



OPEN

## New evidence on the rural poverty and energy choice relationship

Salvatore Di Falco<sup>1✉</sup> & Gary Lynam<sup>2</sup>

We combine a global micro level dataset that includes 17 different rural Sub-Saharan countries with satellite information about precipitation during the growing season to estimate the impact of economic conditions on energy choice. Differently from the existing literature, we aim to causally estimate the impact of household welfare variation on the likelihood of choosing a specific energy source. It is found, consistent with theory, that increases in income do determine an increase in the likelihood of using relatively cleaner and more efficient sources of fuel. We find, however, that this impact is quantitatively very small. Results hold conditional on assets, wealth and a large battery of controls and fixed effects. Policy implications are developed.

Around 3 billion people in the world are “energy poor”, meaning they rely upon harmful fuel sources like biomass-generated fire for their cooking and heating. The reduction of the pervasive use of wood-based combustibles implies two particularly important environmental and health objectives. First, less pressure on forests for the provision of energy sources. Second, less detrimental effects of indoor pollution as the use of these fuel sources have negative impacts on health through the inhalation of fine particles. These particles result in over 2.8 million premature deaths worldwide<sup>1,2</sup>. Transition to clean and affordable energy sources is a fundamental objective of sustainable development (Sustainable Development Goal no 7).

Economic forces can play an important role in facilitating energy choice and potential transition. A common conjecture is that the choice of a more efficient and clean energy source is a direct, or at least a natural, consequence of an increase in living standards. Energy source choices, as they reflect different levels of environmental quality, can be considered a luxury good. Increases in income, among other things, implies greater control over resources and more awareness of the effect of energy choice<sup>3,4</sup>. This conjecture is behind the widely studied energy ladder hypothesis<sup>5–10</sup>. In the developing and emerging world that means going from traditional inefficient sources (fuel wood, charcoal, kerosene etc.) to more efficient modern sources (liquid gas, electricity), once higher levels of income are attained. This paper tests this hypothesis by presenting evidence from households in 17 rural Sub-Saharan countries. Differently from the existing literature using multi-country datasets, our aim is to causally estimate the impact of income changes on energy choice<sup>11–14</sup>. To this end we use random variation in rainfall during the growing season of the main harvested crops. In rural environments, rainfall variations during this period provide a clean source of variation (also called natural experiment<sup>15</sup>) that is not systematically related to other variables. For instance, households experiencing less than average rainfall will obtain lower harvests and thus income. Using random variation in rainfall during the growing season, therefore allows us to make causal inference and circumvent the critical issue of accurate income measurement in developing countries. Our study focuses on such an environment, as we only use observations of rural farm households in the Demographic Health Survey (DHS). Table 1 shows that 80% of the sample relies on rural assets such as agricultural land for their living. Given the very low levels of irrigation, the welfare of these farm households heavily depends on the rainfall they obtain during the growing season.

We then use a multinomial logit to estimate the causal impact of random household income variation on the likelihood of moving up the energy ladder. To further address omitted variables concerns, a large battery of controls is added. These include household’s assets, wealth, education, and other socio-economic characteristics of the household head as well as access to electricity and infrastructure. A set of time and geographic fixed effects is also included to control for sources of unobservable heterogeneity that are space and time invariant.

### Data sources and variables

The dataset was generated by combining USAID’s Demographic Health Survey (DHS) country level household surveys with publicly available satellite information on precipitation accessed through the Google Earth Engine tool. The DHS survey information contains a set of data from representative national household surveys. Several

<sup>1</sup>Institute of Economics and Econometrics, Geneva School of Economics and Management, University of Geneva, Geneva, Switzerland. <sup>2</sup>Institute for Environmental Studies, University of Geneva, Geneva, Switzerland. ✉email: salvatore.difalco@unige.ch

Variable name	Description	Mean	Sd	Min	Max
Household owns farmland	1—owns farmland, 0—otherwise	0.8		0	1
Household owns livestock	1—owns livestock, 0—otherwise	0.7		0	1
# Household members	Count of full-time members in household	5.07	3.16	1	66
Household head age	The age of the household head	45.62	16.49	11	97
Household head education	Highest achievement of education of household head: none, (in)complete primary, (in)complete secondary and tertiary	3.88	4.39	0	24
Wealth index	The wealth index is calculated using data on a household's ownership of selected assets, such as electrical appliances; materials used for housing construction; and types of water access and sanitation facilities. Generated with a statistical procedure known as principal components analysis, the wealth index places individual households on a continuous scale of relative wealth. DHS separates all interviewed households into five wealth quintiles				
Precipitation seasonal mean	The wet seasonal precipitation—mean of the cluster (100 mm)—derived from the Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) <sup>16</sup>	10.40	6.43	0.009	42.82
Precipitation seasonal anomaly	The cluster standard deviation of the wet seasonal precipitation—mean from the 20-year seasonal precipitation mean (1997–2017)	– 0.15	1.05	– 3.57	3.03
Precipitation seasonal anomaly category type	Taking – 0.5 to 0.5 standard deviations (sd) to be the normal precipitation seasonal anomaly, four categories are defined as follows: Extreme drought (more than – 2 sd from the 20-year mean) Severe drought (between – 2 and – 0.5 sd from the 20-year mean) Severe flooding (between 0.5 and 2 sd from the 20-year mean) Extreme flooding (more than 2 sd from the 20-year mean)				
Year temperature mean	The yearly temperature mean of the cluster (Celsius)—derived from the MOD11A1 V6 satellite <sup>17</sup>	30.77	4.060	17.36	49.47
Soil moisture	The soil moisture (in the layer of 0–100 cm depth) mean of the cluster (kg/m <sup>2</sup> )—derived from the GLDAS-2 satellite using the 4-soil layer Noah model <sup>18</sup> . This is a measure of the quantity of water moisture in kilograms per square metre of a depth of 100 cm	210.3	55.37	53.27	400.3
Country electricity access (% of pop)	The percentage of the country's population that has access to the electricity grid, as given by The World Bank database—accessed June 2019				

**Table 1.** Variables definition and descriptive statistics.

years between 2006 and 2016 are available for each country, resulting in a pooled cross-section of 460,780 household observations (see Table A1). Filtering for rural households (defined as households residing in the countryside, as opposed to urban households that live in cities and towns), we are left with 310,707 observations across 17 sub-Saharan Africa countries: Burkina Faso, Burundi, Ethiopia, Ghana, Kenya, Lesotho, Liberia, Madagascar, Mali, Malawi, Mozambique, Namibia, Nigeria, Rwanda, Senegal, Sierra Leone, and Zimbabwe. The survey records the primary source of energy used for cooking by the household in each round. These are: biomass (e.g., crop residue, dung, fuel wood), charcoal, kerosene, liquid petroleum gas (LPG), and electricity. Each source represents a rung in the energy ladder, from the least efficient and most detrimental to household health and the environment, to the most efficient and clean. To build our key dependent variable, we organize and categorize these fuels into three distinct rungs of the ladder (as in Van der Kroon et al.<sup>10</sup>). The first rung includes the traditional fuel sources (biomass), the second includes the transitional sources (charcoal and kerosene), and the third encompasses the modern sources such as LPG and electricity. The DHS survey also records a large set of relevant controls. These include household assets (if the household owns any land or livestock), household wealth, the number of household members, the age of the household head, and the household head's level of education. We expand the list of controls by including two important variables at a higher level of aggregation. First, we include a control for the extent of access to electricity at the country level (the percentage of population with electricity access). Second, we include a control for soil quality as this captures the level of natural capital available to the household.

As previously mentioned, we use random variation in rainfall during the growing season as a measure of income variation at the household level. The GPS points of household clusters were used in conjunction with the Google Earth Engine tool to calculate relevant precipitation data for the corresponding year of the country survey. All the geographic data extractions used these publicly available cluster locations provided by the DHS Program surveyors. The GPS location of each cluster centre is recorded during either the fieldwork or listing stage of the survey process. Geo-masking methods are then used to ensure the confidentiality of the DHS respondents, displacing the cluster centres from their true locations by up to 10 kms. Therefore, we take this distance as the diameter of the circular area we calculate our geospatial data in, with the cluster GPS point as its centre. All geospatial data are calculated as annual means and are lagged 1 year prior to the DHS survey interview dates. The one exception for this is the main variable of interest, precipitation, which is calculated across the average planting/sowing month to the average harvest month for the main staple crops (cereals) in each country. The cereal crop calendar is presented in the Appendix.

We obtained the amount of rainfall (mm during the growing season) and calculated a set of precipitation anomaly dummy variables. We first computed a precipitation anomaly as the standardized difference between the observed level of rainfall during the growing season minus its long-term average (20-year seasonal precipitation average). This anomaly variable thus allows us to measure how much (in terms of standard deviation) the rainfall during the growing season is different from the long-term seasonal mean (20 years). These values were then converted into 4 categories: extreme drought (less than 2 standard deviations), mild drought (between less

than 2 standard deviations and 0.5), mild to severe flooding (more than 0.5 standard and less than 2), extreme flooding (torrential wet season with more than 2 standard deviations). This approach allowed us to analyse the seasonal extremes which let us investigate if there are any negative effects on the primary fuel choice. The definitions of the variables used in the analysis and the summary statistics are reported in the Table 1.

## Model

In order to model the household primary fuel choices, several multinomial logit (MNL) regressions were carried out<sup>19</sup>. MNL is a standard technique used in quantitative studies on energy transition which assumes households have a preference of energy type that maximises their perceived indirect utility function (more details are provided in A3). This allows one to assess how a given driver (income in this study) affects the choice of energy for a particular use conditional on a set of controls such as household head age, education, number of household members and so on<sup>11</sup>. The dependent variable is defined over three fuel type categories for their main source of cooking fuel, i.e. traditional, transitional and modern. For the  $i$ th household faced with the  $j$ th choice at time  $t$ , the utility function can be written as:

$$U_{ijt} = \beta' X_{ijt} + \varepsilon_{ijt}$$

where,  $X_{ijt}$  is a vector of explanatory variables for household  $i$ , in the region/country  $j$  at time  $t$ .  $\beta$  is a vector of unknown regression coefficients and  $\varepsilon_{ijt}$  is the error term. More details on the method are reported in the Appendix. For simplicity of exposition, we let the matrix  $X$  represent the key rainfall variable, the controls, and the fixed effects.

## Results

Table 2 reports the estimation results. The omitted baseline is the transitional category. Column (1) and (3) report the estimated marginal effects for the precipitation anomalies without controls. Columns (2) and (4) reports the marginal effects once all the controls are included. All specifications include time and country fixed effects (an alternative specification with regional fixed effects was excluded due to asymmetry in the variance matrix). Results are very consistent across specifications. Extreme drought and flooding have a positive impact on the probability that the household will utilise traditional biomass-based energy sources. Coherently with the energy ladder hypotheses, the same extreme variables have a negative impact on the probability to use modern energy sources. However, it should be noted that the magnitude of this effect is very small. Negative income shocks, such as extreme drought or flooding, increases the probability of going from transitional to traditional sources by less than 1%. Similarly, the same negative variation in income reduces the probability of choosing more modern energy sources by less than 1%. While the estimated income variations are associated with movement up or down the energy ladder consistent with the prior, the changes in probability seem very small. The results are robust to the inclusion of a large battery of controls. These include owning either farmland or livestock (which acts as a proxy for agricultural assets), soil moisture and temperature (which act as controls for agricultural production), a wealth index, education, and household size. Results are also robust to the inclusion of controls for access to electricity and vegetation conditions. We avoid making any interpretation of the controls' coefficient estimates as these may pick up simple correlations.

As further robustness checks, we use alternative metrics for the precipitation during the growing season. First, we use precipitation mean and its quadratic term. Second, we use the log of the precipitation mean. Both parametric specifications allow to further test the hypotheses consider the potential non-linear nature of the relationship between income and fuel choices. Results are again consistent and are reported in the Tables 3 and 4.

## Conclusions

This paper contributes to the energy transition literature by revisiting the energy ladder hypothesis. We expand on the existing literature in two dimensions. First, by analysing a large micro dataset that encompasses 17 countries. Second, by estimating the causal effect of rural income variations on fuel choice. The results show a weak impact of extreme precipitations on the choices of energy, only increasing the probability of choosing more traditional sources by approximately 1%. Given that precipitation captures the income variations derived from agricultural

Variables	(1)	(2)	(3)	(4)
	Trad base	Trad control	Mod base	Mod control
Extreme drought	0.0186*** (0.00475)	0.00565*** (0.00132)	- 0.000616 (0.000827)	- 2.22e-05** (1.13e-05)
Mild wet season drought	0.0107*** (0.00349)	- 0.000368 (0.000980)	- 0.00309*** (0.000695)	- 1.58e-05 (9.86e-06)
Mild to severe flooding	- 0.000185 (0.00397)	0.00160* (0.000952)	0.000560 (0.000869)	- 2.10e-05* (1.14e-05)
Extreme flooding	0.0177*** (0.00620)	0.00558*** (0.00179)	- 0.000835 (0.00101)	- 3.21e-05* (1.67e-05)
Controls	No	Yes	No	Yes
Country + time fixed effects	Yes	Yes	Yes	Yes
Observations	304,010	263,349	304,010	263,349

**Table 2.** Multinomial logits outcomes. Marginal effects. Clustered standard errors in parentheses. \*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.1$ .

Variables	(1)	(2)	(3)	(4)
	Trad base	Trad control	Mod base	Mod control
Precipitation seasonal mean (100 mm)	0.000927 (0.00075)	0.00134*** (0.00028)	- 0.000743*** (0.00014)	- 1.37e-05*** (2.95e-06)
Precipitation seasonal mean quadratic	- 0.00013*** (2.42e-05)	- 4.73e-05*** (8.14e-06)	3.08e-05*** (4.60e-06)	3.92e-07*** (9.44e-08)
Controls	No	Yes	No	Yes
Country + time fixed effects	Yes	Yes	Yes	Yes
Observations	304,010	263,349	304,010	263,349

**Table 3.** Model estimated replacing drought indicator with seasonal rainfall. Clustered standard errors in parentheses \*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.1$ .

Variables	(1)	(2)	(3)	(4)
	Trad base	Trad control	Mod base	Mod control
Log of precipitation seasonal mean	- 0.0243*** (0.00313)	0.000719 (0.00110)	- 0.00286*** (0.000389)	- 4.65e-05*** (6.63e-06)
Controls	No	Yes	No	Yes
Country + time fixed effects	Yes	Yes	Yes	Yes
Observations	304,010	263,349	304,010	263,349

**Table 4.** Model estimated replacing drought indicator with natural log of precipitation. Clustered standard errors in parentheses \*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.1$ .

output, we can observe that the income channel predicted in the energy ladder theory is weak relative to the expected impact. These results highlight that increases in economic conditions need to go beyond economic growth to reach the Sustainable Development Goal 7 of achieving universal access to clean energy.

It should be stressed that our results are consistent with the energy stacking literature<sup>20,21</sup>, where household characteristics (such as education, available labour for collecting firewood and cultural preferences), as well as conditions external to households (local market prices and stable supplies, as well as national policy programmes/ economic performance) are more important factors in determining energy choices among rural households<sup>10</sup>. This literature also states that households do not arbitrarily switch from one fuel to another for all household needs, but instead stack energy sources for use in different tasks. As several studies support that households indeed stack their stove fuel choices, this implies policy interventions should target the adaption of the cleanest stack and consider the local context for successful and long-lasting transition, rather than promoting the use of single fuel-stove combinations that do not meet the targeted community's needs<sup>22,23</sup>. However, to explore this further and thus better inform policy, national survey data needs to start focusing beyond primary fuel and look at energy stacks for household energy needs.

## Data availability

The datasets used and analysed during the current study available from the corresponding author on reasonable request.

Received: 21 April 2022; Accepted: 1 February 2023

Published online: 27 February 2023

## References

- Jeuland, M., Pattanayak, S. K. & Bluffstone, R. The economics of household air pollution. *Annu. Rev. Resour. Econ.* **7**, 81–108 (2015).
- OECD/IEA. Energy Access Outlook: from Poverty to Prosperity, World Energy Outlook Special Report. (2017).
- McCONNELL, K. E. Income and the demand for environmental quality. *Environ. Dev. Econ.* **2**, 383–399 (1997).
- Freeman, A. M. *The Measurement of Environmental Values and Resources: Theory and Methods* (Resources for the Future, 1993).
- Leach, G. The energy transition. *Energy Policy* **20**, 116–123 (1992).
- Heltberg, R. Factors determining household fuel choice in Guatemala. *Environ. Dev. Econ.* **10**, 337–361 (2005).
- Hosier, R. H. & Dowd, J. Household fuel choice in Zimbabwe. *Resour. Energy* **9**, 347–361 (1987).
- Narasimha Rao, M. & Reddy, B. S. Variations in energy use by Indian households: An analysis of micro level data. *Energy* **32**, 143–153 (2007).
- Kowsari, R. & Zerriffi, H. Three-dimensional energy profile. *Energy Policy* **39**, 7505–7517 (2011).
- van der Kroon, B., Brouwer, R. & van Beukering, P. J. H. The energy ladder: Theoretical myth or empirical truth? Results from a meta-analysis. *Renew. Sustain. Energy Rev.* **20**, 504–513 (2013).
- Heltberg, R. Fuel switching: Evidence from eight developing countries. *Energy Econ.* **26**, 869–887 (2004).
- Goldemberg, J. World energy assessment. Energy and the challenge of sustainability. (2001).
- Knight, K. W. & Rosa, E. A. Household dynamics and fuelwood consumption in developing countries: A cross-national analysis. *Popul. Environ.* **33**, 365–378 (2012).
- Burke, P. J. The national-level energy ladder and its carbon implications. *Environ. Dev. Econ.* **18**, 484–503 (2013).
- Rosenzweig, M. R. & Wolpin, K. I. Natural, "Natural Experiments" in Economics. *J. Econ. Lit.* **38**, 827–874 (2000).

16. Funk, C. *et al.* The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Sci. Data* **2**, 150066. <https://doi.org/10.1038/sdata.2015.662015> (2019).
17. Wan, Z., Hook, S. & Hulley, G. MOD11A1 MODIS/Terra Land Surface Temperature/Emissivity Daily L3 Global 1km SIN Grid V006. 2015, distributed by NASA EOSDIS Land Processes DAAC. (2019).
18. Rodell, M. *et al.* The global land data assimilation system. *Bull. Am. Meteorol. Soc.* **85**(3), 381–394 (2004).
19. McFadden, D. *Conditional Logit Analysis of Qualitative Choice Behaviour* (Acad. Press, 1974).
20. Masera, O. R., Saatkamp, B. D. & Kammen, D. M. From linear fuel switching to multiple cooking strategies: A critique and alternative to the energy ladder model. *World Dev.* **28**, 2083–2103 (2000).
21. Gebreegziabher, Z. *et al.* Fuel savings, cooking time and user satisfaction with improved biomass cookstoves: Evidence from controlled cooking tests in Ethiopia. *Resour. Energy Econ.* **52**, 173–185 (2018).
22. Shankar, A. *et al.* Everybody stacks: Lessons from household energy case studies to inform design principles for clean energy transitions. *Energy Policy* **141**, 111468 (2020).
23. Ruiz-Mercado, I. & Masera, O. Patterns of stove use in the context of fuel-device stacking: Rationale and implications. *EcoHealth* **12**, 42–56 (2015).

## Author contributions

Both authors wrote the main manuscript, run the analysis, and reviewed the manuscript in equal parts.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-29285-6>.

**Correspondence** and requests for materials should be addressed to S.F.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

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



**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/>.

© The Author(s) 2023