03 Main Analysis

Nate Breznau

2022-04-15

```
df_orig_complete <- read_csv(file = here::here("data", "df_orig_complete.csv"))
```

```
Tbl5_m1 <- polr(factor(stdliv_next5) ~ factor(cat_age) + factor(wave) + factor(cat_age)*factor
(wave), data = df_orig_complete, Hess = T)

Tbl5_m2 <- polr(factor(stdliv_next5) ~ factor(cat_age) + factor(wave) + factor(cat_age)*factor
(wave) + female + factor(education) + factor(EGP6) + factor(income), data = df_orig_complete,
Hess = T)

Tbl5_m3 <- polr(factor(stdliv_next5) ~ factor(cat_age) + factor(wave) + factor(cat_age)*factor
(wave) + female + factor(education) + factor(EGP6) + factor(income) + pensions + unemployed +
car_owner + factor(cntry), data = df_orig_complete, Hess = T)

tab_model(Tbl5_m1, Tbl5_m2, Tbl5_m3, transform = NULL, p.style = "stars", show.ci = F, show.lo
glik = T,  pred.labels = c("k1","k2","k3","k4","Age 30-44","Age 44-59","Age >60","Year 2007","
Age 30-44*year'07", "Age 44-59*year'07", "Age >60*year'07", "Female", "Educ mid", "Educ high"
,"EGP: routine non-man", "EGP: Self", "EGP: Skilled", "EGP: Unskilled", "EGP: Farmers", "EGP:
Never had a job", "Income mid", "Income high","Income missing", "Pensions and benefits","Unemp
loyed", "Car","bulgaria", "czech", "estonia", "hungary","latvia", "lithuania", "moldova","pola
nd", "romania", "russia","slovakia", "ukraine"), file = here::here("results", "Tbl5_rep_f.doc"
))
```

	factor(stdliv next 5)	factor(stdliv next 5)	factor(stdliv next 5)
Predictors	*Log-Odds*	*Log-Odds*	*Log-Odds*
k1	-2.51 ***	-2.45 ***	-2.21 ***
k2	-1.61 ***	-1.54 ***	-1.29 ***
k3	0.59 ***	0.68 ***	0.99 ***
k4	3.18 ***	3.30 ***	3.64 ***
Age 30-44	-0.41 ***	-0.46 ***	-0.46 ***
Age 44-59	-0.81 ***	-0.82 ***	-0.82 ***
Age >60	-0.97 ***	-0.82 ***	-0.75 ***
Year 2007	0.75 ***	0.73 ***	0.75 ***
Age 30-44*year'07	0.00	0.02	-0.02
Age 44-59*year'07	-0.11	-0.10	-0.11
Age >60*year'07	-0.38 ***	-0.40 ***	-0.39 ***

Female		-0.15 ***	-0.10 ***
Educ mid		0.08 **	0.02
Educ high		0.21 ***	0.17 ***
EGP: routine non-man		-0.08 *	-0.06
EGP: Self		0.22 ***	0.26 ***
EGP: Skilled		-0.16 ***	-0.14 ***
EGP: Unskilled		-0.17 ***	-0.13 ***
EGP: Farmers		-0.18 ***	-0.14 ***
EGP: Never had a job		-0.32 ***	-0.19 ***
Income mid		0.14 ***	0.11 ***
Income high		0.41 ***	0.36 ***
Income missing		0.21 *	0.18
Pensions and benefits			-0.09 **
Unemployed			-0.20 ***
Car			0.27 ***
bulgaria			0.02
czech			0.33 ***
estonia			0.66 ***
hungary			-0.24 ***
latvia			0.72 ***
lithuania			0.58 ***
moldova			-0.27 ***
poland			0.27 ***
romania			0.40 ***
russia			0.15 **
slovakia			0.00
ukraine			-0.12 *
Observations	35648	35648	35648
R^2 Nagelkerke	0.068	0.086	0.120

log-Likelihood	-45584.674	-45258.680	-44644.165

- *p<0.05 ** p<0.01 *** p<0.001*

```
tab_model(Tbl5_m1, Tbl5_m2, Tbl5_m3, transform = NULL, p.style = "stars", show.ci = F, show.loglik = T,  pred.labels = c("k1","k2","k3","k4","Age 30-44","Age 44-59","Age >60","Year 2007","Age 30-44*year'07", "Age 44-59*year'07", "Age >60*year'07", "Female", "Educ mid", "Educ high","EGP: routine non-man", "EGP: Self", "EGP: Skilled", "EGP: Unskilled", "EGP: Farmers", "EGP: Never had a job", "Income mid", "Income high","Income missing", "Pensions and benefits","Unemployed", "Car","bulgaria", "czech", "estonia", "hungary","latvia", "lithuania", "moldova","poland", "romania", "russia","slovakia", "ukraine"), file = here::here("results", "Tbl5_rep_f.htm"))
```

	factor(stdliv next 5)	**factor(stdliv next 5)**	**factor(stdliv next 5)**
Predictors	*Log-Odds*	*Log-Odds*	*Log-Odds*
k1	-2.51 ***	-2.45 ***	-2.21 ***
k2	-1.61 ***	-1.54 ***	-1.29 ***
k3	0.59 ***	0.68 ***	0.99 ***
k4	3.18 ***	3.30 ***	3.64 ***
Age 30-44	-0.41 ***	-0.46 ***	-0.46 ***
Age 44-59	-0.81 ***	-0.82 ***	-0.82 ***
Age >60	-0.97 ***	-0.82 ***	-0.75 ***
Year 2007	0.75 ***	0.73 ***	0.75 ***
Age 30-44*year’07	0.00	0.02	-0.02
Age 44-59*year’07	-0.11	-0.10	-0.11
Age >60*year’07	-0.38 ***	-0.40 ***	-0.39 ***
Female		-0.15 ***	-0.10 ***
Educ mid		0.08 **	0.02
Educ high		0.21 ***	0.17 ***
EGP: routine non-man		-0.08 *	-0.06
EGP: Self		0.22 ***	0.26 ***
EGP: Skilled		-0.16 ***	-0.14 ***
EGP: Unskilled		-0.17 ***	-0.13 ***
EGP: Farmers		-0.18 ***	-0.14 ***
EGP: Never had a job		-0.32 ***	-0.19 ***

Income mid		0.14 ***	0.11 ***
Income high		0.41 ***	0.36 ***
Income missing		0.21 *	0.18
Pensions and benefits			-0.09 **
Unemployed			-0.20 ***
Car			0.27 ***
bulgaria			0.02
czech			0.33 ***
estonia			0.66 ***
hungary			-0.24 ***
latvia			0.72 ***
lithuania			0.58 ***
moldova			-0.27 ***
poland			0.27 ***
romania			0.40 ***
russia			0.15 **
slovakia			0.00
ukraine			-0.12 *
Observations	35648	35648	35648
R^2 Nagelkerke	0.068	0.086	0.120
log-Likelihood	-45584.674	-45258.680	-44644.165

- *p<0.05 ** p<0.01 *** p<0.001*

```
#knitr::include_graphics(here::here("results", "Tbl5_rep_f.htm"))

#webshot(here::here("results", "Tbl5_rep_f.htm"), file = here::here("results", "Tbl5_rep_f.png
"))

#knitr::include_graphics(here::here("results", "Tbl5_rep_f.png"))
```

Marginal Estimates

The coefficients in the table are different from the original results, but they are also different from the Stata results (see 03_Main_Analysis_Stata.do). Clearly the ordred probit models use slightly different estimation techniques. But, the

quesiton is: are the predicted values the same?

Calculate Margins

```
newdat <- data.frame(
  cat_age = rep(1:4, 2),
  wave = c(rep(1993, 4), rep(2007, 4)),
  female = rep(0, 8),
  education = rep(1,8),
  EGP6 = rep(1,8),
  income = rep(1,8),
  pensions = rep(mean(df_orig_complete$pensions, na.rm = T),8),
  unemployed = rep(mean(df_orig_complete$unemployed, na.rm = T),8),
  car_owner = rep(mean(df_orig_complete$car_owner, na.rm = T),8),
  cntry = "belarus") # ref cat

newdat <- cbind(newdat, predict(Tbl5_m3, newdat, type = "probs"))
```

Import Stata Results

```
Tbl5_m3_Stata <- read_csv(here::here("results","Tbl5_m3.csv"))
```

```
## Rows: 40 Columns: 5
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## dbl (5): outcome, age_cat, wave, margin, se
##
## 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.
```

Combine into Margins Comparison Table

```
Tbl5_marg_comp <- newdat %>%
  dplyr::select(wave, cat_age, `1`, `2`, `3`, `4`, `5`)

colnames(Tbl5_marg_comp) <- c("wave", "age_cat", "fall_alot", "fall", "stay_same", "rise", "ri
se_alot")

# add Stata results
Tbl5_marg_comp[9:16,] <- NA
Tbl5_marg_comp$Software <- c(rep("R polr",8), rep("Stata oprobit", 8))
Tbl5_marg_comp[9:16,1:2] <- Tbl5_marg_comp[1:8,1:2]

#sort Stata data
Tbl5_m3_Stata <- Tbl5_m3_Stata[order( Tbl5_m3_Stata[,"outcome"], Tbl5_m3_Stata[,"wave"] ),]

Tbl5_marg_comp$fall_alot[9:16] <- Tbl5_m3_Stata$margin[1:8]
Tbl5_marg_comp$fall[9:16] <- Tbl5_m3_Stata$margin[9:16]
Tbl5_marg_comp$stay_same[9:16] <- Tbl5_m3_Stata$margin[17:24]
Tbl5_marg_comp$rise[9:16] <- Tbl5_m3_Stata$margin[25:32]
```

```
Tbl5_marg_comp$rise_alot[9:16] <- Tbl5_m3_Stata$margin[33:40]
```

Comparison

```
Tbl5_marg_comp %>%
  subset(age_cat == 1 | age_cat == 4) %>%
  ggplot(aes(y = (fall_alot+fall), x = wave, color = interaction(Software,age_cat))) +
  geom_point() +
  geom_line() +
  labs(y = "Predicted Likelihood of Percieved Standard\nof Living Falling in the Next 5 Years"
, x = "Survey Wave") +
  scale_color_manual(name = "Age group &\nSoftware used",
                     values = c("#481567FF",
                                "#238A8DFF",
                                "#453781FF",
                                "#29AF7FFF"),
                     labels = c("Age 19-29\nusing R polr", "Age 19-29\nusing Stata oprobit",
 "Age 60+\nusing R polr", "Age 60+\nusing Stata oprobit")) +
  scale_x_discrete(limits = c(1993,2007)) +
  theme_classic()
```

```
## Warning: Continuous limits supplied to discrete scale.
## Did you mean `limits = factor(...)` or `scale_*_continuous()`?
```

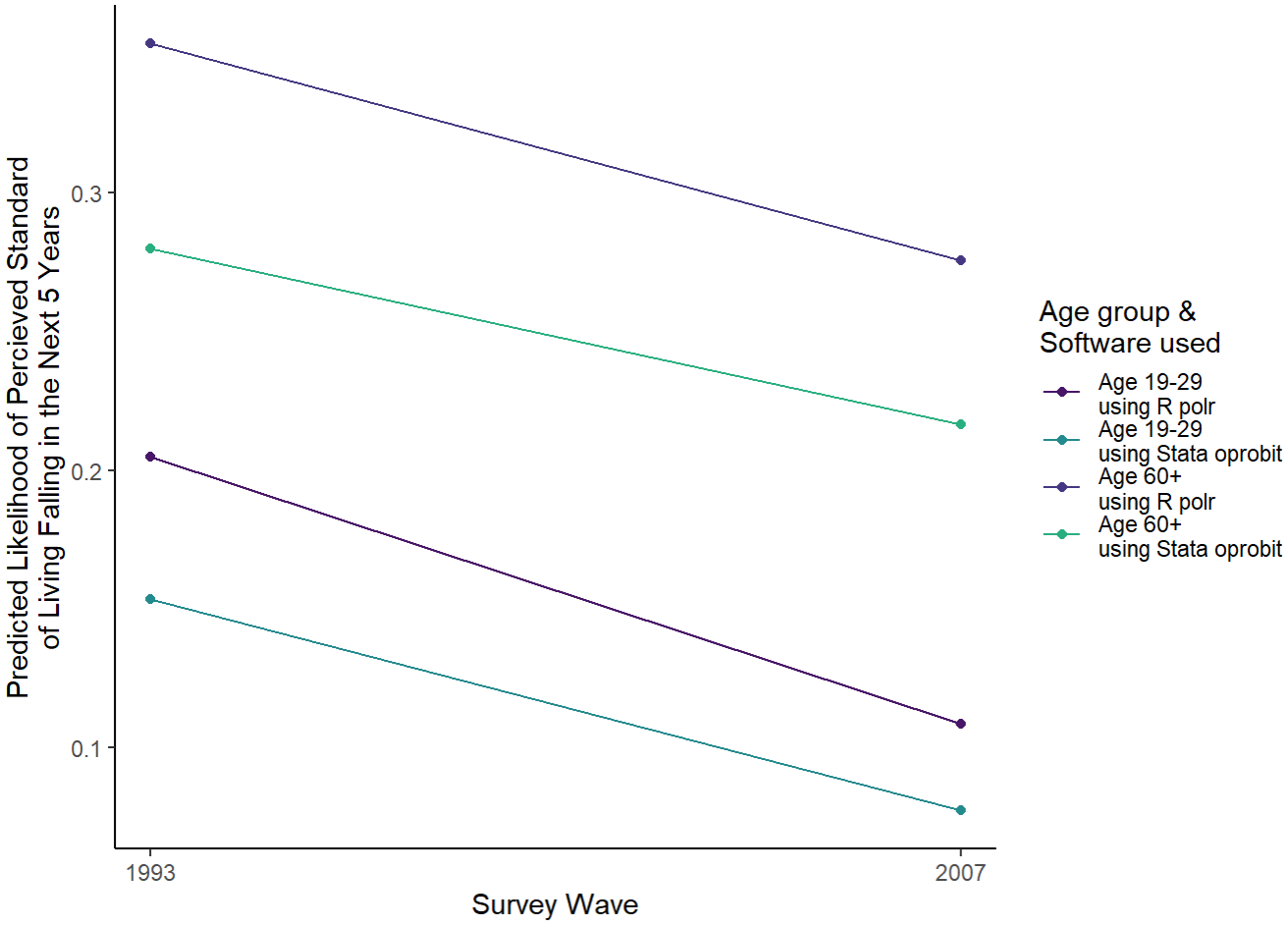

Table Comparing Stata and R Margins

```
kable_styling(kable(Tbl5_marg_comp))
```

wave	age_cat	fall_alot	fall	stay_same	rise	rise_alot	Software
1993	1	0.0927751	0.1120754	0.5111955	0.2567204	0.0272337	R polr
1993	2	0.1396044	0.1505535	0.5098853	0.1826181	0.0173387	R polr
1993	3	0.1888147	0.1808238	0.4819878	0.1362233	0.0121504	R polr
1993	4	0.1784101	0.1752032	0.4890246	0.1443496	0.0130125	R polr
2007	1	0.0460742	0.0624050	0.4351127	0.4004501	0.0559580	R polr
2007	2	0.0722499	0.0917628	0.4935636	0.3069644	0.0354592	R polr
2007	3	0.1088939	0.1264952	0.5154429	0.2262760	0.0228920	R polr
2007	4	0.1311603	0.1443632	0.5127259	0.1931386	0.0186120	R polr
1993	1	0.0626676	0.0909308	0.4790867	0.3246091	0.0427057	Stata oprobit
1993	2	0.1054630	0.1243734	0.5027682	0.2447451	0.0226503	Stata oprobit
1993	3	0.1464445	0.1480962	0.4993305	0.1922632	0.0138655	Stata oprobit
1993	4	0.1366441	0.1430050	0.5015386	0.2033232	0.0154891	Stata oprobit
2007	1	0.0265037	0.0508320	0.3974964	0.4314184	0.0937495	Stata oprobit
2007	2	0.0479969	0.0764672	0.4575851	0.3618393	0.0561114	Stata oprobit
2007	3	0.0808522	0.1064922	0.4943244	0.2864267	0.0319044	Stata oprobit
2007	4	0.0974332	0.1188876	0.5011741	0.2573054	0.0251997	Stata oprobit

Colophon

```
sessionInfo()
```

```
## R version 4.1.3 (2022-03-10)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 19043)
##
## Matrix products: default
##
## locale:
```

```
## [1] LC_COLLATE=English_United States.1252
## [2] LC_CTYPE=English_United States.1252
## [3] LC_MONETARY=English_United States.1252
## [4] LC_NUMERIC=C
## [5] LC_TIME=English_United States.1252
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
##  [1] knitr_1.38       webshot_0.5.2    sjPlot_2.8.10    MASS_7.3-55
##  [5] kableExtra_1.3.4 forcats_0.5.1    stringr_1.4.0    dplyr_1.0.8
##  [9] purrr_0.3.4      readr_2.1.2      tidyr_1.2.0      tibble_3.1.6
## [13] ggplot2_3.3.5    tidyverse_1.3.1
##
## loaded via a namespace (and not attached):
##  [1] nlme_3.1-155       fs_1.5.2           bit64_4.0.5        lubridate_1.8.0
##  [5] insight_0.17.0     httr_1.4.2         rprojroot_2.0.2    tools_4.1.3
##  [9] backports_1.4.1    bslib_0.3.1        utf8_1.2.2         R6_2.5.1
## [13] sjlabelled_1.1.8   DBI_1.1.2          colorspace_2.0-3   withr_2.5.0
## [17] tidyselect_1.1.2   emmeans_1.7.3      bit_4.0.4          compiler_4.1.3
## [21] performance_0.9.0  cli_3.2.0          rvest_1.0.2        pacman_0.5.1
## [25] xml2_1.3.3         labeling_0.4.2     bayestestR_0.11.5  sass_0.4.1
## [29] scales_1.1.1       mvtnorm_1.1-3      systemfonts_1.0.4  digest_0.6.29
## [33] minqa_1.2.4        rmarkdown_2.13     svglite_2.1.0      pkgconfig_2.0.3
## [37] htmltools_0.5.2    lme4_1.1-28        highr_0.9          dbplyr_2.1.1
## [41] fastmap_1.1.0      rlang_1.0.2        readxl_1.4.0       rstudioapi_0.13
## [45] farver_2.1.0       jquerylib_0.1.4    generics_0.1.2     jsonlite_1.8.0
## [49] vroom_1.5.7        magrittr_2.0.2     parameters_0.17.0  Matrix_1.4-0
## [53] Rcpp_1.0.8.3       munsell_0.5.0      fansi_1.0.3        lifecycle_1.0.1
## [57] stringi_1.7.6      yaml_2.3.5         snakecase_0.11.0   grid_4.1.3
## [61] parallel_4.1.3     sjmisc_2.8.9       crayon_1.5.1       lattice_0.20-45
## [65] ggeffects_1.1.1    haven_2.4.3        splines_4.1.3      sjstats_0.18.1
## [69] hms_1.1.1          pillar_1.7.0       boot_1.3-28        estimability_1.3
## [73] effectsize_0.6.0.1 reprex_2.0.1       glue_1.6.2         evaluate_0.15
## [77] modelr_0.1.8       vctrs_0.3.8        nloptr_2.0.0       tzdb_0.3.0
## [81] cellranger_1.1.0   gtable_0.3.0       assertthat_0.2.1   datawizard_0.4.0
## [85] xfun_0.30          xtable_1.8-4       broom_0.7.12       viridisLite_0.4.0
## [89] ellipsis_0.3.2     here_1.0.1
```