

CorruptionPolitEcon Codebook

Jackson Mandel

2025-04-20

Overview

This dataset combines several key indicators around democracy, political openness, informal employment, and corruption. The objective is to facilitate the study of interactions between corruption and political economy on a cross-country comparative basis.

Sources and Methodology

The data in this dataset is derived from three key sources: the V-Dem dataset, Transparency International's Corruption Perceptions Index, and the International Labor Organization's ILOSTAT data on informal unemployment.

V-Dem Data

6 indicators (Polyarchy, LibDem, ParticipDem, EgalDem, FreeFairElect, EqualAccess) are derived from the V-Dem dataset. The V-Dem data is in the form of ratings on a scale from 0 to 1, assessed by country experts. In all of the variables used in this dataset, higher scores represent greater presence of a particular feature (free and fair elections, egalitarian democracy, etc.) V-Dem typically uses five experts per country-year observation, out of a pool of around 4,000 total experts. They use a model to algorithmically estimate the reliability of each expert relative to other experts, as well as the extent to which their assessment differs from those of other experts. These are then used to weight each expert's assessment relative to others in calculating final scores.

Corruption Perceptions Index Data

6 indicators (CPIScore, CPIScoreRank, CPIScoreStdErr, EconIntelRating, SnPGloInsightsRating, WE-FRating) are derived from the Corruption Perceptions Index. This index is maintained by Transparency International, and includes the overall CPI score - which is an average of corruption scores assigned by a group of major international organizations - as well as the original scores used to calculate that number. Our dataset includes the CPI score, as well a selection of the original scores used to calculate it. These variables are on a 0 to 100 scale, with 0 being least corrupt and 100 being more corrupt. For our dataset, we have recalculated the scores to be on a 0 to 1 scale.

ILOSTAT Data

3 indicators (TotInfLab, MaleInfLab, FemaleInfLab) are derived from ILOSTAT's data on informal employment. These variables are on a 0 to 1 scale, with 0 representing 0% of a given population employed in the informal economy, and 1 representing 100% of a given population employed in the informal economy. Employment in the informal economy constitutes employment in "Informal productive activities," which are defined as "all productive activities carried out by persons or economic units that are - in law or in practice - not covered by formal arrangements."

Citations

1. Coppedge, Michael, John Gerring, Carl Henrik Knutsen, Staffan I. Lindberg, Jan Teorell, David Altman, Fabio Angiolillo, Michael Bernhard, Agnes Cornell, M. Steven Fish, Linnea Fox, Lisa Gastaldi, Haakon Gjerlow, Adam Glynn, Ana Good God, Sandra Grahn, Allen Hicken, Katrin Kinzelbach, Joshua Krusell, Kyle L. Marquardt, Kelly McMann, Valeriya Mechkova, Juraj Medzihorsky, Natalia Natsika, Anja Neundorff, Pamela Paxton, Daniel Pemstein, Johannes von Roemer, Brigitte Seim, Rachel Sigman, Svend-Erik Skaaning, Jeffrey Staton, Aksel Sundstroem, Marcus Tannenberg, Eitan Tzelgov, Yi-ting Wang, Felix Wiebrecht, Tore Wig, Steven Wilson and Daniel Ziblatt. 2025. "V-Dem [Country-Year/Country-Date] Dataset v15" Varieties of Democracy (V-Dem) Project. <https://doi.org/10.23696/vdemds25>.
 2. Transparency International. 2025. Corruption Perceptions Index 2024. <https://www.transparency.org/en/cpi/2024>.
 3. International Labour Organization. 2024. Informal Employment by Sex and Public/Private Sector – ILOSTAT. <https://ilostat.ilo.org/topics/informality/>.
-

Variables

Country Name

Variable Name: CountryName

Variable Type: character

Description: Name of country

Year (Excluding Informal Labor Data)

Variable Name: YearExclInfLab

Variable Type: numeric

Description: The year the data was gathered for all variables except those relating to informal labor. All of the non-informal labor data was gathered in 2024.

Polyarchy

Variable Name: Polyarchy

Variable Type: numeric

Description: V-Dem describes this variable as rating “To what extent is the ideal of electoral democracy in its fullest sense achieved?” as evaluated by country experts in the field. Electoral democracy is defined as making rulers accountable to their citizens via electoral procedures. On a scale of 0 to 1.

Table 1: Summary Statistics for Polyarchy

Statistic	Value
Min.	0.015
1st Qu.	0.245
Median	0.496
Mean	0.486
3rd Qu.	0.729
Max.	0.916

Liberal Democracy

Variable Name: LibDem

Variable Type: numeric

Description: V-Dem describes this variable as rating “To what extent is the ideal of liberal democracy achieved?” as evaluated by country experts in the field. Liberal democracy is defined broadly as the protection of minority and individual rights against the tyranny of the majority. On a scale of 0 to 1.

Table 2: Summary Statistics for Liberal Democracy

Statistic	Value
Min.	0.009
1st Qu.	0.125
Median	0.343
Mean	0.375
3rd Qu.	0.617
Max.	0.883

Participatory Democracy

Variable Name: ParticipDem

Variable Type: numeric

Description: V-Dem describes this variable as rating “To what extent is the ideal of participatory democracy achieved?” as evaluated by country experts in the field. Participatory democracy consists of the active involvement of citizens in all aspects of the political process, electoral and non-electoral. On a scale of 0 to 1.

Table 3: Summary Statistics for Participatory Democracy

Statistic	Value
Min.	0.010
1st Qu.	0.130

Statistic	Value
Median	0.293
Mean	0.311
3rd Qu.	0.473
Max.	0.794

Egalitarian Democracy

Variable Name: EgalDem

Variable Type: numeric

Description: V-Dem describes this variable as rating “To what extent is the ideal of egalitarian democracy achieved?” as evaluated by country experts in the field. Egalitarian democracy consists of equal opportunity to exercise political rights and liberties across all social groups. On a scale of 0 to 1.

Table 4: Summary Statistics for Egalitarian Democracy

Statistic	Value
Min.	0.025
1st Qu.	0.178
Median	0.326
Mean	0.366
3rd Qu.	0.571
Max.	0.878

Free and Fair Elections

Variable Name: FreeFairElect

Variable Type: numeric

Description: V-Dem describes this variable as rating what extent the most recent election in a country can be considered free and fair as evaluated by country experts in the field. Free and fair elections are defined as those that reflect the will of the people with minimal interference beyond routine errors. On a scale of 0 to 1.

Table 5: Summary Statistics for Free and Fair Elections

Statistic	Value
Min.	0.000
1st Qu.	0.203
Median	0.542
Mean	0.521
3rd Qu.	0.855
Max.	0.966

Equal Access

Variable Name: EqualAccess

Variable Type: numeric

Description: V-Dem describes this variable as answering the question “How equal is access to power?” as evaluated by country experts in the field. On a scale of 0 to 1.

Table 6: Summary Statistics for Equal Access

Statistic	Value
Min.	0.054
1st Qu.	0.452
Median	0.647
Mean	0.609
3rd Qu.	0.811
Max.	0.970

CPIScore

Variable Name: CPIScore

Variable Type: numeric

Description: The CPI score is calculated by taking the average of a group of corruption index scores by major international publications and organizations. It is originally calculated on a 1 to 100 scale, but we have recalculated it on a 0 to 1 scale. 0 represents least corrupt, and 100 represents most corrupt. For more information on the types of indices used to calculate the CPI score, see the other corruption rating variables (EconIntelRating, etc.)

Table 7: Summary Statistics for Corruption Perceptions Index Score

Statistic	Value
Min.	0.080
1st Qu.	0.270
Median	0.390
Mean	0.423
3rd Qu.	0.550
Max.	0.900
NA's	13.000

CPIRank

Variable Name: CPI Score Rank

Variable Type: numeric

Description: This variable is the rank of CPI scores relative to those of other countries. 1 represents the most corrupt country, 2 the second most corrupt, etc.

CPIStdErr

Variable Name: FreeFairElect

Variable Type: numeric

Description: This variable is the standard error of a country's given CPI score.

Economist Intelligence Unit Rating

Variable Name: EconIntelRating

Variable Type: numeric

Description: The Economist Intelligence Unit CPI rating is a corruption index put together by the Economist Intelligence Unit - a group run by the newspaper The Economist - which ranks countries based on the existence and virulence of various corrupt public sector practices, such as bribery, nepotism, and state capture.

Table 8: Summary Statistics for Economist Intelligence Unit Ratings

Statistic	Value
Min.	0.180
1st Qu.	0.350
Median	0.350
Mean	0.451
3rd Qu.	0.670
Max.	0.830
NA's	54.000

S & P Gloabl Insights Rating

Variable Name: SnPGloInsightsRating

Variable Type: numeric

Description: The S & P Global Insights CPI rating is a corruption index put together by S & P Global Insights which ranks countries based on the existence and virulence of various corrupt public sector practices, such as bribery, nepotism, and state capture.

Table 9: Summary Statistics for S & P Global Insights Ratings

Statistic	Value
Min.	0.060
1st Qu.	0.320
Median	0.320
Mean	0.421
3rd Qu.	0.590
Max.	0.850
NA's	15.000

World Economic Forum Rating

Variable Name: WEFRating

Variable Type: numeric

Description: The World Economic Forum CPI rating is a corruption index put together by the World Economic Forum which ranks countries based on the existence and virulence of various corrupt public sector practices, such as bribery, nepotism, and state capture.

Table 10: Summary Statistics for WEF Rating

Statistic	Value
Min.	0.060
1st Qu.	0.290
Median	0.445

Statistic	Value
Mean	0.465
3rd Qu.	0.613
Max.	1.000
NA's	67.000

Informal Labor Year

Variable Name: InfLabYear

Variable Type: numeric

Description: This variable represents what year the data on informal unemployment was gathered in.

Table 11: Summary Statistics for Informal Labor Year

Statistic	Value
Min.	2012.000
1st Qu.	2018.750
Median	2021.500
Mean	2020.393
3rd Qu.	2023.000
Max.	2024.000
NA's	95.000

Total Informal Employment Rate

Variable Name: TotInfLab

Variable Type: numeric

Description: This variable represents a country's total informal employment rate across all sexes, ages, and sectors. The data was gathered by each country's statistical agency and reported to the International Labor Organization.

Table 12: Summary Statistics for Total Informal Unemployment Rate

Statistic	Value
Min.	11.089
1st Qu.	54.352
Median	74.269
Mean	69.073
3rd Qu.	88.258
Max.	98.490
NA's	95.000

Male Informal Employment Rate

Variable Name: MaleInfLab

Variable Type: numeric

Description: This variable represents a country's informal employment rate among men across all ages and

sectors. The data was gathered by each country's statistical agency and reported to the International Labor Organization.

Table 13: Summary Statistics for Male Informal Labor Employment Rates

Statistic	Value
Min.	13.299
1st Qu.	55.904
Median	75.168
Mean	68.590
3rd Qu.	85.422
Max.	97.947
NA's	95.000

Female Informal Employment Rate

Variable Name: FemaleInfLab

Variable Type: numeric

Description: This variable represents a country's informal employment rate among women across all ages and sectors. The data was gathered by each country's statistical agency and reported to the International Labor Organization.

Table 14: Summary Statistics for Female Informal Labor Employment Rates

Statistic	Value
Min.	8.112
1st Qu.	50.794
Median	74.076
Mean	69.365
3rd Qu.	91.703
Max.	99.054
NA's	95.000