

Testing for market power in agricultural land markets: Evidence from Ukraine

2025-09-10

This README file provides an explanation of how to reproduce the results, including descriptive and regression tables and figures, from the paper titled “Testing for market power in agricultural land markets: Evidence from Ukraine”.

Reproducibility Package

Introduction

This reproducibility package comprises essential elements that support the replication of the results described in the paper. The main components are as follows:

- (i) Raw data: The primary data sources for this project are administrative records obtained from various government agencies (specifically under the Ministry of Agrarian Policy and Food of Ukraine and the Ministry of Justice of Ukraine). Access to these datasets was granted exclusively for the purposes of this research as part of an ongoing collaboration; therefore, these data are confidential and cannot be shared or distributed under any circumstances. However, the package includes non-confidential raw data files from multiple sources, along with all necessary links and code to download publicly available raw datasets that supplement the analysis conducted in this study.
- (ii) Stata codes for processing raw datasets: The package includes Stata codes that merge data from various sources and perform comprehensive cleaning to produce analysis-ready data sets.
- (iii) Stata codes to reproduce tables and figures: Dedicated Stata code files are available for each set of tables and figures, enabling users to reproduce statistical analyses and generate the corresponding outputs.

Computational Requirements

Software requirements: Data cleaning and analysis are done using Stata 18. Raw data extraction and generation are done using ArcMap (used version 10.1) or QGIS.

Memory and Runtime Requirements: Most of the Stata codes were last run on a Windows 10 laptop with 32GB of RAM and i7-1365U processor with 1.8 GHz, but the codes that process cadastral parcel information were run on a server computer with 128GB of RAM and Platinum 8462Y+ 2.80 GHz (2 processors).

File Structure

To ensure that each step of the workflow is clearly documented and reproducible, the package separates raw data, processing scripts, analysis code, and outputs into distinct folders. This organization allows users to easily follow the progression from raw data to final results, making it straightforward to rerun analyses or adapt the workflow as needed. The folder structure of the reproducibility package is outlined below:

/UKRTranReprod/

— README.docx	# Project description and instructions
— Analysis/	# Scripts and results for data analysis
— Code/	# Analysis code files
— Output/	# Output generated from analysis
— DataManipulation/	# Raw data, manipulation and cleaning folders
— CleanData/	# Cleaned and processed datasets
— Code/	# Scripts for data merging and cleaning
— RawData/	# Raw, original datasets
— Temp/	# Temporary or intermediate files
— Paper/	# Manuscript

- **DataManipulation/:**

This folder is dedicated to data preparation, cleaning, and management. It organizes all files and scripts needed to transform raw data into analysis-ready datasets.

- ✓ **RawData/:**
Contains the original, unprocessed datasets as received from various data sources, organized into separate folders by theme and data sources. These files are not modified and serve as the starting point for all data manipulation. Note that these datasets are made available solely for verification purposes and will not be shared with the package.
- ✓ **Code/:**
Includes scripts for importing raw data, merging datasets, cleaning variables, and generating intermediary files. Each script is named to indicate the dataset it processes and, when relevant, its specific function (e.g., regression data for winter and summer seasons, spatial data preparation). The recommended order for running these scripts is provided in the “master.do” file.
- ✓ **CleanData/:**
Stores datasets that have been cleaned and processed, making them ready for analysis. These files are generated by running the scripts in the Code/ subfolder via the “master.do” file, and are used as input for the analysis scripts. Note that these datasets will be shared or distributed with the package.
- ✓ **Temp/:**
Contains temporary or intermediate files created during data processing. These files are used internally by the scripts and are not required for final analysis or output.

- **Analysis/:**

This directory includes all scripts and results (tables and figures) associated with the main data analysis.

- ✓ **Code/:**
Contains analysis scripts that utilize cleaned and processed data to produce results (tables and figures) included in the manuscript. Each script is named according to the corresponding table or

figure caption in the manuscript and can be run independently once the data is prepared. The recommended order for running these scripts is specified in the “master.do” file.

✓ **Output/:**

Contains all output generated by the analysis scripts, including summary statistics, regression results and figures. Each output file is labeled to match the table or figure number in the manuscript.

Data Availability

The analysis utilizes administrative raw data, including cadastral records, registry of rights (including land lease and sales transactions), and entries from the State Agrarian Registry, as provided by various Ukrainian government agencies. Access to these datasets was granted exclusively for the purposes of this research within the framework of an ongoing institutional collaboration with the Ministry of Agrarian Policy and Food of Ukraine and the Ministry of Justice of Ukraine. Consequently, these datasets are confidential and cannot be shared, distributed, or disclosed under any circumstances. Researchers seeking access to the data for the purpose of replicability may contact Daniel Ali (dali@worldbank.org) and Klaus Deininger (kdeininger@worldbank.org). Any such requests will be considered and supported to the extent permitted by the data sharing agreements with the respective institutions.

Outlined below are the sources for raw datasets, the procedures used for data generation and extraction, and, where available, the scripts utilized for data retrieval across various platforms.

Agricultural land sales and lease transactions (DZK_Trans): This dataset serves as the primary source for the analysis and comprises detailed records of land lease and sales prices at both the cadastral parcel and transacting party levels for transactions conducted between January 1, 2021, and December 31, 2024. In addition to transaction prices, the dataset includes supplementary parcel attributes, such as area and land use category. Access to this dataset, which is an extract from the cadastral records and registry of rights, was granted by the Ministry of Agrarian Policy and Food of Ukraine. The dataset is confidential and cannot be shared or distributed.

List of cadastral parcels (Cadaster): This dataset comprises a comprehensive list of cadastral parcels organized by oblast, containing information on parcel sizes. Each transacted parcel can be linked to these parcels using its unique cadastral number. Access to this dataset was provided by the Ministry of Agrarian Policy and Food of Ukraine exclusively for the purposes of this research project and cannot be shared or distributed.

List of farms (individual and legal entities) from State Agrarian Registry (SAR_RRP): This dataset contains list of parcels along with the registered parties (owners or tenants), extracted in Sep. 2024 and Nov. 2024, recorded in the SAR. This dataset is used to establish the holding size of farms that have participated in the agricultural lease or sales markets since 2021. Access to the data was provided by the Ministry of Agrarian Policy and Food for Ukraine for the purpose of this research and cannot be shared or distributed.

NAIS (registry of rights): This dataset consists of a list of legal entities, along with their registered rights, that participated in land lease and sales transactions but are not yet registered in SAR. It supplements the information extracted from SAR to help determine holding size. Access was provided by the Ministry of Justice of Ukraine for the purpose of this research and cannot be shared or distributed.

Sinergise fields (Sinergise): To generate operational agricultural field -level characteristics for each transacted parcels, the cadastral shapefile was intersected with nearly one million field boundaries generated by Sinergise in 2022. Two datasets were produced: (1) the area of overlap between Sinergise fields and cadastral parcels, and (2) the area of fields not yet surveyed or included in the cadaster. The methodology for delineating field boundaries is described at <https://area->

monitoring.sinergise.com/docs/markers/field-delineation/. As the datasets include cadastral information, they cannot be shared or distributed.

Crop cover data: Annual crop maps (both for winter and summer crops) for all of Ukraine from 2019 to 2024 were produced using the machine learning methodology described by Kussul et al. (2025). Training data were collected through two rounds of in-situ sampling along main roads each year, except in 2022, when the conflict limited data collection to a single round. For each agricultural season, all available optical imagery from Sentinel-2 and Synthetic Aperture Radar (SAR) data from Sentinel-1 were used to train either a multilayer perceptron or a random forest model. These models were designed to classify specific crop types, rather than simply identifying the presence of vegetation. The crop maps used in the paper are available from <https://ukraine-cropmaps.com/geoserver/wms>.

Crop maps were overlaid with both cadastral shapefile and the village shapefile (ADMIN_3.shp, available at https://github.com/justinelliotmeyers/official_ukraine_administrative_boundary_shapefile/) to calculate time-series cultivated area at the cadastral parcel and village council level. The dataset with the cadastral numbers is confidential and cannot be shared.

Conflict indicators (Conflict_Damage and Conflict_Public): These datasets provide bi-weekly measures of conflict-induced field damages at the village level (i.e., the area of affected fields hand-labeled using freely available Sentinel-2 (S2) imagery). Additionally, they include bi-weekly binary conflict indicators derived from publicly available situational reports. These datasets are the same as those utilized by Deininger et al. (2025). Comprehensive details regarding the methodology and the raw data can be found at: <https://reproducibility.worldbank.org/catalog/329>.

Distance to frontline (DistanceFL): To calculate the Euclidean distance from transacted parcels to the frontline, a map of conflict-affected amalgamated communities was constructed using publicly available situational reports. This distance dataset was produced by overlaying the cadastral map of leased parcels with the mapped frontline. This dataset cannot be shared or distributed, consistent with the restrictions described above.

OSM: To calculate Euclidean distance from each cadastral parcels to public infrastructure (population centers, elevators, motorways, ports, primary, secondary and tertiary roads, and railway lines), the centroids of cadastral parcels were overlaid with OpenStreetMap (OSM) data. The data were retrieved on December 7, 2022, except for population centers and elevators which were retrieved on March 19, 2024, from <http://export.hotosm.org>.

Soil: Cadastral parcel polygons were overlaid with publicly available SoilGrids raster data (Poggio *et al.* 2021) with a resolution of 250 m to obtain mean pH, soil density, and N as well as organic matter content in the 0-60 cm horizon for each parcel as a measure of soil quality. Soil quality data is at from <https://files.isric.org/soilgrids/latest/data>.

OTG, Village: OTG boundaries that were retrieved from [Ukraine - Subnational Administrative Boundaries | Humanitarian Dataset | HDX](#) (the geodatabase for subnational boundaries was downloaded on June 16, 2022 and the name of the OTG shapefile was “ukr_admbnda_adm3_sspe_20220114”) and village boundaries (available at https://github.com/justinelliotmeyers/official_ukraine_administrative_boundary_shapefile/ with a name ADMIN_3) were overlaid with the centroid of cadastral parcels to create a correspondence table linking parcels to their geographical locations.

ExchangeRate_NBU: The monthly official exchange rate of the Hryvnia to the US dollar, covering the period from January 2021 to December 2024, was downloaded from website of the National Bank of Ukraine (data available at: [Official Exchange Rates](#)).

References

- Klaus W. Deininger; Daniel Ayalew Ali; Kussul Nataliia; Guido Lemoine; Andrii Shelestov; Leonid Shumilo. 2025. Using Remotely Sensed Data to Assess War-Induced Damage to Agricultural Cultivation: Evidence from Ukraine. Policy Research Working Paper 1221, Washington, D.C.: World Bank Group. <https://documentsinternal.worldbank.org/search/40047588>
- Poggio, L., L. M. de Sousa, N. H. Batjes, G. B. M. Heuvelink, B. Kempen, E. Ribeiro and D. Rossiter. 2021. "Soilgrids 2.0: Producing Soil Information for the Globe with Quantified Spatial Uncertainty." *SOIL* 7 (1): 217-240.