Things That Could Cause Weird Estimator Behavior, Particularly When Analyzing by Country

Chris Felton

January 16, 2025

My guess is you know about many or all of these (for instance, I see you ignore recruitment mode in imputation), but I figure you might have missed some and it would be good to have everything in one place...

General Issues

- `POLITICAL_ID` is being treated as numeric but I'm not sure it should be
- `PSU_W1` being treated as numeric, not sure it should be
- `DRINKS_W1` being treated as numeric, not sure whether it should be. Is it actually top-coded at 97 or is that a code for does not apply?... perhaps just a highly skewed distribution
- `CIGARETTES` is often very skewed, and the max is often 96 or 97—are these categories? Because it's currently treated as numeric
- Same with number of children and number in household... I break it down by country but basically I'm not sure whether it's supposed to be numeric or what. But the current extreme values may be causing some problems...
- Often `REL#_W1` is just all `NA`s for a given country and `#`
- Same with `TEACHINGS_#_W1`
- `INCOME_FEELINGS` is supposed to be binary (according to the spreadsheet) but it's 1–4 categorical
- Same with Life Balance
- All binary variables appear to be coded 1–2 instead of 0–1
- `COUNTRY_REL` vars are often all `NA`s for a particular country

Country == 1 (Argentina)

- `NUM_CHILDREN` crazy outliers
- `NUM_HOUSEHOLD` outliers

Country == 2 (Australia)

- Only one recruitment mode

Country == 3 (Brazil)

- `NUM_CHILDREN` crazy outliers... is it supposed to be numeric?
- `NUM_HOUSEHOLD` outliers... is it supposed to be numeric?

Country == 4 (Egypt)

- Only one recruitment mode
- **No** crazy outliers for children and household
- Little variation in political ideology

Country == 5 (Germany)

- `NUM_CHILDREN` crazy outliers... is it supposed to be numeric?
- `NUM_HOUSEHOLD` outliers... is it supposed to be numeric?

Country == 6 (India)

- Only one recruitment mode
- `NUM_CHILDREN` crazy outliers... is it supposed to be numeric?
- `POLITICAL_ID` all `NA`s
- **No** crazy outliers for `NUM_HOUSEHOLD`...

Country == 7 (Indonesia)

- `NUM_CHILDREN` crazy outliers... is it supposed to be numeric?
- `NUM_HOUSEHOLD` outliers... is it supposed to be numeric?

Country == 8 (Israel)

- Only one recruitment mode
- **No** crazy outliers in household or children

Country == 9 (Japan)

- Crazy outliers in household and children
- Little variation in political ideology

Country == 10 (Kenya)

- Only one recruitment mode
- **No** crazy outliers in children and household

Country == 11 (Mexico)

- Kinda crazy outlier in household, none for children

Country == 12 (Nigeria)

- Only one recruitment method
- **No** crazy outliers in household or children

Country == 13 (Philippines)

- Lots of `TEACHING` vars are all 10s
- `CIGARETTES` max is only 60, whereas it's usually 96 or 97...
- Number of children outlier isn't too crazy, same with household

Country == 14 (Poland)

- Lots of non-variation in different teaching questions
- **No** crazy outliers in children or household

Country == 15 (Nothing)

- What is the story behind this lol

Country == 16 (South Africa)

- No crazy outliers in children or household
- Cigarettes max at 60 rather than 96 or 97

Country == 17 (Spain)

- Household outlier, children slight outlier
- Cigarettes max 75

Country == 18 (Tanzania)

- **No** crazy outliers for children or household

Country == 19 (Turkey)

- Drinks max at 80
- **No** crazy outliers for children or household

Country == 20 (UK)

- Cigarettes max at 70
- No crazy outlier for children
- Crazy outlier for household

Country == 21 (Nothing)

- ?

Country == 22 (USA)

- Huge outliers in sampling weight
- Only one recruitment mode
- Huge outliers in children and household, not the usual max for children though (90)

Country == 23 (Sweden)

- Only one recruitment mode
- Crazy outlier in household but not the usual (74)
- No crazy outlier for children

Country == 24 (Hong Kong)

- Only one recruitment mode
- Not the usual max for drinks (56)
- Not the usual max for cigarettes (30)
- No crazy outliers for children or household