

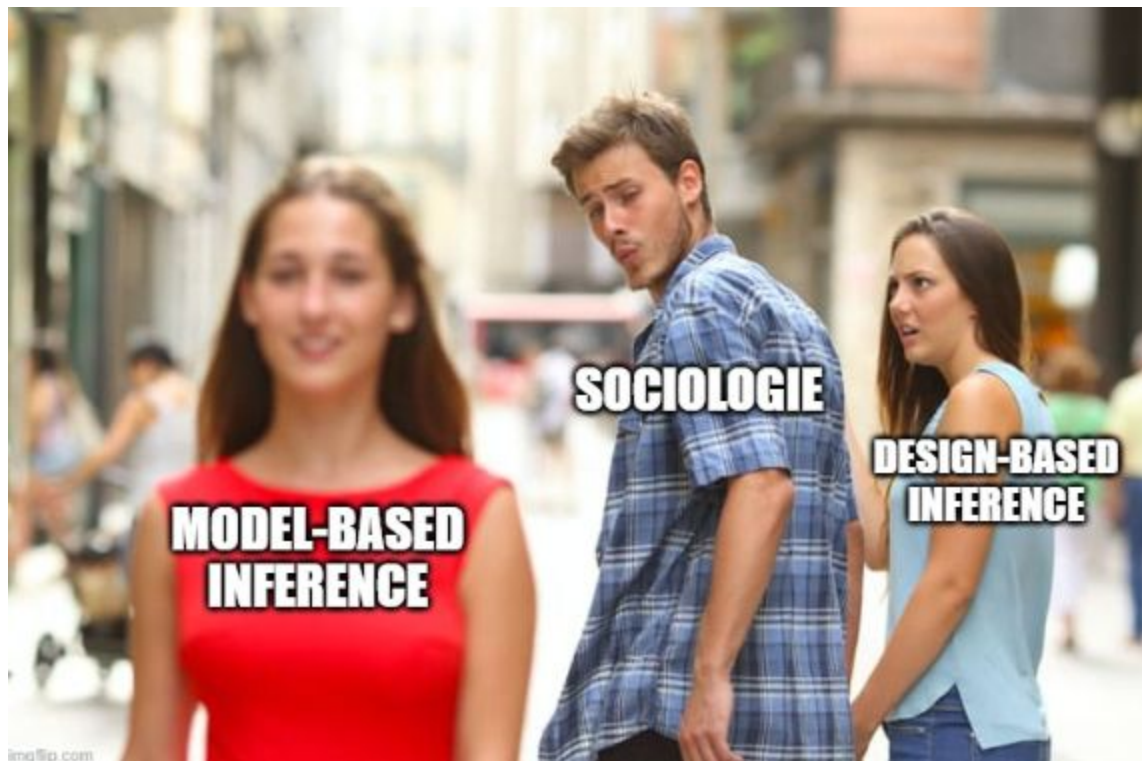
Odhad výsledků voleb z nereprezentativních dat: Víceúrovňová regrese s post-stratifikací

Předvolební výzkumy: Metodická konference SIMAR
1. června 2022
Aleš Vomáčka, Jaromír Mazák, Matouš Pilnáček



FILOZOFICKÁ FAKULTA
Univerzita Karlova





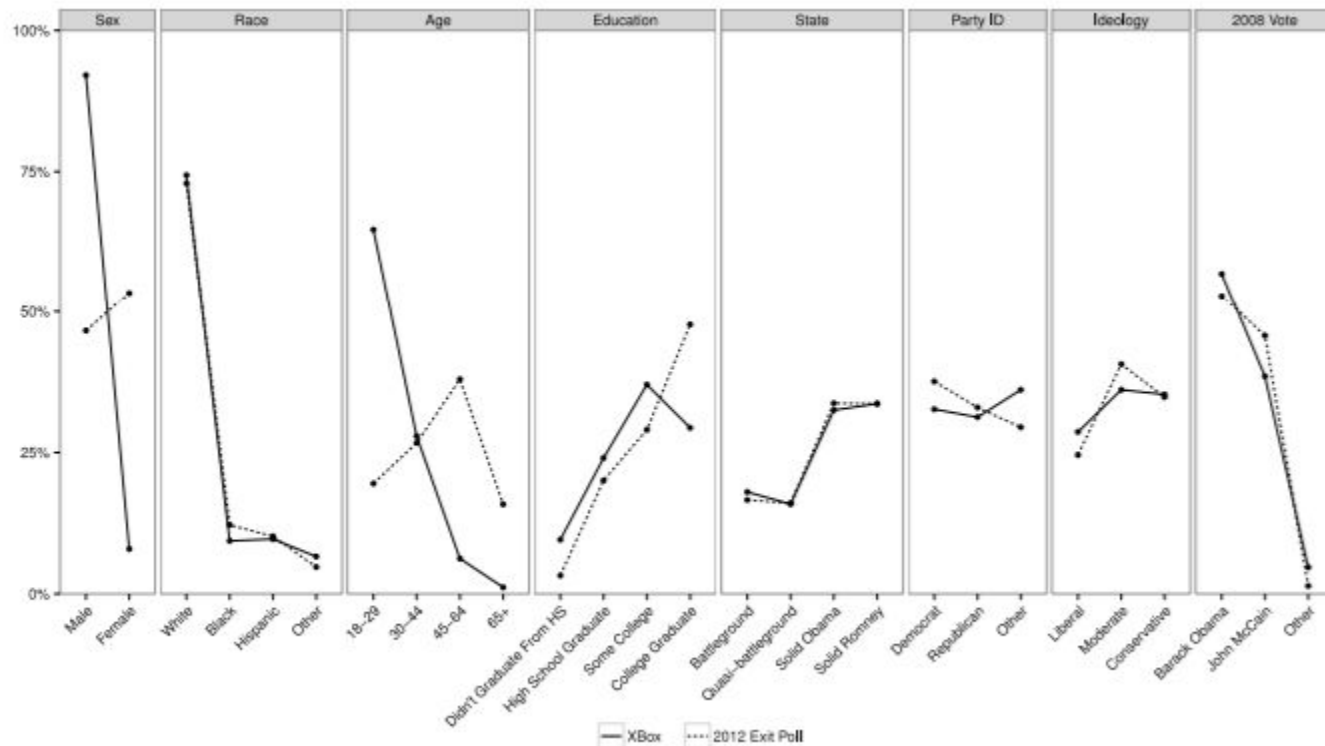
**MODEL-BASED
INFERENCE**

SOCIOLOGIE

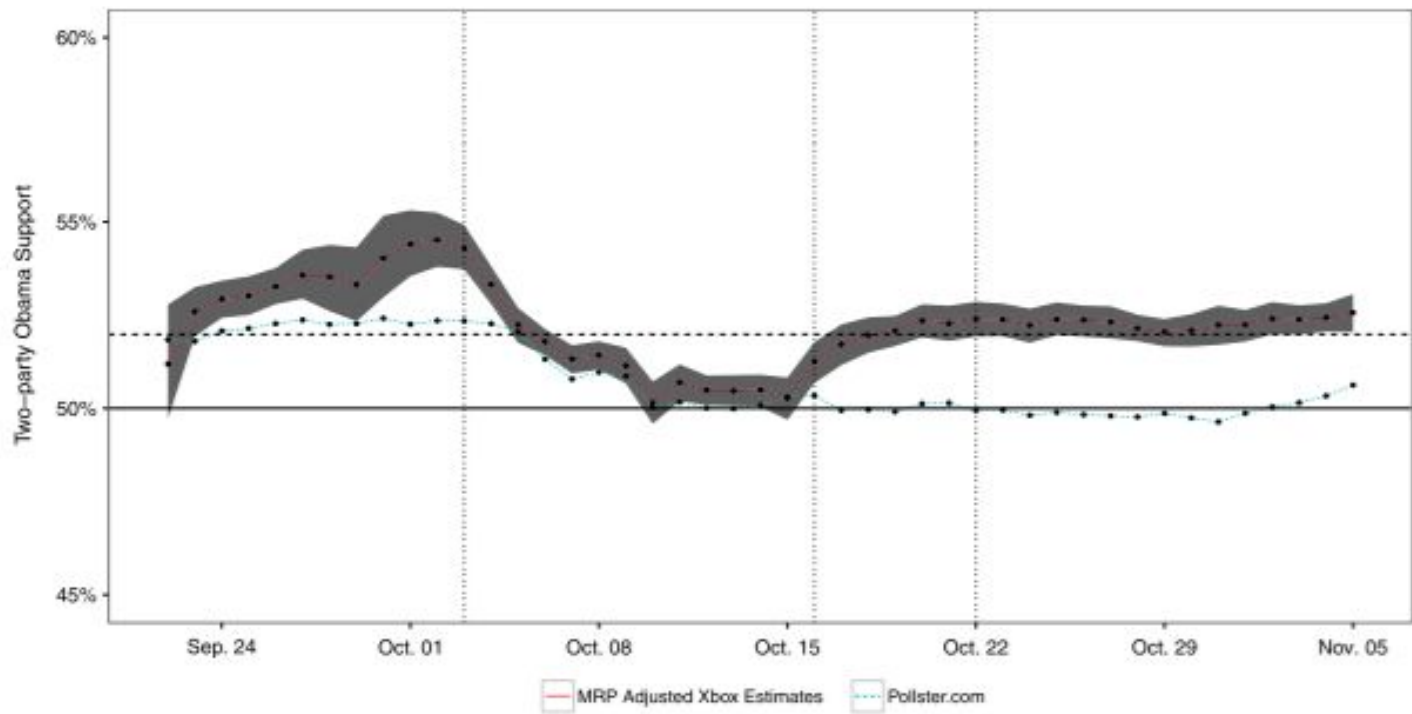
**DESIGN-BASED
INFERENCE**



Nereprezentativita mezi uživateli X-box



Míra přesnosti predikce na základě dat X-box



Vždycky to ale nefunguje tak skvěle...

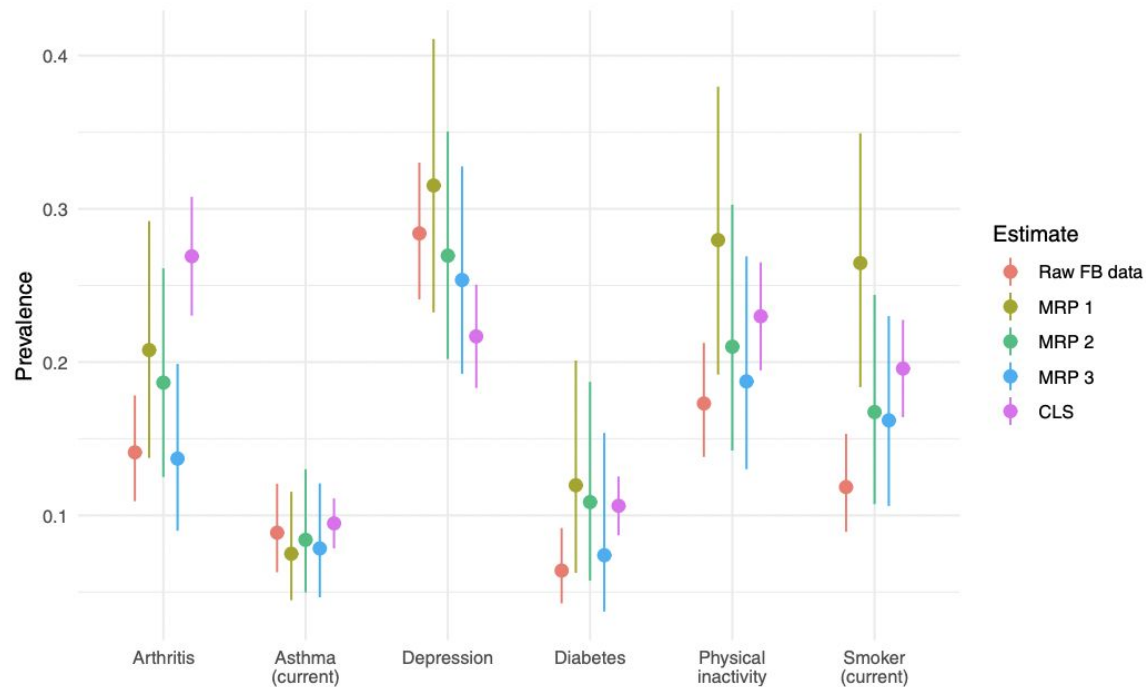


Fig. 2. Estimates of health outcomes with 95% credible intervals. FB: Facebook.

MRP - základní principy

GRÓ: predikce volebních preferencí pro velký počet subpopulací, které se potom agregují.

Kolik? V našem případě 48 356 subpopulací.

Proměnné:

- Pohlaví (2)
- Vzdělání (4)
- Věk (7)
- Okres bydliště (77)
- Minulá volba (12)
- Agregované pro okresy: nezaměstnanost, přírůstek, podíl věřících, naděje na dožití, podíl v exekuci, trestné činy

Víceúrovňová bayesovská regrese s post-stratifikací

1. Víceúrovňová: Pomáhá zpřesnit predikce pro velmi malé subpopulace (Partial pooling)
2. Bayesovská: simulační přístup, větší stabilita než MLE
3. Post-stratifikace: Výsledné pravděpodobnosti vážíme počtem voličů v každé subpopulaci

Sběr dat

Média: Seznam.cz, Deník N, Novinky.cz (n = ca 140 tisíc)

Sociální sítě (Facebook) (n = ca 2000)

Český národní panel - náhodný výběr (n = ca 3 000)

Český národní panel - kvótní výběr (n = ca 1 000)

Partner sběru dat:

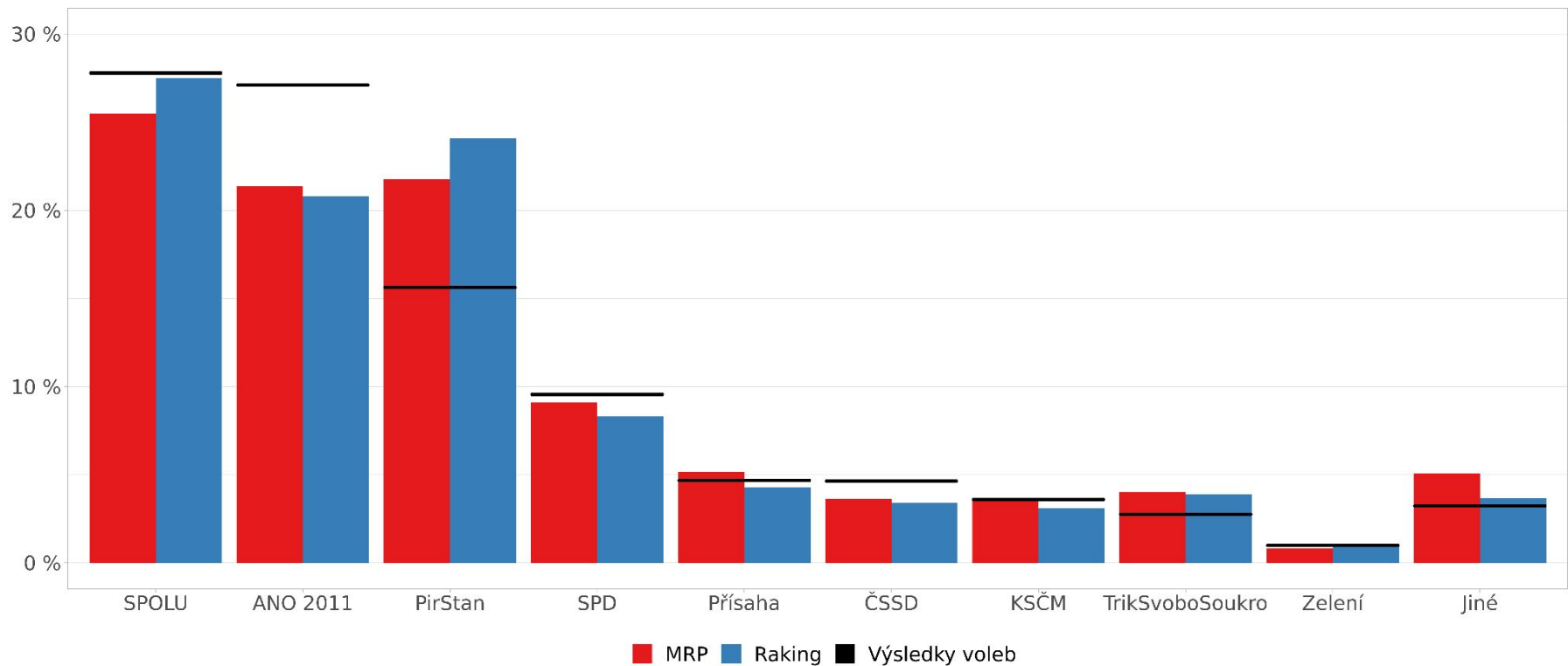


Jak jsme si vedli? Mediální anketa



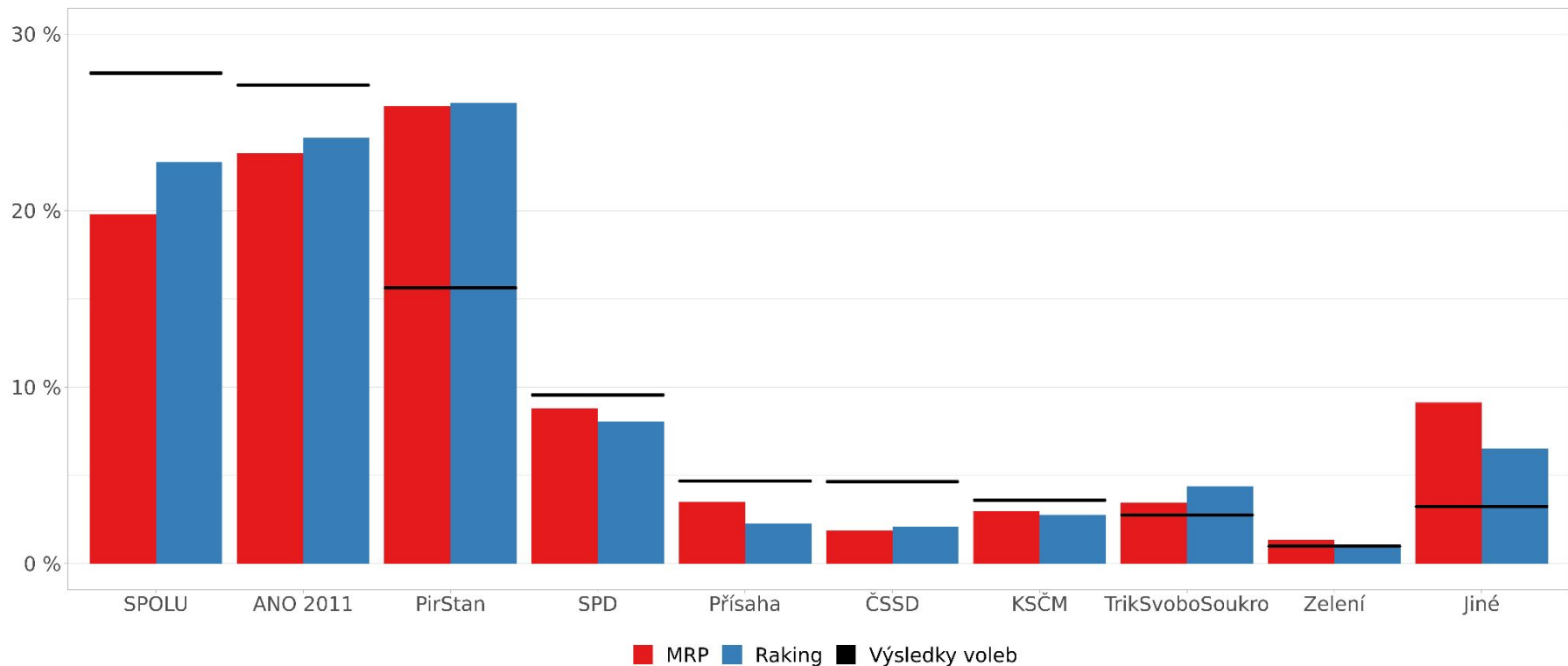
Počet respondentů = 40 000

Mediální anketa: Porovnání s vážením (raking)



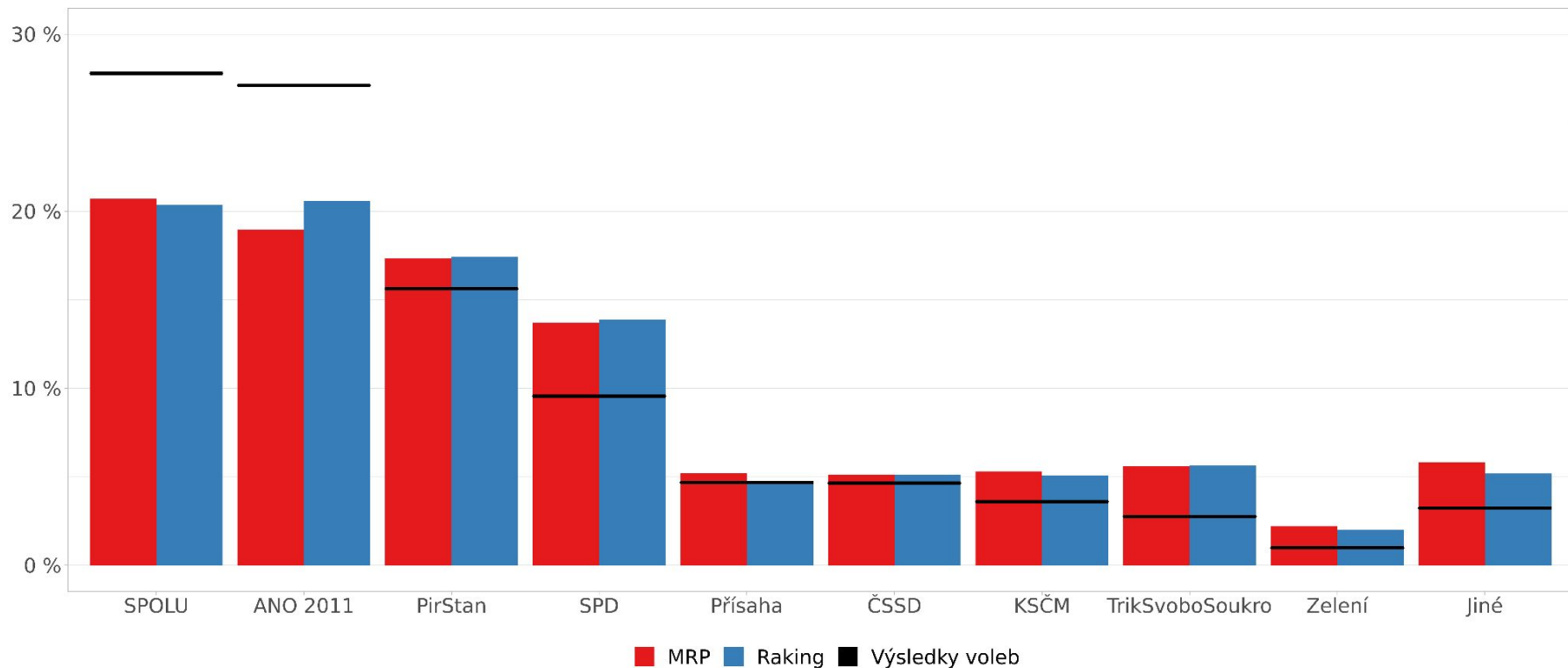
Počet respondentů = 40 000

Facebook: Porovnání s vážením



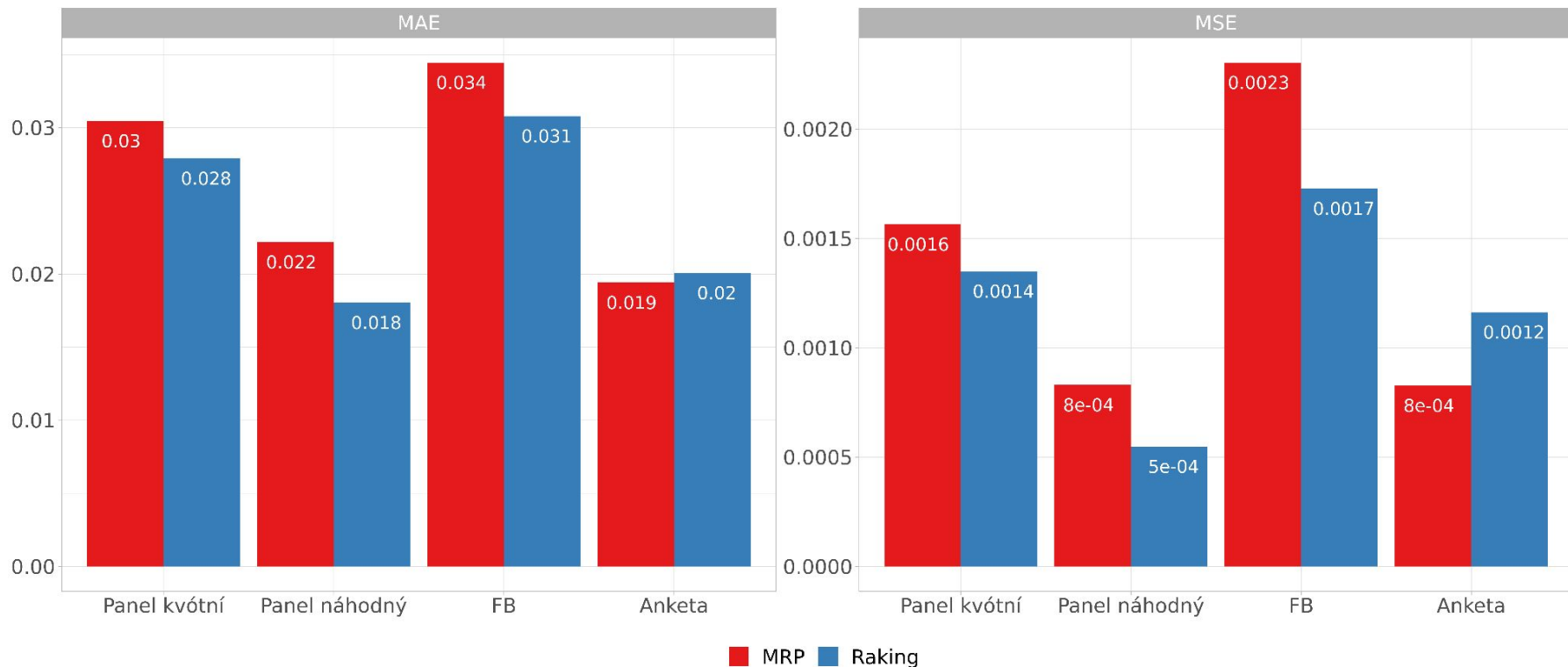
Počet respondentů = 1 861

ČNP - kvótní výběr: Porovnání s vážením



Počet respondentů = 885

Celkový verdikt?



Počty respondentů: Panel kvótní = 885, Panel náhodný = 2 707, FB = 1 861, Anketa = 40 000

Diskuse

- Dramaticky nadreprezentovaní lidé se zájmem o politiku (doplnit o dotazování “měkkých” charakteristik?)
- Dvě ambice: (1) odhad o populaci z nereprezentativních dat, (2) možnost přibližných odhadů pro malé územní celky (vs. velké online panely).



FILOZOFICKÁ FAKULTA
Univerzita Karlova



Odhad výsledků voleb z nereprezentativních dat: Víceúrovňová regrese s post-stratifikací

Aleš Vomáčka, Jaromír Mazák, Matouš Pilnáček
(ales@vomacka.io; jaromir.mazak@ff.cuni.cz;
matous.pilnacek@soc.cas.cz)

