# Supplementary Materials for "The Effect of Legislature Size on Public Spending: A Meta-Analysis"

Huzeyfe Alptekin* Danilo Freire† Umberto Mignozzetti‡

Catarina Roman§

June 13, 2020

# Contents



*Research Associate, Contemporary Brazilian History Research and Documentation Center, School of Social Sciences, Getulio Vargas Foundation, Brazil, huzeyfealptekin@gmail.com.

†Postdoctoral Research Associate, The Political Theory Project, Brown University, Providence, RI 02912, USA, danilofreire@brown.edu, http://danilofreire.github.io.

‡School of International Relations, Fundação Getulio Vargas, São Paulo, SP, Brazil and Wilf Family Department of Politics, NYU, NY, USA, umberto.mig@nyu.edu, http://umbertomig.com.

§Research Associate, Department of International Relations, Getulio Vargas Foundation, Brazil, catarinamroman@gmail.com.

# A Search criteria

The first step in reviews is to gather a sample of records. This initial procedure is the foundation for the entire meta-analytical process: if the goal is to aggregate all the available data on a given topic, the primary pool must contain all of it. At the same time, the sample should be concise enough to optimise the work flow. Upon selecting the online databases, we attempted to do a “term” search, based on a set of words we scouted from the distributive politics literature we already knew. This search produced oversized databases, of which a large share was completely unrelated to our subject of investigation. The strategy we found to be the most thorough and productive was to select the records that cited Weingast, Shepsle and Johnsen’s 1981 paper “The political economy of benefits and costs: A neoclassical approach to distributive politics”. Although Google Scholar reports the article amounts to 2,180 citations[1], as of the 21st of November 2019, our search resulted in a total of 2,664 records.

We webscraped the main databases in Social Sciences: Google Scholar (n = 1001); Microsoft Academic (n = 927); and Scopus (n = 736). The R script we wrote extracted, to a CSV file, the title, abstract, authors, year, journal of publication, and database from which the record originated. We screened these results with an English language and Article restriction – meaning, we excluded all

[1] As of May 11th, 2020.

records written in other languages and all that were not academic papers, such as book chapters, doctoral theses, etc. We set no restriction to unpublished articles.

# B Article Selection

The selection process was conducted by two authors in three phases. In the first round, we excluded all titles that were obviously unrelated to our topic of investigation. For instance, we curiously found articles about automobile motors amidst our sample. We consider this a preliminary step, since we were not able to eliminate a large number of entries. Thus, we read all abstracts. We chose to maintain those that indicated either government expenditure or legislative structures were central matters. For instance, if the paper sought to identify variables that increased government size, it was maintained. Abstracts that indicated the paper discussed or estimated the impacts of representative institutions, elections, or chamber dynamics, for example, were also included. This allowed us to significantly reduce our sample to 376 records.

In the second phase, we assessed full-texts. To proceed, the paper should (i) conduct a quantitative analysis, (ii) report data on the number of legislators, and (iii) also on public expenditure. If the record was marked as positive for all three, it was maintained. Disagreements in this phase were discussed among the authors, and a third was consulted when needed.

The third phase consisted of filling out tables for each of the remaining 50 articles to systematically evaluate their eligibility. Since government spending and the number of parliamentaries assume the form of different variables, we extracted the coefficients that provided this information in all papers, indicating and defining which variables each author used. By observing what is most commonly used, we could decide which variables we would employ in our analysis. In this phase, we also collected information on whether or not the paper had been published, and if it explicitly discussed the *law of 1/n*. Upon choosing the variables, we excluded the non-conforming studies, arriving at our final sample of 26 articles.

## B.1 Exclusion analysis

We selected the final pool of articles based on two criteria regarding their reported coefficients:

1. Matched treatment variable:

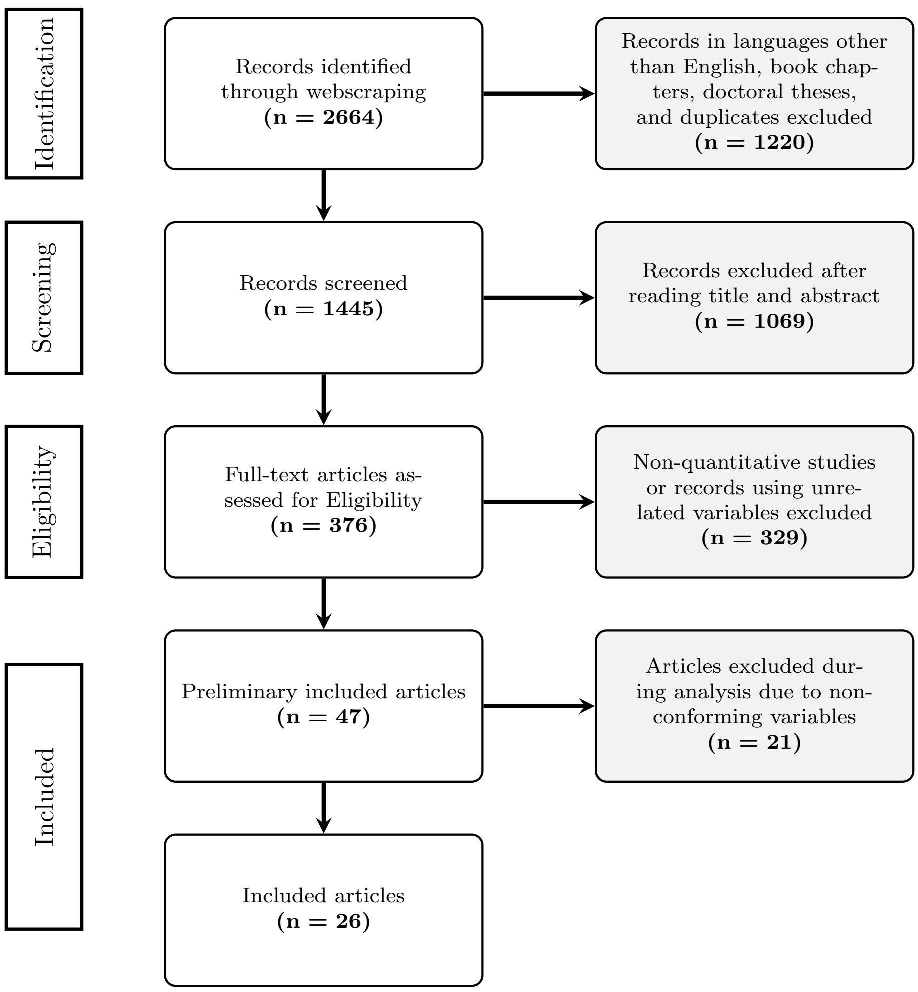
- *N*: Number Legislators Lower House
- *logN*: Log Number Legislators Lower House
- *K*: Number Legislators Upper House

2. Matched outcome variable:

- *ExpPC*: Expenditure Per Capita
- *logExpPC*: Log Expenditure Per Capita
- *PCTGDP*: Percent GDP Public Expenditure

## B.2 Flow Chart

We visually organised the flow of articles across eligibility phases through this diagram[2]. The column to the right depicts the amount of articles excluded in each phase, and the one to the left shows the number of records evaluated.




[2]According to the PRISMA statement for reporting meta-analyses and systematic reviews

# C Meta-analysis dataset

The meta-analytic data is comprised of two datasets. The first dataset has the main coefficients that were reported in the paper. It includes only the most rigorous model from each paper, that is, those estimated with the largest $n$, most control variables, and fixed effects if the authors added them. If the article employed a regression discontinuity design, we chose the coefficient from the optimal bandwidth or from the intermediate one. This sample encompasses 36 estimates, as 10 articles analysed two dependent or independent variables of interest. Our second sample, in contrast, contains all the 126 effect sizes reported in the 26 papers.

In the main text, we focus on the results for our restricted sample as we consider them more robust, but the findings are very similar when we use the extended set. Below is the data extraction process for all relevant coefficients in the selected articles. In the @ref(meta-an) and @ref(meta-reg) sections of this Appendix, you will find all tests performed in both reduced and full samples.

# D Descriptive statistics

In this section, we present the descriptive statistics for our meta-analytic sample. We focus in the following paper characteristics: study year, whether the paper has been published or not, the three dependent variables of interest, the three independent variables of interest, and a few statistics about the coefficients. We also add a descriptive statistics table, similar to the one in the main paper.

## D.1 Study Year

For study year, we have an average of 2009.33, with standard deviation of 6.5. The oldest study included in the paper is dated from 1998, while the most recent paper is dated from 2019. Therefore, we cover 21 years of tests of the 1/n hypothesis.

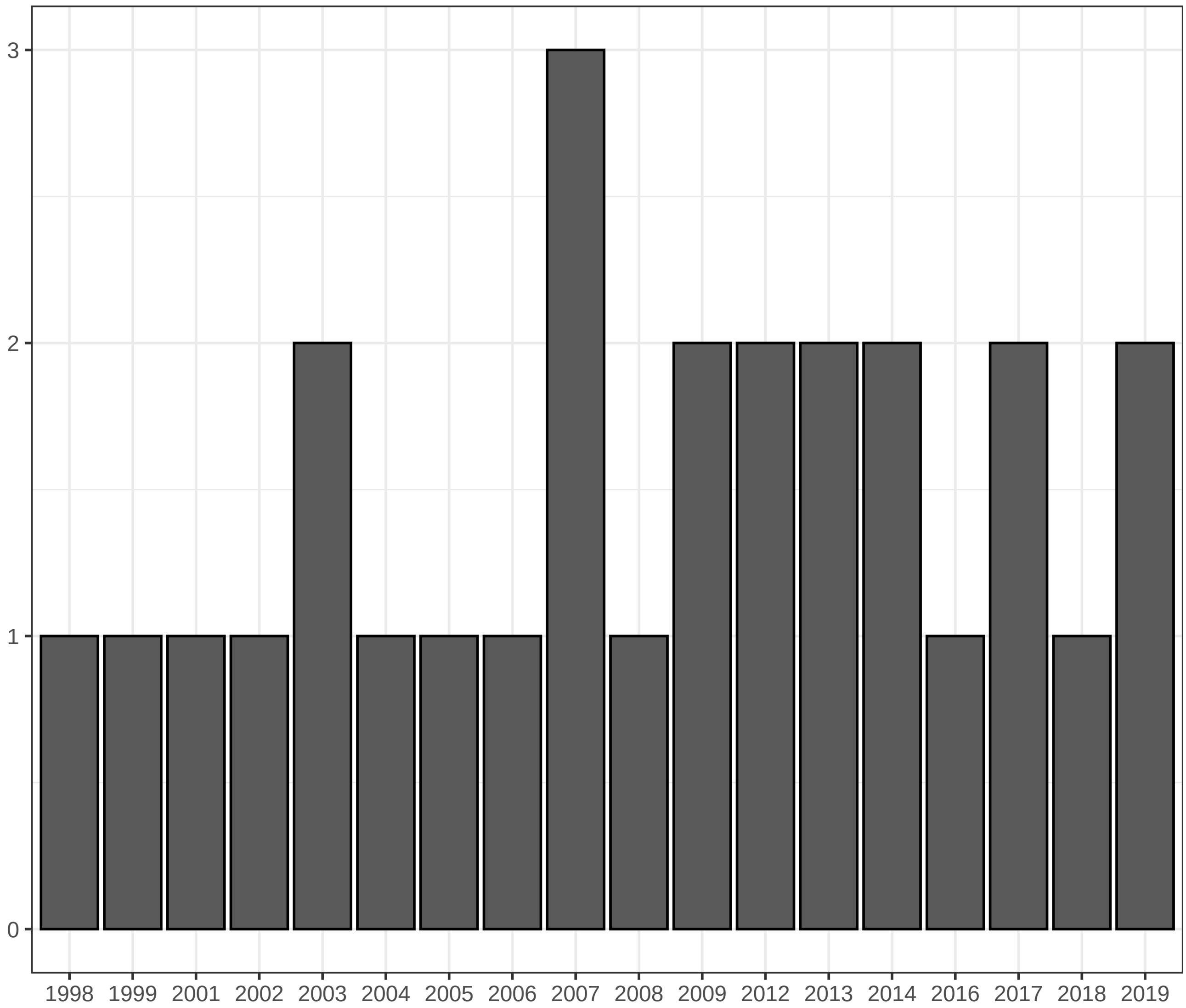


Figure 1: Study Year Frequencies

## D.2 Frequency of Published Papers

We decided to report studies regardless of their publication status. From the 26 papers in the sample, 22 were published while 4 were not published.

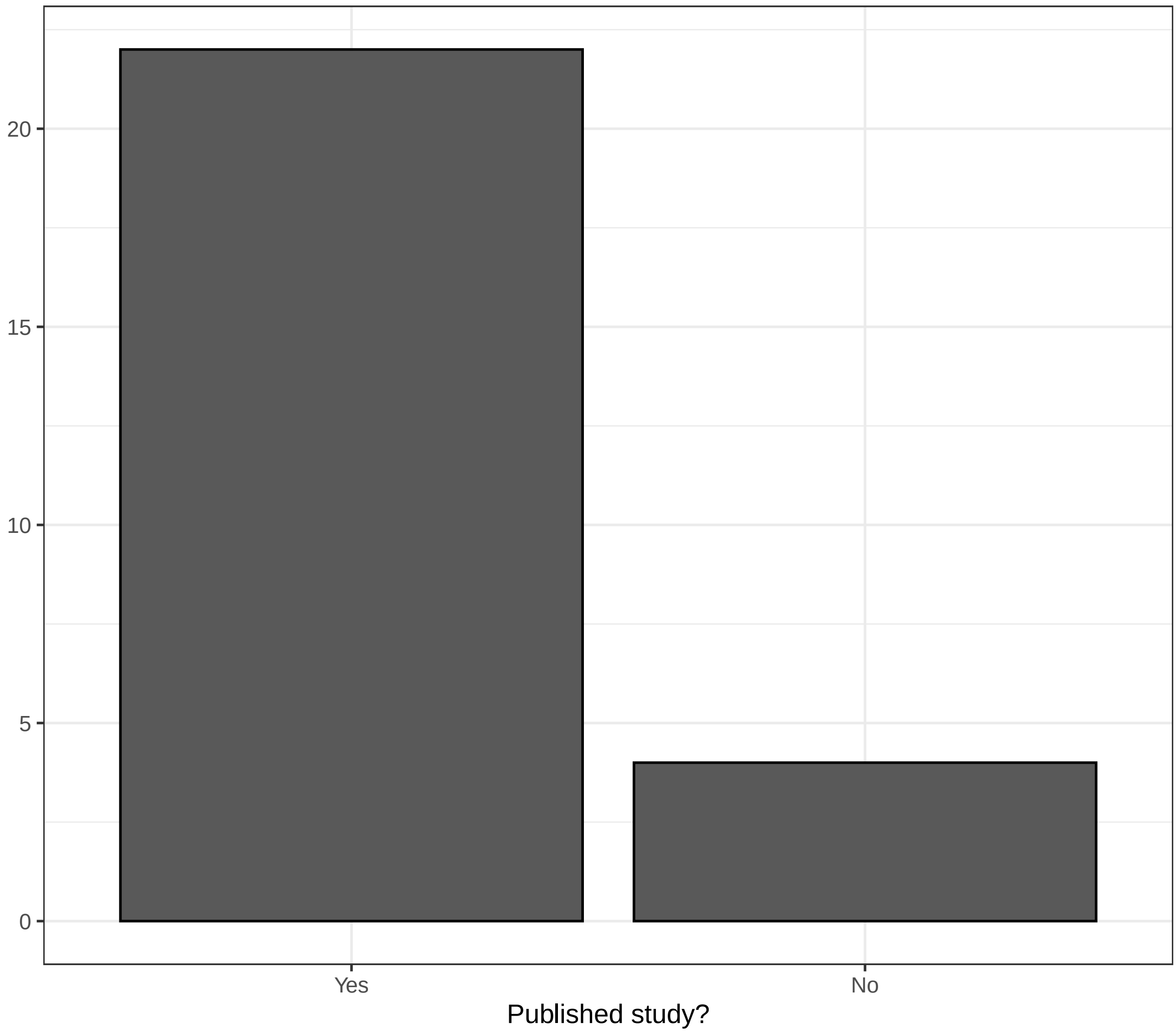



Figure 2: Was the study published?

## D.3 Electoral system

There are many differences in the papers relative to the research design. One remarkable difference is the application of the theory, that was build with majoritarian systems in mind, in non-majoritarian democracies. In our sample, 12 of the papers study *Majoritarian* systems while 14 study *Non-Majoritarian* electoral systems.[3]

[3]Note that the argument for working in a non-majoritarian system, we need to assume that despite the fact that politicians are able to campaign in every place in the district, the votes are geographycally concentrated. The concentration facilitates politicians to use pork-barrel project to captivate their electoral supporters.

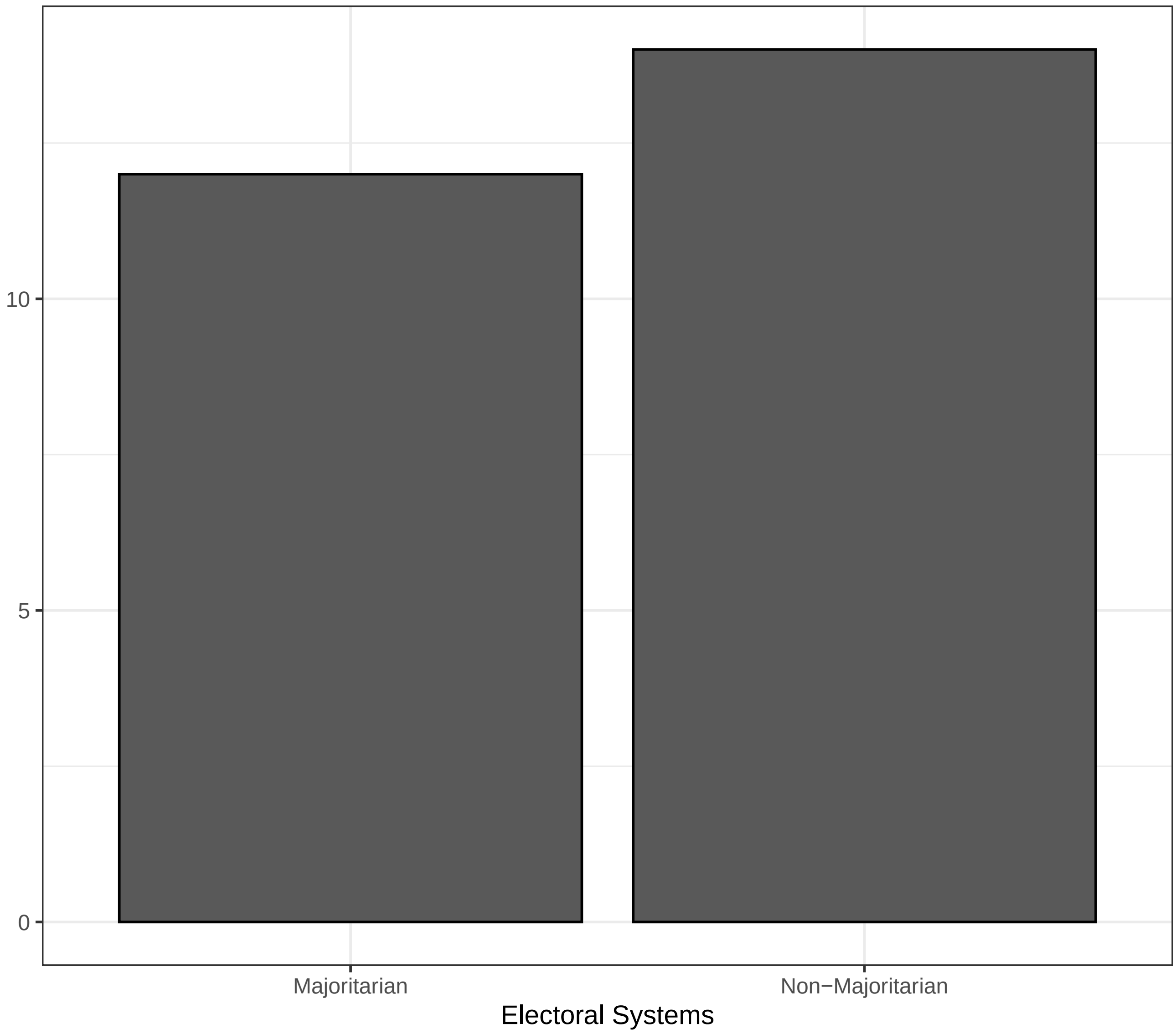



Figure 3: Electoral Systems

## D.4 Aggregation Level

The aggregation level is also an important characteristic of the empirical tests of the law of 1/n. The model was build to explain the dynamics of majoritarian countries, but the theory was tested in municipal, county, states, and country levels. In our sample, 6 studied *local* level (municipalities and counties), 15 studied *State* (or Provincial) levels, and 5 studied *Country* level data.

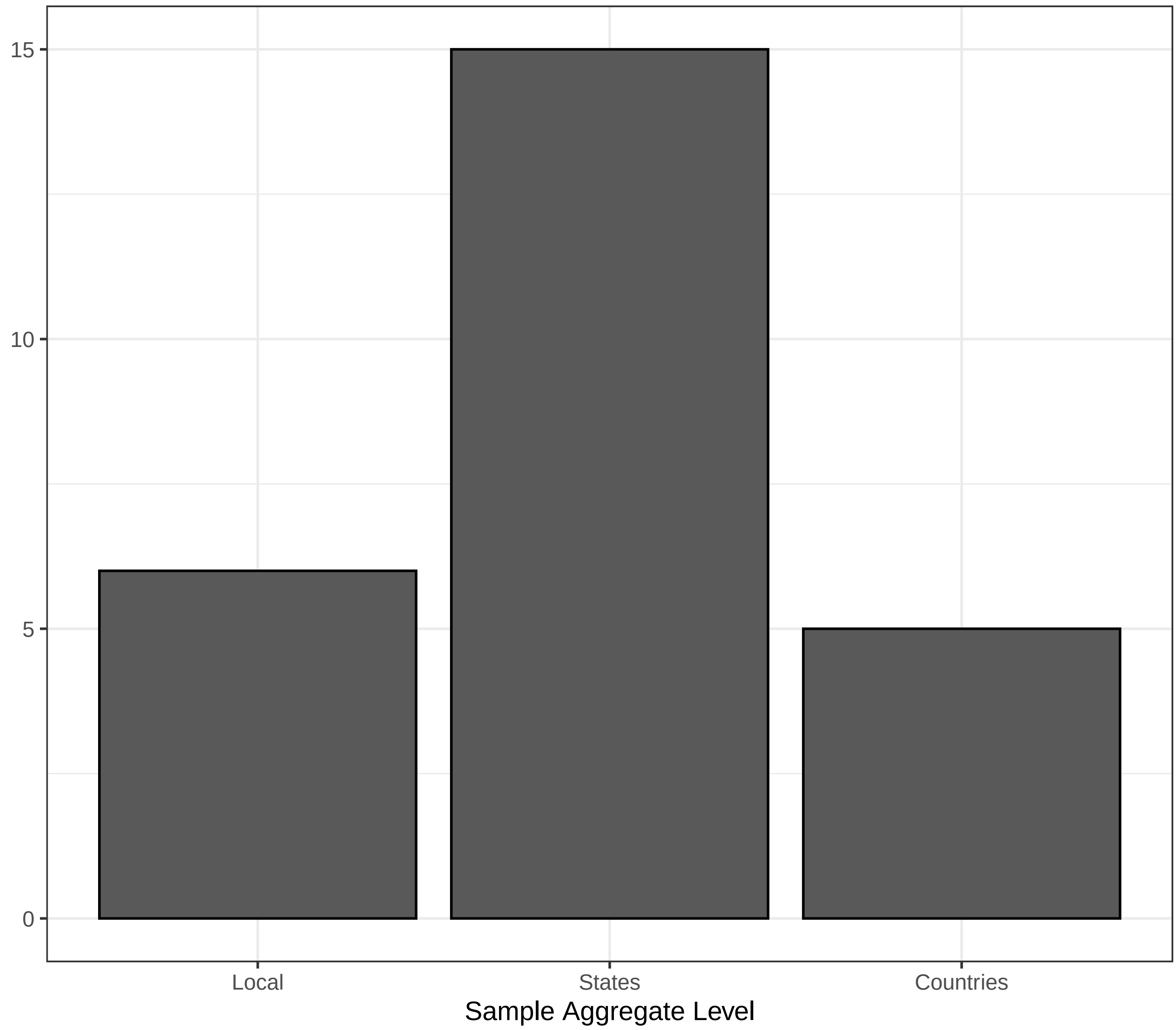



Figure 4: Sample Aggregate Level

## D.5 Dependent variables

The outcome variables included in the paper were:

- 13 Per Capita Expenditure papers
- 7 Natural Log of Per Capita Expenditure papers
- 8 Expenditure as a Percentage of the GDP papers

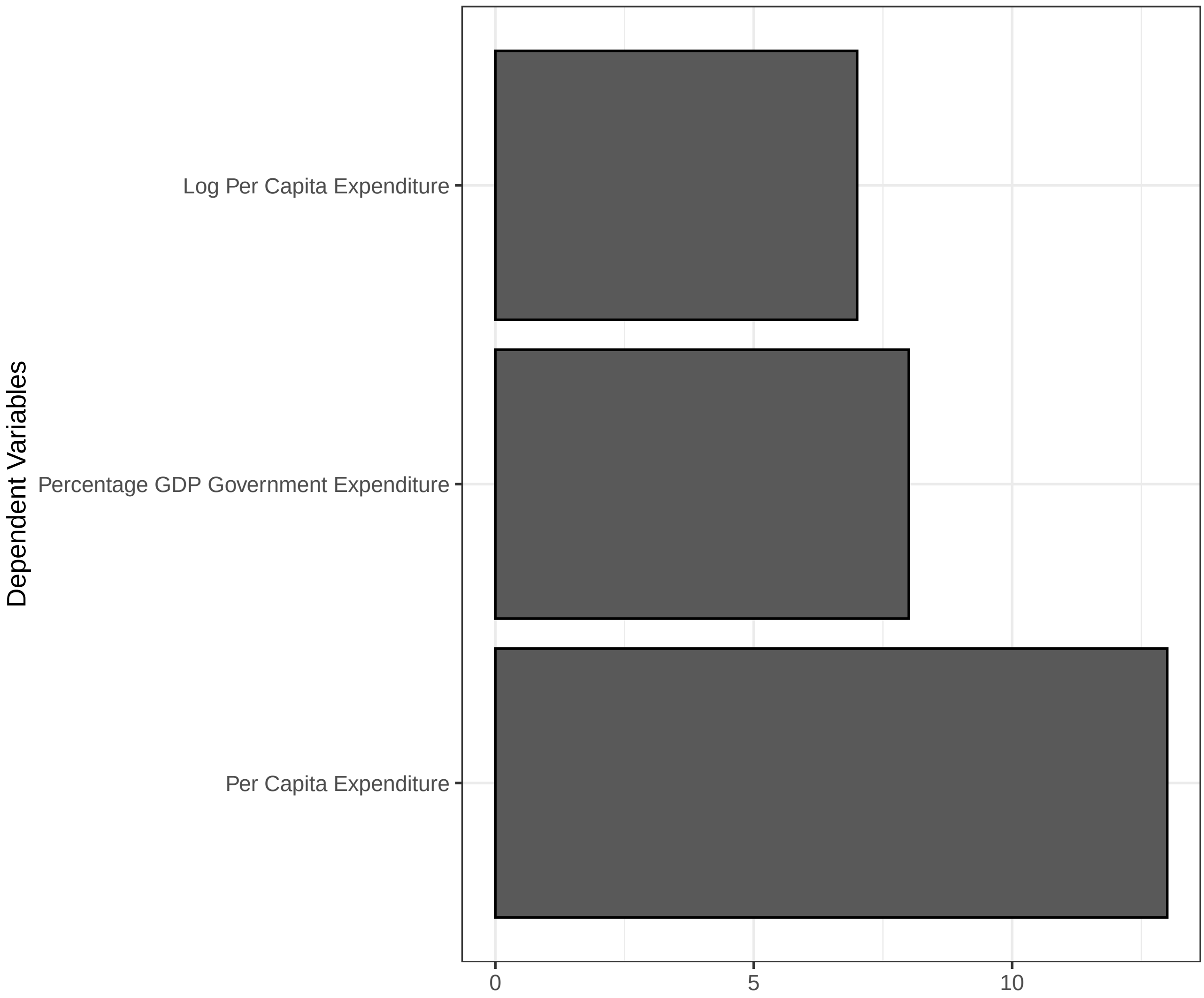


Figure 5: Dependent variables across the law of 1/n studies

## D.6 Independent variables

Most of papers in our meta-analytic sample study the number of legislators in the lower house (20). The second most frequent independent variable is the number of legislators in the upper house (9). Finally, the minority of papers study the Natural Log of the number of legislators in the lower house (5). Some papers had multiple coefficients, and thus the total number of coefficients is 36, while the number of papers is only 26.

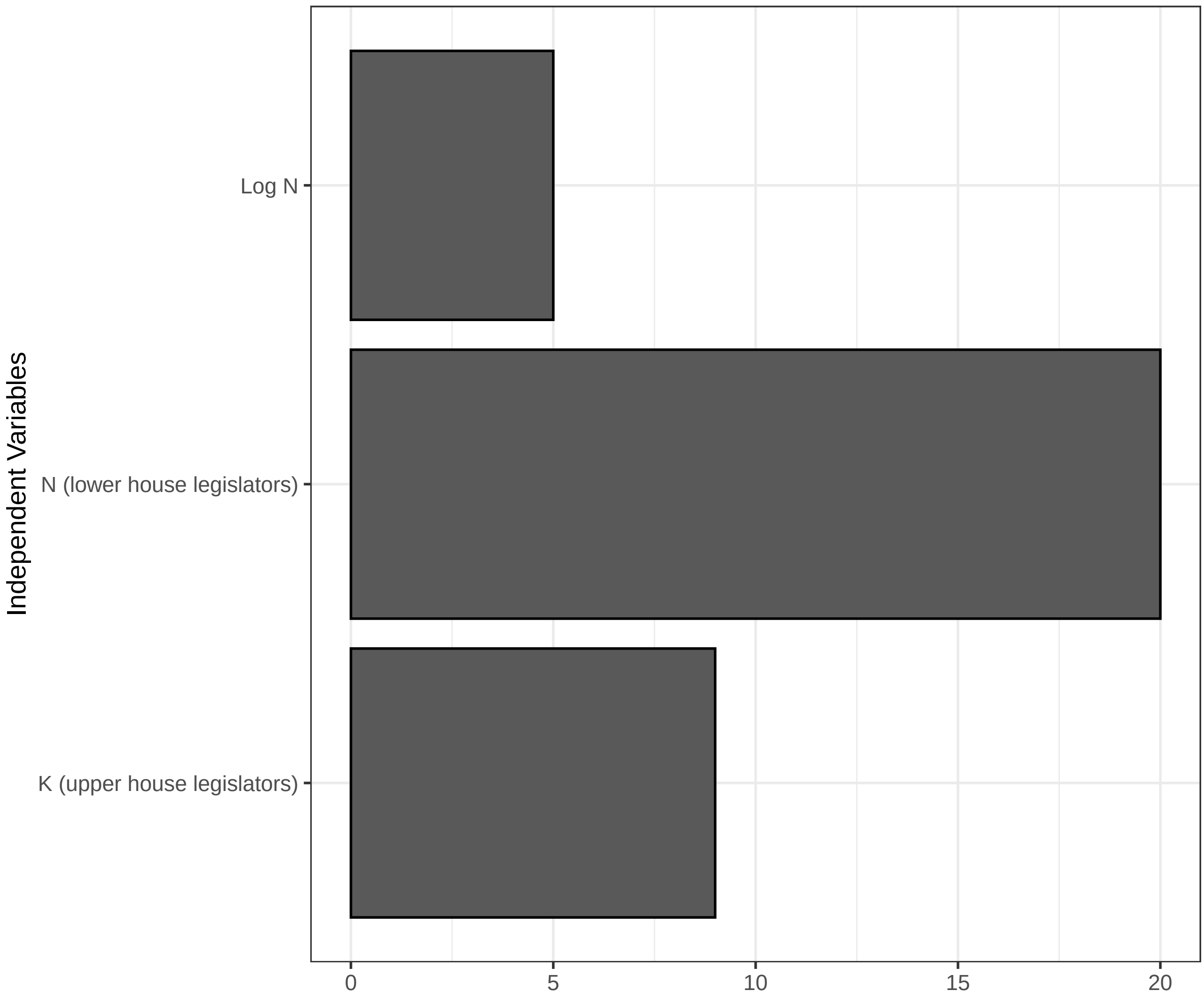
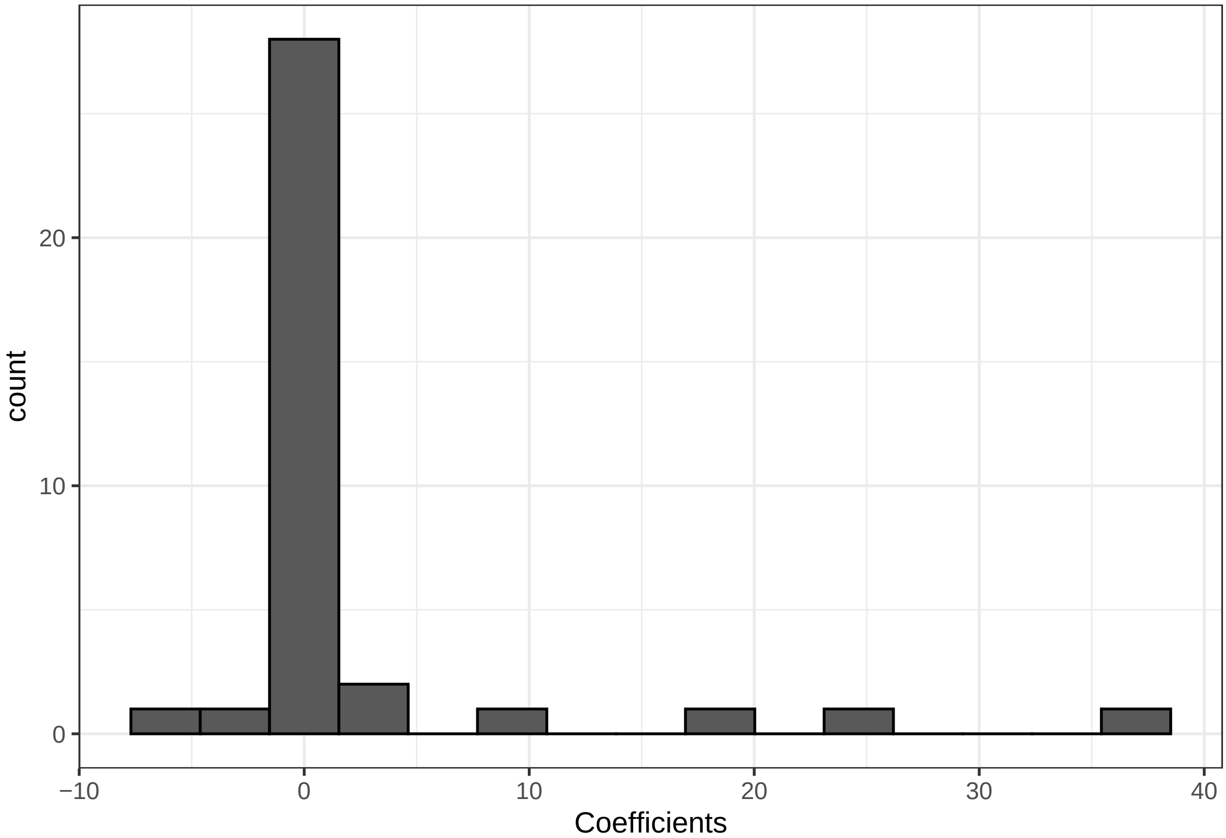
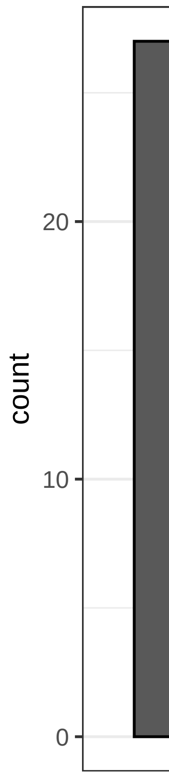


Figure 6: Independent variables across the law of 1/n studies

## D.7 Histogram of the Coefficients and the Standard Errors

The coefficients in the papers present a striking variability. In this section, we plot a histogram of the coefficients for all measurements included in the meta-analytic dataset.




## D.8 Sign Coefficients

One simple statistic that we can compute to access the law of 1/n is the frequency of positive and negative estimates. Below we plot the frequency for all the papers included in the meta-analytic dataset.

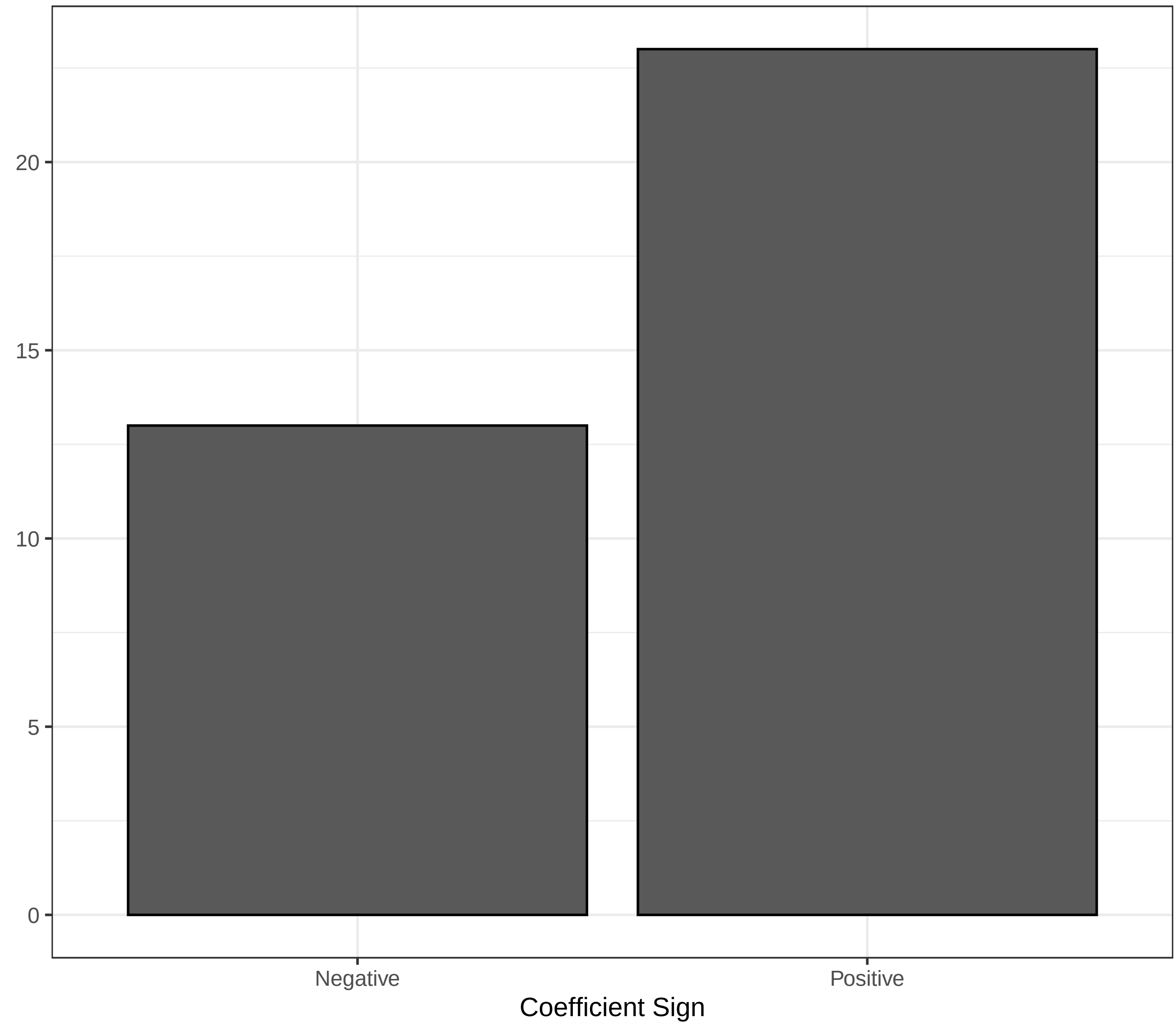



Figure 7: Coefficient Sign

# E Descriptive Statistics of Moderators

We chose a set of moderators that frequently appear in the literature and may help us interpret our results. We included them in our meta-regressions alongside an indicator for the type of independent variable used in the original study ($n$, $\log(n)$, or $k$). The additional moderators are: 1) electoral system; 2) data aggregation level; 3) estimation method; 4) publication year; 5) paper publication in an academic journal. The table below presents descriptive statistics for these moderators in our selection of articles.

```
##
## --------Summary descriptives table by 'usemeta2'---------
```

```
##
## ____________________________________________________________________________
##                                [ALL]    Other Coefficients Main Coefficients
##                                N=128          N=92               N=36
##
## Independent Variables:
##     K                       38 (29.7%)      29 (31.5%)         9 (25.0%)
##     N                       72 (56.2%)      51 (55.4%)        21 (58.3%)
##     logN                    18 (14.1%)      12 (13.0%)         6 (16.7%)
## Year                        2008 (5.98)    2008 (5.75)       2009 (6.50)
## Published work:
##     Yes                     104 (81.2%)     74 (80.4%)        30 (83.3%)
##     No                      24 (18.8%)      18 (19.6%)         6 (16.7%)
## Electoral system:
##     Majoritarian            58 (45.3%)      40 (43.5%)        18 (50.0%)
##     Non-Majoritarian        70 (54.7%)      52 (56.5%)        18 (50.0%)
## Estimation method:
##     OLS                     56 (43.8%)      43 (46.7%)        13 (36.1%)
##     PANEL                   57 (44.5%)      40 (43.5%)        17 (47.2%)
##     IV                       6 (4.69%)      3 (3.26%)          3 (8.33%)
##     RDD                      9 (7.03%)      6 (6.52%)          3 (8.33%)
## Sampling Aggregation Level:
##     Local                   23 (18.0%)      16 (17.4%)         7 (19.4%)
##     States                  70 (54.7%)      49 (53.3%)        21 (58.3%)
##     Countries               35 (27.3%)      27 (29.3%)         8 (22.2%)
##
```

# F Binomial Tests for Coefficient Signs

The law of 1/n poses we should have a positive influence of legislature size on expenditure. A general test of the theory could investigate whether the papers tend to find a higher frequency of positive coefficients. In statistical terms, consider a random variable representing the coefficient sign for the papers. As each sign of the paper is a Bernoulli trial, the aggregate for all papers follow a Binomial distribution, with parameters $n$ equals the number of papers, and $p$ the chance of a positive sign. The law of 1/n can be reformulated as the chance of $p > 0.5$, which facilitates the testing of the theory. The null hypothesis for such a test is that:

- $H_0$: the proportion of positive and negative signs are indistinguishable ($p = 0.5$).

As we are taking an agnostic approach, we acknowledge that either the law of 1/n ($p > 0.5$), or the reverse law of 1/n ($p < 0.5$), could be true. In this case, the alternative hypothesis is $p \neq 0.5$. To perform this test, we run binomial tests in R, using the function `binom.test(.)`.

This test has at least two advantages. First, it is robust to the design of the paper. As papers select different types of journals, countries, samples, and other characteristics, this increases the heterogeneity. This test ignores the design discrepancies and focus on the overall effect. Second, this test has the advantage of be simple and straightfoward. It requires very few assumption and has a direct statistical formulation. The disadvantage is that we can extract more information with meta-regressions.

For the number of legislators in the lower house ($N$), the results follow below.

```
##
##  Exact binomial test
##
## data:  table(aux$scoef)[2] and sum(table(aux$scoef))
## number of successes = 10, number of trials = 21, p-value = 1
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
##  0.2571306 0.7021932
## sample estimates:
## probability of success
##              0.4761905
```

Under the null hypothesis of $p = 0.5$, we find that 10 studies, out of 21, had positive sign. The chance of a distribution with $p = 0.5$ generate this sample is equal to p-value = 1. Therefore, we reject the hypothesis that $p \neq 0.5$.

For the log of the number of legislators in the lower house ($\log(N)$), the results follow below.

```
##
##  Exact binomial test
##
## data:  table(aux$scoef)[2] and sum(table(aux$scoef))
## number of successes = 5, number of trials = 6, p-value = 0.2188
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
##  0.3587654 0.9957893
## sample estimates:
## probability of success
##              0.8333333
```

Under the null hypothesis of $p = 0.5$, we find that 5 studies, out of 6, had positive sign. The chance of a distribution with $p = 0.5$ generate this sample is equal to p-value = 0.219. Therefore, we reject the hypothesis that $p \neq 0.5$.

Finally, for the number of legislators in the upper house ($K$), the results follow below.

```
##
##  Exact binomial test
##
## data:  table(aux$scoef)[2] and sum(table(aux$scoef))
## number of successes = 8, number of trials = 9, p-value = 0.03906
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
##  0.5175035 0.9971909
## sample estimates:
## probability of success
##              0.8888889
```

Under the null hypothesis of $p = 0.5$, we find that 8 studies, out of 9, had positive sign. The chance of a distribution with $p = 0.5$ generate this sample is equal to p-value = 0.039. Therefore, we accept the hypothesis that $p \neq 0.5$. This is the only test that presents evidence of an association between the legislature size and expenditure.

# G Meta-analysis

We combined the three independent ($N$, $\log(N)$, and $K$) with the levels of the three dependent variables (Expenditure Per Capita, Log of Expenditure Per Capita, Expenditure as a Percentage of the GDP). This formed a 3x3 possibility for our analysis.

## G.1 Lower House Size and Expenditure per Capita

```
##                                  SMD             95%-CI %W(random)
## Crowley (2019)               -0.3510 [-1.8112;  1.1092]        5.3
## Lee and Park (2018)          -0.8510 [-3.5851;  1.8831]        2.1
## Lee (2016)                    0.0164 [-2.5570;  2.5898]        2.4
## Kessler (2014)                0.1740 [ 0.0074;  0.3406]       13.1
## Bjedov et al. (2014)         -0.0030 [-0.0226;  0.0166]       13.4
## Baskaran (2013)               0.9740 [-0.1212;  2.0692]        7.3
## Erler (2007)                  3.9300 [ 1.6172;  6.2428]        2.8
## Chen and Malhotra (2007)     -2.0400 [-4.6468;  0.5668]        2.3
## Fiorino and Ricciuti (2007)  0.2130 [ 0.1777;  0.2483]       13.4
## Primo (2006)                 -0.8200 [-1.1924; -0.4476]       12.2
## Matsusaka (2005)             -0.9600 [-1.3128; -0.6072]       12.3
## Schaltegger and Feld (2009)  0.0010 [-0.0010;  0.0030]       13.4
##
## Number of studies combined: k = 12
##
##                           SMD             95%-CI     t p-value
## Random effects model -0.0699 [-0.6712; 0.5314] -0.26  0.8028
## Prediction interval          [-1.5540; 1.4142]
##
## Quantifying heterogeneity:
```

```
##  tau^2 = 0.3690 [0.1794; 4.7570]; tau = 0.6075 [0.4236; 2.1810];
##  I^2 = 94.7% [92.3%; 96.3%]; H = 4.34 [3.61; 5.21]
##
## Test of heterogeneity:
##       Q d.f.  p-value
##  206.92   11 < 0.0001
##
## Details on meta-analytical method:
## - Inverse variance method
## - Restricted maximum-likelihood estimator for tau^2
## - Q-profile method for confidence interval of tau^2 and tau
## - Hartung-Knapp adjustment for random effects model
```

And the forest plot:

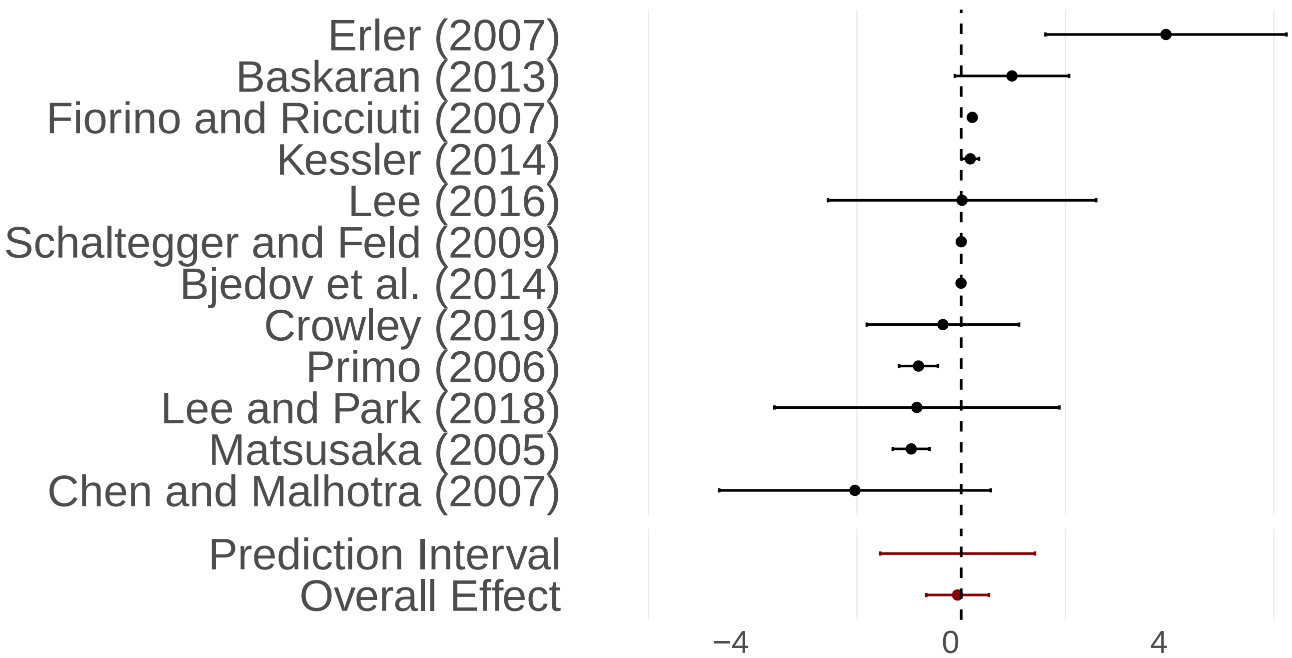



Figure 8: Effect of lower houses size (N) on Per Capita Expenditure (ExpPC)

Highlights:

1. The results are highly heterogeneous: $I^2 = $ 94.68.
2. The Random effects modem SMD estimated is $g = $ -0.07 (*SE* = 0.273).
3. The prediction interval ranges from -1.55 to 1.41. Therefore, it emcompasses zero.

### G.1.1 Electoral system subgroup analysis

The law of 1/n was created for majoritarian systems. In the theoretical section below, we explain why the argument have potential issues when applied to non-majoritarian electoral systems. We estimated a subgroup analysis using a binary electoral system.

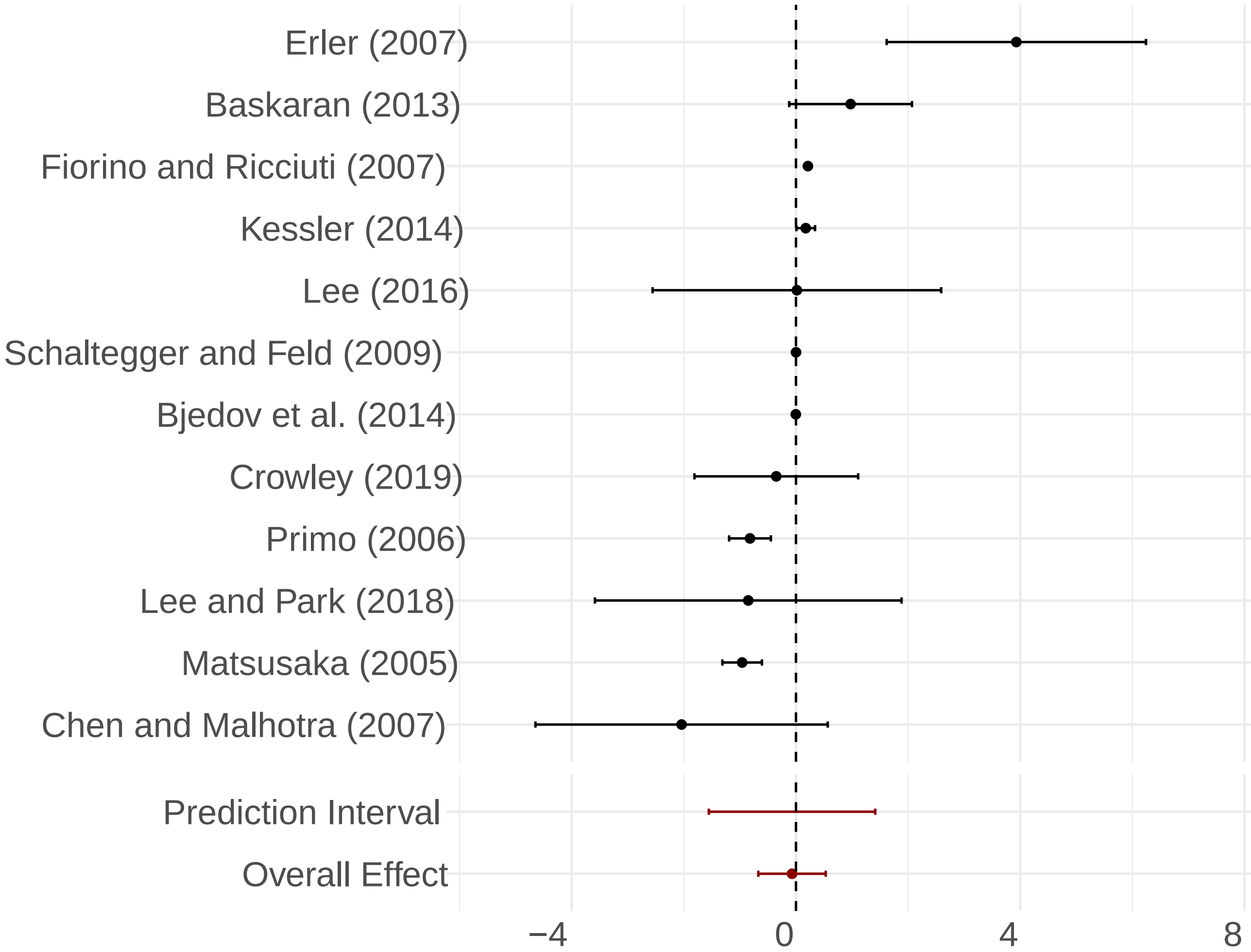



Figure 9: Subgroup Analysis of (N) x (ExpPC), controlling by electoral system

Therefore, we can see that the hypothesis that majoritarian systems produce systematic positive effects was disproved. The majoritarian systems in the sample had a random effects model estimate of -0.25, while the random effects model in the non-majoritarian subgroup fitted a value of 0.08. Both are non-significant, but they reassure us that the absense of effect is not caused by pooling multiple types of electoral systems.

## G.2 Log Lower House Size and Expenditure per Capita

There were no studies that had per capita expenditure in the dependent variable and log of lower house size in the treatment variable.

## G.3 Upper House Size and Expenditure per Capita

Now we are investigating the upper house size (K). In this model, we investigate the effect of upper house size on expenditure per capita (ExpPC).

```
# Pooling effects analysis -- ExpPC x K
aux <- dat %>%
  filter(indepvar2 == 'K',
         depvar2 == 'ExpPC')

mod <- metagen(coef, SE, data=aux,
          studlab=paste(authoryear),
          comb.fixed = FALSE,
          comb.random = TRUE,
          method.tau = "REML",
          hakn = TRUE,
          prediction=TRUE,
          sm="SMD")
mod
```

```
##                                    SMD             95%-CI %W(random)
## Crowley (2019)                  8.2100 [ 0.2702; 16.1498]       20.0
## Lee and Park (2018)            19.7400 [ 3.2645; 36.2155]       13.8
## Lee (2016)                     38.4400 [ 0.7499; 76.1301]        5.1
## Bradbury and Stephenson (2009)  0.6240 [ 0.2295;  1.0185]       23.1
## Chen and Malhotra (2007)       26.0900 [11.4883; 40.6917]       15.1
## Primo (2006)                    0.9700 [-0.4804;  2.4204]       23.0
##
## Number of studies combined: k = 6
##
##                          SMD              95%-CI    t p-value
## Random effects model 10.6134 [ -2.6210; 23.8479] 2.06  0.0943
```

```
## Prediction interval          [-21.1303; 42.3571]
##
## Quantifying heterogeneity:
##  tau^2 = 104.2124 [20.3551; >1042.1236]; tau = 10.2084 [4.5117; >32.2819];
##  I^2 = 79.4% [55.1%; 90.6%]; H = 2.20 [1.49; 3.26]
##
## Test of heterogeneity:
##      Q d.f. p-value
##  24.31    5  0.0002
##
## Details on meta-analytical method:
## - Inverse variance method
## - Restricted maximum-likelihood estimator for tau^2
## - Q-profile method for confidence interval of tau^2 and tau
## - Hartung-Knapp adjustment for random effects model
```

And the forest plot:

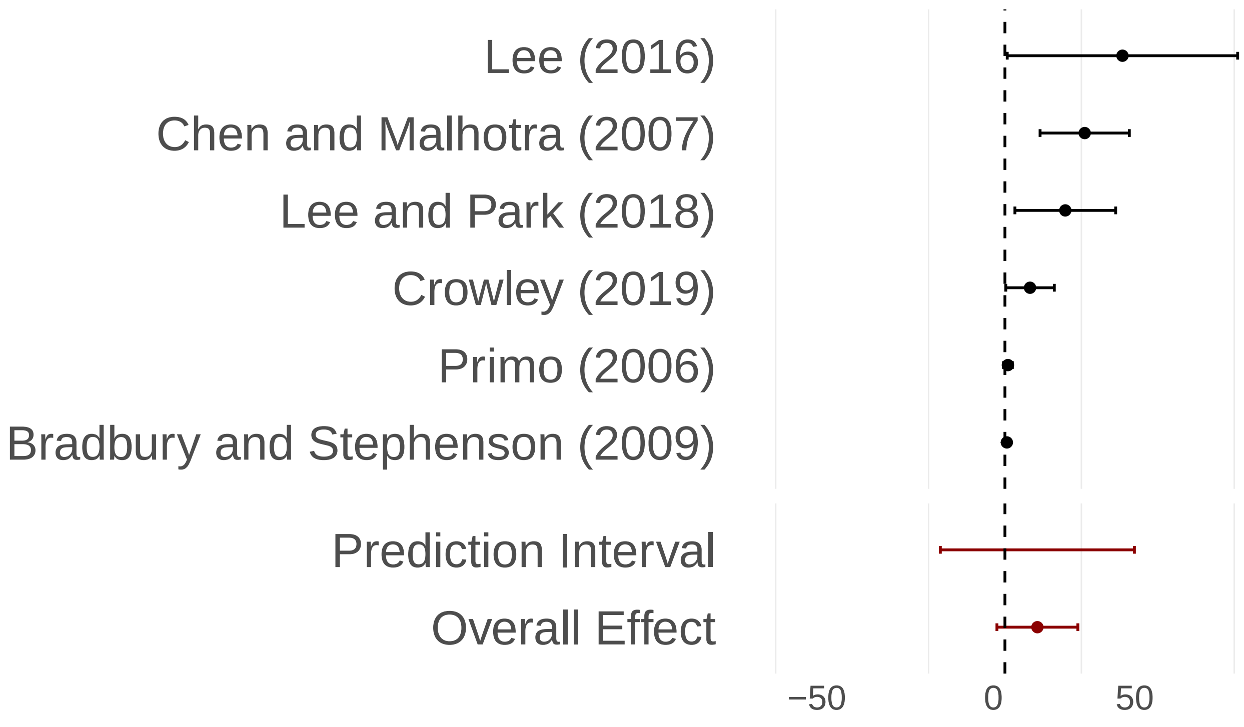



Figure 10: Effect of upper house size (K) on the per capita government expenditure (ExpPC)

Highlights:

1. The results are highly heterogeneous: $Iˆ2 = $ 79.43.
2. The Random effects modem SMD estimated is $g = $ 10.61 (*SE* = 5.148).
3. The prediction interval ranges from -21.13 to 42.36. Therefore, it emcompasses zero.

## G.4 Lower House Size and Log Expenditure Per Capita

This model estimates the Log of Per Capita Expenditure as the dependent variable, and the number of lower house legislators as the treatment variable.

```
# Pooling effects analysis -- logExpPC x N
aux <- dat %>%
  filter(indepvar2 == 'N',
         depvar2 == 'logExpPC')

mod <- metagen(
  coef, SE, data=aux,
  studlab=paste(authoryear),
  comb.fixed = FALSE,
  comb.random = TRUE,
  method.tau = "REML",
  hakn = TRUE,
  prediction = TRUE,
  sm="SMD"
  )

mod
```

```
##                                SMD             95%-CI %W(random)
## Lewis (2019)               -0.1740 [-0.2450; -0.1030]       24.3
## Höhmann (2017)             -0.0300 [-0.0496; -0.0104]       26.6
## Drew and Dollery (2017)     0.0770 [ 0.0221;  0.1319]       25.3
## Pettersson-Lidbom (2012) -0.1590 [-0.2394; -0.0786]       23.7
##
## Number of studies combined: k = 4
##
##                           SMD             95%-CI      t p-value
```

```
## Random effects model -0.0686 [-0.2560; 0.1188] -1.17  0.3282
## Prediction interval          [-0.6179; 0.4807]
##
## Quantifying heterogeneity:
##  tau^2 = 0.0128 [0.0034; 0.1933]; tau = 0.1133 [0.0584; 0.4396];
##  I^2 = 92.5% [84.1%; 96.5%]; H = 3.66 [2.51; 5.34]
##
## Test of heterogeneity:
##      Q d.f.  p-value
##  40.11    3 < 0.0001
##
## Details on meta-analytical method:
## - Inverse variance method
## - Restricted maximum-likelihood estimator for tau^2
## - Q-profile method for confidence interval of tau^2 and tau
## - Hartung-Knapp adjustment for random effects model
```

And the forest plot:

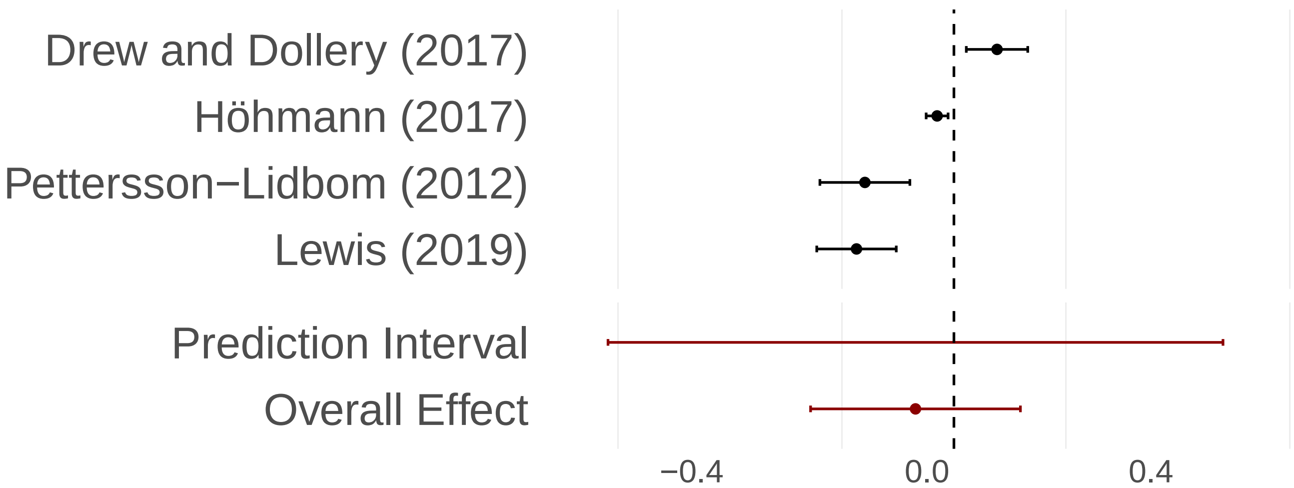



Figure 11: Effect of lower houses size (N) on log of per capita expenditure (logExpPC)

Highlights:

1. The results are highly heterogeneous: $Iˆ2 = $ 92.52.
2. The Random effects modem SMD estimated is $g = $ -0.07 (*SE* = 0.059).

3. The prediction interval ranges from -0.62 to 0.48. Therefore, it emcompasses zero.

## G.5 Log of Lower House Size and Log of Expenditure Per Capita

In this specification, we study the log of per capita expenditure (logExpPC) as a function of the log of lower house size (logN).

```
# Pooling effects analysis -- logExpPC x logN
aux <- dat %>%
  filter(indepvar2 == 'logN',
         depvar2 == 'logExpPC')

mod <- metagen(coef, SE, data=aux,
          studlab=paste(authoryear),
          comb.fixed = FALSE,
          comb.random = TRUE,
          method.tau = "REML",
          hakn = TRUE,
          prediction=TRUE,
          sm="SMD")
mod
```

```
##                       SMD           95%-CI %W(random)
## MacDonald (2008) 0.1360 [0.0447; 0.2273]       31.9
## Baqir (2002)     0.1127 [0.0396; 0.1858]       34.2
## Baqir (1999)     0.3020 [0.2269; 0.3771]       33.9
##
## Number of studies combined: k = 3
##
##                           SMD             95%-CI    t p-value
## Random effects model 0.1844 [-0.0738; 0.4425] 3.07  0.0916
## Prediction interval         [-1.2580; 1.6267]
##
## Quantifying heterogeneity:
```

```
##  tau^2 = 0.0093 [0.0014; 0.4193]; tau = 0.0964 [0.0372; 0.6476];
##  I^2 = 85.9% [59.0%; 95.2%]; H = 2.66 [1.56; 4.54]
##
## Test of heterogeneity:
##      Q d.f. p-value
##  14.18    2  0.0008
##
## Details on meta-analytical method:
## - Inverse variance method
## - Restricted maximum-likelihood estimator for tau^2
## - Q-profile method for confidence interval of tau^2 and tau
## - Hartung-Knapp adjustment for random effects model
```

And the forest plot:

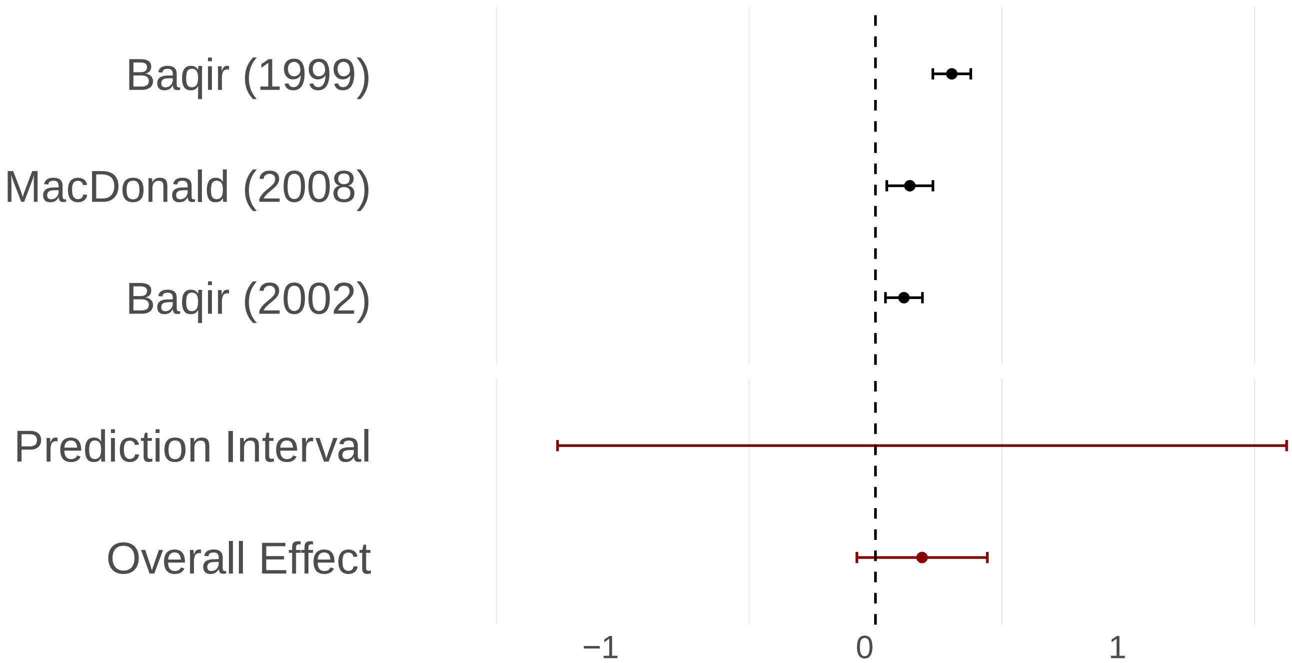



Figure 12: Effect of log lower houses size (logN) on the log of per capita government expenditure (logExpPC)

Highlights:

1. The results are highly heterogeneous: $I^2 = $ 85.9.
2. The Random effects modem SMD estimated is $g = $ 0.18 (*SE* = 0.06). **This model is significant at the 10% confidence level.**

3. The prediction interval ranges from -1.26 to 1.63. Therefore, it emcompasses zero.

## G.6 Log of Upper House Size and Log of Expenditure Per Capita

No studies related the log of per capita expenditure with the size of upper house (K).

## G.7 Lower House Size and Expenditure as Percentage of GDP

This model fits the random effects for the percentage of GDP as public expenditure as the main outcome, and the size of lower house as the main treatment variable.

```r
# Pooling effects analysis -- PCTGDP x N
aux <- dat %>%
  filter(indepvar2 == 'N',
         depvar2 == 'PCTGDP')

mod <- metagen(coef, SE, data=aux,
          studlab=paste(authoryear),
          comb.fixed = FALSE,
          comb.random = TRUE,
          method.tau = "REML",
          hakn = TRUE,
          prediction=TRUE,
          sm="SMD")
mod
```

```
##                                SMD             95%-CI %W(random)
## Bjedov et al. (2014)       -0.0040 [-0.0432;  0.0352]       15.1
## Maldonado (2013)           -0.0609 [-0.0838; -0.0380]       19.5
## Mukherjee (2003)            0.0030 [ 0.0010;  0.0050]       23.0
## Bradbury and Crain (2001)  0.0036 [ 0.0008;  0.0065]       23.0
## Ricciuti (2004)             0.0140 [-0.0095;  0.0375]       19.4
##
## Number of studies combined: k = 5
##
```

```
##                              SMD             95%-CI      t p-value
## Random effects model -0.0083 [-0.0450; 0.0285] -0.62  0.5667
## Prediction interval          [-0.1054; 0.0889]
##
## Quantifying heterogeneity:
##  tau^2 = 0.0008 [0.0002; 0.0072]; tau = 0.0275 [0.0129; 0.0849];
##  I^2 = 87.1% [72.2%; 94.0%]; H = 2.78 [1.90; 4.08]
##
## Test of heterogeneity:
##      Q d.f.  p-value
##  30.97    4 < 0.0001
##
## Details on meta-analytical method:
## - Inverse variance method
## - Restricted maximum-likelihood estimator for tau^2
## - Q-profile method for confidence interval of tau^2 and tau
## - Hartung-Knapp adjustment for random effects model
```

And the forest plot:

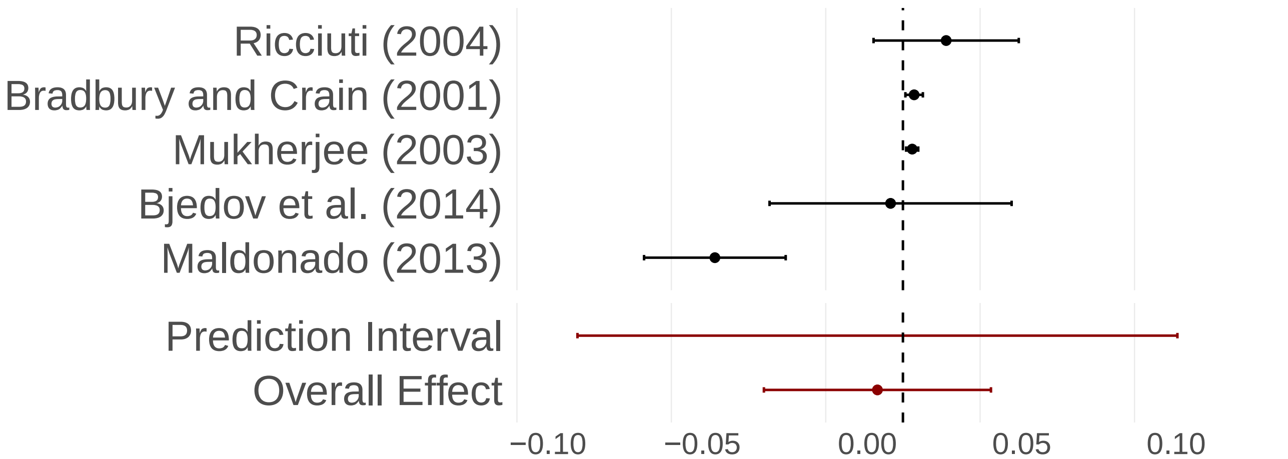



Figure 13: Effect of lower houses size (N) on percentage of public expenditure GDP (PCTGDP)

Highlights:

1. The results are highly heterogeneous: $I^2 = $ 87.08.
2. The Random effects modem SMD estimated is $g = $ -0.01 ($SE$ = 0.013).

3. The prediction interval ranges from -0.11 to 0.09. Therefore, it emcompasses zero.

## G.8 Log Lower House Size and Expenditure as Percentage of GDP

This meta-regression investigates the percentage of GDP as public expenditure as the dependent variable and the log lower house size (logN) as the treatment variable.

```
# Pooling effects analysis -- PCTGDP x logN
aux <- dat %>%
  filter(indepvar2 == 'logN',
         depvar2 == 'PCTGDP')

mod <- metagen(
  coef, SE, data=aux,
  studlab=paste(authoryear),
  comb.fixed = FALSE,
  comb.random = TRUE,
  method.tau = "REML",
  hakn = TRUE,
  prediction=TRUE,
  sm="SMD"
  )

mod
```

```
##                         SMD            95%-CI %W(random)
## Baqir (1999)         2.0660 [ 1.4887; 2.6433]       40.8
## Lledo (2003)        -4.6900 [-9.9427; 0.5627]       17.7
## Stein et al. (1998)  0.0109 [-0.0171; 0.0389]       41.5
##
## Number of studies combined: k = 3
##
##                         SMD              95%-CI    t p-value
## Random effects model 0.0203 [ -7.1961;  7.2367] 0.01  0.9914
```

```
## Prediction interval         [-36.2058; 36.2465]
##
## Quantifying heterogeneity:
##  tau^2 = 5.3156 [0.5756; >100.0000]; tau = 2.3056 [0.7587; >10.0000];
##  I^2 = 96.1% [91.8%; 98.2%]; H = 5.08 [3.48; 7.42]
##
## Test of heterogeneity:
##      Q d.f.  p-value
##  51.65    2 < 0.0001
##
## Details on meta-analytical method:
## - Inverse variance method
## - Restricted maximum-likelihood estimator for tau^2
## - Q-profile method for confidence interval of tau^2 and tau
## - Hartung-Knapp adjustment for random effects model
```

And the forest plot:

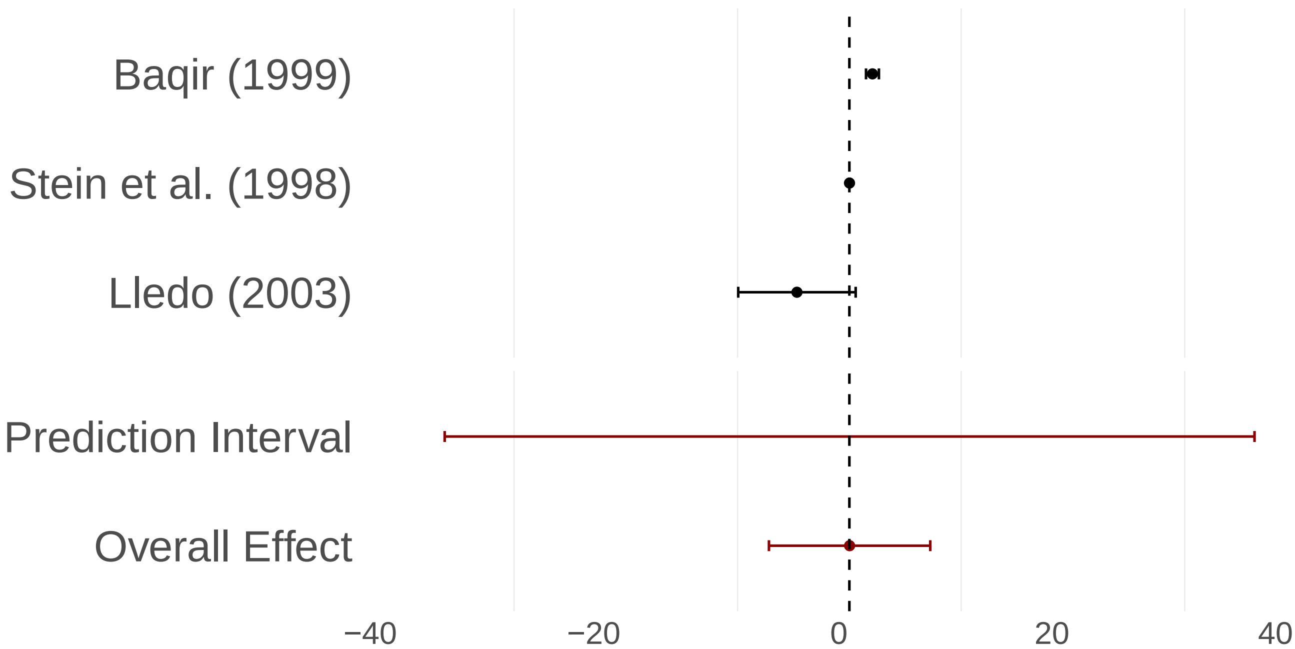



Figure 14: Effect of log lower houses size (logN) on the GDP share of public expenditure (PCTGDP)

Highlights:

1. The results are highly heterogeneous: $I^2 = $ 96.13.

2. The Random effects modem SMD estimated is $g = $ 0.02 ($SE = 1.677$).
3. The prediction interval ranges from -36.21 to 36.25. Therefore, it emcompasses zero.

## G.9 Upper House Size and Expenditure as Percentage of GDP

This model looks into the effect of upper house size (K) on the public expenditure share of the GDP (PCTGDP).

```r
# Pooling effects analysis -- PCTGDP x K
aux <- dat %>%
  filter(indepvar2 == 'K',
         depvar2 == 'PCTGDP')

mod <- metagen(coef, SE, data=aux,
          studlab=paste(authoryear),
          comb.fixed = FALSE,
          comb.random = TRUE,
          method.tau = "REML",
          hakn = TRUE,
          prediction=TRUE,
          sm="SMD")
mod
```

```
##                                SMD             95%-CI  %W(random)
## Maldonado (2012)           -0.0400 [-0.0659; -0.0141]        31.3
## Bradbury and Crain (2001)   0.0126 [ 0.0010;  0.0243]        36.4
## Ricciuti (2004)             0.0160 [-0.0075;  0.0395]        32.3
##
## Number of studies combined: k = 3
##
##                           SMD            95%-CI      t  p-value
## Random effects model -0.0027 [-0.0793; 0.0738] -0.15   0.8915
## Prediction interval          [-0.4284; 0.4229]
##
## Quantifying heterogeneity:
```

```
##  tau^2 = 0.0008 [0.0001; 0.0388]; tau = 0.0284 [0.0101; 0.1970];
##  I^2 = 85.8% [58.6%; 95.1%]; H = 2.65 [1.55; 4.53]
##
## Test of heterogeneity:
##      Q d.f. p-value
##  14.07    2  0.0009
##
## Details on meta-analytical method:
## - Inverse variance method
## - Restricted maximum-likelihood estimator for tau^2
## - Q-profile method for confidence interval of tau^2 and tau
## - Hartung-Knapp adjustment for random effects model
```

And the forest plot:

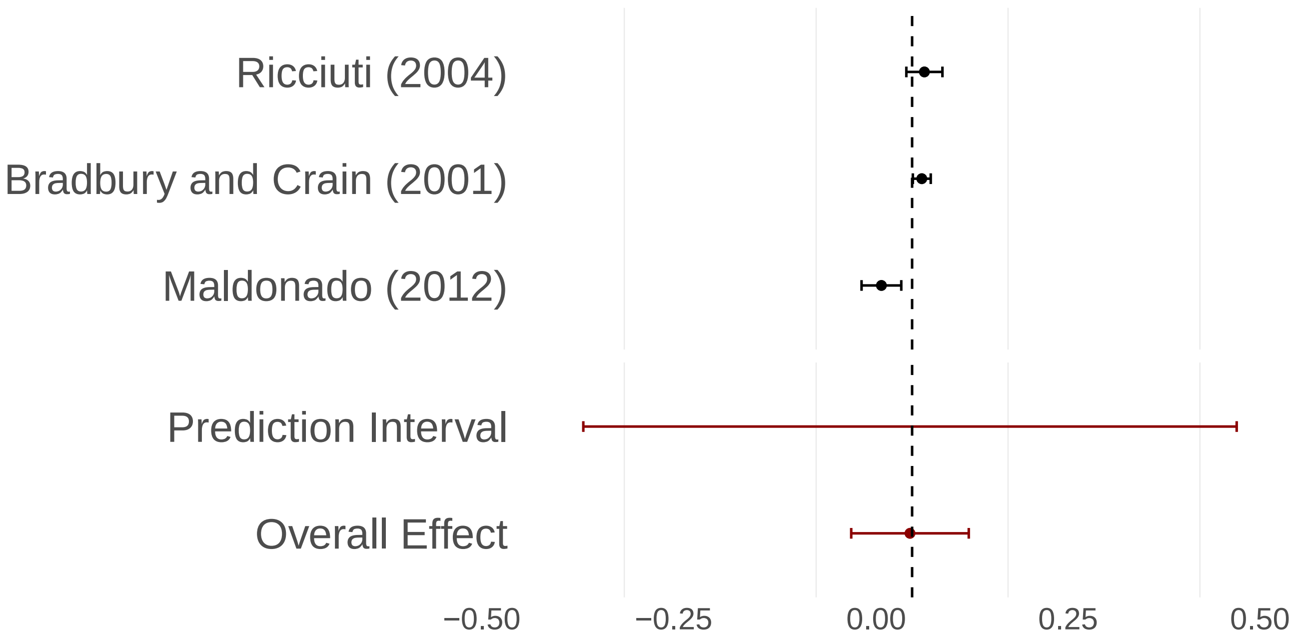



Figure 15: Effect of upper house size (K) on the public expenditure share of the GDP (PCTGDP)

Highlights:

1. The results are highly heterogeneous: $I^2 = $ 85.79.
2. The Random effects modem SMD estimated is $g = $ 0 ($SE$ = 0.018).
3. The prediction interval ranges from -0.43 to 0.42. Therefore, it emcompasses zero.

# H Meta-Analysis (all coefficients)

## H.1 Lower House Size and Expenditure Per Capita

```
##                                   SMD              95%-CI %W(random)
## Crowley (2019)                -0.3510 [-1.8112;  1.1092]        2.0
## Crowley (2019)                 5.9750 [ 0.7889; 11.1611]        0.3
## Crowley (2019)                 7.6580 [-0.0290; 15.3450]        0.2
## Lee and Park (2018)           -0.8510 [-3.5851;  1.8831]        0.9
## Lee and Park (2018)           -1.6890 [-3.0551; -0.3229]        2.1
## Lee and Park (2018)            7.6320 [ 3.1064; 12.1576]        0.4
## Lee (2016)                     0.0164 [-2.5570;  2.5898]        1.0
## Kessler (2014)                 0.1740 [ 0.0074;  0.3406]        3.6
## Kessler (2014)                 0.2230 [ 0.1211;  0.3249]        3.6
## Kessler (2014)                 0.2150 [ 0.0954;  0.3346]        3.6
## Kessler (2014)                 0.1580 [ 0.0522;  0.2638]        3.6
## Bjedov et al. (2014)          -0.0030 [-0.0226;  0.0166]        3.6
## Bjedov et al. (2014)          -0.0060 [-0.0256;  0.0136]        3.6
## Baskaran (2013)                0.9740 [-0.1212;  2.0692]        2.5
## Erler (2007)                   3.9300 [ 1.6172;  6.2428]        1.2
## Chen and Malhotra (2007)      -2.0400 [-4.6468;  0.5668]        1.0
## Chen and Malhotra (2007)      -1.4000 [-2.6544; -0.1456]        2.3
## Fiorino and Ricciuti (2007)  0.2130 [ 0.1777;  0.2483]        3.6
## Fiorino and Ricciuti (2007)  0.2290 [ 0.1565;  0.3015]        3.6
## Fiorino and Ricciuti (2007)  0.4550 [ 0.3805;  0.5295]        3.6
## Fiorino and Ricciuti (2007)  0.4110 [ 0.3150;  0.5070]        3.6
## Fiorino and Ricciuti (2007)  0.2260 [ 0.1221;  0.3299]        3.6
## Fiorino and Ricciuti (2007)  0.2130 [-0.4083;  0.8343]        3.1
## Fiorino and Ricciuti (2007)  0.1850 [-0.4128;  0.7828]        3.2
## Fiorino and Ricciuti (2007)  0.2350 [-0.4235;  0.8935]        3.1
## Fiorino and Ricciuti (2007)  0.3740 [ 0.2486;  0.4994]        3.6
```

```
## Fiorino and Ricciuti (2007)  0.8110 [ 0.4562;  1.1658]        3.4
## Fiorino and Ricciuti (2007)  0.7950 [ 0.4500;  1.1400]        3.5
## Fiorino and Ricciuti (2007)  0.8490 [ 0.3825;  1.3155]        3.3
## Primo (2006)                -0.8200 [-1.1924; -0.4476]        3.4
## Primo (2006)                -1.7000 [-2.3076; -1.0924]        3.2
## Primo (2006)                -2.3700 [-3.0952; -1.6448]        3.0
## Primo (2006)                -2.0300 [-2.7552; -1.3048]        3.0
## Matsusaka (2005)            -0.9600 [-1.3128; -0.6072]        3.4
## Schaltegger and Feld (2009)  0.0010 [-0.0010;  0.0030]        3.6
## Schaltegger and Feld (2009) -0.0010 [-0.0030;  0.0010]        3.6
##
## Number of studies combined: k = 36
##
##                            SMD            95%-CI     t p-value
## Random effects model -0.0169 [-0.4166; 0.3829] -0.09  0.9322
## Prediction interval          [-1.7588; 1.7250]
##
## Quantifying heterogeneity:
##  tau^2 = 0.6959 [0.7202; 4.3553]; tau = 0.8342 [0.8486; 2.0869];
##  I^2 = 95.3% [94.2%; 96.1%]; H = 4.60 [4.16; 5.08]
##
## Test of heterogeneity:
##       Q d.f.  p-value
##  739.53   35 < 0.0001
##
## Details on meta-analytical method:
## - Inverse variance method
## - Restricted maximum-likelihood estimator for tau^2
## - Q-profile method for confidence interval of tau^2 and tau
## - Hartung-Knapp adjustment for random effects model
```

And the forest plot:

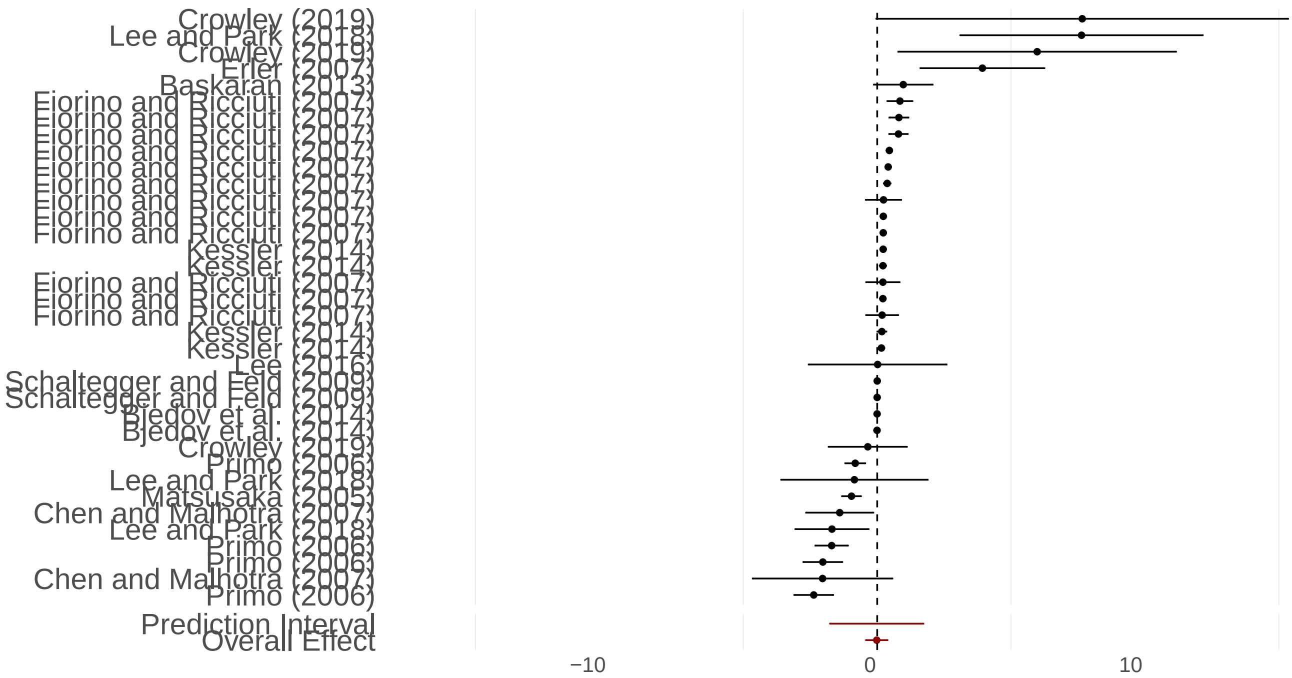



Figure 16: Effect of lower houses size (N) on Per Capita Expenditure (ExpPC)

Highlights:

1. The results are highly heterogeneous: $Iˆ2 = $ 95.27.
2. The Random effects modem SMD estimated is $g = $ -0.02 (*SE* = 0.197).
3. The prediction interval ranges from -1.76 to 1.73. Therefore, it emcompasses zero.

### H.1.1 Electoral system subgroup analysis

The law of 1/n was created for majoritarian systems. In the theoretical section below, we explain why the argument have potential issues when applied to non-majoritarian electoral systems. We estimated a subgroup analysis using a binary electoral system.

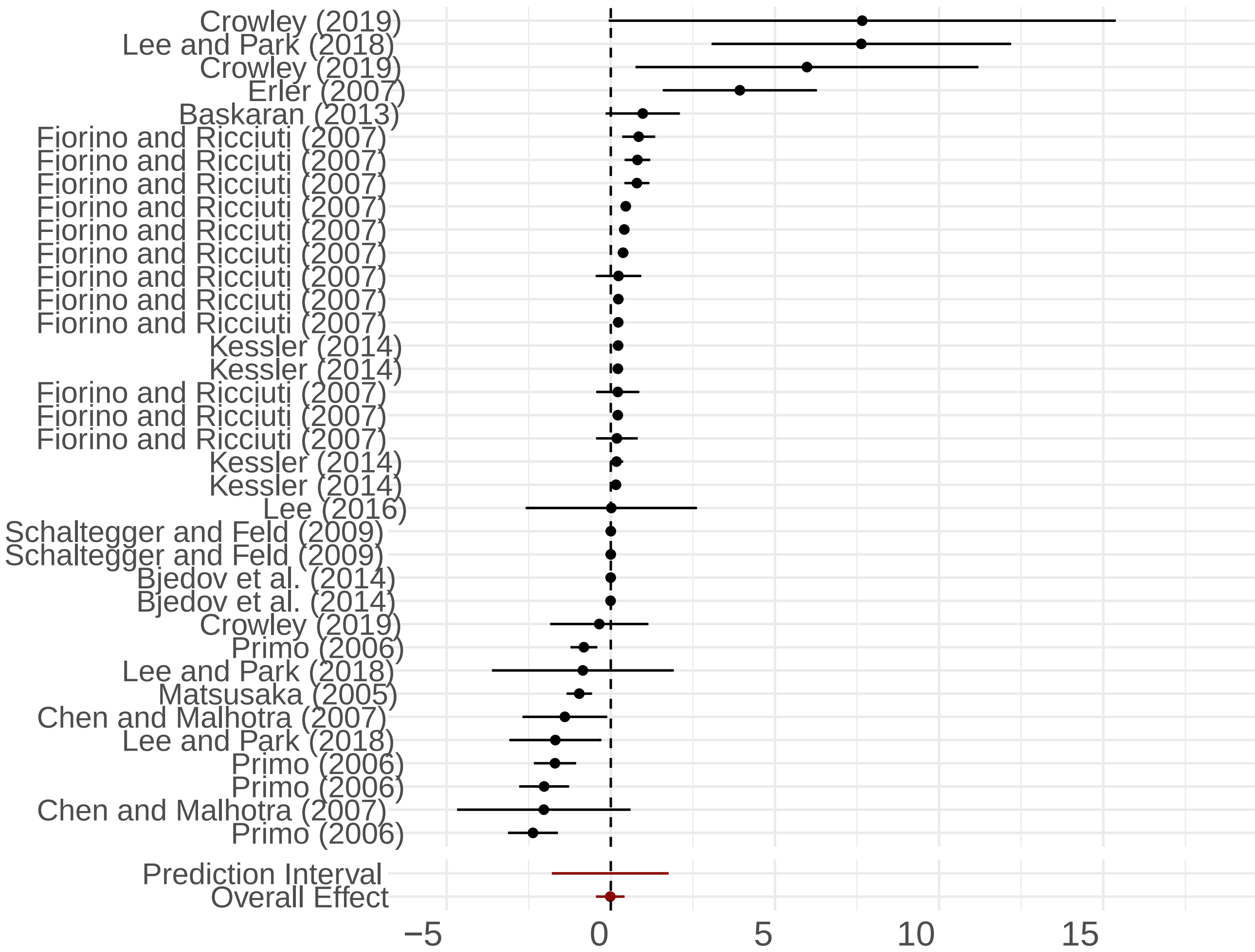



Figure 17: Subgroup Analysis of (N) x (ExpPC), controlling by electoral system

Therefore, we can see that the hypothesis that majoritarian systems produce systematic positive effects was disproved. The majoritarian systems in the sample had a random effects model estimate of -0.25, while the random effects model in the non-majoritarian subgroup fitted a value of 0.08. Both are non-significant, but they reassure us that the absense of effect is not caused by pooling multiple types of electoral systems.

## H.2 Log of Lower House Size and Expenditure Per Capita

There were no studies that had per capita expenditure in the dependent variable and log of lower house size in the treatment variable.

## H.3 Upper House Size and Expenditure Per Capita

Now we are investigating the upper house size (K). In this model, we investigate the effect of upper house size on expenditure per capita (ExpPC).

```
# Pooling effects analysis -- ExpPC x K
aux <- fulldat %>%
  filter(indepvar2 == 'K',
         depvar2 == 'ExpPC')

mod <- metagen(coef, SE, data=aux,
          studlab=paste(authoryear),
          comb.fixed = FALSE,
          comb.random = TRUE,
          method.tau = "REML",
          hakn = TRUE,
          prediction=TRUE,
          sm="SMD")
mod
```

```
##                                  SMD              95%-CI %W(random)
## Crowley (2019)                8.2100 [  0.2702; 16.1498]        4.8
## Crowley (2019)                8.4230 [-27.1895; 44.0355]        0.4
## Crowley (2019)                9.5940 [  2.1383; 17.0497]        5.1
## Lee and Park (2018)          19.7400 [  3.2645; 36.2155]        1.7
## Lee and Park (2018)          10.0600 [  2.2887; 17.8313]        4.9
## Lee and Park (2018)           9.0620 [-30.8821; 49.0061]        0.3
## Lee (2016)                   38.4400 [  0.7499; 76.1301]        0.4
```

```
## Lee (2016)                     37.8500 [  3.0214; 72.6786]        0.4
## Lee (2016)                     25.6100 [ -0.8103; 52.0303]        0.8
## Lee (2016)                      5.9960 [-19.6011; 31.5931]        0.8
## Lee (2016)                     25.5600 [ -0.8799; 51.9999]        0.8
## Lee (2016)                      4.6930 [-19.5126; 28.8986]        0.9
## Bradbury and Stephenson (2009)  0.6240 [  0.2295;  1.0185]       10.0
## Chen and Malhotra (2007)       26.0900 [ 11.4883; 40.6917]        2.1
## Chen and Malhotra (2007)        8.3000 [  3.6941; 12.9059]        7.3
## Chen and Malhotra (2007)        5.1400 [  0.1813; 10.0987]        7.0
## Chen and Malhotra (2007)        4.7800 [ -0.9039; 10.4639]        6.4
## Chen and Malhotra (2007)       20.3800 [  7.6990; 33.0610]        2.6
## Chen and Malhotra (2007)        4.8700 [  1.2833;  8.4567]        8.2
## Chen and Malhotra (2007)       26.7500 [  0.8589; 52.6411]        0.8
## Primo (2006)                    0.9700 [ -0.4804;  2.4204]        9.7
## Primo (2006)                    5.9000 [  2.6857;  9.1143]        8.5
## Primo (2006)                    5.7500 [  2.3593;  9.1407]        8.4
## Primo (2006)                    6.9600 [  2.6089; 11.3111]        7.6
##
## Number of studies combined: k = 24
##
##                          SMD             95%-CI    t  p-value
## Random effects model 7.2162 [ 4.4400;  9.9925] 5.38 < 0.0001
## Prediction interval         [-1.2217; 15.6542]
##
## Quantifying heterogeneity:
##  tau^2 = 14.7532 [5.4141; 111.2304]; tau = 3.8410 [2.3268; 10.5466];
##  I^2 = 77.7% [67.3%; 84.8%]; H = 2.12 [1.75; 2.57]
##
## Test of heterogeneity:
##      Q d.f.  p-value
```

```
##  103.34   23 < 0.0001
##
## Details on meta-analytical method:
## - Inverse variance method
## - Restricted maximum-likelihood estimator for tau^2
## - Q-profile method for confidence interval of tau^2 and tau
## - Hartung-Knapp adjustment for random effects model
```

And the forest plot:

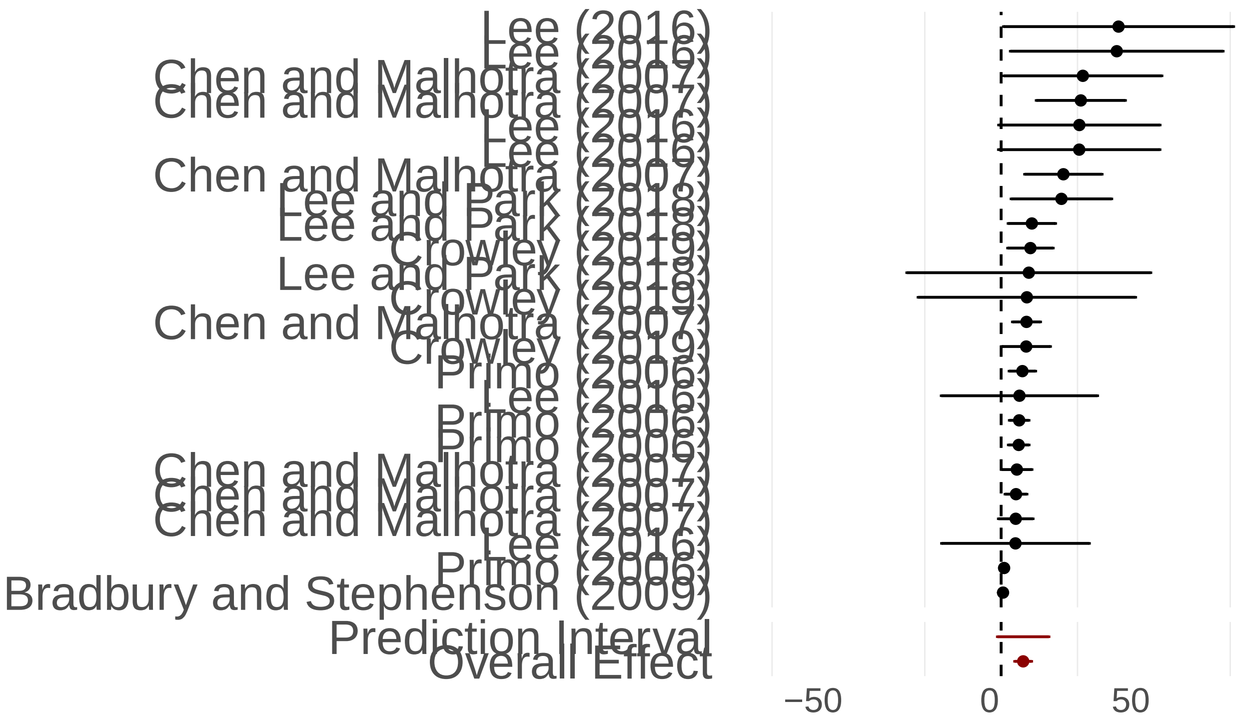



Figure 18: Effect of upper house size (K) on the per capita government expenditure (ExpPC)

Highlights:

1. The results are highly heterogeneous: $I^2 = $ 77.74.
2. The Random effects modem SMD estimated is $g = $ 7.22 ($SE$ = 1.342).
3. The prediction interval ranges from -1.22 to 15.65. Therefore, it emcompasses zero.

## H.4 Lower House Size and Log of Expenditure Per Capita

This model estimates the Log of Per Capita Expenditure as the dependent variable, and the number of lower house legislators as the treatment variable.

```
# Pooling effects analysis -- logExpPC x N
aux <- fulldat %>%
  filter(indepvar2 == 'N',
         depvar2 == 'logExpPC')

mod <- metagen(coef, SE, data=aux,
          studlab=paste(authoryear),
          comb.fixed = FALSE,
          comb.random = TRUE,
          method.tau = "REML",
          hakn = TRUE,
          prediction=TRUE,
          sm="SMD")
mod
```

```
##                                SMD             95%-CI %W(random)
## Lewis (2019)               -0.1740 [-0.2450; -0.1030]        6.6
## Höhmann (2017)             -0.0300 [-0.0496; -0.0104]        7.1
## Höhmann (2017)             -0.0300 [-0.0496; -0.0104]        7.1
## Höhmann (2017)             -0.0400 [-0.0792; -0.0008]        7.0
## Drew and Dollery (2017)     0.0770 [ 0.0221;  0.1319]        6.8
## Drew and Dollery (2017)     0.0310 [-0.0121;  0.0741]        6.9
## Pettersson-Lidbom (2012) -0.1590 [-0.2394; -0.0786]        6.4
## Pettersson-Lidbom (2012) -0.1470 [-0.2274; -0.0666]        6.4
## Pettersson-Lidbom (2012) -0.0900 [-0.1645; -0.0155]        6.5
## Pettersson-Lidbom (2012) -0.0810 [-0.1574; -0.0046]        6.5
## Pettersson-Lidbom (2012) -0.0880 [-0.1625; -0.0135]        6.5
```

```
## Pettersson-Lidbom (2012)  0.2100 [ 0.1649;  0.2551]        6.9
## Pettersson-Lidbom (2012)  0.1570 [ 0.0845;  0.2295]        6.5
## Pettersson-Lidbom (2012) -0.1990 [-0.2774; -0.1206]        6.4
## Pettersson-Lidbom (2012) -0.1690 [-0.2494; -0.0886]        6.4
##
## Number of studies combined: k = 15
##
##                           SMD            95%-CI     t p-value
## Random effects model -0.0463 [-0.1142; 0.0216] -1.46  0.1655
## Prediction interval          [-0.3105; 0.2178]
##
## Quantifying heterogeneity:
##  tau^2 = 0.0139 [0.0070; 0.0364]; tau = 0.1181 [0.0836; 0.1908];
##  I^2 = 93.8% [91.2%; 95.6%]; H = 4.00 [3.38; 4.75]
##
## Test of heterogeneity:
##       Q d.f.  p-value
##  224.56   14 < 0.0001
##
## Details on meta-analytical method:
## - Inverse variance method
## - Restricted maximum-likelihood estimator for tau^2
## - Q-profile method for confidence interval of tau^2 and tau
## - Hartung-Knapp adjustment for random effects model
```

And the forest plot:

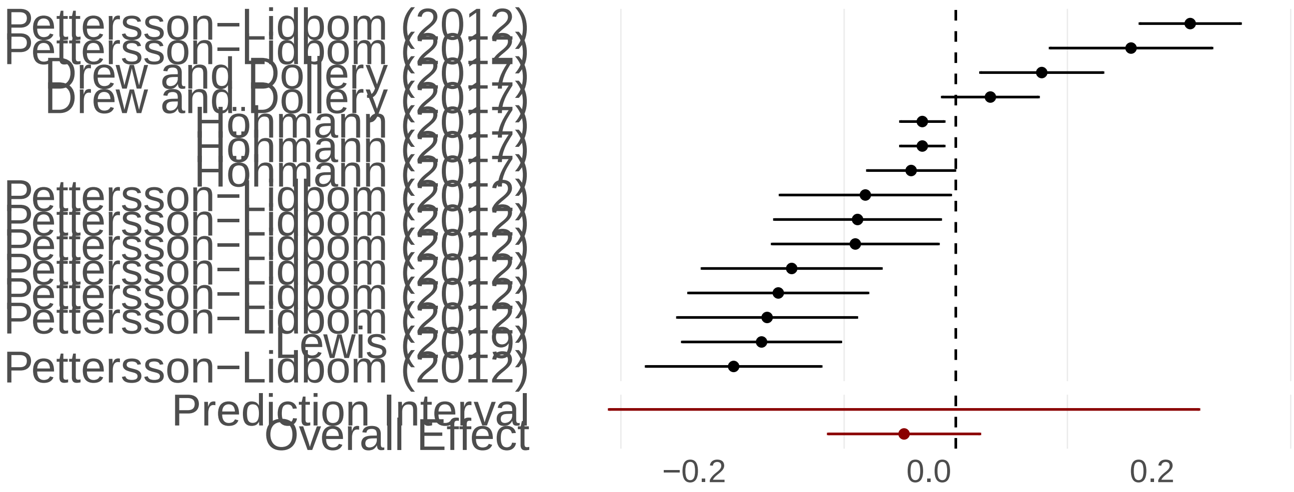



Figure 19: Effect of lower houses size (N) on log of per capita expenditure (logExpPC)

Highlights:

1. The results are highly heterogeneous: $I^2 = $ 93.77.
2. The Random effects modem SMD estimated is $g = $ -0.05 ($SE$ = 0.032).
3. The prediction interval ranges from -0.31 to 0.22. Therefore, it emcompasses zero.

## H.5 Log of Lower House Size and Log of Expenditure Per Capita

In this specification, we study the log of per capita expenditure (logExpPC) as a function of the log of lower house size (logN).

```
# Pooling effects analysis -- logExpPC x logN
aux <- fulldat %>%
  filter(indepvar2 == 'logN',
         depvar2 == 'logExpPC')

mod <- metagen(coef, SE, data=aux,
          studlab=paste(authoryear),
          comb.fixed = FALSE,
          comb.random = TRUE,
          method.tau = "REML",
          hakn = TRUE,
          prediction=TRUE,
          sm="SMD")
mod
```

```
##                       SMD            95%-CI %W(random)
## MacDonald (2008) 0.1360 [0.0447; 0.2273]        7.9
## MacDonald (2008) 0.2319 [0.1322; 0.3316]        7.4
## MacDonald (2008) 0.1443 [0.0471; 0.2415]        7.6
## MacDonald (2008) 0.1594 [0.0667; 0.2521]        7.8
## MacDonald (2008) 0.2259 [0.1163; 0.3355]        6.9
## Baqir (2002)     0.1127 [0.0396; 0.1858]        9.1
## Baqir (2002)     0.2760 [0.2007; 0.3513]        8.9
## Baqir (2002)     0.3021 [0.2270; 0.3772]        8.9
## Baqir (2002)     0.3203 [0.2450; 0.3956]        8.9
## Baqir (1999)     0.3020 [0.2269; 0.3771]        8.9
## Baqir (1999)     0.2760 [0.2007; 0.3513]        8.9
```

```
## Baqir (1999)     0.2950 [0.2165; 0.3735]        8.7
##
## Number of studies combined: k = 12
##
##                          SMD           95%-CI     t  p-value
## Random effects model 0.2346 [0.1864; 0.2828] 10.71 < 0.0001
## Prediction interval         [0.0848; 0.3844]
##
## Quantifying heterogeneity:
##  tau^2 = 0.0040 [0.0011; 0.0145]; tau = 0.0636 [0.0335; 0.1203];
##  I^2 = 70.0% [45.6%; 83.4%]; H = 1.82 [1.36; 2.45]
##
## Test of heterogeneity:
##      Q d.f. p-value
##  36.62   11  0.0001
##
## Details on meta-analytical method:
## - Inverse variance method
## - Restricted maximum-likelihood estimator for tau^2
## - Q-profile method for confidence interval of tau^2 and tau
## - Hartung-Knapp adjustment for random effects model
```

And the forest plot:

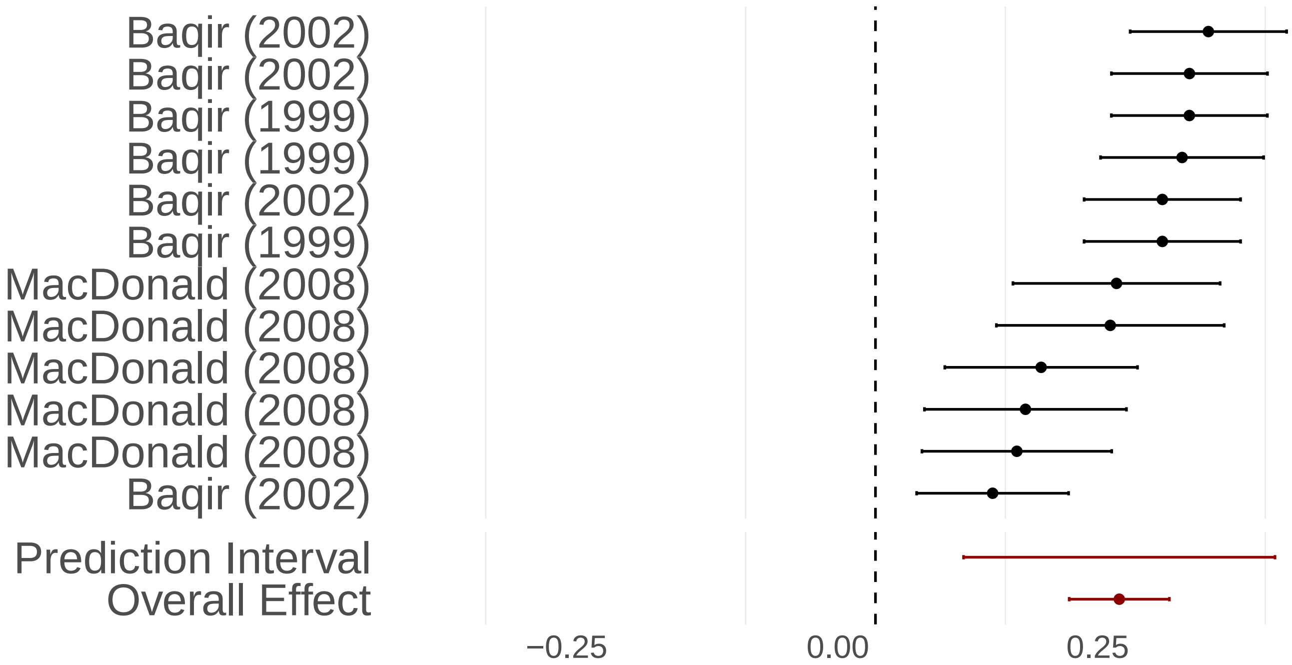



Figure 20: Effect of log lower houses size (logN) on the log of per capita government expenditure (logExpPC)

Highlights:

1. The results are highly heterogeneous: $I^2 = $ 69.96.
2. The Random effects modem SMD estimated is $g = $ 0.23 ($SE$ = 0.022). **This model is significant at the 10% confidence level.**
3. The prediction interval ranges from 0.08 to 0.38. Therefore, it does not emcompasses zero.

## H.6 Upper House Size and Log of Expenditure Per Capita

No studies related the log of per capita expenditure with the size of upper house (K).

## H.7 Lower House Size and Expenditure as Percentage of GDP

This model fits the random effects for the percentage of GDP as public expenditure as the main outcome, and the size of lower house as the main treatment variable.

```
# Pooling effects analysis -- PCTGDP x N
aux <- fulldat %>%
  filter(indepvar2 == 'N',
         depvar2 == 'PCTGDP')

mod <- metagen(coef, SE, data=aux,
          studlab=paste(authoryear),
          comb.fixed = FALSE,
          comb.random = TRUE,
          method.tau = "REML",
          hakn = TRUE,
          prediction=TRUE,
          sm="SMD")
mod
```

```
##                                SMD             95%-CI  %W(random)
## Bjedov et al. (2014)       -0.0040 [-0.0432;  0.0352]        2.1
## Bjedov et al. (2014)       -0.0080 [-0.0472;  0.0312]        2.1
## Maldonado (2013)           -0.0609 [-0.0838; -0.0380]        3.6
## Mukherjee (2003)            0.0030 [ 0.0010;  0.0050]        5.6
## Mukherjee (2003)            0.0090 [ 0.0051;  0.0129]        5.5
## Mukherjee (2003)            0.0110 [ 0.0051;  0.0169]        5.4
## Mukherjee (2003)            0.0050 [-0.0009;  0.0109]        5.4
## Mukherjee (2003)            0.0400 [ 0.0380;  0.0420]        5.6
## Mukherjee (2003)            0.0300 [ 0.0280;  0.0320]        5.6
## Mukherjee (2003)            0.0100 [ 0.0061;  0.0139]        5.5
## Mukherjee (2003)            0.0200 [ 0.0122;  0.0278]        5.3
```

```
## Bradbury and Crain (2001)  0.0036 [ 0.0008;  0.0065]        5.6
## Bradbury and Crain (2001)  0.0005 [-0.0016;  0.0027]        5.6
## Bradbury and Crain (2001)  0.0169 [ 0.0131;  0.0208]        5.6
## Bradbury and Crain (2001)  0.0123 [ 0.0087;  0.0160]        5.6
## Ricciuti (2004)            0.0140 [-0.0095;  0.0375]        3.5
## Ricciuti (2004)           -0.0110 [-0.0286;  0.0066]        4.2
## Ricciuti (2004)            0.0070 [-0.0067;  0.0207]        4.7
## Ricciuti (2004)            0.0050 [-0.0126;  0.0226]        4.2
## Ricciuti (2004)            0.0050 [-0.0126;  0.0226]        4.2
## Ricciuti (2004)            0.0120 [-0.0017;  0.0257]        4.7
##
## Number of studies combined: k = 21
##
##                          SMD             95%-CI    t p-value
## Random effects model 0.0078 [-0.0003; 0.0160] 2.01  0.0579
## Prediction interval         [-0.0259; 0.0416]
##
## Quantifying heterogeneity:
##  tau^2 = 0.0002 [0.0002; 0.0007]; tau = 0.0156 [0.0136; 0.0261];
##  I^2 = 98.5% [98.2%; 98.7%]; H = 8.11 [7.40; 8.88]
##
## Test of heterogeneity:
##        Q d.f.  p-value
##  1314.54   20 < 0.0001
##
## Details on meta-analytical method:
## - Inverse variance method
## - Restricted maximum-likelihood estimator for tau^2
## - Q-profile method for confidence interval of tau^2 and tau
## - Hartung-Knapp adjustment for random effects model
```

And the forest plot:

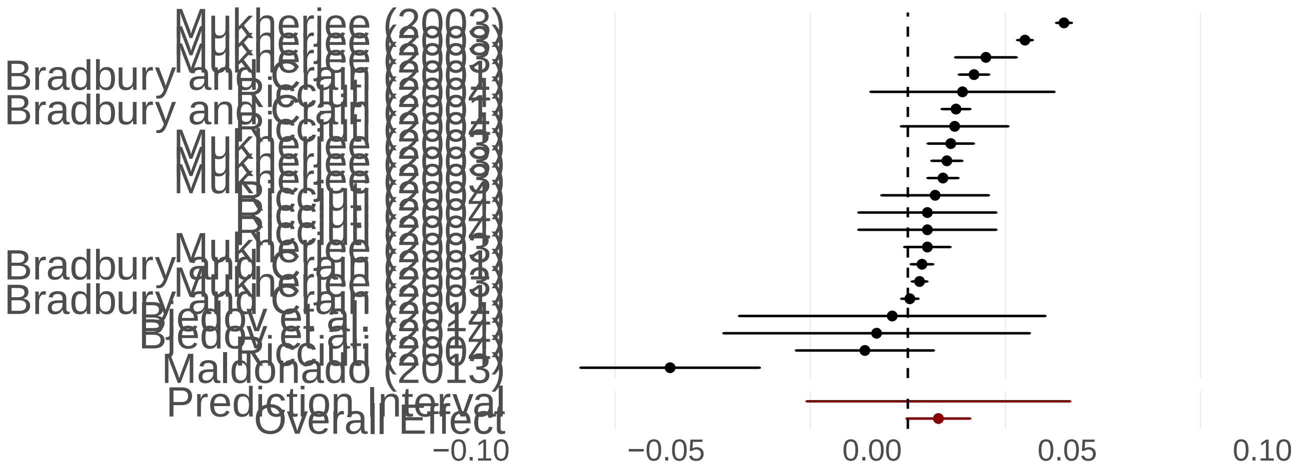



Figure 21: Effect of lower houses size (N) on percentage of public expenditure GDP (PCTGDP)

Highlights:

1. The results are highly heterogeneous: $I^2 = $ 98.48.
2. The Random effects modem SMD estimated is $g = $ 0.01 (*SE* = 0.004).
3. The prediction interval ranges from -0.03 to 0.04. Therefore, it emcompasses zero.

## H.8 Log of Lower House Size and Expenditure as Percentage of GDP

This meta-regression investigates the percentage of GDP as public expenditure as the dependent variable and the log lower house size (logN) as the treatment variable.

```
# Pooling effects analysis -- PCTGDP x logN
aux <- fulldat %>%
  filter(indepvar2 == 'logN',
         depvar2 == 'PCTGDP')

mod <- metagen(coef, SE, data=aux,
          studlab=paste(authoryear),
          comb.fixed = FALSE,
          comb.random = TRUE,
          method.tau = "REML",
          hakn = TRUE,
```

```
           prediction=TRUE,
           sm="SMD")
mod
```

```
##                            SMD            95%-CI %W(random)
## Baqir (1999)            2.0660 [ 1.4887; 2.6433]       18.9
## Baqir (1999)            2.0120 [ 1.4235; 2.6005]       18.8
## Baqir (1999)            2.4680 [ 1.8817; 3.0543]       18.8
## Lledo (2003)           -4.6900 [-9.9427; 0.5627]        3.8
## Stein et al. (1998)  0.0109 [-0.0171; 0.0389]       19.8
## Stein et al. (1998)  0.0135 [-0.0102; 0.0372]       19.8
##
## Number of studies combined: k = 6
##
##                            SMD            95%-CI    t p-value
## Random effects model 1.0619 [-0.7256; 2.8493] 1.53  0.1873
## Prediction interval         [-3.0267; 5.1504]
##
## Quantifying heterogeneity:
##  tau^2 = 1.6850 [0.6497; 38.1618]; tau = 1.2981 [0.8060; 6.1775];
##  I^2 = 96.9% [95.2%; 98.1%]; H = 5.71 [4.55; 7.16]
##
## Test of heterogeneity:
##       Q d.f.  p-value
##  163.00    5 < 0.0001
##
## Details on meta-analytical method:
## - Inverse variance method
## - Restricted maximum-likelihood estimator for tau^2
## - Q-profile method for confidence interval of tau^2 and tau
## - Hartung-Knapp adjustment for random effects model
```

And the forest plot:

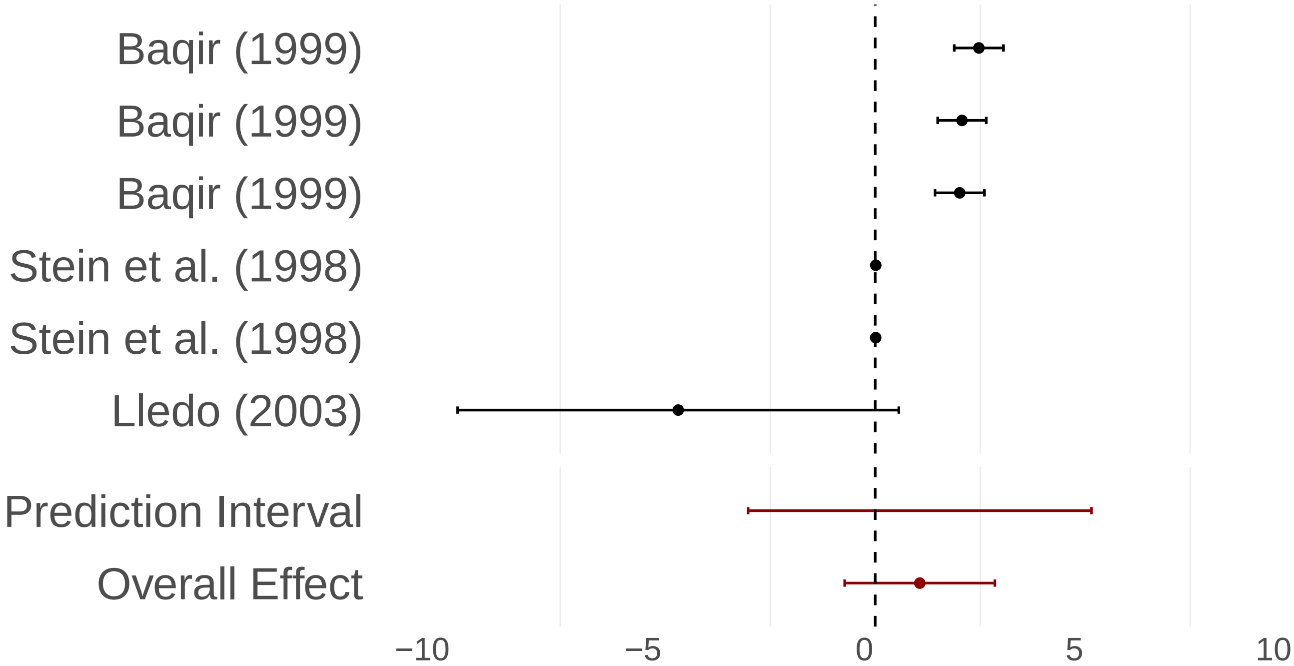



Figure 22: Effect of log lower houses size (logN) on the GDP share of public expenditure (PCTGDP)

Highlights:

1. The results are highly heterogeneous: $I^2 = $ 96.93.
2. The Random effects modem SMD estimated is $g = $ 1.06 (*SE* = 0.695).
3. The prediction interval ranges from -3.03 to 5.15. Therefore, it emcompasses zero.

## H.9 Upper House Size and Expenditure as Percentage of GDP

This model looks into the effect of upper house size (K) on the public expenditure share of the GDP (PCTGDP).

```
# Pooling effects analysis -- PCTGDP x K
aux <- fulldat %>%
  filter(indepvar2 == 'K',
         depvar2 == 'PCTGDP')

mod <- metagen(coef, SE, data=aux,
          studlab=paste(authoryear),
          comb.fixed = FALSE,
```

```r
          comb.random = TRUE,
          method.tau = "REML",
          hakn = TRUE,
          prediction=TRUE,
          sm="SMD")
mod
```

```
##                                 SMD             95%-CI %W(random)
## Maldonado (2012)            -0.0400 [-0.0659; -0.0141]        5.7
## Bradbury and Crain (2001)  0.0126 [ 0.0010;  0.0243]        9.8
## Bradbury and Crain (2001)  0.0050 [ 0.0016;  0.0083]       11.8
## Bradbury and Crain (2001) -0.0113 [-0.0163; -0.0064]       11.5
## Bradbury and Crain (2001) -0.0056 [-0.0102; -0.0010]       11.6
## Ricciuti (2004)              0.0160 [-0.0075;  0.0395]        6.2
## Ricciuti (2004)              0.0210 [-0.0006;  0.0426]        6.7
## Ricciuti (2004)              0.0140 [-0.0036;  0.0316]        7.9
## Ricciuti (2004)              0.0030 [-0.0088;  0.0148]        9.7
## Ricciuti (2004)              0.0300 [-0.0210;  0.0810]        2.2
## Ricciuti (2004)              0.0300 [-0.0210;  0.0810]        2.2
## Ricciuti (2004)              0.0390 [-0.0022;  0.0802]        3.1
## Ricciuti (2004)              0.0127 [-0.0147;  0.0401]        5.3
## Ricciuti (2004)              0.0160 [-0.0075;  0.0395]        6.2
##
## Number of studies combined: k = 14
##
##                          SMD            95%-CI    t p-value
## Random effects model 0.0056 [-0.0042; 0.0155] 1.24  0.2376
## Prediction interval         [-0.0233; 0.0346]
##
## Quantifying heterogeneity:
##  tau^2 = 0.0002 [0.0001; 0.0008]; tau = 0.0125 [0.0109; 0.0279];
```

```
## I^2 = 80.0% [67.3%; 87.8%]; H = 2.24 [1.75; 2.86]
##
## Test of heterogeneity:
##      Q d.f.  p-value
##  65.02   13 < 0.0001
##
## Details on meta-analytical method:
## - Inverse variance method
## - Restricted maximum-likelihood estimator for tau^2
## - Q-profile method for confidence interval of tau^2 and tau
## - Hartung-Knapp adjustment for random effects model
```

And the forest plot:

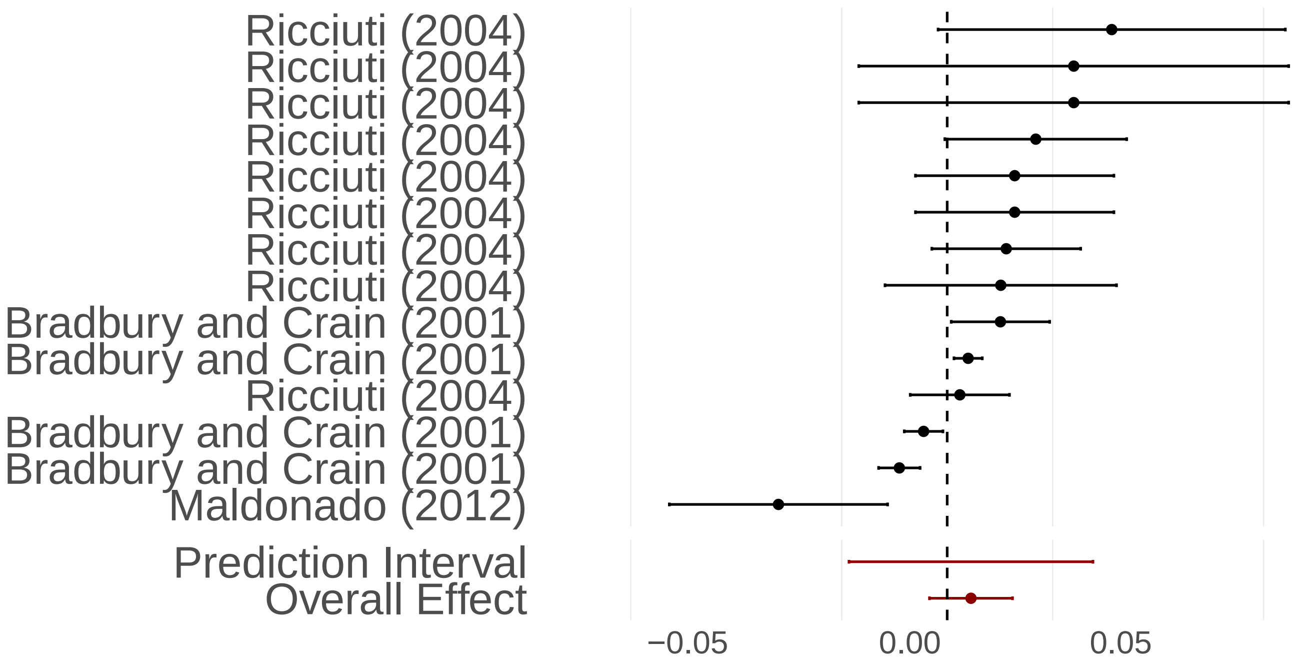



Figure 23: Effect of upper house size (K) on the public expenditure share of the GDP (PCTGDP)

Highlights:

1. The results are highly heterogeneous: $I^2 = $ 80.01.
2. The Random effects modem SMD estimated is $g = $ 0.01 (*SE* = 0.005).
3. The prediction interval ranges from -0.02 to 0.03. Therefore, it emcompasses zero.

# I Meta-regressions

## I.1 Meta-regressions for Expenditure as a Percentage of the GDP

```r
summary(mod)
```

```
##
## Mixed-Effects Model (k = 11; tau^2 estimator: REML)
##
##   logLik  deviance       AIC       BIC      AICc
##   5.2651  -10.5303    7.4697   -0.6428  187.4697
##
## tau^2 (estimated amount of residual heterogeneity):     0.0000 (SE = 0.0001)
## tau (square root of estimated tau^2 value):             0.0003
## I^2 (residual heterogeneity / unaccounted variability): 0.07%
## H^2 (unaccounted variability / sampling variability):   1.00
## R^2 (amount of heterogeneity accounted for):            100.00%
##
## Test for Residual Heterogeneity:
## QE(df = 3) = 3.6665, p-val = 0.2998
##
## Test of Moderators (coefficients 2:8):
## F(df1 = 7, df2 = 3) = 10.9991, p-val = 0.0373
##
## Model Results:
##
##                           estimate      se     tval    pval    ci.lb    ci.ub
## intrcpt                     2.6693  2.0404   1.3082  0.2820  -3.8241   9.1627
## indepvar2N                 -0.0094  0.0061  -1.5371  0.2219  -0.0287   0.0100
## indepvar2logN              -0.0109  0.0194  -0.5608  0.6141  -0.0726   0.0509
## year                       -0.0003  0.0010  -0.3279  0.7646  -0.0035   0.0029
## publishedNo                 0.0633  0.0159   3.9732  0.0285   0.0126   0.1141
```

```
## elecsys2Non-Majoritarian   -2.0556  0.3260  -6.3054  0.0081  -3.0930  -1.0181
## methodPANEL                  0.0557  0.0147   3.7831  0.0324   0.0088   0.1025
## agglevelStates              -0.0036  0.0250  -0.1453  0.8937  -0.0833   0.0760
##
## intrcpt
## indepvar2N
## indepvar2logN
## year
## publishedNo                 *
## elecsys2Non-Majoritarian  **
## methodPANEL                 *
## agglevelStates
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

As we have considerable heterogeneity in our sample, we run a permutation test to ensure the validity of our estimates. The results follow below.

```
##
## Test of Moderators (coefficients 2:8):
## F(df1 = 7, df2 = 3) = 10.9991, p-val* = 0.0050
##
## Model Results:
##
##                           estimate      se     tval   pval*    ci.lb    ci.ub
## intrcpt                     2.6693  2.0404   1.3082  0.3360  -3.8241   9.1627
## indepvar2N                 -0.0094  0.0061  -1.5371  0.0910  -0.0287   0.0100
## indepvar2logN              -0.0109  0.0194  -0.5608  0.5490  -0.0726   0.0509
## year                       -0.0003  0.0010  -0.3279  0.6690  -0.0035   0.0029
## publishedNo                 0.0633  0.0159   3.9732  0.0370   0.0126   0.1141
## elecsys2Non-Majoritarian   -2.0556  0.3260  -6.3054  0.0160  -3.0930  -1.0181
```

```
## methodPANEL                  0.0557  0.0147   3.7831  0.0480   0.0088   0.1025
## agglevelStates              -0.0036  0.0250  -0.1453  0.8640  -0.0833   0.0760
##
## intrcpt
## indepvar2N                 .
## indepvar2logN
## year
## publishedNo                *
## elecsys2Non-Majoritarian   *
## methodPANEL                *
## agglevelStates
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We have the following results for the meta-regressions of Expenditure Per Capita:

1. Compared with `K`, models with `N` and `logN` find significantly negative coefficients.
2. Year has null effect.
3. Unpublished papers tend to have higher coefficients than published papers.
4. Passing from `Majoritarian` to `Non-Majoritarian`, decreases significantly the effects found in our models.
5. In terms of the modeling, passing from `OLS` to `PANEL` increases the detected effects.
6. When passing from `Local` to `State` or `World` levels, it decreases the detected effect size.

Below we also run the meta-regressions adding all coefficients in the papers. The results follow below:

```r
summary(mod)
```

```
##
## Mixed-Effects Model (k = 41; tau^2 estimator: REML)
##
```

```
##    logLik   deviance        AIC        BIC       AICc
##   86.5613  -173.1227  -155.1227  -141.6541  -147.2966
##
## tau^2 (estimated amount of residual heterogeneity):     0.0001 (SE = 0.0000)
## tau (square root of estimated tau^2 value):             0.0102
## I^2 (residual heterogeneity / unaccounted variability): 93.92%
## H^2 (unaccounted variability / sampling variability):   16.44
## R^2 (amount of heterogeneity accounted for):            99.92%
##
## Test for Residual Heterogeneity:
## QE(df = 33) = 1004.7869, p-val < .0001
##
## Test of Moderators (coefficients 2:8):
## F(df1 = 7, df2 = 33) = 31.1106, p-val < .0001
##
## Model Results:
##
##                          estimate      se      tval    pval     ci.lb    ci.ub
## intrcpt                  -10.1105  5.5586   -1.8189  0.0780  -21.4195   1.1985
## indepvar2N                -0.0015  0.0050   -0.2936  0.7709   -0.0116   0.0087
## indepvar2logN              0.0376  0.0195    1.9296  0.0623   -0.0020   0.0772
## year                       0.0061  0.0028    2.2005  0.0349    0.0005   0.0117
## publishedNo                0.1134  0.0261    4.3478  0.0001    0.0603   0.1664
## elecsys2Non-Majoritarian  -2.1629  0.1626  -13.3046  <.0001   -2.4937  -1.8322
## methodPANEL                0.1256  0.0316    3.9796  0.0004    0.0614   0.1898
## agglevelStates            -0.0888  0.0359   -2.4710  0.0188   -0.1619  -0.0157
##
## intrcpt                    .
## indepvar2N
## indepvar2logN              .
```

```
## year                        *
## publishedNo               ***
## elecsys2Non-Majoritarian  ***
## methodPANEL               ***
## agglevelStates              *
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##
## Test of Moderators (coefficients 2:8):
## F(df1 = 7, df2 = 33) = 31.1106, p-val* = 0.0010
##
## Model Results:
##
##                           estimate      se      tval   pval*     ci.lb    ci.ub
## intrcpt                   -10.1105  5.5586   -1.8189  0.0630  -21.4195   1.1985
## indepvar2N                 -0.0015  0.0050   -0.2936  0.7500   -0.0116   0.0087
## indepvar2logN               0.0376  0.0195    1.9296  0.0500   -0.0020   0.0772
## year                        0.0061  0.0028    2.2005  0.0370    0.0005   0.0117
## publishedNo                 0.1134  0.0261    4.3478  0.0090    0.0603   0.1664
## elecsys2Non-Majoritarian   -2.1629  0.1626  -13.3046  0.0010   -2.4937  -1.8322
## methodPANEL                 0.1256  0.0316    3.9796  0.0100    0.0614   0.1898
## agglevelStates             -0.0888  0.0359   -2.4710  0.0400   -0.1619  -0.0157
##
## intrcpt                     .
## indepvar2N
## indepvar2logN               *
## year                        *
## publishedNo                **
## elecsys2Non-Majoritarian  ***
```

```
## methodPANEL                **
## agglevelStates              *
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

For all the coefficients, we have the following results:

1. Compared with `K`, models with `N` and `logN` tend to have significantly negative coefficients.
2. Year has a positive effect: the younger the publication, the higher the detected coefficient.
3. Unpublished papers tend to have higher coefficients than published papers.
4. Passing from `Majoritarian` to `Non-Majoritarian`, decreases significantly the effects found in our models.
5. In terms of the modeling, passing from `OLS` to `PANEL` increases the detected effects.
6. When passing from `Local` to `State` or `World` levels, it decreases the detected effect size.

## I.2 Meta-regressions for Expenditure Per Capita

```r
summary(mod)
```

```
##
## Mixed-Effects Model (k = 18; tau^2 estimator: REML)
##
##    logLik  deviance       AIC       BIC      AICc
## -34.6251   69.2502   85.2502   88.4333  157.2502
##
## tau^2 (estimated amount of residual heterogeneity):     1.8429 (SE = 1.2361)
## tau (square root of estimated tau^2 value):             1.3575
## I^2 (residual heterogeneity / unaccounted variability): 95.05%
## H^2 (unaccounted variability / sampling variability):   20.21
## R^2 (amount of heterogeneity accounted for):            0.00%
##
```

```
## Test for Residual Heterogeneity:
## QE(df = 11) = 45.4940, p-val < .0001
##
## Test of Moderators (coefficients 2:7):
## F(df1 = 6, df2 = 11) = 0.3429, p-val = 0.8998
##
## Model Results:
##
##                           estimate        se     tval    pval      ci.lb
## intrcpt                  -104.0701  318.9300  -0.3263  0.7503  -806.0302
## indepvar2N                 -2.9238    2.0932  -1.3968  0.1900    -7.5309
## year                        0.0525    0.1586   0.3308  0.7470    -0.2967
## elecsys2Non-Majoritarian    0.3458    1.5533   0.2226  0.8279    -3.0730
## methodPANEL                 1.4571    2.2376   0.6512  0.5283    -3.4679
## methodIV                    1.4936    2.6675   0.5599  0.5868    -4.3776
## agglevelStates             -0.0915    2.4255  -0.0377  0.9706    -5.4299
##                              ci.ub
## intrcpt                   597.8900
## indepvar2N                  1.6834
## year                        0.4017
## elecsys2Non-Majoritarian    3.7645
## methodPANEL                 6.3821
## methodIV                    7.3648
## agglevelStates              5.2470
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

As we have considerable heterogeneity in our sample, we run a permutation test to ensure the validity of our estimates. The results follow below.

```
## Error in rma.uni(x$yi, x$vi, weights = x$weights, mods = cbind(X[sample(x$k),  :
```

```
##   Fisher scoring algorithm did not converge. See 'help(rma)' for possible remedies.
## Error in rma.uni(x$yi, x$vi, weights = x$weights, mods = cbind(X[sample(x$k),  :
##   Fisher scoring algorithm did not converge. See 'help(rma)' for possible remedies.
## Error in rma.uni(x$yi, x$vi, weights = x$weights, mods = cbind(X[sample(x$k),  :
##   Fisher scoring algorithm did not converge. See 'help(rma)' for possible remedies.
## Error in rma.uni(x$yi, x$vi, weights = x$weights, mods = cbind(X[sample(x$k),  :
##   Fisher scoring algorithm did not converge. See 'help(rma)' for possible remedies.
## Error in rma.uni(x$yi, x$vi, weights = x$weights, mods = cbind(X[sample(x$k),  :
##   Fisher scoring algorithm did not converge. See 'help(rma)' for possible remedies.
## Error in rma.uni(x$yi, x$vi, weights = x$weights, mods = cbind(X[sample(x$k),  :
##   Fisher scoring algorithm did not converge. See 'help(rma)' for possible remedies.
## Error in rma.uni(x$yi, x$vi, weights = x$weights, mods = cbind(X[sample(x$k),  :
##   Fisher scoring algorithm did not converge. See 'help(rma)' for possible remedies.
## Error in rma.uni(x$yi, x$vi, weights = x$weights, mods = cbind(X[sample(x$k),  :
##   Fisher scoring algorithm did not converge. See 'help(rma)' for possible remedies.
## Error in rma.uni(x$yi, x$vi, weights = x$weights, mods = cbind(X[sample(x$k),  :
##   Fisher scoring algorithm did not converge. See 'help(rma)' for possible remedies.
## Error in rma.uni(x$yi, x$vi, weights = x$weights, mods = cbind(X[sample(x$k),  :
##   Fisher scoring algorithm did not converge. See 'help(rma)' for possible remedies.
## Error in rma.uni(x$yi, x$vi, weights = x$weights, mods = cbind(X[sample(x$k),  :
##   Fisher scoring algorithm did not converge. See 'help(rma)' for possible remedies.
## Error in rma.uni(x$yi, x$vi, weights = x$weights, mods = cbind(X[sample(x$k),  :
##   Fisher scoring algorithm did not converge. See 'help(rma)' for possible remedies.

##
## Test of Moderators (coefficients 2:7):
## F(df1 = 6, df2 = 11) = 0.3429, p-val* = 0.5790
##
## Model Results:
##
##                               estimate        se      tval   pval*      ci.lb
```

```
## intrcpt                    -104.0701  318.9300  -0.3263  0.6170  -806.0302
## indepvar2N                   -2.9238    2.0932  -1.3968  0.0760    -7.5309
## year                          0.0525    0.1586   0.3308  0.6070    -0.2967
## elecsys2Non-Majoritarian      0.3458    1.5533   0.2226  0.7270    -3.0730
## methodPANEL                   1.4571    2.2376   0.6512  0.3340    -3.4679
## methodIV                      1.4936    2.6675   0.5599  0.4020    -4.3776
## agglevelStates               -0.0915    2.4255  -0.0377  0.9520    -5.4299
##                               ci.ub
## intrcpt                    597.8900
## indepvar2N                   1.6834  .
## year                         0.4017
## elecsys2Non-Majoritarian     3.7645
## methodPANEL                  6.3821
## methodIV                     7.3648
## agglevelStates               5.2470
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We have the following results for the meta-regressions of Expenditure Per Capita:

1. Compared with `K`, models with `N` tend to detect significantly smaller effects.
2. Year has null effect.
3. Passing the electoral rules from `Majoritarian` to `Non-Majoritarian`, increases significantly the per capita expenditure found in our models.
4. In terms of the modeling, passing from `OLS` to `PANEL` or `IV` increases the detected effects.
5. When passing from `Local` to `State` level, decreases the detected effects.

Below we also run the meta-regressions adding all coefficients in the papers. The results follow below:

```r
summary(mod)
```

```
##
## Mixed-Effects Model (k = 60; tau^2 estimator: REML)
##
##    logLik   deviance        AIC        BIC       AICc
## -141.1228   282.2456   298.2456   314.0079   301.5183
##
## tau^2 (estimated amount of residual heterogeneity):     1.7264 (SE = 0.4944)
## tau (square root of estimated tau^2 value):             1.3139
## I^2 (residual heterogeneity / unaccounted variability): 99.80%
## H^2 (unaccounted variability / sampling variability):   500.07
## R^2 (amount of heterogeneity accounted for):            39.21%
##
## Test for Residual Heterogeneity:
## QE(df = 53) = 325.8548, p-val < .0001
##
## Test of Moderators (coefficients 2:7):
## F(df1 = 6, df2 = 53) = 5.9441, p-val < .0001
##
## Model Results:
##
##                            estimate        se     tval    pval      ci.lb
## intrcpt                   -296.9072  166.6870  -1.7812  0.0806  -631.2389
## indepvar2N                  -5.4468    0.9692  -5.6201  <.0001    -7.3907
## year                         0.1503    0.0830   1.8117  0.0757    -0.0161
## elecsys2Non-Majoritarian     1.0236    0.7701   1.3293  0.1894    -0.5209
## methodPANEL                 -0.1422    0.8136  -0.1747  0.8620    -1.7739
## methodIV                     0.1907    0.8223   0.2319  0.8175    -1.4587
## agglevelStates              -0.2008    1.0049  -0.1998  0.8424    -2.2164
##                             ci.ub
```

```
## intrcpt                    37.4245    .
## indepvar2N                 -3.5029  ***
## year                        0.3167    .
## elecsys2Non-Majoritarian    2.5682
## methodPANEL                 1.4896
## methodIV                    1.8401
## agglevelStates              1.8149
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##
## Test of Moderators (coefficients 2:7):
## F(df1 = 6, df2 = 53) = 5.9441, p-val* = 0.0010
##
## Model Results:
##
##                             estimate        se     tval   pval*      ci.lb
## intrcpt                    -296.9072  166.6870  -1.7812  0.0230  -631.2389
## indepvar2N                   -5.4468    0.9692  -5.6201  0.0010    -7.3907
## year                          0.1503    0.0830   1.8117  0.0200    -0.0161
## elecsys2Non-Majoritarian      1.0236    0.7701   1.3293  0.0790    -0.5209
## methodPANEL                  -0.1422    0.8136  -0.1747  0.8220    -1.7739
## methodIV                      0.1907    0.8223   0.2319  0.7250    -1.4587
## agglevelStates               -0.2008    1.0049  -0.1998  0.7850    -2.2164
##                              ci.ub
## intrcpt                    37.4245    *
## indepvar2N                 -3.5029  ***
## year                        0.3167    *
## elecsys2Non-Majoritarian    2.5682    .
## methodPANEL                 1.4896
```

```
## methodIV                    1.8401
## agglevelStates              1.8149
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

With all coefficients, the results of the effect sizes on the Expenditure Per Capita Regressions are the following:

1. Compared with `K`, models with `N` tend to detect significantly smaller effects.
2. Year has now a positive effect on coefficient sizes.
3. Passing the electoral rules from `Majoritarian` to `Non-Majoritarian`, increases significantly the effects on per capita expenditure found in our models.
4. In terms of the modeling, passing from `OLS` to `PANEL` decreases the detected effects.
5. All other coeffients were not significant.

## I.3 Meta-regressions for the Log of Expenditure Per Capita

```
summary(mod)
```

```
##
## Mixed-Effects Model (k = 7; tau^2 estimator: REML)
##
##   logLik  deviance       AIC       BIC      AICc
##   0.8657   -1.7315   12.2685   -1.7315  124.2685
##
## tau^2 (estimated amount of residual heterogeneity):     0.0096 (SE = 0.0147)
## tau (square root of estimated tau^2 value):             0.0977
## I^2 (residual heterogeneity / unaccounted variability): 92.15%
## H^2 (unaccounted variability / sampling variability):   12.74
## R^2 (amount of heterogeneity accounted for):            65.22%
##
```

```
## Test for Residual Heterogeneity:
## QE(df = 1) = 12.7408, p-val = 0.0004
##
## Test of Moderators (coefficients 2:6):
## F(df1 = 5, df2 = 1) = 2.9742, p-val = 0.4128
##
## Model Results:
##
##                 estimate       se     tval    pval      ci.lb     ci.ub
## intrcpt           8.9711  47.4747   0.1890  0.8811  -594.2521  612.1943
## indepvar2N       -0.1641   0.3258  -0.5037  0.7029    -4.3043    3.9760
## year             -0.0044   0.0237  -0.1864  0.8827    -0.3053    0.2965
## publishedNo       0.1520   0.1902   0.7993  0.5707    -2.2647    2.5687
## methodPANEL       0.2581   0.1886   1.3680  0.4018    -2.1389    2.6550
## agglevelStates   -0.0875   0.1901  -0.4602  0.7254    -2.5028    2.3278
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

As we have considerable heterogeneity in our sample, we run a permutation test to ensure the validity of our estimates. The results follow below.

```
##
## Test of Moderators (coefficients 2:6):
## F(df1 = 5, df2 = 1) = 2.9742, p-val* = 0.3720
##
## Model Results:
##
##                 estimate       se     tval   pval*      ci.lb     ci.ub
## intrcpt           8.9711  47.4747   0.1890  0.9030  -594.2521  612.1943
## indepvar2N       -0.1641   0.3258  -0.5037  0.7020    -4.3043    3.9760
## year             -0.0044   0.0237  -0.1864  0.9040    -0.3053    0.2965
```

```
## publishedNo       0.1520   0.1902   0.7993  0.5950    -2.2647    2.5687
## methodPANEL       0.2581   0.1886   1.3680  0.3800    -2.1389    2.6550
## agglevelStates   -0.0875   0.1901  -0.4602  0.6990    -2.5028    2.3278
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

| | Estimate | SE | T | P-Value | CI |
|---|---|---|---|---|---|
| Intercept | 8.971 | 47.475 | 0.189 | 0.903 | (-594.2521 ; 612.1943) |
| Indepvar: N | -0.164 | 0.326 | -0.504 | 0.702 | (-4.3043 ; 3.976) |
| Year | -0.004 | 0.024 | -0.186 | 0.904 | (-0.3053 ; 0.2965) |
| Published: No | 0.152 | 0.190 | 0.799 | 0.595 | (-2.2647 ; 2.5687) |
| Method: Panel | 0.258 | 0.189 | 1.368 | 0.380 | (-2.1389 ; 2.655) |
| AggLevel: States | -0.087 | 0.190 | -0.460 | 0.699 | (-2.5028 ; 2.3278) |

We have the following results for the meta-regressions of Log of Expenditure Per Capita:

1. Unpublished papers report a significantly higher coefficient.
2. In terms of the modeling, passing from `OLS` to `PANEL` increases the detected effects.
3. All other coefficients remained insignificant.

Below we also run the meta-regressions adding all coefficients in the papers. The results follow below:

```r
summary(mod)
```

```
##
## Mixed-Effects Model (k = 27; tau^2 estimator: REML)
##
##   logLik  deviance       AIC       BIC      AICc
##  21.9924  -43.9848  -27.9848  -20.0190  -14.8939
##
```

```
## tau^2 (estimated amount of residual heterogeneity):     0.0051 (SE = 0.0021)
## tau (square root of estimated tau^2 value):             0.0716
## I^2 (residual heterogeneity / unaccounted variability): 86.93%
## H^2 (unaccounted variability / sampling variability):   7.65
## R^2 (amount of heterogeneity accounted for):            82.37%
##
## Test for Residual Heterogeneity:
## QE(df = 20) = 98.5701, p-val < .0001
##
## Test of Moderators (coefficients 2:7):
## F(df1 = 6, df2 = 20) = 16.9707, p-val < .0001
##
## Model Results:
##
##                  estimate       se     tval    pval     ci.lb    ci.ub
## intrcpt           -1.6655  15.8337  -0.1052  0.9173  -34.6940  31.3630
## indepvar2N         0.0088   0.1262   0.0701  0.9448   -0.2544   0.2721
## year               0.0009   0.0079   0.1187  0.9067   -0.0155   0.0174
## publishedNo        0.0829   0.0728   1.1387  0.2683   -0.0689   0.2347
## methodPANEL       -0.2436   0.0705  -3.4537  0.0025   -0.3908  -0.0965   **
## methodRDD         -0.2978   0.0656  -4.5398  0.0002   -0.4347  -0.1610  ***
## agglevelStates    -0.0438   0.0673  -0.6505  0.5228   -0.1842   0.0966
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

With all coefficients, the results of the effect sizes on the Log of Expenditure Per Capita Regressions are the following:

1. In terms of the modeling, passing from `OLS` to `PANEL` or `RDD` decreases the detected effects.
2. All other coefficients remained insignificant.

# J Summary of Results

| | ExpPC | | | logExpPC | | | PCTGDP | | |
|---|---|---|---|---|---|---|---|---|---|
| | Estimate | SE | P-Value | Estimate | SE | Signif | Estimate | SE | Signif |
| **Base** | | | | | | | | | |
| Intercept | -104.07 | 318.93 | | 8.971 | 47.475 | | 2.6693 | 2.0404 | |
| Indepvar: N | -2.924 | 2.093 | | -0.164 | 0.326 | | -0.0094 | 0.0061 | |
| Indepvar: logN | | | | | | | -0.0109 | 0.0194 | |
| Year | 0.052 | 0.159 | | -0.004 | 0.024 | | -3e-04 | 0.001 | |
| Published: No | | | | 0.152 | 0.19 | | 0.0633 | 0.0159 | * |
| Elecsys: Non-Majoritarian | 0.346 | 1.553 | | | | | -2.0556 | 0.326 | ** |
| Method: Panel | 1.457 | 2.238 | | 0.258 | 0.189 | | 0.0557 | 0.0147 | * |
| Method: IV | 1.494 | 2.668 | | | | | | | |
| AggLevel: States | -0.091 | 2.425 | | -0.087 | 0.19 | | -0.0036 | 0.025 | |
| **All Coefs** | | | | | | | | | |
| Intercept | -296.907 | 166.687 | | -1.666 | 15.834 | | -10.11 | 5.559 | |
| Indepvar: N | -5.447 | 0.969 | *** | 0.009 | 0.126 | | -0.001 | 0.005 | |
| Indepvar: logN | | | | | | | 0.038 | 0.019 | |
| Year | 0.15 | 0.083 | | 0.001 | 0.008 | | 0.006 | 0.003 | * |

## K Summary of Results (Permutation)

| | ExpPC | | | logExpPC | | | PCTGDP | | |
|---|---|---|---|---|---|---|---|---|---|
| | Estimate | SE | Signif | Estimate | SE | Signif | Estimate | SE | Signif |
| **Base** | | | | | | | | | |
| Intercept | -104.07 | 318.93 | | 8.971 | 47.475 | | 2.669 | 2.04 | |
| Indepvar: N | -2.924 | 2.093 | | -0.164 | 0.326 | | -0.009 | 0.006 | |
| Indepvar: logN | | | | | | | -0.011 | 0.019 | |
| Year | 0.052 | 0.159 | | -0.004 | 0.024 | | 0 | 0.001 | |
| Published: No | | | | 0.152 | 0.19 | | 0.063 | 0.016 | * |
| Elecsys: Non-Majoritarian | 0.346 | 1.553 | | | | | -2.056 | 0.326 | * |
| Method: Panel | 1.457 | 2.238 | | 0.258 | 0.189 | | 0.056 | 0.015 | * |
| Method: IV | 1.494 | 2.668 | | | | | | | |
| AggLevel: States | -0.091 | 2.425 | | -0.087 | 0.19 | | -0.004 | 0.025 | |
| **All Coefs** | | | | | | | | | |
| Intercept | -296.907 | 166.687 | * | -1.666 | 15.834 | | -10.11 | 5.559 | |
| Indepvar: N | -5.447 | 0.969 | ** | 0.009 | 0.126 | | -0.001 | 0.005 | |
| Indepvar: logN | | | | | | | 0.038 | 0.019 | |
| Year | 0.15 | 0.083 | * | 0.001 | 0.008 | | 0.006 | 0.003 | * |

# L The Theory behind the Meta Analysis

There are two main estimators for conducting meta analysis: fixed effects and random effects models. The fixed effects model assumes that there is one true effect in reality, and that all estimates are an attempt to uncover this true effect. The random effects model, on the other hand, assumes that there are a distribution of true effects, that vary based on sample and tests characteristics.

In this paper, we use the random effects model. The empirical papers testing the law of 1/n are very diverse. We tried to capture some of this diversity by considering the main dependent and independent variables separately, but they have at least three other important sources of dispersion:

1. **Subjects**: Counties, Municipalities, States, Provinces, Countries.
2. **Electoral systems**: Majoritarian, PR, Mixed.
3. **Modeling strategies**: Panel data, Standard OLS, IV, RDD.

These sources of heterogeneity have two implications. First, it makes our estimates very disperse. The heterogeneity tests are all but one significant. When the sample sizes are large enough, we removed more heterogeneous studies, but we still had considerable dispersion in our estimates. Second, the amount of heterogeneity makes fixed effects estimates unrealistic and bised. Thus, we opt for random effects model.

Let each study having an effect of $T_i$. In a random effects model, we can decompose this effect in two components, the true effect that the study with the same specifications as $i$ come from, $\theta_i$, and a within-study error $\varepsilon_i$:

$$T_i \ = \ \theta_i + \varepsilon_i$$

And the random effects model assumes that the $\theta_i$ varies from study to study, having a true parameter $\mu$, plus a between-study error, $\xi_i$:

$$T_i \ = \ \mu + \xi_i + \varepsilon_i$$

And the random effects model estimates the parameter $\mu$, under the challenge of estimating both the within-and-between-study sampling errors.

In all empirical estimates, we use the package `meta`, and the package `dmetar`, described in (Doing Meta-Analysis with R)[`https://bookdown.org/MathiasHarrer/Doing_Meta_Analysis_in_`

R/random.html]. To empirically implement the random effects model, we need to choose a method to estimate the true effect size variance, $\tau^2$, which in our formulation, represents the variance of $\xi_i$. We selected the **Restricted Maximum Likelihood Estimator**, as the literature regards it as more precise when we have continuous measures, such as we have on our data (link)[https://www.ncbi.nlm.nih.gov/pubmed/26332144].

# M Robustness: Full model meta-regressions combined

In this section, we aggregate all the coefficients and run a multivariate meta-regression, controlling by:

1. The type of the dependent variable in the study (expenditure per capita, log of the expenditure per capita, and share of government expenditure in the GDP)
2. The type of the independent variable in the stydy (N, K, log of N);
3. The electoral system (Majoritarian, Proportional Representation, and Mixed).

The results follow below, and show null effect for all variables, including the intercept.

```
summary(mod)
```

```
##
## Mixed-Effects Model (k = 36; tau^2 estimator: REML)
##
##   logLik  deviance       AIC       BIC      AICc
## -48.7375   97.4751  125.4751  141.3720  177.9751
##
## tau^2 (estimated amount of residual heterogeneity):     0.2218 (SE = 0.0934)
## tau (square root of estimated tau^2 value):             0.4710
## I^2 (residual heterogeneity / unaccounted variability): 99.94%
## H^2 (unaccounted variability / sampling variability):   1664.72
## R^2 (amount of heterogeneity accounted for):            0.00%
##
## Test for Residual Heterogeneity:
```

```
## QE(df = 23) = 180.2779, p-val < .0001
##
## Test of Moderators (coefficients 2:13):
## F(df1 = 12, df2 = 23) = 0.3638, p-val = 0.9638
##
## Model Results:
##
##                             estimate        se     tval    pval      ci.lb
## intrcpt                       8.8454  102.4317   0.0864  0.9319  -203.0507
## depvar2PCTGDP                 0.3691    0.7171   0.5147  0.6117    -1.1143
## depvar2logExpPC              -0.3558    0.6671  -0.5333  0.5989    -1.7357
## indepvar2N                   -0.4660    0.5062  -0.9206  0.3668    -1.5132
## indepvar2logN                -0.1900    0.9137  -0.2079  0.8371    -2.0801
## year                         -0.0040    0.0508  -0.0792  0.9376    -0.1092
## publishedNo                   0.1640    0.5892   0.2783  0.7833    -1.0549
## elecsys2Non-Majoritarian      0.2772    0.6107   0.4539  0.6541    -0.9860
## methodPANEL                   0.0126    0.6188   0.0204  0.9839    -1.2676
## methodIV                     -0.1279    0.9306  -0.1375  0.8918    -2.0530
## methodRDD                    -0.1604    0.9312  -0.1722  0.8648    -2.0868
## agglevelStates               -0.4388    0.5952  -0.7371  0.4685    -1.6701
## agglevelCountries            -1.2370    1.0390  -1.1906  0.2459    -3.3864
##                                ci.ub
## intrcpt                     220.7414
## depvar2PCTGDP                 1.8524
## depvar2logExpPC               1.0242
## indepvar2N                    0.5812
## indepvar2logN                 1.7002
## year                          0.1011
## publishedNo                   1.3828
## elecsys2Non-Majoritarian      1.5404
```

```
## methodPANEL                  1.2928
## methodIV                     1.7971
## methodRDD                    1.7660
## agglevelStates               0.7926
## agglevelCountries            0.9123
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

As we have considerable heterogeneity in our sample, we run a permutation test to ensure the validity of our estimates. The results follow below.

```
##
## Test of Moderators (coefficients 2:13):
## F(df1 = 12, df2 = 23) = 0.3638, p-val* = 0.5730
##
## Model Results:
##
##                            estimate        se     tval   pval*      ci.lb
## intrcpt                      8.8454  102.4317   0.0864  0.9140  -203.0507
## depvar2PCTGDP                0.3691    0.7171   0.5147  0.4550    -1.1143
## depvar2logExpPC             -0.3558    0.6671  -0.5333  0.4590    -1.7357
## indepvar2N                  -0.4660    0.5062  -0.9206  0.1590    -1.5132
## indepvar2logN               -0.1900    0.9137  -0.2079  0.7420    -2.0801
## year                        -0.0040    0.0508  -0.0792  0.9200    -0.1092
## publishedNo                  0.1640    0.5892   0.2783  0.6800    -1.0549
## elecsys2Non-Majoritarian     0.2772    0.6107   0.4539  0.5040    -0.9860
## methodPANEL                  0.0126    0.6188   0.0204  0.9730    -1.2676
## methodIV                    -0.1279    0.9306  -0.1375  0.8410    -2.0530
## methodRDD                   -0.1604    0.9312  -0.1722  0.8330    -2.0868
## agglevelStates              -0.4388    0.5952  -0.7371  0.2720    -1.6701
## agglevelCountries           -1.2370    1.0390  -1.1906  0.1190    -3.3864
```

```
##                                ci.ub
## intrcpt                     220.7414
## depvar2PCTGDP                 1.8524
## depvar2logExpPC               1.0242
## indepvar2N                    0.5812
## indepvar2logN                 1.7002
## year                          0.1011
## publishedNo                   1.3828
## elecsys2Non-Majoritarian      1.5404
## methodPANEL                   1.2928
## methodIV                      1.7971
## methodRDD                     1.7660
## agglevelStates                0.7926
## agglevelCountries             0.9123
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

In the main text, we selected the coefficients based on the regressions that had most observations and that presented a full model (with fixed effects or intermediate bandwidth in RDD). Below we also run the meta-regressions adding all coefficients in the papers. The results follow below:

```r
summary(mod)
```

```
##
## Mixed-Effects Model (k = 128; tau^2 estimator: REML)
##
##    logLik   deviance        AIC        BIC       AICc
## -192.8810   385.7620   413.7620   452.1911   417.9620
##
## tau^2 (estimated amount of residual heterogeneity):     0.0624 (SE = 0.0107)
## tau (square root of estimated tau^2 value):             0.2499
```

```
## I^2 (residual heterogeneity / unaccounted variability): 99.96%
## H^2 (unaccounted variability / sampling variability):   2820.84
## R^2 (amount of heterogeneity accounted for):            66.55%
##
## Test for Residual Heterogeneity:
## QE(df = 115) = 2096.9691, p-val < .0001
##
## Test of Moderators (coefficients 2:13):
## F(df1 = 12, df2 = 115) = 2.9370, p-val = 0.0014
##
## Model Results:
##
##                           estimate       se     tval    pval     ci.lb
## intrcpt                    52.2976  32.3686   1.6157  0.1089  -11.8184
## depvar2PCTGDP               0.6076   0.2760   2.2016  0.0297    0.0609
## depvar2logExpPC            -0.2159   0.1884  -1.1458  0.2543   -0.5892
## indepvar2N                 -0.1568   0.1444  -1.0858  0.2798   -0.4429
## indepvar2logN              -0.1304   0.2978  -0.4378  0.6623   -0.7203
## year                       -0.0258   0.0161  -1.6016  0.1120   -0.0576
## publishedNo                -0.1154   0.1592  -0.7252  0.4698   -0.4308
## elecsys2Non-Majoritarian    0.5684   0.2168   2.6211  0.0100    0.1388
## methodPANEL                -0.2318   0.1475  -1.5720  0.1187   -0.5239
## methodIV                   -0.1568   0.2357  -0.6650  0.5074   -0.6237
## methodRDD                  -0.3624   0.2287  -1.5846  0.1158   -0.8154
## agglevelStates             -0.6097   0.2094  -2.9118  0.0043   -1.0244
## agglevelCountries          -1.6121   0.3690  -4.3685  <.0001   -2.3431
##                              ci.ub
## intrcpt                   116.4136
## depvar2PCTGDP               1.1542    *
## depvar2logExpPC             0.1574
```

```
## indepvar2N                   0.1293
## indepvar2logN                0.4596
## year                         0.0061
## publishedNo                  0.1999
## elecsys2Non-Majoritarian     0.9979   **
## methodPANEL                  0.0603
## methodIV                     0.3102
## methodRDD                    0.0906
## agglevelStates              -0.1949   **
## agglevelCountries           -0.8811  ***
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
permutest(mod, progbar = F)
```

```
##
## Test of Moderators (coefficients 2:13):
## F(df1 = 12, df2 = 115) = 2.9370, p-val* = 0.0010
##
## Model Results:
##
##                           estimate       se     tval   pval*     ci.lb
## intrcpt                    52.2976  32.3686   1.6157  0.0160  -11.8184
## depvar2PCTGDP               0.6076   0.2760   2.2016  0.0060    0.0609
## depvar2logExpPC            -0.2159   0.1884  -1.1458  0.1020   -0.5892
## indepvar2N                 -0.1568   0.1444  -1.0858  0.0990   -0.4429
## indepvar2logN              -0.1304   0.2978  -0.4378  0.4970   -0.7203
## year                       -0.0258   0.0161  -1.6016  0.0160   -0.0576
## publishedNo                -0.1154   0.1592  -0.7252  0.2820   -0.4308
## elecsys2Non-Majoritarian    0.5684   0.2168   2.6211  0.0010    0.1388
## methodPANEL                -0.2318   0.1475  -1.5720  0.0190   -0.5239
```

```
## methodIV                  -0.1568   0.2357  -0.6650  0.2810   -0.6237
## methodRDD                 -0.3624   0.2287  -1.5846  0.0370   -0.8154
## agglevelStates            -0.6097   0.2094  -2.9118  0.0010   -1.0244
## agglevelCountries         -1.6121   0.3690  -4.3685  0.0010   -2.3431
##                              ci.ub
## intrcpt                   116.4136    *
## depvar2PCTGDP               1.1542   **
## depvar2logExpPC             0.1574
## indepvar2N                  0.1293    .
## indepvar2logN               0.4596
## year                        0.0061    *
## publishedNo                 0.1999
## elecsys2Non-Majoritarian    0.9979  ***
## methodPANEL                 0.0603    *
## methodIV                    0.3102
## methodRDD                   0.0906    *
## agglevelStates             -0.1949  ***
## agglevelCountries          -0.8811  ***
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

# N Auxiliary functions

## N.1 Function for generate the Meta-Analytic Figures

This function receives the meta-analysis results and build a forest plot using `ggplot2`.

```
# Build plot function for forest plots
build_forest <- function(mod, capt, lsize = 22, ttl = NULL) {
  # Build dataset for plot
  mod2 <- tibble(
```

```r
  TE = mod$TE,
  seTE = mod$seTE,
  studlab = mod$studlab,
  lower = mod$lower,
  upper = mod$upper,
  group = "A") %>%
  bind_rows(.,
            aux = tibble(
              TE = c(mod$TE.random, NA),
              seTE = c(mod$seTE.random, NA),
              studlab = c("Overall Effect", "Prediction Interval"),
              lower = c(mod$lower.random, mod$lower.predict),
              upper = c(mod$upper.random, mod$upper.predict),
              group = "B")) %>%
  group_by(studlab) %>%
  mutate(studlab2 = paste0(studlab, "_", 1:n())) %>%
  ungroup()

# Graph limits
limg <- max(abs(c(mod2$lower, mod2$upper)))

# Build plot
p <- mod2 %>%
  ggplot(aes(y = reorder(studlab2,TE),
             x = TE, xmin = lower, xmax = upper)) +
  geom_point(aes(color = group)) +
  geom_errorbarh(aes(color = group), height = 0.1) +
  scale_color_manual(values = c("#000000", "#8b0000")) +
  scale_x_continuous(limits=c(-1.1*limg, 1.1*limg)) +
  scale_y_discrete(
```

```
      labels = function(x) str_replace(x, "_[0-9]*$", "")) +
    geom_vline(xintercept=0,
               color="#000000", linetype="dashed") +
    labs(x = "",
         y = "") +
    facet_grid(group~., scales = "free", space = "free") +
    labs(caption = capt,
         title = ttl) +
    theme_minimal() %+replace%
    theme(strip.text.y = element_blank(),
          legend.position = "none",
          axis.text.y = element_text(size = .8*lsize,
                                     hjust = 1),
          axis.text.x = element_text(size = .6*lsize,
                                     hjust = 1.1),
          plot.caption = element_text(size=lsize),
          plot.title.position = 'plot',
          plot.title = element_text(hjust = 0.5,
                                    face = 'bold',
                                    margin=margin(0,0,10,0)),
          panel.grid.major = element_blank())
  return(p)
}
```

## N.2 Webscraping code

```
#### PACKAGES CONFIGURATION ####

# Needed packages
pkgs <- c('tidyverse', 'rvest', 'RSelenium')
```

```r
# Install if not already installed
installIfNot <- function(x) {
  if(x %in% rownames(installed.packages()) == FALSE)
    install.packages(x, dependencies = T, repos = "http://cran.us.r-project.org")
}
lapply(pkgs, installIfNot)

# Load packs
lapply(pkgs, require, character.only = T)
rm(pkgs, installIfNot)

#### SETTING UP SELENIUM ####

# Alternative 1: Setting up Selenium (head)
rsD <- rsDriver(port = 1114L, browser = c("firefox"))
remDr <- rsD$client
remDr$open()

#### Google scholar ####

# site: https://scholar.google.com
remDr$navigate("https://scholar.google.com/
               scholar?cites=13117579863846712459&as_sdt=2005&sciodt=0,5")

articles_weingast <- tibble(
  value = NA,
  term = NA,
  page = NA
)
```

```r
k <- 0

for (j in 1:213) { # we had to manually choose the number of pages here; Y.M.M.V.

  Sys.sleep(rpois(1, 5))

  # Getting articles basic information
  k <- k + 1

  webElem <- remDr$findElement("css", "body")


  title <- read_html(remDr$getPageSource()[[1]]) %>%
    html_nodes(
      xpath = '//*[contains(concat( " ", @class, " " ),
      concat( " ", "gs_rt", " " ))]'
      ) %>%
    html_text() %>%
    enframe(name = NULL) %>%
    rename("title" = "value") %>%
    mutate(page = k)

  articles_partial <- read_html(remDr$getPageSource()[[1]]) %>%
    html_nodes(xpath = '//*[contains(concat( " ", @class, " " ), concat( " ", "gs_a", " " ))]')
    html_text() %>%
    enframe(name = NULL) %>%
    bind_rows(
      tibble(
        value = "delete",
```

```
        term = NA,
        page = NA
      ),
      .
    ) %>%
    bind_cols(., title)


  # Binding articles
  articles_weingast <- bind_rows(articles_weingast, articles_partial)

  # Changing Pages
  next_button <- remDr$findElement(using = "xpath", "/html/body/div/div[11]/div[2]/div[2]/div[3
                                  div[2]/center/table/tbody/tr/td[12]/a/b")
  next_button$clickElement()
  # Deleting cookies
  remDr$deleteAllCookies()
}

write_csv(articles_weingast, "scholar_weingast_raw.csv")

articles_weingast <- articles_weingast %>%
  select(-term, -page) %>%
  filter(value != "delete") %>%
  slice(2:nrow(.)) %>%
  filter(!grepl("books.google.com", value)) %>%
  filter(!grepl("BOOK", title)) %>%
  separate(., value, into = c("author", "value"),
           sep = " -", remove = T, extra = "merge", fill = "right") %>%
  separate(., value, into = c("journal", "year"),
```

```
           sep = ",", remove = T, extra = "merge", fill = "right") %>%
  mutate(
    year_2 = ifelse(is.na(year), journal, year),
    journal = ifelse(is.na(year), NA, journal),
    year = year_2
  ) %>%
  select(-year_2) %>%
  separate(., year, into = c("year", "site"),
           sep = "-", remove = T, extra = "merge", fill = "right") %>%
  mutate(
    year = gsub("[^0-9 ]", "", value),
    year = gsub("^[0-9]{5,}", "", year),
    year = gsub("^ {1,}", "", year)
  ) %>%
  separate(., year, into = c("year", "junk"),
           sep = " ", remove = T, extra = "merge", fill = "right") %>%
  separate(., value, into = c("journal", "value"),
           sep = "-", remove = T, extra = "merge", fill = "right") %>%
  mutate(
    journal_untidy = year,
    year = gsub("[^0-9 ]", "", year)
  )

articles <- articles %>%
  na.omit() %>%
  distinct(., value, .keep_all = T)

write_csv(articles, "google_scholar_clean.csv")

#### SCOPUS ####
```

```
# url: https://www.scopus.com/home.uri

# Scraping Scopus requires a bit more manual labor.
# You can login on scopus through your university/institution
# and download the metadata of the article(s) you
# want directly from there.
# All we need to do after that is scrape the information of
# every link from the .csv file downloaded previously

scopus <- read_csv("scopus.csv")

scopus <- scopus %>%
  mutate(article = map_chr(Link, ~ {
    remDr$navigate(.x)
    read_html(remDr$getPageSource()[[1]]) %>%
      html_nodes(xpath = '//*[(@id = "abstractSection")]//p') %>%
      html_text() %>%
      paste(., collapse = "\r\n")
  })) %>%
  mutate(
    article = gsub(
      '\r\n\nUse this section.*Topics\n\n\n"',
      "",
      article
    ),
    article = gsub(
      "Topics are unique.*onwards.",
      "",
      article
```

```
    ),
    article = gsub(
      "Use this section.*documents.",
      "",
      article
    ),
    article = gsub(
      "Learn more about these Topics",
      "",
      article
    ),
    article = gsub(
      ". 20.*, Springer Science\\+Business Media, LLC, part of Springer Nature.",
      "",
      article
    ),
    article = gsub("\r", "", article),
    article = gsub("\n", "", article),
    article = gsub(" {2,}", " ", article),
    article = gsub(" {3,}", "", article)
  )

write_csv(scopus, "scopus_clean.csv")

#### Microsoft Academic ####

# url: https://academic.microsoft.com/home
articles <- list()
k <- 0
remDr$navigate("https://academic.microsoft.com/paper/
```

```
               2076316673/citedby/search?q=The%20Political%
               20Economy%20of%20Benefits%20and%20Costs%3A%20A%
               20Neoclassical%20Approach%20to
               %20Distributive%20Politics&qe=RId%253D2076316673&f=&orderBy=0")


# 1) Getting hyperlinks from articles
for (j in 1:100) {
  k <- k + 1
  print(k)

  # Navigating Website

  Sys.sleep(rpois(2, 5))

  # Getting articles' links
  articles[[k]] <- read_html(remDr$getPageSource()[[1]]) %>%
    html_nodes(xpath = "//a") %>%
    html_attr("href") %>%
    enframe(name = NULL) %>%
    filter(
      grepl("paper/", value),
      !grepl("citedby", value)
    ) %>%
    mutate(value = paste0("https://academic.microsoft.com/", value))

  # Changing Page
  next_page_others <- remDr$findElement(using = "xpath", "/html/body/div/div/div/router-view/ro
  compose/div/div[2]/ma-pager/div/i[2]")
  next_page_others$clickElement()
```

```
}



articles <- articles %>%
  reduce(bind_rows) %>%
  distinct(value, .keep_all = T)

# 2) Navigating through articles and scraping them

articles_links <- articles %>%
  mutate(
    abstract = NA,
    title = NA,
    year = NA,
    journal = NA,
    authors = NA,
    tags = NA
  )

for (i in 1:nrow(articles_links)) {
  remDr$navigate(articles_links$value[i])
  Sys.sleep(rpois(1, 4))

  articles_links$abstract[i] <- read_html(remDr$getPageSource()[[1]]) %>%
    html_nodes(., xpath = "//html/body/div/div/div/router-view/compose[1]/
                div/div/ma-entity-detail-info/compose/div/div/div[1]/p") %>%
    html_text() %>%
    paste(., collapse = " ")

  articles_links$title[i] <- read_html(remDr$getPageSource()[[1]]) %>%
```

```
  html_nodes(.,
             xpath = '//*[contains(concat( " ", @class, " " ),
             concat( " ", "name", " " ))]') %>%
  html_text() %>%
  paste(., collapse = " ")

articles_links$year[i] <- read_html(remDr$getPageSource()[[1]]) %>%
  html_nodes(.,
             xpath = '//*[contains(concat( " ", @class, " " ),
             concat( " ", "name-section", " " ))]
             //*[contains(concat( " ", @class, " " ), concat( " ", "year", " " ))]') %>%
  html_text() %>%
  paste(., collapse = " ")

articles_links$journal[i] <- read_html(remDr$getPageSource()[[1]]) %>%
  html_nodes(., xpath = '//*[contains(concat( " ", @class, " " )
             , concat( " ", "pub-name", " " ))]') %>%
  html_text() %>%
  paste(., collapse = " ")

articles_links$authors[i] <- read_html(remDr$getPageSource()[[1]]) %>%
  html_nodes(., xpath = "/html/body/div/div/div/router-view/compose[1]/div/div/
             ma-entity-detail-info/compose/div/div/div[1]/
             ma-author-string-collection") %>%
  html_text() %>%
  paste(., collapse = " ")

articles_links$tags[i] <- read_html(remDr$getPageSource()[[1]]) %>%
  html_nodes(., xpath = "/html/body/div/div/div/router-view/compose[1]/
             div/div/ma-entity-detail-info/compose/
```

```
               div/div/div[1]/ma-tag-cloud/div") %>%
    html_text() %>%
    paste(., collapse = " ")
}


write_csv(articles_links, "microsoft_academic_clean.csv")
```

# O Software Version for this Appendix

```
sessionInfo()
```

```
## R version 3.6.3 (2020-02-29)
## Platform: x86_64-apple-darwin15.6.0 (64-bit)
## Running under: macOS Catalina 10.15.5
##
## Matrix products: default
## BLAS:   /Library/Frameworks/R.framework/Versions/3.6/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/3.6/Resources/lib/libRlapack.dylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] grid      parallel  stats     graphics  grDevices utils     datasets
## [8] methods   base
##
## other attached packages:
##  [1] magick_2.3          kableExtra_1.1.0    ggpubr_0.3.0
##  [4] gridExtra_2.3       gridGraphics_0.5-0  knitr_1.28
##  [7] compareGroups_4.4.1 SNPassoc_1.9-2      mvtnorm_1.1-1
```

```
## [10] survival_3.1-12     haplo.stats_1.7.9   readxl_1.3.1
## [13] metafor_2.4-0       Matrix_1.2-18       meta_4.12-0
## [16] forcats_0.5.0       stringr_1.4.0       dplyr_0.8.5
## [19] purrr_0.3.4         readr_1.3.1         tidyr_1.1.0
## [22] tibble_3.0.1        ggplot2_3.3.1       tidyverse_1.3.0
##
## loaded via a namespace (and not attached):
##   [1] TH.data_1.0-10      minqa_1.2.4         colorspace_1.4-1
##   [4] ggsignif_0.6.0      rio_0.5.16          ellipsis_0.3.1
##   [7] flextable_0.5.10    htmlTable_1.13.3    base64enc_0.1-3
##  [10] fs_1.4.1            rstudioapi_0.11     mice_3.9.0
##  [13] farver_2.0.3        MatrixModels_0.4-1  fansi_0.4.1
##  [16] lubridate_1.7.9     xml2_1.3.2          codetools_0.2-16
##  [19] splines_3.6.3       Formula_1.2-3       jsonlite_1.6.1
##  [22] nloptr_1.2.2.1      broom_0.5.6         cluster_2.1.0
##  [25] dbplyr_1.4.4        png_0.1-7           compiler_3.6.3
##  [28] httr_1.4.1          backports_1.1.7     assertthat_0.2.1
##  [31] cli_2.0.2           acepack_1.4.1       htmltools_0.4.0
##  [34] quantreg_5.55       tools_3.6.3         gtable_0.3.0
##  [37] glue_1.4.1          tinytex_0.23        Rcpp_1.0.4.6
##  [40] carData_3.0-4       cellranger_1.1.0    vctrs_0.3.1
##  [43] writexl_1.3         nlme_3.1-148        xfun_0.14
##  [46] openxlsx_4.1.5      lme4_1.1-23         rvest_0.3.5
##  [49] CompQuadForm_1.4.3  lifecycle_0.2.0     rstatix_0.5.0
##  [52] statmod_1.4.34      polspline_1.1.19    MASS_7.3-51.6
##  [55] zoo_1.8-8           scales_1.1.1        hms_0.5.3
##  [58] sandwich_2.5-1      SparseM_1.78        RColorBrewer_1.1-2
##  [61] HardyWeinberg_1.6.3 curl_4.3            yaml_2.2.1
##  [64] gdtools_0.2.2       rms_6.0-0           rpart_4.1-15
##  [67] latticeExtra_0.6-29 stringi_1.4.6       highr_0.8
```

```
##  [70] checkmate_2.0.0     zip_2.0.4           boot_1.3-25
##  [73] truncnorm_1.0-8     chron_2.3-55        systemfonts_0.2.3
##  [76] rlang_0.4.6         pkgconfig_2.0.3     Rsolnp_1.16
##  [79] evaluate_0.14       lattice_0.20-41     labeling_0.3
##  [82] htmlwidgets_1.5.1   cowplot_1.0.0       tidyselect_1.1.0
##  [85] magrittr_1.5        R6_2.4.1            generics_0.0.2
##  [88] Hmisc_4.4-0         multcomp_1.4-13     DBI_1.1.0
##  [91] pillar_1.4.4        haven_2.3.1         foreign_0.8-75
##  [94] withr_2.2.0         abind_1.4-5         nnet_7.3-14
##  [97] car_3.0-8           modelr_0.1.8        crayon_1.3.4
## [100] uuid_0.1-4          officer_0.3.11      rmarkdown_2.2
## [103] jpeg_0.1-8.1        data.table_1.12.8   blob_1.2.1
## [106] webshot_0.5.2       reprex_0.3.0        digest_0.6.25
## [109] munsell_0.5.0       viridisLite_0.3.0
```