**Supporting Information**

***Methods***

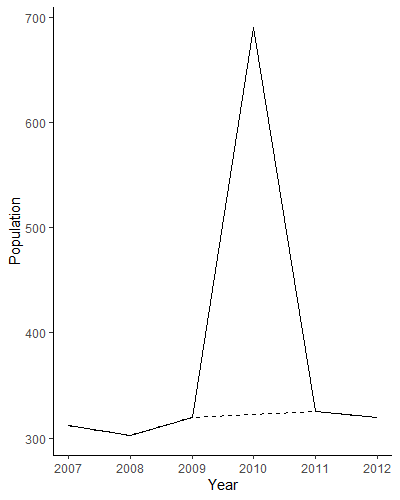
***Socioeconomic data cleaning***

Prior to aggregation to the commune level, village data were checked for missing values. In some cases, villages had data for a subset of years but were missing data for other years. If the missing data were at the start of the study period or the end of the study period it was assumed that the village was either an old or a new village. Villages can be merged with larger villages, or two sub-villages, or “*Kroms*”, can be split into two distinct villages over time for administrative purposes. In these cases, the rows (years) with missing data were deleted, but the years with data were retained as these represent villages that existed in that year. If the missing data were in the middle of the study period (for more than one year), or if data for that village only exists for one or several years in the middle of the study period, then the data were assumed to be incomplete and the village was deleted. If the village had data for all years except one, then the missing values were estimated using linear interpolation. If the village existed in all years, but was missing data from multiple years, the village was deleted. If an entire commune was missing in some years, the commune was deleted. The above cleaning process removed 312 villages (total number of villages = 84,195), or 0.37% of the data. Data were then split into individual years, and the final village-level data were aggregated to the commune- and province level using the operations defined below in Table Sx.

After aggregation, each variable was checked for obvious errors or unlikely outliers via plotting of histograms and trends. Plots were done at the province level first, to identify any communes within a province that had particularly unusual values or trends. If unusual values or trends were identified the commune was investigated in more detail. Outlier values that appeared inconsistent or implausible were removed and replaced with a value estimated via linear interpolation (Figure Sx). In some cases, where data had been converted from raw values to a proportion of the total population, errors in the raw data were discovered. This became clear when the resulting proportion was >1. In these cases the proportion was changed to 1.

**Table Sx. Mathematical operations used to aggregate socioeconomic variables from the village to the commune and province level.**

|  |  |
| --- | --- |
| Variable | Operation |
| Total population | Sum |
| Number of families | Sum |
| Number of males aged 18-64 | Sum |
| Number of females aged 18-64 | Sum |
| Number of people aged over 61 | Sum |
| Total number of indigenous people | Sum |
| Number of families whose main occupation is farming | Sum |
| Number of land conflict cases | Sum |
| Number of in-migrants | Sum |
| Number of out-migrants | Sum |
| Number of criminal cases | Sum |
| Proportion of population that is indigenous | Mean |
| Proportion of females aged 6-24 in full time education | Mean |
| Proportion of males aged 6-24 in full time education | Mean |
| Proportion of females aged 15-45 who are illiterate | Mean |
| Proportion of males aged 15-45 who are illiterate | Mean |
| Proportion of families whose main occupation is farming | Mean |
| Proportion of people who are primarily employed in the primary sector | Mean |
| Proportion of people who are primarily employed in the secondary sector | Mean |
| Proportion of people who are primarily employed in the tertiary sector | Mean |
| Proportion of people who are primarily employed in the quaternary sector | Mean |
| Proportion of families who have less than 1ha of farmland | Mean |
| Proportion of families who have buffalo | Mean |
| Proportion of families who have pigs | Mean |
| Proportion of families who have access to waste collection | Mean |
| Number of infant (<6mo) mortality cases | Mean |
| Number of child (<5 years old) mortality cases | Mean |
| Distance to the nearest school | Median |
| Distance to the Commune Office | Median |
| Distance to the nearest health centre | Median |



**Figure Sx. An example of linear interpolation for a commune with an implausible outlier. The example shows a value for the population of a commune in 2010 which is likely to be an error (solid line), and the resulting correction (dashed line).**

***Correlation***

For both analyses, correlation of predictors was assessed.

For the macroeconomic analysis, the following decisions were made based on high correlation (Table Sx):

* GNI variable dropped due to very high correlation with GDP. Competing theories about drivers of forest loss – national economy (GDP) or socioeconomic status of population (GNI). Because the second half of this chapter was focusing on socioeconomics, I decided that GDP was more interesting in this case.
* Neither population density (pop\_den) or producer price for rubber (prod\_rub) were dropped despite correlation. There is no plausible relationship between these two variables, and they were included to explain different drivers of forest loss. The two variables were in different variable sets, and so both were retained.
* Population density and amount of forest remaining (for\_rem) were positively correlated, which was counterintuitive. Previous studies have highlighted remaining forest as an important control variable, and so both variables were retained (O’Brien 2017).
* Producer price for rubber (prod\_rub) and forest remaining were negatively correlated. Previous studies have highlighted remaining forest as an important control variable, and so both variables were retained (O’Brien 2017).
* Agricultural Raw Materials Index (armi) was correlated with median price for rubber (rub\_med). This was likely to be a genuine correlation. The index was slightly correlated with more than one of the commodity price variables, and I was interested in the individual commodities, and so armi was dropped.
* Agricultural sector proportion of GDP (agr\_gdp) and industrial sector proportion of GDP (ind\_gdp) were correlated, and conceptually I was more interested in the impact of the agricultural sector (as it is more likely to affect forest cover), and so ind\_gdp was dropped.
* Median price of rice (rice\_med) and producer price of rice (prod\_rice) were correlated. These two variables were in different sets, and so were retained for the initial modelling.
* The producer price for rubber (prod\_rub) and the producer price for rice (prod\_rice) were correlated. A large number of the economic land concessions allocated in Cambodia were for rubber, and so my hypothesis was that rubber prices would be more important for predicting forest loss than rice. Therefore prod\_rice was dropped.

For the socioeconomic variables, correlation was assessed within each variable set. If there were incidents of high correlation, a principal component analysis (PCA) was conducted to see which variables explained the most variance. Based on these analyses, the following decisions were made:

* Total population, number of families, number of males, number of females, and population over 61 were all correlated. Following a PCA, total population was selected.
* As expected, all education variables were highly correlated. In this case, the proportion of males aged 6-24 was selected (without a PCA) because in this cultural context, males are far more likely to be engaged in activities that contribute to forest loss.
* As expected, there was a negative correlation between the proportion of people employed in the primary sector and the proportion of people employed in the tertiary and quaternary sectors, and a correlation between the proportion of people employed in the primary sector and the proportion of people whose main occupation was farming. The PCA results suggested that the proportion of people employed in the primary sector (propPrimSec) and secondary sector (propSecSec) were the most valuable predictors.
* Proportion of people with less than 1 hectare of farmland, and proportion of families who keep buffalos, were dropped due to inconsistencies in the data which suggested changes in the data collection or questions over time.
* Distance to the nearest school (dist\_sch) and distance to the nearest health centre (KM\_Heal\_cen) were correlated, and the PCA analysis was inconclusive about which variable to retain. Distance to school was retained based on the theory that forest clearance activities are more likely to be conducted by young males. An absence of accessible education is likely to be more of a driving factor in these activities than an absence of accessible health care.
* Both healthcare variables (infant mortality and child mortality) were dropped due to poor data quality.

A final assessment of correlation between predictor variables (after removal of the above) revealed no major correlations (Table Sx).

Table Sx. Correlation matrix for macroeconomic variables. Values over 0.6 are highlighted in red, and values below -0.6 are highlighted in yellow.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | for\_cov | for\_cov\_perc | gdp | gdp\_gr | gni | fdi | ind\_gdp | agr\_gdp | dev\_agri | dev\_env | pop\_den | armi | cpi | nfi | rice\_med | rub\_med | corn\_med | sug\_med | for\_prod | prod\_rice | prod\_rub | prod\_cass | prod\_corn | prod\_sug | for\_rem |
| for\_cov |  |  | -0.30 | 0.30 | -0.30 | -0.31 | 0.47 | -0.53 | 0.16 | -0.05 | 0.39 | 0.00 | -0.23 | -0.21 | -0.06 | 0.11 | 0.00 | -0.02 | -0.58 | -0.25 | -0.28 | 0.28 | -0.19 | -0.33 | 0.65 |
| for\_cov\_perc | |  | -0.30 | 0.30 | -0.30 | -0.31 | 0.47 | -0.53 | 0.16 | -0.05 | 0.39 | 0.00 | -0.23 | -0.21 | -0.06 | 0.11 | 0.00 | -0.02 | -0.58 | -0.25 | -0.28 | 0.28 | -0.19 | -0.33 | 0.65 |
| gdp | -0.30 | -0.30 |  | 0.40 | 0.99 | 0.25 | 0.12 | -0.22 | -0.01 | -0.01 | -0.51 | 0.30 | 0.35 | 0.00 | 0.19 | 0.22 | 0.23 | -0.15 | 0.14 | 0.41 | 0.63 | 0.26 | 0.57 | 0.19 | -0.60 |
| gdp\_gr | 0.30 | 0.30 | 0.40 |  | 0.39 | 0.20 | 0.12 | -0.30 | 0.27 | 0.14 | 0.00 | 0.30 | 0.28 | -0.06 | -0.20 | 0.37 | 0.02 | -0.22 | 0.12 | -0.02 | 0.11 | 0.52 | 0.60 | -0.09 | -0.03 |
| gni | -0.30 | -0.30 | 0.99 | 0.39 |  | 0.24 | 0.13 | -0.21 | 0.01 | 0.03 | -0.51 | 0.34 | 0.35 | -0.02 | 0.16 | 0.24 | 0.22 | -0.14 | 0.14 | 0.41 | 0.60 | 0.23 | 0.62 | 0.15 | -0.55 |
| fdi | -0.31 | -0.31 | 0.25 | 0.20 | 0.24 |  | -0.48 | 0.23 | 0.01 | 0.41 | -0.31 | 0.03 | 0.30 | 0.09 | 0.03 | 0.09 | 0.24 | -0.12 | 0.26 | 0.24 | 0.19 | 0.06 | 0.46 | 0.54 | -0.26 |
| ind\_gdp | 0.47 | 0.47 | 0.12 | 0.12 | 0.13 | -0.48 |  | -0.61 | -0.16 | 0.02 | 0.21 | 0.07 | -0.30 | -0.50 | -0.25 | 0.08 | -0.01 | -0.01 | -0.48 | -0.21 | -0.07 | 0.23 | -0.20 | -0.33 | 0.16 |
| agr\_gdp | -0.53 | -0.53 | -0.22 | -0.30 | -0.21 | 0.23 | -0.61 |  | -0.04 | 0.15 | -0.24 | -0.08 | 0.55 | 0.51 | 0.14 | -0.11 | -0.09 | 0.37 | 0.19 | 0.02 | -0.06 | -0.38 | -0.01 | 0.20 | -0.20 |
| dev\_agri | 0.16 | 0.16 | -0.01 | 0.27 | 0.01 | 0.01 | -0.16 | -0.04 |  | -0.12 | 0.01 | 0.26 | 0.15 | -0.08 | -0.06 | 0.02 | 0.02 | 0.12 | -0.06 | -0.08 | 0.16 | 0.05 | 0.13 | 0.06 | 0.02 |
| dev\_env | -0.05 | -0.05 | -0.01 | 0.14 | 0.03 | 0.41 | 0.02 | 0.15 | -0.12 |  | -0.03 | 0.05 | 0.30 | -0.24 | -0.35 | 0.11 | 0.03 | 0.32 | -0.01 | -0.28 | -0.04 | 0.03 | 0.08 | 0.16 | 0.01 |
| pop\_den | 0.39 | 0.39 | -0.51 | 0.00 | -0.51 | -0.31 | 0.21 | -0.24 | 0.01 | -0.03 |  | -0.43 | -0.45 | -0.06 | -0.31 | -0.31 | -0.19 | -0.38 | -0.26 | -0.58 | -0.79 | -0.06 | -0.35 | -0.48 | 0.79 |
| armi | 0.00 | 0.00 | 0.30 | 0.30 | 0.34 | 0.03 | 0.07 | -0.08 | 0.26 | 0.05 | -0.43 |  | 0.54 | -0.01 | 0.26 | 0.89 | 0.57 | 0.56 | 0.03 | 0.41 | 0.59 | 0.23 | 0.48 | -0.27 | -0.20 |
| cpi | -0.23 | -0.23 | 0.35 | 0.28 | 0.35 | 0.30 | -0.30 | 0.55 | 0.15 | 0.30 | -0.45 | 0.54 |  | 0.33 | 0.33 | 0.43 | 0.41 | 0.48 | 0.09 | 0.22 | 0.50 | -0.06 | 0.42 | 0.08 | -0.32 |
| nfi | -0.21 | -0.21 | 0.00 | -0.06 | -0.02 | 0.09 | -0.50 | 0.51 | -0.08 | -0.24 | -0.06 | -0.01 | 0.33 |  | 0.39 | 0.02 | -0.15 | 0.13 | -0.01 | 0.10 | -0.02 | -0.41 | 0.02 | 0.21 | 0.03 |
| rice\_med | -0.06 | -0.06 | 0.19 | -0.20 | 0.16 | 0.03 | -0.25 | 0.14 | -0.06 | -0.35 | -0.31 | 0.26 | 0.33 | 0.39 |  | 0.24 | 0.60 | 0.16 | 0.00 | 0.67 | 0.47 | 0.01 | 0.20 | 0.13 | -0.09 |
| rub\_med | 0.11 | 0.11 | 0.22 | 0.37 | 0.24 | 0.09 | 0.08 | -0.11 | 0.02 | 0.11 | -0.31 | 0.89 | 0.43 | 0.02 | 0.24 |  | 0.56 | 0.48 | 0.05 | 0.48 | 0.40 | 0.48 | 0.49 | -0.39 | -0.07 |
| corn\_med | 0.00 | 0.00 | 0.23 | 0.02 | 0.22 | 0.24 | -0.01 | -0.09 | 0.02 | 0.03 | -0.19 | 0.57 | 0.41 | -0.15 | 0.60 | 0.56 |  | 0.15 | -0.10 | 0.46 | 0.48 | 0.23 | 0.40 | -0.15 | -0.04 |
| sug\_med | -0.02 | -0.02 | -0.15 | -0.22 | -0.14 | -0.12 | -0.01 | 0.37 | 0.12 | 0.32 | -0.38 | 0.56 | 0.48 | 0.13 | 0.16 | 0.48 | 0.15 |  | -0.19 | 0.11 | 0.28 | -0.01 | -0.12 | -0.07 | -0.14 |
| for\_prod | -0.58 | -0.58 | 0.14 | 0.12 | 0.14 | 0.26 | -0.48 | 0.19 | -0.06 | -0.01 | -0.26 | 0.03 | 0.09 | -0.01 | 0.00 | 0.05 | -0.10 | -0.19 |  | 0.27 | 0.15 | 0.11 | 0.25 | 0.11 | -0.41 |
| prod\_rice | -0.25 | -0.25 | 0.41 | -0.02 | 0.41 | 0.24 | -0.21 | 0.02 | -0.08 | -0.28 | -0.58 | 0.41 | 0.22 | 0.10 | 0.67 | 0.48 | 0.46 | 0.11 | 0.27 |  | 0.63 | 0.36 | 0.47 | 0.15 | -0.46 |
| prod\_rub | -0.28 | -0.28 | 0.63 | 0.11 | 0.60 | 0.19 | -0.07 | -0.06 | 0.16 | -0.04 | -0.79 | 0.59 | 0.50 | -0.02 | 0.47 | 0.40 | 0.48 | 0.28 | 0.15 | 0.63 |  | 0.05 | 0.39 | 0.41 | -0.73 |
| prod\_cass | 0.28 | 0.28 | 0.26 | 0.52 | 0.23 | 0.06 | 0.23 | -0.38 | 0.05 | 0.03 | -0.06 | 0.23 | -0.06 | -0.41 | 0.01 | 0.48 | 0.23 | -0.01 | 0.11 | 0.36 | 0.05 |  | 0.36 | -0.37 | -0.07 |
| prod\_corn | -0.19 | -0.19 | 0.57 | 0.60 | 0.62 | 0.46 | -0.20 | -0.01 | 0.13 | 0.08 | -0.35 | 0.48 | 0.42 | 0.02 | 0.20 | 0.49 | 0.40 | -0.12 | 0.25 | 0.47 | 0.39 | 0.36 |  | 0.01 | -0.29 |
| prod\_sug | -0.33 | -0.33 | 0.19 | -0.09 | 0.15 | 0.54 | -0.33 | 0.20 | 0.06 | 0.16 | -0.48 | -0.27 | 0.08 | 0.21 | 0.13 | -0.39 | -0.15 | -0.07 | 0.11 | 0.15 | 0.41 | -0.37 | 0.01 |  | -0.51 |
| for\_rem | 0.65 | 0.65 | -0.60 | -0.03 | -0.55 | -0.26 | 0.16 | -0.20 | 0.02 | 0.01 | 0.79 | -0.20 | -0.32 | 0.03 | -0.09 | -0.07 | -0.04 | -0.14 | -0.41 | -0.46 | -0.73 | -0.07 | -0.29 | -0.51 |  |

Table Sx. Correlation matrix for the socioeconomic variables. There were no coefficients greater than 0.6 or less than -0.6

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | tot\_pop | prop\_ind | pop\_den | M6\_24\_sch | propPrimSec | propSecSec | Les1\_R\_Land | pig\_fam | dist\_sch | garbage | KM\_Comm | land\_confl | crim\_case | Pax\_migt\_in | Pax\_migt\_out | mean\_elev | dist\_border | dist\_provCap |
| tot\_pop |  | -0.33 | 0.42 | 0.19 | -0.31 | 0.04 | 0.15 | -0.28 | -0.36 | 0.10 | 0.00 | 0.32 | -0.10 | 0.39 | 0.35 | -0.34 | 0.14 | -0.10 |
| prop\_ind | -0.33 |  | -0.22 | -0.36 | 0.15 | -0.04 | -0.16 | 0.20 | 0.43 | -0.04 | 0.21 | -0.08 | 0.13 | -0.12 | -0.14 | 0.47 | -0.18 | 0.06 |
| pop\_den | 0.42 | -0.22 |  | 0.30 | -0.46 | 0.16 | 0.20 | -0.26 | -0.35 | 0.35 | -0.21 | -0.03 | -0.12 | 0.11 | 0.05 | -0.31 | 0.01 | -0.39 |
| M6\_24\_sch | 0.19 | -0.36 | 0.30 |  | -0.09 | 0.01 | 0.11 | 0.07 | -0.37 | 0.03 | -0.21 | 0.04 | -0.09 | -0.04 | -0.06 | -0.20 | 0.07 | -0.21 |
| propPrimSec | -0.31 | 0.15 | -0.46 | -0.09 |  | -0.26 | 0.09 | 0.50 | 0.24 | -0.32 | 0.14 | 0.05 | 0.01 | -0.18 | -0.21 | 0.02 | 0.08 | 0.27 |
| propSecSec | 0.04 | -0.04 | 0.16 | 0.01 | -0.26 |  | 0.02 | -0.14 | -0.08 | 0.05 | -0.05 | -0.02 | -0.02 | 0.04 | 0.03 | -0.02 | 0.04 | -0.09 |
| Les1\_R\_Land | 0.15 | -0.16 | 0.20 | 0.11 | 0.09 | 0.02 |  | 0.01 | -0.23 | -0.08 | -0.11 | -0.01 | -0.09 | 0.07 | 0.01 | -0.21 | 0.20 | -0.14 |
| pig\_fam | -0.28 | 0.20 | -0.26 | 0.07 | 0.50 | -0.14 | 0.01 |  | 0.19 | -0.14 | 0.06 | -0.02 | -0.03 | -0.21 | -0.21 | 0.02 | -0.10 | 0.15 |
| dist\_sch | -0.36 | 0.43 | -0.35 | -0.37 | 0.24 | -0.08 | -0.23 | 0.19 |  | -0.07 | 0.36 | -0.06 | 0.07 | -0.12 | -0.15 | 0.36 | -0.14 | 0.38 |
| garbage | 0.10 | -0.04 | 0.35 | 0.03 | -0.32 | 0.05 | -0.08 | -0.14 | -0.07 |  | -0.06 | -0.03 | 0.04 | 0.05 | 0.02 | 0.00 | -0.05 | -0.13 |
| KM\_Comm | 0.00 | 0.21 | -0.21 | -0.21 | 0.14 | -0.05 | -0.11 | 0.06 | 0.36 | -0.06 |  | 0.09 | 0.03 | 0.04 | 0.00 | 0.11 | -0.05 | 0.24 |
| land\_confl | 0.32 | -0.08 | -0.03 | 0.04 | 0.05 | -0.02 | -0.01 | -0.02 | -0.06 | -0.03 | 0.09 |  | 0.27 | 0.13 | 0.05 | -0.06 | 0.04 | 0.07 |
| crim\_case | -0.10 | 0.13 | -0.12 | -0.09 | 0.01 | -0.02 | -0.09 | -0.03 | 0.07 | 0.04 | 0.03 | 0.27 |  | -0.03 | -0.05 | 0.16 | -0.13 | 0.02 |
| Pax\_migt\_in | 0.39 | -0.12 | 0.11 | -0.04 | -0.18 | 0.04 | 0.07 | -0.21 | -0.12 | 0.05 | 0.04 | 0.13 | -0.03 |  | 0.42 | -0.10 | 0.01 | 0.01 |
| Pax\_migt\_out | 0.35 | -0.14 | 0.05 | -0.06 | -0.21 | 0.03 | 0.01 | -0.21 | -0.15 | 0.02 | 0.00 | 0.05 | -0.05 | 0.42 |  | -0.12 | 0.02 | -0.03 |
| mean\_elev | -0.34 | 0.47 | -0.31 | -0.20 | 0.02 | -0.02 | -0.21 | 0.02 | 0.36 | 0.00 | 0.11 | -0.06 | 0.16 | -0.10 | -0.12 |  | -0.26 | 0.15 |
| dist\_border | 0.14 | -0.18 | 0.01 | 0.07 | 0.08 | 0.04 | 0.20 | -0.10 | -0.14 | -0.05 | -0.05 | 0.04 | -0.13 | 0.01 | 0.02 | -0.26 |  | -0.05 |
| dist\_provCap | -0.10 | 0.06 | -0.39 | -0.21 | 0.27 | -0.09 | -0.14 | 0.15 | 0.38 | -0.13 | 0.24 | 0.07 | 0.02 | 0.01 | -0.03 | 0.15 | -0.05 |  |

***Modelling***

*Macroeconomic models*

Each of the predictors in the three model sets were selected because they were hypothesised to be potential drivers or effective predictors of forest loss (Table Sx).

**Table Sx. Hypothesised relationships between macroeconomic variables and forest loss**

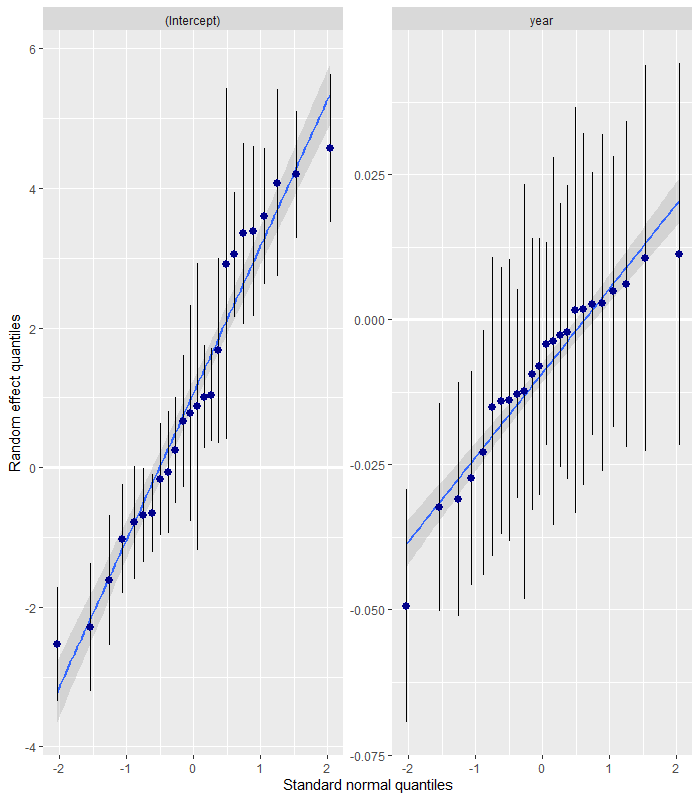
|  |  |
| --- | --- |
| **Variable** | **Hypothesis** |
| *Economic development* |  |
| GDP | Increases in national economic development and wealth will increase forest loss |
| GDP growth | The rate of GDP growth will affect the rate of forest loss |
| FDI | Increased foreign investment will increase forest loss (e.g. through economic land concessions) |
| Agricultural sector proportion of GDP | As the agricultural sector’s contribution to GDP increases, so will forest loss (reflecting increases in agro-industrial concessions). |
| Alternative hypothesis: as the agricultural sector’s contribution to GDP decreases forest loss will increase (reflecting urbanisation and urban expansion) |
| Development flows to agriculture | Increased investment into the agricultural sector will increase forest loss (agricultural expansion) |
| Alternative hypothesis: Increased investment into the agricultural sector will decrease forest loss (increased productivity and intensification of existing agricultural land) |
| Development flows to the environment | Increased investment into the environment sector will decrease forest loss |
| Population density | Increases in population density will increase forest loss |
| *Commodities* |  |
| Crop production index | Increases in crop production will increase forest loss |
| Non-food production index | Increases in non-food agricultural production will increase forest loss |
| Median rice price | Increases in the price of rice will increase forest loss |
| Median rubber price | Increases in the price of rubber will increase forest loss |
| Median corn price | Increases in the price of corn will increase forest loss |
| Median sugar price | Increases in the price of sugar will increase forest loss |
| Production value from forestry | Increases in the production of forestry products will increase forest loss |
| *Producer prices* |  |
| Producer price, rubber | Increases in the producer price of rubber will increase forest loss |
| Producer price, cassava | Increases in the producer price of cassava will increase forest loss |
| Producer price, corn | Increases in the producer price of corn will increase forest loss |
| Producer price, sugar | Increases in the producer price of sugar will increase forest loss |

*Socioeconomic models*

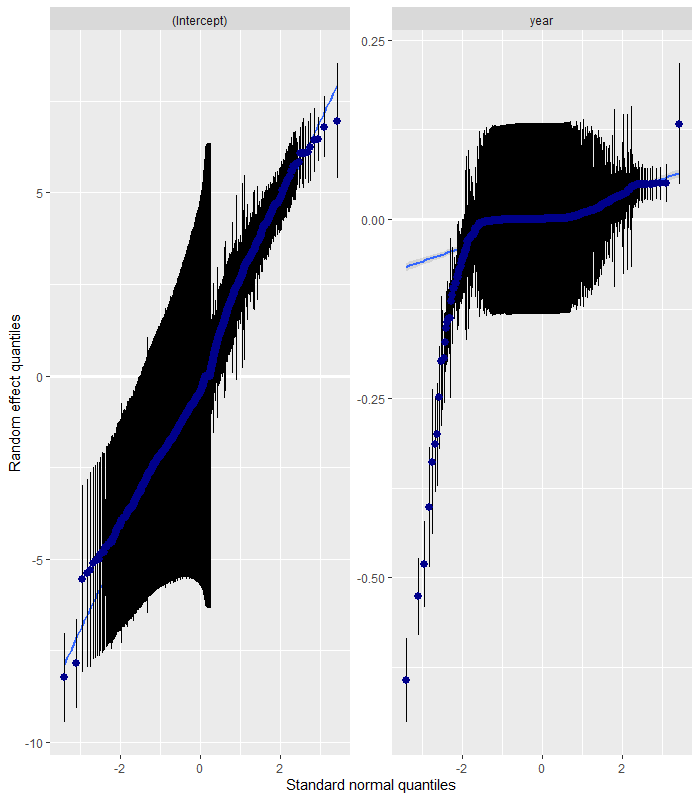
Each of the eight model sets were selected because they were hypothesised to be potential drivers or effective predictors of forest cover (Table Sx). The predictors within each of the sets were selected as proxies for the set because of their relevance, or because they were the best quality data that related to the set.

|  |  |  |
| --- | --- | --- |
| **Set** | **Hypothesis** | **Variable(s)** |
| Demographics | Communes/provinces with higher human populations and higher human population density will have lower forest cover due to urbanisation and agricultural expansion. Communes/provinces with higher indigenous populations will have higher forest cover because areas with high indigenous populations are more remote, and indigenous communities rely more on forests for traditional livelihoods. | Total population |
| Population density |
| Proportion indigenous |
| Education | Communes/provinces with lower levels of education will have lower forest cover because logging and forest clearance is conducted predominantly by young males of school age. Alternative hypothesis: communes/provinces with higher levels of education will have lower forest cover because education levels are likely to be higher in urban areas. | Proportion of males aged 6 – 24 in full time education |
| Employment | Communes/provinces with higher proportions of adults in the primary sector will have higher forest cover because these areas are likely to be more remote and have more natural resources such as forests. Communes/provinces with higher proportions of adults in the secondary sector will have lower forest cover as the secondary sector will be more prominent in urban/developed areas. | Proportion of adults employed in the primary sector |
| Proportion of adults employed in the secondary sector |
| Economic security | Communes/provinces with higher proportions of families with poor economic security (farmland, livestock) will have lower forest cover because rural populations in areas with high forest cover have access to land and livestock, whereas poor families in urban/developed areas do not. | Proportion of families with <1ha of rice land |
| Proportion of families who keep pigs |
| Access to services | Communes/provinces with large distances to schools are likely to be large, remote communes/provinces with high forest cover. Alternative hypothesis: areas with large distances to schools will lead to higher proportions of males out of education and engaging in forest clearing activities. Communes/provinces with higher proportions of families with access to waste collection will be in developed, urban areas and will have lower forest cover. Communes/provinces with larger distances to commune offices will be larger, more remote areas with higher forest cover. Alternative hypothesis: communes/provinces with larger distances to commune offices will have weaker governance and less law enforcement, resulting in lower forest cover. | Distance to nearest school |
| Proportion of families with access to waste collection |
| Distance to the Commune office |
| Social justice | Communes/provinces with a higher number of criminal cases will be more urbanised area and therefore will have lower forest cover. Communes/provinces with a higher number of land conflicts will be in areas of high forest cover where land speculation and land disputes are high. Alternative hypothesis: communes/provinces with a higher number of land conflicts will be in areas with a high number of economic land concessions where forest clearance has occurred, and so will have lower forest cover. | Number of criminal cases |
| Number of land conflicts |
| Migration | Communes/provinces with a high number of in-migrants will be urban areas with large industry (i.e., high job availability) and therefore low forest cover. Alternative hypothesis: communes/provinces with a high number of in-migrants will be areas with new economic land concessions which are often in areas of high forest cover. Communes/provinces with a high number of out-migrants will have higher forest cover because they are rural, remote areas with fewer job opportunities. | Number of in-migrants |
| Number of out-migrants |
| Control | All of these variables have potential to influence forest cover within communes/provinces, yet were not specific targets for investigation. Therefore they were included as control variables. | Mean elevation |
| Distance to international border |
| Distance to provincial capital |
| Presence of economic land concessions |
| Presence of protected areas |
| Protected area category |

**Results**



**Figure x. Quantile-quantile plots for the random effect “Province” of the final socioeconomic model**



**Figure x. Quantile-quantile plots for the random effect “Commune” of the final socioeconomic model. Plots suggest the assumption of normality of deviations of the conditional means of the random effects from the global intercept is violated.**

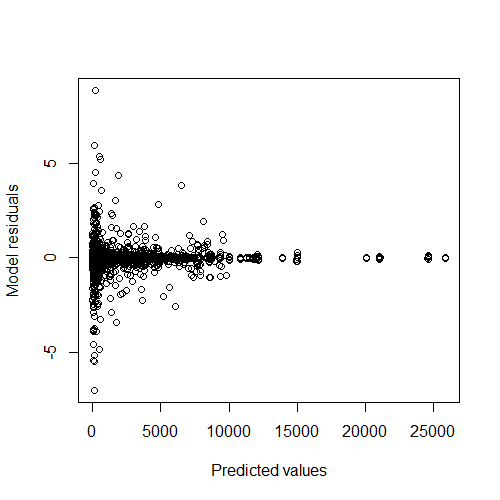


Figure x. Plot of residuals versus fitted values for the final socioeconomic model