Reviewer number 2

Author Response: Thank you for the helpful comments. We have revised the paper in response to your comments. We are adding our responses in purple font in the document below. We have also removed any prior comment that is now resolved to simplify the document.

Author Changes: We have revised the paper in response to your comments. We include a point- by-point description of the changes below.

Thanks for detailed responses. The revised version is much improved. The response to one of the comments needs additional attention. In what follows I highlighted the comment which needs additional work. I also added a few minor new comments on the abstract. My responses are in red fonts in blue background

1. The model clearly uses a nesting logit format, perhaps three nests. Equation 1 of the paper shows only on nest. The formula should be replaced with a formula for the full nest.

Author Response: We have added information on how profit and shares are calculated for the other nests.

Author Changes: We have added additional text and an equation "In the three-level nest version, land allocation at each level is determined by a modified version of equation 1, where Y is replaced by the land allocated to that particular nest. The land allocated to a particular nest is dynamic and varies over time. Profit rates (r_j) at the lowest level of the nest are computed based on price, cost, yield, and subsidy (if included) for land use types (crops, pasture, commercial forest); profit rates for land cover types are input into the model and are based on the value of land. Profit rates for higher levels of the nest (r_{node}) are determined by: $r_{node} = [\sum_{j=1}^{n}(\lambda_j r_j)^{\rho}]^{1/\rho}$

The revision mentioned here is helpful, but it provides a broken formulation and misrepresent your work. To be fully transparent and to make sure that the readers understand that three exponent parameters (distribution parameters) and there sets of share parameters were estimated, I propose to use the following format:

$X_{jt}^{C} = \frac{(\lambda_{jo}^{c} r_{jt}^{c})^{\rho^{C}}}{\sum_j (\lambda_{jo}^{c} r_{jt}^{c})^{\rho^{C}}} \cdot Y_t^{C}$ where "C" represents the crop nest including rice, wheat, ……

$X_{jt}^{A} = \frac{(\lambda_{jo}^{A} r_{jt}^{A})^{\rho^{A}}}{\sum_j (\lambda_{jo}^{A} r_{jt}^{A})^{\rho^{A}}} \cdot Y_t^{A}$ where "A" represents the ag-forestry nest including cropland, forest, ….

$X_{jt}^{R} = \frac{(\lambda_{jo}^{R} r_{jt}^{R})^{\rho^{R}}}{\sum_i (\lambda_{jo}^{R} r_{jt}^{R})^{\rho^{R}}} \cdot Y_t^{R}$ where "R" represents the arable land nest including …..

The members of each nest should be clearly mentioned. In particular, for the crop nest all members of this nest including unused land, cropland pasture, CRP, other idled land should be specified. The information that presented in Table 2 is incomplete and very misleading. Of course, the names of variables should be described fully and maintain this convention through the paper.

Author Response: Please note that the equation did not transfer well from the PDF of the review and thus appears strangely in the comment and response above. We have used the original PDF with the correct formatting when reading and reacting to this comment. The same equation is used in all three places; given the confusion this seems to have caused, we have now repeated it

as suggested by the reviewer. Given the number of crop categories in gcamland, listing all crops in Table 2 made it difficult to read the table, so we have opted to provide the full list in a footnote. We also include a complete nesting diagram with all information in the same figure in the SM. To facilitate readability, we have opted to use plain language descriptors in the paper rather than the abbreviated terminology used with the gcamland code. However, we do agree that we need to specify the mapping between plain language and terminology, so we have added a table to the SM.

Author Changes: We have repeated the logit equation as suggested by the reviewer. We have added information on what is contained in the nests, using the node names (now defined in Table S2) in the main text and the complete list of crops in footnote 6. We have added a nesting diagram to the SM (Figure S1) which indicates which members are in which nests. We have added a table to the SM (Table S2) that maps the plain language descriptors used in the main text of the paper to the precise terminology in the gcamland model.

Sorry that the pdf file including my comments was not presenting the suggested formulas correctly. The pdf version that I posted on the Journal website shows the formulas correctly. Anyway, I see that you correctly specified and defined the proposed equations. Thanks for that and thanks for adding Figure S1 and Tables S2. Your work is now more transparent and traceable.

However, I see that table S2 in defining A, R, and C refers to equation 1 which is wrong. It should refer to equation 2.

Figure S1 represents 7 nests for the supply structure of gcamland. On the other hand, Table S2 shows that your calibration process estimates distribution parameters for only three of them. The rest (including 4 distribution parameters) were given some ad hoc values. For the root nest, it makes sense to use a zero value, as you assumed no change in the areas of members of this nest. However, it is not clear from where the values of 2.7, 0.05 and 1.575 are come from. Explain how those values were determined? The paper should explain the sources of these ad hoc values.

Author Response: We have corrected the reference to equations in table S2. For the other logit exponents, we have used the default values used in GCAM. We have chosen not to vary these in this exercise as they do not directly impact the amount of cropland area, which was the output we focused on in the main text. We have done an additional sensitivity analysis quantifying the impact of these logits on cropland area. In this analysis, we doubled each logit one at a time. The area allocated to each crop changes by less than 1% (the largest change in magnitude is -0.12%). These logits do have a larger impact on other land types. For example, doubling the ForestLand logit results in a shift in the distribution of commercial and non-commercial forest, with commercial forest increasing by as much as 27%. However, total forest is largely unchanged (maximum change of 0.22%). We have added this information to the supplementary material, near Table S2.

For the default values, we realize in re-reading the previous GCAM papers that while we have documented the approach to selecting logit exponents in Wise et al. (2014) and the specific numbers used are available on GitHub (github.com/jgcri/gcam-core) we have not documented specifically how those numbers were chosen. We have now added this to the supplementary material of this paper.

Author Changes: We have corrected the reference to equations in table S2. We have clarified which logit exponents were varied (and which were not) in the methodology section. We have

also added a discussion of the other logit exponents, including why we did not vary them, how the default values were chosen, and the effect of changing those values on the outputs used in this study to the supplemental material:

"Table S2 provides information the gcamland nodes, the total land area for each node in 1990, and logit exponents used in this study. As noted in the main text, three of the logit exponents used in gcamland are varied as part of the analysis in this paper. For the remaining logit exponents (root, Pastureland, Grass/shrubs, Forestland), we use the default values used in GCAM. These values were chosen based on heuristics, where larger values are used for land types that are more substitutable. For the root, this is set to zero, as we do not allow conversion into or out of urban, tundra, or rock/ice/desert. For grass/shrubs, the decision to shift between grassland and shrubland is unlikely to be an economic choice; for this reason, we set the logit exponent to a very low value, effectively preserving the shares of grassland vs shrubland in the initial model year. Both the Forestland and Pastureland logit exponents govern substitution between commercial and non-commercial land types; a shift between these land types is not a land conversion (i.e., it does not require re-planting) but a shift in use (i.e., either moving livestock or engaging in logging activities). For this reason, higher logit exponents are chosen. A higher logit exponent governs Pastureland than Forestland as the shift in use of pastureland is likely to be easier than the change in use of Forestland. We have chosen not to vary these in this exercise as they do not directly impact the amount of cropland area, which was the output we focused on in the main text. We have done an additional sensitivity analysis quantifying the impact of these logits on cropland area. In this analysis, we doubled each logit one at a time. The area allocated to each crop changes by less than 1% (the largest change in magnitude is -0.12%). These logits do have a larger impact on other land types. For example, doubling the ForestLand logit results in a shift in the distribution of commercial and non-commercial forest, with commercial forest increasing by as much as 27%. However, total forest is largely unchanged (maximum change of 0.22%)."

Thanks for the additional sensitivity test. I appreciate your efforts. I also appreciate your note which says you have "not documented specifically how those numbers were chosen". Your response also says: "We have now added this to the supplementary material of this paper". I see that you tried to explain and justify the role of these parameters in explaining Table S2. But you ignored to clearly mention that those values are ad hoc values. I checked Wise et al. (2014). This reference says nothing regarding the selected values under discussion. You have to clearly add the following phrase, or something similar, to the main manuscript not the supporting document:

"The logit values assigned to pastureland, grass/shrubs, and forest land nests have not been obtained from an explicit statistical approach nor a calibration process. They have been selected based on the authors' value judgment".

Additional new comments on abstract:

1) The abstract says: "In this study, we demonstrate a more systematic and empirically-based approach to estimating model parameters for an economic model of land use and land cover change, gcamland".

   I propose the following modification as you only calibrate (not estimate) a few parameters assuming other model parameters are valid:

   "In this study, we demonstrate a more systematic and empirically based approach to calibrating a few selected parameters of an economic model of land use and land cover change, gcamland.

2) The abstract says: “we generate a large set of model parameter perturbations and run gcamland simulations with these parameter sets over the historical period in the United States to quantify land use and land cover, determine how well the model reproduces observations, and identify parameter combinations that best replicate observations”.

I propose the following modifications as you calibrate a small set of parameters assuming other parameters are valid. It is important to say what parameters you calibrated. Your paper is all about it, but the abstract says nothing about that.

“we generate a large set of model parameter perturbations on the selected parameters and run gcamland simulations with these parameter sets over the historical period in the United States to quantify land use and land cover, determine how well the model reproduces observations, and identify parameter combinations that best replicate observations, assuming other model parameters are valid. In particular, 3 parameters out of 7 parameters that govern land allocation in gcanland were calibrated only for the case US.