

# Reforming Land Valuation and Taxation in Ukraine: A path to sustainability and economic fairness

## Contents

1. Overview
2. Data Availability
3. Instructions for Replicators
4. Requirements
5. Code Description

## Overview

This project contains all R code that reproduces analysis conducted for the paper titled “Reforming Land Valuation and Taxation in Ukraine: A path to sustainability and economic fairness” by Klaus Deininger, Daniel Ayalew Ali, Eduard Bukin, and Andrii Martyn.

## Data Availability

Overall: **No data can be made publicly available.**

Reproducing this analysis depends on assembling confidential data from multiple sources. It is based on:

- land sales transactions;
- geo-spatial boundaries of land parcels;
- remotely sensed geographic characteristics about the analyzed land parcels; and
- sub-national administrative boundaries;

Geo-spatial boundaries of land parcels and sales transaction data are not publicly available and cannot be disseminated.

## Data Sources

### *Land sales transactions*

Transaction data used in this analysis was obtained directly from the corresponding registry of property rights, which is not publicly available. This data contains all sales, rental, and other transactions on land. Only sales transactions were used in the analysis, while the normative monetary value (NMV) was obtained from all available transactions.

- **Source:** State Registry of Property Rights on Immovable Assets (confidential) available [at the following link](#). This data was acquired on the request from the World Bank and provided by the Ministry of justice of Ukraine in a tabulated form. It contained a comprehensive list of variables and transactions on land sales, which occurred since 2021.

A similar data set, but with fewer transactions and a sole focus on sales, is prepared by the State Cadastre Service of Ukraine (DZK) and publicly disseminated on their Land Monitoring portal: <https://land.gov.ua/monitorynh-zemelnykh-vidnosyn/>. This alternative data set has a different structure compared to the one used in our analysis; therefore, the data preparataion code must be adjusted accordingly.

- **Source:** Land Relation Monitoring. Data on land relations monitoring (starting from 07.01.2021, bi-weekly) [Моніторинг земельних відносин. Дані моніторингу земельних відносин (починаючи з 01.07.2021р., щодвятижн)]
- **URL:** <https://land.gov.ua/wp-content/uploads/landMon/landTrans.zip>
- **Access year:** November 20, 2024.

### *Geo-spatial boundaries of the transacted land parcels*

To be able to compute the remotely-sensed features for each transacted parcel and for out-of-sample prediction, we use the centroids of our analyzed parcels. This data has not been publicly disclosed since February 2021; however, it was previously accessible through the public cadastre map of Ukraine at <https://map.land.gov.ua/>.

### *SoilGrids — global gridded soil information*

This raster data, disseminated with a 250 m resolution, is used to compute parcel-level soil characteristics. We access layers on pH levels, nitrogen content, organic matter content, and soil density and compute the average of each over a 0-60 cm depth. All computations are made in ArcGIS PRO, while data is downloaded as TIFF images through an API interface.

- **Source:** Poggio, L., de Sousa, L. M., Batjes, N. H., Heuvelink, G. B. M., Kempen, B., Ribeiro, E., and Rossiter, D.: SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty, SOIL, 7, 217–240, 2021. <https://doi.org/10.5194/soil-7-217-2021>
- **URL:** <https://files.isric.org/soilgrids/latest/data>.
- **Access year:** November 20, 2024

Specific variables used: Bulk density of the fine earth fraction,  $\text{cg}/\text{cm}^3$ .

- Files `bdod_{XXX}cm_mean`. Average over depth {XXX} 0-5, 5-15, 15-30, and 30-60 Organic carbon stocks , t/ha.
- File `ocs_0-30cm_mean`. One layer 0-30 cm. Total nitrogen (N),  $\text{cg}/\text{kg}$  .
- Files `nitrogen_{XXX}cm_mean`. Average over depth {XXX} 0-5, 5-15, 15-30, and 30-60 Soil pH,  $\text{pH} \times 10$ .
- Files `phh2o_{XXX}cm_mean`. Average over depth {XXX} 0-5, 5-15, 15-30, and 30-60

### *Land use classification*

This raster imagery provides data on land use at a 10m resolution. It classifies land use in Ukraine following an established methodology and annually updated ground-truth data.

- **Source:**
  - Shelestov A, et al., “Cloud Approach to Automated Crop Classification Using Sentinel-1 Imagery,” in *IEEE Transactions on Big Data*, vol. 6, no. 3, pp. 572-582, 1 Sept. 2020, doi: 10.1109/TBDATA.2019.2940237.
  - Kussul, N., Deininger, K., Shumilo, L., Lavreniuk, M., Ali, D. A., & Nivievskyi, O. (2022). Biophysical Impact of Sunflower Crop Rotation on Agricultural Fields. *Sustainability*, 14(7), 3965. <https://doi.org/10.3390/su14073965>
- **URL:** <https://ukraine-cropmaps.com>
- **Access year:** November 20, 2024

This website is only accessible without the WB VPN. If you disable global protect, you will be able to view it in the browser. We accessed this data directly from the researchers, who make it available online using the GeoServer WMS API.

The WMS API is accessible here: <https://ukraine-cropmaps.com/geoserver/ukraine/wms>

This data is disseminated annually and covers years 2020-2023, relevant for our analysis. Essentially, anyone should be able to download this data from the WMS server using the API and an appropriate software such as QGIS.

### *Macro data*

Data on the consumer price index (CPI) and the exchange rate between Ukrainian Hryvnia and United States Dollars (UAH to USD) is collected from the statistics at the National Bank of Ukraine.

- **URL:** <https://bank.gov.ua/ua/statistic>
- **Access year:** November 20, 2024

Variables: uah\_cpi\_monthly, uah\_cpi\_202312\_100, uah\_usd. Formula to compute: uah\_cpi\_202312\_100 (formula: =value for specific period\*(1+(value for 2023-12)/100)).

### *Administrative boundaries*

To identify the fixed effects relevant to each parcel, we intersect parcel geometries with administrative boundaries at different levels. Administrative boundaries for oblast, rayon, and community levels (admin 1-3) are publicly available, while more granular administrative boundaries are not disseminated publicly.

- **Source:** The Humanitarian Data Exchange. Ukraine - Subnational Administrative Boundaries.
- **URL:** <https://data.humdata.org/dataset/cod-ab-ukr>

- **Access year:** November 20, 2024

This data has not been publicly disclosed since February 2021; however, it was previously accessible through the public cadastre map of Ukraine at <https://map.land.gov.ua/>. Data used in this research was provided by the Cadaster as a geospatial data file with the geographical boundaries of admin 8 level and their corresponding aggregation to the admin 7, and 6 levels.

Currently a 2021 version of this data could be accessed at [https://kadastr.live/?land\\_polygons=false&dzk\\_index\\_map\\_lines=true#12.03/50.66176/32.04554](https://kadastr.live/?land_polygons=false&dzk_index_map_lines=true#12.03/50.66176/32.04554)

## Instructions for Replicators

There are two ways to replicate the analysis:

1. From the pre-cleaned data in the data-clean folder, or
2. From the raw data (*Limited reproducibility verification: code only*)

To reproduce the analysis from the pre-cleaned data:

1. Clone the RStudio project with the code to the local folder (preferably one not synchronized with any cloud drives).
2. Copy the data-clean folder with the pre-processed data into the project directory.
3. Open `0_main.R` script and run it. It will:
  - Install all missing packages
  - Run the `15-analysis.qmd` document, chunk-by-chunk, to produce all the plots and tables and store them as version-controlled “pins” and images in the output folder.
    - Note: Some figures for admin levels 6,7,and 8 are not produced by default because they are based on highly-detailed administrative boundaries. Rendering these figures may take very long!
    - Caution: Please do not “render” this document as most chunks are purposefully disabled to prevent the accidental start of some lengthy computations.
  - Run `20-paper-tables.qmd` by rendering it as a Quarto document to produce all tables and figures used in the analysis.
4. Results should be compiled as a word-document `20-paper-tables.docx` with all tables and plots.

To reproduce the analysis from the raw data, where only limited reproducibility verification is possible (code only) one need to have access to the raw data, which is not possible because of the data confidentiality. Code inspection could be done by:

1. Navigating to `10-data-cleaning.qmd` and investigating it chunk-by-chunk. This document assembles auxiliary data sets and saves them in `data-aux`. Then they are used to generate data in `data-clean`.
2. Follow steps 1-5 as described for reproducing the analysis from the pre-cleaned data.

## Requirements

### Computational Requirements

Any Windows, Linux, or Mac machine with a minimum of 32 GB of RAM should be sufficient to execute this analysis.

### Software Requirements

- **R 4.4.2 or higher**
  - `renv`
- **RStudio 2024.09.1+394 or higher**
- **Quarto 1.6.321.6.32 or higher**