



Predicting household resilience with machine learning: preliminary cross-country tests

Alessandra Garbero¹ · Marco Letta²

Received: 24 July 2021 / Accepted: 3 January 2022
© The Author(s) 2022

Abstract

Using a unique cross-country sample from 10 impact evaluations of development projects, we test the out-of-sample performance of machine learning algorithms in predicting non-resilient households, where resilience is a subjective metrics defined as the perceived ability to recover from shocks. We report preliminary evidence of the potential of these data-driven techniques to identify the main predictors of household resilience and inform the targeting of resilience-oriented policy interventions.

Keywords Resilience · Machine learning · Classification · Targeting · Predictive analytics

JEL Classification C52 · O12

1 Introduction

Following a surge in the interest for the so-called “prediction policy problems” (Kleinberg et al. 2015), the literature on the use of machine learning (ML) in economics and public policy studies is rapidly expanding (Athey 2018; Athey & Imbens 2018; Kleinberg et al. 2018; Mullainathan & Spiess 2017). In parallel, the notion of resilience, defined as the capacity over time of individuals, households, or communities to withstand a myriad of shocks and stressors, is becoming a central paradigm in the development agenda. Its theoretical underpinnings, as well as different empirical methodologies for its measurement, have lately been validated in several scientific

✉ Marco Letta
marco.letta@uniroma1.it

Alessandra Garbero
a.garbero@ifad.org

¹ Research and Impact Assessment Division, International Fund for Agricultural Development, Rome, Italy

² Department of Social Sciences and Economics, Sapienza University of Rome, Rome, Italy

articles belonging to the so-called ‘development resilience’ literature (see, among many, Barrett & Conostas 2014; Brück et al. 2019; Cissé & Barrett 2018; d’Errico et al. 2019; d’Errico et al. 2020; Smith & Frankenberger 2018). This paper is placed at the intersection between these two strands of research.

A separate and nascent body of empirical work has started testing the potential of ML in predicting well-being measures. In development economics, ML has been lately applied to predict and map poverty (Blumenstock et al. 2015; Jean et al. 2016; Kshirsagar et al. 2017; McBride & Nichols 2018; Perez et al. 2019; Steele et al. 2017) as well as food security (Ganguli et al. 2019; Hossain et al. 2019; Lentz et al. 2019) outcomes, highlighting the great potential of these predictive tools to improve the old problematic issue of the (in)effective targeting of development programmes.

A recent and comprehensive review of the flourishing literature devoted to the conceptualization and measurement of development resilience is carried out by Barrett et al. (2021). The authors first highlight the three main conceptualizations of development resilience: (i) resilience defined as a capacity, e.g. the “capacity that ensures stressors and shocks do not have long-lasting development consequences” (Conostas et al. 2014) and is captured as a latent and multidimensional variable combining observable and unobservable features (Alinovi et al. 2008, 2010; Brück et al. 2019; d’Errico et al. 2020; d’Errico & Di Giuseppe 2018; Smith & Frankenberger 2018); (ii) resilience as a normative condition, i.e. the probability of achieving some minimal standard living conditional of many observable characteristics and exposure to shocks (Barrett & Conostas 2014), which implies that resilience is treated as an outcome in impact evaluation studies (Knippenberg et al. 2019; Upton et al. 2016); (iii) resilience as return to equilibrium in the aftermath of a shock, where the focus is on the ex-post effects of the shocks experienced on some well-being outcomes (Conostas et al. 2014; Hoddinott 2014; Knippenberg et al. 2019).

Then, they provide an overview of the empirical quantitative literature on resilience, emphasizing several limitations of current approaches involving theoretical, empirical, and data-related constraints. Importantly, among these, Barrett et al. (2021) stress that concerns have been raised about the current ability of the most popular methodologies for resilience measurement described above (that do *not* make use of ML techniques) in accurately predicting outcomes out-of-sample. This is a task that ML models, which are built to excel at predicting outcomes (Varian 2014), can, in principle, accomplish. Indeed, scholars have recently raised a call to harness the opportunities provided by machine learning algorithmic procedures to identify better predictors of resilience, predict and highlight the presence of vulnerability hotspots (Jones et al. 2021) and, in turn, improve the design of effective early warning mechanisms (McBride et al. 2021). To the best of our knowledge, however, only one paper to date has investigated how ML methods can predict household resilience, e.g. the contribution by Knippenberg et al. (2019). As part of a broader empirical exercise involving a comparison among different methodologies, the authors of this study apply two ML techniques, namely the Least Absolute Shrinkage and Selection Operator (LASSO) and random forests, to identify the best predictors of a resilience measure based on the Coping Strategy Index of Malawian households.

We build on their pioneering work by providing preliminary cross-country evidence on the potential of ML to improve the study of household resilience as well as the

targeting of policy interventions. Importantly, we are interested in accurately predicting resilience status in addition to identifying its best predictors. For this reason, unlike Knippenberg et al. (2019), we tackle resilience prediction as a classification problem rather than a regression one. In addition, we focus on a cross-country context and use a different proxy for household resilience and a broader set of ML algorithmic routines.

Leveraging a large dataset spanning 10 countries and data-driven resilience prediction via ML, we show that: (i) ML algorithms perform well even when studying households from very different contexts and with a limited amount of widely available information; (ii) simpler algorithms perform almost as well as ‘black-box’ methods (i.e. complex predictive techniques that do not produce an understandable model and are thus characterized by scarce or null explainability) and may be preferable because of their transparency and interpretability.

The results shed light on the predictive potential of ML to both improve the allocation of projects’ funding and better target resilience-oriented policy interventions to those most in need, which would, in turn, maximize the beneficial effects of these development policies.

The rest of this paper is structured as follows. Section 2 presents the data and the machine learning approach. Section 3 reports the results of the empirical analysis. Section 4 discusses the main policy implications and concludes.

2 Data and methods

2.1 Data

Our dataset is composed of cross-sectional household-level surveys from micro-level impact evaluations fielded by the International Fund for Agricultural Development (IFAD).¹ These impact assessments evaluated a selection of the Fund’s development projects, closing between 2016 and 2018. Among these studies, we only selected those with available comparable resilience metrics and socio-economic characteristics. This led us to a final sample of 10 countries for more than 14,000 households observations. All the data come from cross-sectional surveys, with the partial exception of the PASIDP-I project in Ethiopia.² The list of projects included in our dataset is listed in Table 3 in the Annex.

Concerning the outcome variable, we employ a subjective metric of resilience, i.e. the ability to recover from shocks (ATR). This metric is constructed based on answers to the following question: “To what extent were you and your household able to recover from shock x ?”. ATR thus represents a self-assessment from the interviewed households and takes the form of a categorical variable which ranges from 1 to 5 according to the following scale:

- a. Did not recover (= 1).
- b. Recovered to some extent, but worse off than before (= 2).

¹ For more information, see <https://www.ifad.org/en/web/knowledge/publication/asset/41489584>.

² PASIDP-I has a longitudinal component since four rounds of data in a 1-year rotating panel have been collected under this project. However, for the purpose of this paper and since all the other data in our pooled cross-country dataset are cross-sectional, we ignore this longitudinal component in our analysis.

- c. Recovered to the same level as before (= 3).
- d. Recovered, and better off than before (= 4).
- e. Experienced the shock but was not significantly affected (= 5).

This question is asked repeatedly for a roster of several different x shocks (droughts, floods, crop diseases, etc.) that the households might have experienced in the last year prior to the survey. We first take an average of the ATR for all shocks experienced by the household and obtain an average ATR for each household. In the following step, we create a binary outcome variable to discriminate between resilient vs non-resilient households. This dummy variable takes value 1 if the average household ATR is below the sample mean and 0 otherwise. A value of one thus indicates non-resilient households and a value of 0 resilient ones.

The use of a binary outcome is dictated by our preference to tackle the resilience prediction problem as a classification one. Our assumption is that discriminations above or below clear cut-offs are more intuitive for practitioners, policymakers and humanitarian agencies that aim at efficiently targeting their policy interventions, and therefore predicting cut-offs rather than continuous values is more useful for practical purposes (Lentz et al. 2019). The choice of a subjective resilience metric is driven by: (i) data availability; (ii) the assumption that households are in the best position to assess the extent of shock impacts on their welfare and their post-shock recovery, as well as existing evidence that self-reported measures of well-being go hand in hand with objective indicators (Knippenberg et al. 2019); (iii) the increasing use in recent studies of subjective approaches and self-evaluations as resilience metrics which represent valid alternatives to objective indicators (Jones & Tanner 2017; Jones & d'Errico 2019).

As far as the features that may predict resilience are concerned, we employ a set of 14 predictors whose list, summary statistics, and details are reported in Table 4 in the Appendix. These are the most relevant variables that were common and comparable across all the surveys in our pooled dataset and include demographic characteristics, income measures, asset-based indices, food security proxies, and shock exposure metrics, represented by the number of shocks experienced and their perceived severity.

Importantly, we do not provide the algorithms with information about the country, region, district, or village of origin of our households, for three reasons: (i) our samples are not nationally representative from a geographic point of view; (ii) these geographic dummies would not provide any useful information for targeting as we aim to scale up these projects in other contexts; (iii) we are interested in providing useful insights based on generalizable socio-economic and demographic characteristics, not in identifying resilience clusters derived from geographically non-representative samples.

Finally, as some variables had a small number of missing observations, and since some machine learning algorithms handle missing variables differently, we imputed missing values via proximity through a random forest algorithm to make the results comparable across different methods.³

³ We use the *rflmpute* package in R. In any case, our findings are insensitive to the imputation method we employ. The results for the non-imputed data are available upon request.

2.2 Methods

We focus on a purely predictive problem, the prediction of non-resilient households. We are thus interested in minimizing the predictive error on previously unseen data (the so-called ‘test error’), not in the causal impact of any of the features.

To this aim, we employ supervised ML techniques. Machine learning is a subfield of artificial intelligence. ML algorithms have been developed in computer science and statistical literature to deal with predictive tasks (Varian 2014). The aim of ML techniques is to minimize the out-of-sample prediction error and generalize well on future data (Athey and Imbens 2019; Mullainathan and Spiess 2017). Supervised ML involves building a statistical model for predicting an output based on one or more inputs (Lantz 2019).

The standard ML routine is to randomly split the original sample into two disjoint sets: the training set, on which ML algorithms are trained, and the testing set, which is used to evaluate the predictive ability of ML models on previously unseen data. This introduces the so-called ‘firewall’ principle: none of the data involved in generating the prediction function is used to evaluate it (Mullainathan and Spiess 2017). The out-of-sample performance of the model on the unseen held-out data then constitutes a reliable and generalizable measure of the ‘true’ performance of the models on future data. Following this scheme, we randomly split our dataset in a training set, consisting of 2/3 of the whole sample, and a testing set, composed of the remaining 1/3.

We test the performance of five supervised ML algorithms: classification trees; two ensemble methods based on decision trees, namely bootstrap aggregating (bagging) and random forests; k-nearest neighbour (k-NN); and support vector machine (SVM). These techniques are characterized by different degrees of flexibility and complexity, ranging from the simpler classification tree to black-box models such as SVM and random forest. Higher flexibility comes at the cost of a loss of interpretability. With the exception of classification trees, none of the other methods produces readily interpretable, easy-to-explain outputs to understand how the features are related to the class.

Classification trees are based on recursive partitioning, also known as the ‘divide and conquer’ approach (Lantz 2019). Via recursive binary splitting, the tree is grown by repeatedly splitting the data into smaller and smaller subsets until sufficient within-subset homogeneity or a stopping criterion is reached. As trees can suffer from high variance, i.e. they are quite sensitive to small changes in the training sample and prone to overfitting, we also apply bagging and random forest to our classification problem. These ensemble methods build x trees from x bootstrapped training sets and take a majority vote among the x predictions (Hastie et al. 2009). The difference is that for each split in the trees, bagging considers all the features as split candidates, whereas each time a split is considered, random forest randomly subsamples m out of all the p features as candidates each time, thus introducing additional layers of randomness that further decorrelate the trees. k-NN is similar to non-parametric analysis and uses information about an example’s k -nearest neighbours to classify unlabelled examples. For each observation in the testing set, the algorithm identifies the k closest observations

from the training sample and assigns a prediction on the basis of a majority rule, taking as prediction the most frequent outcome among those of the nearest neighbours. Finally, SVM creates a boundary called hyperplane to divide the multidimensional feature space into homogeneous partitions and is able to model highly complex relationships. For all our algorithms, we use tenfold cross-validation on the training data to tune key hyperparameters and solve the bias-variance trade-off.⁴

The number of observations is 9420 households in the training sample, of which 47.3% are resilient and 52.7% non-resilient; and 4854 in the testing sample, of which 47% resilient and 53% non-resilient. After training and tuning the algorithms on the training sample, we evaluate out-of-sample performances in the testing set via confusion matrices in which we compare the predicted and actual values of our binary outcome, resilience status.

3 Results

The results are reported in Table 1. All the algorithms have an accuracy rate above 72% and an even higher sensitivity. Sensitivity is the proportion of actual positives correctly identified and is the metrics we are mostly interested in. For all the algorithms, sensitivity is close to or around 80%. Classification trees perform comparatively well, especially in terms of sensitivity. More complex methods based on decision trees, bagging, and random forest perform better than the tree for all the metrics, i.e. specificity (the proportion of actual negative cases, $y = 0$, correctly identified), sensitivity, and overall accuracy, but not significantly so. As for the other two ‘black-box’ methods, k-NN performs slightly worse than the tree in terms of the accuracy rate but leads to a higher sensitivity, while SVM performs better than the tree but worse than bagging and random forest. Overall, the random forest is the best-performing algorithm.

The classification tree is illustrated in Fig. 1. Five features appear: the (perceived) mean severity of shocks,⁵ total gross income, the Household Dietary Diversity Score (HDDS), the agricultural asset index, and household size. Combinations of these five variables produce the tree represented in Fig. 1. For example, if the severity of shocks is higher than 3.4 and household size is lower than 13, the algorithm predicts the household as non-resilient.⁶ Conversely, if the perceived severity of shocks is less than 3.4, the resilience status depends on interactions between additional variables other

⁴ For the classification tree, we use cross-validation to implement cost complexity pruning, i.e. prune the tree and select the best complexity parameter cp . For k-NN, cross-validation is run to derive the optimal neighbour number k . For random forest, we employ cross-validation to select the best value for the parameter m , that is, the number of variables randomly sampled as candidates at each split. For support vector machine, we use the Gaussian Radial Basis Function (RBF) kernel and the best cross-validated sigma (the inverse kernel width) and cost of constraints violation C . For bagging and random forest, we grow a total of 1000 trees. Finally, since the k-NN and SVM are sensitive to the scale of the data, we normalize all the features and run the two algorithms on the standardized data.

⁵ As reported in Table 4, the mean severity of shocks is the average of a self-reported categorical variable indicating the impact level of each shock experienced by the household.

⁶ Above the severity threshold of 3.4, households with a size larger than 13 have a significantly higher income and experience fewer shocks and less severe shocks, compared to households with fewer than 13 members. The algorithm is thus likely identifying these underlying differences.

Table 1 Out-of-sample performance

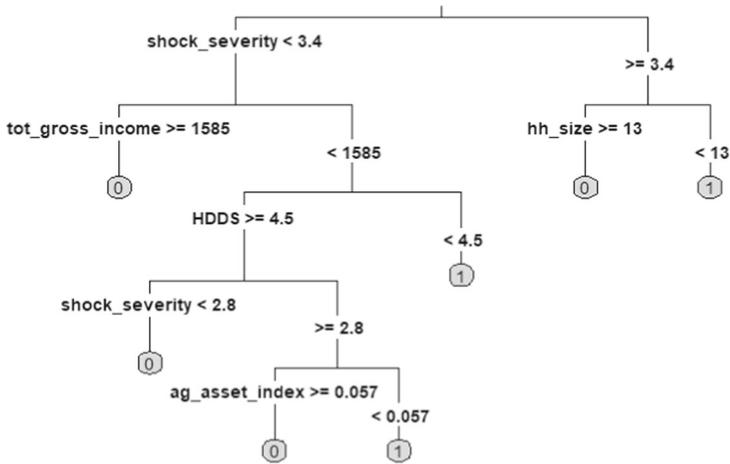
		Real status		
		Resilient	Non-resilient	Total
Classification tree				
Predicted status	Resilient	1535	567	2102
	Non-resilient	746	2006	2752
	Total	2281	2573	4854
	Correctly predicted	67.3%	77.8%	73%
Bagging				
Predicted status	Resilient	1557	525	2075
	Non-resilient	724	2048	2779
	Total	2281	2573	4854
	Correctly predicted	68.3%	79.6%	74.3%
Random forest				
Predicted status	Resilient	1597	502	2109
	Non-resilient	684	2071	2745
	Total	2281	2573	4854
	Correctly predicted	70%	80.5%	75.6%
k-NN				
Predicted status	Resilient	1483	529	2012
	Non-resilient	798	2044	2842
	Total	2281	2573	4854
	Correctly predicted	65%	79.4%	72.7%
Support vector machine				
Predicted status	Resilient	1540	532	2068
	Non-resilient	741	2041	2786
	Total	2281	2573	4854
	Correctly predicted	67.5%	79.3%	73.8%

Bold is arguably the most important part of the table

than shock severity, such as income, food security, and agricultural asset index. For instance, if shock severity is less than 3.4, but total gross income is equal to or higher than 1585 dollars per year, the household is predicted as resilient. If income, instead, is lower than this threshold, and the HDDS is lower than 4.5, the household is predicted as non-resilient. This assignment mechanism goes on until all the observations are placed within one of the nodes.

While no interpretable output is available for k-NN and SVM, bagging and random forest provide a ranking of the predictors. We report the five most important variables according to these algorithms in Table 2.

The score assigned to each variable represents the mean decrease in the Gini Index if that specific variable is excluded from the model. Both bagging and random forest are in agreement with the tree about the predominant importance of the *severity of shocks* and *household income*. The *agricultural asset index* also appears in the top



Legend

shock_severity	Mean severity of the shocks
tot_gross_income	Total gross household income (constant 2018 USD)
HDDS	Household Dietary Diversity Scale
ag_asset_index	Agricultural asset index
hh_size	Household size

Fig. 1 Classification tree for non-resilient household

Table 2 Top 5 most important variables—Ensemble methods

Bagging		Random Forest	
Variable	Mean decrease in Gini Index	Variable	Mean decrease in Gini Index
Mean severity of shocks	901.26781	Mean severity of shocks	831.12757
Total gross income	745.54083	Total gross income	614.47379
Total cultivated land	411.78624	Total cultivated land	419.93563
Asset Index	392.71532	Asset Index	413.41640
Agricultural Asset Index	392.51322	Agricultural Asset Index	394.58717

five. Differently from the tree, bagging and random forest assign a high score to total cultivated land and the household asset index, whereas the HDDS and household size rank lower and are not amongst the most important variables.

In sum, households experiencing more severe shocks and endowed with low levels of income and assets tend to be predicted as non-resilient. The fact that the inability to withstand shocks is associated with such features is of course not unexpected, but it is remarkable that based on such a limited amount of information, the algorithms correctly identify up to four-fifths of previously unseen non-resilient households without even knowing the country, region, district, or village of origin of each household. In turn, this makes data-driven resilience prediction via machine learning an appealing tool for targeting and policy purposes, especially in data-scarce environments that are a frequent and recurrent feature of many developing contexts.

4 Implications and conclusions

Can machine learning be leveraged to predict household resilience? As there is empirical evidence demonstrating that the most common resilience measures have limitations in predicting well-being out-of-sample (Barrett et al. 2021), we deem this a particularly important question.

In this paper, we perform simple and preliminary tests to show that supervised machine learning algorithms can be successfully employed to predict household resilience status as well as identify the main features that drive such predictions. ML techniques were able to identify over three-quarters of the observations and four-fifths of the non-resilient households. We reckon that this is a noteworthy performance, considering that we did not provide the algorithm with the country of origin or other non-generalizable geo-information. The variables we use as features, in fact, are widely available in most of the micro-surveys from developing contexts. The cross-country nature of our dataset provides more external validity to our findings than predictive studies based on a single-country sample.

The implications for policy targeting are evident: policy interventions in the aftermath of covariate shocks such as conflicts, natural disasters, or economic crises could exploit the potential of these techniques to more accurately target non-resilient households based on the features identified and the thresholds indicated as part of the classification algorithm. Specifically, by providing a specific assignment rule determined by the analyses in question, policy implementers can improve the allocation of financial resources by better targeting resilience-enhancement interventions. This would eventually maximize the potential of these development policies to generate beneficial impacts (the so-called ‘treatment effects’) for the most affected portions of rural populations.

Central to the debate on resilience-enhancing development projects is how to effectively target the less resilient with policies that can boost adaptive capacity. The implications of a simple ML predictive exercise such as the one we have conducted in this study suggest that policy implementers could exploit the potential of ML to improve the allocation of projects resources and better target resilience-enhancement interventions to those most in need. This, in turn, would eventually maximize the

potential of these development policies to generate beneficial effects for the most affected portions of rural populations. In addition to their potential to ‘fine-tune’ targeting mechanisms, this type of ML-based predictions can also be employed to refine early warning system development (McBride et al. 2021).

For these policy purposes, ML methods that indeed provide a clear, intuitive, and straightforward assignment mechanism or targeting rule, such as classification trees, may be preferred because of their intrinsic simplicity and resemblance to human decision-making, especially when more complex, black-box methods do not perform significantly better, as was the case in our study.

Our work provides new insights on a key notion in development economics by proposing an empirical approach to tackle the identification of household resilience as a “prediction policy problem” (Kleinberg et al. 2015). While our preliminary evidence provides empirical support to recent calls to leverage ML tools to shortlist variables for targeting purposes and highlight hotspots of vulnerability (Jones et al. 2021; Knippenberg et al. 2019), it is far from being conclusive on the matter.

Further work should compare the performance of ML on subjective resilience with the one on objective metrics. More generally, many different resilience approaches and several cut-offs could and should be tested under the ML lens. A comparison of classification approaches with numeric prediction methods, such as regression trees, can also provide valuable insights, especially on the consistency of the best resilience predictors across different models and methodologies. Another crucial test is to check for the stability of ML-based prediction accuracy, which can be a weakness of ML models (in tracking resilience outcomes over time, especially by using high-frequency longitudinal data, along the lines of Knippenberg et al. (2019)). Finally, it is key to shed light on the effectiveness and accuracy of actual targeting rules of closed resilience-oriented projects through a comparison with ML predictions in rigorous *ex-post* targeting evaluation exercises. All these key issues are deferred to future research.

Acknowledgements We are grateful to the Editor Robert M. Kunst and three anonymous referees for their helpful suggestions and constructive advice. We also thank Emmanuel Flachaire, Lisa Jäckering and Grayson Sakos. This paper was written as part of a project funded under an Innovation Challenge grant by IFAD. The views expressed in the paper are the authors’ ones and do not represent those of the institutions to which they are affiliated.

Author contributions All the authors contributed equally to the research presented in this paper.

Declarations

Conflict of interest This paper was written as part of a project funded under an Innovation Challenge grant by IFAD. The views expressed in the paper are the authors’ ones and do not represent those of the institutions to which they are affiliated. The authors declare that they do not have neither conflicts of interest nor financial or non-financial interests. Finally, this paper is not currently under consideration at any other journal.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line

to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix

See Tables 3 and 4.

Table 3 IFAD Impact Assessment sources

Project name	Country	Year of data collection
Coastal Climate Resilient Infrastructure Project (CCRIP)	Bangladesh	2018
Gente de Valor: Rural Communities Development Project in the Poorest Areas of the State of Bahia	Brazil	2018
Rural Development Support Programme in Guéra (PADER-G)	Chad	2017
Guangxi Integrated Agricultural Development Project (GIADP)	China	2017
Participatory Small-Scale Irrigation Development Programme I (PASIDP I)	Ethiopia	2016–2017
Coastal Community Development Project (CCDP)	Indonesia	2018
Community-based Forestry Development Project in Southern States (DECOFOS)	Mexico	2017
High Value Agriculture Project in Hill in Mountain Areas (HVAP)	Nepal	2018
Participatory Smallholder Agriculture and Artisanal Fisheries Development Programme and Smallholder Commercial Agriculture Project (PAPAFPA/PAPAC)	Sao Tomé & Príncipe	2018
Agriculture Value Chain Support Project (PAFA)	Senegal	2017

Table 4 Summary statistics – Whole sample

Variables	Mean	Var	sd	Obs
<i>Outcome</i>				
Resilience status (0 = resilient; 1 = non-resilient)	0.528	0.249	0.499	14,274
<i>Features</i>				
Treatment	0.489	0.250	0.500	14,274
Age of the household head	48.30	198.9	14.10	14,274
Gender of the household head	0.157	0.132	0.363	14,274
Education level of the household head	0.971	0.796	0.892	14,257
Household size	5.740	13.07	3.616	14,274
Dependency ratio	0.982	0.942	0.971	14,099
Total gross income	2501.3	36,344,330.6	6028.6	14,272
Gross crop income	458.1	2,944,285.9	1715.9	14,274
Total cultivated land	10.72	23,512.5	153.3	13,866
Asset index	0.146	0.0222	0.149	14,268
Agricultural asset index	0.134	0.0274	0.165	14,268
Household Dietary Diversity Score	7.016	6.843	2.616	14,255
Number of shocks	2.412	5.516	2.349	14,253
Mean severity of shocks	3.111	0.958	0.979	14,261

'Resilience status' is a binary variable taking value 1 if household average ability to recover is below the sample mean and 0 otherwise. 'Treatment' is a dummy taking value 1 if the household was in the treatment group and 0 otherwise. 'Gender of the household head' is a dummy taking value 1 if the household head is female and 0 otherwise. 'Education level of the household head' is a categorical variable which can take the following values: 0 = no education; 1 = primary education; 2 = secondary education; 3 = higher education. 'Total gross income' and 'Gross crop income' are annual measures expressed in constant 2018 US dollars. 'Total cultivated land' is measured in hectares. The 'Household Dietary Diversity Score' ranges from 0 to 12 and has a reference period of 7 days for all the country samples except China, for which a reference period of 1 day is used. 'Asset Index' and 'Agricultural Asset Index' are standardized measures of assets which range from 0 to 1 and have been generated for each country sample via factor analysis, using exclusively the assets that were common across all the datasets. 'Mean severity of shocks' is the household average, for all shocks, of a self-reported categorical variable indicating the impact of each shock experienced as assessed by the household. The variable ranges from a score of 1 to 5 as follows: No impact = 1; Slight impact = 2; Moderate impact = 3; Strong impact = 4; Worst ever happened = 5

References

- Alinovi L, Mane E, Romano D (2008) Towards the measurement of household resilience to food insecurity: applying a model to Palestinian household data. *Deriv Food Secu Inf Natl Househ Budg Surv Food Agric Org U N Rome Italy* 137–152
- Alinovi L, D'errico M, Mane E, Romano D (2010) Livelihoods strategies and household resilience to food insecurity: an empirical analysis to Kenya. *Eur Rep Dev* 1–52
- Athey S (2018) The impact of machine learning on economics. *The economics of artificial intelligence: an agenda*. University of Chicago Press, Chicago, pp 507–547
- Athey S, Imbens GW (2019) Machine learning methods that economists should know about. *Ann Rev Econ* 11:685–725

- Barrett CB, Conostas MA (2014) Toward a theory of resilience for international development applications. *Proc Natl Acad Sci* 111(40):14625–14630
- Barrett CB, Ghezzi-Kopel K, Hoddinott J, Homami N, Tennant E, Upton J, Wu T (2021) A scoping review of the development resilience literature: theory, methods and evidence. *World Dev* 146:105612
- Blumenstock J, Cadamuro G, On R (2015) Predicting poverty and wealth from mobile phone metadata. *Science* 350(6264):1073–1076
- Brück T, d’Errico M, Pietrelli R (2019) The effects of violent conflict on household resilience and food security: evidence from the 2014 Gaza conflict. *World Dev* 119:203–223
- Cissé JD, Barrett CB (2018) Estimating development resilience: a conditional moments-based approach. *J Dev Econ* 135:272–284
- Conostas M, Frankenberger T, Hoddinott J (2014) Resilience measurement principles: toward an agenda for measurement design. Food Security Information Network, Resilience Measurement Technical Working Group, Technical Series, 1
- d’Errico M, Letta M, Montalbano P, Pietrelli R (2019) Resilience thresholds to temperature anomalies: a long-run test for rural Tanzania. *Ecol Econ* 164:106365
- d’Errico M, Garbero A, Letta M, Winters P (2020) Evaluating program impact on resilience: evidence from lesotho’s child grants programme. *J Dev Stud* 56(12):2212–2234.
- d’Errico M, Di Giuseppe S (2018) Resilience mobility in Uganda: a dynamic analysis. *World Dev* 104:78–96
- Ganguli S, Dunnmon J, Hau D (2019) Predicting food security outcomes using convolutional neural networks (cnns) for satellite tasking. arXiv preprint [arXiv:1902.05433](https://arxiv.org/abs/1902.05433)
- Hastie T, Tibshirani R, Friedman J (2009) The elements of statistical learning: data mining, inference, and prediction. Springer Science & Business Media, New York
- Hoddinott J (2014) Looking at development through a resilience lens. In Fan S, Pandya-Lorch R, Yosef S (Eds). Resilience for food and nutrition security. *Intl Food Policy Res*
- Hossain M, Mullally C, Asadullah MN (2019) Alternatives to calorie-based indicators of food security: an application of machine learning methods. *Food Policy* 84:77–91
- Jean N, Burke M, Xie M, Davis WM, Lobell DB, Ermon S (2016) Combining satellite imagery and machine learning to predict poverty. *Science* 353(6301):790–794
- Jones L, Tanner T (2017) ‘Subjective resilience’: using perceptions to quantify household resilience to climate extremes and disasters. *Reg Environ Change* 17(1):229–243
- Jones L, Conostas MA, Matthews N, Verkaart S (2021) Advancing resilience measurement. *Nat Sustain* 4(4):288–289
- Jones L, D’Errico M (2019) Resilient, but from whose perspective? Like-for-like comparisons of objective and subjective measures of resilience. *World Dev* 124:104632
- Kleinberg J, Ludwig J, Mullainathan S, Obermeyer Z (2015) Prediction policy problems. *Am Econ Rev* 105(5):491–495
- Kleinberg J, Lakkaraju H, Leskovec J, Ludwig J, Mullainathan S (2018) Human decisions and machine predictions. *Q J Econ* 133(1):237–293
- Knippenberg E, Jensen N, Conostas M (2019) Quantifying household resilience with high frequency data: temporal dynamics and methodological options. *World Dev* 121:1–15
- Kshirsagar V, Wiczorek J, Ramanathan S, Wells R (2017) Household poverty classification in data-scarce environments: a machine learning approach. arXiv preprint [arXiv:1711.06813](https://arxiv.org/abs/1711.06813)
- Lantz B (2019) Machine learning with R: expert techniques for predictive modeling. Packt Publishing Ltd, Birmingham
- Lentz EC, Michelson H, Baylis K, Zhou Y (2019) A data-driven approach improves food insecurity crisis prediction. *World Dev* 122:399–409
- McBride L, Barrett CB, Browne C, Hu L, Liu Y, Matteson DS, Wen J (2021) Predicting poverty and malnutrition for targeting, mapping, monitoring, and early warning. *Appl Econ Perspect Policy* 1–14
- McBride L, Nichols A (2018) Retooling poverty targeting using out-of-sample validation and machine learning. *World Bank Econ Rev* 32(3):531–550
- Mullainathan S, Spiess J (2017) Machine learning: an applied econometric approach. *J Econ Perspect* 31(2):87–106
- Perez A, Ganguli S, Ermon S, Azzari G, Burke M, Lobell D (2019) Semi-supervised multitask learning on multispectral satellite images using wasserstein generative adversarial networks (gans) for predicting poverty. arXiv preprint [arXiv:1902.11110](https://arxiv.org/abs/1902.11110)
- Smith LC, Frankenberger TR (2018) Does resilience capacity reduce the negative impact of shocks on household food security? Evidence from the 2014 floods in Northern Bangladesh. *World Dev* 102:358–376

- Steele JE, Sundsøy PR, Pezzulo C, Alegana VA, Bird TJ, Blumenstock J, Hadiuzzaman KN (2017) Mapping poverty using mobile phone and satellite data. *J R Soc Interface* 14(127):20160690
- Upton JB, Cissé JD, Barrett CB (2016) Food security as resilience: reconciling definition and measurement. *Agric Econ* 47(S1):135–147
- Varian HR (2014) Big data: new tricks for econometrics. *J Econ Perspect* 28(2):3–28

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.