Local GDP Estimates Around the World*

Esteban Rossi-Hansberg
University of Chicago

Jialing Zhang
Princeton University

April 27, 2026

Definition of Columns and Notes

This file provides definitions and descriptions of the columns in the dataset, along with related notes for analysis.

1 Definition of Columns

1. **cell_id**: Refers to 1°x1° cell. The world is divided into 1-degree longitude and 1-degree latitude squares. The *cell_id* is assigned in the order from left to right, and then top to bottom.

2. **subcell_id**: Refers to 0.5°x0.5° cell. It is obtained by dividing the parent 1°x1° cell into four squares, with IDs ranging from 1 to 4.

3. **subcell_id_0_25**: Refers to 0.25°x0.25° cell. It is obtained by dividing the parent 0.5°x0.5° cell into four squares, with IDs ranging from 1 to 4.

4. **iso**: The three-letter ISO code representing the country.

5. **year**: The available year of observation.

6. **predicted_GCP_const_2021_USD**: Predicted GDP of the cell (or polygon for boundary cells) in billion constant 2021 USD.

7. **predicted_GCP_current_USD**: Predicted GDP of the cell (or polygon for boundary cells) in billion current USD.

8. **predicted_GCP_const_2021_PPP**: Predicted GDP of the cell (or polygon for boundary cells) in billion constant 2021 PPP-adjusted dollars (international dollars).

9. **predicted_GCP_current_PPP**: Predicted GDP of the cell (or polygon for boundary cells) in billion current PPP-adjusted dollars (international dollars).

*Esteban Rossi-Hansberg: earossih@uchicago.edu. Jialing Zhang: zhang.jialing@princeton.edu

10. **predicted_GCP_[unit]_q05**: The 5th percentile of predicted GDP across the approximately 1,500 trees in the random forest ensemble, after propagating each tree's prediction through the full post-processing pipeline (censoring, rescaling, and multiplying by national GDP). Reported for each currency unit: *predicted_GCP_const_2021_USD_q05*, *predicted_GCP_current_USD_q05*, *predicted_GCP_const_2021_PPP_q05*, *predicted_GCP_current_PPP_q05*.

11. **predicted_GCP_[unit]_q95**: The 95th percentile of predicted GDP across the random forest trees, after the same post-processing steps as above. Reported for each currency unit: *predicted_GCP_const_2021_USD_q95*, *predicted_GCP_current_USD_q95*, *predicted_GCP_const_2021_PPP_q95*, *predicted_GCP_current_PPP_q95*.

12. **predicted_GCP_[unit]_tree_sd**: The standard deviation of predicted GDP across the random forest trees, after post-processing. Reported for each currency unit: *predicted_GCP_const_2021_USD_tree_sd*, *predicted_GCP_current_USD_tree_sd*, *predicted_GCP_const_2021_PPP_tree_sd*, *predicted_GCP_current_PPP_tree_sd*.

13. **GCP_sd_log_gdp**: The standard deviation of log(GDP) across the random forest trees. Because this measure is in log space, it is invariant to the currency unit and captures proportional uncertainty—a value of 0.10 corresponds to roughly 10% uncertainty in GDP levels.

14. **pop_cell**: Population of the cell (or polygon for boundary cells). This data is extracted from LandScan and rescaled to align with national population data from IMF WEO or the World Bank. See the Appendix Table 3 for sources by country. Please ensure to also cite LandScan when using this column.

15. **cell_GDPC_const_2021_USD**: GDP per capita of the cell (or polygon for boundary cells) in constant 2021 USD, calculated as *predicted_GCP_const_2021_USD / pop_cell*.

16. **cell_GDPC_current_USD**: GDP per capita of the cell (or polygon for boundary cells) in current USD, calculated as *predicted_GCP_current_USD / pop_cell*.

17. **cell_GDPC_const_2021_PPP**: GDP per capita of the cell (or polygon for boundary cells) in constant 2021 PPP-adjusted dollars, calculated as *predicted_GCP_const_2021_PPP / pop_cell*.

18. **cell_GDPC_current_PPP**: GDP per capita of the cell (or polygon for boundary cells) in current PPP-adjusted dollars, calculated as *predicted_GCP_current_PPP / pop_cell*.

19. **is_cell_censored**: A value of 1 indicates that post-adjustments have affected the cell.

20. **method**: Refers to the method used for post-adjustments.

21. **cell_size**: Represents the size of the cell.

22. **national_population**: National population of the country, obtained from IMF WEO or the World Bank. See the Appendix Table 3 for sources by country.

23. **longitude**: Longitude of the bottom-left corner of each cell.

24. **latitude**: Latitude of the bottom-left corner of each cell.

25. **geometry**: Geometry of each cell (or polygon for boundary cells). This column is split into shapefile folder.

2 Notes

1. Sparse population cells may still exhibit overestimated GDP values. To address this, we apply a post-adjustment by censoring predicted cell GDP shares to zero for areas with population densities less than or equal to 0.01, 0.02, or 0.05 individuals per square kilometer. Cells located along country or state borders may form irregular polygons instead of standard grid cells. For these irregular polygons, the same censorship is applied at the polygon level. After aggregating these polygons into standard cells, some cells may still display positive GDP values, even when the *is_cell_censored* column is set to 1.

2. Some cells along country boundaries may overlap multiple countries. In such cases, the cell may not form a regular square but conforms to country boundaries. The columns *cell_id*, *subcell_id*, *subcell_id_0_25*, *cell_size*, *longitude*, and *latitude* pertain to the entire square cell. However, all values relating to GDP, GDP per capita, population, and prediction uncertainty — including *predicted_GCP_[unit]*, *pop_cell*, *cell_GDPC_[unit]*, *predicted_GCP_[unit]_q05/q95*, *predicted_GCP_[unit]_tree_sd*, and *GCP_sd_log_gdp* — relate to the portion within the indicated country in the iso column.

3. Geometry files are in the *shapefile* folder. For boundary cells, geometry is stored as polygons.

4. GDP predictions may slightly differ across resolutions due to the use of distinct models (1°, 0.5°, 0.25°). See Appendix Section 7 for the consistency test of predictions across resolutions.

5. Our cell population data is sourced from LandScan. Before extracting the population for each cell, large inland water areas are removed. The data is then rescaled to align with national population data from the IMF or World Bank. As a result, the population data may differ from the original LandScan data.