APPENDIX 3 Sensitivity analysis methodology and additional results

A3.1 Parameter sampling

We conducted a global sensitivity analysis on the majority of the parameters of the model (see Table A1.1 in the ODD+D description for the selected parameters). To generate perturbed parameter sets we employed the following procedure:

1. Generate a random deviation a_i for each of the P parameters ($\boldsymbol{a} = a_1, \ldots, a_P$), allowing the deviation to be 30% upwards or downwards: $\boldsymbol{a} \sim U(0.7, 1.3)^P$
2. Perturb each parameter from its baseline value X_i ($\boldsymbol{X} = X_1, \ldots, X_P$) by this simulated value, giving a perturbed parameter set: $\boldsymbol{S}'_r = \boldsymbol{aX}$
3. Repeat this procedure 10,000 times, giving $\boldsymbol{S}' = \boldsymbol{S}'_1, \ldots, \boldsymbol{S}'_{10000}$. Here, we used latin hypercube sampling to increase the efficiency of the sampling of the parameter space.

A3.2 Model evaluation

For each set of perturbed parameters $\boldsymbol{S}'_r$ calculate the Quantity of Interest (QoI), where the QoI takes two forms:

(a) QoI_{shock} represents $P(CC \succ Ins)^{shock}$ in Experiment 1 (Table 1) with $T_{assess} = 5$ and $T_{shock} = 10$ and a 10% shock.
(b) QoI_{pov} represents $P(CC \succ Ins)^{pov}$ in Experiment 2 (Table 1) with $T_{pov} = 50$.

The model evaluation procedure results in a "dataset" of sorts, where the independent variables are the parameters ($\boldsymbol{S}'$, with P columns and 10,000 rows) and the dependent variable is the quantity of interest (QoI_{pov} or QoI_{shock} of size 10,000).

A3.3 Gradient-boosted regression forest

The goal of the sensitivity analysis is to assess how changes in the parameters affect the QoI. Hence, we are interested in exploring the function f in the relationship $QoI = f(\boldsymbol{S}')$. This function may be non-linear. We trained a gradient-boosted regression forest (GBRF) to yield a non-parametric representation of f. A GBRF consists of a set of simple regression trees that are fit in a stagewise manner, with each successive tree being fit to the residuals of the previous. GBRFs originated in the machine learning community, and generally exhibit a high predictive performance (Elith et al. 2008). We do not discuss this method in detail here and refer interested readers to Elith et al. (2008).

A3.4 Assessing variable influence

We use partial dependence plots (PDPs) – a common visualization technique for non-parametric models – to visualize the associations between changes in each parameter and the QoI, as assessed by the GBRF. Each point (x, y) on a partial dependence plot for parameter p_i represents the average prediction made by the GBRF (y value) if every instance of p_i is set to x, keeping all other parameters (p_{-i}) at their original values. The slope of the PDP gives an indication of both the magnitude and direction of influence of the parameter on the QoI. A PDP for a linear regression model would show a straight line representing the regression coefficient (β). To generate confidence bounds on our PDPs we bootstrap the "dataset" 100 times, each time re-training the GBRF and re-estimating the PDP.

A3.5 Supplemental results

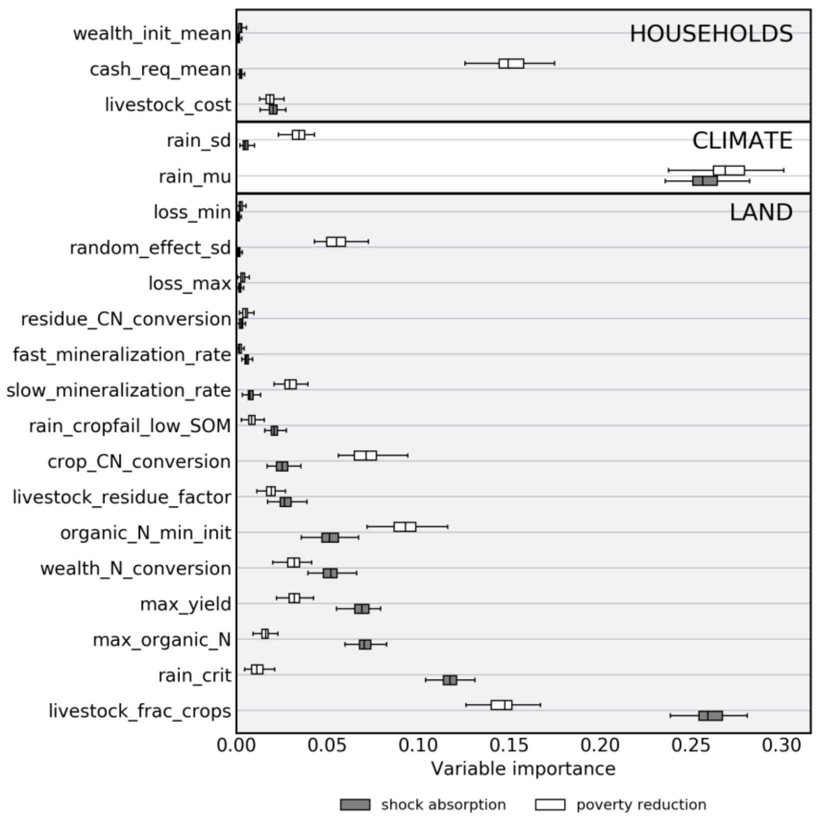

Figure A3.1: Importance of different model parameters in the sensitivity analysis, as calculated by the GBRF. The "variable importance" measure is calculated by scikit-learn in Python (Pedregosa et al. 2011) and is a measure of the amount of variance that each variable explains.

References

Elith, J., J. R. Leathwick, and T. Hastie. 2008. A working guide to boosted regression trees. *Journal of Animal Ecology* 77(4):802–813.

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12:2825–2830.