

## Stephen Haber, “The Rise and Fall of the Resource Curse,” Data Codebook

We follow Stephen Haber and Victor Menaldo, 2011, “Do Natural Resources Fuel Authoritarianism: A Reappraisal of the Resource Curse,” The American Political Science Review 105 (hereafter Haber and Menaldo 2011) by estimating Total Oil Income Per Capita for every country in the world with a Polity2 score from the Polity IV project. We depart from that dataset in three ways. First, the Haber and Menaldo (2011) dataset covers 1800 to 2006. We extend coverage to 2021. Second, we provide estimates for former Soviet states during the period when they were part of pre-1922 Russia and the USSR. Third, we provide estimates of former Yugoslav states during the period when they were part of Yugoslavia. Readers who may wish to re-aggregate these into estimates for the pre-1922 Russia, the USSR, or Yugoslavia simply need to add the data for the constituent states together.

To estimate Total Oil Income Per Capita three values must be known: the volume of crude oil production in each country-year; the real value of crude oil in each year; and the population in each country-year. The formula is simply Crude Oil Production times the Price of Crude Oil in real dollars, divided by Population.

### **Estimating the Volume of Crude Oil Output:**

We follow the general coding rules in Haber and Menaldo (2011) by minimizing the number of sources used to construct country time series to mitigate potential measurement error caused by source changes. We define crude oil output as the quantity of crude oil, including lease condensate. Lease condensates are liquid hydrocarbons that are in gaseous form when under pressure but liquify at the normal temperatures found above ground. They are usually removed at the wellhead, and then enter the crude oil stream. They are not the same as natural gas plant liquids, which have to be captured at specialized natural gas plants through freezing or pressurizing.

We give primacy to three sources that provide both long time series coverage and broad country coverage. For the period covering 1973 to 2021, we give primacy to the US Energy Information Agency (hereafter USEIA), <https://www.eia.gov/international/data/world> For some producers USEIA provides estimates beginning in 1980, rather than 1973. For the period covering 1857 (when commercial production of crude oil on an industrial scale began) to 1966, we give primacy to the American Petroleum Institute, Petroleum Facts and Figures (1971)—hereafter API. Because the footnotes to API indicate that it drew its data from the primary industry trade publication, the Oil and Gas Journal, well as from the U.S. Bureau of Mines, we give primacy to the Oil and Gas Journal Database and to hard copies of the Oil and Gas Journal (hereafter OGJ) for years not covered by USEIA or API. We verify that the API and OGJ data series are in line with one another by comparing overlapping country-years from both sources. We also verify that the USEIA and OGJ series are in line with one another by comparing overlapping country-years from both sources. In cases in which a source provides data in daily units we multiply by 365. If a source provides data in monthly units we multiply by 12.

To produce our OGJ series we rely on three OGJ sources: hardcopies of the OGJ; the Oil and Gas Journal Database (hereafter, OGJD); and the Oil and Gas Journal Databook (hereafter OGJB). We purchased OGJD from Pennwell's, which owns the database, covering the years 1990 to 2006. We retrieved the OGJD for 1970 to 1991 from the Stanford Earth Sciences Library, which has it on floppy disks. When the OGJD provides monthly output data, we sum the values for each month to obtain yearly estimates. We retrieved the OGJD from the Stanford Earth Sciences Library. We retrieved the hardcopies of the OGJ from the Stanford Earth Sciences Library. Typically, the OGJ published its annual country estimates of crude petroleum output for the prior two years in its February 15 issue, although there is some variance in the exact date of publication from year to year.

There are a few country-years (particularly former Soviet Bloc countries in the late 1950s and 1960s) that are missing in both OGJ and API. We therefore turn to another petroleum industry trade journal, World Oil, which produces estimates for those country-years.

We verify that our reliance on API and OGJ for years not covered by USEIA does not miss the pre-1980 data for minor producers by employing pre-1982 volumes of the following publications: World Oil (various issues); International Energy Agency, Oil Information (various issues); and International Energy Agency, Energy Statistics of Non-OECD Countries (various issues). In those few cases in which we find that there is production data for country-years not covered by API or OGJ we draw on these alternative sources.

To estimate crude oil output for the major producers of the former Soviet Union and pre-1922 Russia we draw on the sources enumerated above to first estimate total pre-1917 Russia or USSR production. For the period 1985 to 1991 we then apportion total USSR production among Russia and former Soviet states using output data from BP Statistical Review of World Energy. For years prior to 1985 (or when production was so minor that it is not covered in BP Statistical Review of World Energy), we apportion total USSR or Russian production using country-specific studies. Specifically:

- For Azerbaijan, we rely on Rebecca Hastings, "Oil Capital: Industry and Society in Baku, Azerbaijan, 1870-Present" (Ph.D. Diss, U of Oregon 2020); and Sabit Bagirov, "Azerbaijani Oil: Glimpses of a Long History." For country-years when these sources do not provide data, we estimate the Azerbaijan share of total Russian or USSR output via interpolation.
- For Turkmenistan, we rely on Gregory Ulmishek, "Petroleum Geology and Resources of the Amu-Darya Basin, Turkmenistan, Uzbekistan, Afghanistan, and Iran," US Geological Survey Bulletin 2201-H (2004) and Wikipedia. For country-years when these sources do not provide data, we estimate the Turkmenistan share of total Russian or USSR output via interpolation.
- For Uzbekistan, we rely on Meyileva Kibroyo Oybekovna et. al., 2022, "History of Oil and Gas Industry in Uzbekistan and the Role of Kashkadarya Oil in it," Galaxy International Research Journal 10: 211-213.
- For Belarus, we rely on information in Britannica.com. For country-years when these sources do not provide data, we estimate the Azerbaijan share of total Russian or USSR output via interpolation.

- For Kazakhstan, we rely on Katherine Hardin, 2012, “Kazakhstan’s Energy Sector Since Independence,” Atlantic Council Issue Brief. For country-years when these sources do not provide data, we estimate the Azerbaijan share of total Russian or USSR output via interpolation.

### **Estimating the Price of Crude Petroleum:**

We employ the price series constructed by British Petroleum, [Statistical Review of World Energy](#). We draw on the 2022 edition, which gives crude oil prices in 2021 dollars back to 1861. Our series on Total Oil Output Per Capita thus begins in 1861, even though we have production observations for Romania beginning in 1857 and the United States beginning in 1859.

### **Estimating Populations:**

To mitigate measurement error that might come from source changes we follow the same protocol as for crude oil output by minimizing the number of sources in estimating population series. For the period since 1960, we draw the data from World Bank, World Bank Development Indicators (online edition—hereafter WBDI). For the period 1950 to 1959 (or if WBDI does not provide an estimate) we draw the data from United Nation World Population Prospects (online edition—hereafter UNWPP). For years prior to 1950, we draw on either of two sources: Angus Maddison, [The World Economy: Historical Statistics](#) (Paris, OECD, 2003); or Arthur Banks Cross National Time Series Data Archive website—and we choose among those two sources depending on which provides the longer run of data for that particular country. We note that this leaves some country-years as uncoded, particularly in the late 19th and early 20th centuries. We therefore code those country years from B.R Mitchell, [International Historical Statistics 1750-2003](#), fifth edition (2003). For some country-years we rely on country specific sources, but note that this occurs rarely. When data is missing from these sources we provide an estimate via interpolation. We note that we only provide population estimates if we also have data on crude oil output.

### **Polity2**

We employ the Polity 2 measure created by Monty G. Marshall and Keith Jagers, “Polity IV Project: Political Regime Characteristics and Transitions.” This is Marshall and Jagers combined measure, scaled from -10 (most autocratic) to 10 (most democratic), with politically unstable years interpolated, coded from 1800 to 2017. We normalize their index to run from 0 to 100 by adding 10 and multiplying by five. This transformation makes regression coefficients easier to interpret.

There are cases in which Polity IV treats countries that were protectorates as non-sovereign, and thus does not assign them a Polity2 score. Many of those countries did, however, have continuous regimes that governed the internal affairs of the state. We therefore back-code Polity2 scores to facilitate time series analyses when the historical record indicates that the government had effective control over internal affairs and there was no change in the form of government pre- and post-independence. Kuwait is a case in point: It has been ruled by a single dynasty, the Al Sabah, since 1756. The British established a protectorate over Kuwait in 1899. During World

War I, the British declared Kuwait an independent principality under the protection of the British Empire. In 1962, Kuwait gained formal independence from Britain—but importantly the form of government did not change. We therefore back-code Kuwait to 1800 (the first year of coverage in Polity IV). Specifically:

- We back-code Kuwait from 1962 to 1800.
- We back-code Jordan from 1945 to 1921 because Jordan was founded in 1921 was recognized by the League of Nations as a state under the British mandate in 1922. Jordan remained under British supervision until 1946. The same dynasty ruled Jordan after independence as during the period of the British mandate.
- We back-code Saudi Arabia from 1926 to 1902 because Ottoman rule was by that point nominal and the Al Saud (the post-independence ruling dynasty) had already captured Riyadh and gained control of the country.
- We back-code Bahrain from 1970 to 1820 because as of 1820 Bahrain was a principality run by the Al Khalifa family.
- We back-code Qatar from 1970 to 1878 because as of 1878 it was a principality ruled by the Al Thani family.
- We back-code the United Arab Emirates from 1970 to 1800 because the emirates had been ruled by tribal emirs since the 17<sup>th</sup> century. Since 1970, the emirs pick one of their number to serve as king of the UAE.

### **Harmonizing Country Definitions**

We create a uniform system of country definitions following the conventions in Haber and Menaldo (2011). We code countries following the definitions in the Correlates of War dataset (<http://www.correlatesofwar.org/>), except for the following modifications, which we make to facilitate time series coverage.

When a country is split in two and later reunified, or when it changes name, or when one country absorbs another and takes on a new name, we code that country as having a single country code across time. We assign the country code of the dominant partner. This facilitates time series analyses. We follow similar rules when countries fragment. Specifically:

- We code Czechoslovakia and the Czech Republic as a single country.
- We code pre-1993 and post 1993 Ethiopia as a single country.
- We code Prussia, the German Empire, the Republic of Germany, and the German Federal Republic as a single country. This means that East Germany, as well as the 19<sup>th</sup> century German kingdoms, are not observed in the dataset.
- We code pre-1922 Russia, the Soviet Union, and Post-1991 Russia as a single country.
- We code Vietnam and North Vietnam as a single country. This means that South Vietnam is not observed in the dataset.
- We code Serbia and Yugoslavia as a single country because Serbia was the dominant state in the Yugoslav confederation.
- We code North Yemen and Yemen as a single country. This means that South Yemen is not observed in the dataset.
- We code Korea and South Korea as a single country.
- We code pre-1947 and post-1947 Pakistan as a single country.

- We code Gran Colombia and Colombia as a single country.
- We code Sardinia and Italy as a single country.