Time Mixed Effects Model

Aissata Bah   Brice Laurent   Linh Vu

2023-12-07

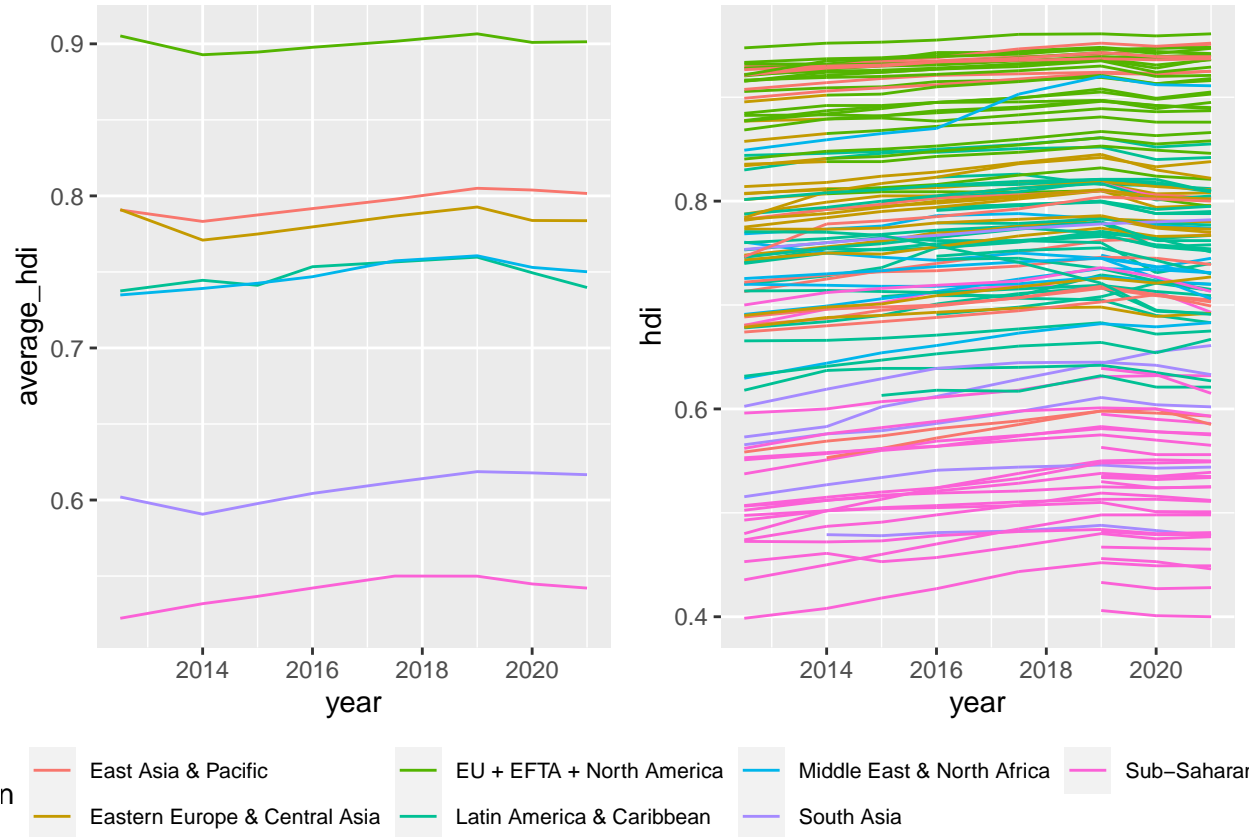## Mixed Effects Linear Model for Time

To include the data points from all the years collected in both datasets, we decided to incorporate time effects into a mixed effects model. For this model, we used the variables manually selected based on correlations: `x1.6`, `x3.2`, `x5.1`, `x6.4`, and `x7.3`. This mixed effects model is based on the equation:

$$HDI_{time,\ country,\ region} = \alpha_{region} + \beta_{region} \times X_{time,\ country,\ region} + a_{country} + b \times X_{time,\ country} + \varepsilon_{time,\ country,\ region}$$

where $\alpha_{region} \sim N(\mu_\alpha, \sigma_\alpha^2)$, $\beta_{region} \sim N(\mu_\beta, \sigma_\beta^2)$, $a_{country} \sim N(\mu_a, \sigma_a^2)$, $b_{country} \sim N(\mu_b, \sigma_b^2)$, and $\varepsilon_{time,\ country,\ region} \sim N(0, \sigma_{HDI}^2)$.

$X_{time,\ country,\ region}$ is the design matrix for the variables `x1.6`, `x3.2`, `x5.1`, `x6.4`, `x7.3`, and intercept, while $X_{time,\ country}$ is the design matrix for the time variable (in years) and intercept.

This equation above assumes independence between random effects. Though this assumption is reasonable for the random effects of Rule of Law variables between regions, it is not as reasonable for the random effects of the time variable between countries. This can be seen from the chart below since the HDI of different countries mostly follow similar trajectories over times, and therefore are associated. There is no closed form solution when we allow correlation between random effects, so the actual mixed effects model will not exactly follow the equation above.

Legend:
- East Asia & Pacific
- Eastern Europe & Central Asia
- EU + EFTA + North America
- Latin America & Caribbean
- Middle East & North Africa
- South Asia
- Sub–Saharan

The coefficients from the fixed effects are as follows:

	Estimate	Std. Error	df	t value	Pr(>|t|)	Bonferroni_Pr(>|t|)
(Intercept)	0.6925175	0.0519539	6.051174	13.3294513	0.0000103	7.22e-05
years_from_2000	0.0019365	0.0002481	125.613090	7.8055328	0.0000000	0.00e+00
x1.6	0.0044611	0.0108274	5.867850	0.4120229	0.6949501	1.00e+00
x3.2	0.0039985	0.0046819	6.165585	0.8540228	0.4250250	1.00e+00
x5.1	-0.0065979	0.0106958	5.189926	-0.6168717	0.5633750	1.00e+00
x6.4	-0.0088892	0.0082408	4.211854	-1.0786782	0.3385920	1.00e+00
x7.3	0.0172646	0.0141687	6.099454	1.2185040	0.2680548	1.00e+00

Although overly conservative, We have included Bonferroni adjused p-values, since multiple hypothesis tests were performed. The only fixed variable that was significant (both unadjusted and adjusted) in this model was the `years_from_2000` variable. The other variables were very far from significance.

```
## Data: selected_columns
## Models:
## time_mixed_effects_noRegion: hdi ~ years_from_2000 + x1.6 + x3.2 + x5.1 + x6.4 + x7.3 + (1 + years_f:
## time_mixed_effects: hdi ~ years_from_2000 + x1.6 + x3.2 + x5.1 + x6.4 + x7.3 + (1 + x1.6 + x3.2 + x5
##                              npar    AIC    BIC logLik deviance  Chisq Df
## time_mixed_effects_noRegion    10 -5356.9 -5308.7 2688.4  -5376.9
## time_mixed_effects             31 -5457.8 -5308.5 2759.9  -5519.8 142.89 21
```

```
##                              Pr(>Chisq)
## time_mixed_effects_noRegion
## time_mixed_effects            < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


## Data: selected_columns
## Models:
## time_mixed_effects_noYear: hdi ~ years_from_2000 + x1.6 + x3.2 + x5.1 + x6.4 + x7.3 + (1 + x1.6 + x3
## time_mixed_effects: hdi ~ years_from_2000 + x1.6 + x3.2 + x5.1 + x6.4 + x7.3 + (1 + x1.6 + x3.2 + x5
##                            npar     AIC      BIC logLik deviance   Chisq Df
## time_mixed_effects_noYear    30 -5086.1 -4941.6 2573.1  -5146.1
## time_mixed_effects           31 -5457.8 -5308.5 2759.9  -5519.8 373.69  1
##                              Pr(>Chisq)
## time_mixed_effects_noYear
## time_mixed_effects           < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We also analyzed the importance of including random effects of the variables x1.6, x3.2, x5.1, x6.4, and
x7.3 based on region, and the random effects of the time variable based on country using an extra-sum-of-
squared F test. These F-tests