Short Summary of Bachelor Thesis Methodology

Max Heinze

July 13, 2022

Data

Mining Data

Mining Polygons are taken from Maus et al. and then further prepared together with Nightlight Data to construct a mining activity index (see below).

Nightlight Data

VIIRS Nightlight Data is taken from the Earth Observation group (see https://eogdata.mines.edu/products/vnl/#annual_v2). I chose annually aggregated masked median values and then cropped the GeoTIFF files for further preparation and construction of the mining index (see below).

Land Cover Data

Land Cover Data is taken from Copernicus (see https://cds.climate.copernicus.eu/cdsapp#!/dataset/satellite-land-cover?tab=form). I aggregated values to the following categories: cropland, tree cover, urban, water, and other vegetation. In principle, this data is given in percentages.

Socioeconomic Data

Socioeconomic data is taken from Indonesia's national statistics office. They publish city/regency level figures of expenditure per capita, expected schooling, mean schooling and life expectancy (since they also publish a HDI figure for each city/regency).

Construction of Mining Index

I extracted night light data from polygons that equal the mining polygons from the mining data, plus a 10 arc second buffer zone. From all data points that fall within each buffered polygon, I take the median and then divide the data set by its standard deviation. For all means that have night light activity of 0 for all years, I apply the median of all mines that *do* have nightlight activity, per year, as a proxy for their night light activity. I then calculate a mining index for each mine by multiplying the mining area from the mining dataset with the night light figure for every year that I have aggregated.

I then assign each mine to a city/regency, for which I use the centroid as reference point in order to be able to uniquely assign them. The individual mining indices for all mines are then added together at the city/regency level to get one aggregated value per city/regency. In total, there are 554 c/r-year observations with a mining index and 4052 without. I assign a mining value of 0 to the latter. In addition, I construct a mining dummy that is 1 if the mining index $\neq 0$ and 0 otherwise.

Modeling

I have up to now run the following models:

Model 1: Tree cover area as dependent, mining value as independent variable

```
## Twoways effects Within Model
##
## Call:
## plm(formula = area_tree_cover ~ mining_value + expenditure +
##     exp_schooling + life_exp_birth, data = data_full, effect = "twoway",
##     model = "within", index = c("cr_code", "year"))
##
## Unbalanced Panel: n = 514, T = 7-11, N = 5587
##
## Residuals:
##        Min.     1st Qu.      Median     3rd Qu.        Max.
## -0.68942694 -0.01692659 -0.00026263  0.01863375  0.67700849
##
## Coefficients:
##                 Estimate Std. Error t-value  Pr(>|t|)
## mining_value   3.8022e-04 4.6333e-04  0.8206    0.4119
## expenditure    3.6829e-05 4.2756e-06  8.6138 < 2.2e-16 ***
## exp_schooling  1.5131e-03 3.9411e-03  0.3839    0.7011
## life_exp_birth 2.5194e-02 3.9436e-03  6.3885 1.825e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    22.688
## Residual Sum of Squares: 22.244
## R-Squared:      0.019567
## Adj. R-Squared: -0.082565
## F-statistic: 25.2416 on 4 and 5059 DF, p-value: < 2.22e-16
```

Model 2: Tree cover area as dependent, mining dummy as independent variable

```
## Twoways effects Within Model
##
## Call:
## plm(formula = area_tree_cover ~ mining_dummy + expenditure +
##     exp_schooling + life_exp_birth, data = data_full, effect = "twoway",
##     model = "within", index = c("cr_code", "year"))
##
## Unbalanced Panel: n = 514, T = 7-11, N = 5587
##
## Residuals:
##        Min.     1st Qu.      Median     3rd Qu.        Max.
## -0.69139954 -0.01604765 -0.00038608  0.01803155  0.62959476
##
## Coefficients:
##                   Estimate  Std. Error t-value  Pr(>|t|)
## mining_dummy   -6.3366e-02  6.8331e-03 -9.2734 < 2.2e-16 ***
## expenditure     3.6058e-05  4.2396e-06  8.5050 < 2.2e-16 ***
## exp_schooling   1.5508e-03  3.9055e-03  0.3971    0.6913
## life_exp_birth  2.5449e-02  3.9104e-03  6.5080 8.354e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    22.688
## Residual Sum of Squares: 21.875
## R-Squared:      0.035826
## Adj. R-Squared: -0.064612
## F-statistic: 46.9949 on 4 and 5059 DF, p-value: < 2.22e-16
```

Model 3: Tree cover percentage as dependent, mining value as independent variable

```
## Twoways effects Within Model
##
## Call:
## plm(formula = tree_cover ~ mining_value + expenditure + exp_schooling +
##     life_exp_birth, data = data_full, effect = "twoway", model = "within",
##     index = c("cr_code", "year"))
##
## Unbalanced Panel: n = 514, T = 7-11, N = 5587
##
## Residuals:
##        Min.     1st Qu.      Median     3rd Qu.        Max.
## -5.1715e-02 -3.7455e-03 -6.7821e-06  3.7337e-03  7.6101e-02
##
## Coefficients:
##                  Estimate  Std. Error t-value  Pr(>|t|)
## mining_value   -1.5056e-04  7.1448e-05 -2.1073   0.03514 *
## expenditure     3.2258e-06  6.5933e-07  4.8926 1.026e-06 ***
## exp_schooling   2.7575e-03  6.0775e-04  4.5372 5.832e-06 ***
## life_exp_birth  1.2413e-03  6.0813e-04  2.0412   0.04128 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    0.53353
## Residual Sum of Squares: 0.52896
## R-Squared:      0.0085499
## Adj. R-Squared: -0.09473
## F-statistic: 10.9067 on 4 and 5059 DF, p-value: 8.3521e-09
```

Model 4: Tree cover percentage as dependent, mining dummy as independent variable

```
## Twoways effects Within Model
##
## Call:
## plm(formula = tree_cover ~ mining_dummy + expenditure + exp_schooling +
##     life_exp_birth, data = data_full, effect = "twoway", model = "within",
##     index = c("cr_code", "year"))
##
## Unbalanced Panel: n = 514, T = 7-11, N = 5587
##
## Residuals:
##        Min.     1st Qu.      Median     3rd Qu.        Max.
## -5.1902e-02 -3.5580e-03 -6.9228e-05  3.6535e-03  7.6815e-02
##
## Coefficients:
##                  Estimate  Std. Error t-value  Pr(>|t|)
## mining_dummy   -9.1118e-03  1.0553e-03 -8.6345 < 2.2e-16 ***
## expenditure     3.0665e-06  6.5475e-07  4.6834 2.895e-06 ***
## exp_schooling   2.6966e-03  6.0315e-04  4.4708 7.963e-06 ***
## life_exp_birth  1.3070e-03  6.0391e-04  2.1642   0.03049 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:    0.53353
## Residual Sum of Squares: 0.52174
## R-Squared:      0.022091
## Adj. R-Squared: -0.079779
## F-statistic: 28.5707 on 4 and 5059 DF, p-value: < 2.22e-16
```

There is a significant effect of the mining index on tree cover percentage, and there is a significant effect of the mining dummy on both tree cover area and tree cover percentage. I am still not sure on which combination to settle – or if I should report all of them, as they yield different interpretations. For the "plain" panel model part, this would be all that I have in terms of models. Ideally, I had planned to then try a spatial model building on these, but I do not know how labor intensive this step is and if this is still viable given the time constraint.