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

Huzeyfe Alptekin* Danilo Freire† Umberto Mignozzetti‡ Catarina Roman§

6 December 2020

Contents

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

†Independent Researcher, danilofreire@gmail.com, http://danilofreire.github.io. Corresponding author.

‡Visiting Assistant Professor, Quantitative Theory and Methods, Emory University, umberto.mignozzetti@emory.edu, http://umbertomig.com.

§Research Associate, School of International Relations, Fundação Getulio Vargas, São Paulo, SP, Brazil, catarinamroman@gmail.com.

A Search Criteria

The first step in our systematic review consisted in gathering a study sample. We started our data collection with a manual search based on a set of keywords we scouted from the distributive politics literature. This search produced a database with many entries that were unrelated to our subject of investigation. To reduce the number of false positives in our sample, we restricted our search to studies that cited Weingast, Shepsle and Johnsen's 1981 paper "The Political Economy of Benefits and Costs: A Neoclassical Approach to Distributive Politics", which is a seminal contribution in the field. Although Google Scholar reports the article has received 2,180 citations, our search resulted in 2,664 records on the 21st of November 2019.

We webscraped three large academic databases: Google Scholar (n = 1001); Microsoft Academic (n = 927); and Scopus (n = 736). The `R` script we wrote extracted the article title, abstract, authors, year, journal of publication, and database from which the record originated. Our code is available in section L.2 below. We screened these results with an English language and article restriction, that is, we excluded all records written in other languages and all that were not academic papers, such as book chapters or doctoral theses. We set no restriction to unpublished articles.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 clearly unrelated to our topic of interest. 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. Then, we read all abstracts. We chose to maintain those which indicated that either government expenditure or legislative structures were the main subject of the paper. 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 were also included. This allowed us to significantly reduce our sample to 376 records.

In the second phase, we assessed full texts. To remain in our sample, the paper should (i) conduct a quantitative analysis, (ii) report data on the number of legislators, and (iii) also include data on public expenditure. If the publication had all three, it was maintained. Disagreements in this phase were discussed among the authors, and a third investigator 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 authors use different measures for government spending and the number of lower/upper house members, we extracted all coefficients that provided this information. We decided which variables to keep by following the current practices of the literature. 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 29 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 in the Lower House
- *logN*: Log Number Legislators in the Lower House
- *K*: Number Legislators in the Upper House

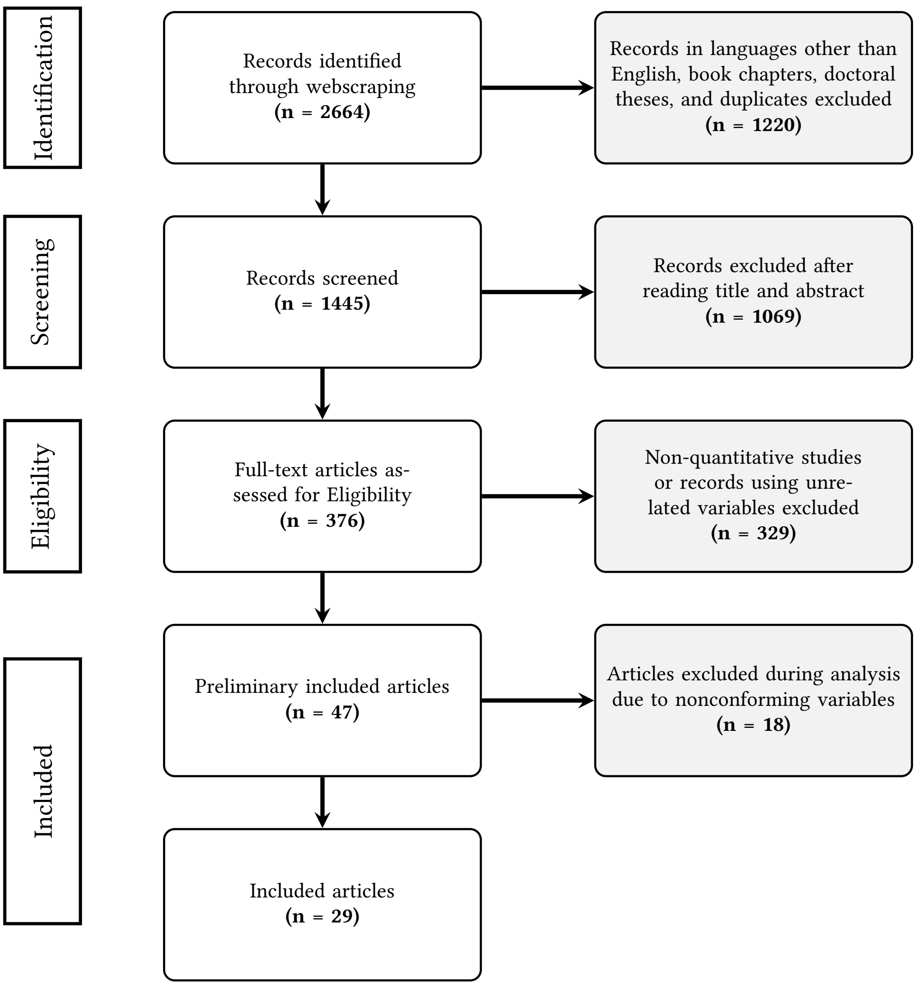
2. Matched outcome variable:

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

B.2 Flow Chart

The diagram below shows each step of our article selection process. We followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement to conduct our study[1]. 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.

[1] More information about the PRISMA statement is available at http://www.prisma-statement.org.



C Meta-Analysis Dataset

Our data are comprised of two datasets. The first dataset has the main coefficients reported in the studies. These data include only the most rigorous model from each paper, that is, those estimated with the largest sample size, 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 42 estimates, as 13 articles analysed two dependent or independent variables of interest. Our second sample, in contrast, contains all the 142 effect sizes reported in the 29 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 dataset. Below is the data extraction process for all relevant coefficients in the selected articles. Here we present the results of all tests performed in both reduced and full samples.

D Descriptive Statistics

In this section, we show the descriptive statistics for our sample. We focus on the following paper characteristics: study year, whether the paper has been published or not, the electoral system of the country discussed in the original study, the data aggregation level, as well as the distribution of the dependent and independent variables of interest. 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 2008.93, with standard deviation of 6.56. The oldest study included in the paper is from 1998, while the most recent paper was written in 2019. Therefore, we cover 21 years of tests of the *law of 1/n*.

```
dat %>%
  select(id, year) %>%
  unique() %>%
  ggplot(aes(x = as.factor(year))) +
    geom_bar(color = "black") +
  labs(x = "",
       y = "") +
  theme_bw()
```

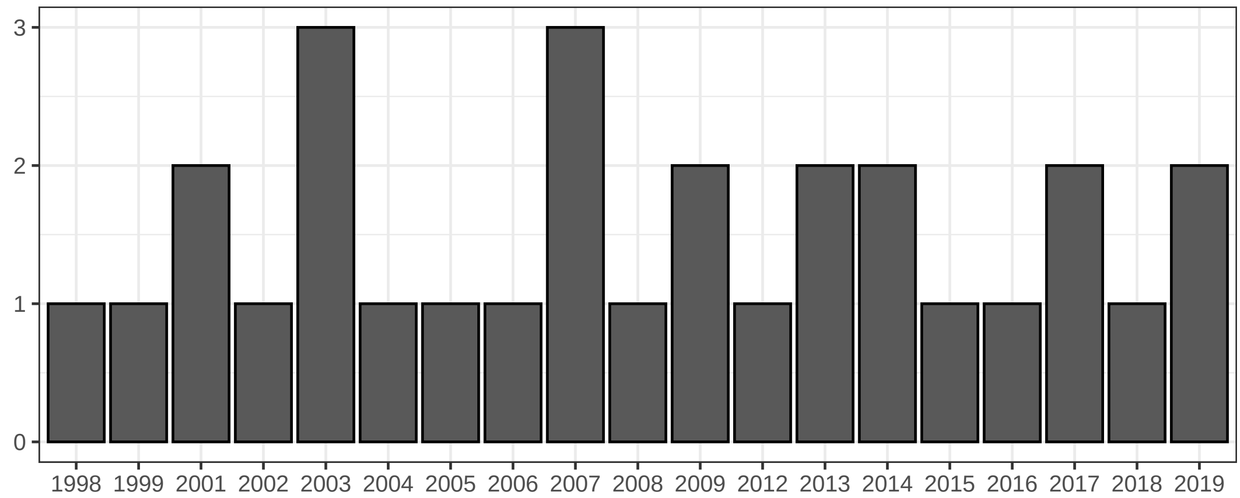


Figure 1: Study Year Frequencies

D.2 Frequency of Published Papers

Studies were included in our sample regardless of their publication status. From the 29 papers in the sample, 25 were published while 4 were not published.

```
dat %>%
  select(id, published) %>%
  unique() %>%
  ggplot(aes(x = as.factor(published))) +
    geom_bar(color = "black") +
  labs(x = "Published Study?",
       y = "") +
  theme_bw()
```

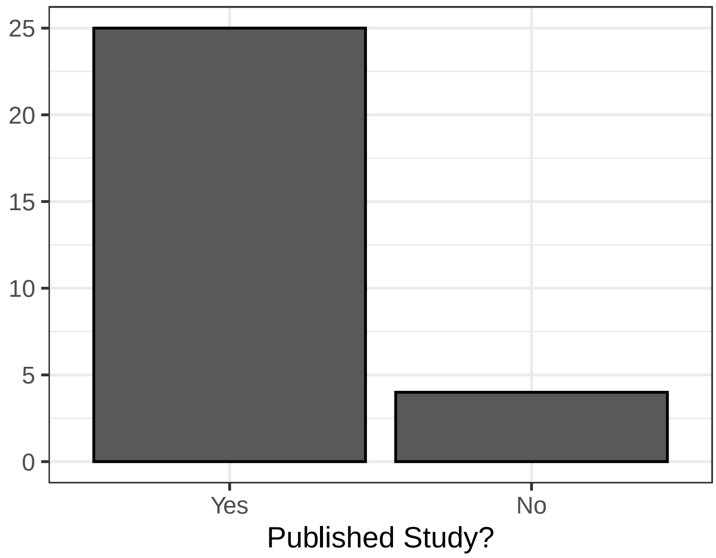


Figure 2: Was the study published?

D.3 Electoral System

Our sample differs considerably in regards to research design. One remarkable difference is that several authors apply the logics of the *law of 1/n*, which was built with majoritarian systems in mind, to non-majoritarian democracies. In the sample, 14 of the papers study *Majoritarian* systems while 15 study *Non-Majoritarian* electoral systems.[2]

[2]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 geographically concentrated. The concentration facilitates politicians to use pork-barrel projects to captivate their electoral supporters.

```r
dat %>%
  select(id, elecsys2) %>%
  unique() %>%
  ggplot(aes(x=as.factor(elecsys2))) +
    geom_bar(color = "black") +
  labs(x = "Electoral Systems",
       y = "") +
  theme_bw()
```

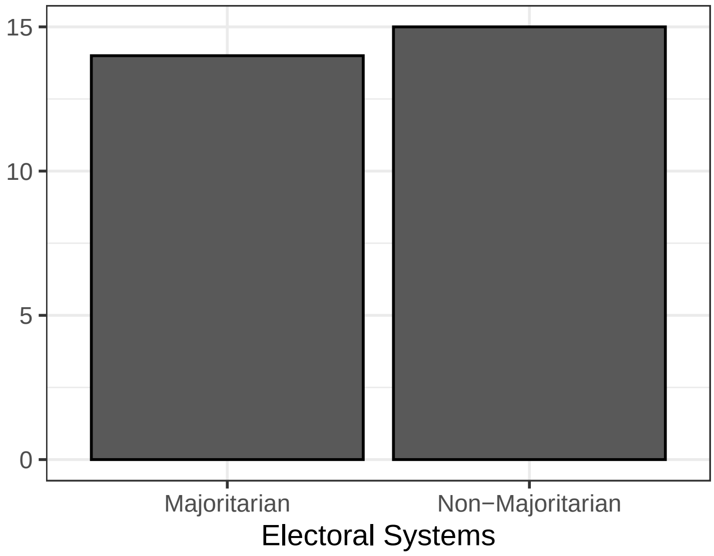


Figure 3: Electoral Systems

D.4 Dependent Variables

The outcome variables included in the paper are:

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

```r
dat %>%
  select(id, depvar2) %>%
  unique() %>%
  mutate(depvar2 = factor(depvar2,
                          labels = c("Per Capita Expenditure",
                                     "Percentage GDP Government Expenditure",
```

```
                                        "Log Per Capita Expenditure"))) %>%
  ggplot(aes(x = depvar2)) +
    geom_bar(color = "black") +
  labs(x = "Dependent Variables",
       y = "") +
  coord_flip() +
  theme_bw()
```

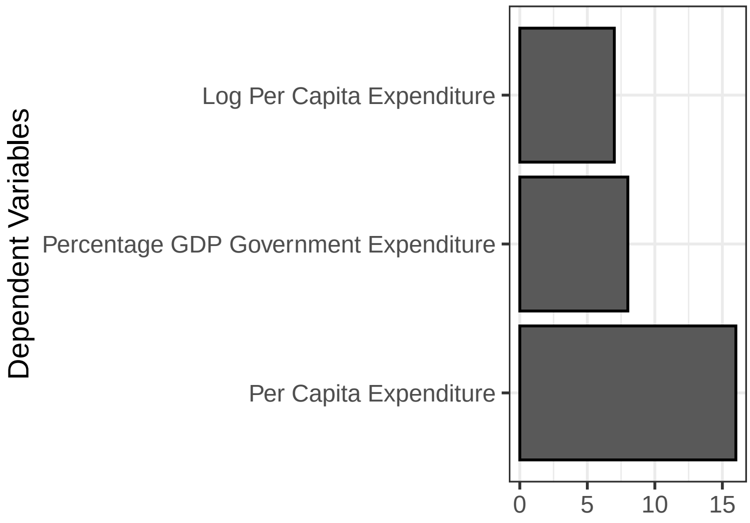


Figure 4: Dependent variables across the law of 1/n studies

D.5 Independent Variables

Most of papers in our sample analyse the number of legislators in the lower house (23). The second most frequent independent variable is the number of legislators in the upper house (12). Finally, the minority of papers use the natural log of the number of legislators in the lower house as an independent variable (5). As we noted above, some papers had multiple coefficients, and thus the total number of coefficients is 32, while the number of papers is only 29.

```
dat %>%
  select(id, indepvar2) %>%
  unique() %>%
  mutate(indepvar2 = factor(indepvar2, labels = c("K (upper house legislators)",
                                                  "N (lower house legislators)",
                                                  "Log N"))) %>%
  ggplot(aes(x = indepvar2)) +
```

```
  geom_bar(color = "black") +
 labs(x = "Independent Variables",
      y = "") +
 coord_flip() +
 theme_bw()
```

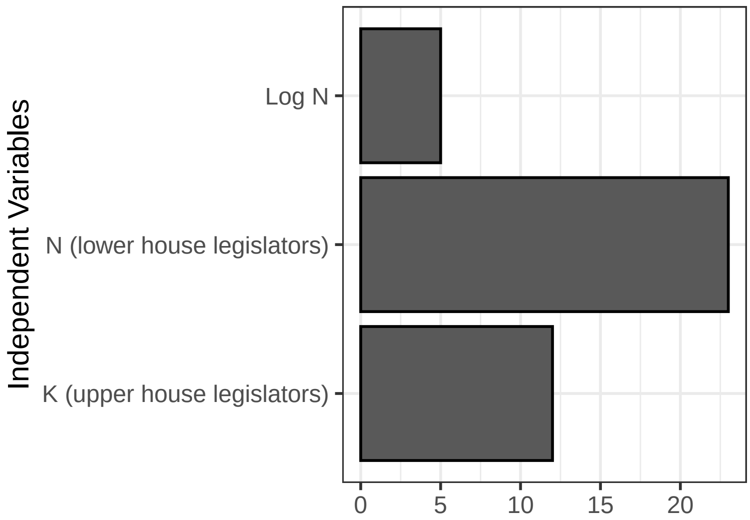


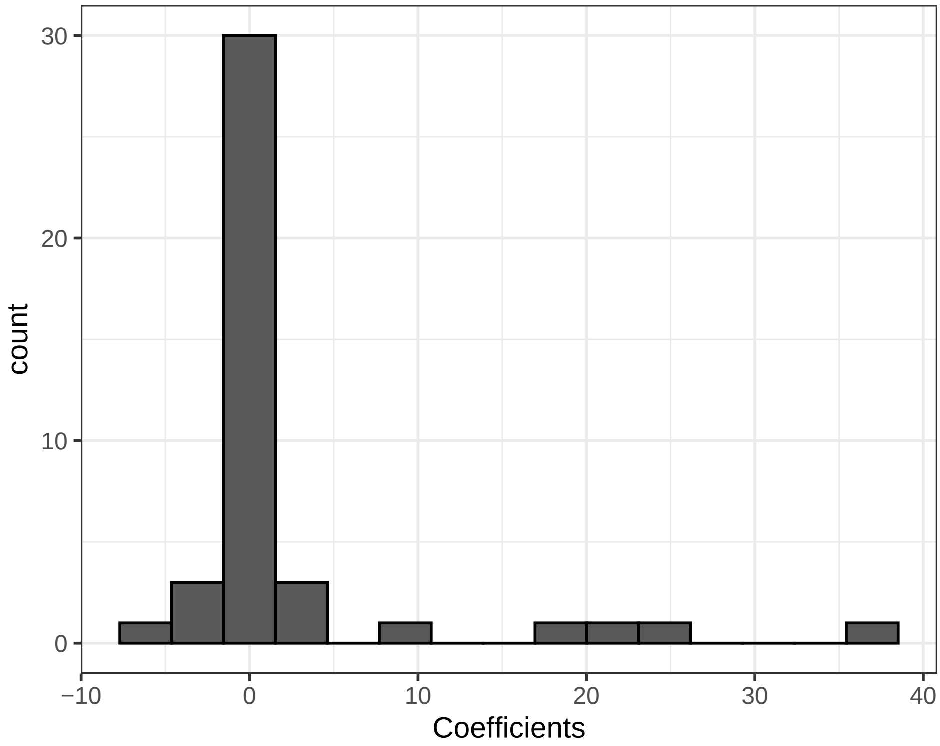
Figure 5: Independent variables across the law of 1/n studies

D.6 Histogram of the Coefficients and the Standard Errors

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

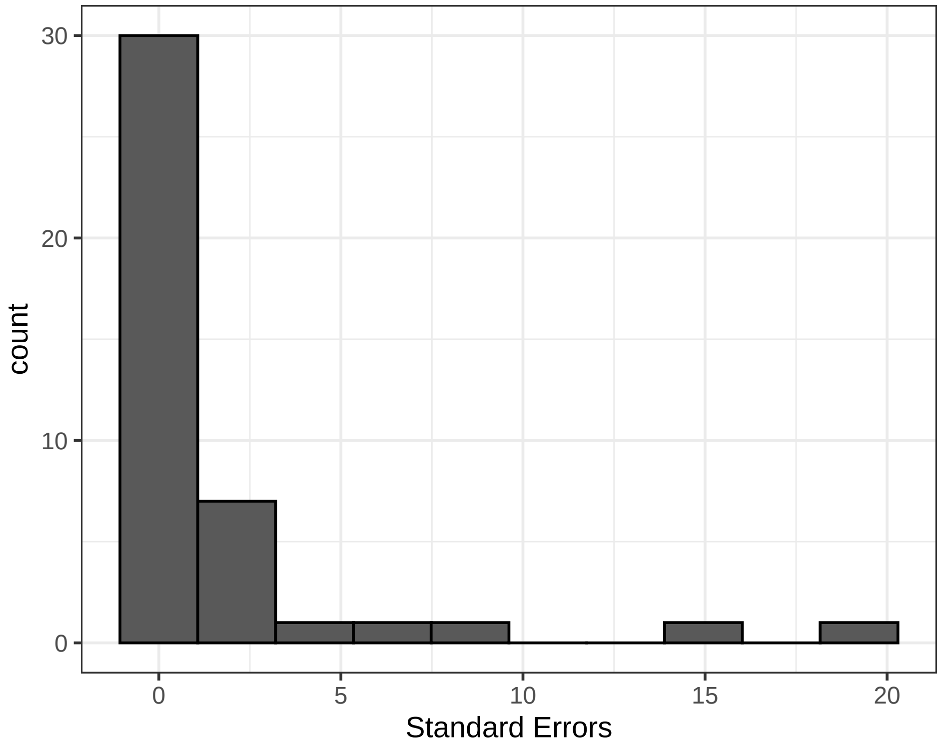
Coefficients:

```
dat %>%
 ggplot(aes(x = coef)) +
 geom_histogram(bins = 15, color = "black") +
 labs(x = "Coefficients") +
 theme_bw()
```



Standard errors:

```
dat %>%
  ggplot(aes(x = SE)) +
  geom_histogram(bins = 10, color = "black") +
  labs(x = "Standard Errors") +
  theme_bw()
```



D.7 Sign Coefficients

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

```r
dat %>%
  ggplot(aes(x=as.factor(scoef))) +
  geom_bar(color = "black") +
  labs(x = "Coefficient Sign",
       y = "") +
  theme_bw()
```

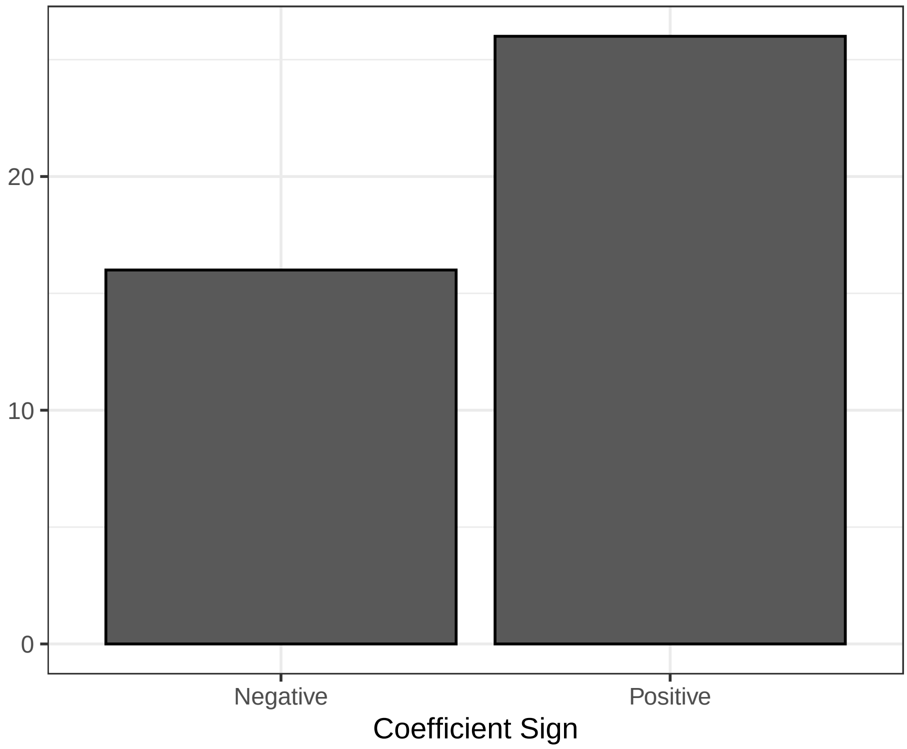


Figure 6: Coefficient Sign

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.

```r
fulldat$usemeta2 <- factor(fulldat$usemeta)
levels(fulldat$usemeta2) <- c("Other Coefficients", "Main Coefficients")
aux <- select(fulldat, usemeta2, indepvar2, year, published,
              elecsys2, method) %>%
  rename(`Independent Variables` = indepvar2,
         `Year` = year,
         `Published work` = published,
         `Electoral system` = elecsys2,
         `Estimation method` = method)

descrTable(~.-usemeta2, aux, y = aux$usemeta2,
           show.p.overall = F, show.all = T)
```

```
##
## --------Summary descriptives table by 'usemeta2'---------
##
## _______________________________________________________________________
##                          [ALL]    Other Coefficients Main Coefficients
##                          N=142          N=100              N=42
## ¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯¯
## Independent Variables:
##     K                   45 (31.7%)     33 (33.0%)        12 (28.6%)
##     N                   79 (55.6%)     55 (55.0%)        24 (57.1%)
##     logN                18 (12.7%)     12 (12.0%)         6 (14.3%)
## Year                    2008 (6.04)   2008 (5.82)        2009 (6.56)
## Published work:
##     Yes                 129 (90.8%)    93 (93.0%)        36 (85.7%)
##     No                  13 (9.15%)     7 (7.00%)          6 (14.3%)
## Electoral system:
##     Majoritarian        68 (47.9%)     46 (46.0%)        22 (52.4%)
##     Non-Majoritarian    74 (52.1%)     54 (54.0%)        20 (47.6%)
## Estimation method:
##     OLS                 47 (33.1%)     36 (36.0%)        11 (26.2%)
##     PANEL               74 (52.1%)     51 (51.0%)        23 (54.8%)
```

```
##     IV                  12 (8.45%)      7 (7.00%)          5 (11.9%)
##     RDD                  9 (6.34%)      6 (6.00%)          3 (7.14%)
## ----------------------------------------------------------------------------------------
```

F Binomial Tests for Coefficient Signs

The *law of 1/n* posits that we should expect a positive influence of legislature size on public expenditures. A general test of the theory could investigate whether the papers tend to find a higher frequency of positive coefficients in their estimations. 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 result for all papers follows 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 two advantages. First, it is robust to the design of the paper. This is an important feature as papers analyse different countries, samples, and have distinct characteristics, such as whether they were published or not. All these factors increase the levels of study heterogeneity. The binomial test ignores the design discrepancies and focuses on the overall reported effect. Second, this test has the advantage of being straightforward and easy to interpret. It requires very few assumption and has a direct statistical formulation. The disadvantage is that we can extract more information from the articles with meta-regressions, as we see in the next sections.

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

```r
aux <- filter(dat, indepvar2 == "N")
aux2 <- binom.test(table(aux$scoef)[2], sum(table(aux$scoef)), p = 0.5)
aux2
```

```
##
##  Exact binomial test
##
## data:  table(aux$scoef)[2] and sum(table(aux$scoef))
```

```
## number of successes = 11, number of trials
## = 24, p-value = 0.8388
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
##  0.2555302 0.6717919
## sample estimates:
## probability of success
##              0.4583333
```

Under the null hypothesis of $p = 0.5$, we find that 11 studies, out of 24, had a positive sign. The chance of a distribution with $p = 0.5$ generate this sample is equal to p-value = 0.839. 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.

```
aux <- filter(dat, indepvar2 == "logN")
aux2 <- binom.test(table(aux$scoef)[2], sum(table(aux$scoef)), p = 0.5)
aux2
```

```
##
##  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
```

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

Finally, for the number of legislators in the upper house (K), the results are:

```r
aux <- filter(dat, indepvar2=='K')

aux2 <- binom.test(table(aux$scoef)[2], sum(table(aux$scoef)), p=0.5)

aux2
```

```
##

##  Exact binomial test

##

## data:  table(aux$scoef)[2] and sum(table(aux$scoef))

## number of successes = 10, number of trials

## = 12, p-value = 0.03857

## alternative hypothesis: true probability of success is not equal to 0.5

## 95 percent confidence interval:

##  0.5158623 0.9791375

## sample estimates:

## probability of success

##              0.8333333
```

Here we see that 10 out of 12 had a positive sign. The p-value for this test is 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

G.1 Estimation Method

In general terms, there are two main ways to conduct a meta-analysis, either by using fixed effects or by employing 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, in contrast, assumes that there is a distribution of true effects, and that the coefficients 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. **Modelling strategies**: Panel data, Standard OLS, IV, RDD.

These sources of heterogeneity have two implications. First, they make our estimates notably disperse. All but one of our heterogeneity tests are 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 biased. 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 into two components, the true effect that the study with the same specifications as i comes from, θ_i, and a within-study error ε_i:

$$T_i = \theta_i + \varepsilon_i$$

And the random effects model assumes that the θ_i varies from study to study, having a true parameter μ, plus a between-study error, ξ_i:

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

And the random effects model estimates the parameter μ, 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. To empirically implement the random effects model, we need to choose a method to estimate the true effect size variance, τ^2, which in our formulation, represents the variance of ξ_i. We selected the

Restricted Maximum Likelihood Estimator, as the literature regards it the most precise when analysing continuous measures, such as we have in our data.

We combined the three independent variables (N, $\log(N)$, and K) with our dependent variables of interest (Expenditure Per Capita, Log of Expenditure Per Capita, Expenditure as a Percentage of the GDP). This formed a 3×3 table, and in the following pages we present the results for each of these combinations.

G.2 Lower House Size and Expenditure per Capita

```
# Pooling effects analysis -- ExpPC x N
aux <- dat %>%
  filter(indepvar2 == 'N',
         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
## Crowley (2019)               -0.3510
## Lee and Park (2018)          -0.8510
## Lee (2016)                    0.0164
## Lee (2015)                   -2.0130
## Kessler (2014)                0.1740
## Bjedov et al. (2014)         -0.0030
## Baskaran (2013)               0.9740
## Erler (2007)                  3.9300
## Chen and Malhotra (2007)     -2.0400
## Fiorino and Ricciuti (2007)   0.2130
```

```
## Primo (2006)                  -0.8200
## Matsusaka (2005)              -0.9600
## Gilligan and Matsusaka (2001) -0.0140
## Schaltegger and Feld (2009)    0.0010
## Ricciuti (2003)                1.7650
##                                          95%-CI
## Crowley (2019)                [-1.8112;  1.1092]
## Lee and Park (2018)           [-3.5851;  1.8831]
## Lee (2016)                    [-2.5570;  2.5898]
## Lee (2015)                    [-4.5727;  0.5467]
## Kessler (2014)                [ 0.0074;  0.3406]
## Bjedov et al. (2014)          [-0.0226;  0.0166]
## Baskaran (2013)               [-0.1212;  2.0692]
## Erler (2007)                  [ 1.6172;  6.2428]
## Chen and Malhotra (2007)      [-4.6468;  0.5668]
## Fiorino and Ricciuti (2007)   [ 0.1777;  0.2483]
## Primo (2006)                  [-1.1924; -0.4476]
## Matsusaka (2005)              [-1.3128; -0.6072]
## Gilligan and Matsusaka (2001) [-0.0375;  0.0095]
## Schaltegger and Feld (2009)   [-0.0010;  0.0030]
## Ricciuti (2003)               [ 0.0638;  3.4662]
##                               %W(random)
## Crowley (2019)                       4.4
## Lee and Park (2018)                  1.7
## Lee (2016)                           1.9
## Lee (2015)                           1.9
## Kessler (2014)                      11.0
## Bjedov et al. (2014)                11.2
## Baskaran (2013)                      5.9
## Erler (2007)                         2.3
## Chen and Malhotra (2007)             1.9
## Fiorino and Ricciuti (2007)         11.2
## Primo (2006)                        10.2
```

```
## Matsusaka (2005)                    10.3
## Gilligan and Matsusaka (2001)       11.2
## Schaltegger and Feld (2009)         11.2
## Ricciuti (2003)                      3.6
##
## Number of studies combined: k = 15
##
##                           SMD            95%-CI
## Random effects model -0.0374 [-0.5691; 0.4942]
## Prediction interval           [-1.4238; 1.3490]
##                           t p-value
## Random effects model -0.15  0.8822
## Prediction interval
##
## Quantifying heterogeneity:
##  tau^2 = 0.3504 [0.2108; 4.0970]; tau = 0.5920 [0.4591; 2.0241];
##  I^2 = 93.5% [90.8%; 95.4%]; H = 3.92 [3.30; 4.66]
##
## Test of heterogeneity:
##       Q d.f.  p-value
##  215.10   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
```

The forest plot:

```r
build_forest(mod, NULL)
```

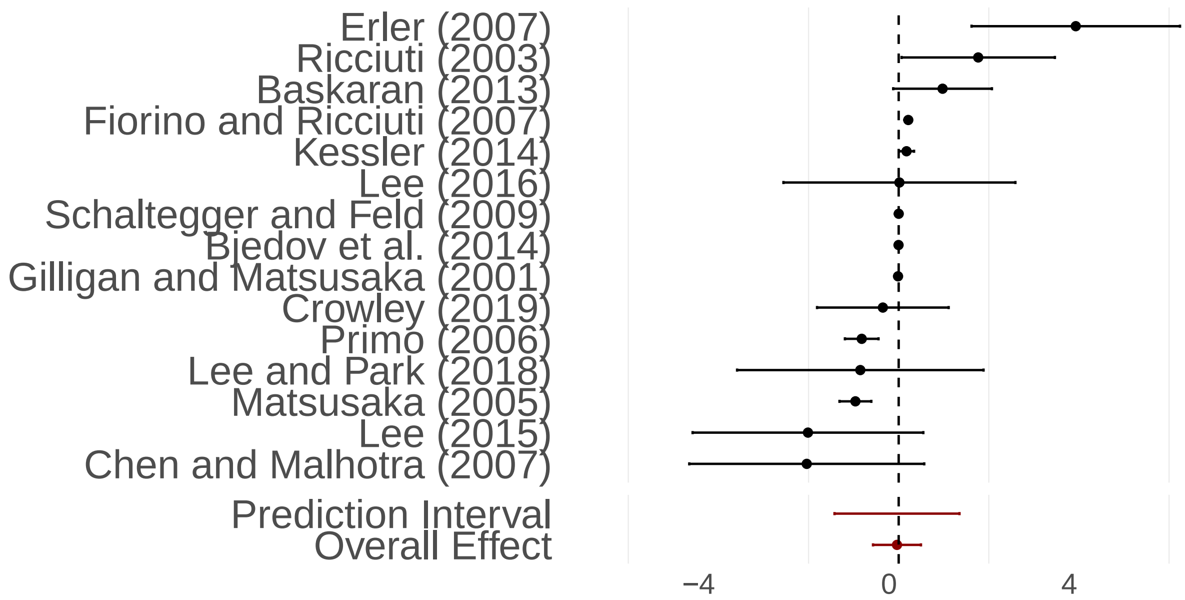


Figure 7: Effect of lower houses size (N) on Per Capita Expenditure (ExpPC)

Highlights:

1. The results are highly heterogeneous: I^2 = 93.49.
2. The estimated SMD in the random effects model is g = -0.04 (SE = 0.248).
3. The prediction interval ranges from -1.42 to 1.35. Therefore, it emcompasses zero.

G.2.1 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.

```
build_forest_het(mod, capt = NULL, hetvar = aux$elecsys2)
```

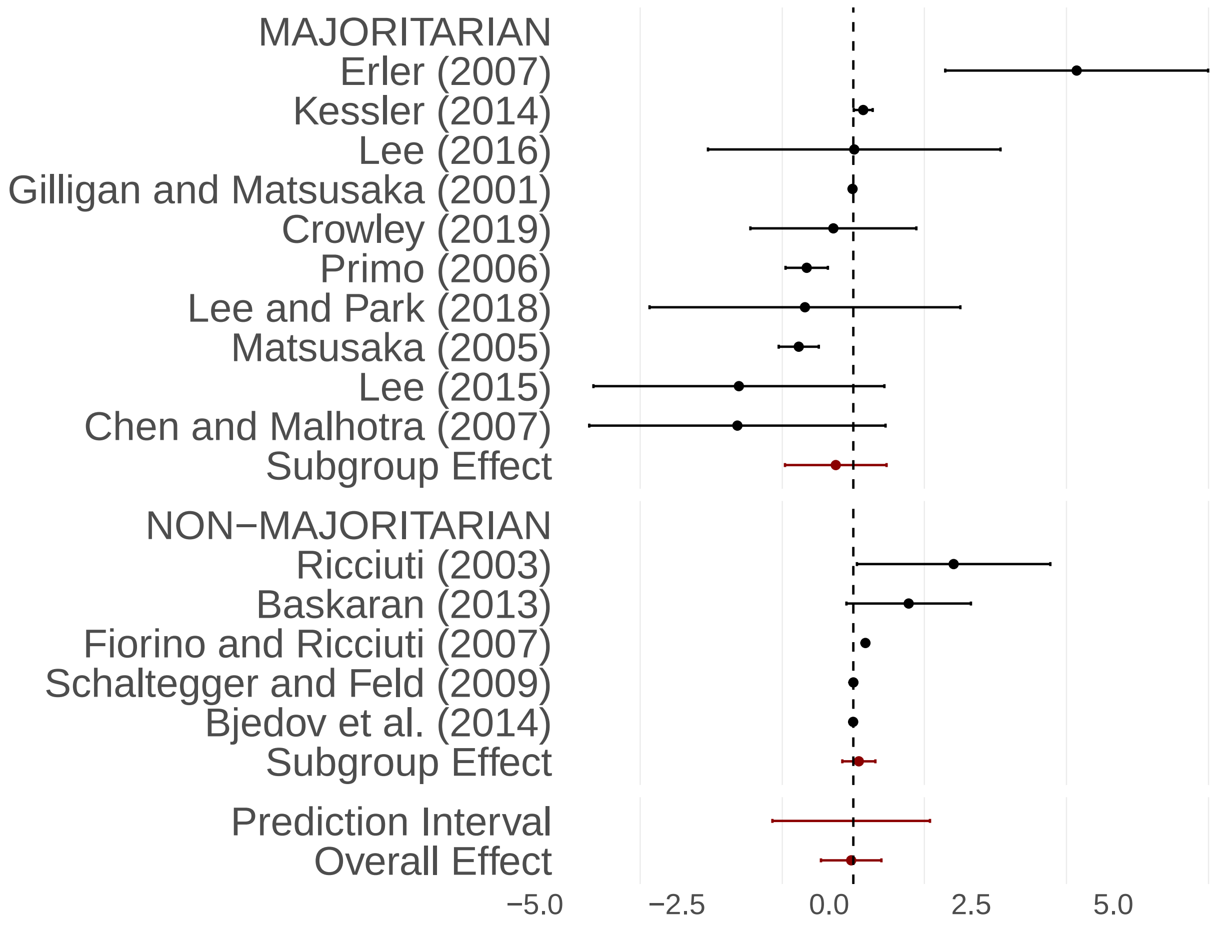


Figure 8: Subgroup Analysis of (N) x (ExpPC), controlling by electoral system

Therefore, we can see that the hypothesis that majoritarian systems produce systematic positive effects was disproved. Both are non-significant, and they reassure us that the absense of effect is not caused by pooling multiple types of electoral systems.

G.3 Log Lower House Size and Expenditure per Capita

There are no studies that have per capita expenditure as the dependent variable and log of lower house size as the treatment variable.

G.4 Upper House Size and Expenditure per Capita

Now, we look into 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',
```

```r
         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
## Crowley (2019)                   8.2100
## Lee and Park (2018)             19.7400
## Lee (2016)                      38.4400
## Lee (2015)                      20.3900
## Bradbury and Stephenson (2009)   0.6240
## Chen and Malhotra (2007)        26.0900
## Primo (2006)                     0.9700
## Gilligan and Matsusaka (2001)    0.1510
## Ricciuti (2003)                 -3.9240
##                                              95%-CI
## Crowley (2019)                  [  0.2702; 16.1498]
## Lee and Park (2018)             [  3.2645; 36.2155]
## Lee (2016)                      [  0.7499; 76.1301]
## Lee (2015)                      [-10.2638; 51.0438]
## Bradbury and Stephenson (2009) [  0.2295;  1.0185]
## Chen and Malhotra (2007)        [ 11.4883; 40.6917]
## Primo (2006)                    [ -0.4804;  2.4204]
## Gilligan and Matsusaka (2001)   [ -0.0136;  0.3156]
## Ricciuti (2003)                 [ -6.4955; -1.3525]
##                                 %W(random)
## Crowley (2019)                        13.1
```

```
## Lee and Park (2018)                   8.4
## Lee (2016)                            2.8
## Lee (2015)                            3.9
## Bradbury and Stephenson (2009)       15.7
## Chen and Malhotra (2007)              9.3
## Primo (2006)                         15.6
## Gilligan and Matsusaka (2001)        15.7
## Ricciuti (2003)                      15.4
##
## Number of studies combined: k = 9
##
##                         SMD              95%-CI
## Random effects model 6.7270 [ -2.2272; 15.6811]
## Prediction interval         [-16.4850; 29.9389]
##                         t p-value
## Random effects model 1.73  0.1214
## Prediction interval
##
## Quantifying heterogeneity:
##  tau^2 = 81.2829 [22.4951; 628.4149]; tau = 9.0157 [4.7429; 25.0682];
##  I^2 = 81.2% [65.4%; 89.8%]; H = 2.31 [1.70; 3.14]
##
## Test of heterogeneity:
##      Q d.f.  p-value
##  42.66    8 < 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:

```r
build_forest(mod, NULL)
```

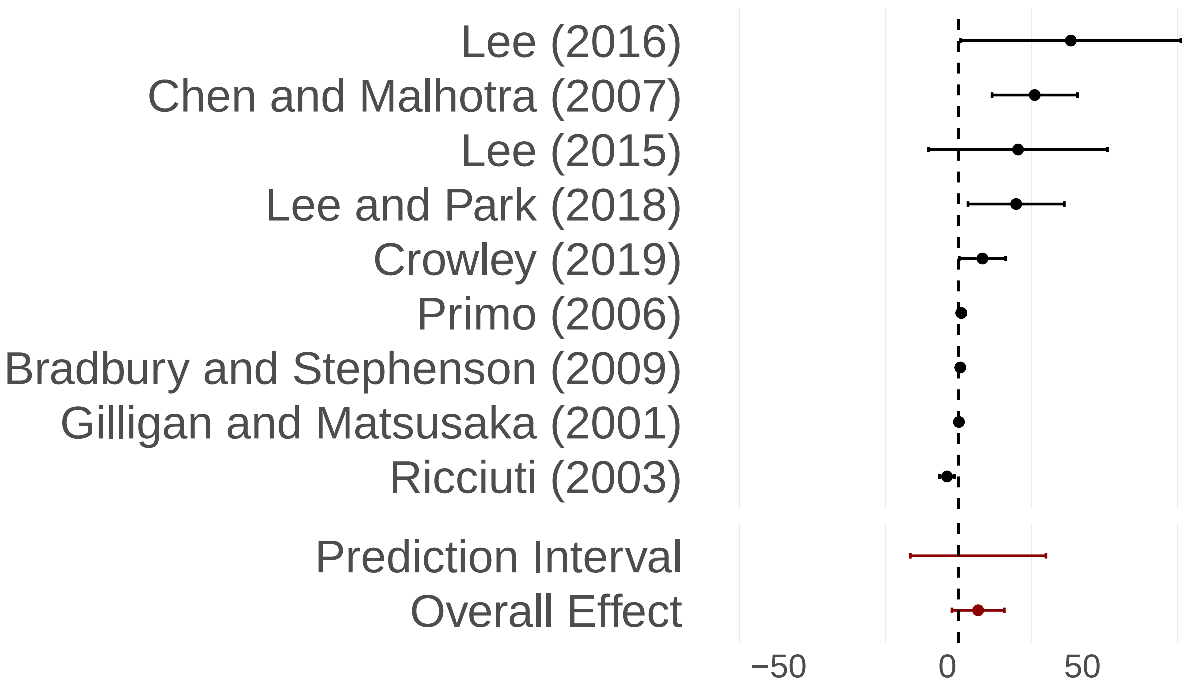


Figure 9: Effect of upper house size (K) on the per capita government expenditure (ExpPC)

Highlights:

1. The results are highly heterogeneous: $I^2 = 81.25$.
2. The estimated SMD in the random effects model is $g = 6.73$ ($SE = 3.883$).
3. The prediction interval ranges from -16.48 to 29.94. Therefore, it emcompasses zero.

G.5 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.

```r
# Pooling effects analysis -- logExpPC x N

aux <- dat %>%

  filter(indepvar2 == 'N',

         depvar2 == 'logExpPC')


mod <- metagen(

  coef, SE, data=aux,

  studlab=paste(authoryear),
```

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

mod
```

```
##                                  SMD
## Lewis (2019)              -0.1740
## Höhmann (2017)            -0.0300
## Drew and Dollery (2017)    0.0770
## Pettersson-Lidbom (2012) -0.1590
##                                          95%-CI
## Lewis (2019)              [-0.2450; -0.1030]
## Höhmann (2017)            [-0.0496; -0.0104]
## Drew and Dollery (2017)   [ 0.0221;  0.1319]
## Pettersson-Lidbom (2012) [-0.2394; -0.0786]
##                           %W(random)
## Lewis (2019)                    24.3
## Höhmann (2017)                  26.6
## Drew and Dollery (2017)         25.3
## Pettersson-Lidbom (2012)        23.7
##
## Number of studies combined: k = 4
##
##                           SMD             95%-CI
## Random effects model -0.0686 [-0.2560; 0.1188]
## Prediction interval          [-0.6179; 0.4807]
##                           t p-value
## Random effects model -1.17  0.3282
## Prediction interval
```

```
##
## 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
```

The forest plot is as follows:

```r
build_forest(mod, NULL)
```

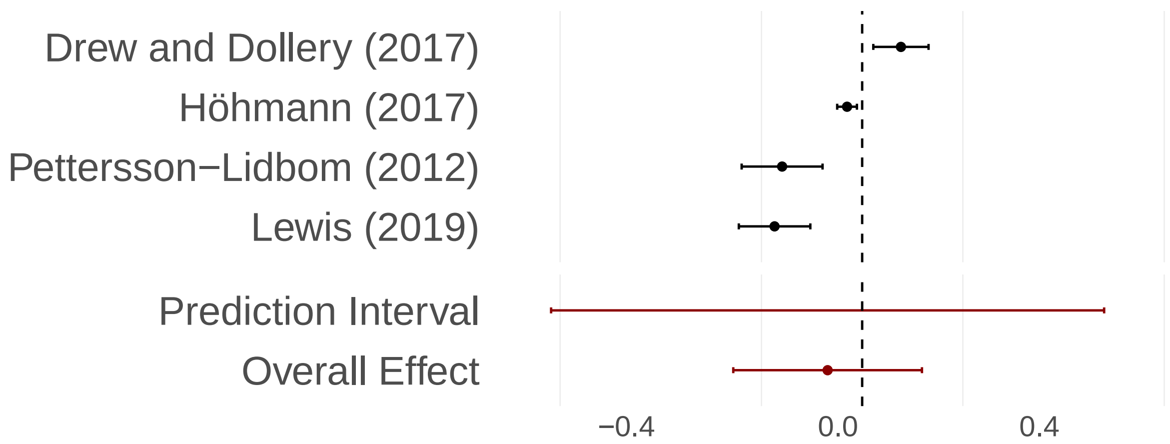


Figure 10: Effect of lower houses size (N) on log of per capita expenditure (logExpPC)

Highlights:

1. The results are highly heterogeneous: I^2 = 92.52.
2. The estimated SMD in the random effects model is g = -0.07 (SE = 0.059).
3. The prediction interval ranges from -0.62 to 0.48. Therefore, it emcompasses zero.

G.6 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
## MacDonald (2008) 0.1360 [0.0447; 0.2273]
## Baqir (2002)     0.1127 [0.0396; 0.1858]
## Baqir (1999)     0.3020 [0.2269; 0.3771]
##                  %W(random)
## MacDonald (2008)       31.9
## Baqir (2002)           34.2
## Baqir (1999)           33.9
##
## Number of studies combined: k = 3
##
##                          SMD            95%-CI
## Random effects model 0.1844 [-0.0738; 0.4425]
## Prediction interval         [-1.2580; 1.6267]
##                          t p-value
```

```
## Random effects model 3.07  0.0916
## Prediction interval
##
## 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
```

The forest plot is available below:

```r
build_forest(mod, NULL)
```

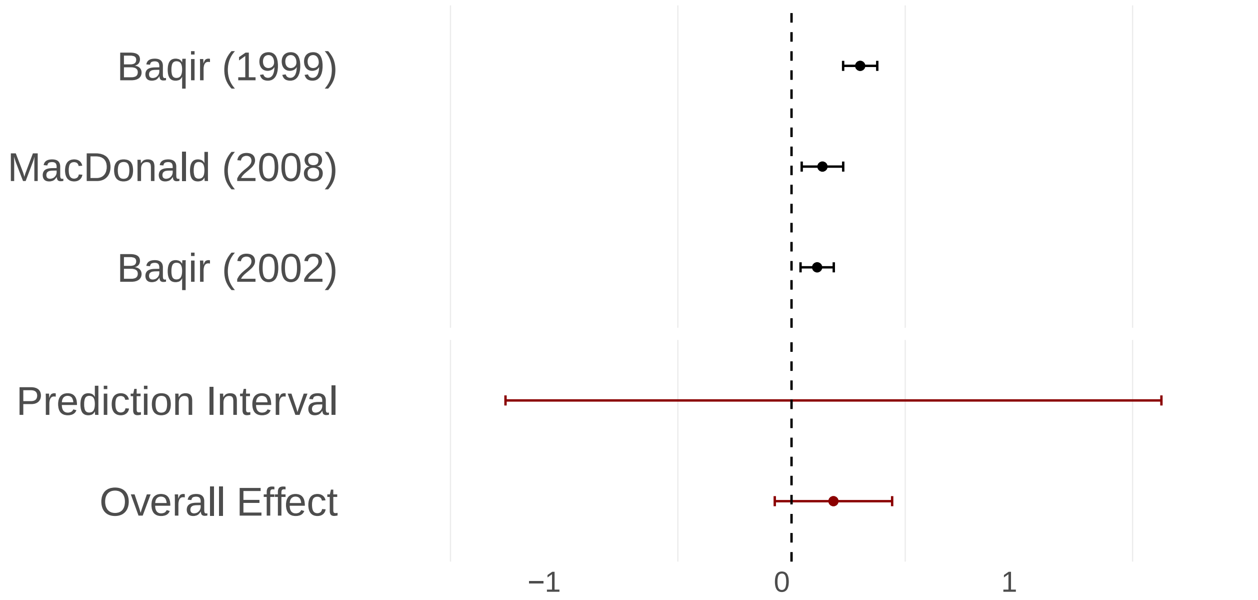


Figure 11: Effect of log lower houses size (logN) on the log of per capita government expenditure (logExpPC)

Highlights:

1. The results are highly heterogeneous: $I^2 = 85.9$.
2. The estimated SMD in the random effects model 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.7 Log of Upper House Size and Log of Expenditure Per Capita

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

G.8 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 treatment variable.

```
# 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
## Bjedov et al. (2014)      -0.0040
## Maldonado (2013)          -0.0609
## Mukherjee (2003)           0.0030
## Bradbury and Crain (2001)  0.0036
## Ricciuti (2004)            0.0140
##                                        95%-CI
```

```
## Bjedov et al. (2014)      [-0.0432;  0.0352]
## Maldonado (2013)          [-0.0838; -0.0380]
## Mukherjee (2003)          [ 0.0010;  0.0050]
## Bradbury and Crain (2001) [ 0.0008;  0.0065]
## Ricciuti (2004)           [-0.0095;  0.0375]
##                           %W(random)
## Bjedov et al. (2014)            15.1
## Maldonado (2013)                19.5
## Mukherjee (2003)                23.0
## Bradbury and Crain (2001)       23.0
## Ricciuti (2004)                 19.4
##
## Number of studies combined: k = 5
##
##                            SMD            95%-CI
## Random effects model -0.0083 [-0.0450; 0.0285]
## Prediction interval          [-0.1054; 0.0889]
##                           t p-value
## Random effects model -0.62  0.5667
## Prediction interval
##
## 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
```

Below, you may find the forest plot:

```r
build_forest(mod, NULL)
```

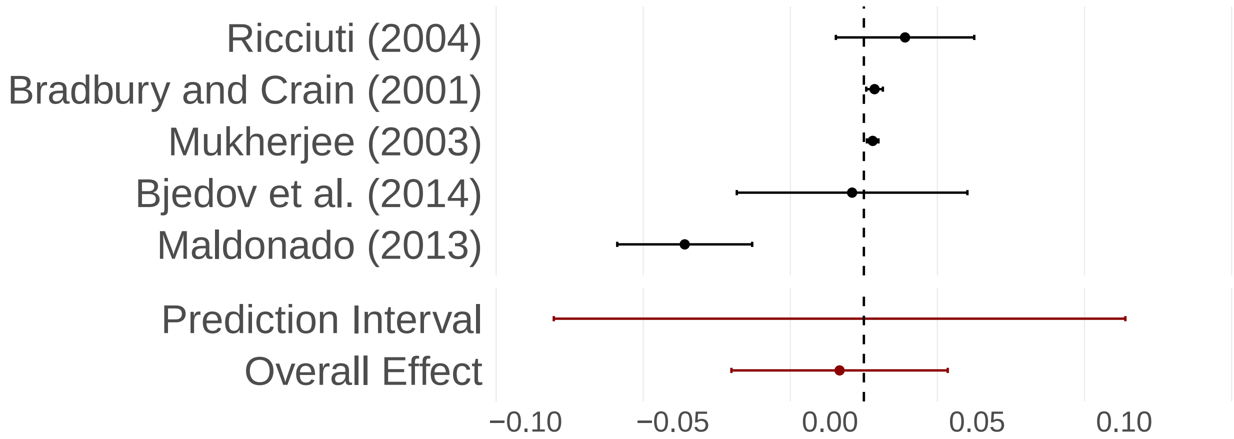


Figure 12: Effect of lower houses size (N) on percentage of public expenditure GDP (PCTGDP)

Highlights:

1. The results are highly heterogeneous: $I^2 = 87.08$.
2. The estimated SMD in the random effects model is g = -0.01 (SE = 0.013).
3. The prediction interval ranges from -0.11 to 0.09. Therefore, it emcompasses zero.

G.9 Log Lower House Size and Expenditure as Percentage of GDP

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

```r
# 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,
```

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

mod
```

```
##                          SMD             95%-CI
## Baqir (1999)          2.0660 [ 1.4887; 2.6433]
## Lledo (2003)         -4.6900 [-9.9427; 0.5627]
## Stein et al. (1998)  0.0109 [-0.0171; 0.0389]
##                      %W(random)
## Baqir (1999)               40.8
## Lledo (2003)               17.7
## Stein et al. (1998)        41.5
##
## Number of studies combined: k = 3
##
##                          SMD               95%-CI
## Random effects model 0.0203 [ -7.1961;  7.2367]
## Prediction interval         [-36.2058; 36.2465]
##                          t p-value
## Random effects model 0.01  0.9914
## Prediction interval
##
## 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
```

The forest plot follows below:

```
build_forest(mod, NULL)
```

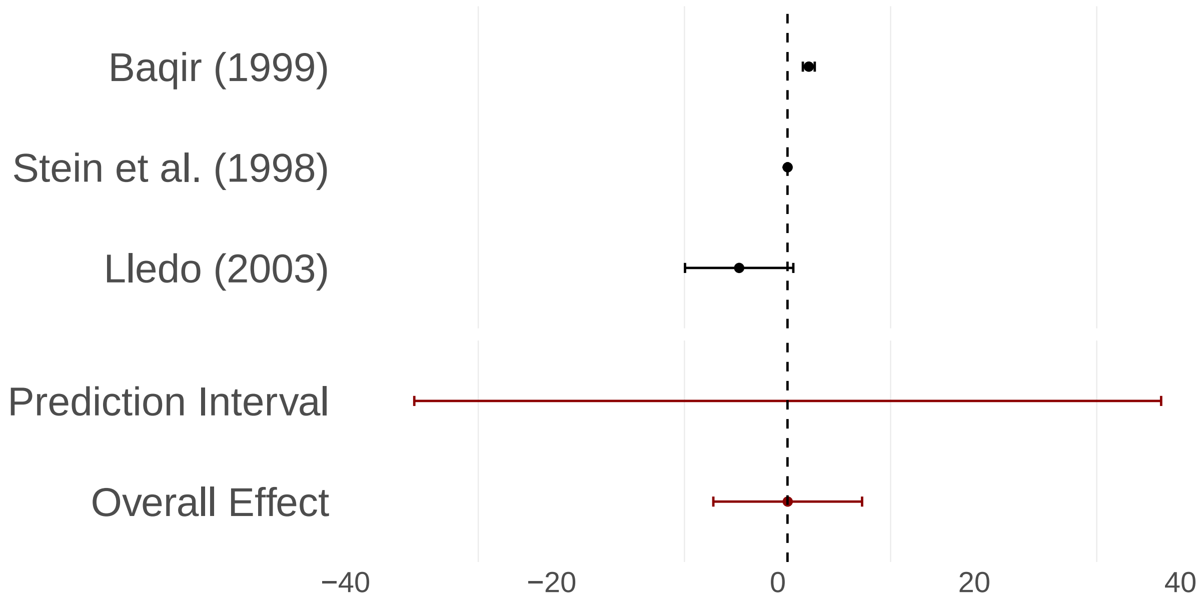


Figure 13: Effect of log lower houses size (logN) on the GDP share of public expenditure (PCTGDP)

Highlights:

1. The results are highly heterogeneous: $I^2 = 96.13$.
2. The estimated SMD in the random effects model is $g = 0.02$ ($SE = 1.677$).
3. The prediction interval ranges from -36.21 to 36.25. Therefore, it emcompasses zero.

G.10 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 <- dat %>%
```

```r
  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
## Maldonado (2013)           -0.0400
## Bradbury and Crain (2001)   0.0126
## Ricciuti (2004)             0.0160
##                                         95%-CI
## Maldonado (2013)           [-0.0659; -0.0141]
## Bradbury and Crain (2001) [ 0.0010;  0.0243]
## Ricciuti (2004)            [-0.0075;  0.0395]
##                           %W(random)
## Maldonado (2013)                31.3
## Bradbury and Crain (2001)       36.4
## Ricciuti (2004)                 32.3
##
## Number of studies combined: k = 3
##
##                            SMD            95%-CI
## Random effects model -0.0027 [-0.0793; 0.0738]
## Prediction interval           [-0.4284; 0.4229]
##                            t p-value
## Random effects model -0.15  0.8915
## Prediction interval
```

```
##
## 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
```

The forest plot follows below:

```r
build_forest(mod, NULL)
```

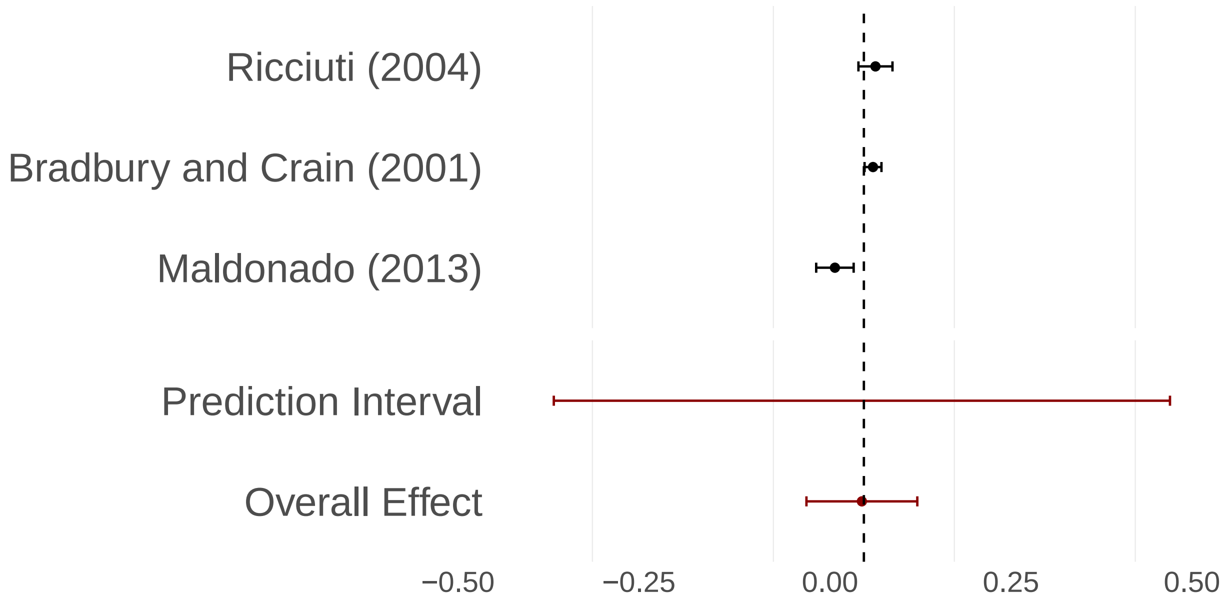


Figure 14: Effect of upper house size (K) on the public expenditure share of the GDP (PCTGDP)

Highlights:

1. The results are highly heterogeneous: $I^2 = 85.79$.

2. The estimated SMD in the random effects model is $g = 0$ ($SE = 0.018$).
3. The prediction interval ranges from -0.43 to 0.42. Therefore, it emcompasses zero.

G.11 Lower House Size and Expenditure per Capita (IV)

```
# Pooling effects analysis -- ExpPC x N (IV only)
aux <- dat %>%
  filter(indepvar2 == 'N',
         depvar2 == 'ExpPC',
         method %in% c('IV'))

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
## Lee (2015)                   -2.0130
## Baskaran (2013)               0.9740
## Fiorino and Ricciuti (2007)   0.2130
## Matsusaka (2005)             -0.9600
##                                          95%-CI
## Lee (2015)                   [-4.5727;  0.5467]
## Baskaran (2013)              [-0.1212;  2.0692]
## Fiorino and Ricciuti (2007) [ 0.1777;  0.2483]
## Matsusaka (2005)             [-1.3128; -0.6072]
##                              %W(random)
## Lee (2015)                         10.8
## Baskaran (2013)                    24.1
```

```
## Fiorino and Ricciuti (2007)       33.2
## Matsusaka (2005)                  31.9
##
## Number of studies combined: k = 4
##
##                          SMD            95%-CI
## Random effects model -0.2192 [-1.9893; 1.5509]
## Prediction interval          [-4.8022; 4.3638]
##                          t p-value
## Random effects model -0.39  0.7198
## Prediction interval
##
## Quantifying heterogeneity:
##  tau^2 = 0.8252 [0.1412; 23.0395]; tau = 0.9084 [0.3758; 4.7999];
##  I^2 = 93.6% [86.8%; 96.9%]; H = 3.95 [2.75; 5.67]
##
## Test of heterogeneity:
##      Q d.f.  p-value
##  46.83    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:

```r
build_forest(mod, NULL)
```

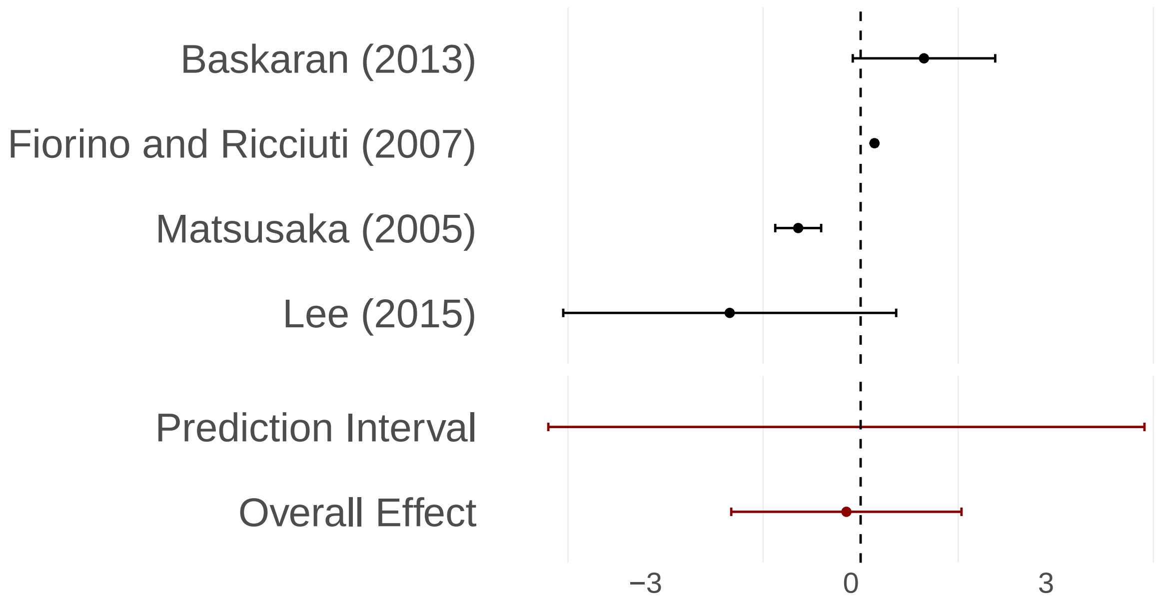


Figure 15: Effect of lower houses size (N) on Per Capita Expenditure (ExpPC)

Highlights:

1. The results are highly heterogeneous: I^2 = 93.49.
2. The estimated SMD in the random effects model is g = -0.04 (SE = 0.248).
3. The prediction interval ranges from -1.42 to 1.35. Therefore, it emcompasses zero.

G.11.1 Regression Method Subgroup Analysis

Over time, the literature evolved to use causally identified techniques for determine the effect of legislature size on the expendutere per capita. To study whether the method had an effect on the estimated coefficients, we fit a subgroup analysis using the method employed in each paper.

```
aux <- dat %>%
  filter(indepvar2 == 'N',
         depvar2 == 'ExpPC')


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

```r
         prediction=TRUE,

         sm="SMD")


build_forest_het(mod, capt = NULL, hetvar = aux$method)
```

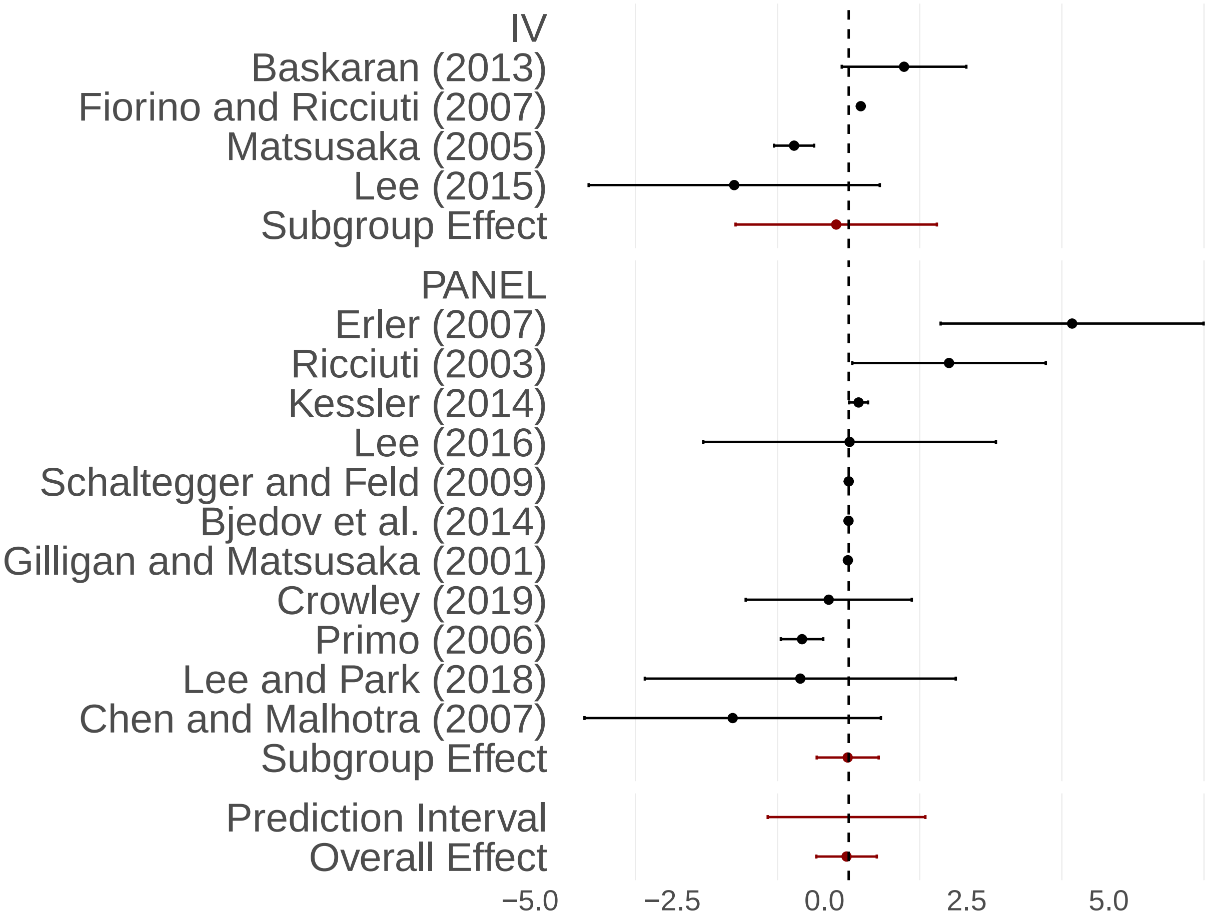


Figure 16: Subgroup Analysis of (N) x (ExpPC), controlling by regression methods

Although all methods generate a null effect, the IV method seems to be well distributed, with two papers with positive effects and two papers negative displaying negative effects. The random effects model for the subgroup is 0.22, which is negative but non-significant. Improve the estimation technique, for the case of IVs, render still a null effect of legislature size on per capita government expenditure.

G.12 Lower House Size and Log of Expenditure per Capita (RDD)

```r
# Pooling effects analysis -- logExpPC x N (RDD only)

aux <- dat %>%

  filter(indepvar2 == 'N',
```

```r
         depvar2 == 'logExpPC',
         method == 'RDD')

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
## Lewis (2019)             -0.1740
## Höhmann (2017)           -0.0300
## Pettersson-Lidbom (2012) -0.1590
##                                      95%-CI
## Lewis (2019)             [-0.2450; -0.1030]
## Höhmann (2017)           [-0.0496; -0.0104]
## Pettersson-Lidbom (2012) [-0.2394; -0.0786]
##                          %W(random)
## Lewis (2019)                   31.8
## Höhmann (2017)                 38.0
## Pettersson-Lidbom (2012)       30.3
##
## Number of studies combined: k = 3
##
##                          SMD             95%-CI
## Random effects model -0.1148 [-0.3174; 0.0878]
## Prediction interval          [-1.2757; 1.0461]
##                          t p-value
## Random effects model -2.44  0.1350
## Prediction interval
```

```
##
## Quantifying heterogeneity:
##  tau^2 = 0.0061 [0.0011; 0.2470]; tau = 0.0783 [0.0339; 0.4970];
##  I^2 = 91.2% [77.1%; 96.6%]; H = 3.36 [2.09; 5.41]
##
## Test of heterogeneity:
##      Q d.f.  p-value
##  22.64    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
```

Forest plot:

```
build_forest(mod, NULL)
```

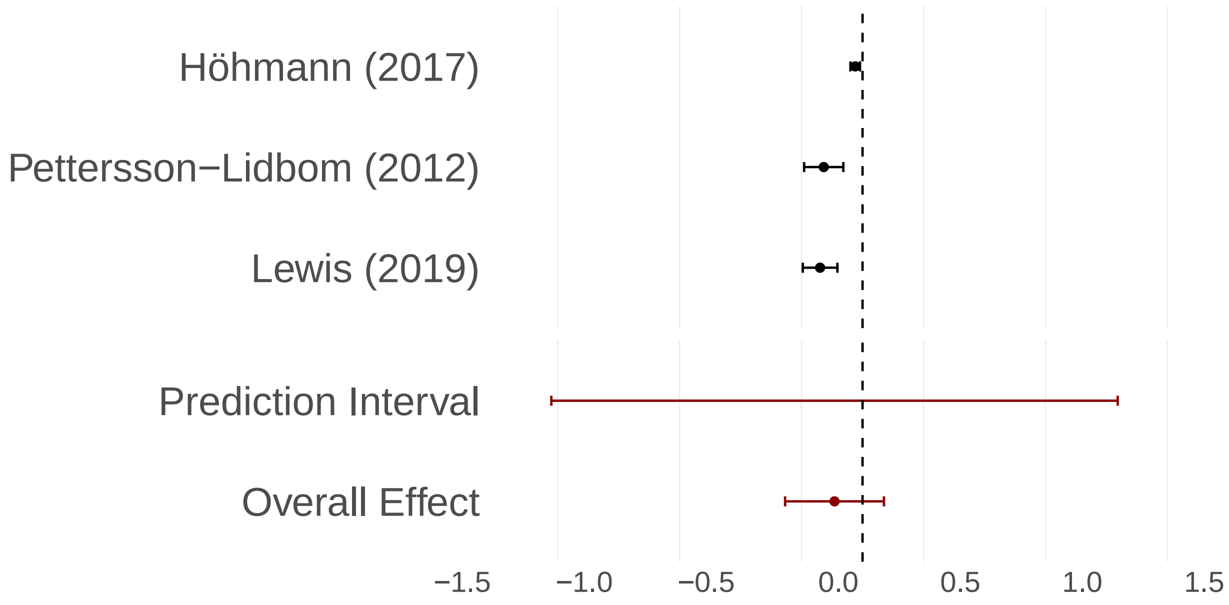


Figure 17: Effect of lower houses size (N) on Per Capita Expenditure (ExpPC)

Highlights:

1. The results are highly heterogeneous: $I^2 = 92.52$.

2. The estimated SMD in the random effects model is g = -0.07 (SE = 0.059).
3. The prediction interval ranges from -0.62 to 0.48. Therefore, it emcompasses zero.

G.12.1 Regression Method Subgroup Analysis

Over time, the literature evolved to use causally identified techniques for determine the effect of legislature size on the log of expendure per capita. To study whether the method had an effect on the estimated coefficients, we fit a subgroup analysis using the method employed in each paper.

```r
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")

build_forest_het(mod, capt = NULL, hetvar = aux$method)
```

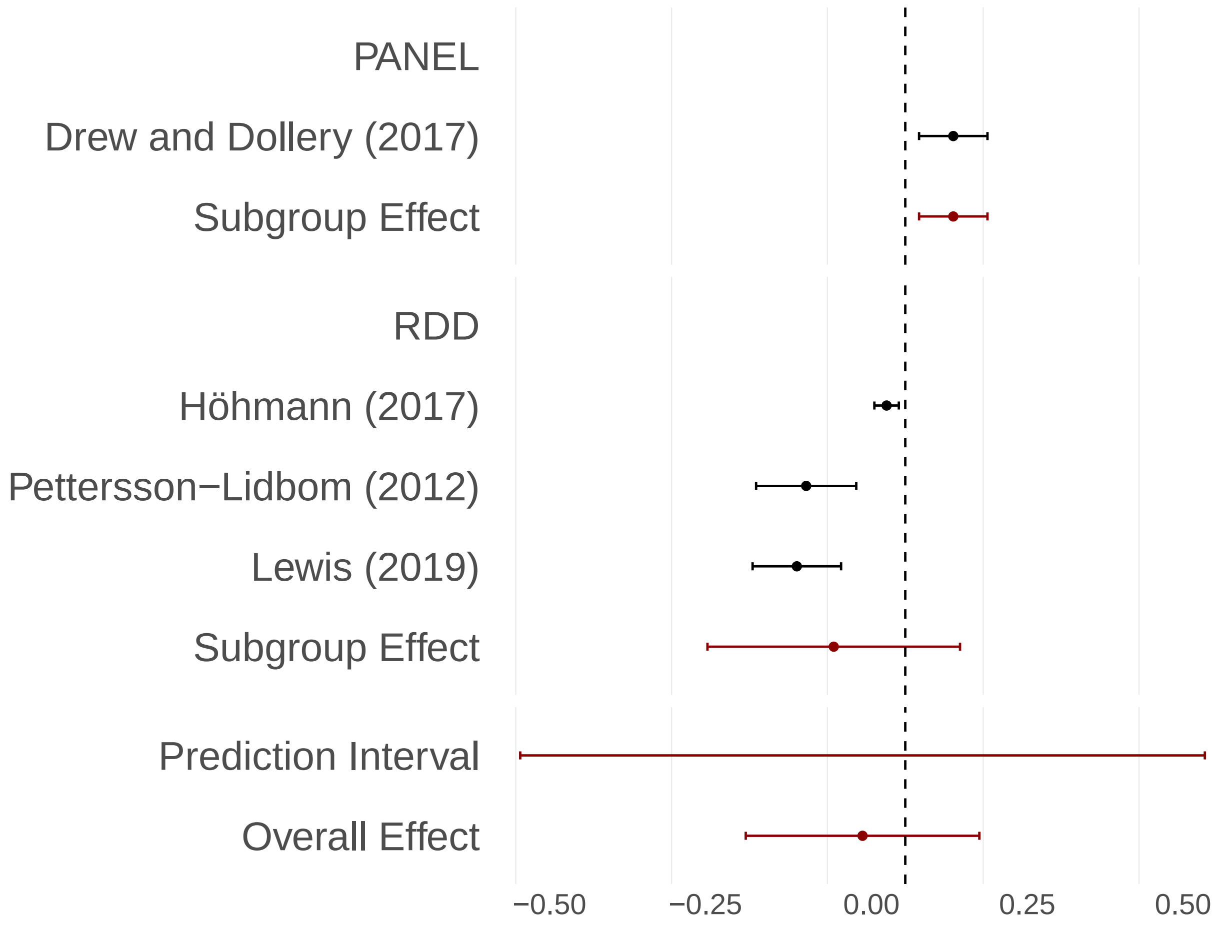


Figure 18: Subgroup Analysis of (N) x (logExpPC), controlling by the regression methods

For the RDD subgroup analysis, we can see a much clearer picture. All estimates are negative, and the subgroup effect is -0.11. This suggests that improving the estimation renders to a negative result that is likely to hold conditional on more papers using the technique displaying the same effect size and sign.

H Code for the Main Graphs

H.1 Figure 1

```
pdf('../graphs/graph1.pdf', width = 16, height = 11)
ggarrange(f1,f3,f5,f4,f6,f2,f7, align = 'hv')
dev.off()
```

```
## pdf
##   2
```

H.2 Figure 2

```
pdf('../graphs/graph2.pdf', width = 12, height = 6)
ggarrange(f8, f9, align = 'hv')
dev.off()
```

```
## pdf
##   2
```

I Meta-Analysis (All Coefficients)

I.1 Lower House Size and Expenditure Per Capita

Here we estimate the relationship between expenditure per capita as a dependent variable, and the lower house size as the independent variable.

```
# Pooling effects analysis -- ExpPC x N
aux <- fulldat %>%
  filter(indepvar2 == 'N',
         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
## Crowley (2019)               -0.3510
## Crowley (2019)                5.9750
## Crowley (2019)                7.6580
## Lee and Park (2018)          -0.8510
```

```
## Lee and Park (2018)           -1.6890
## Lee and Park (2018)            7.6320
## Lee (2016)                     0.0164
## Lee (2015)                    -2.0130
## Lee (2015)                     1.4750
## Lee (2015)                    -0.2680
## Kessler (2014)                 0.1740
## Kessler (2014)                 0.2230
## Kessler (2014)                 0.2150
## Kessler (2014)                 0.1580
## Bjedov et al. (2014)          -0.0030
## Bjedov et al. (2014)          -0.0060
## Baskaran (2013)                0.9740
## Erler (2007)                   3.9300
## Chen and Malhotra (2007)      -2.0400
## Chen and Malhotra (2007)      -1.4000
## Fiorino and Ricciuti (2007)    0.2130
## Fiorino and Ricciuti (2007)    0.2290
## Fiorino and Ricciuti (2007)    0.4550
## Fiorino and Ricciuti (2007)    0.4110
## Fiorino and Ricciuti (2007)    0.2260
## Fiorino and Ricciuti (2007)    0.2130
## Fiorino and Ricciuti (2007)    0.1850
## Fiorino and Ricciuti (2007)    0.2350
## Fiorino and Ricciuti (2007)    0.3740
## Fiorino and Ricciuti (2007)    0.8110
## Fiorino and Ricciuti (2007)    0.7950
## Fiorino and Ricciuti (2007)    0.8490
## Primo (2006)                  -0.8200
## Primo (2006)                  -1.7000
## Primo (2006)                  -2.3700
## Primo (2006)                  -2.0300
## Matsusaka (2005)              -0.9600
```

```
## Gilligan and Matsusaka (2001) -0.0140

## Gilligan and Matsusaka (2001)  0.0040

## Schaltegger and Feld (2009)    0.0010

## Schaltegger and Feld (2009)   -0.0010

## Ricciuti (2003)                1.7650

## Ricciuti (2003)                1.7990

##                                          95%-CI

## Crowley (2019)                [-1.8112;  1.1092]

## Crowley (2019)                [ 0.7889; 11.1611]

## Crowley (2019)                [-0.0290; 15.3450]

## Lee and Park (2018)           [-3.5851;  1.8831]

## Lee and Park (2018)           [-3.0551; -0.3229]

## Lee and Park (2018)           [ 3.1064; 12.1576]

## Lee (2016)                    [-2.5570;  2.5898]

## Lee (2015)                    [-4.5727;  0.5467]

## Lee (2015)                    [-4.8968;  7.8468]

## Lee (2015)                    [-2.9081;  2.3721]

## Kessler (2014)                [ 0.0074;  0.3406]

## Kessler (2014)                [ 0.1211;  0.3249]

## Kessler (2014)                [ 0.0954;  0.3346]

## Kessler (2014)                [ 0.0522;  0.2638]

## Bjedov et al. (2014)          [-0.0226;  0.0166]

## Bjedov et al. (2014)          [-0.0256;  0.0136]

## Baskaran (2013)               [-0.1212;  2.0692]

## Erler (2007)                  [ 1.6172;  6.2428]

## Chen and Malhotra (2007)      [-4.6468;  0.5668]

## Chen and Malhotra (2007)      [-2.6544; -0.1456]

## Fiorino and Ricciuti (2007)   [ 0.1777;  0.2483]

## Fiorino and Ricciuti (2007)   [ 0.1565;  0.3015]

## Fiorino and Ricciuti (2007)   [ 0.3805;  0.5295]

## Fiorino and Ricciuti (2007)   [ 0.3150;  0.5070]

## Fiorino and Ricciuti (2007)   [ 0.1221;  0.3299]

## Fiorino and Ricciuti (2007)   [-0.4083;  0.8343]
```

```
## Fiorino and Ricciuti (2007)   [-0.4128;  0.7828]
## Fiorino and Ricciuti (2007)   [-0.4235;  0.8935]
## Fiorino and Ricciuti (2007)   [ 0.2486;  0.4994]
## Fiorino and Ricciuti (2007)   [ 0.4562;  1.1658]
## Fiorino and Ricciuti (2007)   [ 0.4500;  1.1400]
## Fiorino and Ricciuti (2007)   [ 0.3825;  1.3155]
## Primo (2006)                  [-1.1924; -0.4476]
## Primo (2006)                  [-2.3076; -1.0924]
## Primo (2006)                  [-3.0952; -1.6448]
## Primo (2006)                  [-2.7552; -1.3048]
## Matsusaka (2005)              [-1.3128; -0.6072]
## Gilligan and Matsusaka (2001) [-0.0375;  0.0095]
## Gilligan and Matsusaka (2001) [-0.0156;  0.0236]
## Schaltegger and Feld (2009)   [-0.0010;  0.0030]
## Schaltegger and Feld (2009)   [-0.0030;  0.0010]
## Ricciuti (2003)               [ 0.0638;  3.4662]
## Ricciuti (2003)               [ 0.2957;  3.3023]
##                               %W(random)
## Crowley (2019)                       1.7
## Crowley (2019)                       0.3
## Crowley (2019)                       0.1
## Lee and Park (2018)                  0.8
## Lee and Park (2018)                  1.9
## Lee and Park (2018)                  0.4
## Lee (2016)                           0.9
## Lee (2015)                           0.9
## Lee (2015)                           0.2
## Lee (2015)                           0.9
## Kessler (2014)                       3.2
## Kessler (2014)                       3.2
## Kessler (2014)                       3.2
## Kessler (2014)                       3.2
## Bjedov et al. (2014)                 3.2
```

```
## Bjedov et al. (2014)                3.2
## Baskaran (2013)                     2.2
## Erler (2007)                        1.0
## Chen and Malhotra (2007)            0.9
## Chen and Malhotra (2007)            2.0
## Fiorino and Ricciuti (2007)         3.2
## Fiorino and Ricciuti (2007)         3.2
## Fiorino and Ricciuti (2007)         3.2
## Fiorino and Ricciuti (2007)         3.2
## Fiorino and Ricciuti (2007)         3.2
## Fiorino and Ricciuti (2007)         2.8
## Fiorino and Ricciuti (2007)         2.8
## Fiorino and Ricciuti (2007)         2.7
## Fiorino and Ricciuti (2007)         3.2
## Fiorino and Ricciuti (2007)         3.1
## Fiorino and Ricciuti (2007)         3.1
## Fiorino and Ricciuti (2007)         3.0
## Primo (2006)                        3.0
## Primo (2006)                        2.8
## Primo (2006)                        2.7
## Primo (2006)                        2.7
## Matsusaka (2005)                    3.1
## Gilligan and Matsusaka (2001)       3.2
## Gilligan and Matsusaka (2001)       3.2
## Schaltegger and Feld (2009)         3.2
## Schaltegger and Feld (2009)         3.2
## Ricciuti (2003)                     1.5
## Ricciuti (2003)                     1.7
##
## Number of studies combined: k = 43
##
##                         SMD            95%-CI
## Random effects model 0.0235 [-0.3349; 0.3818]
```

```
## Prediction interval         [-1.6661; 1.7130]
##                         t p-value
## Random effects model 0.13  0.8954
## Prediction interval
##
## Quantifying heterogeneity:
##  tau^2 = 0.6684 [0.6710; 3.4362]; tau = 0.8175 [0.8191; 1.8537];
##  I^2 = 94.4% [93.3%; 95.4%]; H = 4.24 [3.85; 4.66]
##
## Test of heterogeneity:
##       Q d.f.  p-value
##  753.40   42 < 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
```

The forest plot:

```
build_forest(mod, NULL)
```

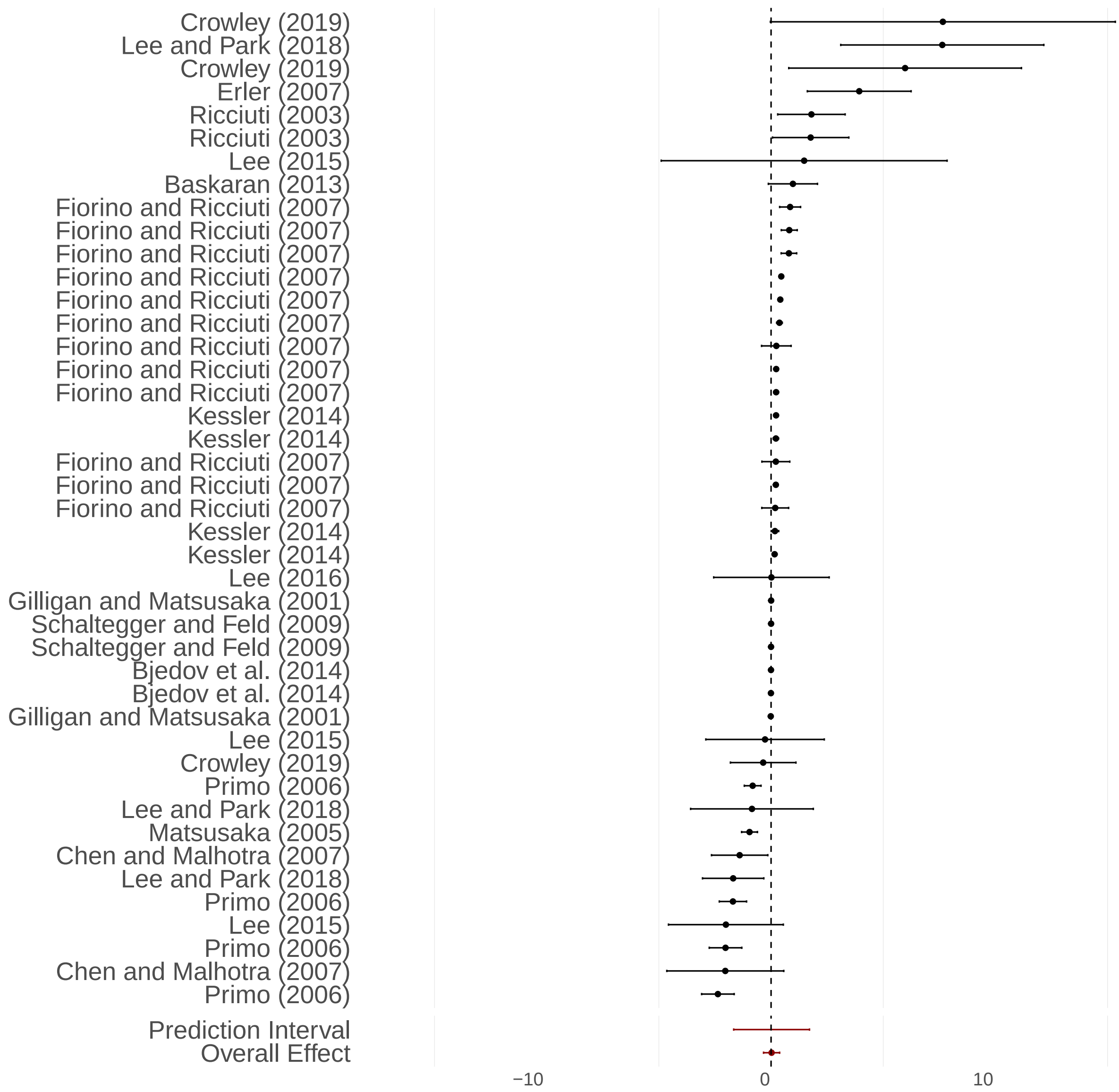


Figure 19: Effect of Lower House Size (N) on Per Capita Expenditure (ExpPC)

Highlights:

1. The results are highly heterogeneous: $I^2 = 94.43$.
2. The estimated SMD in the random effects model is $g = 0.02$ ($SE = 0.178$).
3. The prediction interval ranges from -1.67 to 1.71. Therefore, it emcompasses zero.

I.1.1 Electoral System Subgroup Analysis

The *law of 1/n* was formulated to analyse the budgetary allocation in 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 dummy variable indicating the electoral system included in each model.

```
mod2 <- tibble(
  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()

f8b <- mod2 %>%
  ggplot(aes(y = reorder(studlab2,TE), x = TE, xmin = lower, xmax = upper)) +
  # Studies coefs
  geom_point(aes(color = group)) +
  # Error Bars
  geom_errorbarh(aes(color = group), height = 0.1) +
  # Colors
  scale_color_manual(values = c("#000000", "#8b0000")) +
  # X-axis limit
  scale_x_continuous(limits=c(1.2*(min(mod2$lower)), 1.2*(max(mod2$upper)))) +
  # Y-axis names
  scale_y_discrete(labels = function(x) str_replace(x, "_[0-9]*$", "")) +
  # Vertical dashed line
  geom_vline(xintercept=0, color="#000000", linetype="dashed") +
```

```
  # Labels
  labs(x = "",
       y = "") +
  # Facet - Separating Studies from Overall Effect
  facet_grid(group~., scales = "free", space = "free") +
  # Theme
  theme_minimal() %+replace%
  theme(strip.text.y = element_blank(),
        legend.position = "none",
        axis.text.y = element_text(size = 13, hjust = 1.1),
        axis.text.x = element_text(size = 15, hjust = 1.1))
f8b
```

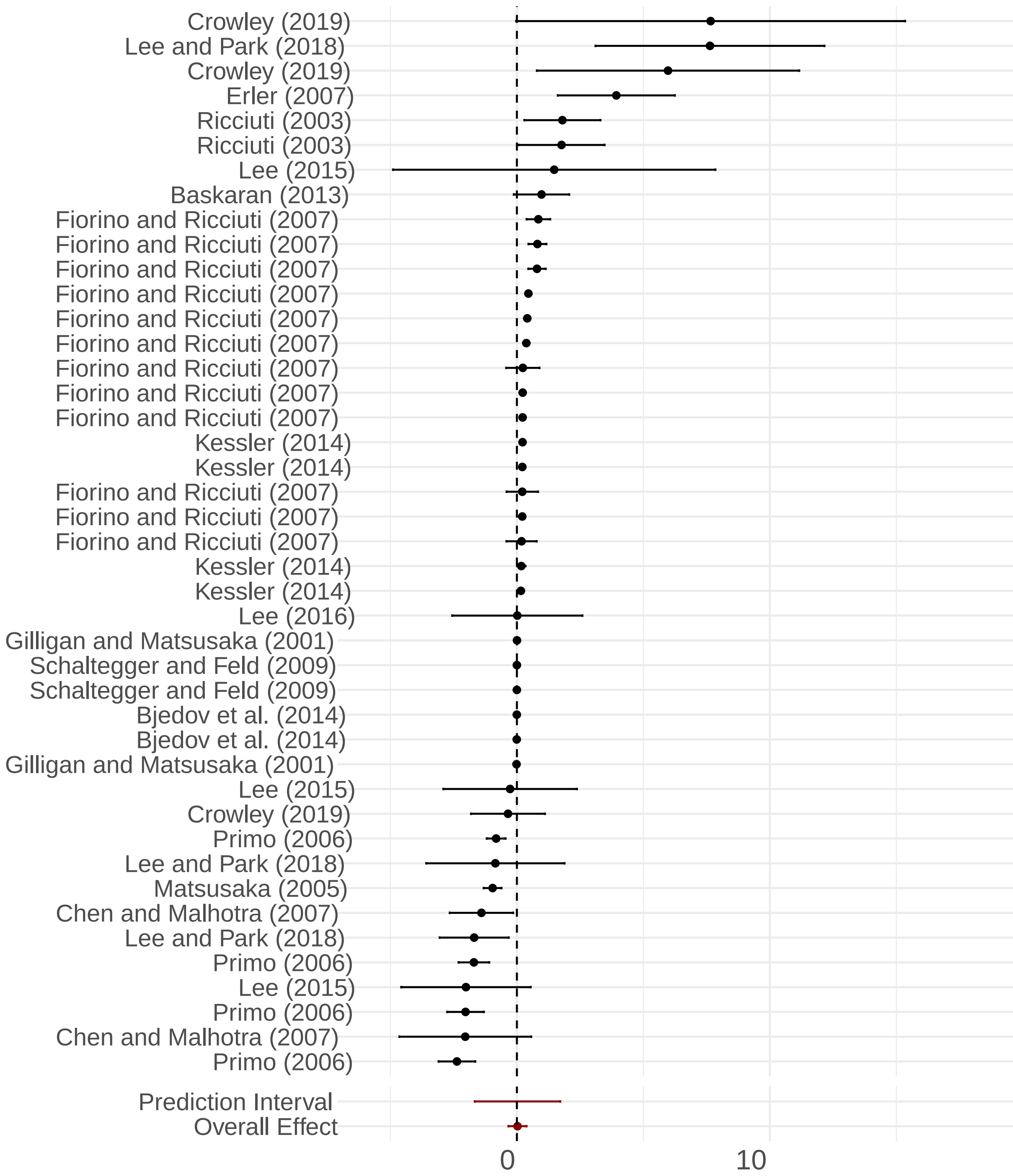


Figure 20: Subgroup Analysis of (N) x (ExpPC), controlling by electoral system

Therefore, we see that majoritarian systems do not have a clear positive effect on budgetary spending. 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 absence of effect is not caused by pooling multiple types of electoral systems.

I.2 Log of Lower House Size and Expenditure Per Capita

There are no studies that have per capita expenditure as the dependent variable and log of lower house size as the treatment variable.

I.3 Upper House Size and Expenditure Per Capita

Now we investigate the effect of the upper house size (K) on government spending. In the model below, we evaluate the relationship between upper house size and 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
## Crowley (2019)                    8.2100
## Crowley (2019)                    8.4230
## Crowley (2019)                    9.5940
## Lee and Park (2018)              19.7400
## Lee and Park (2018)              10.0600
## Lee and Park (2018)               9.0620
## Lee (2016)                       38.4400
## Lee (2016)                       37.8500
## Lee (2016)                       25.6100
## Lee (2016)                        5.9960
```

```
## Lee (2016)                      25.5600
## Lee (2016)                       4.6930
## Lee (2015)                      20.3900
## Lee (2015)                       4.0940
## Lee (2015)                      32.7700
## Bradbury and Stephenson (2009)  0.6240
## Chen and Malhotra (2007)        26.0900
## Chen and Malhotra (2007)         8.3000
## Chen and Malhotra (2007)         5.1400
## Chen and Malhotra (2007)         4.7800
## Chen and Malhotra (2007)        20.3800
## Chen and Malhotra (2007)         4.8700
## Chen and Malhotra (2007)        26.7500
## Primo (2006)                     0.9700
## Primo (2006)                     5.9000
## Primo (2006)                     5.7500
## Primo (2006)                     6.9600
## Gilligan and Matsusaka (2001)   0.1510
## Gilligan and Matsusaka (2001)   0.2140
## Ricciuti (2003)                 -3.9240
## Ricciuti (2003)                 -3.8720
##                                             95%-CI
## Crowley (2019)                  [  0.2702; 16.1498]
## Crowley (2019)                  [-27.1895; 44.0355]
## Crowley (2019)                  [  2.1383; 17.0497]
## Lee and Park (2018)             [  3.2645; 36.2155]
## Lee and Park (2018)             [  2.2887; 17.8313]
## Lee and Park (2018)             [-30.8821; 49.0061]
## Lee (2016)                      [  0.7499; 76.1301]
## Lee (2016)                      [  3.0214; 72.6786]
## Lee (2016)                      [ -0.8103; 52.0303]
## Lee (2016)                      [-19.6011; 31.5931]
## Lee (2016)                      [ -0.8799; 51.9999]
```

```
## Lee (2016)                      [-19.5126; 28.8986]
## Lee (2015)                      [-10.2638; 51.0438]
## Lee (2015)                      [-21.7971; 29.9851]
## Lee (2015)                      [  0.1170; 65.4230]
## Bradbury and Stephenson (2009) [  0.2295;  1.0185]
## Chen and Malhotra (2007)       [ 11.4883; 40.6917]
## Chen and Malhotra (2007)       [  3.6941; 12.9059]
## Chen and Malhotra (2007)       [  0.1813; 10.0987]
## Chen and Malhotra (2007)       [ -0.9039; 10.4639]
## Chen and Malhotra (2007)       [  7.6990; 33.0610]
## Chen and Malhotra (2007)       [  1.2833;  8.4567]
## Chen and Malhotra (2007)       [  0.8589; 52.6411]
## Primo (2006)                   [ -0.4804;  2.4204]
## Primo (2006)                   [  2.6857;  9.1143]
## Primo (2006)                   [  2.3593;  9.1407]
## Primo (2006)                   [  2.6089; 11.3111]
## Gilligan and Matsusaka (2001)  [ -0.0136;  0.3156]
## Gilligan and Matsusaka (2001)  [  0.0709;  0.3571]
## Ricciuti (2003)                [ -6.4955; -1.3525]
## Ricciuti (2003)                [ -6.4474; -1.2966]
##                                %W(random)
## Crowley (2019)                        4.0
## Crowley (2019)                        0.5
## Crowley (2019)                        4.1
## Lee and Park (2018)                   1.8
## Lee and Park (2018)                   4.0
## Lee and Park (2018)                   0.4
## Lee (2016)                            0.5
## Lee (2016)                            0.5
## Lee (2016)                            0.9
## Lee (2016)                            0.9
## Lee (2016)                            0.9
## Lee (2016)                            1.0
```

```
## Lee (2015)                            0.7
## Lee (2015)                            0.9
## Lee (2015)                            0.6
## Bradbury and Stephenson (2009)        6.1
## Chen and Malhotra (2007)              2.2
## Chen and Malhotra (2007)              5.2
## Chen and Malhotra (2007)              5.0
## Chen and Malhotra (2007)              4.8
## Chen and Malhotra (2007)              2.6
## Chen and Malhotra (2007)              5.5
## Chen and Malhotra (2007)              0.9
## Primo (2006)                          6.0
## Primo (2006)                          5.6
## Primo (2006)                          5.5
## Primo (2006)                          5.2
## Gilligan and Matsusaka (2001)         6.1
## Gilligan and Matsusaka (2001)         6.1
## Ricciuti (2003)                       5.8
## Ricciuti (2003)                       5.8
##
## Number of studies combined: k = 31
##
##                          SMD             95%-CI
## Random effects model 6.0634 [ 3.0622;  9.0646]
## Prediction interval         [-5.6093; 17.7361]
##                          t p-value
## Random effects model 4.13  0.0003
## Prediction interval
##
## Quantifying heterogeneity:
##  tau^2 = 30.4136 [15.1455; 126.3725]; tau = 5.5149 [3.8917; 11.2416];
##  I^2 = 80.0% [72.3%; 85.6%]; H = 2.24 [1.90; 2.64]
##
```

```
## Test of heterogeneity:
##       Q d.f.  p-value
##  150.25   30 < 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:

```
build_forest(mod, NULL)
```

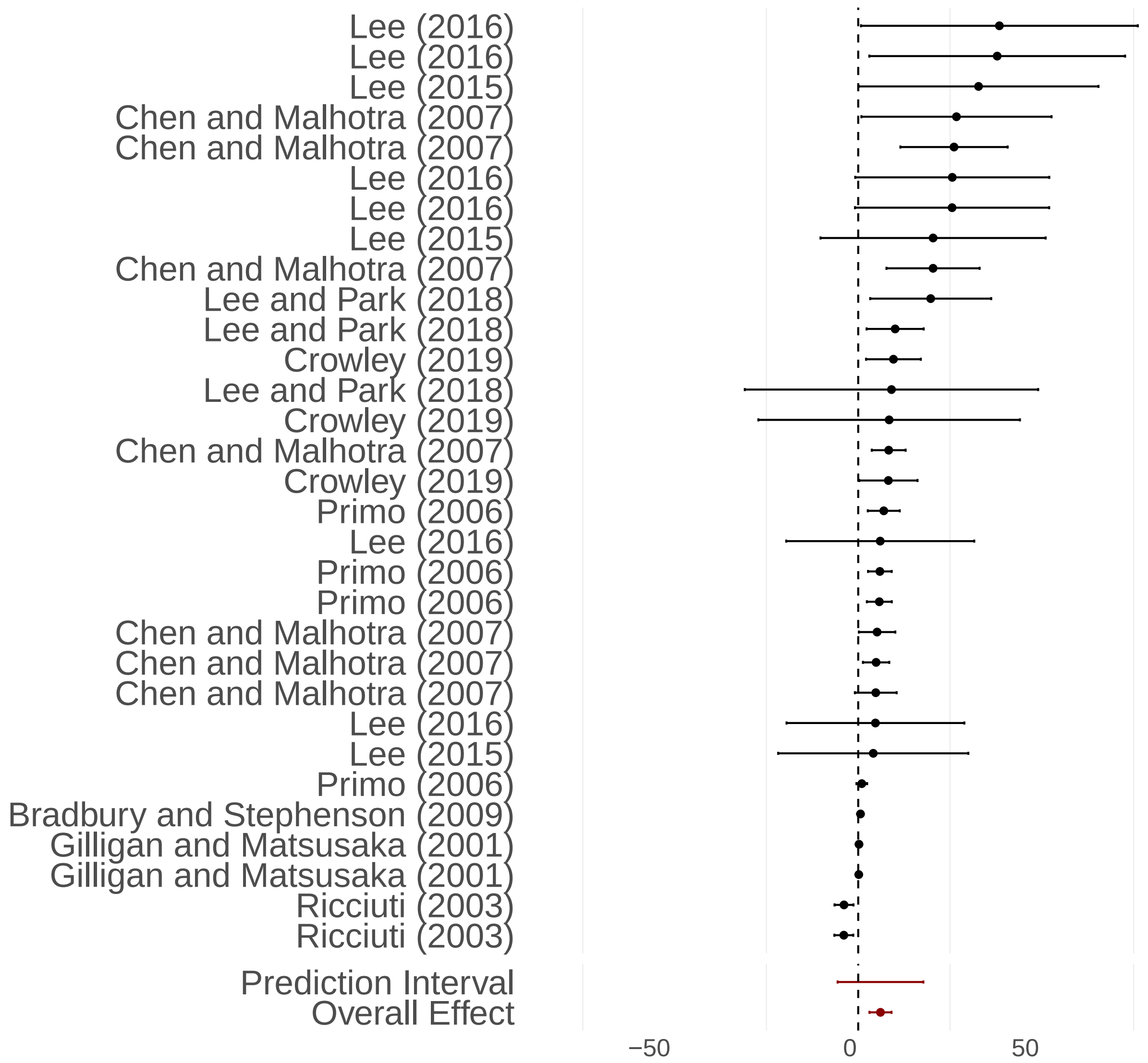


Figure 21: Effect of upper house size (K) on the per capita government expenditure (ExpPC)

Highlights:

1. The results are highly heterogeneous: $I^2 = 80.03$.
2. The estimated SMD in the random effects model is $g = 6.06$ ($SE = 1.47$).
3. The prediction interval ranges from -5.61 to 17.74. Therefore, it emcompasses zero.

I.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
## Lewis (2019)              -0.1740
## Höhmann (2017)            -0.0300
## Höhmann (2017)            -0.0300
## Höhmann (2017)            -0.0400
## Drew and Dollery (2017)    0.0770
## Drew and Dollery (2017)    0.0310
## Pettersson-Lidbom (2012) -0.1590
## Pettersson-Lidbom (2012) -0.1470
## Pettersson-Lidbom (2012) -0.0900
```

```
## Pettersson-Lidbom (2012) -0.0810
## Pettersson-Lidbom (2012) -0.0880
## Pettersson-Lidbom (2012)  0.2100
## Pettersson-Lidbom (2012)  0.1570
## Pettersson-Lidbom (2012) -0.1990
## Pettersson-Lidbom (2012) -0.1690
##                                     95%-CI
## Lewis (2019)             [-0.2450; -0.1030]
## Höhmann (2017)           [-0.0496; -0.0104]
## Höhmann (2017)           [-0.0496; -0.0104]
## Höhmann (2017)           [-0.0792; -0.0008]
## Drew and Dollery (2017)  [ 0.0221;  0.1319]
## Drew and Dollery (2017)  [-0.0121;  0.0741]
## Pettersson-Lidbom (2012) [-0.2394; -0.0786]
## Pettersson-Lidbom (2012) [-0.2274; -0.0666]
## Pettersson-Lidbom (2012) [-0.1645; -0.0155]
## Pettersson-Lidbom (2012) [-0.1574; -0.0046]
## Pettersson-Lidbom (2012) [-0.1625; -0.0135]
## Pettersson-Lidbom (2012) [ 0.1649;  0.2551]
## Pettersson-Lidbom (2012) [ 0.0845;  0.2295]
## Pettersson-Lidbom (2012) [-0.2774; -0.1206]
## Pettersson-Lidbom (2012) [-0.2494; -0.0886]
##                          %W(random)
## Lewis (2019)                    6.6
## Höhmann (2017)                  7.1
## Höhmann (2017)                  7.1
## Höhmann (2017)                  7.0
## Drew and Dollery (2017)         6.8
## Drew and Dollery (2017)         6.9
## Pettersson-Lidbom (2012)        6.4
## Pettersson-Lidbom (2012)        6.4
## Pettersson-Lidbom (2012)        6.5
## Pettersson-Lidbom (2012)        6.5
```

```
## Pettersson-Lidbom (2012)        6.5
## Pettersson-Lidbom (2012)        6.9
## Pettersson-Lidbom (2012)        6.5
## Pettersson-Lidbom (2012)        6.4
## Pettersson-Lidbom (2012)        6.4
##
## Number of studies combined: k = 15
##
##                          SMD            95%-CI
## Random effects model -0.0463 [-0.1142; 0.0216]
## Prediction interval          [-0.3105; 0.2178]
##                          t p-value
## Random effects model -1.46  0.1655
## Prediction interval
##
## 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
```

The forest plot is shown below:

```r
build_forest(mod, NULL)
```

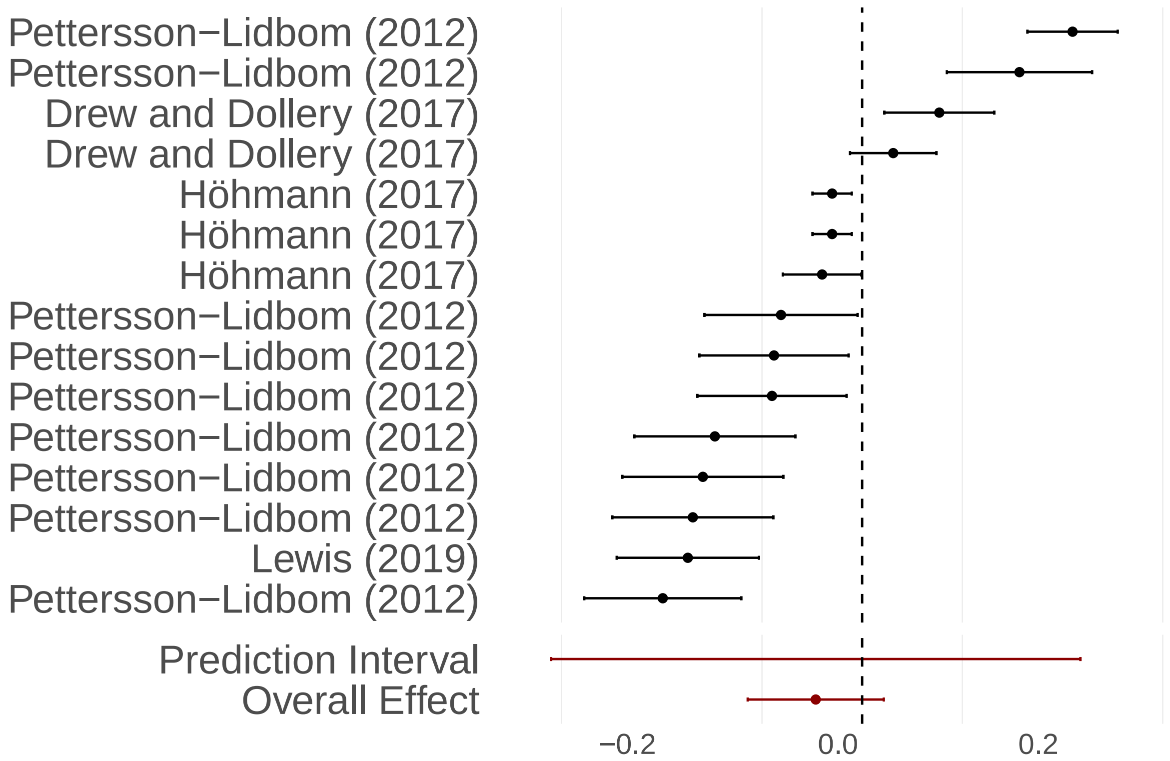


Figure 22: Effect of lower houses size (N) on log of per capita expenditure (logExpPC)

Highlights:

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

I.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).

```r
# 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
## MacDonald (2008) 0.1360 [0.0447; 0.2273]
## MacDonald (2008) 0.2319 [0.1322; 0.3316]
## MacDonald (2008) 0.1443 [0.0471; 0.2415]
## MacDonald (2008) 0.1594 [0.0667; 0.2521]
## MacDonald (2008) 0.2259 [0.1163; 0.3355]
## Baqir (2002)     0.1127 [0.0396; 0.1858]
## Baqir (2002)     0.2760 [0.2007; 0.3513]
## Baqir (2002)     0.3021 [0.2270; 0.3772]
## Baqir (2002)     0.3203 [0.2450; 0.3956]
## Baqir (1999)     0.3020 [0.2269; 0.3771]
## Baqir (1999)     0.2760 [0.2007; 0.3513]
## Baqir (1999)     0.2950 [0.2165; 0.3735]
##                  %W(random)
## MacDonald (2008)        7.9
## MacDonald (2008)        7.4
## MacDonald (2008)        7.6
## MacDonald (2008)        7.8
## MacDonald (2008)        6.9
## Baqir (2002)            9.1
## Baqir (2002)            8.9
## Baqir (2002)            8.9
## Baqir (2002)            8.9
## Baqir (1999)            8.9
## Baqir (1999)            8.9
## Baqir (1999)            8.7
```

```
##
## Number of studies combined: k = 12
##
##                          SMD           95%-CI
## Random effects model 0.2346 [0.1864; 0.2828]
## Prediction interval         [0.0848; 0.3844]
##                           t  p-value
## Random effects model 10.71 < 0.0001
## Prediction interval
##
## 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
```

The forest plot:

```r
build_forest(mod, NULL)
```

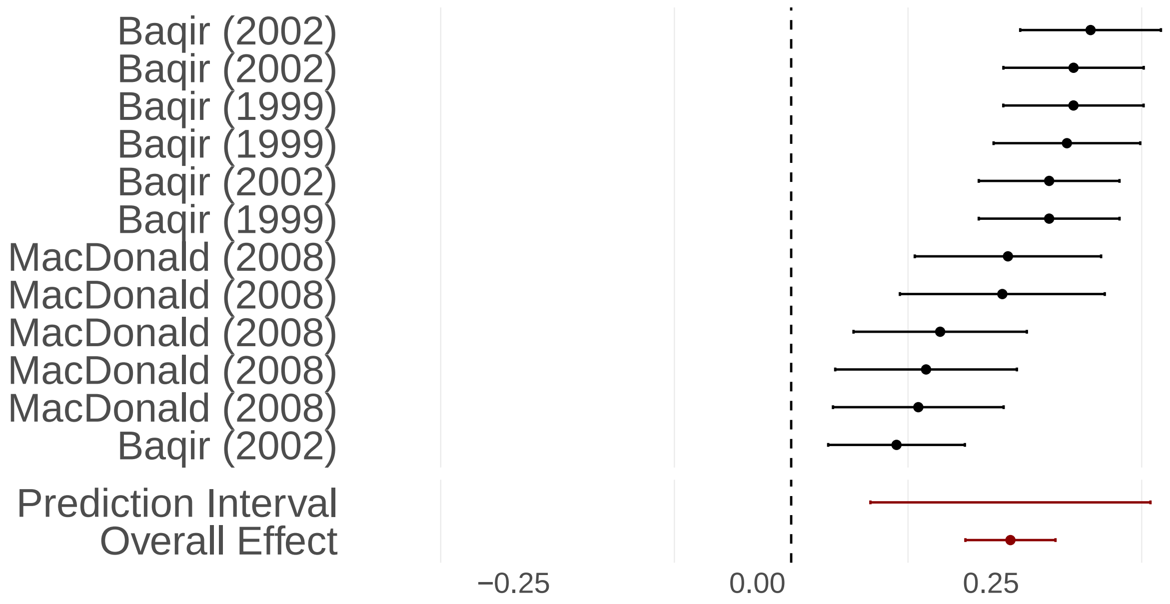


Figure 23: Effect of log lower houses size (logN) on the log of per capita government expenditure (logExpPC)

Highlights:

1. The results are highly heterogeneous: $I^2 = $ 69.96.
2. The estimated SMD in the random effects model 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.

I.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).

I.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 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
## Bjedov et al. (2014)       -0.0040
## Bjedov et al. (2014)       -0.0080
## Maldonado (2013)           -0.0609
## Mukherjee (2003)            0.0030
## Mukherjee (2003)            0.0090
## Mukherjee (2003)            0.0110
## Mukherjee (2003)            0.0050
## Mukherjee (2003)            0.0400
## Mukherjee (2003)            0.0300
## Mukherjee (2003)            0.0100
## Mukherjee (2003)            0.0200
## Bradbury and Crain (2001)   0.0036
## Bradbury and Crain (2001)   0.0005
## Bradbury and Crain (2001)   0.0169
## Bradbury and Crain (2001)   0.0123
## Ricciuti (2004)             0.0140
## Ricciuti (2004)            -0.0110
## Ricciuti (2004)             0.0070
## Ricciuti (2004)             0.0050
## Ricciuti (2004)             0.0050
## Ricciuti (2004)             0.0120
##                                        95%-CI
## Bjedov et al. (2014)       [-0.0432;  0.0352]
## Bjedov et al. (2014)       [-0.0472;  0.0312]
## Maldonado (2013)           [-0.0838; -0.0380]
```

```
## Mukherjee (2003)          [ 0.0010;  0.0050]
## Mukherjee (2003)          [ 0.0051;  0.0129]
## Mukherjee (2003)          [ 0.0051;  0.0169]
## Mukherjee (2003)          [-0.0009;  0.0109]
## Mukherjee (2003)          [ 0.0380;  0.0420]
## Mukherjee (2003)          [ 0.0280;  0.0320]
## Mukherjee (2003)          [ 0.0061;  0.0139]
## Mukherjee (2003)          [ 0.0122;  0.0278]
## Bradbury and Crain (2001) [ 0.0008;  0.0065]
## Bradbury and Crain (2001) [-0.0016;  0.0027]
## Bradbury and Crain (2001) [ 0.0131;  0.0208]
## Bradbury and Crain (2001) [ 0.0087;  0.0160]
## Ricciuti (2004)           [-0.0095;  0.0375]
## Ricciuti (2004)           [-0.0286;  0.0066]
## Ricciuti (2004)           [-0.0067;  0.0207]
## Ricciuti (2004)           [-0.0126;  0.0226]
## Ricciuti (2004)           [-0.0126;  0.0226]
## Ricciuti (2004)           [-0.0017;  0.0257]
##                           %W(random)
## Bjedov et al. (2014)             2.1
## Bjedov et al. (2014)             2.1
## Maldonado (2013)                 3.6
## Mukherjee (2003)                 5.6
## Mukherjee (2003)                 5.5
## Mukherjee (2003)                 5.4
## Mukherjee (2003)                 5.4
## Mukherjee (2003)                 5.6
## Mukherjee (2003)                 5.6
## Mukherjee (2003)                 5.5
## Mukherjee (2003)                 5.3
## Bradbury and Crain (2001)        5.6
## Bradbury and Crain (2001)        5.6
## Bradbury and Crain (2001)        5.6
```

```
## Bradbury and Crain (2001)        5.6
## Ricciuti (2004)                  3.5
## Ricciuti (2004)                  4.2
## Ricciuti (2004)                  4.7
## Ricciuti (2004)                  4.2
## Ricciuti (2004)                  4.2
## Ricciuti (2004)                  4.7
##
## Number of studies combined: k = 21
##
##                         SMD            95%-CI
## Random effects model 0.0078 [-0.0003; 0.0160]
## Prediction interval         [-0.0259; 0.0416]
##                         t p-value
## Random effects model 2.01  0.0579
## Prediction interval
##
## 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
```

Here is the forest plot:

```r
build_forest(mod, NULL)
```

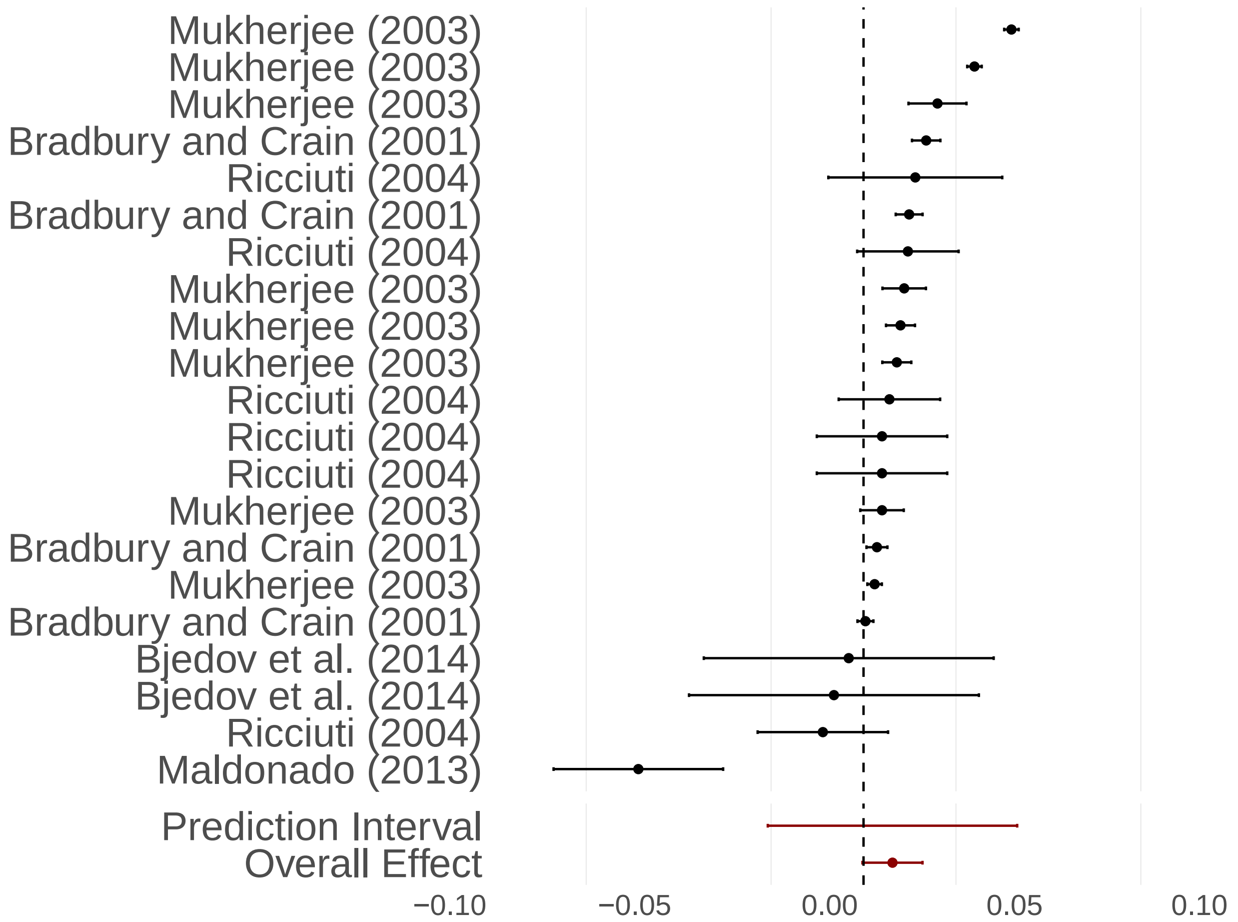


Figure 24: Effect of lower houses size (N) on percentage of public expenditure GDP (PCTGDP)

Highlights:

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

I.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 natural logarithm of lower house size ($\log(N)$) as the treatment variable.

```r
# Pooling effects analysis -- PCTGDP x logN

aux <- fulldat %>%

  filter(indepvar2 == 'logN',

         depvar2 == 'PCTGDP')
```

```r
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
## Baqir (1999)          2.0660 [ 1.4887; 2.6433]
## Baqir (1999)          2.0120 [ 1.4235; 2.6005]
## Baqir (1999)          2.4680 [ 1.8817; 3.0543]
## Lledo (2003)         -4.6900 [-9.9427; 0.5627]
## Stein et al. (1998)  0.0109 [-0.0171; 0.0389]
## Stein et al. (1998)  0.0135 [-0.0102; 0.0372]
##                      %W(random)
## Baqir (1999)               18.9
## Baqir (1999)               18.8
## Baqir (1999)               18.8
## Lledo (2003)                3.8
## Stein et al. (1998)        19.8
## Stein et al. (1998)        19.8
##
## Number of studies combined: k = 6
##
##                          SMD            95%-CI
## Random effects model 1.0619 [-0.7256; 2.8493]
## Prediction interval         [-3.0267; 5.1504]
##                          t p-value
## Random effects model 1.53  0.1873
## Prediction interval
```

```
##
## 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
```

The forest plot:

```
build_forest(mod, NULL)
```

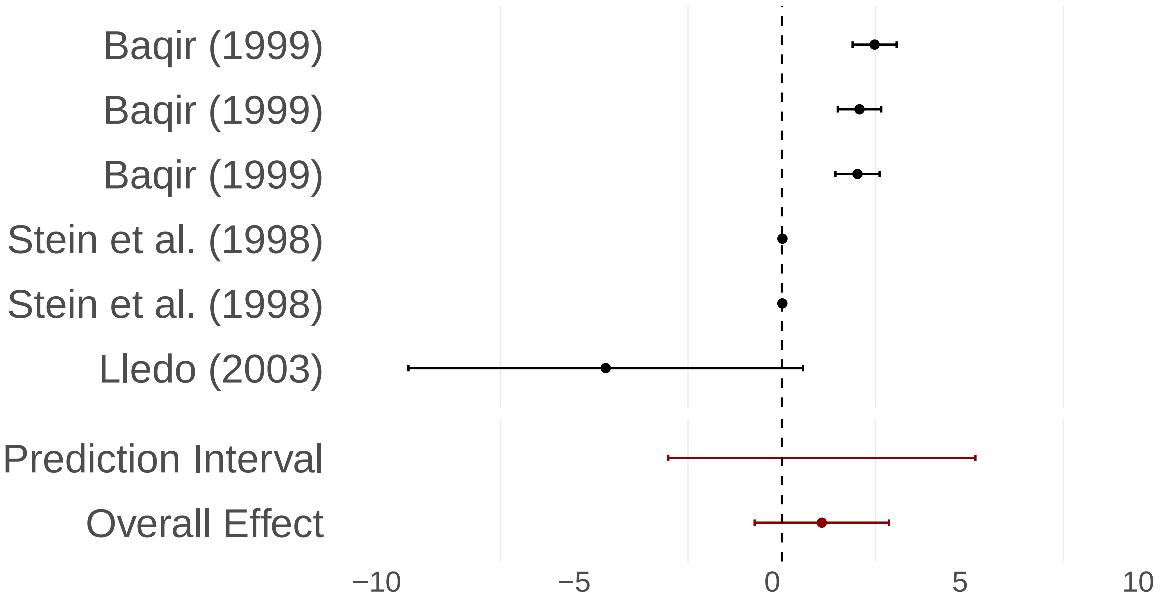


Figure 25: Effect of log lower houses size (logN) on the GDP share of public expenditure (PCTGDP)

Highlights:

1. The results are highly heterogeneous: $I^2 = 96.93$.

2. The estimated SMD in the random effects model is $g = 1.06$ ($SE = 0.695$).
3. The prediction interval ranges from -3.03 to 5.15. Therefore, it emcompasses zero.

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

```
##                                 SMD
## Maldonado (2013)            -0.0400
## Bradbury and Crain (2001)  0.0126
## Bradbury and Crain (2001)  0.0050
## Bradbury and Crain (2001) -0.0113
## Bradbury and Crain (2001) -0.0056
## Ricciuti (2004)             0.0160
## Ricciuti (2004)             0.0210
## Ricciuti (2004)             0.0140
## Ricciuti (2004)             0.0030
## Ricciuti (2004)             0.0300
## Ricciuti (2004)             0.0300
## Ricciuti (2004)             0.0390
```

```
## Ricciuti (2004)            0.0127
## Ricciuti (2004)            0.0160
##                                           95%-CI
## Maldonado (2013)          [-0.0659; -0.0141]
## Bradbury and Crain (2001) [ 0.0010;  0.0243]
## Bradbury and Crain (2001) [ 0.0016;  0.0083]
## Bradbury and Crain (2001) [-0.0163; -0.0064]
## Bradbury and Crain (2001) [-0.0102; -0.0010]
## Ricciuti (2004)           [-0.0075;  0.0395]
## Ricciuti (2004)           [-0.0006;  0.0426]
## Ricciuti (2004)           [-0.0036;  0.0316]
## Ricciuti (2004)           [-0.0088;  0.0148]
## Ricciuti (2004)           [-0.0210;  0.0810]
## Ricciuti (2004)           [-0.0210;  0.0810]
## Ricciuti (2004)           [-0.0022;  0.0802]
## Ricciuti (2004)           [-0.0147;  0.0401]
## Ricciuti (2004)           [-0.0075;  0.0395]
##                           %W(random)
## Maldonado (2013)                 5.7
## Bradbury and Crain (2001)        9.8
## Bradbury and Crain (2001)       11.8
## Bradbury and Crain (2001)       11.5
## Bradbury and Crain (2001)       11.6
## Ricciuti (2004)                  6.2
## Ricciuti (2004)                  6.7
## Ricciuti (2004)                  7.9
## Ricciuti (2004)                  9.7
## Ricciuti (2004)                  2.2
## Ricciuti (2004)                  2.2
## Ricciuti (2004)                  3.1
## Ricciuti (2004)                  5.3
## Ricciuti (2004)                  6.2
##
```

```
## Number of studies combined: k = 14
##
##                          SMD            95%-CI
## Random effects model 0.0056 [-0.0042; 0.0155]
## Prediction interval         [-0.0233; 0.0346]
##                          t p-value
## Random effects model 1.24  0.2376
## Prediction interval
##
## 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:

```
build_forest(mod, NULL)
```

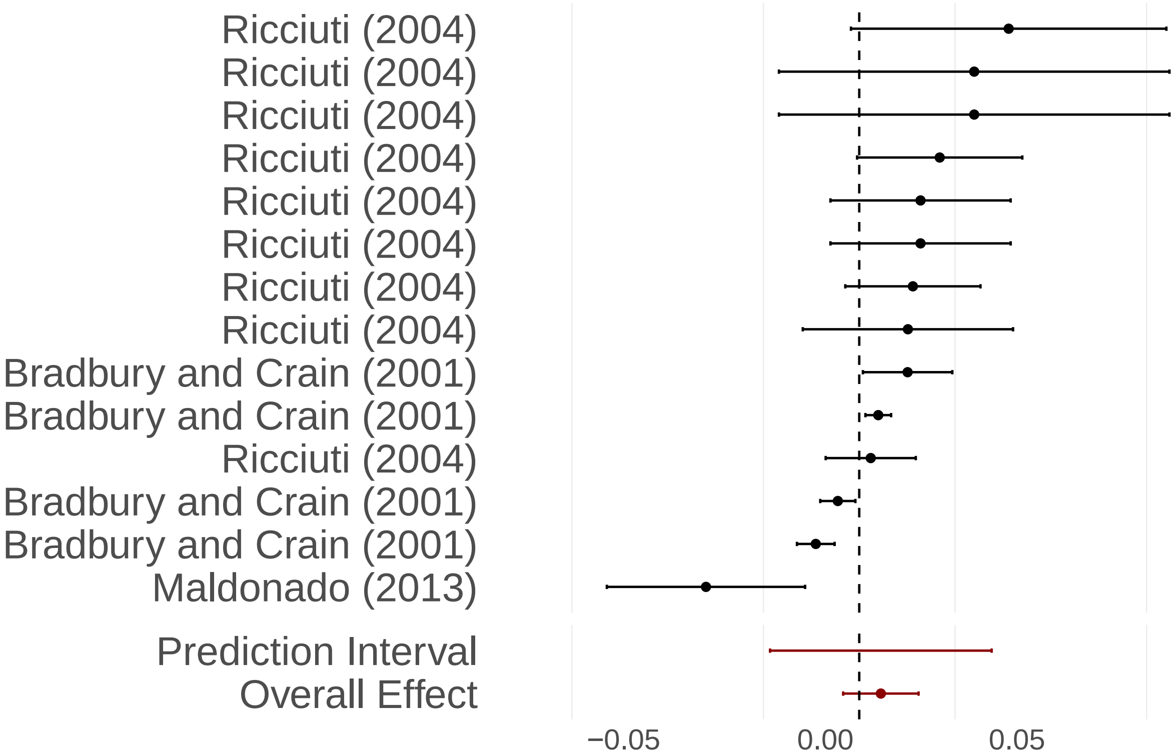


Figure 26: Effect of upper house size (K) on the public expenditure share of the GDP (PCTGDP)

Highlights:

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

I.10 Lower House Size and Expenditure per Capita (IV)

```
# Pooling effects analysis -- ExpPC x N (IV only)

aux <- fulldat %>%

  filter(indepvar2 == 'N',

         depvar2 == 'ExpPC',

         method %in% c('IV'))



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
## Lee (2015)                   -2.0130
## Lee (2015)                    1.4750
## Lee (2015)                   -0.2680
## Baskaran (2013)               0.9740
## Fiorino and Ricciuti (2007)   0.2130
## Fiorino and Ricciuti (2007)   0.2290
## Fiorino and Ricciuti (2007)   0.4550
## Fiorino and Ricciuti (2007)   0.4110
## Matsusaka (2005)             -0.9600
##                                          95%-CI
## Lee (2015)                   [-4.5727;  0.5467]
## Lee (2015)                   [-4.8968;  7.8468]
## Lee (2015)                   [-2.9081;  2.3721]
## Baskaran (2013)              [-0.1212;  2.0692]
## Fiorino and Ricciuti (2007) [ 0.1777;  0.2483]
## Fiorino and Ricciuti (2007) [ 0.1565;  0.3015]
## Fiorino and Ricciuti (2007) [ 0.3805;  0.5295]
## Fiorino and Ricciuti (2007) [ 0.3150;  0.5070]
## Matsusaka (2005)             [-1.3128; -0.6072]
##                              %W(random)
## Lee (2015)                          2.6
## Lee (2015)                          0.5
## Lee (2015)                          2.5
## Baskaran (2013)                     8.6
## Fiorino and Ricciuti (2007)        17.5
## Fiorino and Ricciuti (2007)        17.5
## Fiorino and Ricciuti (2007)        17.5
```

```
## Fiorino and Ricciuti (2007)       17.4
## Matsusaka (2005)                  15.9
##
## Number of studies combined: k = 9
##
##                         SMD            95%-CI
## Random effects model 0.1076 [-0.4199; 0.6351]
## Prediction interval         [-1.3004; 1.5156]
##                         t p-value
## Random effects model 0.47  0.6506
## Prediction interval
##
## Quantifying heterogeneity:
##  tau^2 = 0.3022 [0.0810; 2.2712]; tau = 0.5497 [0.2846; 1.5070];
##  I^2 = 91.5% [86.1%; 94.8%]; H = 3.43 [2.69; 4.39]
##
## Test of heterogeneity:
##      Q d.f.  p-value
##  94.23    8 < 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:

```
build_forest(mod, NULL)
```

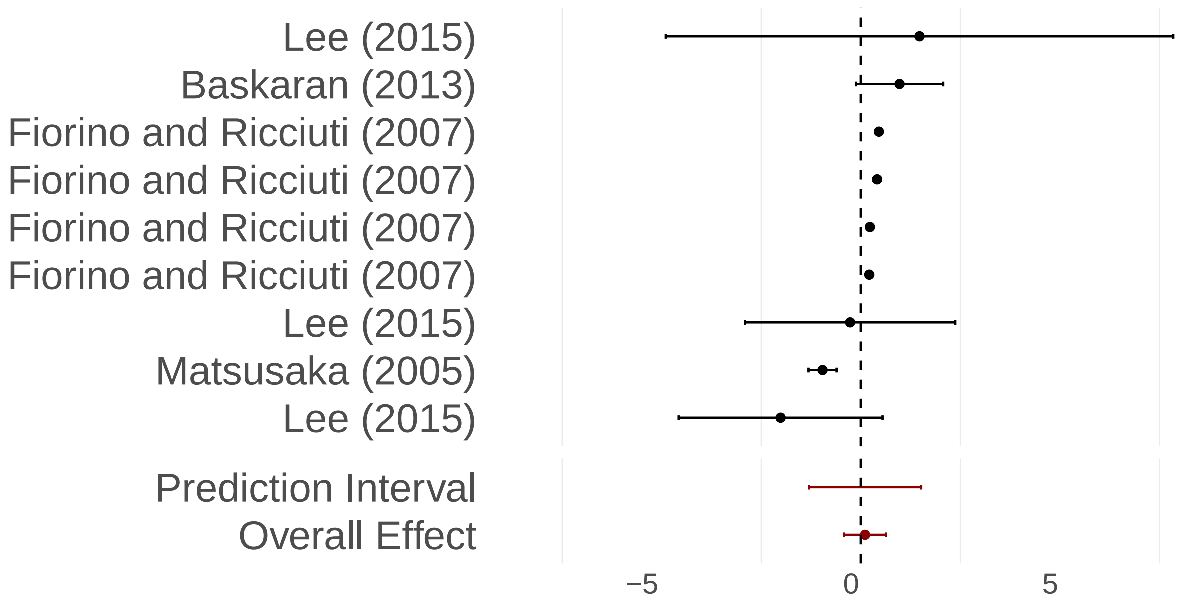


Figure 27: Effect of lower houses size (N) on Per Capita Expenditure (ExpPC)

Highlights:

1. The results are highly heterogeneous: $I^2 = 91.51$.
2. The estimated SMD in the random effects model is $g = 0.11$ ($SE = 0.229$).
3. The prediction interval ranges from -1.3 to 1.52. Therefore, it emcompasses zero.

I.11 Lower House Size and Log of Expenditure per Capita (RDD)

```
# Pooling effects analysis -- ExpPC x N

aux <- fulldat %>%

  filter(indepvar2 == 'N',

         depvar2 == 'logExpPC',

         method == 'RDD')


mod <- metagen(coef, SE, data=aux,

          studlab=paste(authoryear),

          comb.fixed = FALSE,

          comb.random = TRUE,

          method.tau = "REML",

          hakn = TRUE,

          prediction=TRUE,
```

```r
         sm="SMD")
mod
```

```
##                                SMD
## Lewis (2019)             -0.1740
## Höhmann (2017)           -0.0300
## Höhmann (2017)           -0.0300
## Höhmann (2017)           -0.0400
## Pettersson-Lidbom (2012) -0.1590
## Pettersson-Lidbom (2012) -0.1470
## Pettersson-Lidbom (2012) -0.0900
## Pettersson-Lidbom (2012) -0.0810
## Pettersson-Lidbom (2012) -0.0880
##                                   95%-CI
## Lewis (2019)             [-0.2450; -0.1030]
## Höhmann (2017)           [-0.0496; -0.0104]
## Höhmann (2017)           [-0.0496; -0.0104]
## Höhmann (2017)           [-0.0792; -0.0008]
## Pettersson-Lidbom (2012) [-0.2394; -0.0786]
## Pettersson-Lidbom (2012) [-0.2274; -0.0666]
## Pettersson-Lidbom (2012) [-0.1645; -0.0155]
## Pettersson-Lidbom (2012) [-0.1574; -0.0046]
## Pettersson-Lidbom (2012) [-0.1625; -0.0135]
##                          %W(random)
## Lewis (2019)                   10.0
## Höhmann (2017)                 14.9
## Höhmann (2017)                 14.9
## Höhmann (2017)                 13.3
## Pettersson-Lidbom (2012)        9.1
## Pettersson-Lidbom (2012)        9.1
## Pettersson-Lidbom (2012)        9.6
## Pettersson-Lidbom (2012)        9.5
## Pettersson-Lidbom (2012)        9.6
##
```

```
## Number of studies combined: k = 9
##
##                           SMD             95%-CI
## Random effects model -0.0843 [-0.1275; -0.0410]
## Prediction interval          [-0.2075;  0.0390]
##                           t p-value
## Random effects model -4.49  0.0020
## Prediction interval
##
## Quantifying heterogeneity:
##  tau^2 = 0.0024 [0.0006; 0.0104]; tau = 0.0486 [0.0240; 0.1022];
##  I^2 = 76.9% [56.1%; 87.9%]; H = 2.08 [1.51; 2.88]
##
## Test of heterogeneity:
##      Q d.f.  p-value
##  34.70    8 < 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:

```r
build_forest(mod, NULL)
```

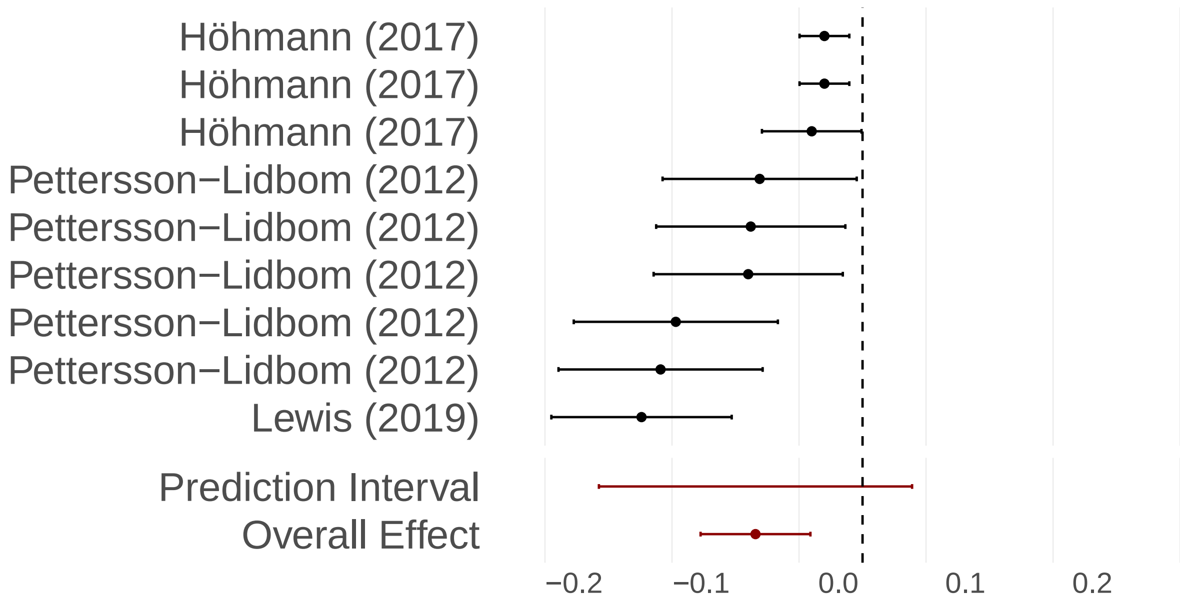


Figure 28: Effect of lower houses size (N) on Per Capita Expenditure (ExpPC)

Highlights:

1. The results are highly heterogeneous: I^2 = 76.94.
2. The estimated SMD in the random effects model is g = -0.08 (SE = 0.019).
3. The prediction interval ranges from -0.21 to 0.04. Therefore, it emcompasses zero.

J Meta-Regressions

J.1 Meta-Regressions for Expenditure as a Percentage of the GDP

In this section, we show the coefficients for our meta-regressions. We start with expenditure as a percentage of GDP as the dependent variable.

```
mod <-  rma(yi = coef,
              sei = SE,
              data = dat,
              method = "REML",
              mods = ~indepvar2+year+published+elecsys2+method,
              test = "knha",
            subset = dat$depvar2=='PCTGDP',
            slab = dat$authoryear)


mod1 <- tibble(
     ` ` = c("Intercept",
             "Indepvar: N",
             "Indepvar: logN",
             "Year",
             "Elecsys: Non-Majoritarian",
             "Method: Panel"),
  Estimate = mod[["beta"]],
  SE = mod[["se"]],
  `T` = mod[["zval"]],
  `P-Value` = mod[["pval"]],
  CI = paste0("(", round(mod[["ci.lb"]], digits = 4), " ; ",
              round(mod[["ci.ub"]], digits = 4), ")"),
  model = "PCTGDP"
) %>%
  mutate_if(is.numeric, list(~round(., digits = 4)))
```

```r
summary(mod)
```

```
##
## Mixed-Effects Model (k = 11; tau^2 estimator: REML)
##
##   logLik  deviance       AIC       BIC
##   6.7485  -13.4970    0.5030   -2.2310
##     AICc
## 112.5030
##
## tau^2 (estimated amount of residual heterogeneity):     0.0002 (SE = 0.0002)
## tau (square root of estimated tau^2 value):             0.0149
## I^2 (residual heterogeneity / unaccounted variability): 92.93%
## H^2 (unaccounted variability / sampling variability):   14.14
## R^2 (amount of heterogeneity accounted for):            99.92%
##
## Test for Residual Heterogeneity:
## QE(df = 5) = 23.7235, p-val = 0.0002
##
## Test of Moderators (coefficients 2:6):
## F(df1 = 5, df2 = 5) = 7.2901, p-val = 0.0240
##
## Model Results:
##
##                           estimate      se
## intrcpt                     9.9763  3.9450
## indepvar2N                 -0.0055  0.0175
## indepvar2logN              -0.0159  0.0359
## year                       -0.0039  0.0020
## elecsys2Non-Majoritarian   -2.0593  0.3804
## methodPANEL                 0.0025  0.0202
##                              tval    pval
## intrcpt                    2.5288  0.0526
## indepvar2N                -0.3123  0.7674
```

```
## indepvar2logN             -0.4420  0.6769
## year                      -2.0202  0.0993
## elecsys2Non-Majoritarian  -5.4132  0.0029
## methodPANEL                0.1245  0.9058
##                             ci.lb    ci.ub
## intrcpt                   -0.1647  20.1174   .
## indepvar2N                -0.0506   0.0396
## indepvar2logN             -0.1083   0.0765
## year                      -0.0090   0.0011   .
## elecsys2Non-Majoritarian  -3.0372  -1.0814  **
## methodPANEL               -0.0493   0.0543
##
## ---
## 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.

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

```
##
## Test of Moderators (coefficients 2:6):
## F(df1 = 5, df2 = 5) = 7.2901, p-val* = 0.0530
##
## Model Results:
##
##                           estimate      se
## intrcpt                     9.9763  3.9450
## indepvar2N                 -0.0055  0.0175
## indepvar2logN              -0.0159  0.0359
## year                       -0.0039  0.0020
## elecsys2Non-Majoritarian   -2.0593  0.3804
## methodPANEL                 0.0025  0.0202
##                               tval   pval*
```

```
## intrcpt                     2.5288  0.0040
## indepvar2N                 -0.3123  0.7550
## indepvar2logN              -0.4420  0.6590
## year                       -2.0202  0.0190
## elecsys2Non-Majoritarian  -5.4132  0.0840
## methodPANEL                 0.1245  0.8940
##                              ci.lb    ci.ub
## intrcpt                    -0.1647  20.1174  **
## indepvar2N                 -0.0506   0.0396
## indepvar2logN              -0.1083   0.0765
## year                       -0.0090   0.0011   *
## elecsys2Non-Majoritarian  -3.0372  -1.0814   .
## methodPANEL                -0.0493   0.0543
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
mod1_permu <- tibble(
     ` ` = c("Intercept",
             "Indepvar: N",
             "Indepvar: logN",
             "Year",
             "Elecsys: Non-Majoritarian",
             "Method: Panel"),
  Estimate = mod[["beta"]],
  SE = mod[["se"]],
  `T` = mod[["zval"]],
  `P-Value` = mod[["pval"]],
  CI = paste0("(", round(mod[["ci.lb"]], digits = 4), " ; ",
              round(mod[["ci.ub"]], digits = 4 ), ")"),
  model = "PCTGDP - Permutation"
) %>%
  mutate_if(is.numeric, list(~round(., digits = 3)))
```

We have the following results for the meta-regressions of Expenditure as a Percentage of GDP:

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 modelling, passing from `OLS` to `PANEL` increases the detected effects.
6. When passing from `Country` to `Cross-Country`, it has no effect on the estimated coefficients.

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

```
mod <- rma(yi = coef,

              sei = SE,

              data = fulldat,

              method = "REML",

              mods = ~ indepvar2+year+published+elecsys2+method,

              test = "knha",

            subset = fulldat$depvar2=='PCTGDP',

            slab = fulldat$authoryear)


mod2 <- tibble(
  ` ` = c("Intercept",

          "Indepvar: N",

          "Indepvar: logN",

          "Year",

          "Elecsys: Non-Majoritarian",

          "Method: Panel"),

  Estimate = mod[["beta"]],

  SE = mod[["se"]],

  `T` = mod[["zval"]],

  `P-Value` = mod[["pval"]],

  CI = paste0("(", round(mod[["ci.lb"]], digits = 4), " ; ",

              round(mod[["ci.ub"]], digits = 4 ), ")"),

  model = "PCTGDP - All coefs"
```

```r
) %>%
  mutate_if(is.numeric, list(~round(., digits = 3)))
```

```r
summary(mod)
```

```
##
## Mixed-Effects Model (k = 41; tau^2 estimator: REML)
##
##    logLik   deviance        AIC        BIC
##   84.0132  -168.0264  -154.0264  -143.1390
##      AICc
## -149.8783
##
## tau^2 (estimated amount of residual heterogeneity):     0.0002 (SE = 0.0001)
## tau (square root of estimated tau^2 value):             0.0129
## I^2 (residual heterogeneity / unaccounted variability): 96.03%
## H^2 (unaccounted variability / sampling variability):   25.20
## R^2 (amount of heterogeneity accounted for):            99.88%
##
## Test for Residual Heterogeneity:
## QE(df = 35) = 1217.8992, p-val < .0001
##
## Test of Moderators (coefficients 2:6):
## F(df1 = 5, df2 = 35) = 33.2523, p-val < .0001
##
## Model Results:
##
##                           estimate      se
## intrcpt                     8.7105  2.0519
## indepvar2N                  0.0047  0.0059
## indepvar2logN              -0.0123  0.0156
## year                       -0.0033  0.0010
## elecsys2Non-Majoritarian   -2.1723  0.1752
## methodPANEL                -0.0041  0.0063
```

```
##                                tval    pval
## intrcpt                      4.2450  0.0002
## indepvar2N                   0.8035  0.4271
## indepvar2logN               -0.7880  0.4360
## year                        -3.1979  0.0029
## elecsys2Non-Majoritarian  -12.4019  <.0001
## methodPANEL                 -0.6490  0.5206
##                              ci.lb    ci.ub
## intrcpt                     4.5448  12.8762  ***
## indepvar2N                 -0.0072   0.0166
## indepvar2logN              -0.0441   0.0194
## year                       -0.0053  -0.0012   **
## elecsys2Non-Majoritarian  -2.5279  -1.8167  ***
## methodPANEL                -0.0169   0.0087
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

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

```
##
## Test of Moderators (coefficients 2:6):
## F(df1 = 5, df2 = 35) = 33.2523, p-val* = 0.0010
##
## Model Results:
##
##                            estimate      se
## intrcpt                      8.7105  2.0519
## indepvar2N                   0.0047  0.0059
## indepvar2logN               -0.0123  0.0156
## year                        -0.0033  0.0010
## elecsys2Non-Majoritarian    -2.1723  0.1752
## methodPANEL                 -0.0041  0.0063
```

```
##                               tval    pval*
## intrcpt                     4.2450  0.0010
## indepvar2N                  0.8035  0.3460
## indepvar2logN              -0.7880  0.3050
## year                       -3.1979  0.0060
## elecsys2Non-Majoritarian  -12.4019  0.0010
## methodPANEL                -0.6490  0.4380
##                             ci.lb     ci.ub
## intrcpt                    4.5448   12.8762  ***
## indepvar2N                -0.0072    0.0166
## indepvar2logN             -0.0441    0.0194
## year                      -0.0053   -0.0012   **
## elecsys2Non-Majoritarian  -2.5279   -1.8167  ***
## methodPANEL               -0.0169    0.0087
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
mod2_permu <- tibble(
     ` ` = c("Intercept",
             "Indepvar: N",
             "Indepvar: logN",
             "Year",
             "Elecsys: Non-Majoritarian",
             "Method: Panel"),
  Estimate = mod[["beta"]],
  SE = mod[["se"]],
  `T` = mod[["zval"]],
  `P-Value` = mod[["pval"]],
  CI = paste0("(", round(mod[["ci.lb"]], digits = 4), " ; ",
              round(mod[["ci.ub"]], digits = 4 ), ")"),
  model = "PCTGDP - All coefs - Permutation"
) %>%
  mutate_if(is.numeric, list(~round(., digits = 3)))
```

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` significantly decreases the effects found in our models.
5. Regarding statistical models, passing from `OLS` to `PANEL` increases the detected effects.

J.2 Meta-Regressions for Expenditure Per Capita

Here we do the same exercise with expenditure per capita as the main outcome.

```
mod <-  rma(yi = coef,

              sei = SE,

              data = dat,

              method = "REML",

              mods = ~indepvar2+year+published+elecsys2+method,

              test = "knha",

            subset = dat$depvar2=='ExpPC',

            slab = dat$authoryear,

            control=list(maxiter=1000))


mod3 <- tibble(

  ` ` = c("Intercept",

          "Indepvar: N",

          "Year",

          "Published: No",

          "Elecsys: Non-Majoritarian",

          "Method: Panel",

          "Method: IV"),

  Estimate = mod[["beta"]],

  SE = mod[["se"]],

  `T` = mod[["zval"]],

  `P-Value` = mod[["pval"]],

  CI = paste0("(", round(mod[["ci.lb"]], digits = 4), " ; ",
```

```r
             round(mod[["ci.ub"]], digits = 4 ), ")"),
  model = "ExpPC"
) %>%
  mutate_if(is.numeric, list(~round(., digits = 3)))
```

```r
summary(mod)
```

```
##
## Mixed-Effects Model (k = 24; tau^2 estimator: REML)
##
##    logLik  deviance       AIC       BIC
## -51.2403  102.4807  118.4807  125.1464
##     AICc
## 136.4807
##
## tau^2 (estimated amount of residual heterogeneity):     2.4577 (SE = 1.2115)
## tau (square root of estimated tau^2 value):             1.5677
## I^2 (residual heterogeneity / unaccounted variability): 99.71%
## H^2 (unaccounted variability / sampling variability):   343.92
## R^2 (amount of heterogeneity accounted for):            0.00%
##
## Test for Residual Heterogeneity:
## QE(df = 17) = 124.7931, p-val < .0001
##
## Test of Moderators (coefficients 2:7):
## F(df1 = 6, df2 = 17) = 0.1403, p-val = 0.9886
##
## Model Results:
##
##                             estimate        se
## intrcpt                     -64.6307  239.9857
## indepvar2N                   -1.0416    1.7620
## year                          0.0325    0.1194
## publishedNo                  -1.4630    2.6010
```

```
## elecsys2Non-Majoritarian    0.4147    1.4962
## methodPANEL                 0.5012    2.7942
## methodIV                   -0.0682    3.1931
##                              tval     pval
## intrcpt                   -0.2693   0.7909
## indepvar2N                -0.5911   0.5622
## year                       0.2719   0.7890
## publishedNo               -0.5625   0.5811
## elecsys2Non-Majoritarian   0.2772   0.7850
## methodPANEL                0.1794   0.8598
## methodIV                  -0.0214   0.9832
##                               ci.lb      ci.ub
## intrcpt                   -570.9562   441.6948
## indepvar2N                  -4.7590     2.6758
## year                        -0.2195     0.2845
## publishedNo                 -6.9506     4.0247
## elecsys2Non-Majoritarian    -2.7420     3.5713
## methodPANEL                 -5.3940     6.3964
## methodIV                    -6.8050     6.6686
##
## ---
## 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.

```r
mod <- permutest(mod, progbar = F)
mod
```

```
##
## Test of Moderators (coefficients 2:7):
## F(df1 = 6, df2 = 17) = 0.1403, p-val* = 0.9180
##
## Model Results:
##
```

```
##                             estimate        se
## intrcpt                     -64.6307  239.9857
## indepvar2N                   -1.0416    1.7620
## year                          0.0325    0.1194
## publishedNo                  -1.4630    2.6010
## elecsys2Non-Majoritarian      0.4147    1.4962
## methodPANEL                   0.5012    2.7942
## methodIV                     -0.0682    3.1931
##                                tval    pval*
## intrcpt                     -0.2693   0.7040
## indepvar2N                  -0.5911   0.4180
## year                         0.2719   0.6980
## publishedNo                 -0.5625   0.4620
## elecsys2Non-Majoritarian     0.2772   0.6600
## methodPANEL                  0.1794   0.8130
## methodIV                    -0.0214   0.9740
##                                 ci.lb      ci.ub
## intrcpt                     -570.9562   441.6948
## indepvar2N                    -4.7590     2.6758
## year                          -0.2195     0.2845
## publishedNo                   -6.9506     4.0247
## elecsys2Non-Majoritarian      -2.7420     3.5713
## methodPANEL                   -5.3940     6.3964
## methodIV                      -6.8050     6.6686
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
mod3_permu <- tibble(
  ` ` = c("Intercept",
          "Indepvar: N",
          "Year",
          "Published: No",
          "Elecsys: Non-Majoritarian",
```

```
          "Method: Panel",
          "Method: IV"),
  Estimate = mod[["beta"]],
  SE = mod[["se"]],
  `T` = mod[["zval"]],
  `P-Value` = mod[["pval"]],
  CI = paste0("(", round(mod[["ci.lb"]], digits = 4), " ; ",
              round(mod[["ci.ub"]], digits = 4 ), ")"),
  model = "ExpPC - Permutation"
) %>%
  mutate_if(is.numeric, list(~round(., digits = 3)))
```

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. `Majoritarian`, when compared to `Non-Majoritarian` electoral systems, significantly increase per capita expenditure.
4. Regarding statistical models, passing from `OLS` to `PANEL` or `IV` increases the detected effects.

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

```
mod <- rma(yi = coef,
            sei = SE,
            data = fulldat,
            method = "REML",
            mods = ~ indepvar2+year+published+elecsys2+method,
            test = "knha",
          subset = fulldat$depvar2=='ExpPC',
          slab = fulldat$authoryear)


mod4 <- tibble(
  ` ` = c("Intercept",
          "Indepvar: N",
          "Year",
```

```r
          "Published: No",
          "Elecsys: Non-Majoritarian",
          "Method: Panel",
          "Method: IV"),
  Estimate = mod[["beta"]],
  SE = mod[["se"]],
  `T` = mod[["zval"]],
  `P-Value` = mod[["pval"]],
  CI = paste0("(", round(mod[["ci.lb"]], digits = 4), " ; ",
              round(mod[["ci.ub"]], digits = 4 ), ")"),
  model = "ExpPC - All coefs"
) %>%
  mutate_if(is.numeric, list(~round(., digits = 3)))
```

```r
summary(mod)
```

```
##
## Mixed-Effects Model (k = 74; tau^2 estimator: REML)
##
##    logLik   deviance        AIC        BIC
## -189.5073   379.0146   395.0146   412.6521
##      AICc
##  397.4973
##
## tau^2 (estimated amount of residual heterogeneity):     3.6491 (SE = 0.8732)
## tau (square root of estimated tau^2 value):             1.9103
## I^2 (residual heterogeneity / unaccounted variability): 99.94%
## H^2 (unaccounted variability / sampling variability):   1765.58
## R^2 (amount of heterogeneity accounted for):            0.00%
##
## Test for Residual Heterogeneity:
## QE(df = 67) = 568.5327, p-val < .0001
##
## Test of Moderators (coefficients 2:7):
```

```
## F(df1 = 6, df2 = 67) = 3.3232, p-val = 0.0063
##
## Model Results:
##
##                             estimate        se
## intrcpt                    -298.4345  150.5649
## indepvar2N                   -3.7466    0.9420
## year                          0.1502    0.0751
## publishedNo                  -2.3118    1.7208
## elecsys2Non-Majoritarian      0.7810    1.0298
## methodPANEL                   0.4206    1.1576
## methodIV                     -0.0401    1.1343
##                             tval    pval
## intrcpt                  -1.9821  0.0516
## indepvar2N               -3.9771  0.0002
## year                      2.0004  0.0495
## publishedNo              -1.3434  0.1837
## elecsys2Non-Majoritarian  0.7584  0.4509
## methodPANEL               0.3634  0.7175
## methodIV                 -0.0354  0.9719
##                             ci.lb    ci.ub
## intrcpt                  -598.9634   2.0944
## indepvar2N                 -5.6270  -1.8663
## year                        0.0003   0.3001
## publishedNo                -5.7466   1.1230
## elecsys2Non-Majoritarian   -1.2746   2.8366
## methodPANEL                -1.8899   2.7312
## methodIV                   -2.3041   2.2239
##
## intrcpt                     .
## indepvar2N               ***
## year                        *
## publishedNo
```

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

```r
mod <- permutest(mod, progbar = F, iter = 100)
mod
```

```
##
## Test of Moderators (coefficients 2:7):
## F(df1 = 6, df2 = 67) = 3.3232, p-val* = 0.0100
##
## Model Results:
##
##                              estimate        se
## intrcpt                     -298.4345  150.5649
## indepvar2N                    -3.7466    0.9420
## year                           0.1502    0.0751
## publishedNo                   -2.3118    1.7208
## elecsys2Non-Majoritarian       0.7810    1.0298
## methodPANEL                    0.4206    1.1576
## methodIV                      -0.0401    1.1343
##                                 tval    pval*
## intrcpt                      -1.9821   0.0100
## indepvar2N                   -3.9771   0.0100
## year                          2.0004   0.0100
## publishedNo                  -1.3434   0.1200
## elecsys2Non-Majoritarian      0.7584   0.3700
## methodPANEL                   0.3634   0.6700
## methodIV                     -0.0354   0.9500
##                                ci.lb     ci.ub
## intrcpt                     -598.9634    2.0944   **
```

```
## indepvar2N                  -5.6270  -1.8663  **
## year                         0.0003   0.3001  **
## publishedNo                 -5.7466   1.1230
## elecsys2Non-Majoritarian    -1.2746   2.8366
## methodPANEL                 -1.8899   2.7312
## methodIV                    -2.3041   2.2239
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
mod4_permu <- tibble(
  ` ` = c("Intercept",
          "Indepvar: N",
          "Year",
          "Published: No",
          "Elecsys: Non-Majoritarian",
          "Method: Panel",
          "Method: IV"),
  Estimate = mod[["beta"]],
  SE = mod[["se"]],
  `T` = mod[["zval"]],
  `P-Value` = mod[["pval"]],
  CI = paste0("(", round(mod[["ci.lb"]], digits = 4), " ; ",
              round(mod[["ci.ub"]], digits = 4 ), ")"),
  model = "ExpPC - All coefs - Permutation"
) %>%
  mutate_if(is.numeric, list(~round(., digits = 3)))
```

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. `Majoritarian`, when compared to `Non-Majoritarian` electoral systems, significantly increase per capita expenditure.

4. Regarding statistical models, passing from `OLS` to `PANEL` decreases the detected effects.
5. All other coeffients were not significant.

J.3 Meta-Regressions for the Log of Expenditure Per Capita

Lastly, we run the model with the natural logarithm of expenditure per capita.

```
mod <- rma(yi = coef,
              sei = SE,
              data = dat,
              method = "REML",
              mods = ~indepvar2+year+published+elecsys2+method,
              test = "knha",
            subset = dat$depvar2=='logExpPC',
            slab = dat$authoryear)


mod5 <- tibble(
  ` ` = c("Intercept",
           "Indepvar: N",
           "Year",
           "Published: No",
           "Method: Panel"),
  Estimate = mod[["beta"]],
  SE = mod[["se"]],
  `T` = mod[["zval"]],
  `P-Value` = mod[["pval"]],
  CI = paste0("(", round(mod[["ci.lb"]], digits = 4), " ; ",
              round(mod[["ci.ub"]], digits = 4 ), ")"),
  model = "logExpPC"
) %>%
  mutate_if(is.numeric, list(~round(., digits = 3)))
```

```
summary(mod)
```

```
##
## Mixed-Effects Model (k = 7; tau^2 estimator: REML)
```

```
##
##    logLik  deviance       AIC       BIC
##    2.1772   -4.3544    7.6456   -0.1956
##      AICc
##   91.6456
##
## tau^2 (estimated amount of residual heterogeneity):     0.0060 (SE = 0.0071)
## tau (square root of estimated tau^2 value):             0.0776
## I^2 (residual heterogeneity / unaccounted variability): 87.73%
## H^2 (unaccounted variability / sampling variability):   8.15
## R^2 (amount of heterogeneity accounted for):            78.05%
##
## Test for Residual Heterogeneity:
## QE(df = 2) = 21.6330, p-val < .0001
##
## Test of Moderators (coefficients 2:5):
## F(df1 = 4, df2 = 2) = 5.7815, p-val = 0.1529
##
## Model Results:
##
##              estimate       se     tval    pval
## intrcpt       -7.3634  25.4584  -0.2892  0.7996
## indepvar2N    -0.2806   0.1621  -1.7309  0.2256
## year           0.0037   0.0127   0.2941  0.7964
## publishedNo    0.2001   0.1269   1.5766  0.2556
## methodPANEL    0.1884   0.0928   2.0296  0.1795
##                  ci.lb     ci.ub
## intrcpt      -116.9019  102.1751
## indepvar2N     -0.9780    0.4169
## year           -0.0509    0.0584
## publishedNo    -0.3459    0.7461
## methodPANEL    -0.2110    0.5879
##
```

```
## ---
## 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.

```r
mod <- permutest(mod, progbar = F)
mod
```

```
##
## Test of Moderators (coefficients 2:5):
## F(df1 = 4, df2 = 2) = 5.7815, p-val* = 0.1380
##
## Model Results:
##
##              estimate       se     tval
## intrcpt       -7.3634  25.4584  -0.2892
## indepvar2N    -0.2806   0.1621  -1.7309
## year           0.0037   0.0127   0.2941
## publishedNo    0.2001   0.1269   1.5766
## methodPANEL    0.1884   0.0928   2.0296
##               pval*      ci.lb     ci.ub
## intrcpt      0.7720  -116.9019  102.1751
## indepvar2N   0.2250    -0.9780    0.4169
## year         0.7680    -0.0509    0.0584
## publishedNo  0.2330    -0.3459    0.7461
## methodPANEL  0.1870    -0.2110    0.5879
##
## intrcpt
## indepvar2N
## year
## publishedNo
## methodPANEL
##
## ---
```

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

```
mod5_permu <- tibble(
  ` ` = c("Intercept",
          "Indepvar: N",
          "Year",
          "Published: No",
          "Method: Panel"),
  Estimate = mod[["beta"]],
  SE = mod[["se"]],
  `T` = mod[["zval"]],
  `P-Value` = mod[["pval"]],
  CI = paste0("(", round(mod[["ci.lb"]], digits = 4), " ; ",
            round(mod[["ci.ub"]], digits = 4 ), ")"),
  model = "logExpPC - Permutation"
) %>%
  mutate_if(is.numeric, list(~round(., digits = 3)))

mod5_permu %>%
  select(-model) %>%
  kable(booktabs = T,align = "c",linesep ='') %>%
  kable_styling(c("striped","bordered"), position = "center")
```

	Estimate	SE	T	P-Value	CI
Intercept	-7.363	25.458	-0.289	0.772	(-116.9019 ; 102.1751)
Indepvar: N	-0.281	0.162	-1.731	0.225	(-0.978 ; 0.4169)
Year	0.004	0.013	0.294	0.768	(-0.0509 ; 0.0584)
Published: No	0.200	0.127	1.577	0.233	(-0.3459 ; 0.7461)
Method: Panel	0.188	0.093	2.030	0.187	(-0.211 ; 0.5879)

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

1. Unpublished papers report significantly higher coefficients.
2. Moving 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:

```
mod <- rma(yi = coef,
              sei = SE,
              data = fulldat,
              method = "REML",
              mods = ~ indepvar2+year+published+elecsys2+method,
              test = "knha",
            subset = fulldat$depvar2=='logExpPC',
            slab = fulldat$authoryear)


mod6 <- tibble(
  ` ` = c("Intercept",
          "Indepvar: N",
          "Year",
          "Published: No",
          "Method: Panel",
          "Method: RDD"),
  Estimate = mod[["beta"]],
  SE = mod[["se"]],
  `T` = mod[["zval"]],
  `P-Value` = mod[["pval"]],
  CI = paste0("(", round(mod[["ci.lb"]], digits = 4), " ; ",
              round(mod[["ci.ub"]], digits = 4), ")"),
  model = "logExpPC - All coefs"
) %>%
  mutate_if(is.numeric, list(~round(., digits = 3)))
```

```
summary(mod)
```

```
##
## Mixed-Effects Model (k = 27; tau^2 estimator: REML)
##
```

```
##   logLik  deviance       AIC       BIC
##  23.4038  -46.8076  -32.8076  -25.4959
##     AICc
## -24.1922
##
## tau^2 (estimated amount of residual heterogeneity):     0.0049 (SE = 0.0019)
## tau (square root of estimated tau^2 value):             0.0703
## I^2 (residual heterogeneity / unaccounted variability): 86.48%
## H^2 (unaccounted variability / sampling variability):   7.40
## R^2 (amount of heterogeneity accounted for):            83.01%
##
## Test for Residual Heterogeneity:
## QE(df = 21) = 102.3018, p-val < .0001
##
## Test of Moderators (coefficients 2:6):
## F(df1 = 5, df2 = 21) = 20.8366, p-val < .0001
##
## Model Results:
##
##              estimate       se     tval    pval
## intrcpt       -8.2960  12.0371  -0.6892  0.4982
## indepvar2N    -0.0577   0.0733  -0.7879  0.4396
## year           0.0042   0.0060   0.7070  0.4874
## publishedNo    0.1030   0.0651   1.5833  0.1283
## methodPANEL   -0.2524   0.0682  -3.6997  0.0013
## methodRDD     -0.2851   0.0617  -4.6212  0.0001
##                 ci.lb    ci.ub
## intrcpt      -33.3285  16.7365
## indepvar2N    -0.2102   0.0947
## year          -0.0082   0.0167
## publishedNo   -0.0323   0.2383
## methodPANEL   -0.3942  -0.1105   **
## methodRDD     -0.4134  -0.1568  ***
```

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

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


mod6_permu <- tibble(
  ` ` = c("Intercept",
         "Indepvar: N",
         "Year",
         "Published: No",
         "Method: Panel",
         "Method: RDD"),
  Estimate = mod[["beta"]],
  SE = mod[["se"]],
  `T` = mod[["zval"]],
  `P-Value` = mod[["pval"]],
  CI = paste0("(", round(mod[["ci.lb"]], digits = 4), " ; ",
              round(mod[["ci.ub"]], digits = 4), ")"),
  model = "logExpPC - All coefs - Permutation"
) %>%
  mutate_if(is.numeric, list(~round(., digits = 3)))
```

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 modelling, passing from `OLS` to `PANEL` or `RDD` decreases the detected effects.
2. All other coefficients remained insignificant.

K Robustness: Meta-Regressions (All Coefficients)

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

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 study (N, K, $log(N)$);
3. The electoral system (Majoritarian, Proportional Representation, and Mixed).

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

```
mod <-  rma(yi = coef,
            sei = SE,
            data = dat,
            method = "REML",
            mods = ~ depvar2+indepvar2+year+published+elecsys2+method,
            test = "knha")
```

```
summary(mod)
```

```
##
## Mixed-Effects Model (k = 42; tau^2 estimator: REML)
##
##   logLik  deviance       AIC       BIC
## -66.5234  133.0468  157.0468  174.2547
##     AICc
## 174.3802
##
## tau^2 (estimated amount of residual heterogeneity):     0.2406 (SE = 0.0872)
## tau (square root of estimated tau^2 value):             0.4905
## I^2 (residual heterogeneity / unaccounted variability): 99.97%
## H^2 (unaccounted variability / sampling variability):   2868.63
## R^2 (amount of heterogeneity accounted for):            0.00%
##
## Test for Residual Heterogeneity:
## QE(df = 31) = 254.3094, p-val < .0001
```

```
##
## Test of Moderators (coefficients 2:11):
## F(df1 = 10, df2 = 31) = 0.2187, p-val = 0.9926
##
## Model Results:
##
##                            estimate
## intrcpt                    -29.2446
## depvar2PCTGDP                0.0220
## depvar2logExpPC             -0.2734
## indepvar2N                  -0.1759
## indepvar2logN                0.2902
## year                         0.0147
## publishedNo                  0.2990
## elecsys2Non-Majoritarian    -0.1018
## methodPANEL                  0.0189
## methodIV                    -0.1654
## methodRDD                    0.0879
##                                 se     tval
## intrcpt                    69.9886  -0.4178
## depvar2PCTGDP               0.5698   0.0387
## depvar2logExpPC             0.7105  -0.3848
## indepvar2N                  0.4574  -0.3846
## indepvar2logN               0.9295   0.3122
## year                        0.0348   0.4211
## publishedNo                 0.7197   0.4154
## elecsys2Non-Majoritarian    0.4822  -0.2112
## methodPANEL                 0.5043   0.0375
## methodIV                    0.7825  -0.2113
## methodRDD                   0.8490   0.1035
##                              pval
## intrcpt                    0.6789
## depvar2PCTGDP              0.9694
```

```
## depvar2logExpPC           0.7030
## indepvar2N                0.7032
## indepvar2logN             0.7570
## year                      0.6766
## publishedNo               0.6807
## elecsys2Non-Majoritarian  0.8341
## methodPANEL               0.9703
## methodIV                  0.8340
## methodRDD                 0.9182
##                              ci.lb
## intrcpt                  -171.9872
## depvar2PCTGDP              -1.1402
## depvar2logExpPC            -1.7226
## indepvar2N                 -1.1087
## indepvar2logN              -1.6055
## year                       -0.0564
## publishedNo                -1.1688
## elecsys2Non-Majoritarian   -1.0852
## methodPANEL                -1.0097
## methodIV                   -1.7613
## methodRDD                  -1.6437
##                             ci.ub
## intrcpt                  113.4980
## depvar2PCTGDP              1.1842
## depvar2logExpPC            1.1757
## indepvar2N                 0.7569
## indepvar2logN              2.1858
## year                       0.0857
## publishedNo                1.7667
## elecsys2Non-Majoritarian   0.8816
## methodPANEL                1.0475
## methodIV                   1.4306
## methodRDD                  1.8195
```

```
##
## ---
## 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.

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

```
##
## Test of Moderators (coefficients 2:11):
## F(df1 = 10, df2 = 31) = 0.2187, p-val* = 0.8730
##
## Model Results:
##
##                            estimate
## intrcpt                    -29.2446
## depvar2PCTGDP                0.0220
## depvar2logExpPC             -0.2734
## indepvar2N                  -0.1759
## indepvar2logN                0.2902
## year                         0.0147
## publishedNo                  0.2990
## elecsys2Non-Majoritarian    -0.1018
## methodPANEL                  0.0189
## methodIV                    -0.1654
## methodRDD                    0.0879
##                                 se     tval
## intrcpt                    69.9886  -0.4178
## depvar2PCTGDP               0.5698   0.0387
## depvar2logExpPC             0.7105  -0.3848
## indepvar2N                  0.4574  -0.3846
## indepvar2logN               0.9295   0.3122
## year                        0.0348   0.4211
## publishedNo                 0.7197   0.4154
```

```
## elecsys2Non-Majoritarian   0.4822  -0.2112
## methodPANEL                0.5043   0.0375
## methodIV                   0.7825  -0.2113
## methodRDD                  0.8490   0.1035
##                            pval*
## intrcpt                   0.5680
## depvar2PCTGDP             0.9550
## depvar2logExpPC           0.5520
## indepvar2N                0.5900
## indepvar2logN             0.6430
## year                      0.5650
## publishedNo               0.5350
## elecsys2Non-Majoritarian  0.7490
## methodPANEL               0.9440
## methodIV                  0.7560
## methodRDD                 0.8760
##                              ci.lb
## intrcpt                   -171.9872
## depvar2PCTGDP               -1.1402
## depvar2logExpPC             -1.7226
## indepvar2N                  -1.1087
## indepvar2logN               -1.6055
## year                        -0.0564
## publishedNo                 -1.1688
## elecsys2Non-Majoritarian    -1.0852
## methodPANEL                 -1.0097
## methodIV                    -1.7613
## methodRDD                   -1.6437
##                             ci.ub
## intrcpt                   113.4980
## depvar2PCTGDP               1.1842
## depvar2logExpPC             1.1757
## indepvar2N                  0.7569
```

```
## indepvar2logN              2.1858
## year                       0.0857
## publishedNo                1.7667
## elecsys2Non-Majoritarian   0.8816
## methodPANEL                1.0475
## methodIV                   1.4306
## methodRDD                  1.8195
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

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

```
mod <- rma(yi = coef,
              sei = SE,
              data = fulldat,
              method = "REML",
              mods = ~ depvar2+indepvar2+year+published+elecsys2+method,
              test = "knha")
```

```
summary(mod)
```

```
##
## Mixed-Effects Model (k = 142; tau^2 estimator: REML)
##
##    logLik   deviance        AIC
## -246.6859   493.3717   517.3717
##       BIC       AICc
##  551.8741   520.0158
##
## tau^2 (estimated amount of residual heterogeneity):     0.1403 (SE = 0.0225)
## tau (square root of estimated tau^2 value):             0.3746
## I^2 (residual heterogeneity / unaccounted variability): 99.98%
```

```
## H^2 (unaccounted variability / sampling variability):   5885.15
## R^2 (amount of heterogeneity accounted for):            21.72%
##
## Test for Residual Heterogeneity:
## QE(df = 131) = 2259.6537, p-val < .0001
##
## Test of Moderators (coefficients 2:11):
## F(df1 = 10, df2 = 131) = 1.3302, p-val = 0.2209
##
## Model Results:
##
##                           estimate
## intrcpt                   -32.5954
## depvar2PCTGDP              -0.1128
## depvar2logExpPC            -0.3454
## indepvar2N                 -0.1581
## indepvar2logN               0.3610
## year                        0.0164
## publishedNo                 0.3905
## elecsys2Non-Majoritarian    0.1048
## methodPANEL                -0.1230
## methodIV                   -0.0445
## methodRDD                  -0.0611
##                                se     tval
## intrcpt                   33.6402  -0.9689
## depvar2PCTGDP              0.2148  -0.5252
## depvar2logExpPC            0.2664  -1.2963
## indepvar2N                 0.1844  -0.8573
## indepvar2logN              0.3876   0.9312
## year                       0.0168   0.9759
## publishedNo                0.3095   1.2617
## elecsys2Non-Majoritarian   0.2270   0.4616
## methodPANEL                0.1631  -0.7542
```

```
## methodIV                   0.3029  -0.1468
## methodRDD                  0.3044  -0.2006
##                             pval     ci.lb
## intrcpt                   0.3344  -99.1437
## depvar2PCTGDP             0.6003   -0.5379
## depvar2logExpPC           0.1971   -0.8724
## indepvar2N                0.3929   -0.5228
## indepvar2logN             0.3535   -0.4059
## year                      0.3309   -0.0168
## publishedNo               0.2093   -0.2218
## elecsys2Non-Majoritarian  0.6451   -0.3443
## methodPANEL               0.4521   -0.4458
## methodIV                  0.8835   -0.6437
## methodRDD                 0.8413   -0.6633
##                             ci.ub
## intrcpt                   33.9528
## depvar2PCTGDP              0.3122
## depvar2logExpPC            0.1817
## indepvar2N                 0.2067
## indepvar2logN              1.1278
## year                       0.0495
## publishedNo                1.0027
## elecsys2Non-Majoritarian   0.5538
## methodPANEL                0.1997
## methodIV                   0.5548
## methodRDD                  0.5411
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

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

```
##
## Test of Moderators (coefficients 2:11):
```

```
## F(df1 = 10, df2 = 131) = 1.3302, p-val* = 0.0030
##
## Model Results:
##
##                             estimate
## intrcpt                     -32.5954
## depvar2PCTGDP                -0.1128
## depvar2logExpPC              -0.3454
## indepvar2N                   -0.1581
## indepvar2logN                 0.3610
## year                          0.0164
## publishedNo                   0.3905
## elecsys2Non-Majoritarian      0.1048
## methodPANEL                  -0.1230
## methodIV                     -0.0445
## methodRDD                    -0.0611
##                                  se     tval
## intrcpt                     33.6402  -0.9689
## depvar2PCTGDP                0.2148  -0.5252
## depvar2logExpPC              0.2664  -1.2963
## indepvar2N                   0.1844  -0.8573
## indepvar2logN                0.3876   0.9312
## year                         0.0168   0.9759
## publishedNo                  0.3095   1.2617
## elecsys2Non-Majoritarian     0.2270   0.4616
## methodPANEL                  0.1631  -0.7542
## methodIV                     0.3029  -0.1468
## methodRDD                    0.3044  -0.2006
##                              pval*     ci.lb
## intrcpt                     0.1580  -99.1437
## depvar2PCTGDP               0.4100   -0.5379
## depvar2logExpPC             0.0530   -0.8724
## indepvar2N                  0.1870   -0.5228
```

```
## indepvar2logN             0.1660   -0.4059
## year                      0.1540   -0.0168
## publishedNo               0.0680   -0.2218
## elecsys2Non-Majoritarian  0.5170   -0.3443
## methodPANEL               0.2490   -0.4458
## methodIV                  0.8270   -0.6437
## methodRDD                 0.7750   -0.6633
##                             ci.ub
## intrcpt                  33.9528
## depvar2PCTGDP             0.3122
## depvar2logExpPC           0.1817  .
## indepvar2N                0.2067
## indepvar2logN             1.1278
## year                      0.0495
## publishedNo               1.0027  .
## elecsys2Non-Majoritarian   0.5538
## methodPANEL               0.1997
## methodIV                  0.5548
## methodRDD                 0.5411
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

L Auxiliary Functions

L.1 Function to Generate Meta-Analytic Figures

This function receives the meta-analysis results and builds 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(

    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)
}


# Build plot function for forest plots to heterogeneous subgroup analysis
build_forest_het <- function(mod, capt, lsize = 22, ttl = NULL, hetvar = NULL) {
  mod <- update(mod, byvar = hetvar, print.byvar = F)
```

```
# Build dataset for plot
mod2 <- tibble(
  byvar = mod$byvar,
  TE = mod$TE,
  seTE = mod$seTE,
  studlab = mod$studlab,
  lower = mod$lower,
  upper = mod$upper,
  group = "A") %>%
  arrange(byvar)
auxmod <- tibble()
for (i in rev(mod$bylevs)){
  auxmod <- rbind(auxmod, tibble(byvar = i,
                  TE = NA,
                  seTE = NA,
                  studlab = toupper(i),
                  lower = NA,
                  upper = NA,
                  group = "B"))
  auxmod <- rbind(auxmod,
                  filter(mod2, byvar==i) %>%
                    arrange(desc(TE)))
  auxmod <- rbind(auxmod, tibble(
    byvar = i,
    TE = mod$TE.random.w[which(mod$bylevs==i)],
    seTE = mod$seTE.random.w[which(mod$bylevs==i)],
    studlab = 'Subgroup Effect',
    lower = mod$lower.random.w[which(mod$bylevs==i)],
    upper = mod$upper.random.w[which(mod$bylevs==i)],
    group = "B"))
}
auxmod <- rbind(auxmod, tibble(
  byvar = NA,
```

```r
  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"))
mod2 <- data.frame(auxmod)
mod2$byvar <- toupper(mod2$byvar)
TEaux <- mod2$TE
TEaux[mod2$studlab== 'Subgroup Effect'] = TEaux[mod2$studlab== 'Subgroup Effect'] - 100

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

# Build plot
p <- mod2 %>%
  ggplot(aes(y = reorder(studlab, TEaux),
             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(byvar~., 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)
}
```

L.2 Webscraping Code

We used the code below to download the articles cited in our paper.

```
# Required packages
pkgs <- c("tidyverse", "rvest", "RSelenium")


# Install the packages if necessary
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 packages
lapply(pkgs, require, character.only = T)
rm(pkgs, installIfNot)
```

```r
# 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
)

k <- 0

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

  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 ####
```

```r
# 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/router-view/
  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")
```

M Session Information

```r
sessionInfo()
```

```
## R version 4.0.2 (2020-06-22)
## Platform: x86_64-apple-darwin17.0 (64-bit)
## Running under: macOS Catalina 10.15.7
##
## Matrix products: default
## BLAS:   /Library/Frameworks/R.framework/Versions/4.0/Resources/lib/libRblas.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/4.0/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      stats     graphics  grDevices
## [5] utils     datasets  methods   base
##
## other attached packages:
##  [1] compareGroups_4.4.6
##  [2] magick_2.5.2
##  [3] kableExtra_1.3.1
##  [4] ggpubr_0.4.0
##  [5] gridExtra_2.3
##  [6] gridGraphics_0.5-0
##  [7] knitr_1.30
##  [8] data.table_1.13.2
##  [9] devtools_2.3.1
## [10] usethis_1.6.3
## [11] readxl_1.3.1
## [12] metafor_2.4-0
## [13] Matrix_1.2-18
## [14] meta_4.15-1
```

```
## [15] forcats_0.5.0
## [16] stringr_1.4.0
## [17] dplyr_1.0.2
## [18] purrr_0.3.4
## [19] readr_1.3.1
## [20] tidyr_1.1.2
## [21] tibble_3.0.4
## [22] ggplot2_3.3.2
## [23] tidyverse_1.3.0
## [24] rmarkdown_2.5
## [25] nvimcom_0.9-102
##
## loaded via a namespace (and not attached):
##  [1] minqa_1.2.4
##  [2] colorspace_2.0-0
##  [3] ggsignif_0.6.0
##  [4] ellipsis_0.3.1
##  [5] rio_0.5.16
##  [6] rprojroot_1.3-2
##  [7] flextable_0.6.0
##  [8] base64enc_0.1-3
##  [9] fs_1.5.0
## [10] mice_3.12.0
## [11] rstudioapi_0.13
## [12] farver_2.0.3
## [13] remotes_2.2.0
## [14] fansi_0.4.1
## [15] lubridate_1.7.9
## [16] xml2_1.3.2
## [17] codetools_0.2-16
## [18] splines_4.0.2
## [19] pkgload_1.1.0
## [20] jsonlite_1.7.1
```

```
## [21] nloptr_1.2.2.2

## [22] broom_0.7.2

## [23] dbplyr_1.4.4

## [24] compiler_4.0.2

## [25] httr_1.4.2

## [26] backports_1.2.0

## [27] assertthat_0.2.1

## [28] cli_2.2.0

## [29] htmltools_0.5.0

## [30] prettyunits_1.1.1

## [31] tools_4.0.2

## [32] gtable_0.3.0

## [33] glue_1.4.2

## [34] tinytex_0.27

## [35] Rcpp_1.0.5

## [36] carData_3.0-4

## [37] cellranger_1.1.0

## [38] vctrs_0.3.5

## [39] writexl_1.3.1

## [40] nlme_3.1-148

## [41] xfun_0.19

## [42] ps_1.4.0

## [43] openxlsx_4.2.3

## [44] testthat_2.3.2

## [45] lme4_1.1-26

## [46] rvest_0.3.6

## [47] CompQuadForm_1.4.3

## [48] lifecycle_0.2.0

## [49] statmod_1.4.35

## [50] rstatix_0.6.0

## [51] MASS_7.3-51.6

## [52] scales_1.1.1

## [53] hms_0.5.3
```

```
## [54] parallel_4.0.2
## [55] HardyWeinberg_1.6.8
## [56] yaml_2.2.1
## [57] curl_4.3
## [58] memoise_1.1.0
## [59] gdtools_0.2.2
## [60] stringi_1.5.3
## [61] highr_0.8
## [62] desc_1.2.0
## [63] boot_1.3-25
## [64] pkgbuild_1.1.0
## [65] zip_2.1.1
## [66] truncnorm_1.0-8
## [67] chron_2.3-56
## [68] systemfonts_0.3.2
## [69] rlang_0.4.9
## [70] pkgconfig_2.0.3
## [71] Rsolnp_1.16
## [72] evaluate_0.14
## [73] lattice_0.20-41
## [74] labeling_0.4.2
## [75] processx_3.4.5
## [76] tidyselect_1.1.0
## [77] magrittr_2.0.1
## [78] R6_2.5.0
## [79] generics_0.1.0
## [80] DBI_1.1.0
## [81] pillar_1.4.7
## [82] haven_2.3.1
## [83] foreign_0.8-80
## [84] withr_2.3.0
## [85] survival_3.1-12
## [86] abind_1.4-5
```

```
## [87] modelr_0.1.8

## [88] crayon_1.3.4

## [89] car_3.0-10

## [90] uuid_0.1-4

## [91] officer_0.3.15

## [92] blob_1.2.1

## [93] callr_3.5.1

## [94] reprex_0.3.0

## [95] digest_0.6.27

## [96] webshot_0.5.2

## [97] munsell_0.5.0

## [98] viridisLite_0.3.0

## [99] sessioninfo_1.1.1
```