# Proper Replication in the Study of Ticket Splitting

Barry C. Burden
Harvard University
Department of Government
Littauer Center 228
Cambridge, MA 02138-3001
burden@fas.harvard.edu

David C. Kimball
Southern Illinois University
Department of Political Science
Mailcode 4501
Carbondale, Illinois 62901-4501
dkimball@siu.edu

1/2/01

We thank Wendy Tam Cho and Brian Gaines for providing data buffer files. We expect that Cho and Gaines will house their files at the ICPSR publication-related archive. Data from our 1998 *Review* article remain available there. We thank Micah Altman, Herb Weisberg, and Sid Verba for comments.

We appreciate Cho and Gaines' (hereafter CG) extraordinary interest in our study. CG's paper reads like a long referee's report, and certainly the longest ever published in a political science journal. They attempt to find fault with all of our work, especially our data and computer code, and they reason at all points from the least representative and worst-case examples they can find. Yet after pages of adversarial cross-examination like this, *they do not challenge or propose to change a single substantive conclusion from our original article*. Had CG followed our replication instructions correctly, or if they had even done what they report doing in their article, they would have reproduced our results exactly. *In fact, in the only computer run in which they correctly replicated our model, their results are identical to ours*. The "instability" and other problems they claim to have found are a result of changing the model, statistical options, and even the data from one run to the next, but not informing readers of that fact. We were able to discover this because (as the software manual documents) King's *EzI* software stores the full specification along with the results in the same file. Rather than undermining our substantive arguments about ticket splitting, CG have actually strengthened them.

CG offer several critiques of our study. First, they criticize some simplifying assumptions that we make in modeling ticket splitting. We show that these assumptions are minor, entirely warranted, and in fact are common in the voting behavior literature. Second, CG argue that our data are "uninformative" for ecological inference. As we indicate below, the data are much more informative than they assert. Third, CG argue that King's software program is "unstable," generating vastly different estimates in repeated runs using the same data. In fact, the instability they report is due to their own (unreported) changes in model specification and their inability to consistently follow our replication instructions. The technique produces stable estimates, the reliability of which can be increased by simply raising the number of simulations. Fourth, CG argue that aggregation bias in our data make them ill-suited for King's ecological inference technique. While we agree that aggregation bias is likely present in our data, this does not rule out using King's technique. In fact, King's method is designed to address aggregation

bias and has produced accurate estimates in cases with severe aggregation bias (King 1997, chapters 11-13). Fifth, CG argue that an independent variable used in our subsequent analyses is insufficient as a test of balancing theories of ticket splitting. While we cannot test every prediction of balancing theories in a journal-length manuscript, our examination of the ideological positions of congressional candidates is more consistent with a proximity model of voting than a balancing model. Finally, CG identify some data discrepancies, which are minor and do not affect our conclusions.

More generally, CG question the validity of our ticket splitting estimates without acknowledging that our original study included several tests comparing our estimates to other data sources such as surveys and election outcomes, and that these tests consistently validated the new estimates. CG offer no verification like this, or of any other kind. In addition, they fail to demonstrate how subsequent analyses and our substantive conclusions might change if one were to measure ticket splitting differently or use different estimates of ticket splitting as dependent variables. Even if we use the most anomalous estimates of ticket splitting generated by CG, the substantive conclusions reported in our original article remain unchanged. Their critique focuses on method at the expense of substance, ignoring several important findings in the voting literature that inform the study of ticket splitting and bolster confidence in our estimates and conclusions.

If you are primarily interested in the substantive findings from our 1998 *Review* paper, read no further as you will gain no new insights other than to be more confident that our results are robust. Even if every word that CG wrote were accurate, not a single substantive conclusion in our work would change. The evidence overwhelmingly indicates that variation in the levels and direction of ticket splitting across House districts and states is driven primarily by the relative qualities of congressional candidates. Both straight- and split-ticket voters are largely responding in reasonable ways to the strengths of candidates' messages. Because candidates and campaigns are more balanced in presidential than congressional contests, most ticket splitting is locally induced by the lopsided campaigns that so frequently determine the outcomes of House elections.

Ideological preferences matter too, but their effects are more consistent with the traditional proximity model of voting than with newer balancing arguments offered to explain divided government. Congressional incumbents with moderate roll call voting records attract more ticket splitting than ideological extremists. This means that divided government ought not be interpreted as a national mandate purposely chosen by moderate voters for the parties to work together. Ticket splitting is largely a response to local candidate offerings. Even though some voters surely try to balance strategically, the real-world constraints of imbalanced congressional campaigns often restrict this possibility. Americans have surely become comfortable with divided government, in part because party loyalties have weakened, thus removing an important barrier to such behavior.

We made this case in our article using the 1988 elections as a case study and ecological data as evidence. We have since extended the data analysis, estimation techniques, and theoretical argument to a book-length treatment (Burden and Kimball Forthcoming). We do not deny that "deriving insight into why individual split their ballots" using any kind of data is difficult. This explains in part why debates about the causes of ticket splitting and divided government remain hot in the literature. Our approach is reasonable for testing balancing and other theories with aggregate data. Our book goes further by using different sources of both individual and aggregate data from a half-century of presidential and congressional elections.

**Our Simplifying Assumptions are Modest**

We now turn to the specific issues CG raise with our article. First, they are concerned about three simplifying assumptions we made. The first of these is disregarding votes for third-party candidates. To demonstrate the consequences of eliminating minor party candidates, they choose Vermont's at-large House district as a running example. This choice is a peculiar one and they admit that "Vermont is quite unusual in this regard, as significant candidates running for neither major party are rare in recent American elections" (3). Though Socialist Bernie Sanders, who finished second in Vermont that year, makes for an interesting case study, third parties

played a small role in the 1988 elections. We are quite aware of the importance of third parties in American electoral politics and have independently done research on them (Lacy and Burden 1999; Nichols, Kimball, and Beck 1999). Selecting a representative district rather than the most unusual House contest in the country would make a fairer examination of our exclusion of third-party voters. Such criticisms are not lodged at those who study congressional elections and ticket splitting in the same manner (e.g., Jacobson 1990; Brody *et al*. 1994). In 1988, 98.7% of presidential votes, 98.3% of Senate votes, and 98.5% of House votes went to the two major parties. We thus feel justified in limiting our analysis to Democratic and Republican candidates.

CG also criticize our allowance of ballot roll-off in House races without allowing for the possibility that voters may cast a vote in the House contest while abstaining from the presidential contest. We assume for simplicity that presidential non-voters also abstain from the congressional race. Though this assumption is occasionally violated, it captures much of the real world in a statistical model. While ballot roll-off is frequent, surveys indicate that only about 1% of voters "roll-on" by abstaining from the presidential contest while voting in lower level races. In 1988, House turnout exceeds presidential turnout in only about 1 in 10 districts and even then by small amounts. Figure 1 plots major-party turnout in House and presidential contests for each district in 1988. A large majority of the points fall below the 45-degree line, but none of them stray far above the diagonal. While turnout in the presidential contest often greatly exceeds House turnout, House turnout never exceeds presidential turnout by much. While others are free to expand our analysis to a $3 \times 3$ table that allows for House roll-off *and* presidential undervoting, CG offer no evidence that relaxing this assumption would substantially improve our substantive inferences about the nature and causes of ticket splitting.

*Figure 1 about here*

CG believe that we have made "some very big simplifications" by considering President/House and President/Senate voting patterns separately rather than studying them jointly along with many other ballot contests. Other forms of ticket splitting deserve closer scrutiny

(Beck *et al*. 1992; Campbell and Miller 1957; Soss and Canon 1995), and we encourage further research on the subject, but limited our focus to president-Congress ticket splitting for two theoretical reasons. First, divided national government is caused by different parties controlling Congress and the White House. Second, policy-balancing theories emphasize the separation of executive and legislative powers in the United States. Because CG urge us to pay equal attention to ticket splitting that involves votes for candidates from the obscure Small is Beautiful, Peace and Freedom, and Liberty Union Parties and require that we consider the combinations of President/House/Senate votes as a three-way split- or straight-ticket package, they are demanding that any aggregate or ecological inference analysis of voting patterns such as ours decompose every district into tables as large as $10 \times 10 \times 10$. This requires indiscriminately estimating and understanding the meanings of about 1,000 cells per district rather than the six we believe are most important. No evidence exists that paying attention to these nonentities on the national scene would help us learn anything important about American politics. Our efforts to simplify the study of president-Congress ticket splitting are entirely reasonable and follow the same procedures as virtually every prior study of ticket-splitting and every other use of the aggregate congressional and presidential electoral results.

**We Have Informative Data for Ecological Inference**

We sought the best available method for drawing inferences about ticket splitting in districts and states. As we explain in our article, neither survey data nor an aggregate measure of ticket splitting within electoral units is adequate for our purposes. For ecological inference, researchers often use Goodman's (1959) regression, but this is problematic because it assumes that ticket splitting rates are constant across districts and it allows predicted values to take on nonsensical values below 0 and above 1. King's method is attractive because it relaxes this constancy assumption and combines Goodman's regression with information known to be true using the "method of bounds." It seems to us hard to argue that adding information known to be true to the standard method in the field could possibly hurt.

An important part of King's method is the assumption that the two estimated quantities of interest – in this case, the frequency of ticket splitting by Bush and Dukakis voters – follow a truncated bivariate normal distribution. CG believe that our House data are not informative enough and suggest that King's distributional assumption does not apply because the mode of the distribution is not evident in a tomography plot. Though our real-world data do not match the artificial data they contrive for Figure 1, we show here that the real data are more informative than they acknowledge. Even having data that are not 100% informative does not render King's method unusable: it simply increases the uncertainty (i.e., standard errors) of the estimates produced (King 1997, 129). Finally, since we modeled different voting patterns in the South than in the rest of the country, we never assumed a single mode (or a single truncated bivariate normal distribution) applied to our data.

Let us begin with what is known for certain about these data. The widest possible bounds on any quantities of interest are [0,1]. That is, no less than 0% and no more than 100% of voters can split their tickets. Using deterministic information available in the data, which King shows can be visualized with a tomography plot, the bounds can often be reduced (Duncan and Davis 1953). The margins of a cross-tabulation like the one presented in Table 1 of our article (Burden and Kimball 1998, 535-6) reduces the range of possible values of the internal cells. These bounds are different for every district. Before any statistics are applied, and without *any* assumptions, we know that the aggregate bounds can be roughly cut in half in our data. Many of the district bounds are even narrower. We thus eliminate half of the ecological inference problem up front.

CG state that the aggregate bounds in our data "are too wide to imply any sort of substantive conclusion." This statement is incorrect; even if it were true, many individual district bounds are even narrower than the aggregate bounds. The standard errors of the district-level estimates of ticket splitting reflect the amount of information in the district-level bounds. We incorporated the standard errors in subsequent analyses by giving greater weight to districts with more informative bounds, another feature that CG fail to acknowledge.

As King (1997, chapter 7) explains, interpreting tomography plots involves more than just looking for a single point where the lines intersect.[1] For example, King (1997, 129) points out that informative aggregate data often have a wide range of values on the *X* variable (here Dukakis' share of the district-level vote in 1988). Variation in the *X* variable produces different sloping lines in a tomography plot, increasing the chances that the lines will intersect in a common area. The Dukakis share of the vote ranges from .23 to .96, with a standard deviation of .12. As a result, the tomography plot includes some lines that are almost vertical (where Dukakis ran well ahead of Bush), some lines that are close to horizontal (where Dukakis ran far behind Bush), and many diagonal lines (where Dukakis ran close to Bush). CG incorrectly compare a tomography plot of our data (their Figure 3) to their example of uninformative data (their Figure 2), in which there is much less variation in the slopes of the lines. Near-vertical lines give maximally informative estimates for the fraction of Bush voters who vote for the Democratic House candidate, whereas near-horizontal lines give maximally informative estimates for the fraction of Dukakis voters who cast their ballots for the Republican House candidate. CG also fail to note that lines that cut off the bottom left corner or top right corner (lines which constitute many congressional districts) give narrow bounds on *both* quantities of interest.

We also relied on some common sense in interpreting the tomography plot and diagnosing the ticket splitting estimates produced by King's method. Given the importance of partisanship and the (albeit weakening) presence of presidential coattails in congressional elections, it is safe to assume that the Democratic House candidate's share of the vote in contested races should be higher among Dukakis voters than among Bush voters by a non-trivial amount ($b^b_i > b^w_i$). Thus, if one imagines a diagonal line from the lower left corner to the upper right corner of the tomography plot, the truth for each district (and the nation) must fall on the

[1]CG state that "the primary purpose of examining a tomography plot is to assess whether the assumption of an underlying truncated bivariate normal distribution is reasonable for the data." This is one purpose, but not the main one, which is to assess the degree of information in the bounds. CG clearly miss the fact that many of the districts have highly informative bounds.

corresponding line but below that diagonal. This cuts the bounds further, and it indicates that the mode of the truncated bivariate normal distribution (the area where the lines tend to intersect) should fall toward the lower right-hand corner of the tomography plot (where $b^b_i$ is large and $b^w_i$ is small).[2] While we do not impose this assumption in our application of King's method, the tomography plot locates a mode above the lower right corner of the unit square. The lines do not literally intersect at single point but most cluster in that area (just outside the unit square).[3] The procedure estimated more ticket splitting among Bush than Dukakis voters as we expected, though it was not required to do so. In our original estimates and all of our replications, King's method placed the mode in the same area of the tomography plot below the diagonal. A closer inspection thus indicates that our data are dramatically more informative than CG admit.

**What We Have Here is a Failure to Replicate**

CG condemn King's software and assert that the program suffers from "instability," producing quite different estimates using the same data and model specifications. This is a serious charge that deserves careful consideration. It turns out to be false.

In the replication materials we archived with the ICPSR, we provided two datasets (one for the House and another for the Senate), specific directions for using the data with King's software, and a few other details about the data and estimation so that researchers could replicate all eight tables in our article.[4] King's *EzI* software allows the user to save all data, model specifications, and estimates from an ecological inference run in a "data buffer" file. The

[2]Incorporating this assumption into King's method (by adjusting the prior values of the quantities to be estimated or by culling simulated values that violate the assumption) is a possible avenue for updating our analysis. We simply used the assumption as a diagnostic device to make sure that the district-level estimates and the simulated values did not appear above the diagonal in the tomography plot. Our estimates never violated the assumption.

[3] As King (1997, 202, 238) points out, the mode need not fall within the unit square. In cases where the mode falls outside the unit square, much of the truncated bivariate normal distribution masses along the edge of the square, thereby reducing the variance in the estimated quantities of interest.

[4] The data and replication instructions were archived on May 15, 1998 as ICPSR Publication-Related Archive Study Number 1140 (http://www.icpsr.umich.edu/cgi/ab.prl?file=1140). King's ecological inference software has been available at http://gking.harvard.edu since 1995, two years before the publication of *A Solution to the Ecological Inference Problem.* CG did not archive their data but the

program automatically prompts the user to save the buffer file after estimation. Buffer files can then be imported into *EzI* later to view estimates, model specifications, program settings, and diagnostic graphs and statistics.[5]

*In reviewing the buffer files from CG's replication of our study, we discovered that they neither followed our replication directions nor used consistent model specification and program settings from one run to the next, contrary to their statement that the specification, data, and program "were exactly the same."* Changing the settings, variables, and even the data themselves from one run to the next is not a test of a program's stability, and doing so does not allow a researcher to conclude anything about the original results. The "instability" that CG say they found is really due to their own inconsistency, though they failed to inform readers of this fact.

Let us start with CG's first attempts to replicate our estimates of President-House ticket splitting in 1988 (Table A-1 and Figures 4 and 5 in their critique). Our House dataset contained 424 observations after removal of 11 districts where data were not available. Uncontested House races pose a challenge. On the one hand, these districts are clearly different than the others. For example, when a Republican runs unopposed for Congress, all Democratic presidential voters must either abstain from the House race or split their tickets. Making ecological inferences in these districts is not challenging. In addition, mean ballot roll-off in uncontested House races was 29%, while mean roll-off in contested House races was less than 6%. On the other hand, uncontested House races certainly contribute to overall levels of ticket splitting. As a result, we used the "_Eselect" option in *EzI*, which allows users to include only contested districts in the

---

*Review's* editor required CG to provide them to us; we plan to submit them along with our corrected data and replication files to the ICPSR. On the replication standard in general, see King (1995).

[5]The documentation for King's ecological inference software, whether one is using EI (the Gauss version) or *EzI* (the stand-alone version), come with public domain software and is freely available on King's web site (http://gking.harvard.edu).

likelihood maximization stage of ecological inference.[6] The idea was to base the distribution on the large majority of districts that featured House candidates of both major parties. This is not the same as removing the uncontested districts from the dataset. As the *EzI* documentation clearly explains, observations that are "selected out" of the likelihood maximization stage are included in the simulation stage and thus are included in all aggregate estimates. This allowed us to set aside the uncontested contests without dropping them from the data, so that the uncontested House races still figured into aggregate (i.e., national) estimates of ticket splitting.

Despite clear instructions in the documentation for our replication data, CG selected out the uncontested races in only 3 of their 10 first-stage runs and in only 1 of their 10 second-stage runs. This is neither what we said was required to replicate our results nor what CG reported doing. In addition, CG fail to tell readers that they altered the "_Eselect" setting in *EzI* across their 10 replication runs. *In fact, in only one of their 10 runs did CG follow our directions, and in that case their results match ours exactly.* "Run 4" in Table A-1 of their paper is the only run where case selection was done properly at both stages and is consequently the only run that replicates our published findings. CG's Run 4 estimates the proportions of Bush splitters and Dukakis splitters to be .3303 and .1985, respectively, while we estimate .331 and .198. There is no statistical or substantive difference between 33.03% and 33.10% of Bush voters splitting their tickets. We are not surprised to see that doing the same thing twice produces essentially the same results.[7] The variability demonstrated in the other nine runs is due to their changing the

[6]The documentation accompanying our replication data at ICPSR instructed researchers to "use CONTEST to E_Select out uncontested House races." The CONTEST variable, we explained, is a dummy variable indicating whether a House seat is contested by both parties or not. King (1997, 283) recommends this option when there are "unanimous" districts whose values are likely to be generated by a different process than governs the rest of the data.

[7]The small difference between the two numbers is due to simulation variability, which is exactly as expected. For our purposes, we used only enough simulations to guarantee the same answer in repeated runs to about a percentage point, since smaller differences are substantively irrelevant in studies of split-ticket voting. If one were sufficiently fastidious to want answers to be identical to two decimal points, they would merely increase the number of simulations. CG misrepresent the concept and role of modern statistical simulation, a venerable technique that is increasingly used and completely uncontroversial across a dozen scientific disciplines.

program's parameters and even the data across runs.[8] Rather than heed our replication instructions and rerun our model 10 times, as they claim to have done, CG ran 10 different specifications, changing the samples and model options each time without informing readers of what was happening behind the scenes.

As an honest test of the stability of *EzI*, and as a comparison to the estimates reported by CG, we conducted ten identical runs of the president-House ticket splitting estimates. These are genuine replications in that the data and model settings remain the same in all ten runs. Our results are presented in Table 1 and Figure 2. Table 1 replicates CG's Table A-1 while Figure 2 replicates their Figure 4 with one presentational difference. The scale in the figure has been set to the full unit interval rather than the smaller intervals of length roughly 0.14 that they arbitrarily choose to magnify differences, a violation of the most basic rules of graphic presentation (Cleveland 1985; Tufte 1983). Our results demonstrate much less variability in estimates and standard errors than reported by CG for both stages, and no variability of substantive import. As the clustering of points around (.198,.331) in Figure 2 demonstrates, none of the estimates differs substantively from the estimates we published three years ago.

*Table 1 and Figure 2 about here*

Because CG altered their execution of *EzI* each time it was run, we are likewise apt to infer that the "instability" they observe is their own doing. We do not claim to have identified the ideal way to study ticket splitting and divided government or that our models are specified perfectly, but part of a replication study's goal should be to simply repeat the original analysis to

[8] For example, the procedure used to estimate the area of the bivariate normal distribution (a setting labeled Ecdfbvn by EI) was changed across runs by CG. There are multiple ways to estimate the distribution, but the same one should always be used when researchers are merely interested in duplicating results. Changing this parameter affects both the speed and accuracy of the estimation and, as the EI manual explains, "this global should not be changed…unless you have a good reason to do so" (see http://gking.harvard.edu/ei/node14.html). In 2 of the first-stage runs Ecdfbvn was set to 5, but it was set at 2 in the remaining 8. In the second-stage runs, the changes were exactly opposite: set at 2 in 8 of the second-stage runs and 5 in the others. In the final first-stage run the nonparametric version of *EzI* (EnonPar) was used (an alternative to the model's assumption of bivariate normality that CG never acknowledge). In the covariate runs reported in Table 5 of CG (2001), they simply deleted the uncontested

verify that the outcome is reliable. CG have not done this in good faith despite the fact that our data, method, and contact information have been public since before our article was published.[9]

We are surprised that CG say nothing about the district-level estimates of ticket splitting, which are more important to our theory than the aggregate estimates upon which they focus. We spent most of our article – Tables 2, 3, 6, 7, and 8 – analyzing and modeling the district estimates. We aimed at explaining the differences in levels and directions of ticket splitting across districts (and later across time) rather than describing a single national estimate. Despite the unorthodox way in which CG attempted to replicate our estimates of ticket splitting, we are buoyed by the fact that all of their district-level estimates are close to ours.

We correlated the district-level estimates of ticket splitting generated by CG with one another and our original estimates. The *lowest* of these correlations is 0.9823, and the correlations *averaged* above 0.99. This verifies that the ordering and spacing of districts relative to one another (in terms of ticket splitting) did not differ at all substantively, even across the flawed replication runs generated by CG.

**Aggregation Bias and the Use of Covariates**

The ecological inference problem is a difficult one, which explains why nearly a century of scholarship has been devoted to it. King's approach is imperfect of course, but a better technique is not yet available. It is a real advance over the standard techniques used by social scientists from Goodman (1959) even up to the mid-1990s (Achen and Shively 1995). No method does better given aggregate data and most impose less realistic assumptions or require external data that often do not exist. When using King's method, there are decisions for analysts

---

districts from the dataset without indicating before estimating their covariate runs (see previous footnote). Further, the second covariate run has 25 rather than the default 100 simulations.

[9] Johnston and colleagues (Johnston and Hay 1984; Johnston and Pattie 2000) have used an alternative ecological inference method to estimate the frequency of ticket splitting. Their "entropy-maximizing" procedure incorporates the method of bounds (like King) but also uses survey data to further narrow the district-level bounds. Johnston has used the same method on our ICPSR data to estimate president-Congress ticket splitting in 1988, and his estimates are also highly correlated with ours (personal correspondence, 7/3/00).

to make, as his book explains. They are like the kinds of decisions researchers make when running regression models. Though such processes are prone to human error, theory and previous experience usually guide researchers' choices.

CG correctly note that aggregation bias is probably present in our data (as their Figure 6 suggests). Aggregation bias occurs when an unknown quantity of interest ($b^b_i$ or $b^w_i$) is correlated with $X_i$. This problem causes Goodman's regression to produced biased ecological inferences. However, CG incorrectly argue that aggregation bias dooms our efforts to estimate ticket splitting. They state that "it is possible to obtain *reasonable* estimates of individual-level parameters given only aggregate data if the aggregated data set contains no aggregation bias. This holds true regardless of how much information seems to appear in the data set" (emphasis added). This is false. As King (1997, chapters 10-12) demonstrates, if the bounds suitably constrain the estimates, there can be enormous aggregation bias yet inferences will be accurate. There are other ways to control aggregation bias too.

Among the choices users of King's method must make is selection of covariates. This is a specification issue much like decisions about which variables to include and exclude on the right-hand side of a regression equation. Covariates help manage aggregation bias, a potential problem for all ecological inference techniques. Though King's method has been shown to be more robust to this than previous methods, aggregation bias still needs to be taken seriously. CG criticize our decision to use a South dummy variable as a covariate and even argue that the choice of covariates is such a quandary as to render King's method essentially useless. Covariate selection is always ambiguous and this is problematic because different choices lead to different results. The same can be said of specification choices for users of regression and other models if one assumes their premise that substantive knowledge is useless. In our case we have both substantive knowledge and interest in issues of ticket splitting, divided government, and American electoral politics generally and made a careful covariate choice after reading and

theorizing about what discrete variables might contribute to step changes in levels of ticket splitting. The idea that southern politics is distinctive is hardly controversial.

When using standard linear regression models, one depends on several assumptions. For example, the errors are homoscedastic and normally distributed and observations are not serially correlated. Such formal assumptions are needed to make the model easier to derive but may be violated to some degree without jeopardizing the results seriously. One rarely believes that the assumption of perfect homoscedasticity holds; we are usually satisfied with data that are not "too" heteroscedastic. It is not the presence or absence as much of the degree of a violation. An attraction of OLS is that it is robust to modest violations of assumptions. A prerequisite for using King's estimator should not be that aggregation bias can be shown to be totally absent, but that aggregation bias is not "too bad," for King's procedure is robust in the face of modest violations of this assumption (King 1997). Moreover, narrow bounds, a heavily truncation distribution of the quantities of interest, and covariates further minimize aggregation bias problems.

*CG assume that researchers have no substantive knowledge of the problem they are studying and instead use a "kitchen sink" approach to model specification.* They introduce each of the variables in our dataset (intended for other purposes) as covariates. Because the aggregate estimates change using different covariates, they conclude that proper covariate selection is impossible. This is technically true just as it is true that one will never know if a regression model is correctly specified. Estimates are only accurate to the degree that model specification is right. One response is to throw up one's hands and decide never to run regression models. Alternatively, one might be guided by theory, the existing literature, and first-hand experience to choose regressors.[10] We adopted the latter approach and chose a covariate theorized, based on work by people like Frymer (1994; Frymer, Kim, and Bimes 1997) and our own experiences with the data and substantive knowledge, that a covariate indicating southern states seemed a

reasonable way to address potential aggregation bias problems. We thus find CG's Table 5 reassuring rather than alarming. They believe that "erratic performance" shows that covariate choice is an unresolvable quandary, as we suppose they would believe about variable selection for a regression model. We think that substantive knowledge and theory are helpful and that different specifications should lead to different results, whether one is using OLS, ecological inference techniques, or any other estimator.[11]

Most of the alternative covariates CG use are not substantively justifiable. Substantive knowledge is useful in a lopsided kind of way: it might not identify the perfect covariate but it can certainly rule out a large number of inappropriate covariates. Using NOMINATE scores and the experienced challenger measure as covariates are especially troubling to us. The effects of both variables on ticket splitting should vary depending on the party of the incumbent and the type of ticket splitting being estimated (DR splitting versus RD splitting). We argue and find that the effects of ideology on ticket splitting vary by the party of the incumbent and the direction of ticket splitting (DR splitting versus RD splitting). A conservative incumbent increases RD splitting if the incumbent is a Democrat but reduces DR splitting if the incumbent is a Republican. This finding is well-supported in the literature (Frymer 1994; Frymer, Kim, and Bimes 1997; Grofman *et al*. 2000; Petrocik and Doherty 1996).

The challenger variable is expected to operate in a similar fashion. An experienced challenger may limit RD splitting in districts with a Democratic incumbent (and hence a Republican challenger) but increase RD splitting in districts where the incumbent is a Republican and the challenger is a Democrat. Using either variable as a covariate to estimate ticket splitting in all districts obscures this important distinction. In addition, while we find both measures useful as independent variables in models where the observations are limited to Democratic *or*

[10] King (1997, 280) explains that covariates are often helpful, "But variables that are candidates for the second stage should only be included as covariates in the original model if they help avoid aggregation bias and do not create other problems."

Republican incumbents, neither variable is meant to apply to open seat contests. The challenger variable is coded as 0 for all open seat races, and the NOMINATE variable contains the retiring incumbent's ideological position for open seat contests. Thus, including either variable as a covariate to estimate ticket splitting across *all* contests is certain to cause more confusion. CG do more than assume a lack of substantive knowledge about covariate choice (a kind of diffuse prior). They actually make choices that are contrary to a vast substantive literature.

The only alternative covariates used by CG that can be justified on substantive grounds are ballot format and incumbency. In both cases, the ticket splitting estimates CG report using these alternative covariates are quite close to and are not even statistically different from the estimates we report using South as a covariate. In addition, CG fail to determine whether subsequent analyses would differ when using ticket splitting estimates based on a different covariate than the one we used.

**Measurement and Theory Testing**

The dominant theory in the study of ticket splitting in the United States before our article appeared falls under the rubric of "policy balancing." Balancing has been the subject of heated debate in the literatures on split-ticket voting and divided government. The idea is that some moderate voters choose to balance two extreme parties by choosing a liberal Democrat for one office and a conservative Republican for the other. Thus, voters think about the joint outcome of congressional and presidential races and seek to offset the outcome of one contest with the results of the other. Fiorina (1996) is most associated with balancing theories, but his description of exactly what candidate or party distances should be used in operationalizing the theory is unstated. His clearest prediction is that ticket splitting should increase as parties polarize, as moderate voters become less satisfied with either extreme. In contrast, we believe that candidates not voters play the largest role in producing divided government. Most voters (either by choice

[11] Note that Figure 7 misleadingly exaggerates the amount of variability due to covariates just as Figure 4 did for the 10 "replication" runs.

or constraints imposed by the supply of candidates) choose candidates individually based on ideological proximity to the degree that such positions matter at all.

Since the publication of our article in 1998, many other studies have tested predictions of balancing theories of divided government, some supportive (Smith *et al*. 1999; Lacy and Paolino 1998), others not (Born 2000; Grofman *et al*. 2000; Mattei and Howes 2000). It is not possible to test all of the theory's predictions in one journal-length article. Thus, we chose to contrast policy balancing against a simple proximity model of voting. As CG note, voters following a proximity rule might cast a split ballot if it happens that they are closer to the Democratic candidate in one contest but closer to the Republican candidate in another contest. We agree and made the same point in our article (542). In contrast, balancing theories posit a different decision process in which voters select a presidential candidate from one side of the ideological spectrum and a congressional candidate from the other side to moderate government policy.

To test this specific prediction in a way tractable for a journal article, we measured the ideological distance between candidates of different parties in two ways. While neither is ideal because ideological positions are inherently unobservable and researchers disagree about how best to measure them, our measures do an adequate job of assessing the hypothesis from different directions. In the book that builds upon our article, we expand the range of measures used and always find evidence consistent with our theory.

In the House models, we use the incumbent's first dimension NOMINATE score as an indicator of their ideology. Because the comparison made in balancing models is between the congressional and presidential candidates, we assume that the presidential candidate is fixed, so the scores could be subtracted from any arbitrary constant as long as the Democratic candidate for House, for example, is always to the left of Republican presidential candidate Bush (see fn. 15 in our article). This assumption seemed reasonable to us given that the theoretical literature we are in part testing makes it (Fiorina 1996) and that empirical literature supports it (Ansolabehere, Snyder, and Stewart 2001; Erikson 1990). Because the comparison with the presidential

candidate changes depending on the party of the incumbent, we analyze Democratic and Republican incumbents and RD and DR voting separately.

If voters want to balance, they should be inclined to offset a vote for Bush with a vote for a liberal Democratic candidate for Congress. Our Table 8 explains levels of RD splitting using the Democratic incumbent's NOMINATE score as a measure of their distance from Bush. The positive and significant coefficient indicates that RD ticket splitting is more common in districts where the Democratic incumbent is conservative than in districts where the Democrat is liberal, precisely the opposite of balancing theory but consistent with race-specific proximity voting.

For President/Senate ticket splitting, we drew estimates of both presidential and senatorial candidates in each state from the 1988 wave of the Senate Election Study (SES). The SES asked representative samples of voters in each state to place the four candidates on the traditional 7-point ideological scale. Following Erikson (1990), we used the mean placements for each candidate in each state. The distance between Dukakis and the Republican Senate candidate was used in the DR models and the distance between Bush and the Democratic Senate candidates was using the RD models. These data offer some advantages over incumbent-only measures and so we combine our inferences from both kinds of measures to help build grander conclusions.

CG managed to find *two* SES respondents where the assumptions required by these tests do not hold. A respondent in Connecticut and another in New Mexico happened to place the Republican Senate candidate to the left of the Democratic Senate candidate (i.e., $D_s > R_s$).[12] They are only a concern in that they violate one of the (implied or common) assumptions of balancing theory. Fortunately, voters who place Republicans left of Democrats are rare. In the 1988 SES, only 7% of all Senate election voters had such an ordering. Thus, the assumption fits 93% of the time at the individual-level and almost 100% of the time of the time at the aggregate level, where balancing's "as if" assumptions are most likely to hold (see fn. 10 in our article). This kind of

decision-making-based-on-outliers is akin to their choice of Vermont as evidence that a focus on major-party candidates is misguided. We do not find analyses based only on outlying cases to be particularly useful.

A major argument in our paper was that ticket splitting is better described as an unintentional rather than intentional effort to create divided government via balancing (where intentionality is carefully defined). When ideology matters in ticket splitting and congressional elections, it tends to be consistent with the standard spatial model rather than balancing. CG use a broader definition of intentionality to include any case where policy considerations influence the voting decision, even if the balancing decision process does not come into play. In CG's terms, we favor "candidate-proximity voting" over "expected-policy voting."

**Data Discrepancies**

Finally, CG scrutinized every datum in our article to identify possible mistakes in our datasets. The discrepancies they found have two origins. One is that our source, the *Almanac of American Politics*, is sometimes incorrect or at least gives different numbers than other sources. We should note that many of these discrepancies are on the order of dozens of votes in districts with about half a million eligible voters, so they make little difference. Sometimes the truth is not clear, as when the *Almanac*, *Politics in America*, and other sources report slightly different vote totals. The second source of these discrepancies is human error in entering the data. We have adjusted the dataset as CG recommend in their Appendix. Table 2 compares our original estimates of 1988 president-House ticket splitting with the results from three replications using updated data. None is substantively or significantly different from another. It is clear that the

[12] As an aside, we note that both voters, however unusual, still voted on the basis of straight race-by-race proximity, always choosing the candidates nearest them, which is consistent with our argument about the role of ideology in ticket splitting.

data discrepancies in our original study hardly affected the aggregate estimates of ticket splitting. Adjusting the dataset does not alter the substantive conclusions of our study.[13]

## Conclusion

We never stated or implied that "ticket splitting is *solely* the result of lopsided campaigns" as CG contend (1, emphasis added). Like any social phenomenon, there are multiple causes and we seek to identify the most important among them and their relationships. Reacting to existing work in this area, we concluded that "unintentional forces…exert a tremendous influence" (543) and emphasize the importance of the strengths of candidates' messages, "defined by the cues of incumbency and campaign spending, the two largest influences on split-ticket voting" (542) while allowing room for other factors, both systematic and idiosyncratic to play a role. After carefully considering CG's extensive critique, we are more reassured that our initial argument and results stand firm.

[13] The same is true of the coefficients estimated in the second-stage regressions, which are the real quantities of interest.

## References


Ansolabehere, Stephen, James M. Snyder, Jr., and Charles Stewart III. 2001. "Candidate Positioning in U.S. House Elections." *American Journal of Political Science* 45(January):160-78.

Beck, Paul Allen, Lawrence Baum, Aage R. Clausen, and Charles E. Smith. 1992. "Patterns and Sources of Ticket Splitting in Subpresidential Voting." *American Political Science Review* 86(December):916-28.

Born, Richard. 2000. "Policy Balancing Models and the Split-Ticket Voter, 1972-1996." *American Politics Quarterly* 28(April):131-62.

Brody, Richard A., David W. Brady, and Valerie Heitshusen. "Accounting for Divided Government: Generational Effects on Party and Split-Ticket Voting." In *Elections at Home and Abroad: Essays in Honor of Warren E. Miller*, M. Kent Jennings and Thomas E. Mann eds. Ann Arbor: University of Michigan Press.

Burden, Barry C., and David C. Kimball. Forthcoming. *Why Americans Split Their Tickets: Campaigns, Competition, and Divided Government.* Ann Arbor, MI: University of Michigan Press.

Burden, Barry C., and David C. Kimball. 1998. "A New Approach to the Study of Ticket Splitting." *American Political Science Review* 92(September):533-44. (To be reprinted in Richard G. Niemi and Herbert F. Weisberg, ed. *Controversies in Voting Behavior*, 4th ed. Washington, DC: Congressional Quarterly Press.)

Campbell, Angus, and Warren E. Miller. 1957. "The Motivational Basis of Straight and Split Ticket Voting." *American Political Science Review* 51(June):293-312.

Cho, Wendy K. Tam, and Brian J. Gaines. 2001. "Reassessing the Study of Split-Ticket Voting." *American Political Science Review* Forthcoming.

Cleveland, William S. 1985. *The Elements of Graphing Data*. Monterey, CA: Wadsworth.

Duncan, Otis Dudley, and Beverly Davis. 1953. "An Alternative to Ecological Correlation." *American Sociological Review* 18(December):665-6.

Fiorina, Morris P. 1996. *Divided Government.* 2nd ed. Needham, MA: Allyn and Bacon.

Frymer, Paul. 1994. "Ideological Consensus within Divided Party Government." *Political Science Quarterly* 109(Summer):287-311.

Frymer, Paul, Thomas Paul Kim, and Terri S. Bimes. 1997. "Party Elites, Ideological Voters, and Divided Party Government." *Legislative Studies Quarterly* 22(May):195-216.

Goodman, Leo. 1959. "Some Alternatives to Ecological Correlation." *American Journal of Sociology* 64(May):610-24.

Grofman, Bernard, William Koetzle, Michael McDonald, and Thomas Brunell. 2000. "A New Look at Split Ticket Outcomes for House and President: The Comparative Midpoints Model." *Journal of Politics* 62(February):34-50.

Jacobson, Gary C. 1990. *The Electoral Origins of Divided Government.* Boulder, CO: Westview Press.

Johnston, Ronald J., and Charles Pattie. 2000. "Ecological Inference and Entropy-Maximizing: An Alternative Estimation Procedure for Split-Ticket Voting." *Political Analysis* 8(Autumn):333-45.

Johnston, Ronald J., and A. M. Hay. 1984. "The Geography of Ticket-Splitting: A Preliminary Study of the 1976 Elections Using Entropy-Maximizing Methods." *The Professional Geographer* 36:291-6.

King, Gary. 1999. "Replication, Replication." *PS: Political Science and Politics*, with comments from 19 authors and a response, 28(September):443-99.

King, Gary. 1997. *A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data*. Princeton, NJ: Princeton University Press.

Lacy, Dean, and Barry C. Burden. 1999. "The Vote-Stealing and Turnout Effects of Ross Perot in the 1992 U.S. Presidential Election." *American Journal of Political Science* 43(January):233-55.

Lacy, Dean, and Philip Paolino. 1998. "Downsian Voting and the Separation of Powers." *American Journal of Political Science* 42(October):1180-99.

Mattei, Franco, and John S. Howes. 2000. "Competing Explanations of Split-Ticket Voting in American National Elections." *American Politics Quarterly* 28(July):379-407.

Nichols, Stephen, David C. Kimball, and Paul Allen Beck. 1999. "Voter Turnout in the 1996 Election: Resuming the Downward Spiral?" In *Re-Election 1996: How Americans Voted*, Herbert F. Weisberg and Janet M. Box-Steffensmeier, eds. Chatham, NJ: Chatham House.

Petrocik, John R., and Joseph Doherty. 1996. "The Road to Divided Government: Paved without Intention." In *Divided Government: Change, Uncertainty, and the Constitutional Order*, ed. Peter F. Galderisi. Lanham, MD: Rowman and Littlefield Publishers.

Smith, Charles E., Robert D. Brown, John M. Bruce, and L. Marvin Overby. 1999. "Policy Balancing and Voting for Congress in the 1996 National Election." *American Journal of Political Science* 43(July):737-64.

Tufte, Edward R. 1983. *The Visual Display of Quantitative Information*. Cheshire, CT: Graphics Press.

**Table 1: Proper Replication of Burden and Kimball (1988)**

| | First Stage Model | | Second Stage Model | |
|---|---|---|---|---|
| | $b^b$ | $b^w$ | Bush Splitters | Dukakis Splitters |
| | *Not reported* | | 0.331 | 0.198 |
| | | | (0.006) | (0.007) |
| Run 1 | 0.905 | 0.923 | 0.344 | 0.224 |
| | (0.036) | (0.031) | (0.042) | (0.051) |
| Run 2 | 0.911 | 0.918 | 0.325 | 0.205 |
| | (0.043) | (0.036) | (0.007) | (0.009) |
| Run 3 | 0.908 | 0.920 | 0.326 | 0.206 |
| | (0.039) | (0.033) | (0.007) | (0.008) |
| Run 4 | 0.912 | 0.917 | 0.334 | 0.216 |
| | (0.038) | (0.032) | (0.030) | (0.035) |
| Run 5 | 0.910 | 0.919 | 0.323 | 0.203 |
| | (0.038) | (0.032) | (0.007) | (0.008) |
| Run 6 | 0.907 | 0.922 | 0.326 | 0.206 |
| | (0.038) | (0.032) | (0.006) | (0.007) |
| Run7 | 0.904 | 0.924 | 0.329 | 0.206 |
| | (0.039) | (0.033) | (0.007) | (0.008) |
| Run 8 | 0.911 | 0.918 | 0.324 | 0.207 |
| | (0.038) | (0.032) | (0.007) | (0.009) |
| Run 9 | 0.909 | 0.920 | 0.326 | 0.207 |
| | (0.037) | (0.031) | (0.014) | (0.017) |
| Run 10 | 0.914 | 0.915 | 0.323 | 0.206 |
| | (0.037) | (0.031) | (0.007) | (0.008) |

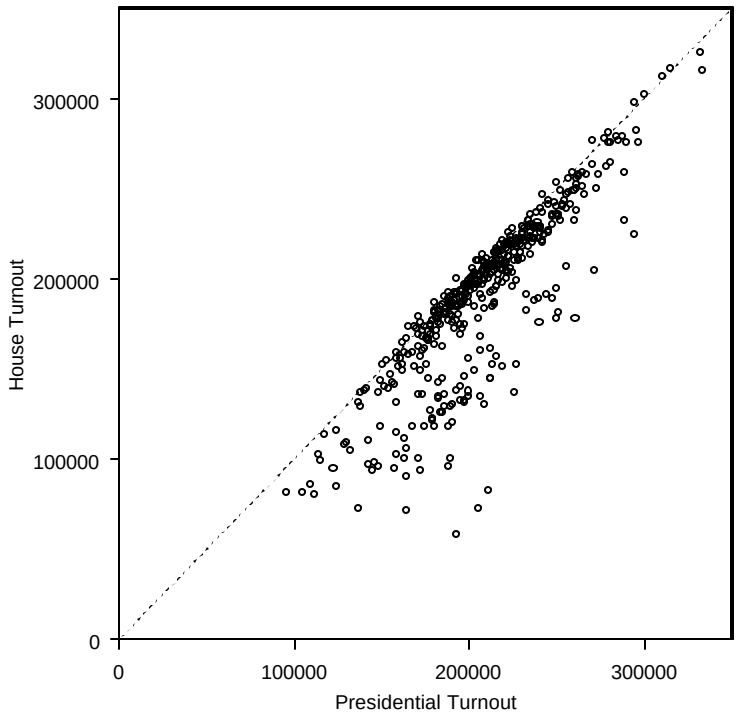
*Notes:* Standard errors are in parentheses. Note the small variation across runs in estimates of both the coefficients and associated standard errors. In the first stage (House ballot completion), $\beta^b$ ranges from 0.904 to 0.914 (with errors ranging from 0.036 to 0.043) and $\beta^w$ ranges from 0.915 to 0.924 (with errors ranging from 0.031 to 0.036). In stage 2 (ticket splitting), Bush splitters range from 0.323 to 0.344 while Dukakis splitters range from 0.203 to 0.224. *EzI* version 2.3 was used.

**Table 2: Adjusting the Dataset Does Not Change Estimates**

| | Bush Splitters | Dukakis Splitters |
|---|---|---|
| | .331 | .198 |
| | (.006) | (.007) |
| Updated Data (Run 1) | .338 | .196 |
| | (.006) | (.009) |
| Updated Data (Run 2) | .336 | .193 |
| | (.006) | (.008) |
| Updated Data (Run 3) | .336 | .193 |
| | (.006) | (.008) |

*Note:* Entries are aggregate estimates and standard errors from three replication runs using the corrected dataset. *EzI* version 1.55 was used.

**Figure 1: House Turnout Rarely Exceeds Presidential Turnout**



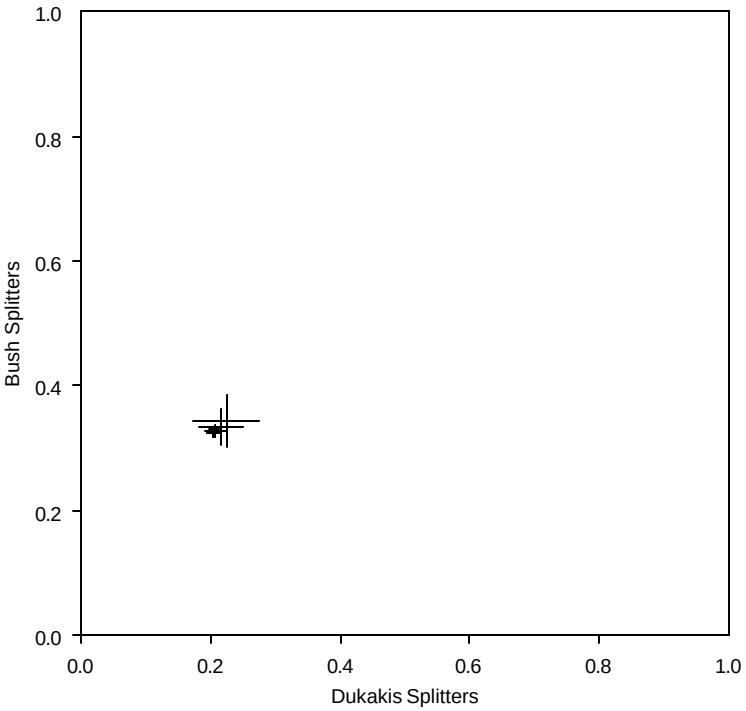


**Figure 2: Results from 10 Replication Runs Do Not Differ Meaningfully**