

Data for Besley and Persson "The Origins of State Capacity ..."

The data used to generate the tables and create the figure are in the Stata data file AER.dta

Each variable in the data set is clearly labelled.

All variables are scaled to lie in the unit interval by dividing with the largest value in the data set.

The do file used in creating the data does this by replacement as follows:

```
replace mincometaxrevenuegdp=mincometaxrevenuegdp/37.15
replace mtradetax_share=1-(mtradetax_share/0.683)
replace mtradeindirecttax_share = 1-(mtradeindirecttax_share/0.81)
replace mtaxrevenue=mtaxrevenue/50.89
replace mcred = 1-(mcred/174)
replace mpcy=mpcy/123.33
replace mprot=1-(mprot/173)
```

The commands to generate the tables are:

Table 1:

```
reg mpcy $politics75 $legal, robust
reg mcred $politics75 $legal, robust
reg mprot $politics75 $legal, robust
reg mgadp $politics75 $legal, robust
```

Table 2:

```
reg mtradetax_share $politics75 $legal, robust
reg mtradeindirecttax_share $politics75 $legal, robust
reg mincometaxrevenuegdp $politics75 $legal, robust
reg mtaxrevenue $politics75 $legal, robust
```

Each line corresponds to a column of each table in order

These use macros created as follows:

```
macro define politics75 "avg_war75 avg_dem75 avg_demparl75"
```

```
macro define legal " l_eng l_soc l_ger l_sca"
```

The graph is created with:

```
twoway (scatter newshareabove pcy if median=="Above Median Income",  
ytitle(Share of Income Taxes in GDP) msymbol(Oh) )  
(scatter newsharebelow pcy if median=="Below Median Income") (lfit  
incometaxrevenuegdp pcy, title(Figure 1))
```