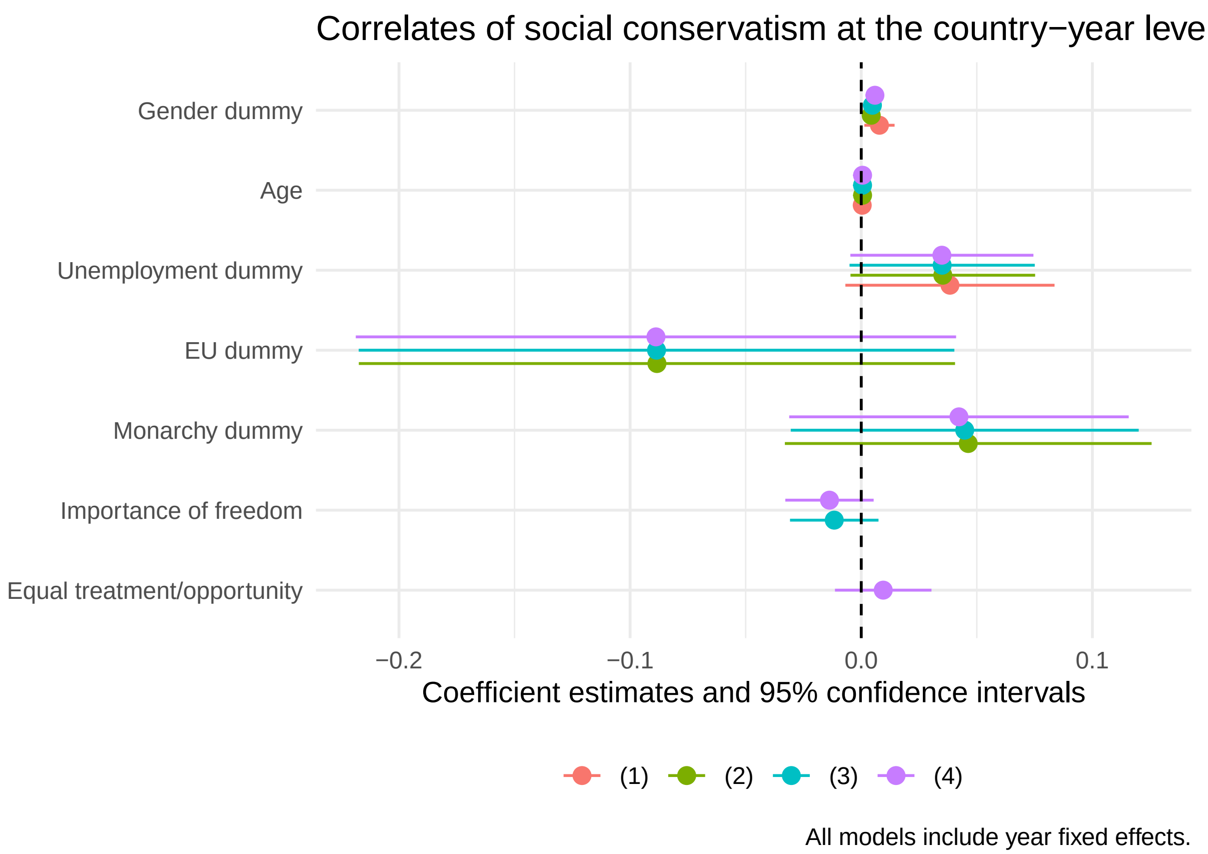
Re-estimate the model above using year-level fixed effects. This time, however, use a different dependent variable: the level of country's conservativism.

I estimate four specifications with `cons` as the dependent variable. I include fixed effects via the `feols()` function from the `fixest` package, which is computationally more efficient than the `plm()` function, and clusters the standard errors at the level of the fixed effects (here the year level).

The justification of all additional covariates is, for the most part, analogous to that offered in 1.3. The only exception is the inclusion of `impfree_recoded` and `ipeqopt_recoded` in the final two models. These variables are only contained in these models because they may be strongly multi-collinear with other variables, thereby inflating the standard errors of the coefficient estimates and increasing the risk of type II errors. To mitigate this risk, I estimate specifications with and without these two variables.

```r
socio_cons_year_fe1 <- feols(cons ~ gndr_dummy + agea + uemp5yr_factor | year, data = ess789_mod)
socio_cons_year_fe2 <- feols(cons ~ gndr_dummy + agea + uemp5yr_factor +
                              eummbr_factor + mnrchy_factor | year, data = ess789_mod)
socio_cons_year_fe3 <- feols(cons ~ gndr_dummy + agea + uemp5yr_factor +
                              eummbr_factor + mnrchy_factor + impfree_recoded | year, data = ess789_mod
socio_cons_year_fe4 <- feols(cons ~ gndr_dummy + agea + uemp5yr_factor +
                              eummbr_factor + mnrchy_factor + impfree_recoded + ipeqopt_recoded | year,

# coefficient plot
modelplot(list(socio_cons_year_fe1, socio_cons_year_fe2,
                socio_cons_year_fe3, socio_cons_year_fe4),
          coef_map = c("ipeqopt_recoded" = "Equal treatment/opportunity",
                       "impfree_recoded" = "Importance of freedom",
                       "mnrchy_factor1" = "Monarchy dummy",
                       "eummbr_factor1" = "EU dummy",
                       "uemp5yr_factor2" = "Unemployment dummy",
                       "agea" = "Age",
                       "gndr_dummy1" = "Gender dummy")) +
  geom_vline(xintercept = 0, linetype = "dashed") +
  labs(title = "Correlates of social conservatism at the country-year level",
       caption = "All models include year fixed effects.") +
  theme(legend.position = "bottom")
```



The coefficient plot shows that only gender is a significant predictor of social conservatism once year fixed effects are taken into account, with males being slightly more likely than females to be socially conservative.

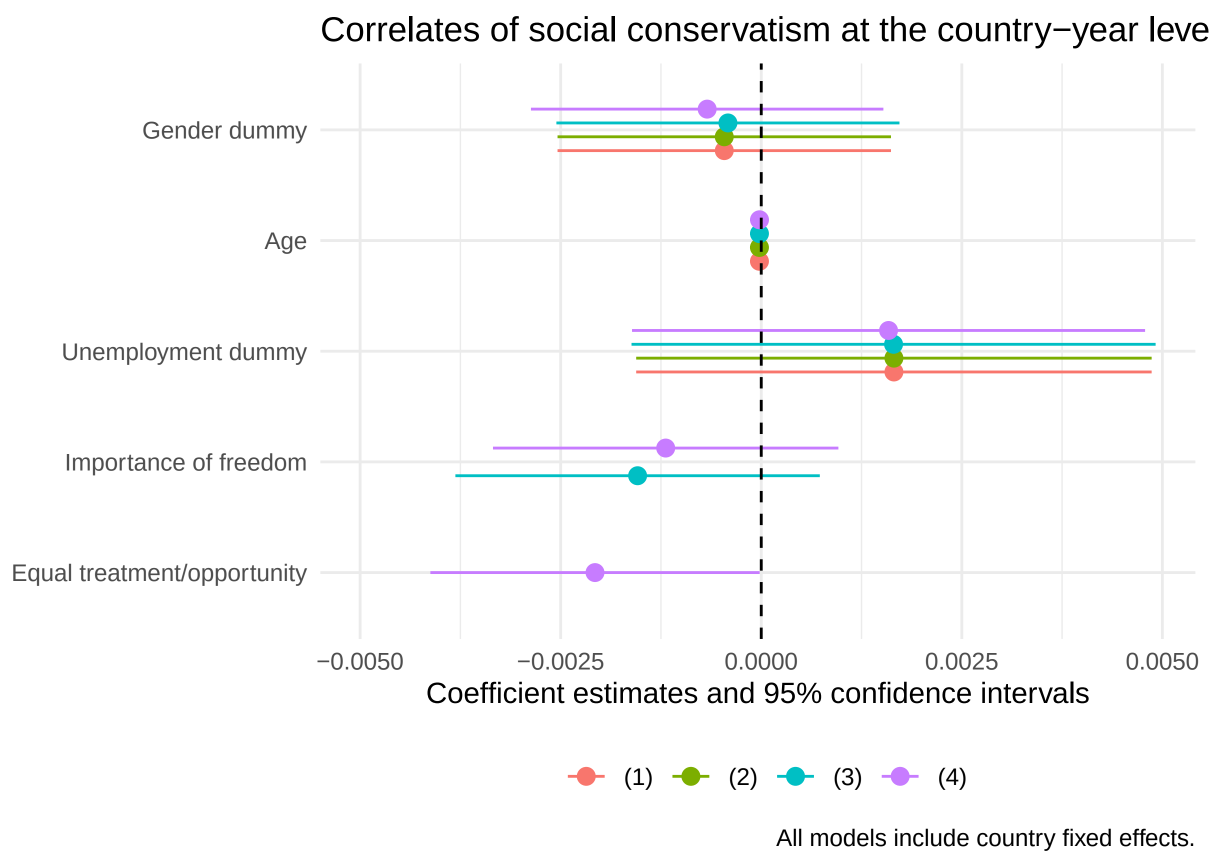
Re-estimate the model above using country-level fixed effects. (Hint: what class is the variable for country? Is it the most appropriate?) Plot the coefficients: What does it change with respect with the model with year fixed effects? Why?

The logic of the four models below is analogous to the previous exercise, save for country fixed effects replacing year fixed effects. As discussed above, country fixed effects net out all (un)observed, country-specific factors that are constant over time. To illustrate this, I have included `eummbr_factor` and `mnrchy_factor`, which are constant over time within countries. R automatically drops these variables since they are already accounted for by means of the country fixed effects, which is why they are not represented in the coefficient plot below.

```
socio_cons_cntry_fe1 <- feols(cons ~ gndr_dummy + agea + uemp5yr_factor | cntry, data = ess789_mod)
socio_cons_cntry_fe2 <- feols(cons ~ gndr_dummy + agea + uemp5yr_factor +
                              eummbr_factor + mnrchy_factor | cntry, data = ess789_mod)
socio_cons_cntry_fe3 <- feols(cons ~ gndr_dummy + agea + uemp5yr_factor +
                              eummbr_factor + mnrchy_factor + impfree_recoded | cntry, data = ess789_mo
socio_cons_cntry_fe4 <- feols(cons ~ gndr_dummy + agea + uemp5yr_factor +
                              eummbr_factor + mnrchy_factor + impfree_recoded + ipeqopt_recoded | cntry


# models
modelplot(list(socio_cons_cntry_fe1, socio_cons_cntry_fe2,
```

```
                socio_cons_cntry_fe3, socio_cons_cntry_fe4),
           coef_map = c("ipeqopt_recoded" = "Equal treatment/opportunity",
                        "impfree_recoded" = "Importance of freedom",
                        "mnrchy_factor1" = "Monarchy dummy",
                        "eummbr_factor1" = "EU dummy",
                        "uemp5yr_factor2" = "Unemployment dummy",
                        "agea" = "Age",
                        "gndr_dummy1" = "Gender dummy")) +
  geom_vline(xintercept = 0, linetype = "dashed") +
  expand_limits(x = -0.005) +
  labs(title = "Correlates of social conservatism at the country-year level",
       caption = "All models include country fixed effects.") +
  theme(legend.position = "bottom")
```



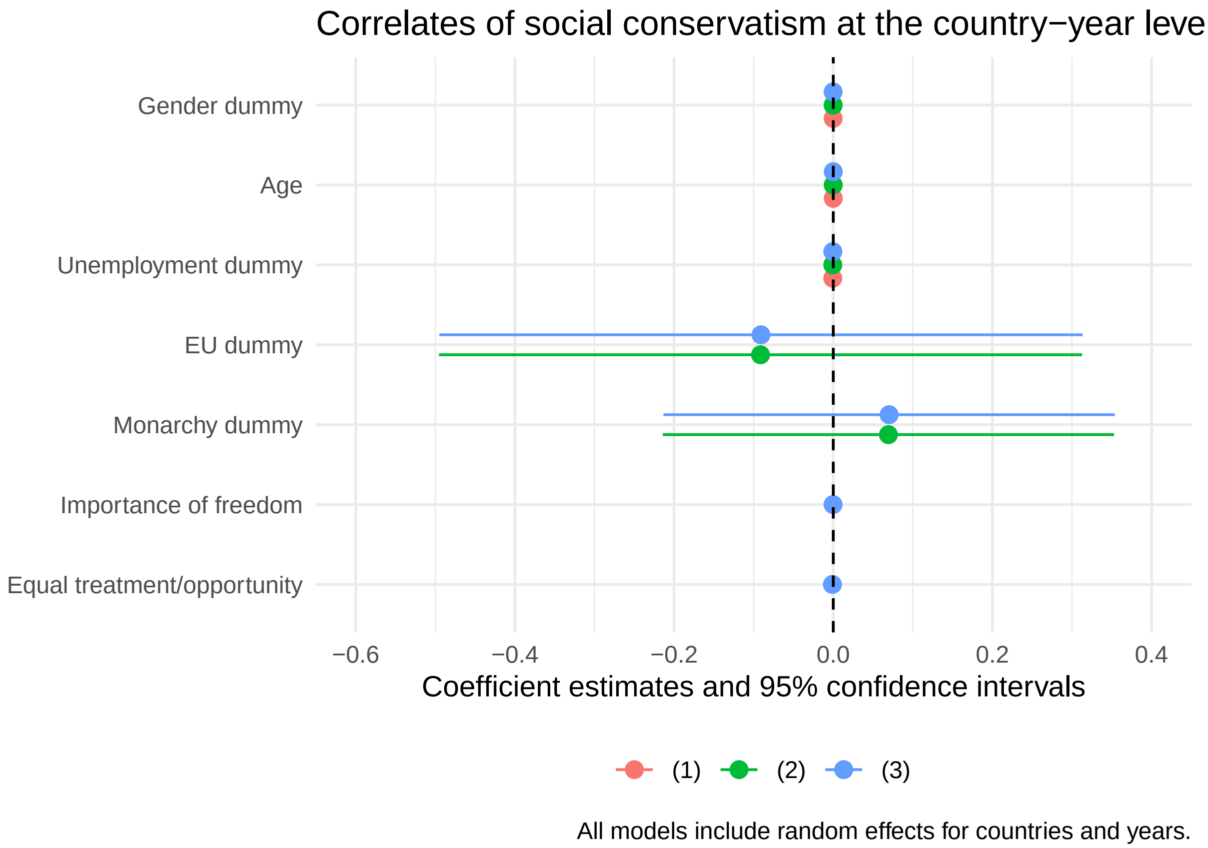
The coefficient plot demonstrates that, within a given country, respondents' belief in equality is significantly and negatively associated with social conservatism in that country in a given year, while all other covariates are insignificant.

2.3

Random Effects: estimate the model using random effects for years and country

To estimate the desired models, I estimate four specifications, where the logic underpinning the choice of covariates is analogous to the previous exercises. The only difference is that I use the `lmer()` to include random effects for years and country.

```
socio_cons_re1 <- lmer(cons ~ gndr_dummy + agea + uemp5yr_factor + (1 + essround | cntry),
                       data = ess789_mod,
                       control = lmerControl(optimizer = "nloptwrap"))
socio_cons_re2 <- lmer(cons ~ gndr_dummy + agea + uemp5yr_factor + eummbr_factor + mnrchy_factor +
                         (1 + essround | cntry),
                       control = lmerControl(optimizer = "nloptwrap"),
                       data = ess789_mod)
socio_cons_re3 <- lmer(cons ~ gndr_dummy + agea + uemp5yr_factor + eummbr_factor + mnrchy_factor +
                        impfree_recoded + ipeqopt_recoded + (1 + essround | cntry),
                       control = lmerControl(optimizer = "nloptwrap"),
                       data = ess789_mod)
# modelsummary
modelplot(list(socio_cons_re1, socio_cons_re2, socio_cons_re3),
          coef_map = c("ipeqopt_recoded" = "Equal treatment/opportunity",
                       "impfree_recoded" = "Importance of freedom",
                       "mnrchy_factor1" = "Monarchy dummy",
                       "eummbr_factor1" = "EU dummy",
                       "uemp5yr_factor2" = "Unemployment dummy",
                       "agea" = "Age",
                       "gndr_dummy1" = "Gender dummy")) +
  geom_vline(xintercept = 0, linetype = "dashed") +
  expand_limits(x = c(-0.6, 0.4)) +
  labs(title = "Correlates of social conservatism at the country-year level",
       caption = "All models include random effects for countries and years.") +
  theme(legend.position = "bottom")
```



- Interpretatio.

2.4 Some Theory

What do fixed effects account for? Specify: for years and countries/geographic regions, What do random effects account for?, Following Schimdt-Catran and Fairbrother, illustrate the structure of fixed effects.

References

Anduiza, Eva, and Guillem Rico. 2022. "Sexism and the Far-Right Vote: The Individual Dynamics of Gender Backlash." *American Journal of Political Science*.

Inglehart, Ronald, and Christian Welzel. 2010. "Changing Mass Priorities: The Link Between Modernization and Democracy." *Perspectives on Politics* 8 (2): 551–67.

Oshri, Odelia, Liran Harsgor, Reut Itzkovitch-Malka, and Or Tuttnauer. 2022. "Risk Aversion and the Gender Gap in the Vote for Populist Radical Right Parties." *American Journal of Political Science*.