

THE PRACTICAL RESEARCHER

State Legislative Elections, 1967–2003: Announcing the Completion of a Cleaned and Updated Dataset

Thomas M. Carsey, *University of North Carolina*

Richard G. Niemi, *University of Rochester*

William D. Berry, *Florida State University*

Lynda W. Powell, *University of Rochester*

James M. Snyder, Jr., *Massachusetts Institute of Technology*

More than 15 years—nine election cycles—have passed since a comprehensive set of state legislative election data was compiled and made available to researchers and practitioners: the Inter-University Consortium for Political and Social Research's (ICPSR) *State Legislative Election Returns in the United States* dataset (Study #8907) collected by Malcolm Jewell (Jewell 1991) and containing observations from 1967 to 1988.¹ With this hiatus in mind, we set out at various times—initially in three independent efforts (Berry and Carsey; Niemi, and Powell; Snyder)—to gather legislative election data for all states and elections since 1988. In addition, Berry and Carsey (2005) cleaned the original dataset to make it more accurate and usable; their corrections led to the release of a revised ICPSR dataset (Study #3938). The culmination of these efforts is a dataset containing information about general elections for state legislative seats from 1967 to 2003, now available through ICPSR (Study 21480).²

In announcing the availability of the newly cleaned and updated dataset, we thought it useful to present trends for three of the most prominent measures related to state legislative elections: incumbent reelection percentages, open seat percentages, and the percentage of competitive races. We present these trends largely without analysis. While some correlates and tentative explanations are obvious—such as the apparent impact of legislative professionalism (Niemi et al. 2006)—developing full explanatory models of each of

these processes requires a major undertaking (e.g., Berry, Berkman, and Schneiderman 2000), and we leave that to ourselves or others in future work.

THE UPDATED STATE LEGISLATIVE ELECTION DATASET

The dataset begins with incomplete data for 1967. Over time, data is available for every state, although individual years or districts are occasionally missing. For example, no data was compiled for Illinois in the 1970s when it used a unique system of cumulative voting, and data for Vermont is available only from 1986 on. The original dataset collected by Jewell includes information on primary elections for several states, but our updated version of the dataset is limited to general elections only. This limitation is important as scholars consider the nature of intra- versus inter-party competition in legislative elections. The trends we present below in incumbency reelection, open-seat races, and closely-contested races do not consider the nature of the contests within primaries. For example, a seat might be open in a general election either because an incumbent retired or because an incumbent was defeated in the primary.³

The dataset is constructed at the candidate level. It provides information on the year of the election, the chamber, the type of district—single-member or multimember with positions, staggered elections, or free-for-all—a unique district number (or district name when states employ names rather than numbers), the candidates' names and their party labels, and the raw vote totals. Vote totals are also shown as percentages of the total vote for those running in single-member districts. At least two methods for calculating vote percentages for candidates running in free-for-all multimember districts have been proposed. Niemi et al. (1991) suggest pairing candidates in such races across the two major parties, such that the winner with the highest number of votes from one party is paired with the losing candidate from the other major party with the fewest votes, and so forth. Berry, Berkman, and Schneiderman (2000) instead compare the votes received by each winner to the number of votes received by the losing candidate who came closest to winning, regardless of party. The dataset leaves these variables blank for multimember free-for-all races, thus allowing individual researchers to make informed choices for their own research regarding how to measure vote percentages for such contests.

An important feature of the original ICPSR dataset is that candidate names were recorded as they appeared on the election records being used, a reasonable choice from the perspective of archiving historical records. However, it results in complications for scholars seeking to track specific candidates over time because of inconsistencies in how candidate names were presented from

one election cycle to the next. Thus, in one year a candidate's name might be recorded as John Smith but in the next year as John H. Smith or J. Smith. These types of inconsistencies made it nearly impossible to use the dataset to reliably compare candidate continuity from one year to the next. Of course, typographical errors only compound the problem. Berry and Carsey estimated that roughly 35 percent of all candidates who appeared in the original dataset more than once had their names recorded in at least two different ways. Berry and Carsey "cleaned" the names in the original dataset to address this problem. Understandably, this was a time-consuming task that required some judgments along the way. For example, the J. Smith who ran at time $t + 1$ could have been the son or daughter of the John Smith who ran at time t . Berry and Carsey developed a set of formal decision rules to guide a computer-automated name recoding process. That process was followed by a full review of all names and name changes by a team of (mostly) graduate students who themselves followed a formal set of decision rules. For difficult cases, students examined additional materials, such as state records and legislative district maps, to determine whether to change the name. This process resulted in reducing the error rate from the 35 percent noted above to approximately 2 percent. Although some errors remain, the cleaning process makes it feasible to use the names to track candidates over time with a high degree of confidence. The same process for cleaning names was applied to the updated data and used when merging new observations with the originals. Interested readers should consult the codebook for ICPSR Study #3938 for more detailed information on the name-cleaning process.

Owing to its origins as three separate projects, the new dataset is not entirely consistent with respect to its treatment of minor parties. In some cases, votes for all candidates were recorded separately, including write-in candidates. In other cases, votes for candidates that individually received less than 5 percent of the vote were combined in an "other" category. We strongly encourage anyone using the dataset to read carefully the accompanying codebook to learn more about these and other oddities that necessarily emerge when trying to record data across states that employ many different electoral systems and many different methods of reporting results, some of which change within a given state during the period covered.

The variance that makes analysis of state-level processes so attractive to scholars also makes data collection efforts at the state level difficult. Each state has its own process for conducting elections and for reporting and archiving results. Simply locating the basic information required several phone calls, library searches, and hours on state government websites. Nonetheless, the availability of more recent election results online greatly facilitates our efforts. Of course, states adopt different types of legislative districts, different con-

ventions for naming or numbering those districts, and some states changed their type of districts during the time period covered by our dataset. States differ in how they report (or whether they report) write-in candidates. A few states—notably, New York—witness numerous candidates who appear on the ballot more than once, as they are endorsed by different political parties. These are only a few examples of the tremendous institutional and political variance that scholars exploit in studying the states, which also makes collecting state-level data more challenging.

We have made numerous checks on the data, but with many thousands of elections and more than a quarter of a million cases included in the dataset, there are bound to be errors. A few may be in the original data transmitted to us, but others surely occurred when we entered the data. Omissions also remain, as some sets of returns do not report full information. The errors and omissions are not likely to have an appreciable effect on analyses aggregated across states, especially when states and years are combined. Nevertheless, when researchers come across cases that are far out of line with other data points or cases they have reason to question, they should check the accuracy of the data for that case before using it to make a substantive point or creatively “explaining” it.

INCUMBENCY, OPEN SEATS, AND COMPETITION

Although the dataset includes results from 1967 through 2003, we limit our presentation of trends to 1972 through 2002. The importance of redistricting to legislative elections suggests beginning after a decennial Census; in any event, one of our measures—that of incumbency reelection rates—uses results from the prior election year, and 1970 is the first year in which the data is largely complete. Similarly, the dataset includes limited results from 2003.

Providing even a cursory overview, such as we do here, requires a surprising number of decisions. First, we decided on separate displays for single-member districts and multimember districts, in part because of the different meaning of competition in free-for-all elections.⁴ We also decided on separate displays for lower and upper houses, as the greater prestige of senate seats, the larger size of senate districts, and the frequent movement of members of lower houses to upper houses suggest that separate treatment is warranted.⁵ Because only a few states hold elections in odd-numbered years, we combined those elections with the ones held in the immediately following year.⁶

For each figure, we first calculate the mean value of the respective variable for each state for each election year. We then averaged those state means to give an overall average rate for each year. The 20th and 80th percentile scores presented in the figures represent the 20th and 80th percentiles from among the state means. Throughout the analysis, we deleted cases if data was avail-

able for fewer than five districts (i.e., a state-year combination). Since the smallest legislatures contain 20 or more seats, so few recorded cases meant that there was a large amount of missing data, making estimates for such cases unreliable. We report the number of states used to calculate each mean in each figure in parentheses under each year labeled on the horizontal axis.⁷ We do not show results for upper-house multimember districts; in the 1990s, only two states fell into this category.

Incumbent Reelection

We begin with incumbent reelection rates—simply the percentage of winners among incumbents running for reelection (Figures 1–3). In each figure, we show the path of the mean along with the 20th and 80th percentiles. In lower house single-member districts (SMDs), the mean reelection rate was more than 90 percent in every election except 1974; even so, the rate has increased over the past three decades and is now on the order of 95 percent. Perhaps due to strategic retirements, the rate does not dip in the years immediately following redistricting. Of course, in any given year there are states with somewhat lower reelection rates, although the 20th percentile has also generally risen over time, and only rarely has the rate in any given state dropped below 80 percent. In multimember districts (MMDs) reelection rates are slightly lower, but they exhibit a similar increase over time and are now on the order of 90 percent (Figure 3).

Figure 1. Average Percentage of Incumbents Re-Elected, Single-Member Districts, State Lower Houses

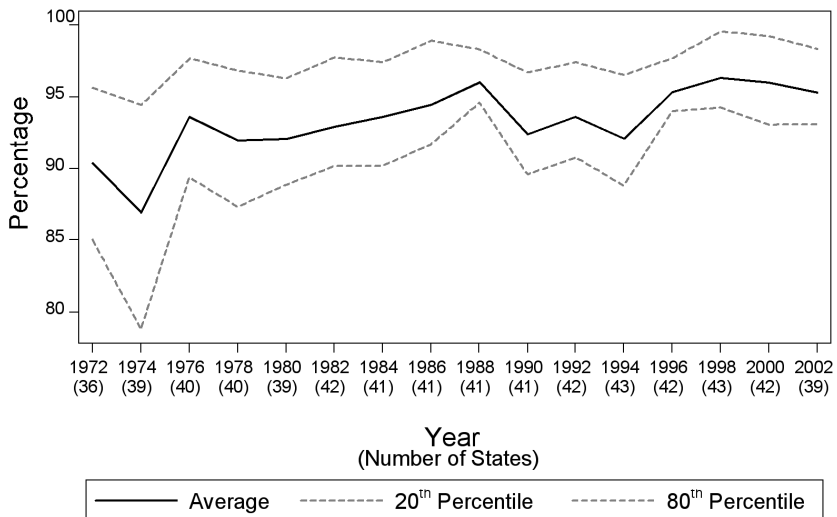


Figure 2. Average Percentage of Incumbents Re-Elected, Single-Member Districts, State Upper Houses

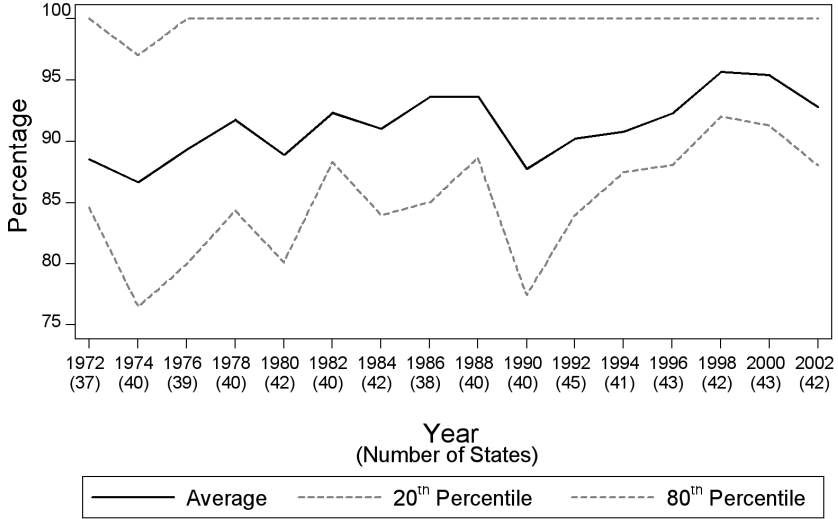
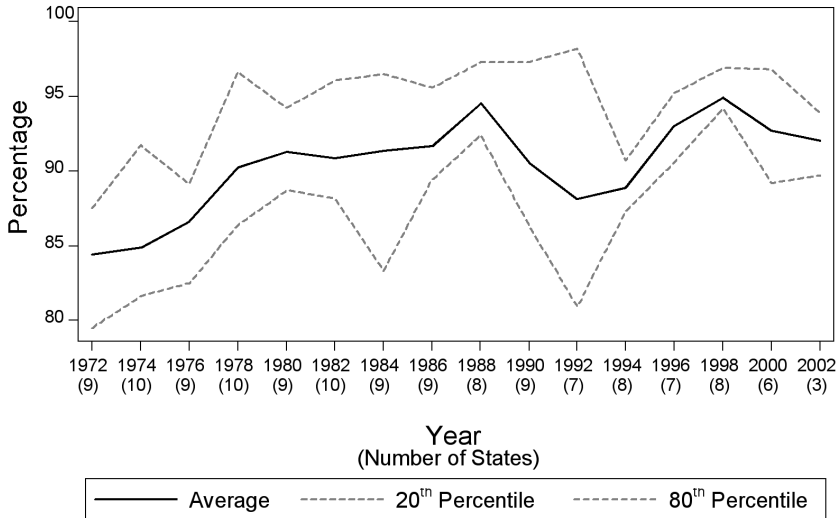


Figure 3. Average Percentage of Incumbents Re-Elected, Multimember Districts, State Lower Houses



In state senates, as noted, we deal only with SMDs. We anticipate that incumbent reelection rates will be slightly lower, as they are for the U.S. senate compared to the U.S. House (Stanley and Niemi 2007, 55–56). Presumably, this occurs because the prestige of the higher office and a smaller number of seats result in more qualified challengers competing for those seats. As it turns out, the differences between upper and lower chambers are very slight and not entirely consistent (Figure 2). For most years, reelection rates are still above 90 percent, and they have risen over time but with the same dip that is seen in lower house elections in the early 1990s. Nonetheless, the within-year variability is greater in the senates, as shown by the wider percentile bands.⁸ At the high end, there are many states (almost half in the two most recent elections) in which not a single incumbent senator was defeated. Yet there were also more instances in the upper house where the reelection rate dropped below 85 percent as compared to the lower house.

Open Seats

High rates of incumbency reelection are more or less consequential for a state depending on the number of incumbents who run for reelection. In the U.S. House, this percentage has been on the order of 90 percent at least since the end of World War II (Stanley and Niemi 2007, 55); consequently, only about 10 percent of the races in any given year are open-seat contests. In state legislatures, the mean percentage of open seats is typically more than twice as high, at least in SMDs (Figures 4, 5).⁹ The percentage of open seats also varies more over time in the states. In redistricting years, the average is usually close to, or above, 30 percent. Even in the intervening years, 15–20 percent or more of the general election contests in many states do not include an incumbent. Multimember districts are again slightly different, with the mean percentages of open seats about 4 percentage points lower than in SMDs (cf. Figures 4, 6).

Within any given year, there is considerable variability, which can be seen in two ways. First, there are occasional instances in which more than half of the seats were open (in addition to recent cases that are due to term limits). Second, some states routinely have more open seat races than others. For example, in Pennsylvania and New York, it is rare for more than 20 percent of the seats to be open in a given year, but in Colorado and Maine it is rare for that few to be open.

Elections in upper house SMDs show higher average percentages of open seats. The mean percentage in state senates never dips below 26 percent; it is typically 5–10 percentage points higher than in lower houses (cf. Figures 4–5). The range for upper houses is also wider; New York averaged only 11

Figure 4. Average Percentage of Open Seat Races, Single-Member Districts, State Lower Houses

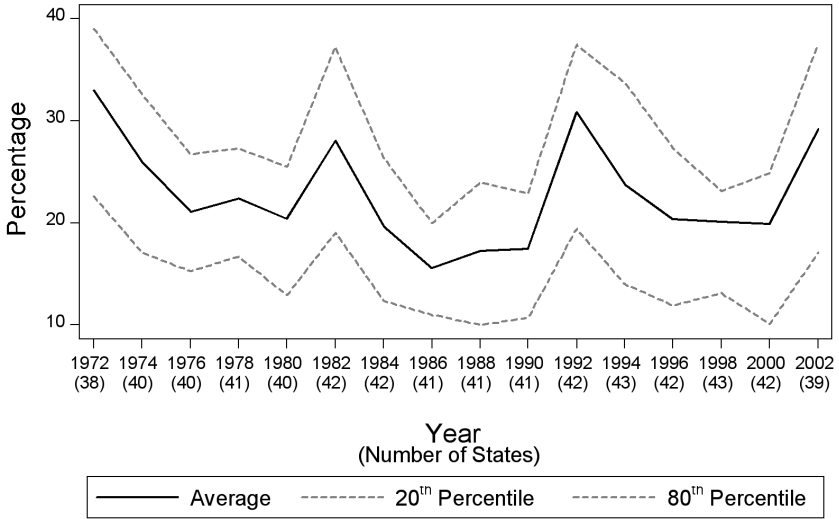


Figure 5. Average Percentage of Open Seat Races, Single-Member Districts, State Upper Houses

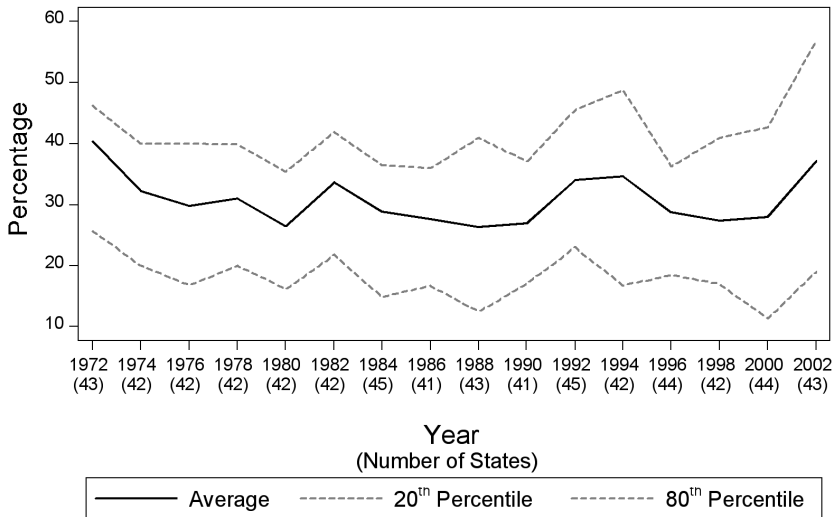
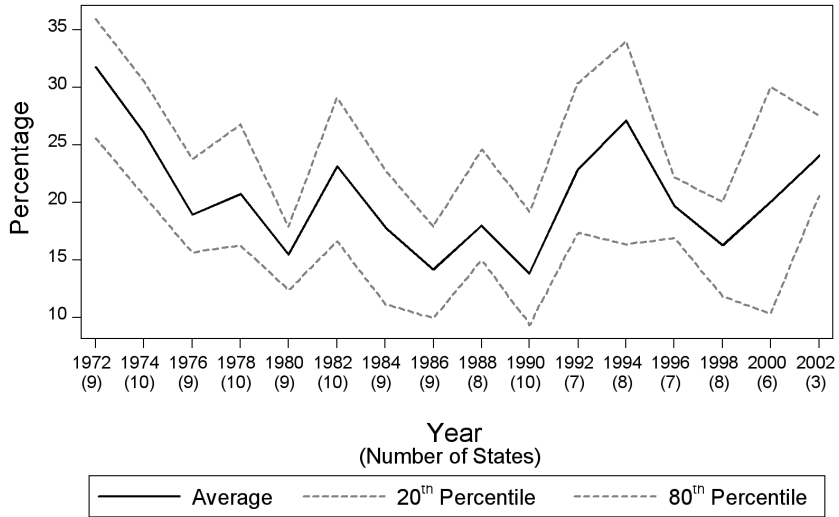


Figure 6. Average Percentage of Open Seat Races, Multimember Districts, State Lower Houses



percent open seats across the three decades, while a number of senates averaged more than 40 percent open seats.

Competition Levels

Declining marginals—i.e., a decline in the number of competitive seats—have been a major concern in legislative studies, both at the congressional and state legislative levels. Fifteen years ago, Weber, Tucker, and Brace (1991, 45) analyzed data from 20 states, concluding that “marginal seats are vanishing in some U.S. state legislatures.” The measure of competitiveness they and others used is whether the winning candidate received less than 60 percent of the vote. Because this is a fairly generous definition of competitive, researchers sometimes also look at numbers of “very” competitive races—those in which the winning candidate received less than 55 percent of the vote.

Percentages of competitive elections, based on the 60 percent cutoff value, are shown in Figures 7–9.¹⁰ The relentless pace of decline observed by Weber, Tucker, and Brace (1991) was disrupted in the election immediately after their observations ended, with competition in lower house SMDs climbing back to the levels observed in the early 1970s (Figure 7). After two relatively high years, however, a decline began anew, so that the average percentage of competitive districts is once again below 30 percent. Unlike in 1982 and 1992, there was no up-tick in competition in the redistricting year of 2002.

At the same time, the degree of competition varies greatly, as indicated by the wide percentile bounds in Figure 7. In some states, as many as 55–60 percent of the districts are competitive in a given year, while in other states, no more than 10–15 percent are competitive.

What drives this connection, of course, is the percentage of open seats. Thus, it is not surprising that competition is somewhat higher in lower house MMDs (Figure 9).¹¹ The means are more variable, but they hover in the 35–45 percent range, some 10 percentage points higher than in SMDs. Competition in SMD upper houses is similar to that in SMD lower houses, declining through 1988, rising sharply, and then declining again (Figure 8).¹²

CONCLUSION

We now have in hand a 50–state dataset covering a time period half again as long as was previously available. Furthermore, a weakness of the original dataset—inconsistencies and unreliability in the recording of names across election years—has been rectified, with the same process for name cleaning being applied to the updated data. Studies of both state legislative elections and state-level election-related phenomena (such as of divided government) should benefit.

What does the new data tell us? In part, they reaffirm the trends that

Figure 7. Average Percentage of Closely Contested Races, Single-Member Districts, State Lower Houses^a

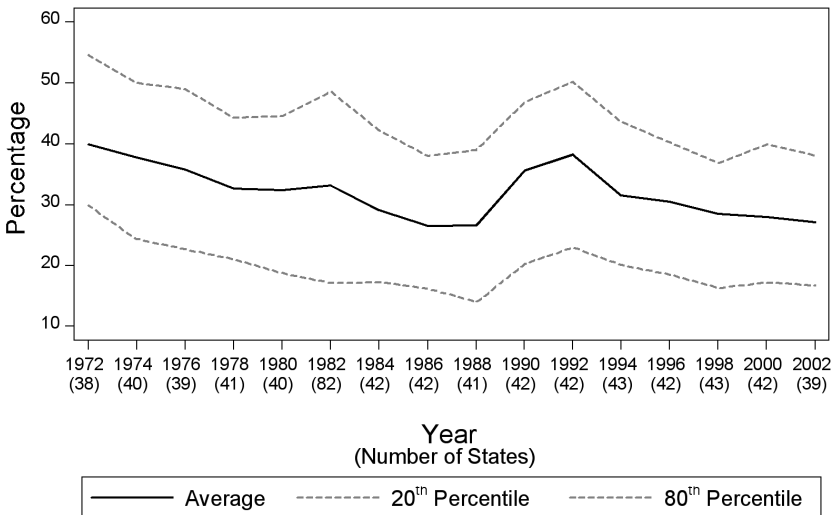


Figure 7a. Average Percentage of Very Closely Contested Races, Single-Member Districts, State Lower Houses^b

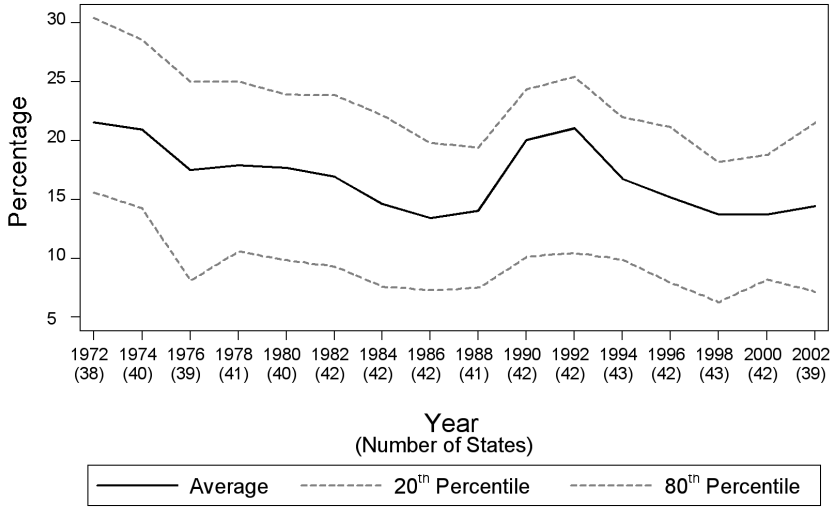


Figure 8. Average Percentage of Closely Contested Races, Single-Member Districts, State Upper Houses

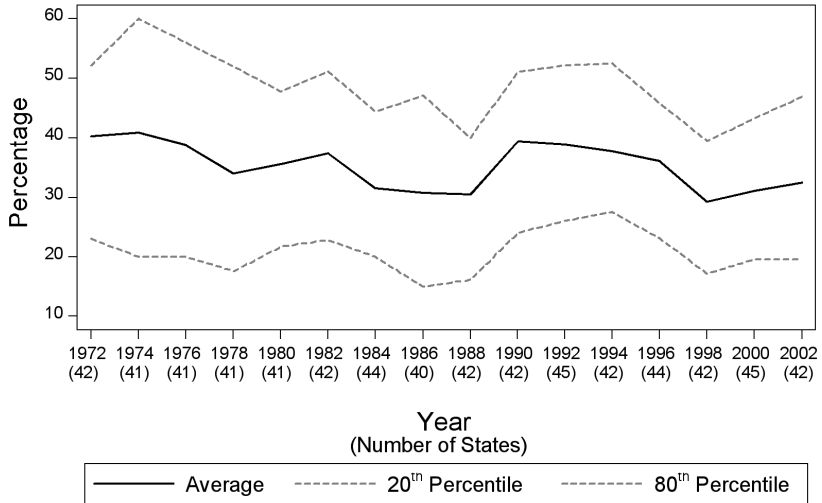
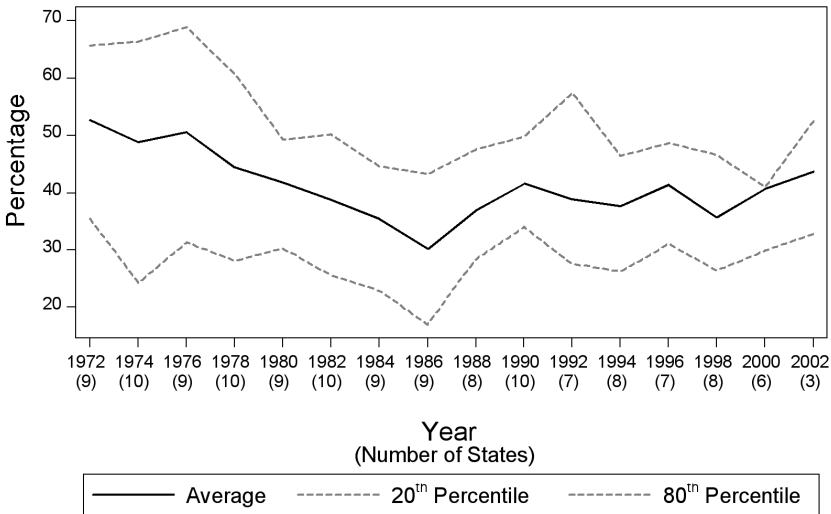


Figure 9. Average Percentage of Closely Contested Races, Multimember Districts, State Lower Houses



we observed using data from the 1970s and 1980s. Incumbents continue to be re-elected at very high rates. Open seat races are more frequent in state legislatures than in congressional elections, but rise in redistricting years and are higher in multimember than in single-member districts. Competition in the form of close election outcomes remains limited to roughly 30 to 40 percent of legislative races across states. But some things did change. Most notable was the increase in the number of closely contested races and the number of open seat races in 1990 and 1992. As a consequence, even though a downward trend was renewed, the level of competition in 2002 was at about the same level as in the late 1980s, precisely when the previous dataset stopped. Had we experienced an extension of trends in the earlier period, we would have a very different situation than currently exists. The level of competition is still below what many would regard as optimal, but its free-fall was interrupted, if only temporarily.

What explains the long-term patterns as well as the abundant cross-sectional variability evident in these data? Even though we have not yet created and tested multivariate models, it seems clear that some of this variability is explained by much-studied institutional factors: type of district (SMD or MMD), level of professionalization of the legislature, possibly chamber (lower vs. upper, or perhaps term length), and certainly the proximity of

the election year to redistricting. Of course, term limits have been added to the mix of institutional factors that shape legislative elections, creating new bumps and perhaps new long-term patterns as well. It is also likely that some of the variability we have reported stems from changes in voter behavior that are much harder to detect with these data. In any case, we look forward to analyses of the causes of what we have observed as well as to new, creative uses of the state legislative election data.

ENDNOTES

The authors gratefully acknowledge support from the National Science Foundation (Grants SBR-9422375, SES-0136526, and SES-0317924) for collecting and processing the data on state legislative elections.

1. *Legislative Studies Quarterly* published a special issue (Volume 16, Issue 1, February 1991) guest edited by Malcolm Jewell dedicated to articles making use of this data.

2. We ask that this dataset be cited as follows: State Legislative Election Returns, 1967–2003. Thomas M. Carsey, William D. Berry, Richard G. Niemi, Lynda W. Powell, and James M. Snyder, ICPSR #21480.

3. Of course, in several states more recently, a seat might be open due to term limits.

4. We combine single-member districts, multimember districts with posts, and multimember districts with staggered terms into a single category, which we refer to for short as SMDs. In districts with posts and staggered terms, constituents are represented by more than one candidate, but candidates run for a single, specific seat.

5. Carey, Niemi, and Powell (2000, 684) suggest that theory supports the significance of term length rather than chamber, although the two are highly correlated.

6. At least recently, states with elections in the year immediately following the Census (e.g., 1991) accomplished their redistricting quickly, so it makes sense to combine these elections with the more usual first-after-the-Census elections (e.g., 1992) and then to continue this approach throughout the decade. Occasionally, but more so in the 1970s, states were required to redistrict between Censuses due to court orders involving equal population or discrimination suits.

7. States with four-year staggered terms are included in every two-year interval if we have sufficient data for five or more of the races being conducted that year.

8. This may be due in part to the smaller number of seats up for election in the Senate compared to the House.

9. The calculations in this section are based on general elections only. An estimate based on national surveys of state legislators (Carey, Niemi, and Powell 2000; 2006) suggests that as many as 35 percent of incumbent defeats occur in primary elections.

10. Technically, we looked at the winning vote margin rather than at the winning candidate's vote percentage. That is, if the winning candidate won by 20 percentage points or less over the second-place candidate, we classified the district as competitive. We included independents and minor party votes. However, there are few independent or minor party candidates who win a significant percentage of the vote, so excluding them would change the results very little. Uncontested districts are included (as a 100–0 result or 100 point vote

margin); this included districts in which a state does not record vote totals in uncontested districts, as happens in recent years in Arkansas, Florida, Louisiana, and Oklahoma. In New York, where multiple parties often endorse the same candidate, the votes from all parties for a given candidate were combined before calculating competition.

11. Competition in multimember free-for-all districts was calculated using the “pairing” method developed by Niemi, Jackman, and Winsky (1991) and noted above. See Berry et al. (2000, Appendix A) for an alternative approach and Jewell and Breaux (1991) for a measure based simply on the number of candidates running in such districts.

12. Patterns for very competitive elections (defined as a margin between the winning candidate and the second place finisher of 10 percentage points or less, rather than 20 percentage points or less) are the same, although the rates are naturally lower. For example, in lower house SMDs, mean levels of competition range from about 13 percent to about 21 percent, and Figure 7a shows a very similar pattern over time to that presented in Figure 7.

REFERENCES

- Berry, William D., Michael B. Berkman, and Stuart Schneiderman. 2000. “Legislative Professionalism and Incumbent Reelection: The Development of Institutional Boundaries.” *American Political Science Review* 94:859–74.
- Berry, William D., and Thomas M. Carsey. 2005. “Revised Candidate-Level State Legislative Returns in the United States with Adjusted Candidate Names, 1968–1989.” ICPSR dataset 3938.
- Carey, John, Richard G. Niemi, and Lynda W. Powell. 2000. “Incumbency and the Probability of Reelection in State Legislative Election.” *The Journal of Politics* 62:671–700.
- Carey, John, Richard G. Niemi, Lynda W. Powell, and Gary Moncrief. 2006. “The Effects of Term Limits on State Legislatures.” *Legislative Studies Quarterly* 31:105–34.
- Jewell, Malcolm. 1991. “Editor’s Introduction: State Legislative Elections.” *Legislative Studies Quarterly* 16:1–5.
- Jewell, Malcolm E., and David Breaux. 1991. “Southern Primary and Electoral Competition and Incumbent Success.” *Legislative Studies Quarterly* 16:129–43.
- Niemi, Richard G., Simon Jackman, and Laura R. Winsky. 1991. “Candidacies and Competitiveness in Multimember Districts.” *Legislative Studies Quarterly* 16:91–109.
- Niemi, Richard G., Lynda W. Powell, William D. Berry, Thomas M. Carsey, and Thomas M. Snyder, Jr. 2006. “Competition in State Legislative Elections, 1992–2002.” In *The Marketplace of Democracy: Electoral Competition and American Politics*, eds. Michael P. McDonald and John Samples. Washington, DC: Brookings.
- Stanley, Harold W., and Richard G. Niemi. 2007. *Vital Statistics on American Politics, 2007–2008*. Washington, DC: CQ Press.
- Weber, Ronald E., Harvey J. Tucker, and Paul Brace. 1991. “Vanishing Marginals in State Legislative Elections.” *Legislative Studies Quarterly* 16:29–47.