American Political Science Review. Forthcoming.

REASSESSING THE STUDY OF SPLIT-TICKET VOTING

Wendy K. Tam Cho*

Brian J. Gaines

Abstract

Burden and Kimball (1998) report that, by using the King estimation procedure for inferring individual-level behavior from aggregate data, they are the first to produce accurate estimates of split-ticket voting rates in congressional districts. There are, however, several reasons to doubt that their analysis produced accurate cross-level inferences. We show that the estimation technique is highly suspect in general and especially unhelpful with these particular data. Hence, their conclusions about split-ticket voting in 1988 should be regarded as dubious. We delineate the key steps and main barriers in performing an unimpeachable analysis of ticket-splitting using aggregate data.

*Wendy K. Tam Cho is Assistant Professor of Political Science and Statistics at the University of Illinois at Urbana-Champaign. Brian J. Gaines is Assistant Professor of Political Science at the University of Illinois at Urbana-Champaign. We would like to thank Bear Braumoeller, Bruce Cain, Michael Caldwell, Lawrence Cho, Christophe Crombez, Susan Jellissen, Masaru Kohno, Peter Nardulli, Brian Sala, and Jasjeet Sekhon for helpful comments. Cho also thanks the National Science Foundation (Grant No. SBR–9806448) for research support.

In a past issue of the *Review*, Burden and Kimball report estimates of ticket-splitting at the district level. They employ an ecological inference method proposed by King (1997), and their analysis of the estimates obtained from this method leads them to conclude that, contrary to some previous findings in the split-ticket voting literature (e.g. Alesina and Rosenthal 1995, Fiorina 1996), "voters are not intentionally splitting their tickets to produce divided government and moderate politics" (Burden and Kimball 1998, 533). Instead, they claim, ticket-splitting is solely the result of lopsided congressional campaigns in which well-funded, high-quality incumbents tend to run against unknown, underfunded challengers.

After working through their analysis of the ticket-splitting problem, we find that there are six reasons to doubt the generality and veracity of their conclusions. First, for tractability, they greatly simplify the ticket-splitting phenomenon by limiting their focus to only three offices (U.S. Representatives, Senators, and President), by discarding all votes not cast for major-party candidates and all ballots lacking presidential-race votes, and by ignoring House-Senate splitting. Second, even if these simplifications could all be justified, King's own checklist for determining which data are amenable to his estimation procedure suggests that these election data are unlikely to be informative. Third, were the assumptions justified and the data promising, problems with the King estimation procedure itself (hereafter, for convenience, referred to as "EI") would remain. In particular, the EI program is sufficiently unstable to produce significantly different results on separate runs using the same data. Fourth, even if the EI program operated flawlessly, the statistical model underlying EI is inappropriate for analysis of the Burden and Kimball data set because its key assumptions are violated by these data. Fifth, the principal substantive claim in their article follows from an OLS model in which the EI estimates serve as the dependent variable. Even if these estimates were correct, however, the key *independent* variable in that OLS model is constructed in such a way that their claims about voter intentions are never substantiated. Lastly, we note in passing that their data set is flawed.[1] In short, much work remains to be done before the phenomenon of split-ticket voting becomes well-understood.

In this article, we revisit Burden and Kimball's analysis, highlighting the critical decisions at each stage, and focusing on how an ideal treatment of the split-ticket voting problem would differ. Our primary goal is to demonstrate that deriving insight into why individuals split their ballots by examining only aggregate data is far more difficult than their article implies. Because our main point is to focus on the extreme difficulty in deriving estimates of individual behavior using only

[1] To maximize consistency, the estimates reported in this article have been obtained from the original Burden-Kimball data set. A catalog of 26 errors in the Burden-Kimball data on US House data is posted at the *APSR* web site.

aggregate data, we do not provide a revised, competing analysis of district-level split-ticket voting estimates. Instead, we precisely identify the major barriers to such an analysis and explicate what remains to be done before accurate estimates can be produced. Accordingly, this paper endeavors to provide a blueprint for future research.

CONCEPTUALIZING THE PROBLEM

Where ballots are secret, objective voting data are usually available only in aggregated form. Thus, the problem of estimating the amount and types of split-ticket voting in an American election can be summarized by a set of $a_1 \times a_2 \ldots \times a_C$ tables, one for each unit of aggregation (e.g. precinct, congressional district, state, etc.), where a_1 represents the number of ways to vote (including abstention) in the contest for office 1, a_2 the number in a second contest, and so on, through all C contests on the ballot. For each table, we know most of the row and column totals (marginals), but we must make assumptions to complete the table and to estimate the cell entries. Given C contests, and a_c voting options for contest c, there are $\prod_{c=1}^{C} a_c$ possible ways to cast a ballot. Statistical estimation quickly becomes difficult as this number grows, or impossible if it exceeds the number of observations / districts.[2]

Burden and Kimball simplify the full ticket-splitting problem. First, they analyze Presidential and House votes and then, separately, Presidential and Senate votes, ignoring House-Senate splits and all other races on the ballot. Thus, they consider not the C-dimensional tables corresponding to complete ballots, but only two sets of 2-way tables (formed from House-President and Senate-President voting). They further simplify the analysis by discarding all votes not cast for one of the two major parties and by assuming (falsely, as they recognize) that there are *no* ballots featuring choices in congressional contests but not choices in the presidential contest. They thereby reduce the size of the table describing each House district from $a_{p_i} \times a_{h_i}$ (where $p_i - 1$ and $h_i - 1$ are the numbers of presidential and House candidates, respectively, who received votes in district i) to 2×3. To ensure comparability across observations, they discard all districts in which no candidate from one major party ran. Table 1 illustrates a House-President observation using data for the state of Vermont.

Table 1 shows, for Vermont, the full House-President case (not the full ballot-splitting problem), while Table 2 shows the simplified version analyzed in two steps by Burden and Kimball.

[2] An additional complication in the American context is that C varies across states, so that the district tables are not all of the same dimension (i.e. same number of offices at stake) let alone the same size (i.e. same number of candidates for each office).

Table 1. The Complete Vermont Presidential-House Vote-Splitting Problem

		House of Representatives Vote								
		A	*B*	*C*	*D*	*E*	*F*	*G*	*none*	*TOTAL*
Presidential Electors Vote	0	v_{0A}	v_{0B}	v_{0C}	v_{0D}	v_{0E}	v_{0F}	v_{0G}	v_{0n}	124331
	1	v_{1A}	v_{1B}	v_{1C}	v_{1D}	v_{1E}	v_{1F}	v_{1G}	v_{1n}	115775
	2	v_{2A}	v_{2B}	v_{2C}	v_{2D}	v_{2E}	v_{2F}	v_{2G}	v_{2n}	1000
	3	v_{3A}	v_{3B}	v_{3C}	v_{3D}	v_{3E}	v_{3F}	v_{3G}	v_{3n}	275
	4	v_{4A}	v_{4B}	v_{4C}	v_{4D}	v_{4E}	v_{4F}	v_{4G}	v_{4n}	205
	5	v_{5A}	v_{5B}	v_{5C}	v_{5D}	v_{5E}	v_{5F}	v_{5G}	v_{5n}	189
	6	v_{6A}	v_{6B}	v_{6C}	v_{6D}	v_{6E}	v_{6F}	v_{6G}	v_{6n}	164
	7	v_{7A}	v_{7B}	v_{7C}	v_{7D}	v_{7E}	v_{7F}	v_{7G}	v_{7n}	142
	8	v_{8A}	v_{8B}	v_{8C}	v_{8D}	v_{8E}	v_{8F}	v_{8G}	v_{8n}	113
	9	v_{9A}	v_{9B}	v_{9C}	v_{9D}	v_{9E}	v_{9F}	v_{9G}	v_{9n}	1134
	None	v_{NA}	v_{NB}	v_{NC}	v_{ND}	v_{NE}	v_{NF}	v_{NG}	v_{Nn}	?
	TOTAL	98937	90026	45330	3110	1455	1070	203	?	≥ 243328

Note: Presidential Candidates: 0. Republican (Bush); 1. Democrat (Dukakis); 2. Libertarian (Paul); 3. National Economic Recovery / Independent (Larouche); 4. New Alliance (Fulani) ; 5. Populist (Duke); 6. Peace and Freedom (Lewin); 7. Socialist / Liberty Union (Kenoyer); 8. Socialist Workers (Warren); 9. Scattering; A. Republican (Smith); B. Independent / Socialist (Sanders); C. Democrat (Poirier); D. Libertarian (Hedbor); E. Liberty Union (Diamondstone); F. Small is Beautiful (Earle); G. Scattering

Table 2. A 2×3 Simplification of the Vermont Presidential-House Vote-Splitting Problem

	House of Representatives Vote			
Presidential Vote Choice	Republican	Democrat	none	Total
Bush (R)	v_{RR}	v_{RD}	v_{Rn}	124331
Dukakis (D)	v_{DR}	v_{DD}	v_{Dn}	115775
Number of Voters	98937	45330	95839	240106

Clearly, in the case of Vermont eliminating non-major-party House candidates distorts the result, since the Independent candidate garnered a full 37% of the vote while the Democratic candidate finished a distant third. Vermont is quite unusual in this regard, as significant candidates running for neither major party are rare in recent American elections.[3] But another simplifying assumption, that v_{N+} must be zero, is more generally problematic. The two-party House vote actually exceeded the two-party presidential vote in 44 districts in 1988. In any event, once the table is reduced to a 2×3, two iterations of King's EI program (which is designed to handle 2×2 tables only) can produce estimates of the parameters of interest. Burden and Kimball's Table A–1 is a general form of our Table 2, except that it transforms the vote frequencies into proportions.[4]

In short, given the complexity of the American ballot, a full analysis of ticket-splitting is a huge job. For the sake of tractability, Burden and Kimball make some small simplifications (disregarding minor party candidates), some big simplifications (disregarding abstention from the presidential contest), and some very big simplifications (examining only two contests at a time). While these choices make the problem manageable, and may be justified on this basis, they do limit the applicability and validity of the eventual substantive conclusions about who splits tickets and why. Hence, at minimum, the authors should have been more cautious about the generality of their conclusions.

Bearing these limitations in mind, do the data on congressional and presidential voting in 1988

[3] In fact, Burden and Kimball inadvertantly juxtaposed the Independent and the Democrat in the Vermont case, so Table 2 does not describe an observation in their analysis. There are 435 House districts, but 7 Louisiana districts had no contests in November 1988, 9 other races were uncontested and produced no vote count, and 64 more were missing a candidate from one of the major parties. Of the 355 districts open to analysis, 154 saw *some* votes won by non-major-party candidates. But in only 12 cases did that vote exceed 4%. The next highest figures after Vermont were 16.8% (NY 19th), 10.0% (NJ 10th), and 8.7% (CA 1st). "Extra" candidates were more frequent in earlier American elections.

[4] Hereafter, we retain Burden and Kimball's notation. They use β_i^b to indicate the proportion of Dukakis voters who cast a Democratic House vote, and β_i^w to indicate the proportion of Bush voters who split their ballot by casting a Democratic House vote. The superscripts b and w are mnemonics for "black" and "white," left over from the main running example in King's book, race and voting.

Table 3. A Reduced Split-Ticket Problem for District i

	House Vote		
Presidential Vote	Democrat	Republican	Vote
Dukakis (D)	β_i^b	$1-\beta_i^b$	λ_i^b
Bush (R)	β_i^w	$1-\beta_i^w$	λ_i^w
Fraction of Voters	T_i	$1-T_i$	V_i

nonetheless reveal interesting or novel information about ticket splitting? Following King's own advice, one should begin an ecological inference analysis by assessing how much information is deterministically available in the aggregate data (King 1997, 277–291). Although the authors report no results from this step of King's procedure, it is a crucial stage of analysis. Thus, we now turn our attention to the specific methods King advocates for aggregate data researchers (see King 1997, Chapter 16).

Assessing Initial Diagnostics

It is cumbersome to sort through large numbers of tables like Table 3. Instead, provided that one has already simplified the problem, there is a method of succinctly summarizing all the known information in an aggregate data problem with a clever plot, originally proposed by Achen and Shively (1995, 208–209), and later given the name "tomography plot" by King (1997). If an estimation problem can be simplified so that it can be represented by a 2×2 table, these tomography plots allow a researcher to assess the scope of the problem quickly.

Consider the 2×2 representation of a ticket-splitting problem in Table 3. There are two unknown parameters for each district, β^b and β^w. Of those voters who voted for the Democratic presidential candidate, the proportion that voted for the Republican House candidate is $1-\beta^b$. The proportion of Bush voters who split their tickets by voting for the Democratic House candidate is β^w. Both β^b and β^w are unknown. The variables T and $1-T$, λ^b and λ^w are either known or were estimated by a previous run of EI. Though there are two unknown parameters, knowledge of one of these parameters determines the value of the other parameter—if we know the value of β^b, then the value of β^w is simply $\frac{T-\beta^b\lambda^b}{\lambda^w}$. The possible values of β^b and β^w thus form a line (a "tomography line" in King's terminology). Each district can be represented by one of these lines. Although we do not know the true values of (β^b, β^w), we do have some information on their possible values. The true values must lie on the tomography line.

Table 4. The Link Between Tomography Plots and the Distributional Assumption

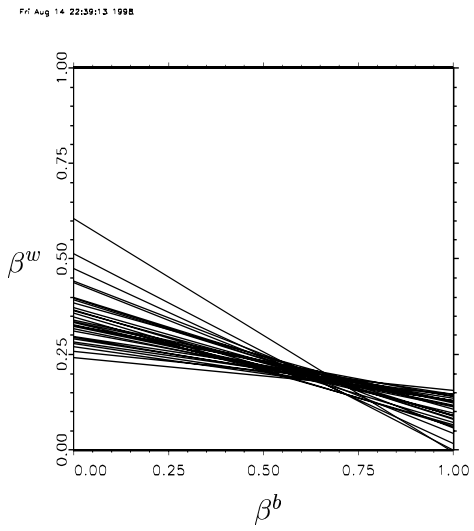
	TBVN Assumption Correct	TBVN Assumption Incorrect
"Informative" Tomography Plot	*Cell 1* *Correct* standard errors Small standard errors	*Cell 2* *Incorrect* standard errors Small and misleading standard errors
"Uninformative" Tomography Plot	*Cell 3* *Correct* standard errors Large standard errors	*Cell 4* *Incorrect* standard errors Large and misleading standard errors

Since tomography plots summarize the available deterministic information in a problem, they constitute a good first diagnostic on how easy or hard the estimation problem will be (King 1997). 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. Since tomography plots are simply diagnostics, the information gleaned from them is merely suggestive and does not allow us to make definitive claims about whether the distributional assumption of the model holds. However, if a tomography plot is "informative," then there is evidence that the distributional assumption may hold. And, if the tomography plot is uninformative, then there is evidence that the distributional assumption may not hold.[5] Whether the distributional assumption holds or not is important because the computation of the standard errors is based on the distributional assumption. So, whether the standard errors are correct or incorrect depends on whether the distributional assumption is correct or incorrect. This logic is summarized in Table 4.

King contends that an informative tomography plot can reasonably be assumed to have been generated by a truncated bivariate normal distribution. Hence, we would attribute a higher probability that the output from our data analysis is summarized by Cell 1 rather than by Cell 2. Similarly, if a tomography plot is uninformative, he claims that we are less likely to believe that the data are generated from the truncated bivariate normal distribution. Hence, an uninformative plot implies that our situation is more likely to be summarized by Cell 4 than by Cell 3. A researcher thus hopes that the tomography plot will be informative: if it is not, the resulting standard errors

[5]In the event that the tomography plot leads one to believe that the distributional assumption is unreasonable, a model that incorporates the truncated bivariate normal distribution might still be adequate provided that one first conditions on the proper set of covariates. In other words, the tomography plot can be thought of as a diagnostic for the necessity of adding covariates to the model.

Figure 1: **Informative Tomography Plot**



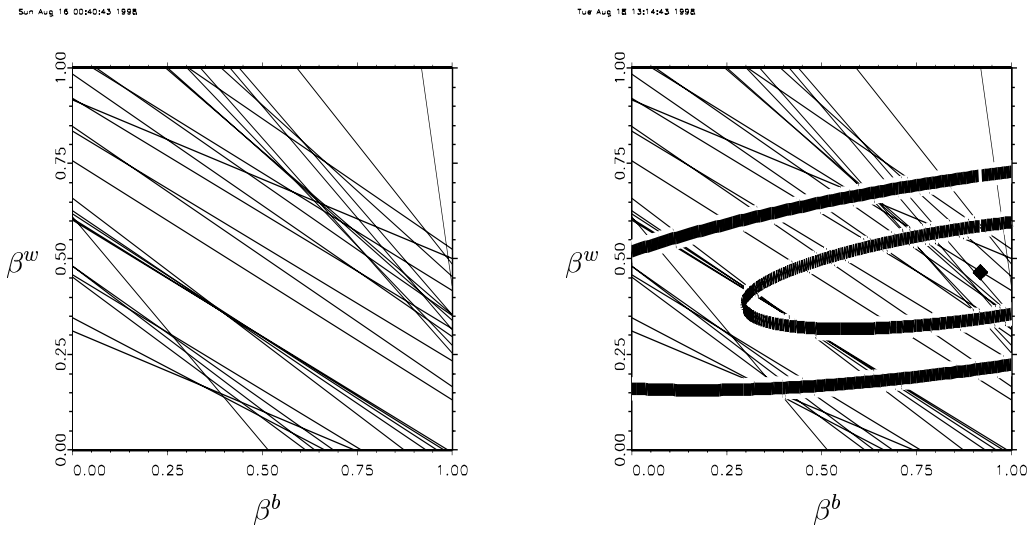
Note: This tomography plot is informative for two reasons. First, all of the lines intersect in one general area of the plot. This gives us confidence in the assumption that all of the lines are related—a key assumption of the EI model. Second, while the bounds on β^b are wide, the bounds on β^w are relatively narrow.

may be too large to be useful. Moreover, the standard errors are likely to be simply incorrect. While any particular set of aggregate data may be described by either Cell 2 or Cell 3, King argues that tomography plots allow us to justify probabilistically that aggregate data are more likely to be described by Cell 4 rather than Cell 3, or by Cell 1 rather than Cell 2 (see King 1997, Chapter 16).

Hence, one's assessment of whether the distributional assumption is correct turns on the nature of the tomography plot. Deciding whether a tomography plot is informative is something of an art, and King does not provide a concrete measure for "informativeness." Consider Figure 1. By King's reasoning, this plot is informative for two reasons. First, we can see from the plot that while the bounds on β^b span the entire permissible $[0, 1]$ range, the bounds on β^w are more narrow and thus supply some information about what the true value of β^w may be.[6] Second, there is a general area of intersection in the tomography plot. If the lines are related (part of the distributional assumption in the EI model), it would not be unreasonable to guess that each line gives us some information about each of the other lines (King 1997, 131). The King estimator incorporates the idea that the true points on each of the lines should fall within the area where the lines

[6]Bounds refer to upper and lower limits on the possible parameter values in light of the known marginals. These can be computed via the method described in Duncan and Davis (1953).

Figure 2: **Uninformative Tomography Plot**

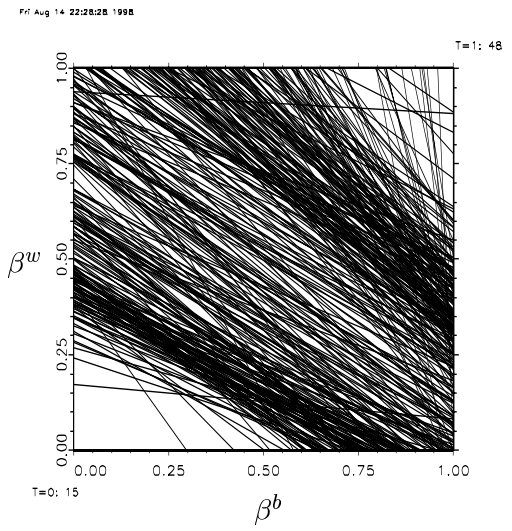


Note: These tomography plots are much less informative than the tomography plot in Figure 1. The lines do not intersect in any one general area of the plot. In addition, the bounds on both β^b and β^w are very wide and span virtually the entire permissible range. The elliptical segments on the right are contour lines that represent the estimated truncated bivariate normal distribution.

generally intersect. In this plot, the area of "general intersection" clearly falls at approximately $(\beta^b, \beta^w) = (0.65, 0.2)$. While this point may not represent the true values, it seems plausible to assume so if we have no other information. Hence, we have gleaned some useful information from this tomography plot. As we shall see, all tomography plots are not as seemingly informative. Sometimes the bounds will not be very informative at all, and, in addition, the tomography lines will not display any sort of commonality. Obviously, in such cases, it is far more difficult to justify forcing a solution to lie in the "general area of intersection" since there is less evidence that the distributional assumption holds.

Contrast the tomography plot in Figure 1 with the tomography plot on the left in Figure 2. It is obvious from this plot that we have very little deterministic information about the underlying data. No "general area of intersection" seems to be present, and the bounds on both parameters are very wide. If one still wants to assume that the distributional assumption is correct, EI can estimate the mode of the distribution. This estimate is indicated by a dot in the plot on the right in Figure 2. However, "if the ultimate conditional distributions are not reasonably close approximations to the truth, incorrect inferences may result" (King 1997, 185). Here, the tomography plot has not given us a good indication that the distributional assumption is correct. Quite the contrary.

Figure 3: **Tomography Plot for Burden-Kimball House Vote-Splitting Data**



Note: This tomography plot is very dense and uninformative, far more similar to the plot in Figure 2 than to the plot in Figure 1.

It is clear from Figures 1 and 2 that all tomography plots are not alike. In some instances, the plot provides a good guess for the parameter estimates (e.g. Figure 1). In other instances, the plot displays seemingly unrelated lines (e.g. Figure 2). One must make distinctions between these two cases, because in the latter instance, EI is not likely to produce a good estimate for the parameters. The distributional assumption is likely to be violated, and the bounds do not allow us to narrow the solution to a small range. Forcing a "solution," by choosing a general point of commonality in these instances is not advisable (see King 1999, 36).

Now consider Figure 3, which displays the tomography plot from the Burden and Kimball House data set. This plot resembles the one in Figure 2 in that both are very uninformative. The main difference is that the Burden-Kimball tomography plot has even more seemingly unrelated lines than the uninformative tomography plot in Figure 2. Again, the bounds are too wide to imply any sort of substantive conclusion. The bounds on β^b are $[0.28, 0.91]$. The bounds on β^w are $[0.24, 0.75]$. Even in this initial stage of assessing the information inherent in the data, it would appear that this split-ticket voting data set does not contain much information about the parameters of interest: the bounds are not much narrower than [0,1], and no general area of intersection is evident. Hence, any inferences made from these data are not likely to be very reliable (King 1997, 185). If the standard errors indicate otherwise, they can only be incorrect.[7]

[7]Burden and Kimball (1998, 536) state that "Because these are maximum likelihood estimates, King's method also

Note that we have not yet begun to estimate the parameters of interest. At this initial stage, we are merely assessing how much information is available for the EI estimation procedure. However, our initial analyses do not portend success in making correct individual-level inferences based on these aggregate data: the bounds are not informative and no mode is apparent. There are some very special situations wherein the method of bounds is truly uninformative yet we are still able to make correct inferences to the individual-level data. Before we move to a discussion of these special circumstances (i.e. when aggregation bias is absent), we discuss a preliminary issue, the accuracy of the computer code which implements King's model. If we cannot trust the computer code, then even if the model is appropriate for the data, the results one obtains are obviously untrustworthy.

THE INSTABILITY OF THE EI PROGRAM

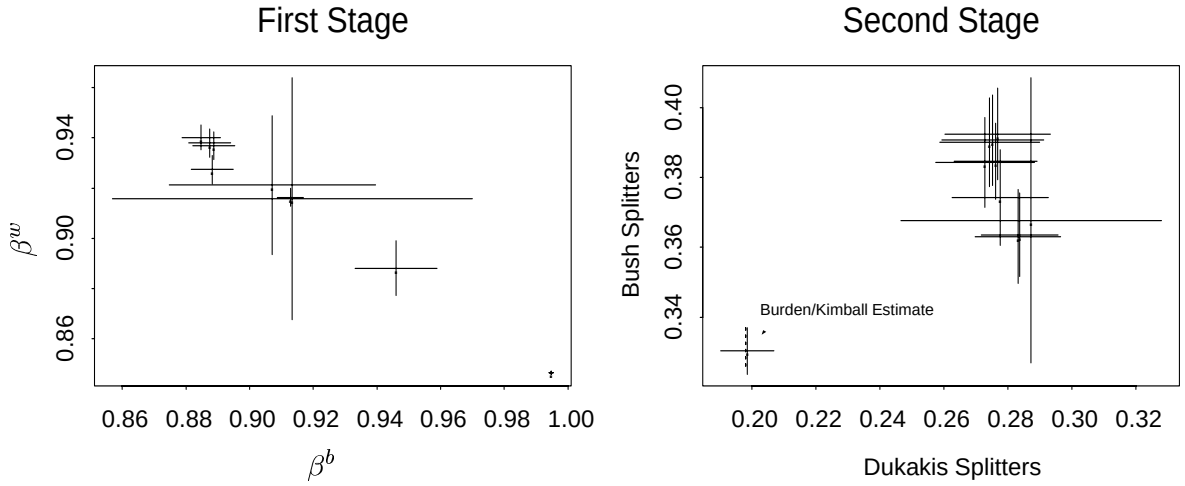
Others have noted apparent software errors in the EI program (Freedman et al. 1998, 1520). Our re-analysis of the Burden and Kimball data reveals some disturbing instability in the results that the program produces. Consequently, replicating a set of results proves more complicated than one would expect.[8]

A replication data set was obtained from the ICPSR data archives. While we were able to replicate exactly the OLS results reported in Table 4 of the Burden and Kimball article, we were unable to replicate the results from EI, though we did eventually come close. In general, one should not expect to produce an exact replication of any results from EI since EI produces slightly (and sometimes not so slightly) different estimates on each run with the exact same data. King acknowledges in the documentation for the program that one should expect to obtain slightly different results on each run because EI uses random simulation to approximate various statistics. He elaborates "this is not a compromise, and it makes computations easy that would otherwise be

produces standard errors for the two ticket-splitting parameters for each congressional district. In this case it yields fairly precise estimates of ticket-splitting." They base their assessment of precision on the small standard errors. We shall see that these standard errors are erroneously calculated, and hence, do not signal genuine precision. Another problem with Burden and Kimball's final estimates of error is that their analysis proceeded through several stages of analysis, each of which produced its own set of errors. Since each stage builds on the previous stage, the errors are compounded. The final set of standard errors assumes away the errors of all of the previous stages. Hence, while their final estimates report small standard errors, these errors do not accurately portray the truth.

[8]In our analysis / replication, we employed the *EzI* Program, versions 1.33, 1.35, 2.01, and 2.20. All of these version showed signs of instability. They are all more recent releases than version 1.21 of *EzI*, the version that Burden and Kimball used to obtain their results. Obviously, it should not matter which version of the program is used. They should all be estimating the same model. However, as we show, it is unclear what model the program is estimating since it will output significantly different results for the exact same data. Moreover, the computational instability was not limited to coefficients and standard errors. For instance, for the ten runs shown in Table A–1, the smallest reported likelihood was -20639.4773 while the largest was 1058.8442. Obviously, the likelihood value should not fluctuate when the model specification is unchanged.

Figure 4: **Instability of the EI Program**



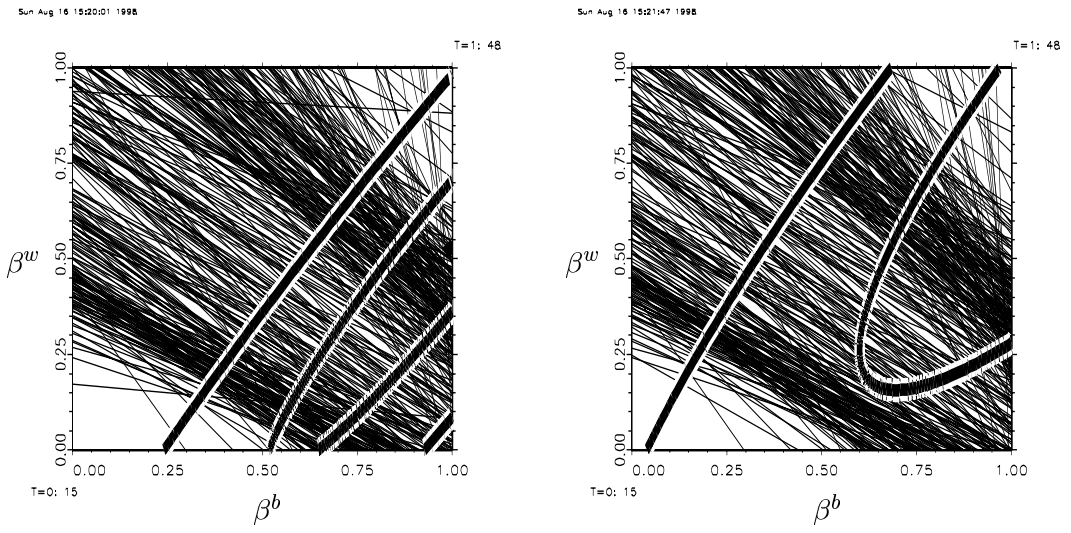
Note: Each cross represents one independent run of the EI program on the Burden and Kimball House data set. The center of the cross is the point estimate. The arms of the cross extend one standard error in each direction. Note that the precision associated with the first-stage estimate where $\beta^b \approx 0.995$ (the very small cross in lower right corner) is not duplicated for any of the other runs. The actual numerical values are listed in Appendix A.

impossible" (1998, 33). However, when these results are not even within a standard error of one another, then proper substantive interpretation becomes seriously impaired.[9].

Recall that the first part of the Burden-Kimball analysis proceeds in two separate stages. In the first stage, they use EI to estimate what proportion of voters abstained from voting in the Congressional election. The bounds are relatively narrow for this stage of the analysis, so one might imagine that EI would have little difficulty producing accurate estimates. Now consider the results depicted in Figure 4 (also presented in Table A–1) of just ten runs of the first stage (all using the same data, of course). The bounds for β^b are $[0.8199, 0.9966]$. The bounds for β^w are $[0.8448, 0.9956]$. The parameter estimates from all runs are within the bounds, but it is clear that many of the runs are significantly and substantively different. Of note is the great variance in standard errors: many runs report a low degree of uncertainty that is incompatible with other runs. These standard errors suggest a precision that cannot be justified. As we noted in footnote

[9]King additionally notes that if one wants to obtain the same results each time, one can set the Gauss random number seed before running EI. However, the issue here is not whether one "wants" the same results each time. One *should* obtain consistent results each time. If the results are not consistent, it is clear that the random simulation or some other aspect of the code is problematic, either incorrectly conceptualized or incorrectly coded. Both of these explanations are equally unsatisfying for the purposes of estimation

Figure 5: **Tomography Plots for the Burden-Kimball Data on two separate runs**



Note: Although the two plots are extremely dense, one can also notice another source of inconsistency in that the tomography lines are not identical. Note the lower left-hand and upper right-hand portions of the plots. In the first plot, there is a tomography line that is close to horizontal at $\beta^w \approx 0.17$. In addition, there is another line that is close to horizontal at the top of the plot at $\beta^w \approx 0.93$. Both of these tomography lines are curiously absent from the second tomography plot even though they have been created from the same data set. There are many other discrepancies between these supposedly identical plots.

7, Burden and Kimball were misled by this apparent precision and their failure to double check EI's performance. Unfortunately, given that we have purportedly estimated the same model on the same data, we now have no reason to believe any one set of these results over the others. Since the circumstances, the data, and the specification were exactly the same, there is no objective basis to make a choice as to which estimates are more correct. The inconsistency is alarming, as these discrepancies are inconsistent with a reliable estimation procedure.[10] Despite these problems, Burden and Kimball made use of these estimates in the second round of their analyses. We now proceed to the second-stage analysis as well.

The results from multiple attempts at the second-stage analyses are not likely to be consistent, given that they begin from different initial premises. Indeed, as we can see from Figure 4 and Table A–1, the second-stage results are substantively different. In Figure 4, the Burden and Kimball

[10]In addition to the problems already cataloged, numerous other disconcerting warnings were printed during the estimation stage. For instance, EI claimed to have computed a Hessian matrix that was not positive definite. (A Hessian matrix *must* be positive definite.) Furthermore, a large number of warning messages were displayed with the caution: "Warning: Some bounds are very far from distribution mean. Forcing xx simulations to their closest bound," where "xx" appeared as a different number in each warning.

result is marked in the second plot with dashed lines (appearing in the lower left-hand corner). From the plot, one can see that we were almost able to replicate this result once (out of 10 total attempts). However, the other nine runs of the EI program on these data produced numbers which clustered away from Burden and Kimball's estimate. Lastly, the tomography plots for two separate runs, with the contour lines, are displayed in Figure 5. It is impossible to tell which plot has detected a better mode since we have already surmised from previously examining the tomography plots that it is debatable whether there is any reasonable mode in these data at all. It is also difficult to determine how the contour lines could differ as substantially as they do between these two plots. Moreover, close inspection reveals many other discrepancies between these supposedly identical tomography plots.

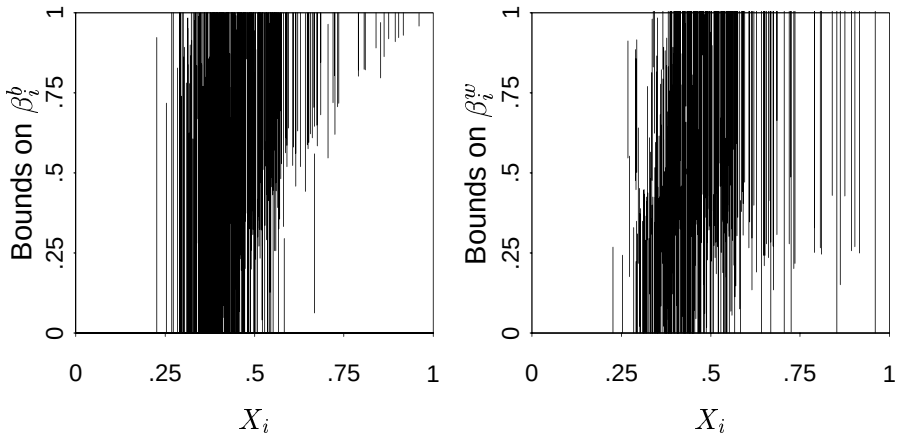
Aggregation Bias and Covariate Selection

As we mentioned previously, it is possible to obtain reasonable estimates of individual-level parameters given only aggregated data if the aggregated data set contains no aggregation bias.[11] This holds true regardless of how much information seems to appear in the data set. In fact, if no aggregation bias exists in the data, simple OLS will provide reasonable, unbiased, and consistent estimates of the parameters in question (Goodman 1953). EI should perform likewise, provided that there are no errors in the code implementing the procedure.

Hence, the question of whether we are able to make reasonable ecological inferences turns on the issue of aggregation bias (Cho 1998). Though King claims that his method is "robust" to violations of the aggregation bias assumption, the evidence strongly suggests otherwise. King's claim originates in (and holds only for) an unorthodox definition of "robustness." He contends that EI is robust because it will never produce estimates which are outside the $[0, 1]$ bounds. While this property is nice, it does not follow that the estimates from EI are close to the truth, or even within a standard error or two of the actual values. Indeed, King cannot claim that his model produces unbiased or consistent estimates if aggregation bias is present. If there is aggregation bias, the estimates from EI are biased, and the discrepancy between the estimates and the true values does not converge in distribution to zero as the sample of data points becomes large (Cho 1998). A fundamental issue, then, is whether aggregation bias exists in the Burden and Kimball

[11] The assumption of no aggregation bias holds if the parameters (β^b and β^w) are not correlated with the regressors, i.e. the X variables. In this application, that would mean that levels of Democratic President-Republican Representative and Republican President-Democratic Representative voting are not correlated with levels of support for the Presidential candidates.

Figure 6: **Aggregation Bias Diagnostic**



Note: Aggregation bias exists if X and β are correlated. The only deterministic information available for β is contained in the bounds. Each line in these plots indicates the range on the bounds for a district. These bounds are generally wide. To the extent that there is any pattern, it indicates a correlation between X and β.

data set. If it does, then the King model (without covariates) is untrustworthy and should not be used.

Assessing the degree to which aggregation bias exists is a daunting task, one that is replete with uncertainty. There are, however, some methods that shed some insight into this problem. One method is to examine the aggregation bias diagnostic plot suggested by King (1997, 238). This plot is shown in Figure 6. Aggregation bias exists if there is a relationship between X and β^b or β^w. Since β^b and β^w are unknown, we are able to plot only the bounds for these two parameters.[12] Clearly, the vast majority of the bounds cover the entire permissible range from 0 to 1. However, we can see that the first plot suggests that the aggregation bias for β^b may be severe, since β^b and X appear to be strongly correlated.

A second method for testing whether aggregation bias exists is simply to run the OLS model. If no aggregation bias exists, the assumptions of OLS are met, and so OLS will yield consistent and unbiased estimates. The OLS model for these data yields

$$T = 0.0331 + 1.075X \tag{1}$$

[12] And, in this instance, these are not genuine bounds because they incorporate the error from a previous stage of the analysis.

where X = the proportion of voters who voted for Dukakis, and T = the proportion of Dukakis voters who also voted for the Democratic House candidate. Clearly, OLS does not yield the correct solution: the model estimates that 110% of the Dukakis voters supported Democratic House candidates, a value that is outside the logical bounds. The assumptions of OLS seem to be violated. Producing out-of-bounds estimates is thus a very useful feature of the linear probability model. While out-of-bounds estimates are clear signals of a misspecified model, the converse of this statement is false: estimates which are within the bounds do not signify a correctly specified model. Since EI *always* produces parameter estimates which are within the $[0, 1]$ bounds, it has no such diagnostic value for assessing whether the specification is correct.

Since OLS produces correct estimates if no aggregation bias exists in the data set, King would conclude from Equation (1) that there is a high probability of aggregation bias in the data set.[13] Given that we then have a prior that aggregation bias exists in the data set, and since EI is not robust to violations of the aggregation bias assumption, its estimates are immediately suspect. It is possible that EI will provide reasonable estimates despite the presence of aggregation bias. However, this result would be the exception, not the rule, since EI is a biased and inconsistent estimator in the presence of aggregation bias. The exception occurs when the bounds are informative. Nonetheless, it is clear from both Figures 3 and 5 that the bounds are far from informative in this data set—the vast majority of the bounds span the entire range of possibilities.

One method for mitigating the effects of aggregation bias is to include covariates in the model (King 1997, 288). If these covariates control aggregation bias by accounting for the correlation between the parameters and the regressors, then the model will produce the correct estimates. Burden and Kimball included one covariate in their model specification, a dummy variable that indicates whether or not a district is located in the South. They included this variable based on their belief that individuals who live in the South are unlike individuals who do not live in the South when it comes to decisions about ticket splitting. They offered no indirect evidence (e.g. survey data) to support this contention. Regardless of whether their intuitions about the South are correct or not, including this covariate does not affect the estimates. Their justification for its inclusion was "to account for possible aggregation bias and to improve the estimates" (1998, 536). However, since the two models produce indistinguishable estimates, there is no reason to believe that a South dummy has any desirable effect in mitigating the aggregation bias. If the specification

[13] Checking the results from Goodman's regression line is a diagnostic suggested by King. He states that "If Goodman's regression line does not cross both the left and the right vertical axes within the [0,1] interval, there is a high probability of aggregation bias. If the line does cross both axes within the interval, we have less evidence of whether aggregation bias exists" (King 1997, 282).

with no covariates is ill-advised, so too is the specification with only the variable "South."

Burden and Kimball were correct that there is a need to alleviate the aggregation bias in their data set, and that incorporating the correct covariates would achieve this end. The problem they encountered is that EI does not provide a test for whether one specification is better than another specification. EI users thus find themselves in a truly problematic situation: they cannot determine which specification is correct, but different specifications can produce very different and inconsistent results.

Consider Table 5, which shows the results from different model specifications (the covariates are from the set Burden and Kimball use in their OLS analysis of their EI-estimated split-ticket voting levels). In these different specifications, the estimated percentages of ticket-splitters vary widely. The values for the standard errors are large for some specifications and extremely small for other specifications. This erratic performance is illustrated in Figure 7. Each rectangle is centered at the point estimate and extends one standard error in each direction. The rectangles would overlap if the models were consistent, but they do not. Even after accounting for the standard error, few of the point estimates are consistent. Substantively, this is a problem because the alternative specifications imply different types of voting behavior. In addition, the computer program, citing various errors, was not able to compute estimates for certain other specifications. To settle on the best available model of the split-ticket vote, one must somehow choose from among these different specifications. Finding a proper specification is always a major step, but in the aggregate-data context, there are enormous barriers (see Achen and Shively 1995, Chapter 4 for an extensive discussion; see also Erbring 1990, 264–265, and Haitovsky 1973).

The advice that King offers is that one should include covariates that can "be justified with specific reference to prior substantive knowledge about a problem" (King 1997, 173). He provides no empirical test for choosing covariates, but only this admonition to exercise one's belief about what may be true. This would be unproblematic if different researchers always reached common substantive conclusions after imposing their own beliefs on the model specification. As this congruence virtually never occurs, however, it is obvious that a formal method is needed to determine which covariates are likely to belong in a properly specified model. After all, "including the wrong variables does not help with aggregation bias" (King 1997, 173).

Although the problem of identifying proper covariates with formal tests is not solved, and may not even admit a "solution" in the sense of a universally optimal test, there is at least one start on this problem. Tam (1997) notes that the statistical literature on changepoints and parameter

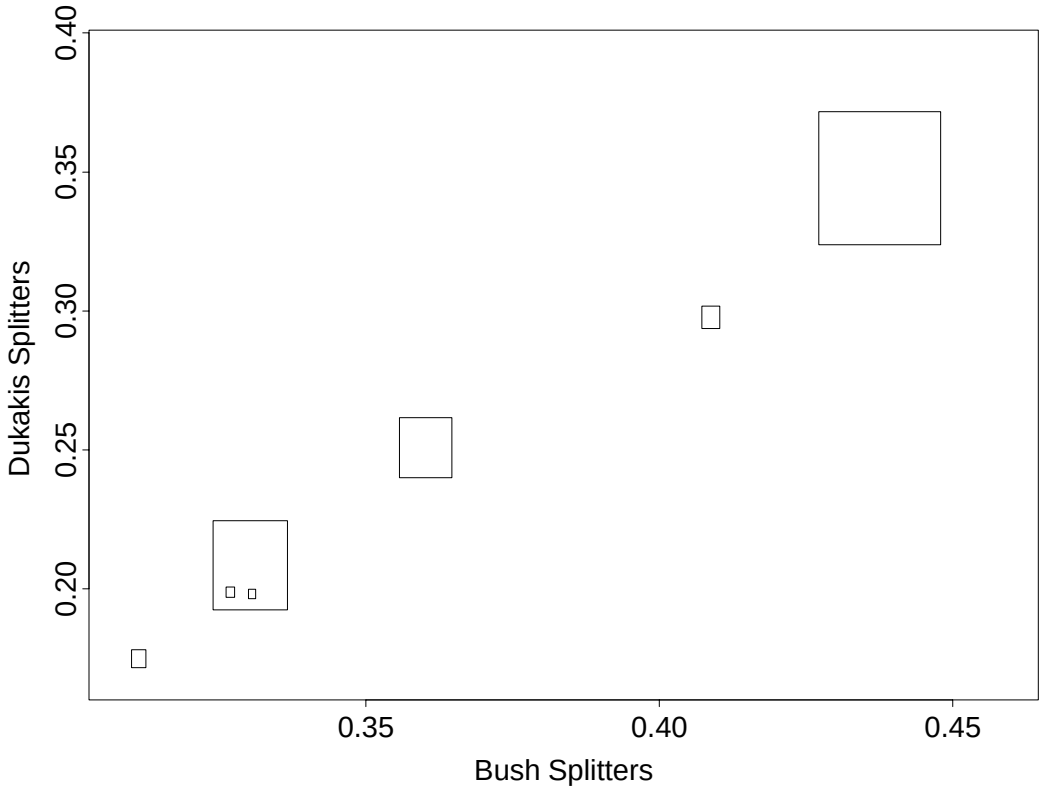
Table 5. The Effect of Different Covariates

	Bush Splitters	Dukakis Splitters
No covariates	0.3306	0.1982
	(0.0058)	(0.0074)
South*	0.3310	0.1980
	(0.0060)	(0.0070)
NOMINATE Score^{bw}	0.4376	0.3477
	(0.0986)	(0.1075)
NOMINATE Score^{w}	0.4088	0.2977
	(0.0146)	(0.0180)
Experienced Challenger and Money	0.3602	0.2508
	(0.0421)	(0.0488)
Experienced Challenger	0.3303	0.2085
	(0.0601)	(0.0719)
Democratic Incumbent^{b}	0.3113	0.1749
Republican Incumbent^{w}	(0.0115)	(0.0143)
Ballot^{bw}	0.3269	0.1988
	(0.0068)	(0.0084)
NOMINATE Score^{b}	EI could not estimate	
Experienced Challenger, Democratic and Republican Incumbent	EI could not estimate	

A superscript ^{b} indicates the covariate was used for β^{b}.
A superscript ^{w} indicates the covariate was used for β^{w}.
* Burden and Kimball's specification.

Standard errors in parentheses

Figure 7: **The Effect of Different Covariates**



Note: Each rectangle represents the results of a particular specification of the EI model. The rectangles are centered on the point estimate and are one standard error wide and high. If the separate estimates were consistent with one another, the rectangles would all overlap.

constancy is addressed to an analogous problem, and so is very promising as a source for guidance on how to pick covariates in the aggregate data context. The reason to introduce covariates, recall, is because the parameters of interest are not constant throughout the data set. Hence, a useful empirical test should discriminate between covariates that do divide the sample into subgroups in which parameters are nearly constant and those that do not (Cho 2001).

In a general changepoint problem, a random process generates independent observations indexed by some non-random factor, often, but not exclusively, time. One may wish to test whether a change occurred in the random process by searching over partitions that divide the data into subsets appearing to have different distribution functions. Again, the subsets can be sorted chronologically, or can be generated from an ordering of some other measured property. The literature is large and diverse: some tests assume that the number of changepoints is unknown, while others assume a fixed number of changepoints; some fix the variance of the distribution, while others estimate the variance as a parameter; some assume that the different distributions take similar forms, while others allow more flexibility in this regard. Tests vary in nature as well: some are Bayesian (e.g. Smith 1975, Schulze 1982, Carlin, Gelfand, and Smith 1992), some are parametric (e.g. Ritov 1990, Andrews 1993, Andrews, Lee, and Ploberger 1996), some are nonparametric (e.g. Wolfe 1984, Carlstein 1986), and some are related to time series analysis (e.g. Brown, Durbin, and Evans 1975). In short, there are diverse means by which one can draw inferences about changepoints and constancy, or lack thereof, of parameters. For present purposes, what is important is that the general object in this literature is to find a means for partitioning data sets into subsets within which there is some degree of parameter constancy. Since this is precisely the goal for the researcher choosing covariates to remove aggregation bias in an ecological inference problem, the application of changepoint tests to aggregate data problems seems extremely promising.

Cho (2001) introduces one formal covariate-selection test adapted from time-series analogs. There is not likely to be *one* covariate-selection test that is optimal for all aggregate data problems, but employing an empirical test is clearly preferable to imposing subjective beliefs. It is important that aggregate data analysts have some standard by which to judge whether one specification is superior to another. Further development of well-specified statistical tests for covariate selection should be the priority for aggregate data research.

Burden and Kimball were in need of just such a test, since their data exhibited aggregation bias. Lacking any means by which to compare covariates that might alleviate the problem, they settled on one covariate chosen on qualitative grounds. Unfortunately, this covariate did not perform the

necessary function of removing aggregation bias, and their analysis suffered accordingly.

Moreover, poor covariate selection was not the sole specification problem in Burden and Kimball's investigation of split-ticket voting. Their analysis proceeded through multiple stages: they first estimated abstention (with EI), then estimated ticket-splitting rates, conditional on the abstention estimates (again with EI), and finally, modeled these district-level estimates of split-ticket voting as a function of candidate, institutional, and constituency traits (with OLS). Specification issues arise at every stage of the analysis, and, naturally, the accuracy and validity of each stage depend strongly on the accuracy and validity of the preceding stages. Since the errors from each of the stages compound, the final results are highly prone to indeterminacy.

Intention and Ticket-splitting

Burden and Kimball contend that their research makes two distinct contributions to the study of ticket-splitting. First, as pioneers in applying King's EI methods, they purport to provide the first accurate estimates of the extent of ticket-splitting. Second, their analysis of splitting (as estimated by EI) reveals that it is primarily an unintentional rather than intentional activity. Americans simultaneously support different parties at a given moment not because they prefer to see power balanced or shared, but because strategic choices by candidates and parties induce splitting. We have already demonstrated that the first of these innovations is more apparent than real—EI certainly does not produce new levels of accuracy in estimating ticket-splitting behavior. To conclude, we will briefly re-evaluate the second point.

The term "intentional" could be ambiguous in this context, but the authors clarify that their interest lies in assigning primary responsibility for ticket-splitting to either candidates or voters (1998, 533). If levels of tickets-splitting seem to respond to candidate traits such as incumbency, spending differentials, or candidate experience, they propose, it is not the case that the masses deliberately divide their support between parties. Thus, they conclude, the candidates, not the voters, move first. It is, of course, already very well known that contemporary American elections feature a substantial incumbency advantage. To verify that some ticket-splitting seems to originate in incumbents' skills at drawing cross-party support, however, is not to rule out that voters are quite consciously spreading support across parties or ideologies. Burden and Kimball did not test whether incumbents are helped or hindered in drawing non-party-based support by the expected fates of their parties' presidential candidates. In that respect, they do not give "intentional" ticket-splitting much chance to surface.

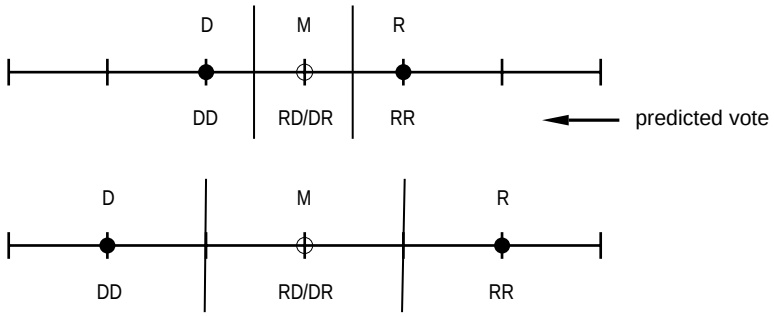
The critical result for their claims about balancing and intent, ultimately, is an insignificant coefficient on the variable they label "ideological distance." They operationalize this variable as the mean distance on a 7-point ideology scale between Bush (R_p) and the Democratic Senate candidate (D_s), as assigned by a state's Senate Election Study (SES) respondents. There are some problems with this construction. Aggregation to state means is noisy: in most states, standard deviations of R_p and D_s span about a quarter of the entire interval. More importantly, the idea that greater spread between these two candidates might yield more ticket-splitting relies on some strong, unstated assumptions about intra-party homogeneity, voter distributions, and the origin of vote splitting.

Figure 8A illustrates the logic whereby spatial party differentials might lead to vote splitting. If both Democrats (e.g., D_s and D_p, where s and p denote "senate" and "presidential" candidates) are located at D, both Republicans at R, and if expected policy outcomes for unified government are, thus, D and R, but for divided government are some weighted average of D and R, say M, then standard proximity theory identifies cutpoints defining zones in which voters should prefer to vote straight tickets (DD or RR) or split tickets (DR or RD). Then, as $|D_s - R_p|$ grows, the central region containing split-ticket voters grows. Burden and Kimball's "ideological distance" variable is thus constructed on three assumptions: first, that voters react to expected policy outcomes, not candidates per se; second, that the two Democrats and two Republicans in question are ideologically very similar, if not identical; and, third, that a substantial portion of the electorate resides in the center of the ideological spectrum, so that enlargement of the middle region does result in more split-ticket voting occurring.

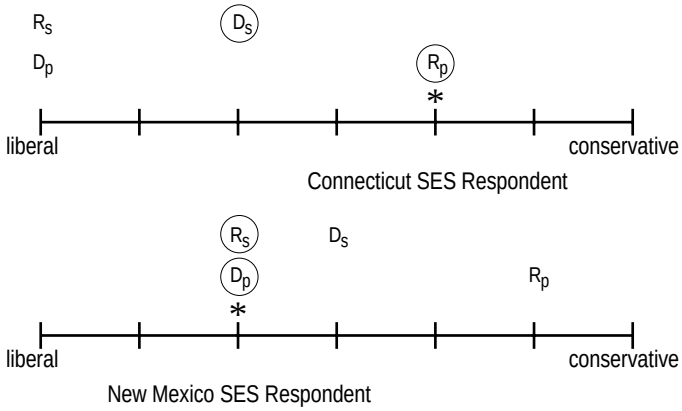
Note, then, that the logic of their test fails if voters do not perceive the two Republicans and two Democrats to be ideological twins or if the district's electorate is bipolar. On the first point, consider Figure 8B, and suppose that Connecticut has a symmetrical (e.g. uniform) voter distribution. As R_p moves right, $|D_s - R_p|$ increases. Under the same assumptions about voting as just applied to Figure 8A, though, the amount of split ticket voting should *decrease* with this increase in $|D_s - R_p|$, since the ticket splitters are now those in the outer regions, given this particular arrangement of candidates. This is the *exact opposite* effect from that shown by Figure 8A and assumed to apply everywhere by Burden and Kimball. And, although they point out in their footnote 11 (1998, 539) that the SES set uniquely provides the necessary data to test a balancing thesis since it includes placements of respondents and *all four candidates* on a common scale, only half of this information is actually incorporated in their construction of "ideological distance." Either large

Figure 8: **Alternative Spatial Theories of Split-Ticket Voting**

A. Expected-Policy Voting



B. Candidate-Proximity Voting



Note: Predicted and actual votes cast are circled in part B. For these two respondents, actual votes are identical to votes as predicted by expected-policy and candidate-proximity theories.

ideological variation between different party nominees or a non-centrally distributed electorate thwarts interpretation of their regression results.

There is, moreover, a plausible variety of intentional vote splitting not captured by their logic. Voters who select candidates only according to ideological proximity, without making projections about policy outcomes that will result from the various permutations of candidate victories, can intentionally split tickets if they perceive there to be large differences in the positions of candidates from the same party. Figure 8B shows two actual SES respondents (marked with asterisks) whose split-ticket votes exactly match the predictions of simple candidate-proximity theory. Note that $|D_s - R_p|$ is identical in the two cases, even though one is a vote of $R_p D_s$ and the other a vote of $D_p R_s$. For both of these respondents, in fact, reported votes are consistent with either expected-policy voting or candidate-proximity voting. This observational equivalence also undercuts strong claims about voter intentions.

More generally, spatial theories of voting are many and varied, and even if one posits that voters choose according to policy outcomes, the proper econometric specification to test for "intent" will depend critically on the underlying formal model. Merrill and Grofman's recent work (1999) unifying directional and proximity models is an excellent blueprint in this regard. Burden and Kimball's analysis, by contrast, is not nearly flexible enough to support the strong conclusion that voters do not consciously choose to split tickets.

CONCLUSION

Burden and Kimball's search for "intention" in vote-splitting is more accurately a test for a particular kind of balancing behavior, under key assumptions about ideological homogeneity within parties and distributions of district electorates. That analysis is not general enough to support strong conclusions about voting behavior. But this is a secondary issue: the biggest flaw in the analysis lies in the dependent variable. The authors claim as their principal achievement to have developed the first-ever accurate district-level estimates of vote-splitting. However, their faith in the EI procedure is misplaced, and their findings are not truly findings at all. Excitement over King's contributions to the ecological inference problem needs to be tempered in the light of EI's clear limitations, and the many potential pitfalls in any aggregate data problem. It is unreasonable to declare new-found knowledge when the novel findings depend critically on very strong and unverifiable assumptions about the underlying individual-level data. Split-ticket voting behavior remains a fascinating topic, and it also remains a topic plagued by severe data-analysis barriers.

Appendix: Program Instability

Table A–1. Results from the First Two Estimation Stages

	First Stage Model		Second Stage Model	
	β^b	β^w	Bush Splitters	Dukakis Splitters
Estimates from				
Burden and Kimball	*not reported*		0.3310	0.1980
			(0.0060)	(0.0070)
Run 1	0.9947	0.8463	0.3677	0.2873
	(0.0007)	(0.0006)	(0.0409)	(0.0408)
Run 2	0.9947	0.8463	0.3631	0.2832
	(0.0007)	(0.0006)	(0.0135)	(0.0134)
Run 3	0.8848	0.9402	0.3636	0.2837
	(0.0060)	(0.0050)	(0.0120)	(0.0120)
Run 4	0.9134	0.9158	0.3303	0.1985
	(0.0565)	(0.0482)	(0.0067)	(0.0083)
Run 5	0.9460	0.8880	0.3846	0.2762
	(0.0128)	(0.0109)	(0.0109)	(0.0130)
Run 6	0.8887	0.9369	0.3743	0.2776
	(0.0066)	(0.0056)	(0.0137)	(0.0151)
Run 7	0.8874	0.9380	0.3907	0.2753
	(0.0066)	(0.0056)	(0.0129)	(0.0159)
Run 8	0.8882	0.9274	0.3925	0.2769
	(0.0066)	(0.0056)	(0.0132)	(0.0165)
Run 9	0.9071	0.9212	0.3901	0.2743
	(0.0324)	(0.0277)	(0.0127)	(0.0156)
Run 10	0.9128	0.9163	0.3843	0.2729
	(0.0041)	(0.0035)	(0.0129)	(0.0154)

Standard errors in parentheses.
Note the wide discrepancies across runs in the estimates of both the coefficients *and* the associated standard errors. In the first-stage, $\beta^b \in [0.8848, 0.9947]$ while the errors range from 0.0007 to 0.0565. And $\beta^w \in [0.8463, 0.9402]$ with errors ranging from 0.0006 to 0.0482. In stage two, Bush splitters range from 0.3310 to 0.3925 while Dukakis splitters range from 0.1980 to 0.2873. In any single run (e.g. Run 4), one might be led to believe that the standard error is minuscule even though such precision is not verified by subsequent runs.

References

Achen, Christopher H. and W. Phillips Shively. 1995. *Cross-Level Inference.* Chicago, IL: University of Chicago Press.

Alesina, Alberto and Howard Rosenthal. 1995. *Partisan Politics, Divided Government, and the Economy*. Cambridge: Cambridge University Press.

Andrews, Donald W.K., 1993. "Test for Parameter Instability and Structural Change with Unknown Change Point." *Econometrica* 61 (July): 821–856.

Andrews, Donald W.K., Inpyo Lee, and Werner Ploberger. 1996. "Optimal Changepoint Tests for Normal Linear Regression." *Journal of Econometrics* 70 (January): 9–38.

Brown, R. L., J.Durbin, and J.M. Evans, 1975. "Techniques for Testing the Constancy of Regression Relationships Over Time." *Journal of the Royal Statistical Society, Series B* 37 (2): 149–192.

Burden, Barry C. and David C. Kimball. 1998. "A New Approach to the Study of Ticket Splitting." *American Political Science Review* 92 (September): 533–544.

Carlin, Bradley P., Alan E. Gelfand, and Adrian F.M. Smith. 1992. "Hierarchical Bayesian Analysis of Changepoint Problems." *Applied Statistics* 41 (2): 389–405.

Carlstein, E. 1988. "Nonparametric Change-Point Estimation." *The Annals of Statistics* 16 (March): 188–197.

Cho, Wendy K. Tam. 1998. "Iff the Assumption Fits: A Comment on the King Ecological Inference Solution." *Political Analysis* 7: 143–163.

Cho, Wendy K. Tam. 2001. "Latent Groups and Cross-Level Inferences." *Electoral Studies.* Forthcoming.

Duncan, Otis Dudley, and Beverly Davis. 1953. "An Alternative to Ecological Correlation," *American Sociological Review* 18 (December): 665–666.

Erbring, Lutz. 1990. "Individuals Writ Large: An Epilogue on the 'Ecological Fallacy.' *Political Analysis* 1: 235–69.

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

Freedman, D.A., S.P. Klein, M. Ostland, and M.R. Roberts. 1998. "On 'Solutions' to the Ecological Inference Problem." *Journal of the American Statistical Association* 93 (December): 1518–1522.

Freedman, D.A., M. Ostland, M.R. Roberts, and S.P. Klein. 1999. "Response to King's Comment." *Journal of the American Statistical Association* 94 (March): 355–357.

Goodman, Leo A. 1953. "Ecological Regressions and Behavior of Individuals." *American Sociological Review* 18 (December): 663–664.

Haitovsky, Yoel. 1973. *Regression Estimation from Grouped Observations.* New York: Hafner.

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

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

King, Gary. 1998. "EI: A Program for Ecological Inference." Version 1.18, February 20. Harvard University.

King, Gary. 1999. "The Future of Ecological Inference Research: A Reply to Freedman et al." Letter to the Editor of the *Journal of the American Statistical Association* 94, 445 (March): 352–355.

Merrill, Samuel III and Bernard Grofman. 1999. *A Unified Theory of Voting: Directional and Proximity Spatial Models*. Cambridge: Cambridge University Press.

Ritov, Y. 1990. "Asymptotic Efficient Estimation of the Change Point with Unknown Distributions." *The Annals of Statistics* 18 (December): 1829–1839.

Schulze, U. 1982. "Estimation in Segmented Regression: Known Number of Regimes." *Mathematische Operationsforschung und Statistik, Series Statistics* 13 (2): 295–316.

Smith, A.F.M. 1975. "A Bayesian Approach to Inference About a Change-point in a Sequence of Random Variables." *Biometrika* 62 (August): 407–416.

Tam, Wendy K. 1997. "Structural Shifts and Deterministic Regime Switching in Aggregate Data Analysis." Master's Essay. Department of Statistics. University of California at Berkeley.

Wolfe, Douglas A. and Edna Schechtman. 1984. "Nonparametric Statistical Procedures for the Changepoint Problem." *Journal of Statistical Planning and Inference* 9 (May): 389–396.