Historical GDP/GDPPC/Population

☐

Contents

Summary

The historical GDP, population, and GDPPC data is pulled from the article **"New Estimates of Over 500 Years of Historic GDP and Population Data"** by Fariss et al. (2022). The article proposed a dynamic latent variable model to address three major issues in historical GDP, population, and GDP per capita data. Namely 1) missing data, 2) measurement uncertainty, 3) systematic bias across sources.

The authors built a dynamic latent variable model that combined multiple historical and contemporary datasets. The work was supported by the Security and Political Economy (SPEC) Lab at the University of Southern California.

For population and GDP per capita, the latent trait model parameters were constructed based on following formula:

Parameter	Prior
Country i latent population estimate in first year t	$\theta_{pop[it=1]} \sim \mathcal{N}(0, 1)$
Country i latent population estimate in all other years	$\theta_{pop[it]} \sim \mathcal{N}(\theta_{pop}[t-1], \sigma_{pop})$
Latent population standard deviation	$\sigma_{pop} \sim \mathcal{HN}(0, 1)$
Country i latent GDPPC estimate in first year t	$\theta_{gdppc[it=1]} \sim \mathcal{N}(0, 1)$
Country i latent GDPPC estimate in all other years	$\theta_{gdppc[it]} \sim \mathcal{N}(\theta_{gdppc}[t-1], \sigma_{gdppc})$
Latent GDP standard deviation	$\sigma_{gdppc} \sim \mathcal{HN}(0, 1)$
Country i latent GDP estimate in all years	$\theta_{gdp[it]} \sim \mathcal{N}(\theta_{gdppc[it]} + \theta_{pop[it]}, \sigma_{gdp})$
Latent GDP standard deviation	$\sigma_{gdp} \leftarrow \sqrt{\sigma_{pop}^2 + \sigma_{gdppc}^2}$
Model j intercept "difficulty parameter"	$\alpha_j \sim \mathcal{N}(\bar{y}_{it,j}, 4)$
Item-type cross-sectional conversion parameter	$\beta_{k[i]} \sim \mathcal{HN}(1, 0.5)$
Item-type cross-sectional random effect (ppp)	$\gamma_{k[i]} \sim \mathcal{N}(0, 1)$
Item-type cross-sectional random effect (exchange rate)	$\gamma_{k[i]}^* \sim \mathcal{N}(\gamma_{k[i]}, \sigma_\gamma)$
Uncertainty for exchange rate cross-sectional random effect	$\sigma_\gamma \sim \mathcal{HN}(0, 1)$
Temporal random effect	$\lambda_t \sim \mathcal{N}(0, 1)$
Model uncertainty for all population items Model	$\tau_{pop} \sim \mathcal{HN}(0, 1)$
uncertainty for all GDP per capita items Model	$\tau_{gdppc} \sim \mathcal{HN}(0, 1)$
uncertainty for all GDP items	$\tau_{gdp} \sim \mathcal{HN}(0, 1)$

(Farris et al., 2022)

Tables in IFs

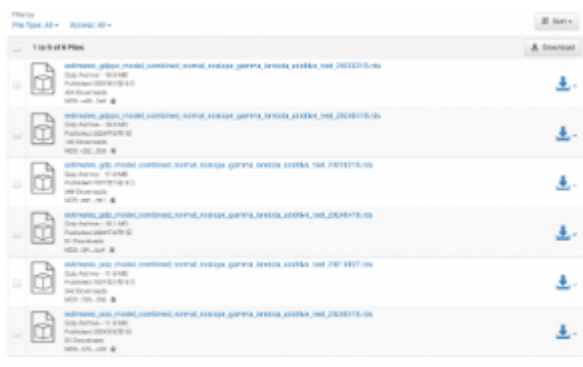| Variable	Table	Definition	Last IFs Update	UsedInPreprocessor
HistoricalGDP	SeriesHistoricalGDP	Mean of historical GDP values extended with a Latent Variable Modeling Framework.	2025/08/21	0
HistoricalGDPPC	SeriesHistoricalGDPPC	Mean of historical GDP per capita values extended with a Latent Variable Modeling Framework.	2025/08/21	0
HistoricalPOP	SeriesHistoricalPOP	Mean of historical population values extended with a Latent Variable Modeling Framework.	2025/08/21	0

Data Pulling Instructions

Step 1:

To pull data from this paper, go to:

https://journals.sagepub.com/doi/full/10.1177/00220027211054432

In the Acknowledgements section, there is a link to the predicted data:

https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/FALCGS

Step 2:

Download the most updated datafiles (in this example, in 2024).

Data Notes

1.   The paper was published in 2022. In 2024, the author shared an updated version of the datafile.

2.   The data files are in .rds format, if using R, open the file and it will automatically import; if using python, use the *read_r* function in the *pyreadr* package.

3.   The datafile included predicted values as well as datasets used in the modeling, slice "latent_gdp", "latent_gdppc", "latent_pop" from the indicator column of corresponding datafiles for the predicted values.

4.   GWNO codes and corresponding country names were provided in the supplementary files of the original paper. However, the supplementary only provided corresponding codes

for 217 countries/regions and the dataset had 225 countries/regions (missing: 89,99,327, 396,397, 563,564,711). For the additional gwno codes, use this link for corresponding country names: http://ksgleditsch.com/data-4.html.

5. For Yeman, there are *Yemen (Arab Republic of Yemen)* and *Yemen, People's Republic of;* and for Viet Nam, there are *Vietnam (Annam/Cochin China/Tonkin)* and *Vietnam, Democratic Republic of* in the original dataset:

5.1 for the years that only one of them had value, we took the value from whichever that is available for population, GDP, and GDPPC

5.2 for the years that both of them had value, we summed the values for population and GDP; and the values for GDPPC are then calculated from population and GDP