Dataset used in the article "Shifting tax burden to top income earners: What is the best way to reduce inequality?" by Jorge Onrubia, Fidel Picos and María del Carmen Rodado.

The empirical analysis of this article has been carried out using microdata from the 2011 Spanish Personal Income Tax Return Sample (this sample is officially named in Spanish "Muestra IRPF 2011 IEF-AEAT – Declarantes") disseminated by the Spanish Institute of Fiscal Studies (IEF). Microdata are collected by the Spanish Tax Administration Agency (AEAT). This database contains more than 2 million observations representative of more than 19 million tax returns, the entire Spanish population of PIT taxpayers (except the Basque Country and Navarre, since both Autonomous Communities have their own personal income taxes, which are administered by their respective regional governments).

The Spanish Personal Income Tax Return Sample (for every year) is an official database included in the National Statistics Plan, that is available to researchers and analysts, free of charge, upon individual request to the IEF. Access to information is subject to the signing of a Protocol that sets the conditions for the reception and use of anonymous microdata files of tax information provided by the IEF. According to this Protocol, the data must be used exclusively for academic or scientific purposes. Once requested data are received by the researcher, these cannot be transferred, under any pretext, to third parties

Please find the Protocol for requesting IEF's tax databases in:
http://www.ief.es/Investigacion/Est_peticion.vbhtml

Programs and other details of the computations sufficient to permit replication

---

- Both data management and statistical analysis has been carried out using Stata 13.

- Gini and Concentration indexes have been calculated using Stata do-file "ginis.ado".

- Stata do-file "1_stylized_progressive_tax.do" contains the Stata programming code to simulate the stylized tax $T = t(y)$ with the same revenue and redistribution impact as the Spanish Personal Income Tax actually applied in year 2011.

- Stata do-file "2_topreform.do" contains the programming code to simulate each of the reforms that increase the relative participation in tax liabilities of the richest, defined in the dataset as nivel=1 (nivel=0 are the "poor").

- Stata do-file "3_top1.do" contains the programming code that shows basic aggregates for each group (1-99 and 100).

- Stata do-file "4_reforms.do" contains the programming codes to define the different scenarios of reform, based on the values of ℓ (shifting $100\times\ell\%$ of the overall revenue from group 99 to group 100).

- Stata do-file "5_ginis_table.do" contains the programming code to compute table with Gini coefficient used in the article.