

Reviewer#1

This paper addresses a well-known tradeoff between development and environment related SDG dimensions, namely the impact of road infrastructure expansion on land cover change at tropical forest margins in the context of the Brazilian Amazon. The paper's conceptual model departs from von Thünen's theory of decreasing agricultural rents as distance from market centers increases. The authors expand the classic model by considering varying levels of prior development at forest frontiers along the distance gradient. This leads to the prediction (and empirical confirmation) that deforestation outcomes of road construction are highest at intermediate levels of development with implications for SDG sensitive infrastructure planning; for example, siting new roads in areas with high prior development levels can sometimes minimize the development-environment tradeoff. I have enjoyed reading the paper and recommend it for publication.

Thank you very much for this clear characterization and for your support.

Can we think of the distance-development gradient as a cross-sectional representation of the forest transition curve? If so, I am wondering what mechanisms are at play when the hypothesized inverted U-shaped relationship with deforestation is strong (e.g. national context characterized by low development, commodity export orientation, weak environmental governance) versus weak (e.g. high development, domestic market dependent agriculture, strong environmental governance). I suggest expanding the discussion section a bit to characterize settings in which the proposed conceptual model is most useful to inform transport infrastructure planning. My gut feeling is that a country's general development level (i.e. reliance on agriculture) and quality (including ambition) of environmental governance are the key mediators of the tradeoff relationship between road building and forest cover.

This is a good idea per understanding contexts. We have taken your suggestion to add to our Discussion. (While you make your point at a national scale, as we now point out too surely it also holds for regions.)

The Brazilian model of Amazon colonization differs quite a bit from that of other tropical forest margins, i.e. important primary roads where build to connect urban centers, including for geopolitical reasons, with suitability for agriculture often not being among the top priorities. What would be the implications of higher levels of endogeneity of roads' locations for the model and the spatial patterns depicted in Figure 2 – perhaps also something to expand on in the discussion.

We fully agree that the spatial nature of the process of development affects both impacts – as highlighted by our results – and how challenging it is to see the impacts. Your speculative process would result in an average road impact that could more closely resemble the high-prior-development result we get for roads, i.e., it could look as if all or average roads have little impact. In fact, a focus near cities, where as you say 'endogenous roads' spring up, is exactly the core issue that we believe drove one paper's average results. As you suggest, we added to our Discussion per how different processes affect impact and its estimation.

The regression equation could be justified with a bit more detail: Why is "prior deforestation" not entering the model in % terms like the dependent variable or vice versa? Why does "prior roads" enter in t-1 and "prior deforestation" in t? Given that the dependent variable contains deforestation in t, wouldn't a t-1 form be preferable? Does the subscript, "t, t-1" represent the variable in first difference?

Thank you for catching this lack of clarity. Indeed, prior deforestation is in % terms and involves the past. Further, we agree that the communication of the first difference was not clear. We shifted it to "t-1 to t".

What is the benefit of running a WLS regression as opposed to including the size of census tracts in an OLS approach? Doesn't this result in parameter estimates being biased towards relationships in large tracts? Also, using a dependent variable in % form may have unfavorable distributional implications? Why are you not using the first difference of absolute changes in deforested areas?

The weighting aims to focus on the reality of the average parcel in the region, as opposed to the average political unit. Very much along the lines of your speculation above – but in a mechanical way – the fact that political units are smaller and more numerous near cities can easily generate empirical results based upon realities of urban areas that already have high prior development. On the other hand, most forest is not actually in such areas, thus weighting by unit area better represents effects on average forest parcels. Stepping back to the big picture, our focus is to distinguish frontiers and then to provide results for each. Thus within our main results, we are not blending the most isolated areas with the most developed areas.

Concerning distributional implications, we may not yet follow which unfavorable issues you are referring to yet, without question, we agree that for a variable such as a percentage, one worries considerably about the denominator. For instance, while any given loss of forest is about the same fraction of forest within a 92% forested area versus a 91% forested area, it looks different as a fraction of a 2% versus a 1% forested area. Indeed, we have seen analyses in highly forested locations that suffer from such issues when using ‘the flip side of forest’, agricultural area. Any loss of forest area, considered as a gain in agricultural area, is a distortedly large fraction of a smaller initial agricultural area when studying the highly forested areas. Bringing this back to our analyses, we examine relatively small forest shifts and in highly forested areas. That is to say, we do not face the distortion of having small differences in small denominators dominate.

Please number equations

Thank you for this suggestion, we have now done so.

Regression tables: Please include standard errors in parenthesis (instead of p-values) and indicate significance levels using stars. Where robust or clustered standard errors calculated?

As you suggest, we have now adjusted Table 2, Table 3 and Table 4 to present them with standard errors, using stars to show significance levels. As our observational units are the census tracts, which are large, and as we are estimating tracts' impacts with county effects to strip out common elements for the tracts within a given county (those being quite large too), we had not felt that we needed to cluster the errors.

Some of the references are incomplete, e.g. Foster and Rosenzweig (2003)

With apologies, we have now fixed this too, checking all of the references.

Reviewer#2

The authors model the impacts of road construction on deforestation depending on prior development. This is a very interesting topic and from my point of view is worth to be published.

We thank you kindly for your support.

The authors justify their study as providing the means for a more sustainable road planning. However, I think that in areas where we have low prior development and according to their model rather low impacts on deforestation, deforestation might be more critical to the natural functioning of the ecosystem. In these areas, which are the very valuable in terms of ecosystem functioning, road effects (apart from deforestation) would be much more devastating than in the intermediate developed regions. This should be discussed at the end of the paper.

Absolutely we agree that interested parties are likely to apply different benefits and costs weights, by the location of benefits and costs. That point was our focus in 2.2 of the literature review. In our submission, we emphasized varied species densities. Now, however, to be more general, we also included your point.

Whatever weights one applies to impacts by location, our results concerning how much impact arises within each location are critical. But we could well imagine that, for instance, your speculation about higher forest damages in remote areas, along with higher political weights upon economic benefits in urban areas, could complement the patterns in our impacts results to suggest a spatial focus for roads. Following your suggestion, we made sure to communicate such combinations within our Discussion.

The paper would certainly benefit from some language editing and proof reading as there are some grammatical mistakes and incorrect or unclear sentences

With apologies, we have now tried to improve on this.

In the introduction the authors often mention tradeoffs but without explicitly naming the tradeoffs

We should be more clear and have now addressed this.

- “Rural roads may generate less of the kinds of benefits that tend to be counted officially, yet more of the benefits that tend not to be” Can you give an example?

That perspective comes from the van de Walle 2002 paper being reviewed within the paragraph you cite. The paper considers Vietnam and in particular rural road projects, as you highlight, and asserts that some significant social benefits arise that are hard to quantify and thus often omitted. Its emphasis is that while the idea of comparing costs and benefits is clear enough, implementation or measures is another matter. Given that, without question, this was not clear from what we wrote, we now call them “social benefits”.

Please use equation numbers and refer to them

Thank you for this suggestion, we have now done so.

In the deforestation formula some words are crossed out

That is a puzzle ... we believe you mean what is now numbered equation (3) and it should be ok now. We believe that it probably resulted from a grammar checker, on a machine different from that used to write. (While perhaps that could happen again, this time we have made sure to also send along a PDF version.)

Try to avoid starting sentences like “Table 4’s second column...” Make a statement what the results show, refer to the actual content of the columns and write (Table 4) behind the statement.

We have now fixed all these and it both communicates and flows better. Thank you for the suggestion.

Overall the captions of tables and figures are not self-explaining. Please provide more information here

Again with apologies, we have tried to improve these.

In the tables you write yes in many cells but it is unclear what that means? Is it telling us that it is significant? Please provide the info in the table header or/and more detailed values in the table cells

This did need to be clearer, since it was ambiguous per significance, as you say. We have updated this.

Figure 3: What does high or low variation mean? Variation of what?

As an example of lack of captions now we have indicated with Figure 3, but should have previously, that the predictions come directly from what is now the numbered equation (2). The term “variance” refers to the epsilon term within equation (2), an equation which was derived directly from the initial equation (1).