

# CLARE: A CAUSAL MACHINE LEARNING APPROACH TO RESILIENCE ESTIMATION

Talip Kilic<sup>‡</sup>, Marco Letta<sup>†</sup>, **Pierluigi Montalbano**<sup>†</sup>, Federica  
Petruccelli<sup>†</sup>

<sup>†</sup> *Department of Social Sciences and Economics, Sapienza University of Rome*

<sup>‡</sup> *Development Data Group, World Bank*

**BETTER DATA FOR BETTER JOBS AND LIVES**  
**DECEMBER 8–9, 2025**  
**WORLD BANK, WASHINGTON, D.C.**



# CONTRIBUTION

- **New resilience index** based on **causal machine learning** techniques applied to **multi-topic longitudinal household data**;



# CONTRIBUTION

- **New resilience index** based on **causal machine learning** techniques applied to **multi-topic longitudinal household data**;
- Model-agnostic, **data-driven**, scalable, and normatively anchored to wellbeing thresholds;



# CONTRIBUTION

- **New resilience index** based on **causal machine learning** techniques applied to **multi-topic longitudinal household data**;
- Model-agnostic, **data-driven**, scalable, and normatively anchored to wellbeing thresholds;
- It can be developed either as a **shock-specific** or a **general-purpose** resilience metric;



# CONTRIBUTION

- **New resilience index** based on **causal machine learning** techniques applied to **multi-topic longitudinal household data**;
- Model-agnostic, **data-driven**, scalable, and normatively anchored to wellbeing thresholds;
- It can be developed either as a **shock-specific** or a **general-purpose** resilience metric;
- It serves a **dual purpose**, aligning with two key definitions of resilience:



# CONTRIBUTION

- **New resilience index** based on **causal machine learning** techniques applied to **multi-topic longitudinal household data**;
- Model-agnostic, **data-driven**, scalable, and normatively anchored to wellbeing thresholds;
- It can be developed either as a **shock-specific** or a **general-purpose** resilience metric;
- It serves a **dual purpose**, aligning with two key definitions of resilience:
  1. as **explanatory (or mediating) variable** that synthetically connects shocks to development outcomes. It aligns with resilience as *“the capacity that ensures stressors and shocks do not have long-lasting adverse development consequences”* (Constas et al.; 2014);



# CONTRIBUTION

- **New resilience index** based on **causal machine learning** techniques applied to **multi-topic longitudinal household data**;
- Model-agnostic, **data-driven**, scalable, and normatively anchored to wellbeing thresholds;
- It can be developed either as a **shock-specific** or a **general-purpose** resilience metric;
- It serves a **dual purpose**, aligning with two key definitions of resilience:
  1. as **explanatory (or mediating) variable** that synthetically connects shocks to development outcomes. It aligns with resilience as *“the capacity that ensures stressors and shocks do not have long-lasting adverse development consequences”* (Constas et al.; 2014);
  2. as **normative outcome** to classify households for targeting and evaluation purposes. It aligns with resilience as *“the ability to achieve and maintain an acceptable standard of wellbeing even in the face of shocks and stressors”* (Barrett and Constas; 2014);



## CONTRIBUTION (CONT.)

- **Operationalizable:** The index is decomposable, enabling **causal identification** of the resilience components that are most effective in protecting populations from shocks;



## CONTRIBUTION (CONT.)

- **Operationalizable:** The index is decomposable, enabling **causal identification** of the resilience components that are most effective in protecting populations from shocks;
- **Shock/Risk-sensitive:** Shocks are embedded in the methodology, and weights are data-driven, derived from a fully nonparametric model;



## CONTRIBUTION (CONT.)

- **Operationalizable:** The index is decomposable, enabling **causal identification** of the resilience components that are most effective in protecting populations from shocks;
- **Shock/Risk-sensitive:** Shocks are embedded in the methodology, and weights are data-driven, derived from a fully nonparametric model;
- **Policy-relevant:** CLARE supports the design, targeting, and monitoring of **resilience-building** policies and investments;



## CONTRIBUTION (CONT.)

- **Operationalizable:** The index is decomposable, enabling **causal identification** of the resilience components that are most effective in protecting populations from shocks;
- **Shock/Risk-sensitive:** Shocks are embedded in the methodology, and weights are data-driven, derived from a fully nonparametric model;
- **Policy-relevant:** CLARE supports the design, targeting, and monitoring of **resilience-building** policies and investments;
- **Transferable:** enables **transfer learning** from data-rich to data-poor settings (**out-of-sample performance**)



# EMPIRICAL APPLICATION

- **Data** We present CLARE application to **drought resilience**, using 35,000 household observations from a 19-wave panel of four SSA countries (28,112 excl. baseline for forecasting).



# EMPIRICAL APPLICATION

- **Data** We present CLARE application to **drought resilience**, using 35,000 household observations from a 19-wave panel of four SSA countries (28,112 excl. baseline for forecasting).
- **Main findings** CLARE **outperforms** existing resilience metrics **out-of-sample**, i.e.:
  - future outcomes (dynamic forecasting)
  - held-out countries (cross-sectional prediction)



# DATA DETAILS

**Table 2: Sources of Household Data**

Country	Survey name	Years	Final n
Malawi	Integrated Household Panel Survey (IHPS)	2010/2011	<i>baseline</i>
		2013	1,282
		2016-2017	1,823
		2019	1,719
Nigeria	General Household Survey (GHS)	2010/2011	<i>baseline</i>
		2012/2013	3,619
		2015/2016	3,650
		2018/2019	1,219
Tanzania	Tanzania National Panel Survey (TZNPS)	2010/2011	<i>baseline</i>
		2012/2013	317
		2014/2015	687
		2019/2020	654
		2009/2010	<i>baseline</i>
Uganda	Uganda National Panel Survey (UNPS)	2010/2011	2,214
		2011/2012	2,211
		2013/2014	977
		2015/2016	2,607
		2018/2019	2,666
		2019/2020	2,467
<b>Total</b>	<b>4 countries</b>	<b>19 waves</b>	<b>28,112</b>

*Notes:* The table presents a summary of household-level data for each country, based on the LSMS Basic Information Documents.



# DATA DETAILS

[▶ DETAILS](#)
**Table A.1: Variable descriptions**

Variable name	Definition
Food Consumption Score (FCS)	Food Consumption Score
Food insecurity	FCS ≤ 35
Share of drought months	Share of growing season months experiencing drought (GS-EDDI > +1)
Drought (GS-EDDI)	= 1 if share of drought GS-EDDI > 75th percentile of the country-specific distribution
Multiple shock	= 1 if household experienced any shock <sup>23</sup>
Age of hh head	Age of household head
Gender of hh head	= 1 if Female household head
Household size	Household size
% members male ≤ 15 yrs	% members male ≤ 15 yrs
% members male 16-65 yrs	% members male 16-65 yrs
% members female 16-65 yrs	% members female 16-65 yrs
% members male > 65 yrs	% members male > 65 yrs
% members female > 65 yrs	% members female > 65 yrs
Rural	= 1 if household lives in a rural area
No. of hh members literate	No. of hh members literate
No. of hh members with primary education	No. of hh members with primary education
No. of hh members with secondary education or higher	No. of hh members with secondary education or higher
Household asset index	Household asset index
Mobile owned	= 1 if household own a mobile phone
TLU today	Tropical Livestock Units as of the time of survey
Number of key services within village	No. of key services within community (of 5)
Distance to market	Distance to market (km)

*Notes:* This table provides definitions for the key variables used in the analysis.



# INTUITION

Resilience indicator constructed as a **data-driven weighted average** of its standard sub-components.



# INTUITION

Resilience indicator constructed as a **data-driven weighted average** of its standard sub-components.

**Causal framework:** resilience estimation stems from the causal relationships between wellbeing, shocks, and resilience drivers.



# INTUITION

Resilience indicator constructed as a **data-driven weighted average** of its standard sub-components.

**Causal framework:** resilience estimation stems from the causal relationships between wellbeing, shocks, and resilience drivers.

**The idea in a nutshell:**



# INTUITION

Resilience indicator constructed as a **data-driven weighted average** of its standard sub-components.

**Causal framework:** resilience estimation stems from the causal relationships between wellbeing, shocks, and resilience drivers.

**The idea in a nutshell:**

- **Compute:** Use ML algorithms to derive **data-driven importance weights** of resilience subcomponents and aggregate them into a **composite** indicator.



# INTUITION

Resilience indicator constructed as a **data-driven weighted average** of its standard sub-components.

**Causal framework:** resilience estimation stems from the causal relationships between wellbeing, shocks, and resilience drivers.

**The idea in a nutshell:**

- **Compute:** Use ML algorithms to derive **data-driven importance weights** of resilience subcomponents and aggregate them into a **composite** indicator.
- **Test** its out-of-sample performance in predicting wellbeing both **out-of-country** and in future periods (**forecasting**).



# INTUITION

Resilience indicator constructed as a **data-driven weighted average** of its standard sub-components.

**Causal framework:** resilience estimation stems from the causal relationships between wellbeing, shocks, and resilience drivers.

**The idea in a nutshell:**

- **Compute:** Use ML algorithms to derive **data-driven importance weights** of resilience subcomponents and aggregate them into a **composite** indicator.
- **Test** its out-of-sample performance in predicting wellbeing both **out-of-country** and in future periods (**forecasting**).
- **Compare** CLARE out-of-sample performance to the most common resilience measures (C&B, Realized measures: RIMA & Tango, Naïve lagged outcome); .



## RELATED LITERATURE

- We draw on the theoretical underpinnings and conceptual framework of the **quantitative resilience literature** (Barrett and Constas; 2014; Barrett et al.; 2021; Cissé and Barrett; 2018; d'Errico and Di Giuseppe; 2018; Smith and Frankenberger; 2018; Alinovi et al.; 2010).



## RELATED LITERATURE

- We draw on the theoretical underpinnings and conceptual framework of the **quantitative resilience literature** (Barrett and Conostas; 2014; Barrett et al.; 2021; Cissé and Barrett; 2018; d'Errico and Di Giuseppe; 2018; Smith and Frankenberger; 2018; Alinovi et al.; 2010).
- We follow up on the **comparative assessment** of resilience indicators recently conducted by Upton et al. (2022).



## RELATED LITERATURE

- We draw on the theoretical underpinnings and conceptual framework of the **quantitative resilience literature** (Barrett and Conostas; 2014; Barrett et al.; 2021; Cissé and Barrett; 2018; d'Errico and Di Giuseppe; 2018; Smith and Frankenberger; 2018; Alinovi et al.; 2010).
- We follow up on the **comparative assessment** of resilience indicators recently conducted by Upton et al. (2022).
- We contribute to the **fresh wave** of empirical approaches for measuring and estimating resilience, e.g., Alloush and Carter (2024); Knippenberg et al. (2019); Lee et al. (2025); Scognamillo et al. (2023).



# MACHINE LEARNING

- Supervised machine learning algorithms have been specifically developed for predictive tasks and excel at predicting **out of sample**.



# MACHINE LEARNING

- Supervised machine learning algorithms have been specifically developed for predictive tasks and excel at predicting **out of sample**.
- In development economics these have been employed, often in conjunction with non-conventional data sources, to predict and map poverty, food security, and resilience (e.g., [Aiken et al. \(2022, 2023\)](#); [Blumenstock et al. \(2015\)](#); [Knippenberg et al. \(2019\)](#)).



# MACHINE LEARNING

- Supervised machine learning algorithms have been specifically developed for predictive tasks and excel at predicting **out of sample**.
- In development economics these have been employed, often in conjunction with non-conventional data sources, to predict and map poverty, food security, and resilience (e.g., [Aiken et al. \(2022, 2023\)](#); [Blumenstock et al. \(2015\)](#); [Knippenberg et al. \(2019\)](#)).
- They allow for fully flexible and **non-parametric** estimation of the relationship between the outcome and the predictors.



# Causal MACHINE LEARNING

- Standard predictive algorithms, however, do not uncover causal relationships but only **associations** between outcome and predictors.



# Causal MACHINE LEARNING

- Standard predictive algorithms, however, do not uncover causal relationships but only **associations** between outcome and predictors.
- But in resilience analysis, causality **matters**, as the primary focus should be on assessing:



# Causal MACHINE LEARNING

- Standard predictive algorithms, however, do not uncover causal relationships but only **associations** between outcome and predictors.
- But in resilience analysis, causality **matters**, as the primary focus should be on assessing:
  1. the impacts of **shocks** on wellbeing;



# Causal MACHINE LEARNING

- Standard predictive algorithms, however, do not uncover causal relationships but only **associations** between outcome and predictors.
- But in resilience analysis, causality **matters**, as the primary focus should be on assessing:
  1. the impacts of **shocks** on wellbeing;
  2. the **protective** role of resilience components, i.e., how these determinants **intervene** in this causal relationship by driving **heterogeneous** effects.



# Causal MACHINE LEARNING

- Standard predictive algorithms, however, do not uncover causal relationships but only **associations** between outcome and predictors.
- But in resilience analysis, causality **matters**, as the primary focus should be on assessing:
  1. the impacts of **shocks** on wellbeing;
  2. the **protective** role of resilience components, i.e., how these determinants **intervene** in this causal relationship by driving **heterogeneous** effects.
- To this end, we leverage *causal* machine learning techniques, specifically developed to estimate **heterogeneous causal effects** rather than merely predict outcomes.



# Causal MACHINE LEARNING

- Standard predictive algorithms, however, do not uncover causal relationships but only **associations** between outcome and predictors.
- But in resilience analysis, causality **matters**, as the primary focus should be on assessing:
  1. the impacts of **shocks** on wellbeing;
  2. the **protective** role of resilience components, i.e., how these determinants **intervene** in this causal relationship by driving **heterogeneous** effects.
- To this end, we leverage *causal* machine learning techniques, specifically developed to estimate **heterogeneous causal effects** rather than merely predict outcomes.



# CAUSAL FORESTS

To illustrate our empirical application we use here **Causal Forest** (Wager and Athey; 2018; Athey et al.; 2019) [For robustness we used an alternative CML method (**the X-Learner**), with fully consistent results]



# CAUSAL FORESTS

To illustrate our empirical application we use here **Causal Forest** (Wager and Athey; 2018; Athey et al.; 2019) [For robustness we used an alternative CML method (**the X-Learner**), with fully consistent results]

- **Causal forests** are an adaptation of **random forests** (Breiman; 2001).



# CAUSAL FORESTS

To illustrate our empirical application we use here **Causal Forest** (Wager and Athey; 2018; Athey et al.; 2019) [For robustness we used an alternative CML method (**the X-Learner**), with fully consistent results]

- **Causal forests** are an adaptation of **random forests** (Breiman; 2001).
- They have been specifically developed for the estimation of **Conditional Average Treatment Effects** (CATEs).



# CAUSAL FORESTS

To illustrate our empirical application we use here **Causal Forest** (Wager and Athey; 2018; Athey et al.; 2019) [For robustness we used an alternative CML method (**the X-Learner**), with fully consistent results]

- **Causal forests** are an adaptation of **random forests** (Breiman; 2001).
- They have been specifically developed for the estimation of **Conditional Average Treatment Effects** (CATEs).
- CATEs constitute non-parametric and granular estimates of treatment effect heterogeneity based on a set of chosen covariates, named **treatment effect modifiers** (Athey and Imbens; 2016).

▶ Technical details



# CAUSAL FORESTS

To illustrate our empirical application we use here **Causal Forest** (Wager and Athey; 2018; Athey et al.; 2019) [For robustness we used an alternative CML method (**the X-Learner**), with fully consistent results]

- **Causal forests** are an adaptation of **random forests** (Breiman; 2001).
- They have been specifically developed for the estimation of **Conditional Average Treatment Effects** (CATEs).
- CATEs constitute non-parametric and granular estimates of treatment effect heterogeneity based on a set of chosen covariates, named **treatment effect modifiers** (Athey and Imbens; 2016).

▸ Technical details

- In our setting:



# CAUSAL FORESTS

To illustrate our empirical application we use here **Causal Forest** (Wager and Athey; 2018; Athey et al.; 2019) [For robustness we used an alternative CML method (**the X-Learner**), with fully consistent results]

- **Causal forests** are an adaptation of **random forests** (Breiman; 2001).
- They have been specifically developed for the estimation of **Conditional Average Treatment Effects** (CATEs).
- CATEs constitute non-parametric and granular estimates of treatment effect heterogeneity based on a set of chosen covariates, named **treatment effect modifiers** (Athey and Imbens; 2016).

▶ Technical details

- In our setting:
  1. The **outcome** is a binary wellbeing variable **W** based on a normative threshold (e.g., FCS);



# CAUSAL FORESTS

To illustrate our empirical application we use here **Causal Forest** (Wager and Athey; 2018; Athey et al.; 2019) [For robustness we used an alternative CML method (**the X-Learner**), with fully consistent results]

- **Causal forests** are an adaptation of **random forests** (Breiman; 2001).
- They have been specifically developed for the estimation of **Conditional Average Treatment Effects** (CATEs).
- CATEs constitute non-parametric and granular estimates of treatment effect heterogeneity based on a set of chosen covariates, named **treatment effect modifiers** (Athey and Imbens; 2016).

▶ Technical details

- In our setting:
  1. The **outcome** is a binary wellbeing variable **W** based on a normative threshold (e.g., FCS);
  2. The treatment is a **shock** of interest **S** (e.g., draught);



# CAUSAL FORESTS

To illustrate our empirical application we use here **Causal Forest** (Wager and Athey; 2018; Athey et al.; 2019) [For robustness we used an alternative CML method (**the X-Learner**), with fully consistent results]

- **Causal forests** are an adaptation of **random forests** (Breiman; 2001).
- They have been specifically developed for the estimation of **Conditional Average Treatment Effects** (CATEs).
- CATEs constitute non-parametric and granular estimates of treatment effect heterogeneity based on a set of chosen covariates, named **treatment effect modifiers** (Athey and Imbens; 2016).

▶ Technical details

- In our setting:
  1. The **outcome** is a binary wellbeing variable **W** based on a normative threshold (e.g., FCS);
  2. The treatment is a **shock** of interest **S** (e.g., draught);
  3. The shock effect modifiers are the **resilience subcomponents Z**, measured **before** the shock.



# CLARE CONSTRUCTION

▶ Full pipeline

1. **Estimation** of shock effects and variable weights;



# CLARE CONSTRUCTION

▶ Full pipeline

1. **Estimation** of shock effects and variable weights;
2. **Aggregation** of resilience determinants using the estimated weights;



# CLARE CONSTRUCTION

▶ Full pipeline

1. **Estimation** of shock effects and variable weights;
2. **Aggregation** of resilience determinants using the estimated weights;
3. **Evaluation** of the performance of CLARE scores.



# ESTIMATION

$$W_{it} = \tau(Z_{it-1})S_{it} + f(X_{it-1}, \bar{X}_{i\cdot}, \bar{X}_{\cdot t-1}) + \varepsilon_{it} \quad (1)$$

where:

- $W_{it}$  represents a binary well-being outcomes anchored to a normative threshold (e.g., **Food Consumption Score**  $\leq 35$  ((Upton et al.; 2022)));
- $S_{it}$  is a binary variable capturing exposure to shocks;
- $X_{it-1}$  is a vector of time-varying and *Mundlak-type* confounders and controls for both observed and unobserved heterogeneity (i.e., household and time fixed effects);



# ESTIMATION

$$W_{it} = \tau(Z_{it-1})S_{it} + f(X_{it-1}, \bar{X}_{i\cdot}, \bar{X}_{\cdot t-1}) + \varepsilon_{it} \quad (1)$$

where:

- $W_{it}$  represents a binary well-being outcomes anchored to a normative threshold (e.g., **Food Consumption Score**  $\leq 35$  ((Upton et al.; 2022)));
- $S_{it}$  is a binary variable capturing exposure to shocks;
- $X_{it-1}$  is a vector of time-varying and *Mundlak-type* confounders and controls for both observed and unobserved heterogeneity (i.e., household and time fixed effects);
- $\tau(\mathbf{Z}_{it-1})$  are the pre-shock resilience components accounting for the heterogeneous effects of shocks;



# CONDITIONAL UNCONFOUNDEDNESS

$$W_{it} = \tau(Z_{it-1})S_{it} + f(X_{it-1}, \bar{X}_i, \bar{X}_{\cdot t-1}) + \varepsilon_{it} \quad (2)$$



# CONDITIONAL UNCONFOUNDEDNESS

$$W_{it} = \tau(Z_{it-1})S_{it} + f(X_{it-1}, \bar{X}_i, \bar{X}_{\cdot,t-1}) + \varepsilon_{it} \quad (2)$$

- The shock variable is assumed to be independent of potential outcomes conditional on controls (**conditional unconfoundedness**).



# CONDITIONAL UNCONFOUNDEDNESS

$$W_{it} = \tau(Z_{it-1})S_{it} + f(X_{it-1}, \bar{X}_{i\cdot}, \bar{X}_{\cdot t-1}) + \varepsilon_{it} \quad (2)$$

- The shock variable is assumed to be independent of potential outcomes conditional on controls (**conditional unconfoundedness**).
- To reinforce this assumption we follow a **double-orthogonalization approach** as follows:



# CONDITIONAL UNCONFOUNDEDNESS

$$W_{it} = \tau(Z_{it-1})S_{it} + f(X_{it-1}, \bar{X}_{i\cdot}, \bar{X}_{\cdot t-1}) + \varepsilon_{it} \quad (2)$$

- The shock variable is assumed to be independent of potential outcomes conditional on controls (**conditional unconfoundedness**).
  - To reinforce this assumption we follow a **double-orthogonalization approach** as follows:
1. **Step 1:** We non parametrically estimate  $W^*$  and  $S^*$  by regressing both outcomes and shocks on  $X_{it-1}$ ,  $\bar{X}_{i\cdot}$ ,  $\bar{X}_{\cdot t-1}$ , including any high-dimensional set of potential confounders;



# CONDITIONAL UNCONFOUNDEDNESS

$$W_{it} = \tau(Z_{it-1})S_{it} + f(X_{it-1}, \bar{X}_{i\cdot}, \bar{X}_{\cdot t-1}) + \varepsilon_{it} \quad (2)$$

- The shock variable is assumed to be independent of potential outcomes conditional on controls (**conditional unconfoundedness**).
- To reinforce this assumption we follow a **double-orthogonalization approach** as follows:
  1. **Step 1:** We non parametrically estimate  $W^*$  and  $S^*$  by regressing both outcomes and shocks on  $X_{it-1}$ ,  $\bar{X}_{i\cdot}$ ,  $\bar{X}_{\cdot t-1}$ , including any high-dimensional set of potential confounders;
  2. **Step 2:** We then regress the residualized shocks  $S^*$  on the residualized outcomes  $W^*$ , conditioning on the shock effect modifiers  $\mathbf{Z}_{it-1}$ ;



# CONDITIONAL UNCONFOUNDEDNESS

$$W_{it} = \tau(Z_{it-1})S_{it} + f(X_{it-1}, \bar{X}_{i\cdot}, \bar{X}_{\cdot t-1}) + \varepsilon_{it} \quad (2)$$

- The shock variable is assumed to be independent of potential outcomes conditional on controls (**conditional unconfoundedness**).
- To reinforce this assumption we follow a **double-orthogonalization approach** as follows:
  1. **Step 1:** We non parametrically estimate  $W^*$  and  $S^*$  by regressing both outcomes and shocks on  $X_{it-1}$ ,  $\bar{X}_{i\cdot}$ ,  $\bar{X}_{\cdot t-1}$ , including any high-dimensional set of potential confounders;
  2. **Step 2:** We then regress the residualized shocks  $S^*$  on the residualized outcomes  $W^*$ , conditioning on the shock effect modifiers  $\mathbf{Z}_{it-1}$ ;
  3. **Step 3:** We cluster at the household level to remove further heteroskedasticity in  $\varepsilon_{it}$ ;



# CONDITIONAL UNCONFOUNDEDNESS

$$W_{it} = \tau(Z_{it-1})S_{it} + f(X_{it-1}, \bar{X}_{i\cdot}, \bar{X}_{\cdot t-1}) + \varepsilon_{it} \quad (2)$$

- The shock variable is assumed to be independent of potential outcomes conditional on controls (**conditional unconfoundedness**).
- To reinforce this assumption we follow a **double-orthogonalization approach** as follows:
  1. **Step 1:** We non parametrically estimate  $W^*$  and  $S^*$  by regressing both outcomes and shocks on  $X_{it-1}$ ,  $\bar{X}_{i\cdot}$ ,  $\bar{X}_{\cdot t-1}$ , including any high-dimensional set of potential confounders;
  2. **Step 2:** We then regress the residualized shocks  $S^*$  on the residualized outcomes  $W^*$ , conditioning on the shock effect modifiers  $\mathbf{Z}_{it-1}$ ;
  3. **Step 3:** We cluster at the household level to remove further heteroskedasticity in  $\varepsilon_{it}$ ;

At the end of the process, our residualized outcomes and shocks are both filtered from time-invariant, time-variant, and time-trend further confounders;



# RETRIEVING THE RELATIVE IMPORTANCE OF RESILIENCE COMPONENTS



# RETRIEVING THE RELATIVE IMPORTANCE OF RESILIENCE COMPONENTS

- Explore household-specific heterogeneity of resilience components;



# RETRIEVING THE RELATIVE IMPORTANCE OF RESILIENCE COMPONENTS

- Explore household-specific heterogeneity of resilience components;
- Retrieve the **relative importance** of each resilience component **Z** in driving heterogeneity in the estimated causal effects, by measuring the **split frequency** of treatment effects;



# RETRIEVING THE RELATIVE IMPORTANCE OF RESILIENCE COMPONENTS

- Explore household-specific heterogeneity of resilience components;
- Retrieve the **relative importance** of each resilience component  $Z$  in driving heterogeneity in the estimated causal effects, by measuring the **split frequency** of treatment effects;
- Formally, the **importance weight**  $w_j$  is calculated as the proportion of trees in which the component is used to make a split (and the process is repeated across all trees in the forest);



# RETRIEVING THE RELATIVE IMPORTANCE OF RESILIENCE COMPONENTS

- Explore household-specific heterogeneity of resilience components;
- Retrieve the **relative importance** of each resilience component  $Z$  in driving heterogeneity in the estimated causal effects, by measuring the **split frequency** of treatment effects;
- Formally, the **importance weight**  $w_j$  is calculated as the proportion of trees in which the component is used to make a split (and the process is repeated across all trees in the forest);
- This ensures that the entire process is fully flexible and nonparametric, capturing possible nonlinear resilience features (as argued by the specialized literature);

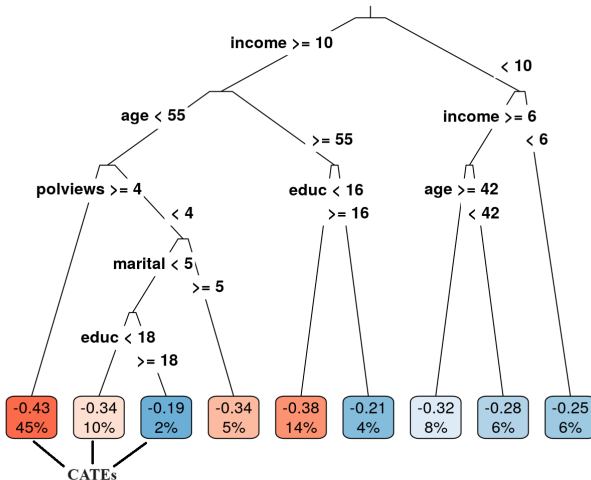


# RETRIEVING THE RELATIVE IMPORTANCE OF RESILIENCE COMPONENTS

- Explore household-specific heterogeneity of resilience components;
- Retrieve the **relative importance** of each resilience component  $Z$  in driving heterogeneity in the estimated causal effects, by measuring the **split frequency** of treatment effects;
- Formally, the **importance weight**  $w_j$  is calculated as the proportion of trees in which the component is used to make a split (and the process is repeated across all trees in the forest);
- This ensures that the entire process is fully flexible and nonparametric, capturing possible nonlinear resilience features (as argued by the specialized literature);
- Although weights are in principle sensitive to the number of covariates, here the set of components is established by the theory (and we can thus be confident on the computed  $w_j$ );



# CAUSAL TREE EXAMPLE



Example of a causal tree



# AGGREGATION

- By normalizing the relative importance measures produced by the forest, we get variable **importance weights**  $\omega_j$  summing up to 1.



# AGGREGATION

- By normalizing the relative importance measures produced by the forest, we get variable **importance weights**  $\omega_j$  summing up to 1.
- We then take a simple **weighted average** of the **standardized** resilience components using the estimated weights:

$$\text{CLARE}_{it} = \left( \sum_{j=1}^J \omega_j Z_{it-1} \right) \cdot (1 - \Pr(W_{it} \leq \underline{W} \mid S_{it}, X_{it-1}, \bar{X}_i, \bar{X}_{\cdot t-1})) \quad (3)$$



## AGGREGATION

- By normalizing the relative importance measures produced by the forest, we get variable **importance weights**  $\omega_j$  summing up to 1.
- We then take a simple **weighted average** of the **standardized** resilience components using the estimated weights:

$$\text{CLARE}_{it} = \left( \sum_{j=1}^J \omega_j Z_{it-1} \right) \cdot (1 - \Pr(W_{it} \leq \underline{W} \mid S_{it}, X_{it-1}, \bar{X}_i, \bar{X}_{\cdot t-1})) \quad (3)$$

- We rescale this weighted average by the **probability of being above the threshold under the shock**—also estimated out-of-sample by the causal forest—and then multiply by 100 [Note that shocks are embedded in CLARE procedure];



## AGGREGATION

- By normalizing the relative importance measures produced by the forest, we get variable **importance weights**  $\omega_j$  summing up to 1.
- We then take a simple **weighted average** of the **standardized** resilience components using the estimated weights:

$$\text{CLARE}_{it} = \left( \sum_{j=1}^J \omega_j Z_{it-1} \right) \cdot (1 - \Pr(W_{it} \leq \underline{W} \mid S_{it}, X_{it-1}, \bar{X}_i, \bar{X}_{t-1})) \quad (3)$$

- We rescale this weighted average by the **probability of being above the threshold under the shock**—also estimated out-of-sample by the causal forest—and then multiply by 100 [Note that shocks are embedded in CLARE procedure];
- The resulting quantity represents the household-specific indicator of resilience to the shock of interest—**CLARE**—and ranges 0-100.



## AGGREGATION

- By normalizing the relative importance measures produced by the forest, we get variable **importance weights**  $\omega_j$  summing up to 1.
- We then take a simple **weighted average** of the **standardized** resilience components using the estimated weights:

$$\text{CLARE}_{it} = \left( \sum_{j=1}^J \omega_j Z_{it-1} \right) \cdot (1 - \Pr(W_{it} \leq \underline{W} \mid S_{it}, X_{it-1}, \bar{X}_i, \bar{X}_{\cdot t-1})) \quad (3)$$

- We rescale this weighted average by the **probability of being above the threshold under the shock**—also estimated out-of-sample by the causal forest—and then multiply by 100 [Note that shocks are embedded in CLARE procedure];
- The resulting quantity represents the household-specific indicator of resilience to the shock of interest—**CLARE**—and ranges 0-100.
- This continuous indicator can then be binarized using various criteria/thresholds (e.g., median).



## AGGREGATION

- By normalizing the relative importance measures produced by the forest, we get variable **importance weights**  $\omega_j$  summing up to 1.
- We then take a simple **weighted average** of the **standardized** resilience components using the estimated weights:

$$\text{CLARE}_{it} = \left( \sum_{j=1}^J \omega_j Z_{it-1} \right) \cdot (1 - \Pr(W_{it} \leq \underline{W} \mid S_{it}, X_{it-1}, \bar{X}_i, \bar{X}_{t-1})) \quad (3)$$

Change in  $W$  due to the shock

- We rescale this weighted average by the **probability of being above the threshold under the shock**—also estimated out-of-sample by the causal forest—and then multiply by 100 [Note that shocks are embedded in CLARE procedure];
- The resulting quantity represents the household-specific indicator of resilience to the shock of interest—**CLARE**—and ranges 0-100.
- This continuous indicator can then be binarized using various criteria/thresholds (e.g., median).

## AGGREGATION

- By normalizing the relative importance measures produced by the forest, we get variable **importance weights**  $\omega_j$  summing up to 1.
- We then take a simple **weighted average** of the **standardized** resilience components using the estimated weights:

$$\text{CLARE}_{it} = \left( \sum_{j=1}^J \omega_j Z_{it-1} \right) \cdot (1 - \Pr(W_{it} \leq \underline{W} \mid S_{it}, X_{it-1}, \bar{X}_i, \bar{X}_{t-1})) \quad (3)$$

Change in W due to the shock

Shock impact on wellbeing levels

- We rescale this weighted average by the **probability of being above the threshold under the shock**—also estimated out-of-sample by the causal forest—and then multiply by 100 [Note that shocks are embedded in CLARE procedure];
- The resulting quantity represents the household-specific indicator of resilience to the shock of interest—**CLARE**—and ranges 0-100.
- This continuous indicator can then be binarized using various criteria/thresholds (e.g., median).



# EVALUATION

- Estimation of the data-driven weights is carried out exclusively using **training set data**.



# EVALUATION

- Estimation of the data-driven weights is carried out exclusively using **training set data**.
- We then **'export'** the weights to other held-out data, **plug in** new values for the resilience components of these new data, and compute CLARE for the **held-out** observations.



# EVALUATION

- Estimation of the data-driven weights is carried out exclusively using **training set data**.
- We then '**export**' the weights to other held-out data, **plug in** new values for the resilience components of these new data, and compute CLARE for the **held-out** observations.
- Finally we test CLARE's out-of-sample performance exclusively on these **testing data**.



# EVALUATION [▶ FULL PIPELINE](#)

On the testing set, we assess the predictive ability of:



# EVALUATION

[▶ FULL PIPELINE](#)

On the testing set, we assess the predictive ability of:

- 1) the **continuous** CLARE index (ranging from 0 to 100) by comparing it to the wellbeing outcome from the testing sets (**Regression task**);



# EVALUATION

[▶ FULL PIPELINE](#)

On the testing set, we assess the predictive ability of:

- 1) the **continuous** CLARE index (ranging from 0 to 100) by comparing it to the wellbeing outcome from the testing sets (**Regression task**);
- 2) the **binary** version that identifies non-resilient households (scoring 1 if the CLARE index is below the **median**, and 0 otherwise) against the observed binary wellbeing status in the testing set (**Classification task**).



# EVALUATION

[▶ FULL PIPELINE](#)

On the testing set, we assess the predictive ability of:

- 1) the **continuous** CLARE index (ranging from 0 to 100) by comparing it to the wellbeing outcome from the testing sets (**Regression task**);
- 2) the **binary** version that identifies non-resilient households (scoring 1 if the CLARE index is below the **median**, and 0 otherwise) against the observed binary wellbeing status in the testing set (**Classification task**). NB: **Illustrative choice**.



# EVALUATION

[▶ FULL PIPELINE](#)

On the testing set, we assess the predictive ability of:

- 1) the **continuous** CLARE index (ranging from 0 to 100) by comparing it to the wellbeing outcome from the testing sets (**Regression task**);
- 2) the **binary** version that identifies non-resilient households (scoring 1 if the CLARE index is below the **median**, and 0 otherwise) against the observed binary wellbeing status in the testing set (**Classification task**). NB: **Illustrative choice**.

We **benchmark** CLARE performance against:



# EVALUATION

▶ FULL PIPELINE

On the testing set, we assess the predictive ability of:

- 1) the **continuous** CLARE index (ranging from 0 to 100) by comparing it to the wellbeing outcome from the testing sets (**Regression task**);
- 2) the **binary** version that identifies non-resilient households (scoring 1 if the CLARE index is below the **median**, and 0 otherwise) against the observed binary wellbeing status in the testing set (**Classification task**). NB: **Illustrative choice**.

We **benchmark** CLARE performance against:

- The **Cissé and Barrett (2018)**'s **Resilience Score**, emerged as the **best-performing** indicator in the study by **Upton et al. (2022)**;



# EVALUATION

▶ FULL PIPELINE

On the testing set, we assess the predictive ability of:

- 1) the **continuous** CLARE index (ranging from 0 to 100) by comparing it to the wellbeing outcome from the testing sets (**Regression task**);
- 2) the **binary** version that identifies non-resilient households (scoring 1 if the CLARE index is below the **median**, and 0 otherwise) against the observed binary wellbeing status in the testing set (**Classification task**). NB: **Illustrative choice**.

We **benchmark** CLARE performance against:

- The **Cissé and Barrett (2018)**'s **Resilience Score**, emerged as the **best-performing** indicator in the study by **Upton et al. (2022)**;
- **'Realized resilience'** (RIMA's & TANGO's binary indicator);



# EVALUATION

[▶ FULL PIPELINE](#)

On the testing set, we assess the predictive ability of:

- 1) the **continuous** CLARE index (ranging from 0 to 100) by comparing it to the wellbeing outcome from the testing sets (**Regression task**);
- 2) the **binary** version that identifies non-resilient households (scoring 1 if the CLARE index is below the **median**, and 0 otherwise) against the observed binary wellbeing status in the testing set (**Classification task**). NB: **Illustrative choice**.

We **benchmark** CLARE performance against:

- The **Cissé and Barrett (2018)**'s **Resilience Score**, emerged as the **best-performing** indicator in the study by **Upton et al. (2022)**;
- '**Realized resilience**' (RIMA's & TANGO's binary indicator);
- A **naïve** approach using lagged wellbeing to predict future wellbeing;



# EVALUATION

▶ FULL PIPELINE

On the testing set, we assess the predictive ability of:

- 1) the **continuous** CLARE index (ranging from 0 to 100) by comparing it to the wellbeing outcome from the testing sets (**Regression task**);
- 2) the **binary** version that identifies non-resilient households (scoring 1 if the CLARE index is below the **median**, and 0 otherwise) against the observed binary wellbeing status in the testing set (**Classification task**). NB: **Illustrative choice**.

We **benchmark** CLARE performance against:

- The **Cissé and Barrett (2018)**'s **Resilience Score**, emerged as the **best-performing** indicator in the study by **Upton et al. (2022)**;
- **'Realized resilience'** (RIMA's & TANGO's binary indicator);
- A **naïve** approach using lagged wellbeing to predict future wellbeing;
- Several simpler estimation approaches.



# A COMPARATIVE PERFORMANCE TEST

We comparatively tested CLARE's performance in predicting outcomes across two key dimensions:

- **Forecasting:** Train the model and estimate weights on **earlier** (1-3) waves of all countries, predict and test out-of-sample performance on **later** (4-7) waves of all countries.



# A COMPARATIVE PERFORMANCE TEST

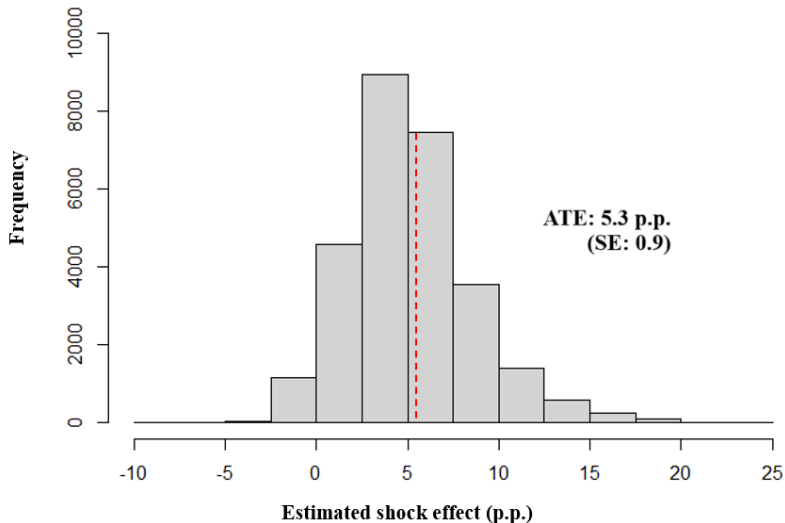
We comparatively tested CLARE's performance in predicting outcomes across two key dimensions:

- **Forecasting:** Train the model and estimate weights on **earlier** (1-3) waves of all countries, predict and test out-of-sample performance on **later** (4-7) waves of all countries.
- **Out-of-country** prediction: **iteratively** train the model and estimate weights on a **rotating** 3-country dataset, predict on the held-out country, and test out-of-sample on the **pooled** dataset of all 4 held-out countries.

**Caveat:** ex-ante predictions and ex-post outcomes are different metrics



# SHOCK EFFECT DISTRIBUTION - TRAINING SET



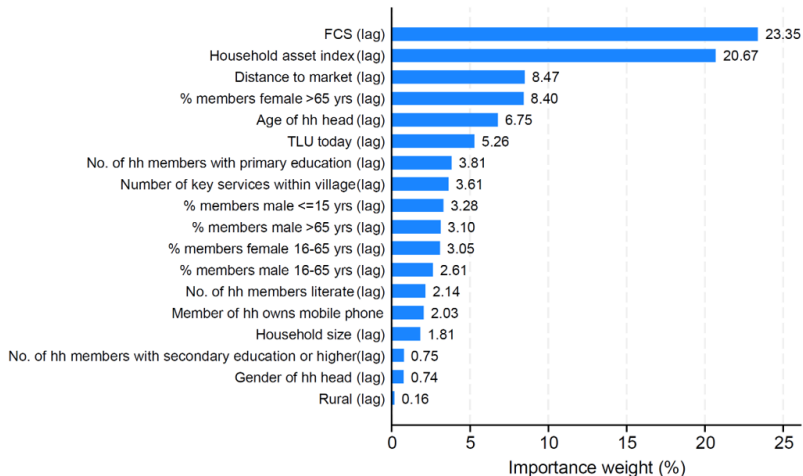
*Notes:* the vertical dotted red line corresponds to the ATE.



# IMPORTANCE WEIGHTS - TRAINING SET

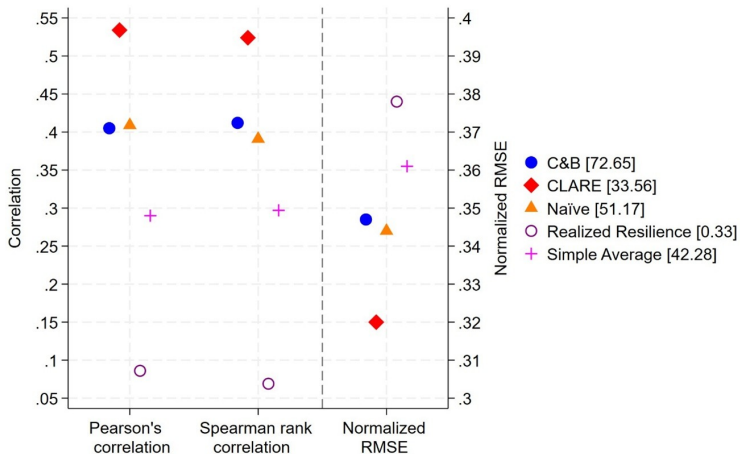
▶ WEIGHTS2

## Variable importance weights for the forecasting task



# COMPARATIVE ASSESSMENT - FORECASTING

▶ FULL



## Out-of-sample performances for the regression problem

▶ Classification

▶ OOC

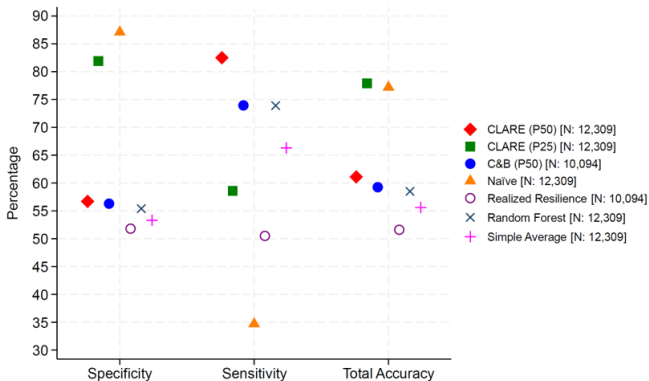


# COMPARATIVE ASSESSMENT - FORECASTING

▶ FULL

## Summary comparative assessment for the forecasting classification task

All countries, waves 4 to 7

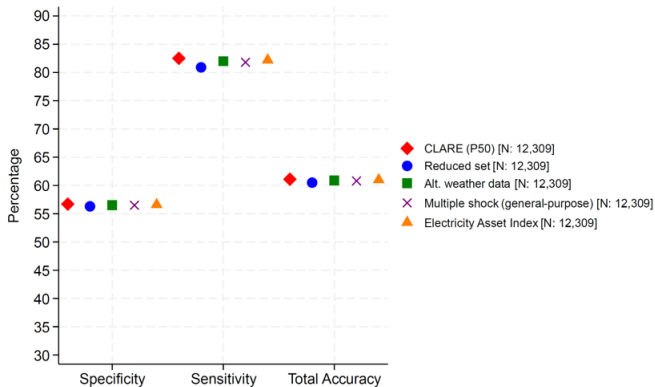


*Notes:* This figure provides a summary comparison of the performance of binary CLARE (using both the median and 25<sup>th</sup> percentile cutoffs) relative to several alternative measures. The testing set comprises observations from all countries in waves 4 to 7. The number of observations (shown in parentheses) differs across methods due to variations in data requirements and model structures. Detailed results for the alternative measures are reported in the Online Appendix.



# SENSITIVITY IN FORECASTING

Summary comparative assessment among different CLARE versions for the forecasting classification task – All countries, waves 4 to 7



*Notes:* This figure presents performance metrics across different CLARE versions in the forecasting regression task. The testing set comprises observations from all countries in waves 4 to 7. The red diamond labeled 'CLARE' in the legend represents the baseline CLARE indicator used for the forecasting analysis. Values in parentheses indicate the average measure score.

## WRAP UP

- Using a data-driven weighting scheme based on estimated causal relationships, CLARE addresses the issue of **aggregating** different resilience subcomponents into a **sound composite** indicator.



## WRAP UP

- Using a data-driven weighting scheme based on estimated causal relationships, CLARE addresses the issue of **aggregating** different resilience subcomponents into a **sound composite** indicator.
- An important feature of CLARE is its **scalability**.



## WRAP UP

- Using a data-driven weighting scheme based on estimated causal relationships, CLARE addresses the issue of **aggregating** different resilience subcomponents into a **sound composite** indicator.
- An important feature of CLARE is its **scalability**.
- Once the weights and probabilities have been estimated, computing the index on new data is straightforward, conditional on:



## WRAP UP

- Using a data-driven weighting scheme based on estimated causal relationships, CLARE addresses the issue of **aggregating** different resilience subcomponents into a **sound composite** indicator.
- An important feature of CLARE is its **scalability**.
- Once the weights and probabilities have been estimated, computing the index on new data is straightforward, conditional on:
  1. the **assumption** of stability/invariance in the data-generating process;



## WRAP UP

- Using a data-driven weighting scheme based on estimated causal relationships, CLARE addresses the issue of **aggregating** different resilience subcomponents into a **sound composite** indicator.
- An important feature of CLARE is its **scalability**.
- Once the weights and probabilities have been estimated, computing the index on new data is straightforward, conditional on:
  1. the **assumption** of stability/invariance in the data-generating process;
  2. the availability of resilience components data for at least **one** wave.



## WRAP UP

- Using a data-driven weighting scheme based on estimated causal relationships, CLARE addresses the issue of **aggregating** different resilience subcomponents into a **sound composite** indicator.
- An important feature of CLARE is its **scalability**.
- Once the weights and probabilities have been estimated, computing the index on new data is straightforward, conditional on:
  1. the **assumption** of stability/invariance in the data-generating process;
  2. the availability of resilience components data for at least **one** wave.
- Data-driven weights enable the establishment of an **objective hierarchy** among resilience components.



## WRAP UP

- Using a data-driven weighting scheme based on estimated causal relationships, CLARE addresses the issue of **aggregating** different resilience subcomponents into a **sound composite** indicator.
- An important feature of CLARE is its **scalability**.
- Once the weights and probabilities have been estimated, computing the index on new data is straightforward, conditional on:
  1. the **assumption** of stability/invariance in the data-generating process;
  2. the availability of resilience components data for at least **one** wave.
- Data-driven weights enable the establishment of an **objective hierarchy** among resilience components.
- **NOTE:** The outcomes are, of course, subject to the threshold we choose to use to discretize the continuous CLARE indicator. This offers the opportunity to **prioritize** the most important **targetable** characteristics.



# WRAP UP

- Unlike for other indicators, CLARE is shock/risk-sensitive (i.e., **shocks play a central role**).



# WRAP UP

- Unlike for other indicators, CLARE is shock/risk-sensitive (i.e., **shocks play a central role**).
- CLARE can be easily **adapted** to **any** kind of shock affecting wellbeing and **any** wellbeing outcome.



# WRAP UP

- Unlike for other indicators, CLARE is shock/risk-sensitive (i.e., **shocks play a central role**).
- CLARE can be easily **adapted** to **any** kind of shock affecting wellbeing and **any** wellbeing outcome.
- Such flexibility allows leveraging it for **multiple purposes**:



# WRAP UP

- Unlike for other indicators, CLARE is shock/risk-sensitive (i.e., **shocks play a central role**).
- CLARE can be easily **adapted** to **any** kind of shock affecting wellbeing and **any** wellbeing outcome.
- Such flexibility allows leveraging it for **multiple purposes**:
  1. Informing resilience-building interventions;



# WRAP UP

- Unlike for other indicators, CLARE is shock/risk-sensitive (i.e., **shocks play a central role**).
- CLARE can be easily **adapted** to **any** kind of shock affecting wellbeing and **any** wellbeing outcome.
- Such flexibility allows leveraging it for **multiple purposes**:
  1. Informing resilience-building interventions;
  2. Facilitating transfer learning;



# WRAP UP

- Unlike for other indicators, CLARE is shock/risk-sensitive (i.e., **shocks play a central role**).
- CLARE can be easily **adapted** to **any** kind of shock affecting wellbeing and **any** wellbeing outcome.
- Such flexibility allows leveraging it for **multiple purposes**:
  1. Informing resilience-building interventions;
  2. Facilitating transfer learning;
  3. Prioritizing data collection efforts in data-poor environments;



# WRAP UP

- Unlike for other indicators, CLARE is shock/risk-sensitive (i.e., **shocks play a central role**).
- CLARE can be easily **adapted** to **any** kind of shock affecting wellbeing and **any** wellbeing outcome.
- Such flexibility allows leveraging it for **multiple purposes**:
  1. Informing resilience-building interventions;
  2. Facilitating transfer learning;
  3. Prioritizing data collection efforts in data-poor environments;
  4. Implementing early-warning systems in shock-prone areas.



# OPEN ISSUES AND RESEARCH AVENUES



# OPEN ISSUES AND RESEARCH AVENUES

- Sensitivity to non-stationarity and distribution shifts  
([Constenla-Villoslada et al.; 2025](#));



# OPEN ISSUES AND RESEARCH AVENUES

- Sensitivity to non-stationarity and distribution shifts ([Constenla-Villoslada et al.; 2025](#));
- Incorporation of ex-ante risks and stressors;



# OPEN ISSUES AND RESEARCH AVENUES

- Sensitivity to non-stationarity and distribution shifts ([Constenla-Villoslada et al.; 2025](#));
- Incorporation of ex-ante risks and stressors;
- Integration with multidimensional outcomes ([Lee et al.; 2025](#));



# OPEN ISSUES AND RESEARCH AVENUES

- Sensitivity to non-stationarity and distribution shifts ([Constenla-Villoslada et al.; 2025](#));
- Incorporation of ex-ante risks and stressors;
- Integration with multidimensional outcomes ([Lee et al.; 2025](#));
- Identifying a minimal variable set for reliable out-of-sample resilience measurement and expanding country coverage (ideally embedded in LSMS, e.g. climate resilience);



# OPEN ISSUES AND RESEARCH AVENUES

- Sensitivity to non-stationarity and distribution shifts ([Constenla-Villoslada et al.; 2025](#));
- Incorporation of ex-ante risks and stressors;
- Integration with multidimensional outcomes ([Lee et al.; 2025](#));
- Identifying a minimal variable set for reliable out-of-sample resilience measurement and expanding country coverage (ideally embedded in LSMS, e.g. climate resilience);
- Expanding the database to include high-frequency phone survey data;



# OPEN ISSUES AND RESEARCH AVENUES

- Sensitivity to non-stationarity and distribution shifts ([Constenla-Villoslada et al.; 2025](#));
- Incorporation of ex-ante risks and stressors;
- Integration with multidimensional outcomes ([Lee et al.; 2025](#));
- Identifying a minimal variable set for reliable out-of-sample resilience measurement and expanding country coverage (ideally embedded in LSMS, e.g. climate resilience);
- Expanding the database to include high-frequency phone survey data;
- Developing ready-to-use software for computation.



**THANK YOU FOR THE ATTENTION!**

For further information: [\\_pierluigi.montalbano@uniroma1.it](mailto:pierluigi.montalbano@uniroma1.it)



# REFERENCES

- Aiken, E., Bellue, S., Karlan, D., Udry, C. and Blumenstock, J. E. (2022). Machine learning and phone data can improve targeting of humanitarian aid, Nature **603**(7903): 864–870.
- Aiken, E. L., Bedoya, G., Blumenstock, J. E. and Coville, A. (2023). Program targeting with machine learning and mobile phone data: Evidence from an anti-poverty intervention in afghanistan, Journal of Development Economics **161**: 103016.
- Alinovi, L., D'errico, M., Mane, E., Romano, D. et al. (2010). Livelihoods strategies and household resilience to food insecurity: An empirical analysis to kenya, European report on development **1**(1): 1–52.
- Alloush, M. and Carter, M. (2024). On the definition and estimation of economic resilience using counterfactuals, National Bureau of Economic Research (NBER) Working Paper **33290**.
- Athey, S. and Imbens, G. (2016). Recursive partitioning for heterogeneous causal effects, Proceedings of the National Academy of Sciences **113**(27): 7353–7360.
- Athey, S., Tibshirani, J. and Wager, S. (2019). Generalized random forests, Annals of Statistics **47**(2): 1148–1178.
- Barrett, C. B. and Conostas, M. A. (2014). Toward a theory of resilience for international development applications, Proceedings of the National Academy of Sciences **111**(40): 14625–14630.
- Barrett, C. B., Ghezzi-Kopel, K., Hoddinott, J., Homami, N., Tennant, E., Upton, J. and Wu, T. (2021). A scoping review of the development resilience literature: Theory, methods and evidence, World Development **146**: 105612.
- Blumenstock, J., Cadamuro, G. and On, R. (2015). Predicting poverty and wealth from mobile phone metadata, Science **350**(6264): 1073–1076.
- Breiman, L. (2001). Random forests, Machine Learning **45**: 5–32.
- Cissé, J. D. and Barrett, C. B. (2018). Estimating development resilience: A conditional moments-based approach, Journal of Development Economics **135**: 272–284.
- Constat, M., Frankenberger, T. and Hoddinott, J. (2014). Resilience measurement principles: Toward an agenda for measurement design, Food Security Information Network, Resilience Measurement Technical Working Group, Technical Series No. 1.
- Constenla-Villoslada, S., Liu, Y., McBride, L., Ouma, C., Mutanda, N. and Barrett, C. B. (2025). High-frequency monitoring enables machine learning-based forecasting of acute child malnutrition for early warning, Proceedings of the National Academy of Sciences **122**(23): e2416161122.
- d'Errico, M. and Di Giuseppe, S. (2018). Resilience mobility in uganda: A dynamic analysis, World Development **104**: 78–96.
- Knippenberg, E., Jensen, N. and Conostas, M. (2019). Quantifying household resilience with high frequency data: Temporal dynamics and methodological options, World Development **121**: 1–15.
- Lee, S., Abay, K. A., Barrett, C. B. and Hoddinott, J. (2025). Estimating multidimensional development resilience, Journal of Development Economics p. 103583.
- Scognamiglio, A., Song, C. and Ignaciuk, A. (2023). No man is an island: A spatially explicit approach to measure development resilience, World Development **171**: 106358.
- Smith, L. C. and Frankenberger, T. R. (2018). Does resilience capacity reduce the negative impact of shocks on household food security? evidence from the 2014 floods in northern bangladesh, World Development **102**: 358–376.
- Upton, J., Constenla-Villoslada, S. and Barrett, C. B. (2022). Caveat utilitor: A comparative assessment of resilience measurement approaches, Journal of Development Economics **157**: 102873.
- Wager, S. and Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests, Journal of the American Statistical Association **113**(523): 1228–1242.



# HOW DO CAUSAL FORESTS WORK?

[▶ BACK](#)

- The causal forest algorithm is an **ensemble of causal trees**.
- Each of these trees is defined by data-driven sample splits generating leaves, which are followed by a prediction of the causal effect over a set of conditioning characteristics  $Z$ . [▶ Example](#)
- The aim of a causal forest is to split the data so as to **maximize** treatment effect heterogeneity across leaves.
- Each individual tree explicitly searches for the subgroups where the treatment effects differ most, and the final prediction is a weighted average over the predictions across all trees.

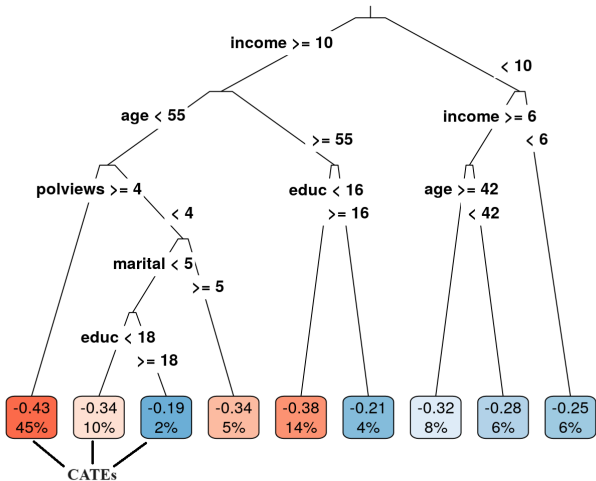


# TECHNICAL DETAILS [▶ BACK](#)

- To prevent overfitting, the algorithm uses the **“honest approach”** (Athey and Imbens; 2016), which randomly splits the sample in two equal parts.
- Half of the sample (the *prediction* sample) is employed to define the sample splits (leaves), while the *estimation* sample is used for estimating the predicted CATE.
- Such a procedure is repeated as many times as there are trees in the forest (2,000 trees, in our case).



# CAUSAL TREE EXAMPLE ▶ BACK



Example of a causal tree



# ORTHOGONALIZATION

- The main identifying assumption underlying the causal forest approach is **conditional unconfoundedness**.



# ORTHOGONALIZATION

- The main identifying assumption underlying the causal forest approach is **conditional unconfoundedness**.
- To make the assumption credible, we adopt a **double orthogonalization approach**, exploiting the Frisch-Waugh-Lovell theorem to non-parametrically control for confounding ([Wager and Athey; 2018](#)).



# ORTHOGONALIZATION

- The main identifying assumption underlying the causal forest approach is **conditional unconfoundedness**.
- To make the assumption credible, we adopt a **double orthogonalization approach**, exploiting the Frisch-Waugh-Lovell theorem to non-parametrically control for confounding ([Wager and Athey; 2018](#)).
- We separately orthogonalize the outcome and shock variables with respect to a host of potential confounders by running two preliminary and fully non-parametric regression forests.



# FULL PIPELINE

[▶ BACK](#)

**Table 1: The CLARE pipeline**

<p><b>1. Preliminary – Data splitting</b></p> <p>We split the full dataset into a training set, on which we will compute resilience weights and conditional probabilities of falling below the normative threshold under the shock scenario, and a testing set, on which we will later test the out-of-sample performance. The definition of the training and testing sets depends on the prediction goal: if the goal is forecasting, the training set should include earlier waves, and the testing set should include only later ones; if the goal is cross-sectional prediction—e.g., out-of-country—the training set should be a rotating one that includes all-but-one countries, and the testing set should be the left-out country.</p>
<p><b>2. Training phase – <u>Use only the training set</u></b></p> <p>In this step, we use causal forests to estimate the heterogeneous effects of the shock (e.g., droughts) on a binary wellbeing outcome (e.g., food insecurity status), conditional on the pre-shock values of the resilience subcomponents. Non-parametric estimation based on double orthogonalization removes confounding effects, while the causal forest estimation of heterogeneous effects is based on a residual-on-residual regression conditional on the resilience subcomponents. At the end of this step, we derive the data-driven weights from the estimated relative variable importance for the resilience subcomponents produced by the forest.</p>
<p><b>3. Prediction phase</b></p> <p>Armed with the model estimated on the training set, we predict the wellbeing outcome on the full sample, which also includes testing set observations. This allows to derive conditional probabilities of falling below the normative threshold under the shock scenario, as well as the heterogeneous, household-specific effects of the shock on the outcome.</p>
<p><b>4. Aggregation phase</b></p> <p>We then compute CLARE as follows: i) <b>Weight aggregation</b>: We take a household-specific weighted average of the normalized resilience subcomponents, with weights derived from step 2; ii) <b>Rescaling</b>: We rescale this weighted average by the probability of falling below the threshold under the shock, estimated in step 3; iii) <b>Index scaling</b>: We multiply by 100 to obtain a 0-100 index; iv) <b>Discretization</b>: the continuous indicator can be discretized into a binary one using different criteria. This process is conducted for the full sample to estimate the resilience indicator for all observations; however, out-of-sample performance testing must be carried out exclusively on the testing set data points.</p>
<p><b>5. Out-of-sample testing – <u>Use only the testing set</u></b></p> <p>We assess the out-of-sample performance of CLARE in predicting the wellbeing outcome on the left-out testing set (later waves in forecasting; countries in out-of-country prediction) using a variety of common performance metrics for both regression and classification tasks using, respectively, the continuous and binary pairs of outcome and indicator. We also compare CLARE's performance with that of other mainstream measures of resilience and alternative approaches.</p>



# DATA DETAILS

[▶ BACK](#)

**Table 2: Sources of Household Data**

Country	Survey name	Years	Final n
Malawi	Integrated Household Panel Survey (IHPS)	2010/2011	<i>baseline</i>
		2013	1,282
		2016-2017	1,823
		2019	1,719
Nigeria	General Household Survey (GHS)	2010/2011	<i>baseline</i>
		2012/2013	3,619
		2015/2016	3,650
		2018/2019	1,219
Tanzania	Tanzania National Panel Survey (TZNPS)	2010/2011	<i>baseline</i>
		2012/2013	317
		2014/2015	687
		2019/2020	654
		2009/2010	<i>baseline</i>
Uganda	Uganda National Panel Survey (UNPS)	2010/2011	2,214
		2011/2012	2,211
		2013/2014	977
		2015/2016	2,607
		2018/2019	2,666
		2019/2020	2,467
<b>Total</b>	<b>4 countries</b>	<b>19 waves</b>	<b>28,112</b>

*Notes:* The table presents a summary of household-level data for each country, based on the LSMS Basic Information Documents.



# DATA DETAILS

[▶ BACK](#)

**Table A.1: Variable descriptions**

Variable name	Definition
Food Consumption Score (FCS)	Food Consumption Score
Food insecurity	FCS ≤ 35
Share of drought months	Share of growing season months experiencing drought (GS-EDDI > +1)
Drought (GS-EDDI)	= 1 if share of drought GS-EDDI > 75th percentile of the country-specific distribution
Multiple shock	= 1 if household experienced any shock <sup>23</sup>
Age of hh head	Age of household head
Gender of hh head	= 1 if Female household head
Household size	Household size
% members male ≤ 15 yrs	% members male ≤ 15 yrs
% members male 16-65 yrs	% members male 16-65 yrs
% members female 16-65 yrs	% members female 16-65 yrs
% members male > 65 yrs	% members male > 65 yrs
% members female > 65 yrs	% members female > 65 yrs
Rural	= 1 if household lives in a rural area
No. of hh members literate	No. of hh members literate
No. of hh members with primary education	No. of hh members with primary education
No. of hh members with secondary education or higher	No. of hh members with secondary education or higher
Household asset index	Household asset index
Mobile owned	= 1 if household own a mobile phone
TLU today	Tropical Livestock Units as of the time of survey
Number of key services within village	No. of key services within community (of 5)
Distance to market	Distance to market (km)

*Notes:* This table provides definitions for the key variables used in the analysis.



# DATA DETAILS

▶ BACK

Table A.3: Household asset index components

Malawi	Nigeria	Tanzania	Uganda
Mortar/pestle	furniture (3/4 piece sofa set)	Radio and Radio Cassette	House
Bed	furniture (chairs)	Telephone(landline)	Other buildings
Table	furniture (tables)	Telephone(mobile)	Land
Chair	mattress	Refrigerator or freezer	Furniture/furnishings
Fan	bed	Sewing Machine	Household appliances
Air conditioner	mat	Television	Television
Radio ('wireless')	sewing machine	Video / DVD	Radio/Cassette
Tape or CD/DVD player, HiFi	gas cooker	Chairs	Generators
Television	stove (electric)	Sofas	Solar panel/electric inverters
VCR	stove gas (table)	Tables	Bicycle
Sewing machine	stove (kerosene)	Watches	Motor cycle
Kerosene/paraffin stove	fridge	Beds	Motor vehicle
Electric or gas stove; hot plate	freezer	Cupboards, chest-of- drawers, boxes, wardrobes, bookcases	Boat
Refrigerator	air conditioner	Lanterns	Other Transport
Washing machine	washing machine	Computer	Jewelry and Watches
Bicycle	electric clothes dryer	Cooking pots, Cups, other kitchen utensils	Mobile phone
Motorcycle/scooter	bicycle	Mosquito net	Computer
Car	motorbike	Iron (Charcoal or electric)	Internet Access
Mini-bus	cars and other vehicles	Electric/gas stove	Other electronic equipment
Lorry	generator	Other stove	Other household assets (eg lawn mower)
Beer-brewing drum	fan	Water-heater	
Upholstered chair, sofa set	radio	Record/cassette player, tape recorder	
Coffee table (for sitting room)	cassette recorder	Complete music system	
Cupboard, drawers, bureau	hi-fi (sound system)	Books (not school books)	
Lantern (paraffin)	microwave	Motor Vehicles	
Desk	iron	Motor cycle	
Clock	tv set	Bicycle	
Iron (for pressing clothes)	computer	Carts	
Computer equipment & accessories	dvd player	Animal Cart	
Satellite dish	satellite dish	Boat/canoe	
Solar panel	musical instrument	Wheel barrow	
Generator		House(s)	
		Air-conditioned	
		Dish antenna/decoder	



## DESCRIPTIVE STATISTICS

▶ BACK

Table A.2: Summary statistics – Full sample

Variable	Obs	Mean	Std. Dev.	Min	Max
Food Consumption Score (FCS)	28,112	51.049	19.848	0	126
FCS (lag)	28,112	50.78	19.864	0	126
Food insecurity	28,112	.203	.402	0	1
Food insecurity (lag)	28,112	.21	.408	0	1
Share of drought months	28,112	.086	.093	0	.6
Drought (GS-EDDI)	28,112	.25	.433	0	1
Multiple shock	28,112	.597	.491	0	1
Age of hh head	28,047	49.224	15.278	10	113
Age of hh head (lag)	28,112	47.539	15.537	14	150
Gender of hh head	28,112	.262	.44	0	1
Gender of hh head (lag)	28,112	.245	.43	0	1
Household size	28,112	6.27	3.298	1	35
Household size (lag)	28,112	6.021	3.164	1	35
% members male <=15 yrs	28,112	.19	.373	0	.857
% members male <=15 yrs (lag)	28,112	.2	.379	0	.857
% members male 16-65 yrs	28,112	.23	.416	0	1
% members male 16-65 yrs (lag)	28,112	.241	.404	0	1
% members female 16-65 yrs	28,112	.247	.411	0	1
% members female 16-65 yrs (lag)	28,112	.255	.413	0	1
% members male >65 yrs	28,112	.053	.229	0	1
% members male >65 yrs (lag)	28,112	.044	.212	0	1
% members female >65 yrs	28,112	.067	.258	0	1
% members female >65 yrs (lag)	28,112	.052	.242	0	1
Rural	28,112	.737	.441	0	1
Rural (lag)	28,112	.737	.44	0	1
No. of hh members literate	28,099	3.127	2.358	0	22
No. of hh members literate (lag)	28,112	3.003	2.333	0	18
No. of hh members with primary education	28,099	1.006	1.282	0	13
No. of hh members with primary education (lag)	28,112	1.181	1.445	0	15
No. of hh members with secondary education or higher	28,099	.851	1.387	0	13
No. of hh members with secondary education or higher (lag)	28,112	.798	1.322	0	15
Household asset index	28,112	-.018	.978	-1.23	8.532
Household asset index (lag)	28,112	-.012	.96	-1.174	8.107
Mobile owned	28,094	.69	.462	0	1
Mobile owned (lag)	28,112	.668	.471	0	1
TLU today	27,719	1.203	5.607	0	480.9
TLU today (lag)	28,112	1.201	5.864	0	490.26
Number of key services within village	27,261	3.632	1.509	0	5
Number of key services within village (lag)	28,112	3.668	1.473	0	5
Distance to market	24,899	36.965	37.914	0	214.36
Distance to market (lag)	28,112	36.118	38.149	0	214.36

Notes: This table reports descriptive statistics for the variables used in the analysis, including both current and lagged values. Summary measures are based on the full pooled dataset across countries and waves.



# IMPORTANCE WEIGHTS - FORECASTING

▶ BACK

Resilience component	Importance weight (%)
FCS (lag)	23.347
Household asset index (lag)	20.668
Distance to market (lag)	8.472
% members female >65 yrs (lag)	8.399
Age of hh head (lag)	6.749
TLU today (lag)	5.265
No. of hh members with primary education (lag)	3.808
Number of key services within village (lag)	3.614
% members male <=15 yrs (lag)	3.282
% members male >65 yrs (lag)	3.098
% members female 16-65 yrs (lag)	3.054
% members male 16-65 yrs (lag)	2.614
No. of hh members literate (lag)	2.136
Member of hh owns mobile phone	2.029
Household size (lag)	1.812
No. of hh members with secondary education or higher (lag)	0.752
Gender of hh head (lag)	0.741
Rural (lag)	0.161



# IMPORTANCE WEIGHTS - SPEI DATA

[▶ BACK](#)

Resilience component	Importance weight (%)
Household asset index (lag)	18.067
FCS (lag)	13.773
Distance to market (lag)	12.537
Number of key services within village (lag)	10.867
% members male <=15 yrs (lag)	8.579
No. of hh members literate (lag)	6.015
Age of hh head (lag)	5.270
TLU today (lag)	4.571
Household size (lag)	3.996
N. of hh members with primary education (lag)	3.005
% members male 16-65 yrs (lag)	2.818
% members female 16-65 yrs (lag)	2.564
No. of hh members with secondary education or higher (lag)	2.510
% members female >65 yrs (lag)	2.335
% members male >65 yrs (lag)	1.830
Member of hh owns mobile phone	0.754
Rural (lag)	0.347
Gender of hh head (lag)	0.161



# IMPORTANCE WEIGHTS - MULTIPLE SHOCKS

▶ BACK

Resilience component	Importance weight (%)
FCS (lag)	27.827
Distance to market (lag)	13.035
Household asset index (lag)	8.493
% members female >65 yrs (lag)	7.419
% members female 16-65 yrs (lag)	7.110
Age of hh head (lag)	6.300
Household size (lag)	4.658
% members male 16-65 yrs (lag)	4.180
% members male >65 yrs (lag)	3.396
TLU today (lag)	3.384
% members male <=15 yrs (lag)	2.857
No. of hh members literate (lag)	2.714
Number of key services within village (lag)	2.622
No. of hh members with secondary education or higher (lag)	2.500
No. of hh members with primary education (lag)	2.250
Member of hh owns mobile phone	0.594
Gender of hh head (lag)	0.475
Rural (lag)	0.187



## RESULTS - FORECASTING

▶ ROBUSTNESS

▶ BACK

Metrics	Value
Pearson's correlation coefficient	0.534
Spearman rank correlation coefficient	0.524
Normalized RMSE (RMSE over FCS sample mean)	0.320
Average CLARE score	33.558

## Regression problem - All countries, waves 4-7.

▶ Descriptives

		Food insecure (FCS $\leq$ 35)		
		Food insecure = 0	Food insecure = 1	Total
Binary CLARE (median cutoff)	Non-resilient = 0	5,788	366	6,154
	Non-resilient = 1	4,427	1,728	6,155
	Total	10,215	2,094	12,309
	Correctly predicted	56.7%	82.5%	61.1%

## Classification problem - All countries, waves 4-7.

▶ Alt. cutoff



# WHO ARE THE RESILIENT?

▶ BACK

**Table C.15: t-test differences between households above and below the median CLARE score**

Variable	Resilient>median (mean)	Non-resilient<median (mean)	Difference in means
Food Consumption Score (FCS)	59.932	43.460	***
Food insecurity	0.059	0.281	***
Drought (GS-EDDI)	0.174	0.183	-
Age of hh head (lag)	45.158	49.676	***
Gender of hh head (lag)	0.230	0.365	***
Household size (lag)	6.636	5.455	***
% members male <=15 yrs (lag)	0.209	0.173	***
% members male 16-65 yrs (lag)	0.752	0.771	***
% members female 16-65 yrs (lag)	0.261	0.234	***
% members male >65 yrs (lag)	0.029	0.064	***
% members female >65 yrs (lag)	0.029	0.085	***
Rural (lag)	0.633	0.860	***
No. of hh members literate (lag)	3.971	2.275	***
No. of hh members with primary education (lag)	1.227	0.581	***
No. of hh members with secondary education or higher (lag)	0.983	0.373	***
Household asset index (lag)	0.370	-0.425	***
Mobile owned (lag)	0.907	0.522	***
TLU today (lag)	1.492	0.813	***
Number of key services within village (lag)	4.132	3.852	***
Distance to market (lag)	18.859	25.485	***

N: 28112

**Resilient vs. Non-resilient HHs - Forecasting, full sample, median cutoff.**



# ROBUSTNESS - FORECASTING

▶ BACK

Metrics	Reduced set	Alternative weather data	Multiple shock	Electricity Asset Index	Baseline CLARE
Pearson's correlation	0.501	0.515	0.533	0.523	0.534
Spearman rank correlation	0.501	0.504	0.522	0.513	0.524
Normalized RMSE	0.328	0.324	0.320	0.323	0.320
Average measure score	32.207	32.427	37.072	35.780	33.558

## Regression problem - All countries, waves 4-7, summary results.

Specification	Specificity	Sensitivity	Accuracy	N. of Observations
Reduced set (3 most important variables)	56.3%	80.9%	60.5%	12,309
Alternative weather data (SPEI)	56.5%	82%	60.9%	12,309
Multiple shock (general-purpose)	56.5%	81.8%	60.8%	12,309
Electricity Asset Index	56.6%	82.2%	61%	12,309
Baseline CLARE (median cutoff)	56.7%	82.5%	61.1%	12,309

## Classification problem - All countries, waves 4-7, summary results.



# FORECASTING WITH ALTERNATIVE CUTOFF

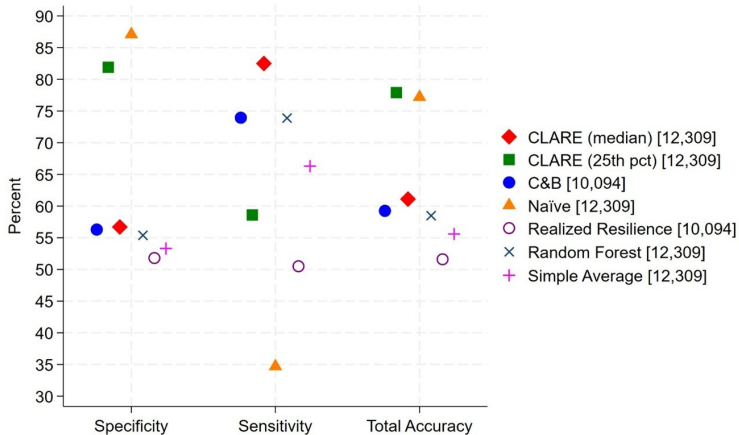
[▶ BACK](#)

		Food insecure (FCS $\leq$ 35)		
		Food insecure = 0	Food insecure = 1	Total
<b>Binary CLARE (25<sup>th</sup> percentile cutoff)</b>	<b>Non-resilient = 0</b>	8,365	866	9,231
	<b>Non-resilient = 1</b>	1,850	1,228	3,078
	<b>Total</b>	10,215	2,094	12,309
<b>Correctly predicted</b>		<b>81.9%</b>	<b>58.6%</b>	<b>77.9%</b>

**Classification problem - All countries, waves 4-7, 25<sup>th</sup> cutoff.**



# COMPARATIVE ASSESSMENT - FORECASTING



Out-of-sample performances for the classification problem

[▶ Back](#)



# RESULTS - OUT OF COUNTRY

▶ BACK

Metrics	Value
Pearson's correlation coefficient	0.567
Spearman rank correlation coefficient	0.562
Normalized RMSE (RMSE over FCS sample mean)	0.320
Average CLARE score	52.990

## Regression problem - All countries, all waves.

▶ Cissé and Barrett

		Food insecure (FCS $\leq$ 35)		
		Food insecure = 0	Food insecure = 1	Total
Binary CLARE (median cutoff)	Non-resilient = 0	13,089	967	14,056
	Non-resilient = 1	9,315	4,741	14,056
	Total	22,404	5,708	28,112
	Correctly predicted	58.4%	83.1%	63.4%

## Classification problem - All countries, all waves.

▶ Cissé and Barrett



# CISSÉ & BARRETT (2018) - OUT OF COUNTRY

▶ BACK

Metrics	Value
Pearson's correlation coefficient	0.327
Spearman rank correlation coefficient	0.318
Normalized RMSE (RMSE over FCS sample mean)	0.367
Average C&B RS	67.82

## Regression problem - All countries, all waves.

		Food insecure (FCS $\leq$ 35)		
		Food insecure = 0	Food insecure = 1	Total
Binary C&B RS (median cutoff)	Non-resilient = 0	12,598	1,458	14,056
	Non-resilient = 1	9,806	4,250	14,056
	Total	22,404	5,708	28,112
	Correctly predicted	<b>56.23%</b>	<b>74.46%</b>	<b>59.93%</b>

## Classification problem - All countries, all waves.

