README for the Reproducibility Package for Pollution intensity of consumption: Exploring the Environmental Engel Curve based on micro-data from 109 countries

Overview

Abstract: Environmental Engel curves (EECs) characterize the relationship between the embodied environmental impact of household consumption choices and their respective income levels. Using detailed microdata from 109 countries across all income levels, we study household expenditure shares on 11 fuel types that differ in their air pollution intensity. The findings show that wealthier households tend to shift away from dirty fuels and toward cleaner ones, although this transition is not guaranteed. In low-income countries, limited infrastructure and poor access to clean fuels slow this process, demonstrating that income alone cannot drive energy transitions. By documenting systematic variation in EECs across income groups and national contexts, the paper emphasizes the joint role of income growth and infrastructure development. The analysis also reveals a dual burden of pollution in which richer households are the primary contributors to outdoor air pollution, while poorer households remain reliant on polluting fuels that increase their exposure to indoor air pollution.

Data Availability Statement

Dataset Name	File Name(s)	Link/Access Instuctions	Access Date	Data Access
DHS Data - Household Environmental Engle Curves	data/DHS Data Process/idhs_00009; idhs_00008.	https://www.idhsdata.org/idhs/	8/7/2023	Limited
CPAT Data - Household expenditure surveys	HH-EECs-V3.csv	https://www.worldbank.org/en/topic/climatechange/brief/climate-policy-assessment-tool	4/8/2023	Restricted

HH-EECs.csv

The data were obtained from the Climate Policy Assessment Tool page. The dataset was directly sourced from CPAT, and a subset showing the share of household expenditures on different fuel types was extracted from the CPAT datasets.

CPAT is a spreadsheet-based model covering more than 200 countries, developed jointly by the IMF and the World Bank to help policymakers evaluate climate-mitigation policies. The tool is primarily designed for use by governments and their technical advisors, rather than as a publicly downloadable product. The latest version of CPAT and its companion Multiscenario Tool can be requested by contacting the relevant World Bank country office. Further information and detailed instructions for obtaining the data are provided on this page:
https://www.worldbank.org/en/topic/climatechange/brief/climate-policy-assessment-tool

DHS Engel Curve data

Demographic and Health Surveys (DHS) data were obtained from:
https://www.idhsdata.org/idhs/

IPUMS-DHS, which harmonizes and facilitates access to DHS data collected in low- and middle-income countries since the 1980s. To access the data, users must create an account on the IPUMS-DHS website and, once approved, can download the datasets of interest. Detailed information and step-by-step instructions for accessing the data are available on the IPUMS-DHS webpage.

After obtaining the necessary files from IPUMS-DHS, the reproducibility package includes the code for merging men's and women's datasets and performing additional steps to construct the final dataset.

Statement about Rights

I certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.
I certify that the author(s) of the manuscript have documented permission to redistribute/publish the data contained within this replication package. Appropriate permissions are documented in the LICENSE.txt file.

Instructions for Replicators

1- Acquire the datasets from the Climate Policy Assessment Tool (CPAT) at the World Bank
2- Acquire the datasets from IPUMS-DHS for Environmental Engle Curve figures
3- Copy the data with the file names specified in the README to the folder "data".
4- Recreate the programming environment specified in the "renv.lock" file by installing the R package renv and running renv::restore() from the reproducibility package's working directory, or alternatively, manually install all the R libraries listed in the "renv.lock" file that are required by this reproducibility package.
Run the script " DHS_DataProcess.R"
Run the script " Figures 4-7.do"
Run the script " Figure3.R"
Given that full access to the reproducibility package data might be challenging for some users due to one dataset not being publicly available, we are including the code outputs generated by the code reviewers in the reproducibility verification in this reproducibility package. These are in the folder "outputs".

Computational Requirements and Runtime

Total runtime: 10 minutes
Some data requires purchase and is not included in the reproducibility package.

Software Requirements

R
STATA

Citation

Elizabeth Heger Boyle, Miriam King, and Matthew Sobek. IPUMS-Demographic and Health Surveys: Version 9 [dataset]. IPUMS and ICF, 2022. https://doi.org/10.18128/D080.V9