Data Description for *Taxing the Rich*

Annual Dataset
File Name: Scheve_Stasavage_TaxingtheRich_2016_annual.dta

Description: This is the 1800-2013 annual dataset in Stata format needed to replicate the main analyses included in the book and in the online appendix.

Country (country) is the observation country name in English. This dataset includes the following countries: Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Ireland, Italy, Japan, the Netherlands, New Zealand, Norway, South Korea, Spain, Sweden, Switzerland, the United Kingdom, and the United States.

COW Code (ccode) is the unique country identifier for the observation country from the Correlates of War project. Please note the following coding decisions, selected to maintain consistency throughout the series:

> Austria – We use the COW Code for modern Austria (305) throughout the series and do not use the distinct COW Code for Austria-Hungary (300) prior to 1919.
>
> Germany – We use the COW Code for unified Germany (255) throughout the series and do not switch to the distinct COW Code for the German Federal Republic (260) throughout the Cold War.

Year (year) is the observation year. The dataset includes all years from 1800-2010 for which a given country was independent.

Half Decade (hdecadec) is a count variable that identifies five-year periods throughout the data. The variable has a minimum value of 1, corresponding to the period from 1816 to 1820, and a maximum value of 39, corresponding to the period from 2006 to 2010.

Independence (independence) is a binary variable that is set equal to 1 in years in which the country was an independent political entity and 0 otherwise. The principle source for this variable is Caramani (2000). *Independence* is coded 1 for the following country-years: Australia: 1901-2010; Austria: 1800-2010; Belgium: 1830-2010; Canada: 1867-2010; Denmark: 1800-2010; Finland: 1917-2010; France: 1800-2010; Germany: 1871-2010; Ireland: 1922-2010; Italy: 1861-2010; Japan: 1800-2010; the Netherlands: 1800-2010; New Zealand: 1856-2010; Norway: 1905-2010; South Korea: 1945-2010; Spain: 1800-2010; Sweden: 1800-2010; Switzerland: 1800-2010; the United Kingdom: 1800-2010; the United States: 1800-2010.

Occupied (occupied) is a binary variable that is set equal to 1 in years in which the country was under foreign occupation and 0 otherwise. The principle source for this variable is Caramani (2000). *Occupied* is coded 1 for the following country-years: Austria: 1939-1954; Belgium: 1941-1943; Denmark: 1941-1944; France: 1941-1943; Germany: 1946-1948; Japan: 1946-1951; the Netherlands: 1941-1944; Norway: 1941-1944.

Top Income Tax Rate (top_incrate_n) is the top marginal income tax rate levied by the national government on individuals in the highest income category. The sample covers all 20 of the countries listed above and is limited to states that have a modern income tax system, defined as an independent national government levying taxes yearly on comprehensive and directly assessed forms of personal income. The sources for the data are described in greater detail in our *Comparative Income Tax Database* codebook, excel dataset, and scanned copies of sources archived at http://data.stanford.edu/citd at Stanford University (Genovese, Scheve, and Stasavage 2014).

Top Income Tax Rate – Local & National (top_incrate_nl) is the combined top marginal income tax rate levied by national and subnational governments on individuals in the highest income category. The sample covers all 20 of the countries listed above and is limited to states that have a modern income tax system, defined as an independent national government levying taxes yearly on comprehensive and directly assessed forms of personal income. For local data, we assume that the taxpayer lived in the largest city in the country. The sources for the data are described in greater detail in our *Comparative Income Tax Database* codebook, excel dataset, and scanned copies of sources archived at http://data.stanford.edu/citd at Stanford University (Genovese, Scheve, and Stasavage 2014).

Top Inheritance Tax Rate (topitaxrate2) is the top marginal rate of inheritance taxation for a single direct descendant based on a cash inheritance. The sample covers all of the countries listed above except for Spain. The sources for the data are described in detail in online Supplementary Materials of our *APSR* paper (Scheve and Stasavage 2012). Further information about our *Comparative Inheritance Taxation Database* is available in the codebook, excel dataset, and scanned copies of sources archived at Institution for Social and Policy Studies at Yale University (Plagge, Scheve, and Stasavage 2011).

Top Effective Income Tax Rate (effective_p99_99) is the effective rate of income tax paid (after credits and bonuses) by individuals in the top 0.01% of the national income distribution. (Top 0.05% for the Netherlands and the United Kingdom.) We have constructed this variable for six countries: Canada, France, the Netherlands, Sweden, the United Kingdom, and the United States. The effective rates for France, the Netherlands, the United Kingdom, and the United States are for national taxes only. The effective rates for Canada and Sweden include local and national taxes. These data are based on income tax returns and are from the *World Top Incomes Project*, which is the work of a wide number of scholars led by Tony Atkinson, Thomas Piketty, and Emmanuel Saez. The principal source for each country comes from contributions to Atkinson and Piketty (2007, 2010). For further details about the sources for each country, see footnote 9 to chapter 3 in Scheve and Stasavage (2016).

Top Effective Income Tax Rate – IRS (irs_soi_top1) is an alternative measure of the effective rate of income tax paid by individuals in the top 1% of the national income distribution in the United States from 1916 to 1995. The data are from each of the annual Internal Revenue Statistics of Income Reports (IRS various years).

War Mobilization (himobpopyear2p) is a binary variable set equal to 1 if a country was engaged in an interstate war and at least 2 percent of the population was serving in the military and 0

otherwise. Our primary data for incidents of interstate war comes from the Correlates of War Project, Militarized Interstate Disputes Data, Version 3.0 (Ghosn, Palmer, and Bremer 2004). We define an interstate war as an interstate war (MID with hostility coded as level 5) in which there are 1,000 or more deaths (fatalities coded as a 6). Therefore, if the conflict involved 1,000 or more deaths, we code there being an interstate war for that country during the entire duration of the conflict. Our primary data on mobilization comes from the Correlates of War Project, National Material Capabilities Data, Version 3.0 (Singer 1987). We used this data to calculate the ratio of military personnel to total population and if this ratio was greater than 0.02 in a given year—and the country was at war as defined above in that year—we coded *War Mobilization* a 1, otherwise it was coded 0.

War Mobilization – 5% (himobpopyearp) is a binary variable set equal to 1 if a country was engaged in an interstate war and at least 5 percent of the population was serving in the military and 0 otherwise. Our primary data for incidents of interstate war comes from the Correlates of War Project, Militarized Interstate Disputes Data, Version 3.0 (Ghosn, Palmer, and Bremer 2004). We define an interstate war as an interstate war (MID with hostility coded as level 5) in which there are 1,000 or more deaths (fatalities coded as a 6). Therefore, if the conflict involved 1,000 or more deaths, we code there being an interstate war for that country during the entire duration of the conflict. Our primary data on mobilization comes from the Correlates of War Project, National Material Capabilities Data, Version 3.0 (Singer 1987). We used this data to calculate the ratio of military personnel to total population and if this ratio was greater than 0.05 in a given year—and the country was at war as defined above in that year—we coded *War Mobilization – 5%* a 1, otherwise it was coded 0. See also the description of *War Mobilization*, above.

In what follows, we discuss missing data and coding issues for each case for the *War Mobilization* measure and analogous measure using the 5 percent cutoff.

Australia – Mobilization data missing from independence (1901) to 1919 and for the year 1942. We filled in this information using additional sources (Clodfelter 2008; Ellis and Cox 1993; Gray and Argyle 1991). As a result, Australia is coded 1 for 1915-1918, 1941-1945, and 0 for every other year in its sample period of 1901-2010. This coding is also the same when the 5 percent threshold is applied.

Austria – Mobilization data is missing before 1919 because Austria at that time was a part of the Austro-Hungarian Empire and is missing from 1939 to 1954. We filled in this information using various sources (Clodfelter 2008). From 1816 to 1914, we used the Austro-Hungarian mobilization data in COW data. For 1915-1918, we also coded Austria based on Austro-Hungarian Empire but corrected the COW data. Specifically, the Empire was mobilized above both the 2 and 5 percent thresholds for 1915-1918. Austria was coded the same as Germany for 1939 to 1945 and was coded 0 for 1946 to 1954. Overall, Austria is coded a 1 for 1915-1918, 1939-1945, and 0 for every other year for its sample period of 1816-2010. The years 1939 and 1940 are switched to zero when the 5 percent threshold is applied.

Belgium – Mobilization data is missing during the four years during WWII that Belgium was occupied. We filled in this information using various sources (Clodfelter 2008). We have coded the mobilization measures 0 for the missing years of 1941 to 1944 as the country was never mobilized above either threshold during WWII. As a result, Belgium is coded a 1 for 1915-1918 and 0 for the remainder of its sample period of 1830-2010. Belgium is coded 0 for all years when the 5 percent threshold is applied.

Canada – Mobilization data is missing from independence (1867) to 1919. We filled in this information using additional sources (Clodfelter 2008; Ellis and Cox 1993; Gray and Argyle 1991). As a result Canada is coded 1 for 1915-1918, 1941-1945, and 0 for every other year in its sample period of 1867-2010. The years 1941 and 1942 are switched to zero when the 5 percent threshold is applied.

Denmark – Mobilization data is missing during the four years during WWII that Denmark was occupied. We filled in this information using various sources (Clodfelter 2008). We have coded the mobilization measures 0 for the missing years of 1941 to 1944 as the country was never mobilized above either threshold during WWII. As a result, Denmark is coded 0 for all years of its sample period of 1816-2010. This coding is also the same when the 5 percent threshold is applied.

Finland – Mobilization data is missing for 1917-1918 and 1941-1942. We filled in this information using additional sources (Clodfelter 2008; Ellis and Cox 1993; Gray and Argyle 1991). As a result, Finland is coded 1 for 1940-1944 and 0 for every other year in its sample period of 1917-2010. This coding is also the same when the 5 percent threshold is applied.

France – Mobilization data is missing for France 1942 and 1943 during WWII. All forces in France were demobilized during occupation although there was significant mobilization outside of France (Clodfelter 2008). This mobilization, however, did not reach our thresholds and so the mobilization measure is set at 0 for both thresholds for these two years. As a result, France is coded 1 for 1871, 1914-1920, 1940-1941, and 0 for every other year in its sample period of 1816-2010. The years 1871, 1914, 1920, and 1941 are switched to zero when the 5 percent threshold is applied.

Germany – Mobilization data for West Germany is used for 1946-1989. As a result Germany is coded 1 for 1871, 1915-1918, 1939-1945, and 0 for every other year in its sample period of 1871-2010. The years 1871, 1939, and 1940 are switched to zero when the 5 percent threshold is applied.

Ireland – Mobilization data for Ireland are complete from independence in 1922. Ireland is coded 0 for its entire sample period of 1922-2010 for both the 2% and 5% thresholds.

Italy – Mobilization data are missing for Italy in 1941. We filled in this information using additional sources (Clodfelter 2008). As a result, Italy is coded 1 for 1915-1918, 1935, 1940-1943, and 0 for all other years in its sample period of 1861-2010. The years 1935, 1940, and 1943 are switched to zero when the 5 percent threshold is applied.

Japan – Mobilization data are missing for Japan before 1860 and for 1946-1951 when it was occupied. We filled in this information using additional sources (Clodfelter 2008). As a result Japan is coded 1 for 1941-1945 and 0 for all other years in its sample period of 1816-2010. The years 1941-1943 are switched to zero when the 5 percent threshold is applied.

Netherlands – Mobilization data is missing during the four years during WWII that the Netherlands was occupied. We filled in this information using various sources (Clodfelter 2008). We have coded the mobilization measures 0 for the missing years of 1941 to 1944 as the country was never mobilized above either threshold during WWII. As a result, the Netherlands is coded 0 for all years of its sample period of 1816-2010. This coding is also the same when the 5 percent threshold is applied.

New Zealand – Mobilization data is missing for New Zealand in the COW data from independence (1856) to 1921 and for 1943, 1945-49. We filled in this information using additional sources (Clodfelter 2008; Ellis and Cox 1993; Gray and Argyle 1991). As a result New Zealand is coded 1 for 1915-1918, 1941-1945, and 0 for all other years in its sample period of 1856-2010. The years 1915-1918 are switched to zero when the 5 percent threshold is applied.

Norway – Mobilization data is missing during the four years during WWII that Norway was occupied. We filled in this information using various sources (Clodfelter 2008). We have coded the mobilization measures 0 for the missing years of 1941 to 1944 as the country was never mobilized above either threshold during WWII. As a result, Norway is coded 0 for all years of its sample period of 1905-2010. This coding is also the same when the 5 percent threshold is applied.

South Korea – Mobilization data for South Korea starts in 1949 in the COW data. We filled in the missing data from the end of Japanese occupation (1945) through 1948 using additional sources (Clodfelter 2008). As a result South Korea is coded 1 for 1953, 1965, 1967-68, 1970 and 0 for all other years in its sample period of 1945-2010. All years are coded 0 when the 5 percent threshold is applied.

Spain – Mobilization data are available for all years for Spain. Spain is coded 0 for all years of its sample period of 1816-2010. This coding is also the same when the 5 percent threshold is applied.

Sweden – Mobilization data for Sweden are missing for 1816-1818, 1820-1822, 1824-1827, 1829, 1832, 1834-1839, 1841-43, 1845, 1847, 1854, 1874, 1914, 1919, 1934, 1939, 1941, and 1946-1956. We were, nonetheless, able to code *War Mobilization* based on the absence of high fatality conflicts. As a result, Sweden is coded 0 for all years of its sample period of 1816-2010. This coding is also the same when the 5 percent threshold is applied.

Switzerland – Mobilization data is available for all years for Switzerland. Switzerland is coded 0 for all years of its sample period of 1816-2010. This coding is also the same when the 5 percent threshold is applied.

United Kingdom – Mobilization data is missing for the UK in 1916. We filled in this information using Braun and McGrattan (1993). As a result, the UK is coded 1 for 1915-1918, 1940-1945, and 0 for all other years in its sample period of 1816-2010. The years 1940-41 are switched to zero when the 5 percent threshold is applied.

United States – Mobilization data is available for all years for the United States. The United States is coded 1 for 1918, 1942-1945, 1951-1953, and 0 for all other years in its sample period of 1816-2010. The years 1918, 1942, and 1951-1953 are switched to zero when the 5 percent threshold is applied.

WWI Mobilization (wwimassmob) is a binary variable set equal to 1 for all observation years if the country was independent and met the definition of *War Mobilization* (above) at any point during WWI and 0 otherwise. The following countries are coded 1: Australia, Austria, Belgium, Canada, France, Germany, Italy, New Zealand, the United Kingdom, and the United States. The following countries are coded 0: Denmark, Japan, Netherlands, Norway, Spain, Sweden, and Switzerland. The following countries were not independent for the duration of WWI and as a result are coded as missing: Finland, Ireland, and South Korea.

Universal Male Suffrage (unisuffrage) is a dummy variable that is set equal to 1 for years in which all adult males are eligible to vote in national elections and 0 otherwise. As is the case with unitary states, for federal states, such as Germany, our variable takes account only of suffrage laws established at the national level and applying to the national legislature, provided that such laws exist. In the case of the United States we date universal male suffrage as 1965 after the application of the Voting Rights Act. In cases where a country established universal suffrage before becoming fully independent from another power, we use the date of the state's independence to code this variable. The principle source for this variable is Caramani (2000), we consulted Mackie and Rose (1991) for countries not covered by the former source, and we supplemented these sources with Croissant (2002) for South Korea.

Universal Male Suffrage – Time Marker (timemarkeru) is a count variable that runs from -10 to 10 and indicates the temporal distance between the observation-year and the year in which the country introduced *Universal Male Suffrage*.

Competitive Elections (competitive_elections) is a binary variables that is set equal to one if the legislature is elected in free multi-party elections, if the executive is directly or indirectly elected in popular elections and is responsible either directly to voters or to a legislature elected according to the first condition, and finally if at least 50 percent of adult males have the right to vote. This definition and data is from Boix and Rosato (2001). The definition is a modification of the definition used by Przeworski et al. (2000) to a context where the suffrage may be restricted. *Competitive Elections* is coded one for the following years: Australia 1901-2010; Austria 1920-1932, 1946-2010; Belgium 1894-2010; Canada 1867-2010; Denmark 1901-2010; Finland 1917-2010; France 1848-1851, 1870-1939, 1945-2010; Germany 1919-1932, 1949-2010; Ireland

1922-2010; Italy 1946-2010; Japan 1952-2010; South Korea 1960, 1988-2010; Netherlands 1897-2010; New Zealand 1856-2010; Norway 1905-2010; Spain 1931-1936, 1977-2010; Sweden 1911-2010; Switzerland 1848-2010; United Kingdom 1885-2010; United States 1816-2010.

Competitive Elections – WWI (wwidemoc) is a binary variable set equal to 1 for all observation years in which the country was independent and met the definition of *Competitive Elections* (above) for the duration of WWI and 0 otherwise. The following countries are coded 1: Australia, Belgium, Canada, Denmark, France, the Netherlands, New Zealand, Norway, Sweden, Switzerland, the United Kingdom, and the United States. The following countries are coded 0: Austria, Germany, Italy, Japan, and Spain. The following countries were not independent for the duration of WWI and as a result are coded as missing: Finland, Ireland, and South Korea.

Direct Elections (directelec) is a binary variable that is set equal to 1 if a country has direct elections for the lower house and 0 otherwise. This variable was coded using Caramani (2000, 58) as the principal source and as otherwise noted below for the remaining countries. Australia 1901 (Mackie and Rose 1991, 1), Austria 1907, Belgium 1847, Canada 1867 (Mackie and Rose 1991, 65), Denmark 1849, Finland 1917, France 1831, Germany 1871, Ireland 1922 (Mackie and Rose 1991, 181), Italy 1861, Japan 1889 (Mackie and Rose 1991, 223), South Korea 1948 (Croissant 2002), Netherlands 1848, New Zealand 1857 (Mackie and Rose 1991, 289), Norway 1906, Spain 1837, Sweden 1911, Switzerland 1848, United Kingdom prior to 1800, United States prior to 1800.

Electorate 25% (electorate25) is a binary variable that is set equal to 1 if at least 25% of adult (with "adult" being determined as above voting age) males are eligible to vote and 0 otherwise. The source for this data is Flora (1983) for the European cases, the *Statistical History of the American Electorate* for the United States, *New Zealand: A Handbook of Historical Statistics* for New Zealand, Griffin (1965) for Japan, Croissant (2002) for South Korea, and Mackie and Rose (1991) for Australia. The dates for Canada are inferred from data on 1867 voter turnout.

Electorate 50% (electorate50) is a binary variable that is set equal to 1 if at least 50% of adult (with "adult" being determined as above voting age) males are eligible to vote and 0 otherwise. The source for this data is Flora (1983) for the European cases, the *Statistical History of the American Electorate* for the United States, *New Zealand: A Handbook of Historical Statistics* for New Zealand, Griffin (1965) for Japan, Croissant (2002) for South Korea, and Mackie and Rose (1991) for Australia. The dates for Canada are inferred from data on 1867 voter turnout.

Electorate 75% (electorate75) is a binary variable that is set equal to 1 if at least 75% of adult (with "adult" being determined as above voting age) males are eligible to vote and 0 otherwise. The source for this data is Flora (1983) for the European cases, the *Statistical History of the American Electorate* for the United States, *New Zealand: A Handbook of Historical Statistics* for New Zealand, Griffin (1965) for Japan, Croissant (2002) for South Korea, and Mackie and Rose (1991) for Australia. The dates for Canada are inferred from data on 1867 voter turnout.

No Upper (noupper) is a binary variable that is set equal to 1 for the absence of an upper house with veto power for which representatives are either not directly elected, elected by a restricted

constituency, appointed, or who sit by hereditary right and 0 otherwise. More formally, this variable takes a value of 1 if any of the follow three conditions are satisfied: (1) there is no upper house; (2) there is an upper house that cannot veto legislation; or (3) there is an upper house in which members are directly elected through universal male suffrage. Our coding for this variable is based primarily on Marriot ([1910] 1926) and on historical information contained on the websites of the respective upper chambers. Additional sources for specific countries are as noted below. The coding for this variable is as follows: Australia 1 for entire period, Austria 1 beginning in 1920, Belgium, 1 beginning in 1918, Canada 0 for all years, Denmark 1 from 1915, Finland 1 for all years, France 0 from 1815-1847 then 1 from 1848-1851 then 0 from 1852-1945 then 1 from 1946 onwards, Germany 0 for all years, Ireland 1 for all years, Italy 1 from 1948, Japan 1 from 1946, South Korea 1 for all years, Netherlands 0 for all years, New Zealand 1 for all years, Norway 1 for all years, Spain 1 from 1931, Sweden 1 from 1918, Switzerland 1 from 1848, United Kingdom 1 from 1911, United States 1 from 1913. Additional sources consulted: Canada: Senate of Canada, Committees and Private Legislation Directorate (2001); Denmark: Danish Parliament (2011); New Zealand: Christie (1924); Italy: Pasquino (2009).

Left Executive (left_executive) is a binary variable that is set equal to 1 in years in which a country's head of government (President in a presidential system and Prime Minister/Chancellor in a parliamentary system) is a member of a socialist, social democratic, or labor party. Switzerland is the exception to this coding. Because of its collegial executive, the Swiss measure is the proportion of the seven members of the Federal Council that are members of a socialist, social democratic, or labor party. Flora (1983) is the principle source for this variable. Canada, Ireland, South Korea, and the United States are not coded as having a chief executive that is a member of a socialist, social democratic, or labor party. Left Executive is coded as one (or, in the Swiss case, greater than zero) for the following observations: Australia 1904, 1908-1917, 1929-1932, 1941-1949, 1972-1975, 1983-1996, 2007-2013; Austria 1918-1920, 1945, 1970-2000, 2007-2013; Belgium 1938,1945-1949, 1954-1958, 1973-1974, 2010-2013; Denmark 1924-1926, 1929-1945, 1947-1950, 1953-1968, 1971-1973, 1975-1982, 1993-2001, 2011-2013; Finland 1926-1927, 1948-1949, 1956-1959, 1966-1970, 1972-1975, 1977-1987, 1995-2003; France 1936-1938, 1946-1947, 1955-1957, 1981-1986, 1988-1993, 1997-2002, 2012-2013; Germany 1918-1920, 1928-1930, 1969-1982, 1998-2005; Italy 1921-1922, 1944-1945, 1983-1987, 1992-1993, 1998-2001, 2006-2008, 2013; Japan 1947-1948, 1994-1996; Netherlands 1948-1958, 1973-1977, 1994-2002; New Zealand 1935-1949, 1957-1960, 1972-1975, 1984-1990, 1999-2008; Norway 1935-1940, 1945-1965, 1971-1981, 1986-1997, 2000-2001, 2005-2013; Spain 1931-1933, 1936-1939, 1982-1996, 2004-2011; Sweden 1920-1921, 1924-1926, 1932-1976, 1982-1991, 1994-2006; Switzerland 1944-1953, 1960-2013; United Kingdom 1924, 1929-1935, 1945-1951, 1964-1970, 1974-1979, 1997-2010.

Left Executive – Time Marker (timemarkerp) is a count variable that runs from -5 to 5 and indicates the temporal distance between the observation-year and the year in which executive partisanship transitioned to a party identified as leftist in *Left Executive*.

Income Share of Top 1% (Top1incomeshare) is a continuous variable that measures the share of pre-tax income earned by individuals at the top 1% of the income distribution. This data is based on income tax returns and is from the *World Top Incomes Database.* The *World Top Incomes Database* is the work of a wide number of scholars led by Tony Atkinson, Thomas Piketty, and

Emmanuel Saez. The data was accessed from http://topincomes.g-mond.parisschoolofecoomics.eu/ and is discussed at length in Atkinson and Piketty (2007, 2010) and in other publications associated with the project.

Income Share of Top 0.01% (Top001incomeshare) is a continuous variable that measures the share of pre-tax income earned by individuals at the top 0.01% of the income distribution. This data is based on income tax returns and is from the *World Top Incomes Database.* The *World Top Incomes Database* is the work of a wide number of scholars led by Tony Atkinson, Thomas Piketty, and Emmanuel Saez. The data was accessed from http://topincomes.g-mond.parisschoolofecoomics.eu/ and is discussed at length in Atkinson and Piketty (2007, 2010) and in other publications associated with the project.

Wealth Share of Top 1% (wealth_p99_00_2014) is a continuous variable that measures the share of national wealth held by those in the top 1% of the wealth distribution. These data are from Ohlsson, Roine, and Waldström (2007) and Roine and Waldström (2014) and supplemented with data by Turner (2010) for Ireland.

Majoritarian Electoral System (majoritarian1) is a dichotomous variable that is coded as one for country-years in which the national-level (lower house) legislative seats were distributed by majoritarian electoral institutions, and zero in years where seats were distributed by proportional, multi-tier, or mixed electoral systems. Electoral systems are classified as majoritarian if they use one of the following types of electoral rules: single member district plurality, single nontransferable vote, block vote, party block vote, borda count, modified borda count, limited vote, two-round system (majority-plurality and majority-runoff), and alternative vote. In order to code this variable, we consulted Mackie and Rose (1991), Caramani (2000), Golder (2005), Bormann and Golder (2013), and Croissant (2002). *Majoritarian Electoral System* is coded one for the following country years: Australia 1901-2013; Austria 1907-1919; Belgium 1830-1898; Canada 1867-2013; Denmark 1848-1919; France 1800-1918, 1928-1945, 1958-1985, 1988-2013; Germany 1871-1917; Italy 1861-1918; Japan 1889-1945, 1947-1995; Netherlands 1848-1917; New Zealand 1856-1995; Norway 1905-1918; South Korea 1948-1962; Spain 1812-1923; Sweden 1866-1908; Switzerland 1848-1918; United Kingdom 1800-2013; United States 1800-2013.

Inflation Crises (infl_cris) is a binary variable that is coded as 1 in years in which the country experiences an inflation crisis and 0 otherwise. Following Reinhart and Rogoff (2009, 4-5), an inflation crisis is identified as an annual inflation rate of 20% or more. The source of this data is Reinhart and Rogoff (2009).

Currency Crises (currency_cris) is a binary variable that is coded as 1 in years in which the country experiences a currency crisis and 0 otherwise. Following Reinhart and Rogoff (2009, 5-6), a currency crisis is identified as an annual currency depreciation of 15 percent or higher relative to an appropriate anchor currency. The source of this data is Reinhart and Rogoff (2009).

Banking Crises (bank_cris) is a binary variable that is coded as 1 in years in which the country experiences a banking crisis and 0 otherwise. Following Reinhart and Rogoff (2009, 8-10), a banking crisis is identified as "(1) bank runs that lead to the closure, merging, or takeover by the

public sector of one or more financial institutions and (2) if there are no runs, the closure, merging, takeover, or large-scale government assistance of an important financial institution (or group of institutions) that marks the start of a string of similar outcomes for other financial institutions." (Reinhart and Rogoff 2009, 11). The source of this data is Reinhart and Rogoff (2009).

External Debt Crises (debt_cris_ext) is a binary variable that is coded as 1 in years in which the country experiences an external debt crisis and 0 otherwise. Following Reinhart and Rogoff (2009, 10-13), an external debt crisis is identified as cases in which a government fails to pay either a principal payment or an interest payment by the set due date on a loan issued under a foreign jurisdiction. The source of this data is Reinhart and Rogoff (2009).

Domestic Debt Crises (debt_cris_dom) is a binary variable that is coded as 1 in years in which the country experiences a domestic debt crisis and 0 otherwise. Following Reinhart and Rogoff (2009, 13-14), a domestic debt crisis is identified as cases in which a government fails to pay either a principal payment or an interest payment by the set due date on a loan issued under its own jurisdiction. The source of this data is Reinhart and Rogoff (2009).

Stock Crash (stock_crash) is a binary variable that is coded as 1 in years in which the country experiences a stock market crash and 0 otherwise. Following Reinhart and Rogoff (2009, 250), a stock market crash is identified as cases in which there is a 25% or greater cumulative decline in real equity prices. The source of this data is Reinhart and Rogoff (2009).

Economic Crises (any_cris) is a binary variable that is coded as 1 in years in which the country experiences any form of economic crises identified above (*Inflation Crises, Currency Crises, Banking Crises, Domestic Debt Crises, External Debt Crises, and Stock Crash*) and 0 otherwise. This variable is derived from the variables obtained from Reinhart and Rogoff (2009).

Neighbor's Top Income Tax Rate (Neightop_incrate_n) is a continuous variable identifying the average *Top Income Tax Rate* of the country's neighbors in a given year. We identify neighboring states as dyads identified as being contiguous at level four or closer by the Correlates of War project (Sinnett et al 2002). Level four contiguity is defined as dyads that are separated by 150 miles of water or less. Dyads in our sample that meet this definition are as follows: Austria-Germany, Austria-Italy, Austria-Switzerland, Belgium-France, Belgium-Germany, Belgium-Netherlands, Belgium-United Kingdom, Canada-United States, Denmark-Germany, Denmark-Netherlands, Denmark-Norway, Denmark-Sweden, Finland-Norway, Finland-Sweden, France-Germany, France-Italy, France-Spain, France-Switzerland, France-United Kingdom, Germany-Netherlands, Germany-Sweden, Germany-Switzerland, Ireland-United Kingdom, Italy-Switzerland, Japan-South Korea, Netherlands-United Kingdom, and Norway-Sweden. We make an exception to this coding for Australia and New Zealand. While these states do not meet this threshold of contiguity, their relative proximity and otherwise close ties is consistent with our definition of neighboring states.

Trade Openness (trade_openness) is a continuous variable that identifies the sum of a state's exports and imports as a proportion of its gross domestic product. The imports and exports data are from Barbieri, Keshk, and Pollins (2009), while the GDP data come from Maddison (2003)

as updated by Bolt and van Zanden (2014). The Maddison Project data are available online at http://www.ggdc.net/madison/madison-project/home.htm. We have converted the GDP data from real GDP in 1990 international dollars to nominal GDP in order to make these values comparable to the trade data.

Capital Openness (capital) is an index measuring legal restrictions on the movement of capital in or out of the country and was developed by Dennis Quinn (1997). This index ranges from a minimum of 0 to a maximum of 100.

GDP per capita (rgdppc) is a continuous variable identifying real gross domestic product per capita in 1990 Geary-Khamis (international) dollars, as originally compiled by Maddison (2003) and updated by Bolt and van Zanden (2014).

Population (pop) is a continuous variable measuring the national population in thousands. The source Maddison (2003) and updated by Bolt and van Zanden (2014).

GDP (rgdp) is a continuous variable identifying real gross domestic product in 1990 Geary-Khamis (international) dollars, as originally compiled by Maddison (2003) and updated by Bolt and van Zanden (2014).

Growth (growth) is a continuous variable measuring the year-over-year change in the value of real GDP as a proportion of the previous year's real GDP. We calculated *Growth* using *GDP* as identified above.

Recession (recession) is a binary variable that is set equal to 1 if *Growth* is negative and otherwise 0. We calculated *Recession* using *Growth* as identified above.

Tax Revenue (revgdp_imf) is a continuous variable that identifies tax revenues as a percent of GDP. These data are from the IMF.

Tax Revenue – Interpolated (revgdp_imf_ip) is *Tax Revenue,* as identified above, with interpolated data points. Because there is considerable missing data in *Tax Revenue,* we use linear interpolation to estimate these values. We do not, however, extrapolate beyond the bounds of our existing data and interpolate only for country-years in which data are available both before and after the missing observation.

Trend – Australia (australiatrend) is a country-specific time trend by *Half Decade* for use in the five-year analysis.

Trend – Austria (austriatrend) is a country-specific time trend by *Half Decade* for use in the five-year analysis.

Trend – Belgium (belgiumtrend) is a country-specific time trend by *Half Decade* for use in the five-year analysis.

Trend – Canada (canadatrend) is a country-specific time trend by *Half Decade* for use in the five-year analysis.

Trend – Denmark (denmarktrend) is a country-specific time trend by *Half Decade* for use in the five-year analysis.

Trend – Finland (finlandtrend) is a country-specific time trend by *Half Decade* for use in the five-year analysis.

Trend – France (francetrend) is a country-specific time trend by *Half Decade* for use in the five-year analysis.

Trend – Germany (germanytrend) is a country-specific time trend by *Half Decade* for use in the five-year analysis.

Trend – Ireland (irelandtrend) is a country-specific time trend by *Half Decade* for use in the five-year analysis.

Trend – Italy (italytrend) is a country-specific time trend by *Half Decade* for use in the five-year analysis.

Trend – Japan (japantrend) is a country-specific time trend by *Half Decade* for use in the five-year analysis.

Trend – Netherlands (netherlandstrend) is a country-specific time trend by *Half Decade* for use in the five-year analysis.

Trend – New Zealand (newzealandtrend) is a country-specific time trend by *Half Decade* for use in the five-year analysis.

Trend – Norway (norwaytrend) is a country-specific time trend by *Half Decade* for use in the five-year analysis.

Trend – South Korea (koreatrend) is a country-specific time trend by *Half Decade* for use in the five-year analysis.

Trend – Spain (spaintrend) is a country-specific time trend by *Half Decade* for use in the five-year analysis.

Trend – Sweden (swedentrend) is a country-specific time trend by *Half Decade* for use in the five-year analysis.

Trend – Switzerland (switzerlandtrend) is a country-specific time trend by *Half Decade* for use in the five-year analysis.

Trend – United Kingdom (uktrend) is a country-specific time trend by *Half Decade* for use in the five-year analysis.

Trend – United States (usatrend) is a country-specific time trend by *Half Decade* for use in the five-year analysis.

Lü, Scheve, and Stasavage Survey Experiment Results Dataset
File Name: Scheve_Stasavage_TaxingtheRich_2016_LSSsurveyexperiment.dta

Description: This is the survey experiment results dataset in Stata format needed to replicate the analysis in Chapter 2.

ID (ID) is a unique identifier assigned to the respondent.

Weight (weight) is the sampling weight calculated to remove remaining imbalances to the marginal distribution of sociodemographics in the population.

Respondent Birth Year (birthyr) is a continuous variable identifying the respondent's year of birth.

Respondent Education (educ) is a categorical variable identifying the respondent's highest level of education at the time of the survey. Educational categories include less than high school, high school graduate, some college, completed a two-year degree, completed a four-year degree, and completed a post-graduate degree.

Respondent Gender (gender) is a categorical variable identifying the respondent's gender as either male or female.

Plan Choice 1 (taxincrease_dfit1) identifies whether for the first pair of tax policies presented the respondent selected Plan A (a proportional plan) or Plan B (a progressive plan).

Plan Choice 2 (taxincrease_dfit2) identifies whether for the second pair of tax policies presented the respondent selected Plan A (a proportional plan) or Plan B (a progressive plan).

Plan Choice 3 (taxincrease_dfit3) identifies whether for the third pair of tax policies presented the respondent selected Plan A (a proportional plan) or Plan B (a progressive plan).

Plan Choice 4 (taxincrease_dfit4) identifies whether for the fourth pair of tax policies presented the respondent selected Plan A (a proportional plan) or Plan B (a progressive plan).

Plan Choice 4–Explanation (qualreason_choice4) is the explanation given by the respondent for their decision with respect to *Plan Choice 4.* The content of this variable was the basis for coding *Plan Choice 4–Explanation Type,* which is defined below.

Plan Choice 4–Explanation Type (Reasons_DefExp) is a coded categorical variable that identifies the explanation given by the respondent for their decision with respect to *Plan Choice 4,* the content of which is given in *Plan Choice 4–Explanation,* described above. The possible categories of explanation are as follows: self-interest, equal treatment, ability-to-pay, compensatory, economic efficiency, progressive treatment, general fairness, other, and multiple reasons.

Personal Income–Below $25,000 (personalinc_blw25k) is a binary variable that is set equal to one if the respondent's income is below $25,000 and zero otherwise.

Personal Income–$25,000 to $200,000 (personalinc_btw25k_200k) is a binary variable that is set equal to one if the respondent's income is between $25,000 and $200,000 and zero otherwise.

Personal Income–Greater than $200,000 (personalinc_greater_200k) is a binary variable that is set equal to one if the respondent's income is greater than $200,000 and zero otherwise.

Tax Schedules Dataset
File Name: Scheve_Stasavage_TaxingtheRich_2016_schedules.dta

Description: This is the tax schedules dataset in Stata format needed to replicate Figure 3.3.

Country (country) is the observation country name in English. This dataset includes the following countries: France, Germany, New Zealand, Sweden, the United Kingdom, and the United States.

COW Code (ccode) is the unique country identifier for the observation country from the Correlates of War project. Please note the following coding decisions, selected to maintain consistency throughout the series:

Germany – We use the COW Code for unified Germany (255) throughout the series and do not switch to the distinct COW Code for the German Federal Republic (260) throughout the Cold War.

Year (year) is the observation year. The dataset includes observations at six 25-year intervals from 1875 to 2000 as well as 2010.

Percentiles (percentiles) is a continuous variable that identifies the *French* income distribution percentile reported by Piketty (2001). We use Piketty's (2001) data to calculate the pertinent multiples of nominal GDP per capita (see the description of *Multiples*, below), which we then use to calculate comparable measures of progressivity across the six countries included in this dataset. We include the following values of *Percentiles* in this dataset: p10, p50, p90, p95, p99, and p99.99. See also footnote 5 to chapter 3 in the text for more information.

Multiples (multiples) is a continuous variable identifying income groups in terms of multiples of nominal GDP per capita. In the dataset we include the following values: 0.5, 1, 1.5, 2, 4, and 100. The values of *Multiples* correspond to those of *Percentiles* (see description of *Percentiles*, above) such that a *Multiples* value of 0.5 is indicative of incomes at the 10^{th} percentile of the French income distribution, a *Multiples* value of 1 is indicative of incomes at the 50^{th} percentile of the French income distribution, a *Multiples* value of 1.5 is indicative of incomes at the 90^{th} percentile of the French income distribution, a *Multiples* value of 2 is indicative of incomes at the 95^{th} percentile of the French income distribution, a *Multiples* value of 4 is indicative of incomes at the 99^{th} percentile of the French income distribution, and a *Multiples* value of 100 identifies incomes at the 99.99^{th} percentile of the French income distribution.

Multiples – Standardized (multiples_stand) is an ordinal variable that recodes *Multiples* such that a *Multiples – Standardized* value of 1 identifies an income level of 0.5 times the nominal GDP per capita, a *Multiples – Standardized* value of 2 identifies an income level of 1 times the nominal GDP per capita, a *Multiples – Standardized* value of 3 identifies an income level of 1.5 times the nominal GDP per capita, a *Multiples – Standardized* value of 4 identifies an income level of 2 times the nominal GDP per capita, a *Multiples – Standardized* value of 5 identifies an income level of 4 times the nominal GDP per capita, and a *Multiples – Standardized* value of 6

identifies an income level of 100 times the nominal GDP per capita. This variable was coded and included for presentational purposes.

Tax Rate (taxrate) is the marginal income tax rate paid by an individual in a given income bracket, defined as a particular multiple of nominal GDP per capita (see the description of *Multiples*, above) in a given country-year. The sources for each country are as follows: France (Piketty 2001; OECD 2014); Germany (Dell 2008; Bundesministerium der Finanzen 2013; OECD 2014); New Zealand (Land Tax and Income Tax 1892; McAlister 2012; OECD 2014); Sweden (Du Rietz, Johansson, and Stenkula 2010; OECD 2014); United Kingdom (Mitchell 1988; Finance Acts of 1925 and 1948; OECD 2014); United States (IRS various years; OECD 2014).

U.S. Public Opinion during WWII
File Name: Scheve_Stasavage_TaxingtheRich_2016_Gallup.dta

Description: This is the Gallup poll descriptive statistics dataset in Stata format needed to replicate Figure 3.11. These data are from Gallup Poll #1941-0242 and Gallup Poll #1942-0263.

Income (Income) is the income category presented to respondents.

Preferred Effective Tax (Median) – Low SES, 1941 (EffTax1941class1) is the median preferred effective tax rate for families of four in a particular income bracket among qualitatively (interviewer) coded low-socioeconomic status respondents in 1941. Source: Gallup Poll #1941-0242.

Preferred Effective Tax (Median) – Medium SES, 1941 (EffTax1941class2) is the median preferred effective tax rate for families of four in a particular income bracket among qualitatively (interviewer) coded medium-socioeconomic status respondents in 1941. Source: Gallup Poll #1941-0242.

Preferred Effective Tax (Median) – High SES, 1941 (EffTax1941class3) is the median preferred effective tax rate for families of four in a particular income bracket among qualitatively (interviewer) coded high-socioeconomic status respondents in 1941. Source: Gallup Poll #1941-0242.

Preferred Effective Tax (Median) – Low SES, 1942 (EffTax1942Wclass1) is the median preferred effective tax rate for families of four in a particular income bracket among qualitatively (interviewer) coded low-socioeconomic status respondents in 1942. Source: Gallup Poll #1942-0263.

Preferred Effective Tax (Median) – Medium SES, 1942 (EffTax1942Wclass2) is the median preferred effective tax rate for families of four in a particular income bracket among qualitatively (interviewer) coded medium-socioeconomic status respondents in 1942. Source: Gallup Poll #1942-0263.

Preferred Effective Tax (Median) – High SES, 1942 (EffTax1942Wclass3) is the median preferred effective tax rate for families of four in a particular income bracket among qualitatively (interviewer) coded high-socioeconomic status respondents in 1942. Source: Gallup Poll #1942-0263.

Public Spending (Lindert) Dataset
File Name: Scheve_Stasavage_TaxingtheRich_2016_Lindert.dta

Description: This is the 1880-1930 ten-year public spending dataset in Stata format needed to replicate the analysis of Lindert's findings in Chapter 5 and the corresponding section of the online appendix. Note that the coverage for independent variables in this dataset dates back to 1870 in order to estimate models with lagged independent variables.

Country (country) is the observation country name in English. This dataset includes the following countries: Australia, Austria, Belgium, Canada, Denmark, Finland, France, Italy, Japan, the Netherlands, New Zealand, Norway, Spain, Sweden, the United Kingdom, and the United States.

COW Code (ccode) is the unique country identifier for the observation country from the Correlates of War project. Please note the following coding decisions, selected to maintain consistency throughout the series:

> Austria – We use the COW Code for modern Austria (305) throughout the series and do not use the distinct COW Code for Austria-Hungary (300) prior to 1919.

Year (year) is the decade start year for a given country-decade observation. The dataset includes observations every ten years from 1870 to 1930 for countries that obtained independence 1870 or earlier. Decade start years differ for countries that gained independence after 1870, in which case the first decade runs from the year of independence to the end of the standardized decade. (For instance, Australia became independent in 1901 and, as a result, its first decade runs from 1901 to 1909.)

Decade (decade) is a count variable that identifies ten-year periods throughout the data. The variable has a minimum value of 1, corresponding to the period from 1870 to 1879, and a maximum value of 7, corresponding to the period from 1930 to 1939.

Public Spending (spending_sum) is a continuous variable that measures the percent of GDP spent by the government on four categories of social programs: health, housing, pensions, and welfare. The data come from Lindert (2004) and are recorded every ten years from 1880 to 1930.

War Mobilization (himobpopyear2p) is the proportion of the years in the observation-period for which *War Mobilization,* as defined under "Annual Dataset" above, is set equal to one.

Universal Male Suffrage (unisuffrage) is the proportion of the years in the observation-period for which *Universal Male Suffrage,* as defined under "Annual Dataset" above, is set equal to one.

Competitive Elections (competitive-elections) is the proportion of the years in the observation-period for which *Competitive Elections,* as defined under "Annual Dataset" above, is set equal to one.

Speech Coding
File Name: Scheve_Stasavage_TaxingtheRich_2016_speechcoding.dta

Related Files: Scheve_Stasavage_TaxingtheRich_2016_speechcoding1.dta
Scheve_Stasavage_TaxingtheRich_2016_speechcoding2.dta
Scheve_Stasavage_TaxingtheRich_2016_speechcoding3.dta
Scheve_Stasavage_TaxingtheRich_2016_speechcoding4.dta
Scheve_Stasavage_TaxingtheRich_2016_speechcoding5.dta
Scheve_Stasavage_TaxingtheRich_2016_speechcoding6.dta
Scheve_Stasavage_TaxingtheRich_2016_speechcoding7.dta
Scheve_Stasavage_TaxingtheRich_2016_speechcoding8.dta
Scheve_Stasavage_TaxingtheRich_2016_speechcoding9.dta
Scheve_Stasavage_TaxingtheRich_2016_speechcoding10.dta

Description: This is the parliamentary speech dataset in Stata format needed to replicate Figure 6.1. The related files are the ten imputations of missing values described in the online appendix to chapter 6.

Speech Number (spnum) identifies the speech observation.

Year (year) is the observation year. The dataset includes observations from the years 1909, 1914, 1915, 1916, 1917, and 1918.

After War (afterwar) is a binary variable set equal to 1 for speeches that were given after the United Kingdom's entry into World War I and 0 otherwise.

Speech Orientation – Student 1 (spo_student1) is a categorical variable corresponding with the first student coder's classification of the orientation of the speech. Speeches were coded as described below.

1 if the speech overall favors one of the following: a comprehensive, permanent income tax; a progressive rate structure for the income tax; higher tax rates on the rich; or a tax on the rich.

2 if the speech argues against one of the following: a permanent income tax, a tax on the rich, higher tax rates on the rich; or an income tax with a progressive rate structure.

3 if no orientation can be assessed, if the speech is a question or a provocation, or if the speaker explicitly expresses no preference.

Speech Orientation – Student 2 (spo_student2) is a categorical variable corresponding with the first student coder's classification of the orientation of the speech. Speeches were coded as described below.

1 if the speech overall favors one of the following: a comprehensive, permanent income tax; a progressive rate structure for the income tax; higher tax rates on the rich; or a tax on the rich.

2 if the speech argues against one of the following: a permanent income tax, a tax on the rich, higher tax rates on the rich; or an income tax with a progressive rate structure.

3 if no orientation can be assessed, if the speech is a question or a provocation, or if the speaker explicitly expresses no preference.

Speech Orientation – Student 3 (spo_student3) is a categorical variable corresponding with the first student coder's classification of the orientation of the speech. Speeches were coded as described below.

1 if the speech overall favors one of the following: a comprehensive, permanent income tax; a progressive rate structure for the income tax; higher tax rates on the rich; or a tax on the rich.

2 if the speech argues against one of the following: a permanent income tax, a tax on the rich, higher tax rates on the rich; or an income tax with a progressive rate structure.

3 if no orientation can be assessed, if the speech is a question or a provocation, or if the speaker explicitly expresses no preference.

Speech Orientation (spoimp) is a categorical variable identifying the speech orientation. Speeches were coded as described below. We assigned a speech to a particular orientation where two or more of our three student coders agreed on the coding. We left the speech orientation as missing where all three of our student coders disagreed on the coding.

1 if the speech overall favors one of the following: a comprehensive, permanent income tax; a progressive rate structure for the income tax; higher tax rates on the rich; or a tax on the rich.

2 if the speech argues against one of the following: a permanent income tax, a tax on the rich, higher tax rates on the rich; or an income tax with a progressive rate structure.

3 if no orientation can be assessed, if the speech is a question or a provocation, or if the speaker explicitly expresses no preference.

Speech Argument – Student 1 (spt_student1) is a categorical variable corresponding with the first student coder's classification of the argument. Speeches were coded as 1 if identified as equal

treatment, 2 if ability to pay, 3 if compensatory, and 4 if some other argument, including economic efficiency, bureaucratic efficiency, prudence, or some other fairness argument.

Speech Argument – Student 2 (spt_student2) is a categorical variable corresponding with the second student coder's classification of the argument. Speeches were coded as 1 if identified as equal treatment, 2 if ability to pay, 3 if compensatory, and 4 if some other argument, including economic efficiency, bureaucratic efficiency, prudence, or some other fairness argument.

Speech Argument – Student 3 (spt_student3) is a categorical variable corresponding with the third student coder's classification of the argument. Speeches were coded as 1 if identified as equal treatment, 2 if ability to pay, 3 if compensatory, and 4 if some other argument, including economic efficiency, bureaucratic efficiency, prudence, or some other fairness argument.

Speech Argument (sptimp) is a categorical variable identifying the argument used in the speech. Speeches were coded as 1 if identified as equal treatment, 2 if ability to pay, 3 if compensatory, and 4 if some other argument, including economic efficiency, bureaucratic efficiency, prudence, or some other fairness argument. We assigned a speech to a particular argument where two or more of our three student coders agreed on the coding. We left the speech argument as missing where all three of our student coders disagreed on the coding.

Onorato, Scheve, and Stasavage Replication Dataset
File Name: Scheve_Stasavage_TaxingtheRich_2016_OSSdata.dta

Description: This is the 1600-2000 annual dataset in Stata format needed to replicate Figure 7.1.

Country (country) is the observation country name in English. The sample of countries includes those Levy (1983) identified as great powers: Austria-Hungary (1600-1918); China (1949-2000); France (1600-2000); Italy (1861-1943); Japan (1905-1945); the Netherlands (1609-1713); the Ottoman Empire (1600-1699); Prussia/Germany (1740-2000); Russia/Soviet Union (1721-2000); Spain (1600-1808); Sweden (1617-1721); the United Kingdom (1600-2000); and the United States (1898-2000).

COW Code (ccode) is the unique country identifier for the observation country from the Correlates of War project. Please note the following coding decisions, selected to maintain consistency throughout the series:

> Austria-Hungary – *In contrast to the other datasets associated with this project*, we use the COW Code for Austria-Hungary (300) throughout the series because Austria-Hungary's classification as a great power ends with the dissolution of the Austro-Hungarian Empire and thus prior to the establishment of an independent Austria (305).
>
> Germany – We use the COW Code for unified Germany and its predecessor, Prussia (255), throughout the series and do not switch to the distinct COW Code for the German Federal Republic (260) throughout the Cold War.
>
> Russia/Soviet Union – We use the COW Code for Russia (365) throughout the series and do not switch to the occasionally used COW Code for the Soviet Union (364) during the Cold War.

Year (year) is the observation year. The dataset includes all years from 1600-2000 for which a given country was identified by Levy (1983) as a great power.

War Year (waryear) is a binary variable that is set equal to 1 in years in which the country was engaged in war and 0 otherwise.

Military Size (military1) is a continuous variable that identifies the number of troops (in thousands) that the national government has available for use against foreign adversaries. This definition does not include reserve troops, colonial troops, civil defense units, and domestic police forces. These data were obtained from multiple sources. Please see the online appendix to Onorato, Scheve, and Stasavage (2014) for full information.

Population (popul1) is a continuous variable that identifies a state's total population (in thousands). These data were obtained from multiple sources. Please see the online appendix to Onorato, Scheve, and Stasavage (2014) for full information.

Military Mobilization (mobil) is the proportion of the population engaged in the military. This variable was calculated by dividing *Military Size* by *Population.* These data were obtained from multiple sources. Please see the online appendix to Onorato, Scheve, and Stasavage (2014) for full information.

Equal Sacrifice Dataset
File Name: Scheve_Stasavage_TaxingtheRich_2016_equalsacrifice.dta

Description: This is the 1844-2000 annual dataset in Stata format needed to replicate Figure 8.1.

Year (year) is the observation year.

Google Ngrams – MA (ngrams) is the number of references to "equality of sacrifice" in the Google Books database. The data are presented as seven-year moving averages.

United Kingdom – MA (uk_ma) is the number of references to "equality of sacrifice" in the United Kingdom's parliamentary debates as reported in Hansard. The data are reported in the dataset as seven-year moving averages.

United States – MA (us_ma) is the number of references to "equality of sacrifice" in the United States's Congressional debates as reported in the Congressional Record. The data are reported in the dataset as seven-year moving averages.

Ballard-Rosa, Martin, and Scheve Survey Experiment Results Dataset
File Name: Scheve_Stasavage_TaxingtheRich_2016_BRMSsurveyexperiment.dta

Description: This is the survey experiment results dataset in Stata format needed to replicate the analyses in Chapter 9. These data are from Ballard-Rosa, Martin, and Scheve (2015).

ID (ID) is a unique identifier assigned to the respondent.

Weight (weight) is the sampling weight calculated to remove remaining imbalances to the marginal distribution of sociodemographics in the population.

Respondent Age (age) is a continuous variable identifying the respondent's age at the time of the survey.

Respondent Education (educ) is a categorical variable identifying the respondent's highest level of education at the time of the survey. Educational categories include less than high school, high school graduate, some college, completed a two-year degree, completed a four-year degree, and completed a post-graduate degree.

Respondent Gender (gender) is a categorical variable identifying the respondent's gender as either male or female.

Preferred Rate (Less than $10,000) (ideal_rate_num_lt10) is the preferred marginal income tax rate for families making less than $10,000 annually.

Preferred Rate ($10,000-$35,000) (ideal_rate_num_10_35) is the preferred marginal income tax rate for families making between $10,000 and $35,000 annually.

Preferred Rate ($35,000-$85,000) (ideal_rate_num_35_85) is the preferred marginal income tax rate for families making between $35,000 and $85,000 annually.

Preferred Rate ($85,000-$175,000) (ideal_rate_num_85_175) is the preferred marginal income tax rate for families making between $85,000 and $175,000 annually.

Preferred Rate ($175,000-$375,000) (ideal_rate_num_175_375) is the preferred marginal income tax rate for families making between $175,000 and $375,000 annually.

Preferred Rate (Greater than $375,000) (ideal_rate_num_gt375) is the preferred marginal income tax rate for families making more than $375,000 annually.

References

Atkinson, Anthony B. 2007. "The Distribution of Top Incomes in the United Kingdom, 1908-2000." In Anthony B. Atkinson and Thomas Piketty, eds. *Top Incomes over the Twentieth Century: A Contrast between Continental Europeans and English-Speaking Countries.* Oxford and New York: Oxford University Press.

Atkinson, Anthony B., and Thomas Piketty, eds. 2007. *Top Incomes over the Twentieth Century: A Contrast between Continental Europeans and English-Speaking Countries.* Oxford and New York: Oxford University Press.

Atkinson, Anthony B., and Thomas Piketty, eds. 2010. *Top Incomes Over the Twentieth Century: A Global Perspective.* Oxford and New York: Oxford University Press.

Ballard-Rosa, Cameron, Lucy Martin, and Kenneth Scheve. 2015. "The Structure of American Income Tax Policy Preferences." Stanford University Working Paper.

Barbieri, Katherine, Omar M. G. Keshk, and Brian Pollins. 2009. "Trading Data: Evaluating our Assumptions and Coding Rules." *Conflict Management and Peace Science* 26(5):471-491.

Boix, Carles, and Sebastian Rosato. 2001. "A Complete Dataset of Political Regimes, 1800-1999." Department of Political Science, University of Chicago.

Bolt, Jutta, and Jan Luiten van Zanden. 2014. "The Maddison Project: Collaborative Research on Historical National Accounts." *The Economic History Review* 67(3):627-651.

Bormann, Nils-Christian, and Matthew Golder. 2013. "Democratic Electoral Systems Around the World, 1946-2011." *Electoral Studies* 32(2):360-369.

Braun, R. Anton, and Ellen R. McGrattan. 1993. "The Macroeconomics of War and Peace." *NBER Macroeconomics Annual* 8:197-258.

Bundesministerium de Finanzen. 2013. "Kaufkraftvergleiche historischer Geldbeträge." http://www.bundesbank.de/Redaktion/ DE/Standardartikel/Statistiken/kaufkraftvergleiche_ historischer_geldbetraege.html?view =render%5BDruckversion%5D (accessed 24 June 2014).

Caramani, Daniele. 2000. *Elections in Western Europe since 1815: Electoral Results by Constituencies.* London: Macmillan Reference.

Christie, James. 1924. "The Legislative Council of New Zealand." *Journal of Comparative Legislation and International Law* 6(1):19-26.

Clodfelter, Michael. 2008. *Warfare and Armed Conflicts: A Statistical Encyclopedia of Casualty and Other Figures, 1494-2007*. 3rd Edition. McFarland & Company, Inc., Publishers: Jefferson, NC and London.

Croissant, Aurel. 2002. “Electoral Politics in South Korea.” in Aurel Croissant, ed., *Electoral Politics in Southeast and East Asia.* Bonn: Freidrich-Eibert-Stifung.

Danish Parliament. 2011. “The Parliamentary System of Denmark.” http://dipd.dk/wpcontent/uploads/2011/10/The Parliamentary System of Denmark.pdf.

Dell, Fabien. 2008. “L’Allemagne Inégale: Inégalités de revenus et de patrimoine en Allemagne, dynamique d’accumulation du capital et taxation de Bismarck à Schroder, 1870–2005.” Unpublished dissertation. Paris School of Economics.

Du Rietz, Gunnar, Dan Johannson, and Mikael Stenkula. 2010. “The Marginal Tax Wedge of Labor in Sweden from 1861 to 2009.” Stockholm: Research Institute of Industrial Economics.

Ellis, John, and Michael Cox. 1993. *The World War I Data Handbook.* Aurum Press: London.

Finance Act, 1925. 15 & 16 Geo. 5, Ch. 36. (available: http://www.legislation.gov.uk/ukpga/1925/36/pdfs/ukpga_19250036_en.pdf)

Finance Act, 1948. 11 & 12 Geo. 6, Ch. 49. (available: http://www.legislation.gov.uk/ukpga/1948/49/pdfs/ukpga_19480049_en.pdf)

Flora, Peter. 1983. *State, Economy, and Society in Western Europe, 1815-1975: A Data Handbook in Two Volumes.* Chicago, IL: St. James.

Genovese, Federica, Kenneth Scheve, and David Stasavage. 2014. “Comparative Income Taxation Database.” Stanford University and New York University.

Ghosn, Faten, Glenn Palmer, and Stuart Bremer. 2004. “The MID3 Data Set, 1993-2001: Procedures, Coding Rules, and Descriptions.” *Conflict Management and Peace Science* 21(2):133-154.

Golder, Matthew. 2005. “Democratic Electoral Systems Around the World, 1946-2000.” *Electoral Studies* 24(1):103-154.

Gray, Randal, and Christopher Argyle. 1991. *Chronicle of the First World War.* Facts on File: Oxford and New York, NY.

Griffin, Edward Geary. 1965. *The Adoption of Universal Manhood Suffrage in Japan.* New York, NY: Columbia University Press.

IRS. various years. *Internal Revenue Service Annual Statistics of Income Reports.* https://www.irs.gov/uac/SOI-Tax-Stats-Archive.

Land Tax and Income Tax, 1892. 56 Vict, No. 55. (available: http://www.nzlii.org/nz/legis/hist_act/laia189256v1892n55278/)

Levy, Jack. 1983. *War in the Modern Great Power System, 1495-1975.* Lexington, KY: University of Kentucky Press.

Lindert, Peter. 2004. *Growing Public: Social Spending and Economic Growth since the Eighteenth Century.* New York, NY: Cambridge University Press.

Mackie, Thomas T., and Richard Rose. 1991. *The International Almanac of Electoral History*. London: Macmillan.

Maddison, Angus. 2003. *The World Economy: History and Statistics.* Paris: OECD.

Marriott, John A. R. [1910] 1926. *Second Chambers: An Inductive Study in Political Science*. Oxford: Clarendon Press.

McAlister, Fiona. 2012. "Average Marginal Income Tax Rates for New Zealand, 1907-2009." New Zealand Treasury Working Paper 12/04.

Mitchell, B.R. 1988. *British Historical Statistics.* Cambridge: Cambridge University Press.

OECD. 2014. Income Tax Database. http://www.oecd.org/tax/tax-policy/tax-database.htm.

Ohlsson, Henry, Jasper Roine, and Daniel Waldenström. 2007. "Long Run Changes in the Concentration of Wealth: An Overview of Recent Findings." Stockholm School of Economics.

Onorato, Massimiliano, Kenneth Scheve, and David Stasavage. 2014. "Technology and the Era of the Mass Army." *Journal of Economic History* 74(2):449-481.

Pasquino, Gianfranco. 2009. "The Italian Senate." *The Journal of Legislative Studies* 8(3):67-78.

Piketty, Thomas. 2001. *Les hauts revenus en France au Xx^{e} siècle: Inegalities et redistributions 1901-1998.* Paris: Grasset.

Piketty, Thomas, and Emmanuel Saez. 2007. "Income and Wage Inequality in the United States." In Anthony B. Atkinson and Thomas Piketty, eds. *Top Incomes Over the Twentieth Century: A Contrast between Continental Europeans and English-Speaking Countries.* Oxford and New York: Oxford University Press.

Plagge, Arnd, Kenneth Scheve, and David Stasavage. 2011. *Comparative Inheritance Taxation Database*. Yale University, ISPS Data Archive.

Przeworski, Adam, Michael Alvarez, José A. Cheibub, and Fernando Limongi. 2000. *Democracy and Development: Political Institutions and Material Well-Being in the World, 1950-1990.* New York, NY: Cambridge University Press.

Quinn, Dennis. 1997. "The Correlates of Change in International Financial Regulation." *American Political Science Review* 91(3):531-551.

Reinhart, Carmen M., and Kenneth S. Rogoff. 2009. *This Time is Different: Eight Centuries of Financial Folly.* Princeton, NJ: Princeton University Press.

Roine, Jasper, and Daniel Waldenström. 2010. "Top Incomes in Sweden over the Twentieth Century." in Anthony B. Atkinson and Thomas Piketty, eds. *Top Incomes over the Twentieth Century: A Global Perspective.* Oxford and New York: Oxford University Press.

Roine, Jasper, and Daniel Waldenström. 2014. "Long-Run Trends in the Distribution of Income and Wealth." Uppsala Center for Fiscal Studies Working Paper 2014:5.

Saez, Emmanuel, and Michael R. Veall. 2007. "The Evolution of High Incomes in Canada 1920-2000." In Anthony B. Atkinson and Thomas Piketty, eds. *Top Incomes Over the Twentieth Century: A Contrast between Continental Europeans and English-Speaking Countries.* Oxford and New York: Oxford University Press.

Salverda, Wiemer, and Anthony B. Atkinson. 2007. "Top Incomes in the Netherlands over the Twentieth Century." in Anthony B. Atkinson and Thomas Piketty, eds. *Top Incomes over the Twentieth Century: A Contrast between Continental European and English-Speaking Countries.* Oxford and New York: Oxford University Press.

Scheve, Kenneth, and David Stasavage. 2012. "Democracy, War, and Wealth: Lessons from Two Centuries of Inheritance Taxation." *American Political Science Review* 106(1):81-102.

Scheve, Kenneth, and David Stasavage. 2016. *Taxing the Rich: A History of Fiscal Fairness in the United States and Europe.* Princeton, NJ: Princeton University Press.

The Senate of Canada, Committees and Private Legislation Directorate. 2001. "A Legislative and Historical Overview of the Senate of Canada." http://www.parl.gc.ca/about/senate/legisfocus/legislative-e.htm.

Singer, David J. 1987. "Reconstructing the Correlates of War Dataset on Material Capabilities of States, 1816-1985." *International Interactions* 14(2):115-132.

Sinnett, Douglas M., Jiroslav Tir, Phillip Schafer, Paul F. Diehl, and Charles Gochman. 2002. "The Correlates of War Project Direct Contiguity Data, Version 3." *Conflict Management and Peace Science* 19(2:58-66.

Söderberg, Hans. 1996. *Inkomstskattens utveckling under 1900-talet. En vägvisare för skatteberäkningar åren 1921–1996*. Stockholm: Skattebetalarnas förening [Swedish Taxpayers' Association].

Turner, John. 2010. "Wealth Concentration in the European Periphery: Ireland, 1858-2001." *Oxford Economic Papers* 62:625-646.