Reproducible Research Repository

CLARE: A Causal Machine Learning Approach to Resilience Estimation

Third Submission: RR_AFR_2025_484
Maria Reyes Retana
reproducibility@worldbank.org
December 12, 2025

This review verifies the reproducibility of the exhibits included in the paper *"CLARE: A Causal Machine Learning Approach to Resilience Estimation"*.

Contents in this review:

Main findings

- **Every exhibit has been reproduced accurately.**
- The code was successfully executed on a new computer after:
  1. Changing the paths in the do file and scripts, and running the code in the mentioned order.
- The output demonstrates consistent stability across multiple runs. Specifically, executing the code two times consecutively yielded identical results.
- The code takes approximately 30 minutes to run.
- We conducted our reproducibility analysis based on the paper shared by the authors by email on November 6th.
- There were minimal differences between the results produced by the authors and those obtained by the replicators. These differences arise from the `grf` package. As noted in the `grf` documentation: although GRF is designed to produce identical estimates across platforms when using a fixed random seed (via the `seed` option), small discrepancies may still occur across different CPU architectures or compilers, even with the same seed. The minor differences observed are documented below.

  In addition, the authors' results were virtually verified by the replicators during a meeting on December 5. The replicators confirmed that the authors' outputs match exactly those presented in the paper. These verified results are also included in the package under the `outputs_replicators` folder.
- **Verification Process and Data Handling**:
  - The reproducibility package relies on 3 types of data: restricted, open and accesible data.

  - Only intermediate data are included in the reproducibility package. The restricted data are merged with open and accessible data at an early stage of the workflow, which prevents the full execution of the analysis starting from raw data. However, all results reported in the manuscript can be fully replicated using the included intermediate datasets. The code used to generate the intermediate data from the raw sources is also included in the package and was virtually verified by the replicators during a meeting held on November 20th.
  - Reviewers used intermediate data provided directly by the authors to conduct the reproducibility verification, and this is included in the public reproducibility package.
  - *data_hash_report.csv* lists the SHA256 hashes of all files in the Data folder to support data integrity checks.

- **Reproducibility Summary**:
  - **Data**: Some data is restricted and has not been included in the reproducibility package, but intermediate data to produce the results presented in the paper are included in the reproducibility package.
  - **Code**: All code files are included in the reproducibility package, but operate from intermediate data, not raw data.
  - **Outputs**: All outputs are generated by code included in the reproducibility package.
  - **Dependencies environment**: The reviewers created a new environment using the latest versions of dependencies available at the moment of the review.

List of exhibits and reproducibility status

Results in the Main Section of the Paper

- **Table 1** Does not show analysis results.
- **Table 2** Does not show analysis results.
- **Table 3** Reproduced requires manual steps. The results are shown in the Stata console.
- **Table 4** Reproduced requires manual steps. The results are shown in the Stata console. There is a minor difference in one of the observations, but since this is really small, this is considered reproducible. Differences are noted below for transparency.

Code

```
.
. * confusion matrix for the out-of-sample performance of binarized CLARE
. tab nonresilient food_insecurity if wave>=4  // This is Table 4 in the paper

nonresilie |    food_insecurity
        nt |         0          1 |     Total
-----------+----------------------+----------
         0 |     5,788        365 |     6,153
         1 |     4,427      1,729 |     6,156
-----------+----------------------+----------
     Total |    10,215      2,094 |    12,309
```

Paper

Table 4: Forecasting food insecurity status out-of-sample using CLARE

– All countries, waves 4 to 7; median CLARE cutoff

		Food insecure (FCS ≤ 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%**

Notes: This table shows the out-of-sample classification performance of the binary CLARE model in forecasting food insecurity (defined as FCS ≤ 35) across all countries and waves 4 to 7. The classification uses the median

- **Table 5** Reproduced requires manual steps. The results are shown in the Stata console. There is a minor difference in one of the observations, but since this is really small, this is considered reproducible. Differences are noted below for transparency.

Code

```
.
. * confusion matrix for the out-of-sample performance of binarized CLARE
. tab nonresilient_2 food_insecurity if wave>=4  // This is Table 5 in the paper
```

nonresilient_2	food_insecurity 0	1	Total
0	8,363	869	9,232
1	1,852	1,225	3,077
Total	10,215	2,094	12,309

Paper

Table 5: Forecasting food insecurity status out-of-sample using CLARE

– All countries, waves 4 to 7; 25th percentile CLARE cutoff

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

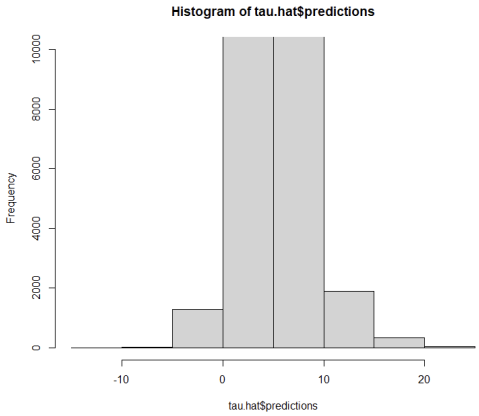
Notes: This table shows the out-of-sample classification performance of the binary CLARE model in forecasting food insecurity (defined as FCS ≤ 35) across all countries and waves 4 to 7. The classification uses the 25th percentile of CLARE as the cutoff. Rows indicate the predicted values (based on CLARE), while columns represent the observed food insecurity status.

- **Table 6** Reproduced requires manual steps. The results are shown in the Stata console.
- **Table 7** Reproduced requires manual steps. The results are shown in the Stata console.
- **Figure 1** Reproduced.
- **Figure 2** Reproduced.
- **Figure 3** Reproduced.
- **Figure 4** Reproduced.

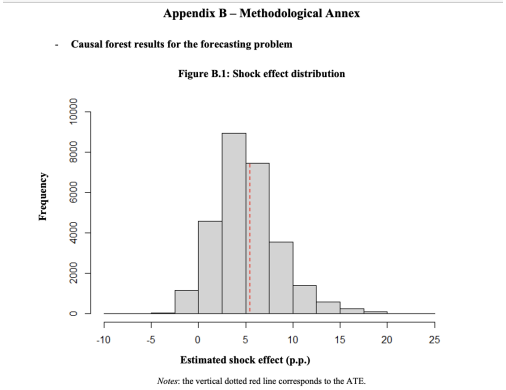
Results in the Annex

- **Table A.1** Does not show analysis results.
- **Table A.2** Reproduced. Verified against *descriptive_statistics.xlsx*
- **Table A.3** Does not show analysis results.
- **Table A.4** Does not show analysis results.
- **Figure B.1** Reproduced. This figure was virtually verified on December 5. The version reproduced by the replicators is shown here for transparency, but the figure presented by the authors during the virtual meeting matched exactly the version published in the paper.

Code



Paper

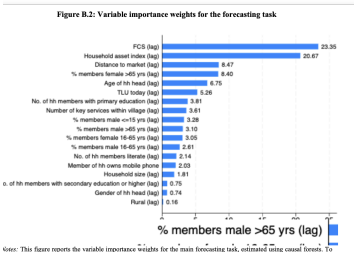


- **Figure B.2** Reproduced. This figure was virtually verified on December 5. The version reproduced by the replicators is shown here for transparency, but the figure presented by the authors during the virtual meeting matched exactly the version published in the paper.

Code

	V1	variable
1	0.212756181	l1hh_asset_index
2	0.210598907	l1fcs
3	0.086152120	l1female_65_share
4	0.082017163	l1dist_market
5	0.069160606	l1age_head
6	0.047439772	l1tlu_today
7	0.040231766	l1fem_16_65_share
8	0.035274289	l1n_services
9	0.033164852	l1n_primary_educ
10	0.031619036	l1male_16_65_share
11	0.031377473	l1male_65_share
12	0.028999994	l1male_15_share
13	0.025230091	l1mobile_owned
14	0.024954972	l1n_literate
15	0.022960750	l1hhsize
16	0.009002418	l1n_secondary_educ
17	0.007172602	l1fem_head
18	0.001887009	l1rural

Paper



- **Table B.1** Reproduced requires manual changes. The results are shown in the R console. There is a minor difference in the numbers, but since this is relatively small, it is considered reproducible. Differences are noted below for transparency.

Code

```
Best linear fit using forest predictions (on held-out data)
as well as the mean forest prediction as regressors, along
with one-sided heteroskedasticity-robust (HC3) SEs:

                           Estimate Std. Error t value    Pr(>t)
mean.forest.prediction      1.00153    0.15129  6.6199 1.855e-11 ***
differential.forest.prediction  0.95883    0.28432  3.3724 0.0003735 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Paper

Table B.1: Causal forest fit

Calibration test	Estimate	Standard error	t-value	Pr(>t)
Mean forest prediction	0.999	0.149	6.690	0.000***
Differential forest prediction	0.995	0.286	3.473	0.000***

Notes: These are the results of the calibration tests recommended by Athey et al. (2019) to assess the quality of the causal

Reproduction Environment

Paper exhibits were reproduced on a computer with the following specifications:

1. Replicator's computer
   - OS: Windows 11 Enterprise
   - Processor: Intel(R) Core(TM) i5-1145G7 CPU @ 2.60GHz
   - Memory available: 15.7 GB
   - Software version: Stata 18.0 MP, R 4.4.1
2. Author's computer
   - OS: Windows 10 x64
   - Memory available: 15.7 GB
   - Software version: Stata 18.0 MP, R 4.2.0