Supplemental Appendix to "Drawing Your Senator From a Jar: Term Length and Legislative Behavior"

Rocío Titiunik*
Assistant Professor
Department of Political Science
University of Michigan

Intended for Online Publication Only

June 6, 2014

Contents

*titiunik@umich.edu, http://www.umich.edu/~titiunik, Department of Political Science, University of Michigan, 5700 Haven Hall, 505 South State Street, Ann Arbor, MI 48109-1045

1 Overview

This supplemental appendix to the paper "Drawing Your Senator From a Jar: Term Length and Legislative Behavior" is intended for online publication only. Section 2 presents the disaggregated data on term length that was used to produce Table 1 in the main body of the paper. Section 3 provides the tables that were used to produce Figures 4 through 7 in the main paper. Next, Section 4 shows power calculations based on a two-sample t-test and standard deviations for various outcomes estimated from the data in each session. Finally, in Section 5 I consider whether some of the institutional features of the state senates considered in the main paper might induce senators to engage in intertemporal strategic behavior, which would induce a violation of the Stable Unit Treatment Value Assumption (SUTVA) and preclude a simple policy recommendation based on the effects reported in the paper.

2 Term length in comparative perspective

This section presents the information presented in Table 1 in the main body of the paper, disaggregated by country and chamber. The countries included are all countries with Polity score ≥ 8, obtained from the Polity IV Dataset (2013) (see also Marshall, Gurr, and Jaggers 2013), for which term length information was available from the Inter-Parliamentary Union (2014).

Table A1: Term Length in National Parliaments – Seventy Countries

Country	Upper or Single Chamber		Lower Chamber	
	Term Length	Staggered	Term Length	Staggered
Albania	4	No	-	-
Argentina	6	Yes	4	Yes
Australia	6	Yes	3	No
Austria	5	No	Variable	No
Belgium	4	No	4	No
Botswana	5	No	-	-
Brazil	8	Yes	4	No
Bulgaria	4	No	-	-
Canada	Continuous	No	4	No
Cape Verde	5	No	-	-
Chile	8	Yes	4	No
Comoros	5	No	-	-
Costa Rica	4	No	-	-
Croatia	4	No	-	-
Cyprus	5	No	-	-
Czech Republic	6	Yes	4	No
Denmark	4	No	-	-
Dominican Republic	4	No	4	No
El Salvador	3	No	-	-
Estonia	4	No	-	-
Finland	4	No	-	-
France	6	No	5	No
Germany	4	No	Variable	No
Ghana	4	No	-	-
Greece	4	No	-	-
Guatemala	4	No	-	-
Hungary	4	No	-	-
India	5	No	6	Yes
Indonesia	5	No	-	-
Ireland	5	No	5	No
Israel	4	No	-	-
Italy	5	No	5	No
Jamaica	5	No	5	No
Japan	4	No	6	Yes
Kenya	5	No	5	No

Continued on next page

Table A1 Continued: Term Length in National Parliaments – Seventy Countries

Country	Upper or Single Chamber		Lower Chamber	
	Term Length	Staggered	Term Length	Staggered
South Korea	4	No	-	-
Latvia	4	No	-	-
Lesotho	5	No	5	No
Lithuania	4	No	-	-
Luxembourg	5	No	-	-
Macedonia	4	No	-	-
Mauritius	5	No	-	-
Mexico	6	No	3	No
Moldova	4	No	-	-
Mongolia	4	No	-	-
Montenegro	4	No	-	-
Netherlands	4	No	4	No
New Zealand	3	No	-	-
Nicaragua	5	No	-	-
Norway	4	No	-	-
Panama	5	No	-	-
Paraguay	5	No	5	No
Peru	5	No	-	-
Philippines	6	Yes	3	No
Poland	4	No	4	No
Portugal	4	No	-	-
Romania	4	No	4	No
Serbia	4	No	-	-
Slovak Republic	4	No	-	-
Slovenia	5	No	4	No
Solomon Islands	4	No	-	-
South Africa	5	No	5	No
Spain	4	No	4	No
Sweden	4	No	-	-
Switzerland	4	No	4	No
Trinidad And Tobago	5	No	5	No
Turkey	4	No	-	-
United Kingdom	Continuous	No	5	No
United States	6	Yes	2	No
Uruguay	5	No	5	No

Note: Countries with Polity score ≥ 8 for which term length information was available. Polity score obtained from the Polity IV Dataset (2013); term length information obtained from the Inter-Parliamentary Union (2014).

3 Outcome tables and randomization inference balance tests

Figures A1 and A2 provide balance tests analogous to those reported in Figures 1 and 2 in the main body of the paper, but using exact (simulation-based) randomization inference p-values instead of the t distributional approximation (see Fisher (1935) and Rosenbaum (2002) for details).

Figure A1: Randomization Inference Balance Tests for Texas

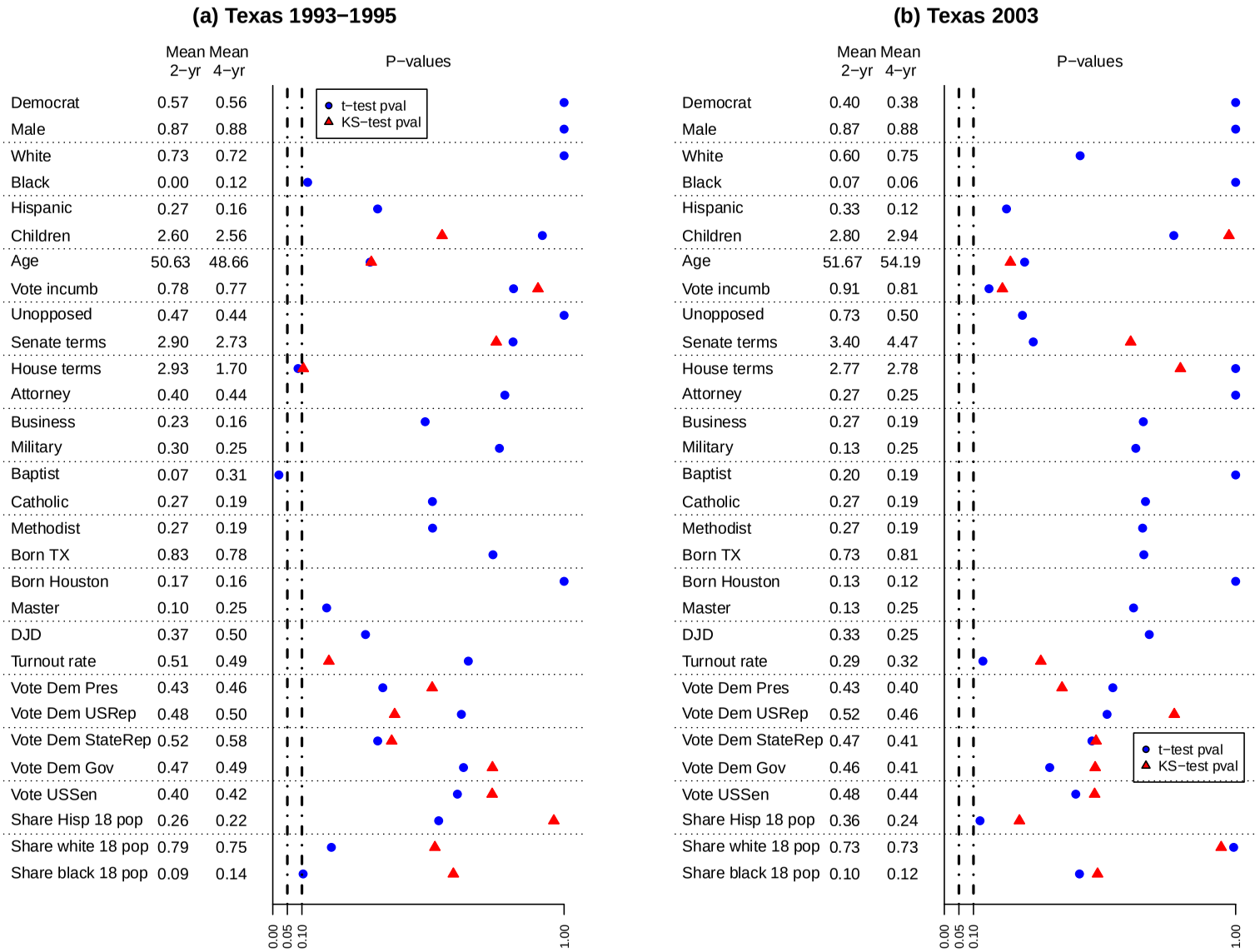
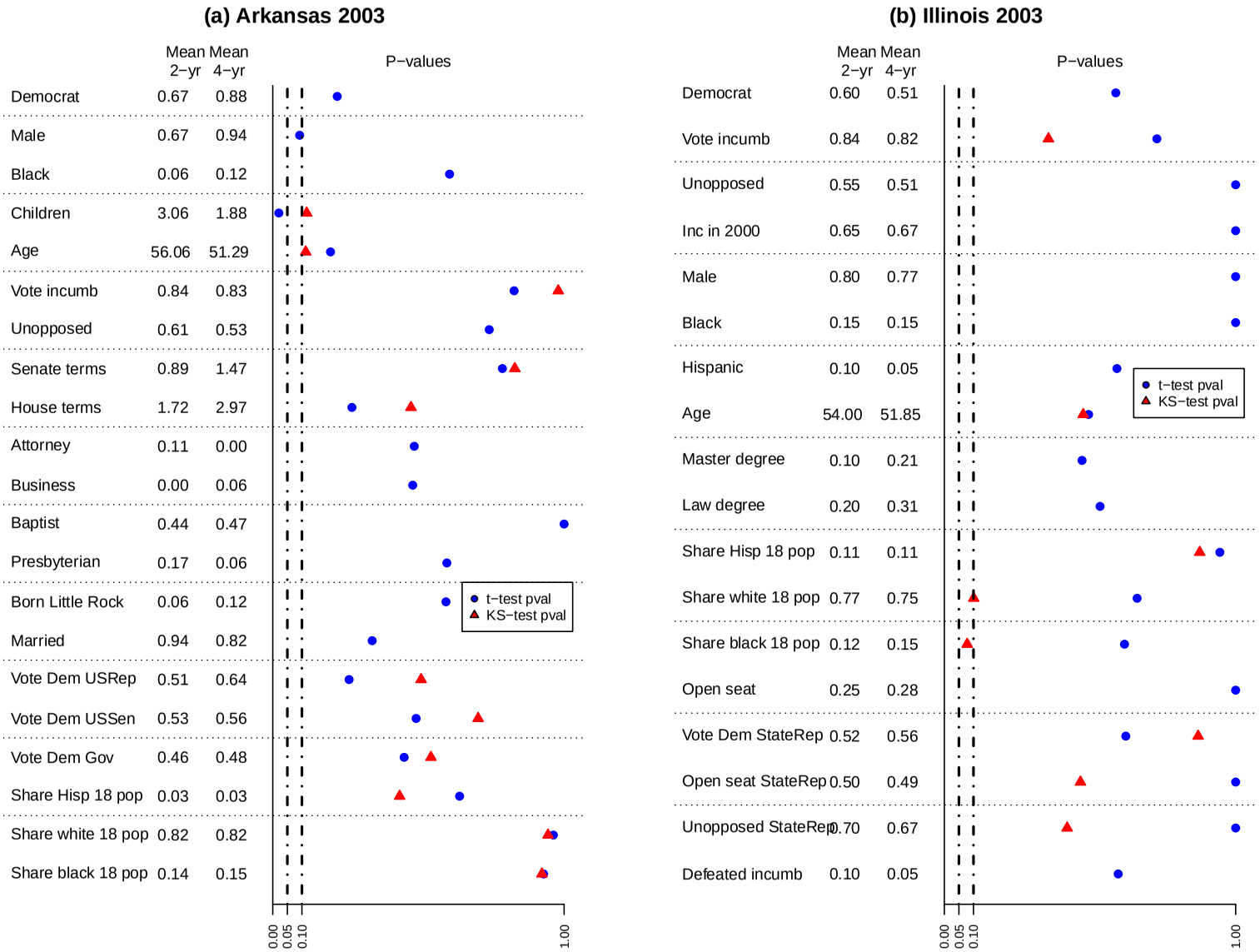


Figure A2: Randomization Inference Balance Tests for Arkansas and Illinois



Tables A2 through A5 below provide more details about the results illustrated in Figures 4 through 7 in the main body of the paper.

Table A2: Effect of Term Length on Various Outcomes, Arkansas 2003 Legislative Session

Outcome	Mean 2yr	Mean 4yr	Diff	SE	P-val	CIl	CIu
Abstention Rate	0.031	0.037	-0.006	0.010	0.542	-0.026	0.014
Bills Introduced	20.610	30.710	-10.090	3.347	0.005	-16.889	-3.300
Nominate Score	0.450	0.370	0.080	0.107	0.436	-0.133	0.303
Legislator-District Distance	6.110	6.590	-0.480	1.535	0.758	-3.594	2.639
Contributions – First Semesters	27.510	2.130	25.380	4.011	0.000	17.235	33.520
Contributions – Last Semester	19.110	2.010	17.100	4.926	0.001	7.104	27.106
Expenditures – First Semesters	19.160	5.700	13.470	4.941	0.010	3.436	23.496
Expenditures – Last Semester	23.750	1.690	22.060	5.922	0.001	10.039	34.084

Note: t-tests of difference in means for several outcomes. Columns contain the following information, respectively: mean among senators serving 2-year terms, mean among senators serving 4-year terms, difference-in-means, standard error of the difference-in-means, p-value corresponding to a t-test of the null hypothesis that means in treated and control groups are equal, lower bound of 95-percent confidence interval associated with the t-test, lower bound of 95-percent confidence interval associated with the t-test. There are 17 senators serving 2 years and 18 senators serving 4 years.

Table A3: Effect of Term Length on Various Outcomes, Illinois 2003 Legislative Session

Outcome	Mean 2yr	Mean 4yr	Diff	SE	P-val	CIl	CIu
Abstention Rate	0.111	0.107	0.004	0.038	0.913	-0.073	0.081
Bills Introduced	79.950	67.620	12.330	17.802	0.491	-23.288	47.957
Nominate Score	0.700	0.770	-0.070	0.045	0.129	-0.159	0.021
Legislator-District Distance	9.450	9.360	0.090	1.997	0.964	-3.904	4.086
Contributions – First Semesters	223.030	195.940	27.090	78.466	0.731	-130.037	184.213
Contributions – Last Semester	238.850	102.380	136.470	69.301	0.054	-2.304	275.244
Expenditures – First Semesters	157.380	164.880	-7.500	45.949	0.871	-99.511	84.512
Expenditures – Last Semester	263.010	96.390	166.620	90.404	0.071	-14.411	347.651

Note: t-tests of difference in means for several outcomes. Columns contain the following information, respectively: mean among senators serving 2-year terms, mean among senators serving 4-year terms, difference-in-means, standard error of the difference-in-means, p-value corresponding to a t-test of the null hypothesis that means in treated and control groups are equal, lower bound of 95-percent confidence interval associated with the t-test, lower bound of 95-percent confidence interval associated with the t-test. There are 19 senators serving 2 years and 40 senators serving 4 years.

Table A4: Effect of Term Length on Various Outcomes, Texas 1993 and 1995 Legislative Sessions (pooled)

Outcome	Mean 2yr	Mean 4yr	Diff	SE	P-val	CIl	CIu
Abstention Rate	0.045	0.024	0.022	0.011	0.061	-0.001	0.044
Bills Introduced	50.230	50.720	-0.490	6.948	0.945	-14.374	13.403
Nominate Score	0.590	0.710	-0.120	0.075	0.115	-0.270	0.030
Legislator-District Distance	4.970	5.160	-0.190	0.869	0.828	-1.926	1.547
Contributions – First Semesters	101.050	81.100	19.950	20.397	0.332	-20.824	60.724
Contributions – Last Semester	168.380	162.520	5.860	35.776	0.870	-65.657	77.373
Expenditures – First Semesters	107.950	96.310	11.640	22.078	0.600	-32.497	55.772
Expenditures – Last Semester	128.920	126.420	2.500	40.315	0.951	-78.093	83.086

Note: t-tests of difference in means for several outcomes. Columns contain the following information, respectively: mean among senators serving 2-year terms, mean among senators serving 4-year terms, difference-in-means, standard error of the difference-in-means, p-value corresponding to a t-test of the null hypothesis that means in treated and control groups are equal, lower bound of 95-percent confidence interval associated with the t-test, lower bound of 95-percent confidence interval associated with the t-test. The analysis pools observations from the 1993 and 1995 Legislative Sessions; there are 30 senators serving 2 years and 32 senators serving 4 years.

Table A5: Effect of Term Length on Various Outcomes, Texas 2003 Legislative Session

Outcome	Mean 2yr	Mean 4yr	Diff	SE	P-val	CIl	CIu
Abstention Rate	0.029	0.011	0.018	0.006	0.005	0.006	0.030
Bills Introduced	60.130	76.880	-16.740	9.470	0.087	-36.056	2.573
Nominate Score	0.740	0.660	0.090	0.062	0.173	-0.040	0.214
Legislator-District Distance	4.930	7.620	-2.690	1.700	0.123	-6.158	0.775
Contributions – First Semesters	259.620	212.120	47.490	44.072	0.290	-42.393	137.378
Contributions – Last Semester	195.770	237.560	-41.790	42.152	0.329	-127.762	44.176
Expenditures – First Semesters	280.770	208.310	72.460	48.206	0.143	-25.852	170.781
Expenditures – Last Semester	105.680	86.430	19.250	27.865	0.495	-37.582	76.079

Note: t-tests of difference in means for several outcomes. Columns contain the following information, respectively: mean among senators serving 2-year terms, mean among senators serving 4-year terms, difference-in-means, standard error of the difference-in-means, p-value corresponding to a t-test of the null hypothesis that means in treated and control groups are equal, lower bound of 95-percent confidence interval associated with the t-test, lower bound of 95-percent confidence interval associated with the t-test. There are 15 senators serving 2 years and 16 senators serving 4 years.

4 Power Calculations

This section presents simple power calculations to provide some information regarding the probability of detecting effects of different size using the small-n experimental data analyzed in the main body of the paper.

All calculations are illustrated graphically in Figures A3 through A11 below. All figures plot power (the probability of rejecting the null hypothesis of no effect given that the alternative hypothesis of a non-zero effect is true) for different values of the alternative hypothesis. In particular, the power calculations assume a two-sample t-test of difference means with a two-sided alternative and non-equal sample sizes. There is one figure corresponding to each combination of outcome and legislative session (with Texas' 1993 and 1995 analyzed jointly as in the main paper). Each of these combinations uses the sample sizes of the particular legislative session considered, and estimates the standard deviation of the outcome from the data corresponding to the particular outcome of the particular legislative session considered.

The effect sizes in the x-axis are displayed in the natural units of the outcomes (number of bills, proportion of abstentions, etc), but in all figures the *range* of the effect sizes goes from 0.1 to 1 standard deviations (SD) of the outcome, in increments of 0.05 SDs. The figures thus show the different effect sizes that are needed for every combination of sample size to produced effects within 0.1 and one standard deviations.

For example, Figure A3 (which corresponds to the scenario in Arkansas for abstention rates) shows that when there are 17 and 18 observations and the SD is equal to 0.03, the power for an effect of 0.1 SD is well below 0.2 (first point in the graph, corresponding to an absolute-value difference in abstention rates below 0.005), but for an effect of 1 SD power is above 0.80 (last point in the graph, roughly an absolute-value difference in abstention rates equal to 0.030). All other figures can be interpreted analogously.

Figure A3: Power Calculations Based on Abstention Rates in Arkansas

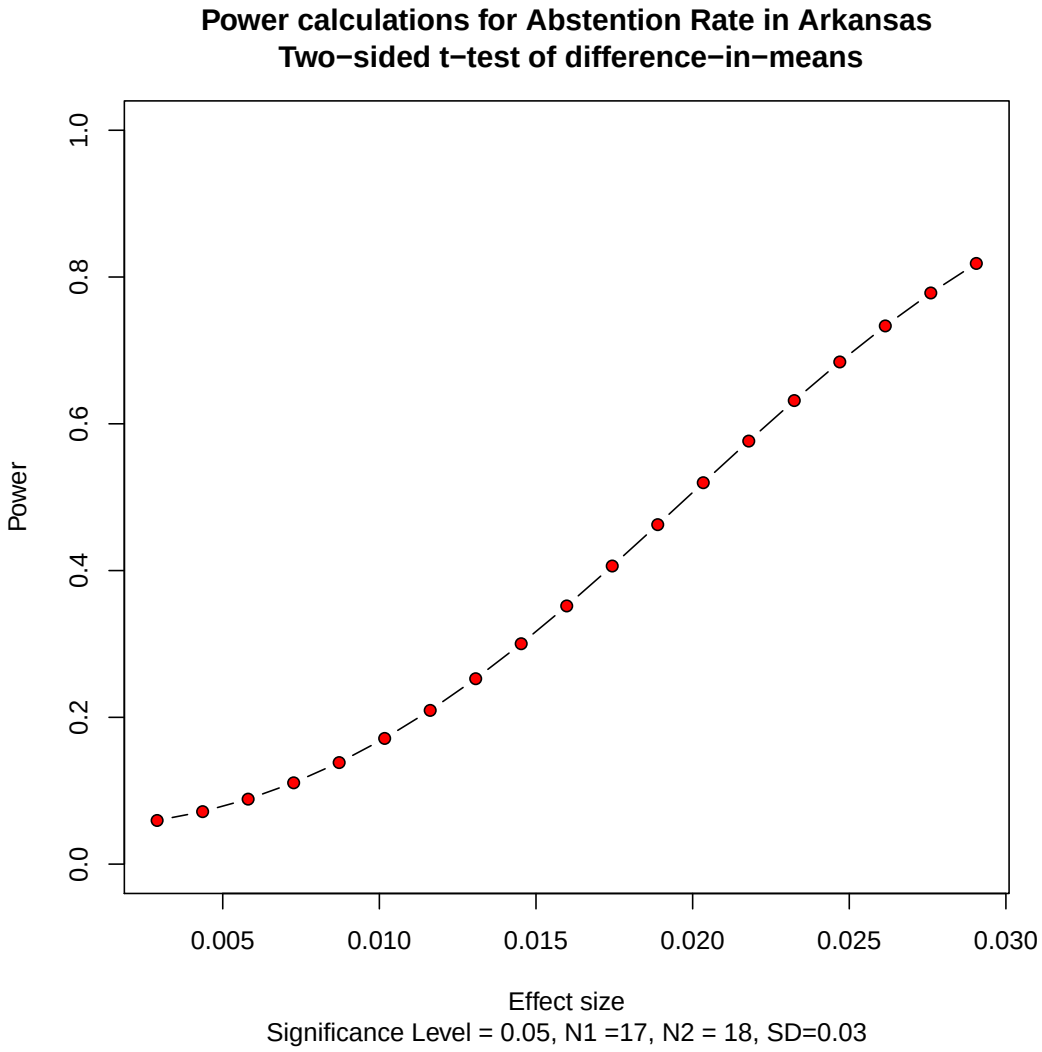


Figure A4: Power Calculations Based on Abstention Rates in Texas

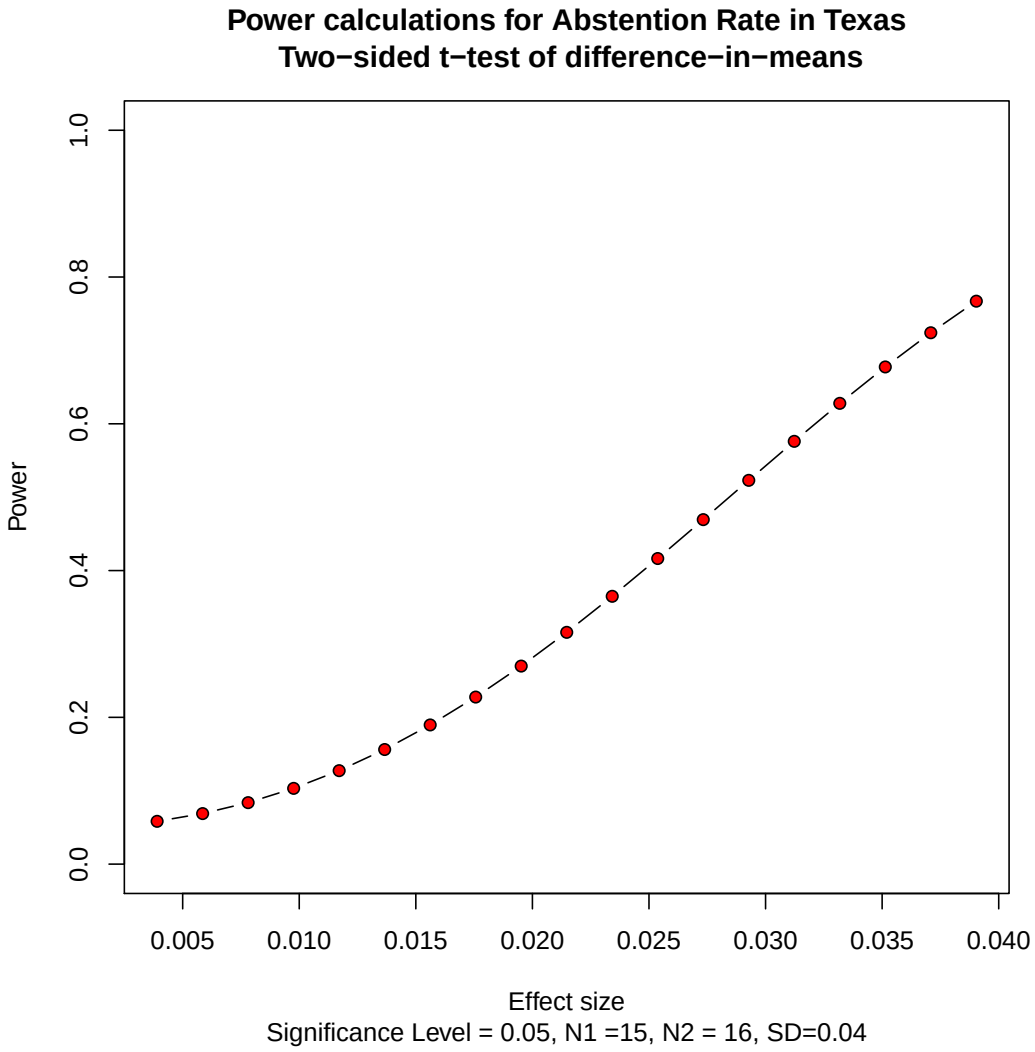


Figure A5: Power Calculations Based on Abstention Rates in Illinois

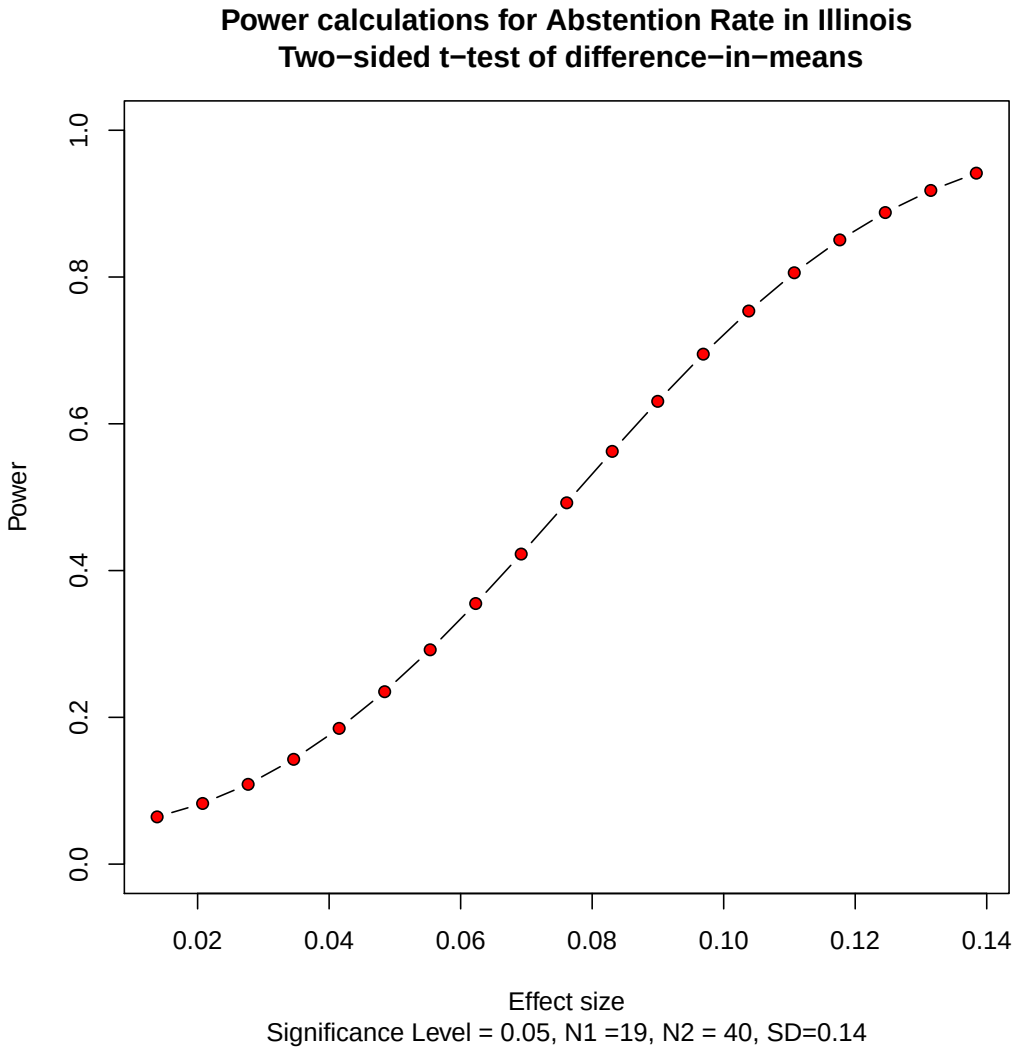


Figure A6: Power Calculations Based on Bills Introduced in Arkansas

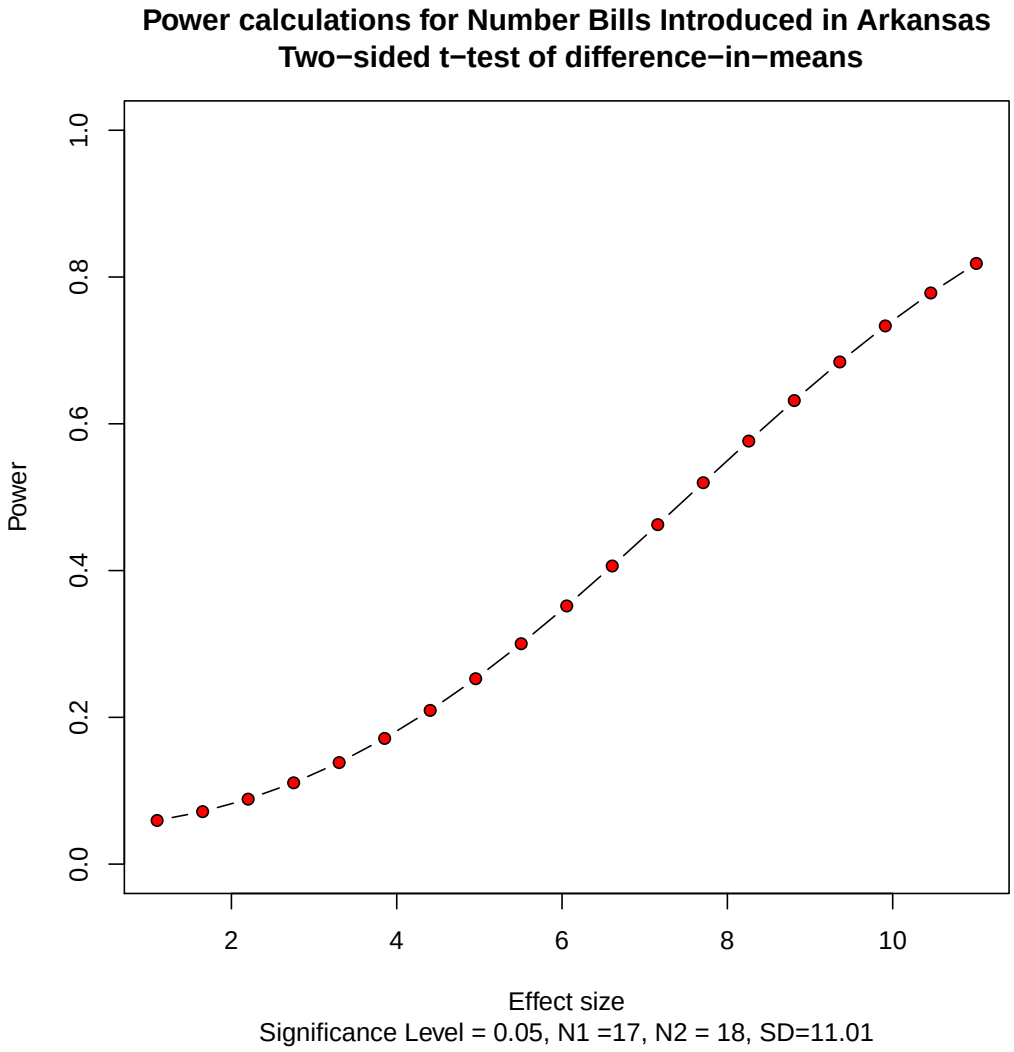


Figure A7: Power Calculations Based on Bills Introduced in Texas

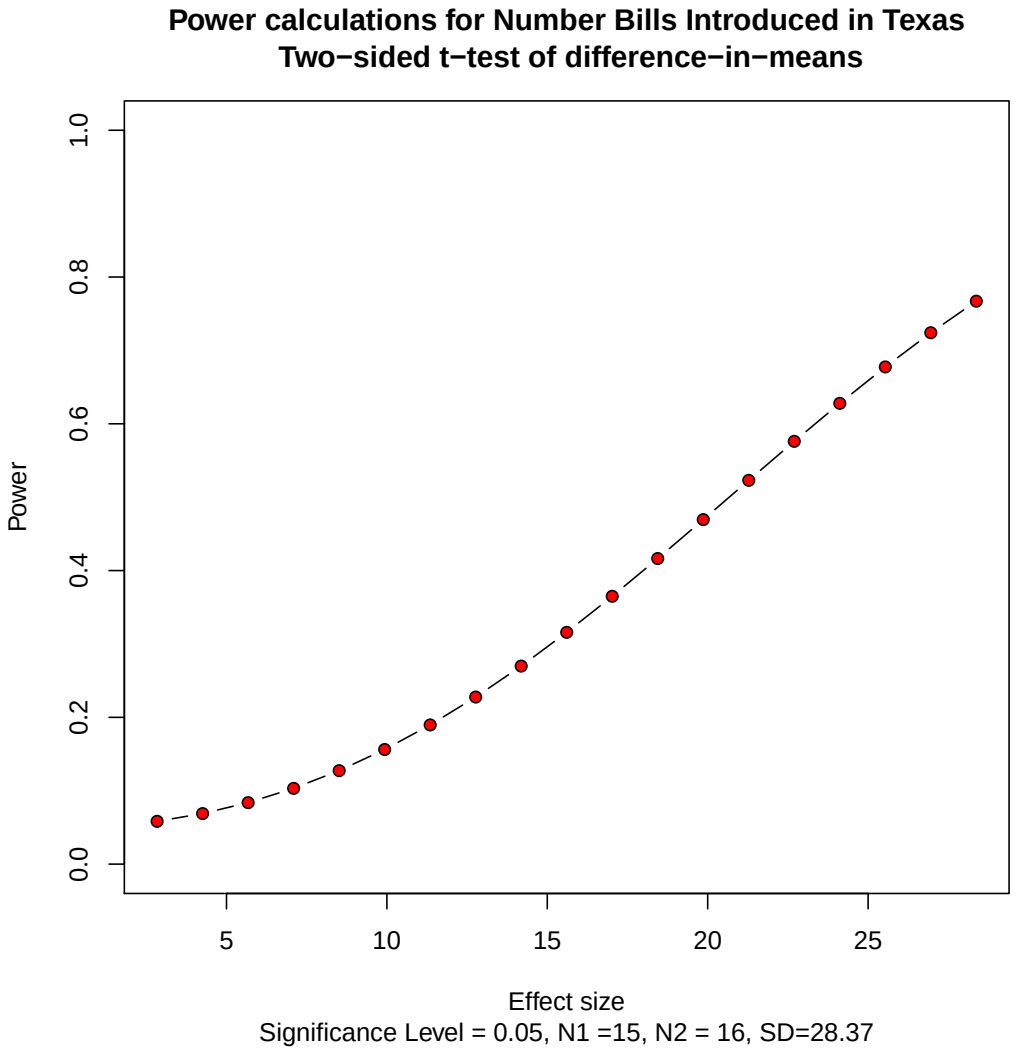


Figure A8: Power Calculations Based on Bills Introduced in Illinois

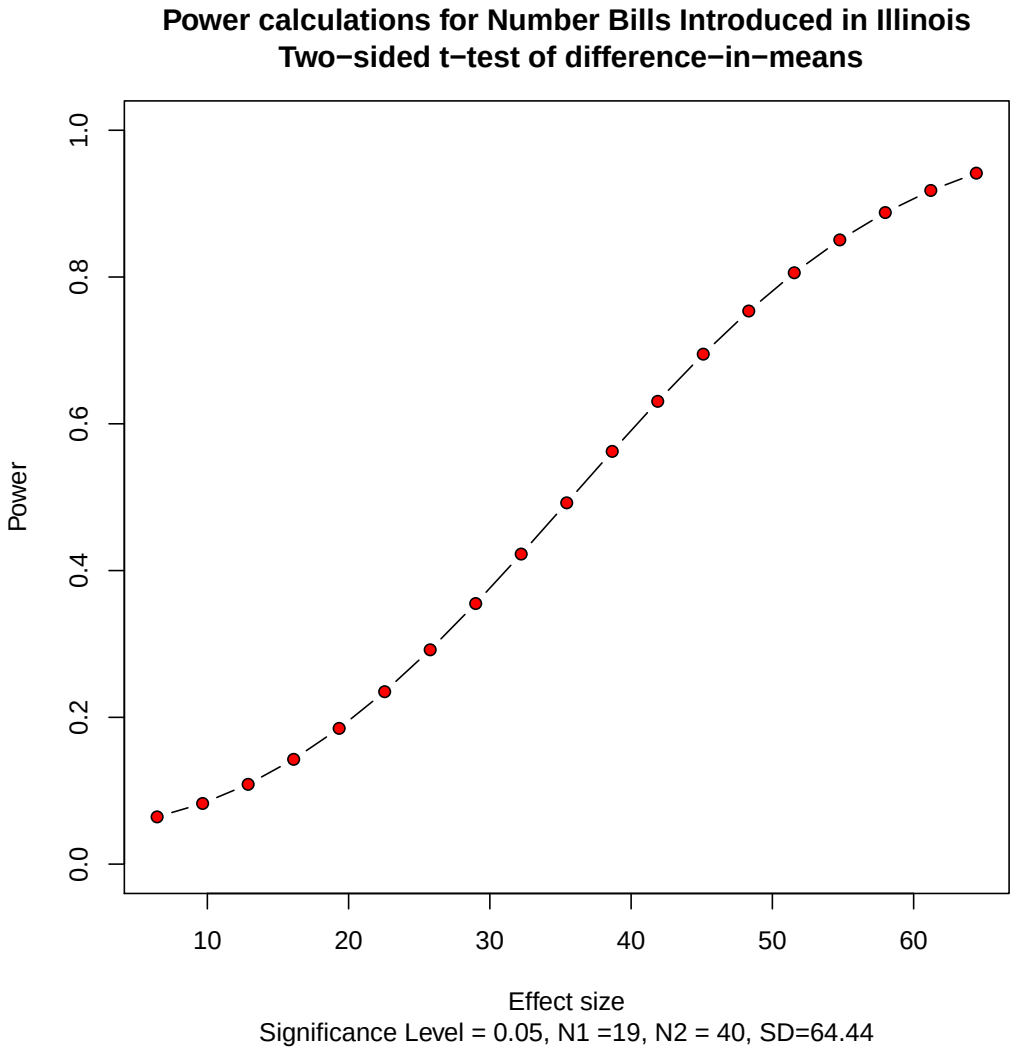


Figure A9: Power Calculations Based on NOMINATE Scores in Arkansas

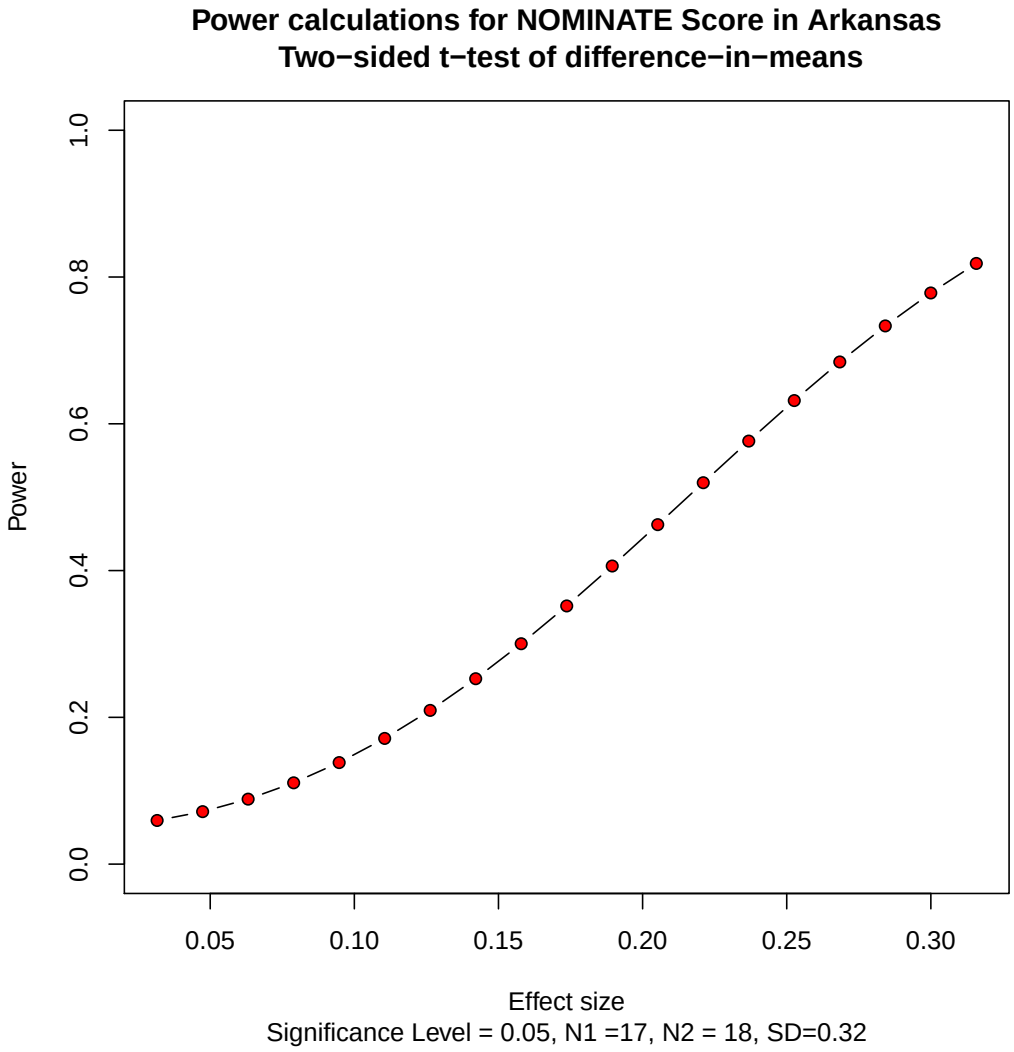


Figure A10: Power Calculations Based on NOMINATE Scores in Texas

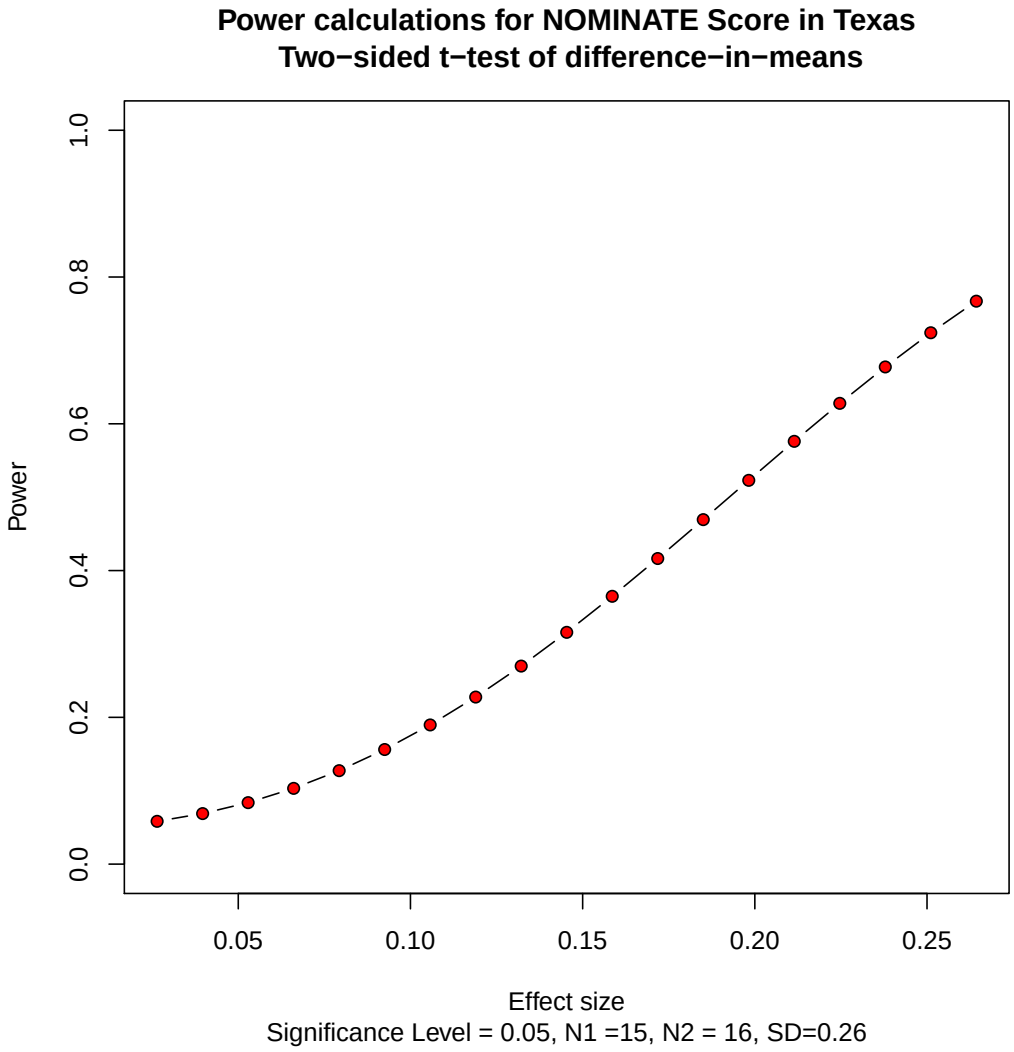
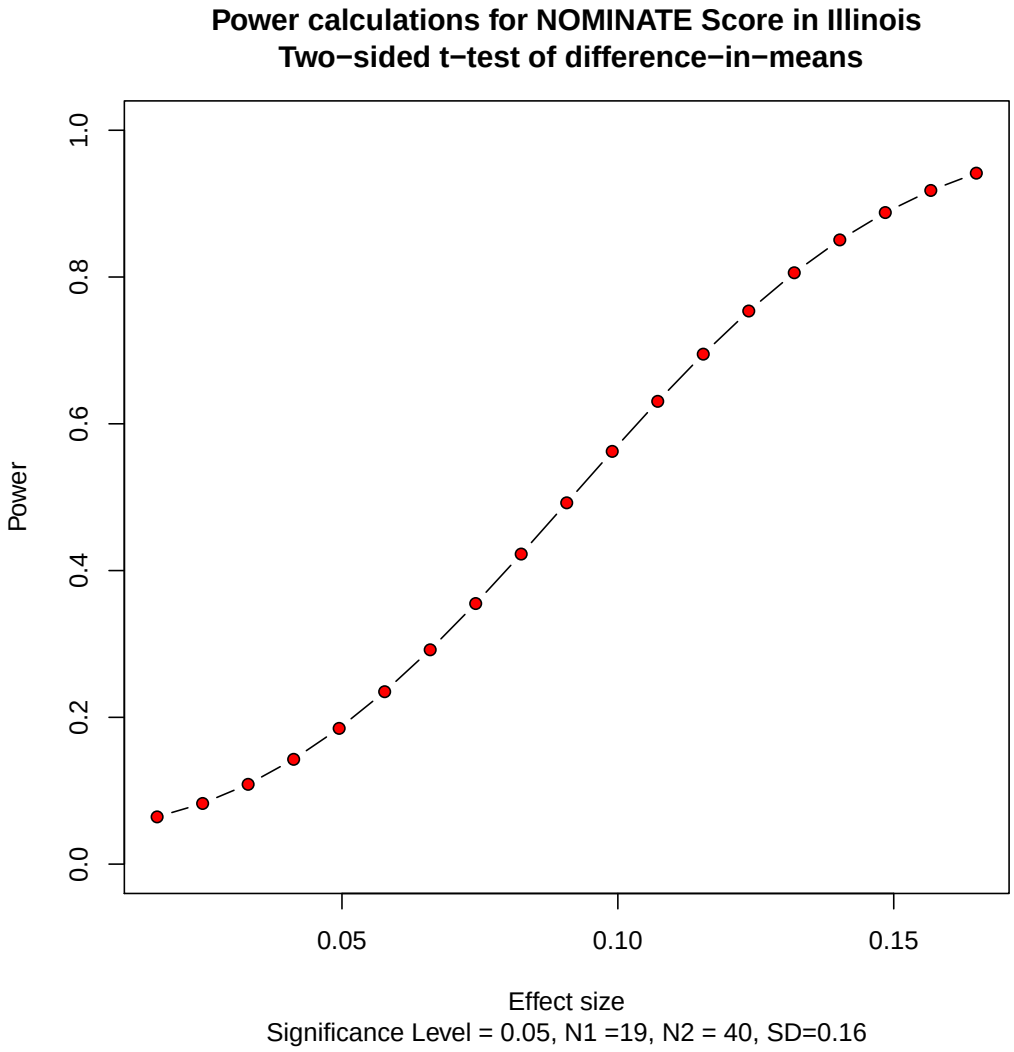


Figure A11: Power Calculations Based on NOMINATE Scores in Illinois



5 SUTVA and the possible limitations imposed by strategic interactions

If senators interacted strategically with each other, the Stable Treatment Unit Value Assumption (SUTVA) might be violated. If this happens, the policy implications of the experimental results presented in the main body of the paper would need to be revised. When SUTVA holds, units do not interfere with each other, in the sense that the outcome of every unit is solely affected by the treatment received by that unit, regardless of the treatment status assigned to the rest of the units participating in the experiment. Formally, letting $\mathbf{T} = (T_1, T_2, \cdots, T_N)$ and $\mathbf{T}' = (T'_1, T'_2, \cdots, T'_N)$ be two N-dimensional vectors that contain the treatment assignment of each of the N units in the experiment, SUTVA is defined as $Y_i^{\mathbf{T}} = Y_i^{\mathbf{T}'}$ if $T_i = T'_i$ (Rubin 1990). In the experiments considered in the main paper, SUTVA requires that a senator assigned a four-year term behave identically whether all other senators are assigned two-year terms, all other senators are assigned four-year terms, or any other possible arrangement of treatment assignment among the remaining senators.

Recent theoretical results developed by Muthoo and Shepsle (2010), Shepsle, Van Houweling, and Dickson (2004) and Shepsle, Van Houweling, Abrams, and Hanson (2009), suggest that the staggered structure of terms might induce strategic intertemporal interactions between legislators. The authors study a divide-the-dollar game in a staggered legislature, where each legislator serves a term of three sessions and a third of the members faces reelection at the end of each session. The model shows that in a staggered legislature, an equilibrium can be sustained in which the different generations of legislators engage in an intertemporal deal according to which in every period the amount to redistribute is shared *only* among those senators who are facing reelection in that period. In this equilibrium, legislators agree to forgo resources early in their terms to aid the generations that are seeking reelection, a behavior that is later reciprocated by their fellow legislators (given a credible punishment regime) and hence allows them to concentrate or "backload" all benefits in their

last period, just before they face reelection.

This theoretical argument is of direct relevance to the term experiments considered in the main body of the paper. If senators engaged in an intertemporal distributional agreement of this sort, the behavior of four-year term senators would not be the behavior that they would have exhibited if the *entire* senate had been elected for four years. The estimated difference between two-year and four-year senators would reflect not only the effect of serving a shorter term, but would also include the compensating effect of the four-year senators, who would be "allowing" two-year senators to behave in the observed way, under the agreement that they would be reciprocated in future sessions. In this case, one could no longer define a *pair* of potential outcomes (Y_{i0}, Y_{i1}) for each senator since individual outcomes would depend on the entire vector of treatment assignments. That is, $Y_i^{\mathbf{T}} \neq Y_i^{\mathbf{T}'}$ with $T_i = T_i'$ and SUTVA would be violated. On the other hand, if these intertemporal deals were not occurring, the results of the experiment would be directly informative of the possible effects decreasing term length for all members of the legislature.

This discussion illustrates that even when the treatment of interest is randomly assigned, understanding the incentives imposed by the institutional structure in which individuals interact is crucial to the correct interpretation of the parameters under study. To approximately test whether senators are behaving according to this model, I take advantage of the "placebo" experiment that occurs in the Texas Senate. As explained in the paper, in 1993, Texas senators were randomly assigned to two-year and four-year terms in the *middle* of the legislative session and, as a consequence, during the first half of the session they ignored who would be running first for reelection. These periods during which *all* senators ignored when their reelection would occur can be used as a baseline to estimate their behavior when intertemporal deals are *not* occurring –under the assumption that when senators ignore whose reelection is coming first an intertemporal deal is impossible to reach. This is a plausible assumption, since before terms are assigned there is no way to know who should be receiving the benefits first, and so "backloading" is undefined.

During these "blind" periods when senators ignore when they will run for reelection, their behavior will not be contaminated by intertemporal deals. Thus, one can observe whether after terms have been assigned both groups deviate from their previous behavior in the blind period. If one observed that, for example, abstention rates increased for the two-year group and simultaneously decreased for the four-year group relative to the blind period abstention rates, then this would provide evidence in favor of the existence of intertemporal logrolling. If, on the other hand, the abstention rates of the two-year group increased but the abstention rates of the other group remained at the same blind-period levels, this would be evidence against the existence of intertemporal deals. In general, finding no difference between the two periods for *only one* of the groups and not the other is the strongest piece of evidence against intertemporal deals, since these deals are by definition zero-sum and require that one group give up resources in favor of the other.

Remarkably, as shown in Table A6, average abstention rates among senators assigned 4-year terms are almost exactly identical before and after the assignment (2.62% versus 2.69%). In contrast, senators assigned 2-year terms more than double their abstention rates after learning that they will only serve for two years. In other words, senators do not change their behavior if they receive long terms, but they do if they receive short terms. This limited piece of evidence is not at all compatible with the kind of SUTVA violation discussed above, according to which both groups should change their behavior to implement an intertemporal deal.

Table A6: Effect of Term Length on Abstention Rate Before and After Terms Are Assigned, Texas 1993 Legislative Session

Outcome	Mean 2yr	Mean 4yr	Diff	SE	P-val	CIl	CIu
Placebo Period	0.0207	0.0262	-0.0054	0.0110	0.6330	-0.0280	0.0170
Non-Placebo Period	0.0447	0.0269	0.0177	0.0090	0.0560	-0.0000	0.0360

Note: t-tests of difference in means for abstention rates in the 1993 Legislative Session before and after the assignment of terms, which was done half way through the session. Placebo period lasts between January and March 1993 before the duration of terms was randomly assigned; Post-Placebo period lasts between assignment of terms and the end of the legislative session. Columns contain the following information, respectively: mean among senators who are later assigned to serve 2-year terms, mean among senators who are later assigned to serve 4-year terms, difference-in-means, standard error of the difference-in-means, p-value corresponding to a t-test of the null hypothesis that means in treated and control groups are equal, lower bound of 95-percent confidence interval associated with the t-test, lower bound of 95-percent confidence interval associated with the t-test. There are 15 senators that are ex-post assgined a 2-year term, and 16 senators that are ex-post assigned a 4-year term.

References

Fisher, Ronald A. 1935. *Design of Experiments*. New York: Hafner.

Inter-Parliamentary Union. 2014. "PARLINE database on national parliaments.". http://www.ipu.org/. Online resource, retrieved May 15, 2014.

Marshall, Monty G., Ted Robert Gurr, and Keith Jaggers. 2013. "Polity IV Project: Political Regime Characteristics and Transitions, 1800-2013. Dataset Users' Manual.".

Muthoo, Abhinay and Keneth A. Shepsle. 2010. "Information, institutions and constitutional arrangements." *Public choice* 144 (1): 1–36.

Polity IV Dataset. 2013. "Polity IV Project.". http://www.systemicpeace.org/polityproject.html. Online resource, retrieved May 15, 2014.

Rosenbaum, Paul R. 2002. *Observational Studies*. New York: Springer-Verlag 2nd edition.

Rubin, Donald B. 1990. "Comment: Neyman (1923) and Causal Inference in Experiments and Observational Studies." *Statistical Science* 5 (4): 472–480.

Shepsle, Keneth A., Robert P. Van Houweling, Samuel J. Abrams, and Peter C. Hanson. 2009. "The Senate Electoral Cycle and Bicameral Appropriations Politics." *American Journal of Political Science* 53 (2): 343–359.

Shepsle, Keneth A., Robert P. Van Houweling, and Eric S. Dickson. 2004. "Bargaining in Legislatures with Overlapping Generations of Politicians." Typescript, Harvard University.