Indodiversity: Analyzing biodiversity effects on rubber and oil palm plots in Jambi, Indonesia

Module: Data Analysis with R in Agricultural Economics

Student Name: Jakob Vincent Latzko

Matr. - Nr: 21344724

Master program: Development Economics

Semester: 5th

Abstract

Biodiversity loss due to transformation of tropical rainforest into utilizable land is one of the most relevant environmental topics of our time. An overall trend of reducing biodiversity at small-scale farm level sustained by the newest round of our dataset points into this direction. Our dataset combines agricultural, economic and ecological problem sets, as it analyzes species richness on rubber and oil palm plots in Jambi, Indonesia. One of the world's regions that is the most affected by tropical land-use change. Further analysis is needed to draw more substantial conclusions from the data.

Keywords: *Rubber, palm oil, Jambi, Indonesia, small-scale farming, species richness, lowland rainforest transformation*

1 Introduction

This seminar paper was written as a final paper of the course *Data Analysis with R in Agricultural Economics* which took place in March, 2019. Not withstanding the other topic options, e.g. Time Series, Panel Data or gravity models in trade, I chose to pursue my own seminar paper project. The data I use is about species richness, more precisely, plant richness in rubber and palm-oil plantations. More details about the data characteristics are provided in the first section of the main part.

Especially Palm-oil has been widely discussed in the media and it is often mentioned as one of the principal causes of rainforest clearing. The data we collected is highly relevant from an interdisciplinary point of view, bringing ecological, agricultural and socioeconomic problem sets together. In a time where movements against climate change become agenda setters, this topic becomes more and more relevant and needs to be discussed. With this seminar paper I will provide a fact-based, scientifical contribution to this discussion.

Throughout the paper, first I will open up the subject by describing the problem set. Subsequently I will show descriptive statistics to show the structure of the data we collected. For a biodiversity analysis, some indices are needed wich are explained in section 2.3 Finally, one of these indices, the Shannon Index, is applied to the data and mapped over the five regencies of Jambi province.

By undertaking all these steps I will try to answer the following question: *Is there a negative trend of species richness described by the data?*

2 Main part

2.1 Literature insight and problem set

Both rubber and oil palm farmers have in common that most of them do not depend entirely on the earnings of their agricultural activity; many of them pursue further economic activities and have e.g. a small retail trade or work as a mechanic.

Sayer et al. (2012) formulates the four truths about palm oil, which "should be acknowledged in any meaningful debate about the topic:

1. the demand for oil palm will continue to increase in response to a growing and increasingly affluent global population
2. oil palm is one of the most profitable land uses in the humid tropics
3. oil palm plantations store more carbon than alternative agricultural land uses
4. native biodiversity within oil palm plantations is far lower than in the natural forests they often replace

Biodiversity is always reducing substantially regarding any type of large scale agriculture which uses the same land than the former rainforest. Also rubber, banana or cocoa plantations have a negative biodiversity balance regarding autochthonous rainforests.

What becomes clear when reading about any of the agricultural technologies in this context, is that there is a clear trade-off. A trade-off between increasing agricultural production by expanding the cultivated area and the environmental protection

However, the productivities of palm oil is not at all at its peak. The same paper of Sayer et al. (2012) states that under favorable conditions 10-11 tons per hectar can be produced, while the average for Indonesia is only at 3.9 $\frac{t}{ha}$ and the best observed yields reaching 7-8.

In this context, rubber tells a different story which might, nonetheless, be considered as the second side of the same coin. In Jambi province, small-scale farmers of rubber and palm oil live usually house to house. And there are many farmers who cultivate both.

Without going into detail, the rubber farming has been disrupted a lot by the price collapse which happened after 2012. Although by July 2019 the price seems to be slightly increasing again, the overall tendence of the rubber price over the last years is clearly negative. Today, many rubber farmers consider changing their technology to oil palm which yields at current prices a more profitable product.

Thus, the first conclusion that can be drawn out of the literature is that there is still a lot of room for improvements. The palm-oil farmers in Indonesia could potentially produce a lot more - even within their current area. Coming back to the trade-off between agriculture and environment that I discussed before: the environmental damage might well be hold at a limited level, while the production per unit of are is increased. This would permit us to hold this trade-off at the lowest possible level.

The biodiversity data collected in the C01 project plays an important role for understanding the reasons behind the inefficiencies that currently exist on the plots of Jambi - both for rubber and for oil palm.

2.2 Data structure

This chapter gives an overview - mostly by descriptive statistics - about the data that we collected last year and that will be analyzed throughout this paper. This is important to familiarize ourself with the data in order to detect weaknesses or sources of bias.

The data was collected in three rounds from 2012 to 2018 in Jambi Province in Sumatera, Indonesia. The last round, 2018, was collected by my colleague Bernhard Dalheimer, our Indonesian assistants and me. The C01 sample represents a subsample within the major C-group's sample of 831 plots of around 750 small-scale farmer households. Its size is around 200 households, whereas in each round c.a. 20 additional plots - which then cannot be used for the panel data analysis - were included.

The C-group sample was designed as a representative sample of the small-scale farmers in Jambi province, Indonesia. $\frac{3}{4}$ of the households are rubber farmers, roughly $\frac{1}{4}$ of the sample produces palm oil. The rubber farmers can be divided into jungle rubber and rubber plantation groups. The distinction of these two groups plays an important role concerning biodiversity on the plots, I will continue pointing this out in the biodiversity section. About the sampling framework, the responsible researchers (Faust et al. 2013: 16) wrote the following:

> *"Five regencies, which comprise most of the lowland transformation systems of Jambi province, were selected purposively. These regencies are Sarolangun, Bungo and Tebo as the ones in proximity of the Bukit Duableas National Park as well as Batanghari and Muaro Jambi, representing the Harapan Rainforest research area. In order to capture geographical disparity and regional diversity, the number of villages per regency and district was fixed. From each of the selected regencies, five districts and four rural villages from each of these districts were selected randomly."*

In this research framework, the C-groups are rather concerned about socioeconomic data, whilst the other supergroups are closer to natural sciences. The data collection for C01 consisted in:

- plant richness data collection
- soil data collection
- soil sample processing

Within this context, the field study lived up to its name, as for the data collection it was necessary to reach a specific spot deep into the rubber or palm oil plantations. We arrived to the exact same spot either thanks to GPS navigation, or to the farmers' memories about the last C01 visit. Subsequently, the 25m² square was built up and all the plants in this quadrant were counted and listed.

In the 2018 round, slightly less than 10% of the sample could not be visited, either because the farmer was unavailable or because he gave up his plot.

Table 1: Data extract

household ID	Number of species
3	8
12	18
13	13
15	6
22	12
24	11
25	9
29	13
34	13
37	14

Figure1: Density of species reported per wave

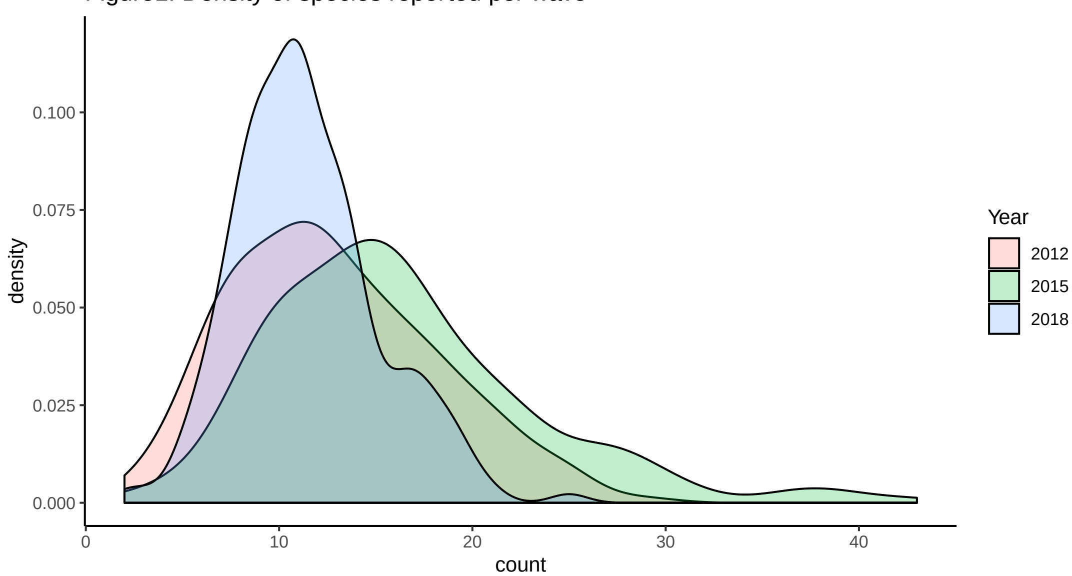


Table 1 shows some examples of how many different species were usually reported per plot. We see that the range is usually from 5 to 15 species. This data is from the last round of 2018. In order to make the reported numbers comparable, Figure 1 provides us a comparison of the densities of the number of species reported. As it can be seen, the peak of the distributions are slightly above a species count of 11. Although the 2015 wave has its peak at 15 species per plot. Furthermore, the 2018 wave has a higher peak which means that the numbers of species reported are distributed closer to its peak of 10. While in 2012, 7.5% of the observations took a value around 11, in 2018 it was more than 10%.

Despite the minor differences of the densities of the data collection, the three distributions are quite comparable. That becomes even clearer when considering the large darkblue area which is covered by all three density graphs. Now, after having performed the first glimpse at the data, we can thus state that there are no severe inconsistencies.

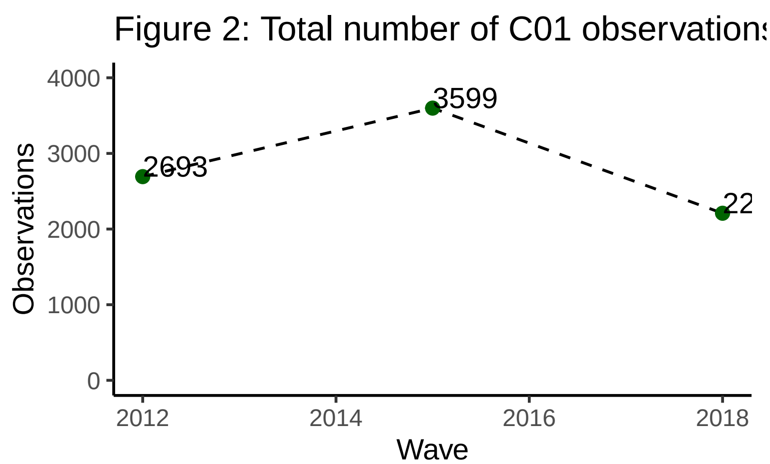


Figure 2 shows us the total number of observations in the data set. The number of observations is increasing in 2015 with respect to 2012 and decreases again in 2018.

Table 2: Total number of species per wave

Wave	Species
2012	541
2015	1477
2018	504

After having seen the distribution of the observations adn the total number of observations, we have a look at the total number of species reported per wave in Table 2. What is eye-catching is the proportionately high number of different species for 2015, while the 2012 and 2018 wave both report numbers that are slightly above 500. After having a look at the dataset, one finds out that the method of reporting unknown plants in the 2015 was different. For each of these an individual name was assigned with the structure "unident_x". The dataset reports more than 1000 observations with these names - and very few of them appear twice or thrice. Which makes the overall number of different species considerably high. In this context it might appear reasonable to talk about an outlier - which is quite likely due to the use of this naming method.

Further ahead in the paper concerning the species richness indicators and the Shannon Index interpretation, this data issue becomes especially relevant. For the data structure we can hitherto state that all three waves might have differences - but they still remain comparable and are suitable for further analysis which will be pursued in the following parts.

2.3 Methodology

Beyond descriptive statistics which are presented in this paper, a more analytical approach is the use of biodiversity measures. For our C01 dataset the use of species richness indices such as the Simpson and the Shannon Index are the most appropriate. The formula of the Shannon index is as follows:

$$H = -\sum_{i=1}^{s} p_i \ln p_i \tag{1}$$

Where s is the total number of species - in our case, we apply the total number of species found in each round. p is the proportion - as long as the formula is applied at plot level - of the species i found in one plot over all species. This proportion is then multiplied by the natural logarithm of the same. The index accounts for both abundance and evenness of the species found and the outcome of this index in one plot is comparable to the one of another plot in the same round.

The formula of the Simpson index is similar:

$$D = \frac{1}{\sum_{i=1}^{s} p_i^2} \tag{2}$$

The main difference regarding the interpretation of both indices is that the Simpson Index takes the dominance of one species on a plot into account. If there is a frequency of 2000 of one plant and four more plants of one each in a plot, the Simpson Index will yield a considerably low index of species richness. Whereas six evenly distributed plants of 500 each would yield a higher Simpson Index.

The Shannon Index is a rather general measure of plant richness and takes rather the total number of species that is reported per plot into account. The higher the number of species - the higher the index. For the sake of comparableness I decided that for the mapping I will use the Shannon index. The mapping will be discussed in the following part.

2.4 Discussion of the findings

2.4.1 Invasives

The first thing we will have a look at in our biodiversity analysis are the most frequent species for the 2018 wave. In Table 3 we see that *Cyrtococcum patens* has a simple majority. Which is not surprising because it is a species from the grass family. *Clidemia hirta*, however, is a perennial shrub, not original from Indonesia and it is an invasive plant. Invasive plants take away the space of autochthonous plants. So a high share of invasive species would generally lead to lower biodiversity scores.

In this *top 10* of the most frequent plants, three are invasive. It would be interesting to see comparable statistics from a virgin rainforest to see how or if these are also affected by invasive species.

Table 3: The most frequent species

Species name	Total frequency	Invasive species
Cyrtococcum patens	31180	no
Clidemia hirta	10714	yes
Centotheca lappacea	9368	no
Asystasia gangetica	8229	yes
Ottochloa nodosa	7247	no
Axonopus compressus	6829	no
Leptaspis urceolata	5118	no
Scleria ciliaris	4523	no
Legazpia polygonoides	3922	no
Dicranopteris linearis	3733	yes

2.4.2 Mapping

The next part is the one that caused by far the most effort, time and work. For each wave, a map of Jambi province is provided. For each map, the means of the Shannon Index over all plots within a regency were calculated. As Jambi has 11 regencies and our sample is only collected in five, there are grey areas. The regencies, from left to right, are *Bungo, Tebo, Sarolangun, Batang Hari and Muaro Jambi*. In the first place, the 2012 and 2018 maps look quite similar regarding the colours and the distribution of the colours over the regency. Yet, the 2018 map is less greener than the one from 2012.

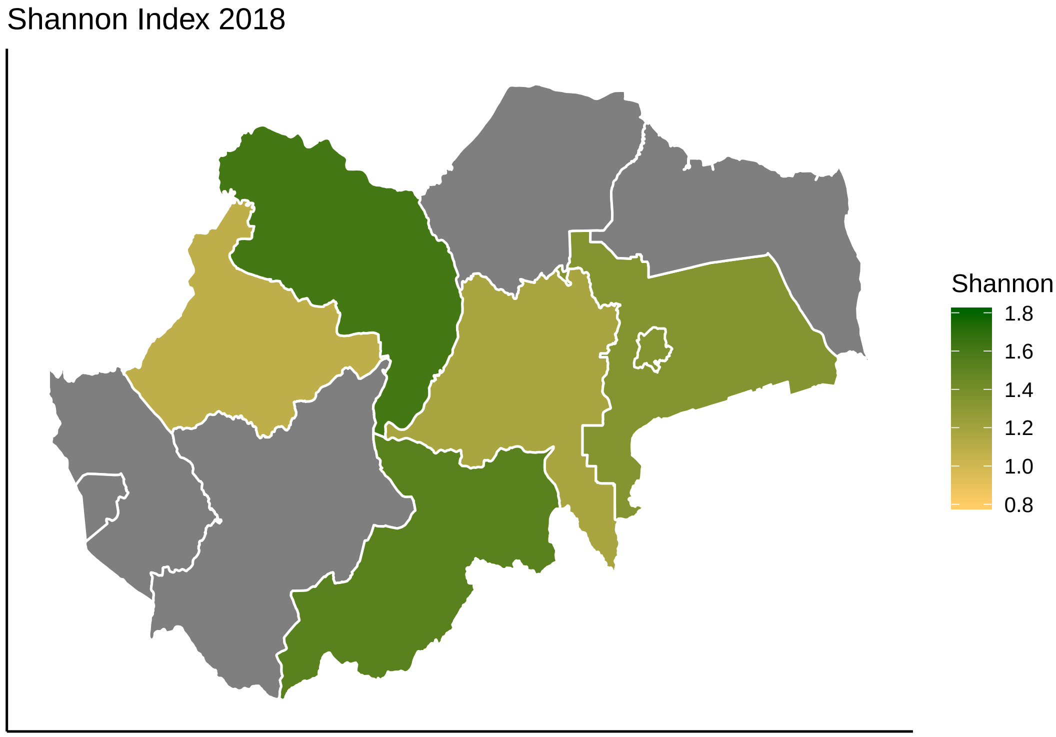
Regarding the 2015 map, it seems that the species richness on the plots is substantially lower than in the other waves. However, we cannot really draw this conclusion. The reason is a result of data problem. As the total number of species for the 2015 wave is as high as described in Table 2, the baseline of the Shannon Index will also increase - and the proportion will be lower than in the other waves. This creates Shannon indizes in a very different magnitude.

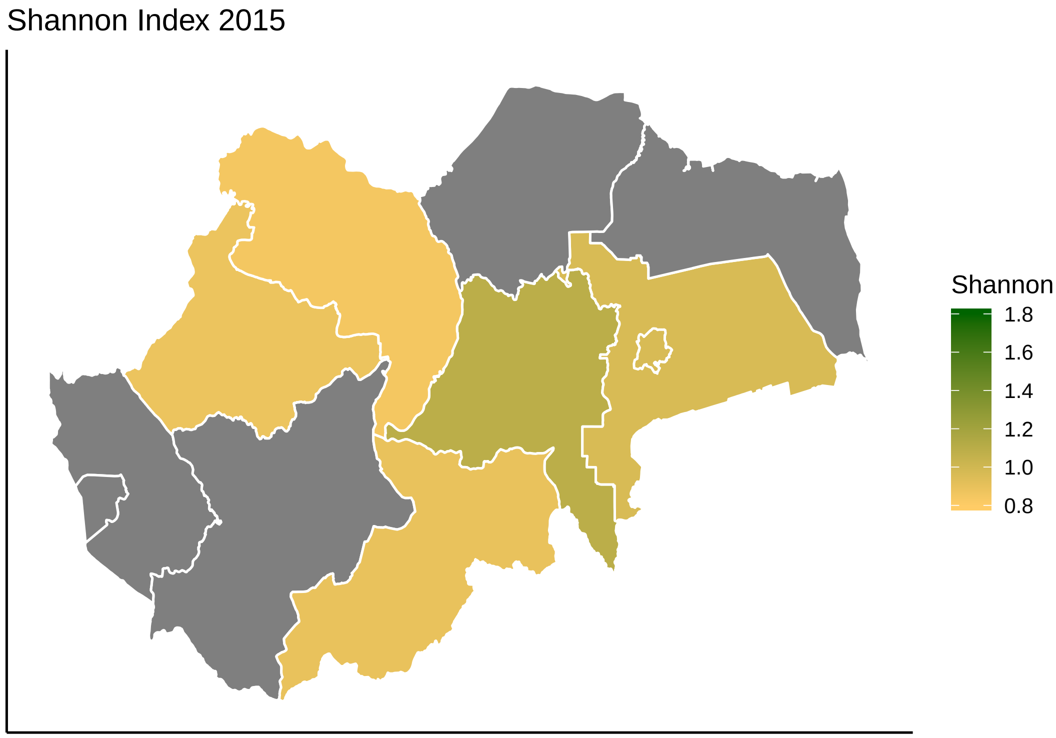
For the moment, we will mostly focus on the 2012 and 2018 maps. The green spot on both of them is the *Tebo* region in the north of Jambi Province. Its Shannon Index is substantially higher compared to those of the other regencies. *Batang Hari* is the only regency that seems to maintain the same colour throughout the three waves. The species richness of *Sarolangun* and *Muaro Jambi* region reduced according to the data.

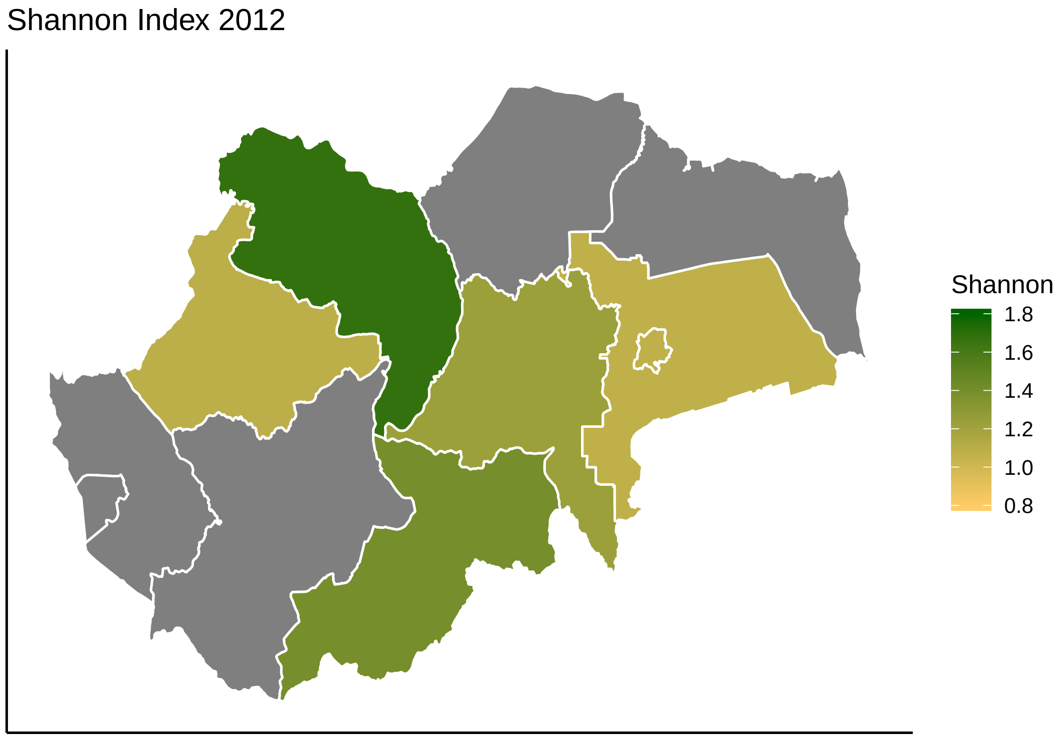
One of the motivations for the mapping was to make these regencies comparable in terms of biodiversity on the plots. Once this is done, it will be possible to look at the reasons why these differences exist. Why is that relevant? Because if there exist biodiversity differences between the regencies which share at least the same climate zone, maybe some farmers work more efficiently than others in terms of avoiding environmental damage. Once this finding can be isolated, the farmers of the other regions might adapt the technology or technique of the environmentally more efficient regency.

In the case of *Tebo* it seems likely that the higher Shannon figure has to do with the high share of rubber plantations - which tend to have a higher biodiversity - in the sample. Still, with the available data I cannot fortify this notion with hard statistical evidence.

In any case, the next step would be to seperate the rubber from the oil palm plots in the sample and look at the different means at the map.







3 Conclusion

As just mentioned by the final part of the last chapter, this paper can only be the beginning of digging deeper into the topic. I touched many topics I considered relevant in the context of the aforementioned production-environment trade-offs. However, for the moment no sustainable conclusions can be drawn that indicate a causal relationship in any direction.

While writing this paper there were more and more problem sets and data visualization schemes popping up in my head. These C01 datasets themselve have a lot more potential, especially when pointing out the differences between oil palm and rubber. Not to mention the potential of the data when merging them with socioeconomic survey data. Especially possible linkages between plant diversity and certain income measures would be very interesting to analyze. Although the sample size is not big, the panel structure let the findings yield some explanatory power. Yet, given the small scale of the seminar, decisions had to be made to the detriment of some measures and in favor of others.

Taking all the aforementioned chapters into account, the total number of species reported and the overall Shannon Index diminished in the 2018 sample with regard to 2012. I would therefore conclude that the data is in fact showing a negative trend of species richness. Still, as I mentioned above, further analysis is the next necessary step.