Codebook: Utility Efficiency, Market Structure, and Access to Electricity

Edgar Aguilar

April 20, 2025

Contents

1 Overview, Sources, and Construction

This codebook accompanies the dataset Efficiency_Markets_Access.csv, which merges utility-level performance metrics with country-level electricity market structure, governance characteristics, and national electricity access rates. The dataset supports a benchmarking analysis of 49 electric utilities across 37 countries and includes 21 cleaned and standardized variables.

The dataset brings together three complementary snapshots:

1. Utility Performance (2012–2022)

Inputs (e.g., employees and assets) and outputs (e.g., net profit margin, customer count, distribution losses) are averaged across 2012–2022 using the UPBEAT Utility Database. This decadal average smooths year-to-year volatility and captures each utility's steady-state performance.

2. Market Structure & Governance (2012)

Characteristics like vertical integration, independent regulation, and IPP presence are fixed at their 2012 values, taken from the World Bank electricity market reform dataset. This aligns with the beginning of the performance window.

3. Electricity Access (2022)

Country-level access data are drawn from the World Development Indicators, anchoring each utility to the most recent year of access data.

All data-processing steps—from cleaning and merging to reshaping and calculating DEA scores—are fully reproducible via the replication script Efficiency_scores-market-access_dataset_creation.R.

Primary Data Sources:

- UPBEAT Utility Database: Utility performance and financial indicators.
- World Bank Market Structure Dataset: Country-level electricity market structures and governance characteristics.
- World Bank - Access to Electricity (WDI): National electricity access rates.

2 Variable Descriptions and Summaries

2.1 Efficiency Score

- **Type:** Numeric
- **Description:** A Data Envelopment Analysis (DEA) efficiency score ranging from 0 to 1, where 1 indicates a fully efficient utility. Scores were computed using an input-oriented CCR model based on selected input (Employees, Total Assets) and output (Net Profit Margin, Customers, Transformed Distribution Losses) variables.
- **Methodology:** The score was computed using an input-oriented CCR DEA model (constant returns to scale), implemented via the Benchmarking package in R. Inputs included Employees and Total Assets; outputs included Net Profit Margin, Customers, and Transformed Distribution Losses. Input/output values were averaged across 2012–2022. Utilities with missing data were excluded from the model.
- **Source:** Constructed from UPBEAT database.

Table 1: Summary - Efficiency Score

Min	Q1	Median	Mean	Q3	Max	NAs
0.0373708	0.2966317	0.5430189	0.5583466	0.8297779	1	0

2.2 Utility Long Name

- **Type:** Character
- **Description:** Full legal name of the electric utility.
- **Source:** UPBEAT

2.3 Utility Short Name

- **Type:** Character
- **Description:** Abbreviated or acronym version of the utility name.
- **Source:** UPBEAT

2.4 Ownership

- **Type:** Factor
- **Description:** Ownership structure of the utility (Public or Private).
- **Source:** UPBEAT

Table 2: Ownership Structure of Utilities

Ownership	Count
Private	11
Public	38

2.5 Type

- **Type:** Factor
- **Description:** Categorical variable describing the operational structure of the utility. All utilities in the dataset include a **distribution component** in their operations, but may also engage in other segments of the power sector:
  - **D (Distribution-only):** Utility is solely responsible for electricity distribution.
  - **G&D (Generation and Distribution):** Utility is involved in both electricity generation and distribution, but not transmission.
  - **VIU (Vertically Integrated Utility):** Utility is active across the entire electricity value chain, including generation, transmission, and distribution.
- **Source:** UPBEAT

Table 3: Distribution of Utility Operational Structures

Type	Count
D	25
G&D	3
T&D	1
VIU	20

2.6 Country

- **Type:** Character
- **Description:** Country of operation.
- **Source:** World Bank Market Structure Dataset

Table 4: List of Countries in the Dataset

Countries
Angola, Argentina, Armenia, Bangladesh, Benin, Bolivia, Bosnia and Herzegovina, Brazil, Burkina Faso, Cameroon, Central African Republic, Colombia, Democratic Republic of the Congo, Dominica, Ecuador, Federated States of Micronesia, Ghana, Indonesia, Ivory Coast, Kazakhstan, Kyrgyzstan, Malaysia, Marshall Islands, Mozambique, Namibia, Niger, Pakistan, Paraguay, Peru, Saint Lucia, Senegal, South Africa, Sudan, Thailand, Uganda, Zambia, Zimbabwe

The dataset includes 37 unique countries.

2.7 Region

- **Type:** Factor
- **Description:** Geographical region classification assigned to each country.
- **Source:** UPBEAT (World Bank classification)

Table 5: Distribution of Utilities by Geographical Region

Region	Count
AFR	19
EAP	6
ECA	5
LAC	13
SAR	6

2.8 Income Group

- **Type:** Factor
- **Description:** Country income classification (Low, Lower-middle, etc.).
- **Source:** UPBEAT (World Bank classification)

Table 6: Distribution of Utilities by Income Group

Income Group	Count
Low income	7
Lower middle income	19
Upper middle income	23

2.9 Employees

- **Type:** Numeric
- **Description:** Total number of employees in utility company.
- **Source:** UPBEAT

Table 7: Summary - Employees

Min	Q1	Median	Mean	Q3	Max	NAs
143.75	607.3636	3076	6679.158	6311.125	51150.36	0

2.10 Total Assets

- **Type:** Numeric
- **Description:** Total assets in USD millions.
- **Source:** UPBEAT

Table 8: Summary - Total Assets

Min	Q1	Median	Mean	Q3	Max	NAs
27.99988	388.9492	1259.999	5182.658	2096.416	89969.89	0

2.11 Net Profit Margin

- **Type:** Numeric
- **Description:** Net profit as a percentage of revenue.
- **Source:** UPBEAT

Table 9: Summary - Net Profit Margin

Min	Q1	Median	Mean	Q3	Max	NAs
-0.766375	-0.0328633	0.01552	-0.0259443	0.0484	0.193	0

2.12 Customers

- **Type:** Numeric
- **Description:** Number of customers served by utility company.
- **Source:** UPBEAT

Table 10: Summary - Customers

Min	Q1	Median	Mean	Q3	Max	NAs
1379.937	413678.7	987979.8	3280708	2203402	68146061	0

2.13 Distribution Losses

- **Type:** Numeric
- **Description:** Share of electricity lost in distribution.
- **Source:** UPBEAT

Table 11: Summary - Distribution Losses

Min	Q1	Median	Mean	Q3	Max	NAs
0.03485	0.0834	0.131	0.1493647	0.1929817	0.3576	0

2.14 Transformed Distribution Losses

- **Type:** Numeric
- **Description:** 1 minus distribution losses.
- **Source:** UPBEAT (calculated)
- **Note:** This is a transformation of "Distribution Losses" variable to fit DEA model appropiately as an output.

Table 12: Summary - Transformed Distribution Losses

Min	Q1	Median	Mean	Q3	Max	NAs
0.6424	0.8070183	0.869	0.8506353	0.9166	0.96515	0

2.15 Operating and Debt Service Cost Recovery

- **Type:** Numeric
- **Description:** Ratio of billed revenue to cost of service, calculated on a billed basis (excluding subsidies and collections). This reflects whether the utility is generating enough billed revenue to cover its operating costs and debt service obligations.
- **Source:** UPBEAT

Table 13: Summary - Cost Recovery

Min	Q1	Median	Mean	Q3	Max	NAs
0.331851	0.8831145	0.9851942	0.932262	1.04524	1.175037	30

2.16 Market Structure

- **Type:** Factor

- **Description:** Categorical variable representing the utility's national power market structure, based on a simplified classification derived from World Bank market reform data. The three categories are:
  - **VIU (Vertically Integrated Utility):** The utility operates within a fully integrated system controlling generation, transmission, and distribution.
  - **SBM (Single Buyer Model):** A central purchasing entity procures electricity from generators and sells it to distributors or end-users.
  - **WRC (Wholesale-Retail Competition):** Market features competition at both the wholesale and retail levels, allowing multiple sellers and/or buyers.
- **Note:** This classification reflects the extent of market liberalization and unbundling reforms in the utility's country of operation as of 2012.
- **Source:** World Bank Market Structure Dataset

Table 14: Distribution of Utilities by Market Structure

Market Structure	Count
Single Buyer Model (SBM)	22
Vertically Integrated Utility (VIU)	15
Wholesale-Retail Competition	12

2.17 Regulator

- **Type:** Logical
- **Description:** TRUE if an independent regulator exists.
- **Source:** World Bank Market Structure Dataset

Table 15: Presence of Independent Regulators

Regulator	Count
FALSE	19
TRUE	30

2.18 Private IPP Established

- **Type:** Logical
- **Description:** TRUE if Independent Power Producer (IPP) exist.
- **Source:** World Bank Market Structure Dataset

Table 16: Existence of Independent Power Producers (IPPs)

IPP Established	Count
FALSE	20
TRUE	29

2.19 Unbundling

- **Type:** Logical
- **Description:** TRUE if transmission is unbundled.
- **Source:** World Bank Market Structure Dataset

Table 17: Transmission Unbundling Status

Unbundling	Count
FALSE	23
TRUE	26

2.20 Out of VIU

- **Type:** Logical
- **Description:** TRUE if the utility is no longer vertically integrated.
- **Source:** World Bank Market Structure Dataset

Table 18: Utilities Exiting Vertically Integrated Utility (Out of VIU)

Out of VIU	Count
FALSE	15
TRUE	34

2.21 Access 2022

- **Type:** Numeric
- **Description:** Percent of national population with electricity access in 2022.
- **Source:** World Bank World Development Indicators (WDI)

Table 19: Summary - Access to Electricity (2022)

Min	Q1	Median	Mean	Q3	Max	NAs
15.7	63.2	96.2	80.45102	100	100	0

3 Notes

- DEA efficiency scores were calculated using the CCR input-oriented model.
- Monetary figures are in millions of USD.
- Logical fields show TRUE/FALSE flags for market conditions.
- This dataset was developed as part of my Substantial Research Paper (SRP) to fulfill the requirements for the Master of Arts in International Economic Relations: Quantiative Methods at American University. The data preparation, analysis, and documentation were completed in Spring 2025.