An Easy and Accurate Regression Model for Multiparty Electoral Data

Michael Tomz
Department of Political Science,
Stanford University, Stanford, CA 94305-6044
e-mail: tomz@stanford.edu

Joshua A. Tucker
Department of Politics and Woodrow Wilson School,
Princeton University, Princeton, NJ 08544
e-mail: jtucker@princeton.edu

Jason Wittenberg
Department of Political Science,
University of Wisconsin, Madison, Madison, WI 53706-1389
e-mail: witty@polisci.wisc.edu

Katz and King have previously proposed a statistical model for multiparty election data. They argue that ordinary least-squares (OLS) regression is inappropriate when the dependent variable measures the share of the vote going to each party, and they recommend a superior technique. Regrettably, the Katz–King model requires a high level of statistical expertise and is computationally demanding for more than three political parties. We offer a sophisticated yet convenient alternative that involves seemingly unrelated regression (SUR). SUR is nearly as easy to use as OLS yet performs as well as the Katz–King model in predicting the distribution of votes and the composition of parliament. Moreover, it scales easily to an arbitrarily large number of parties. The model has been incorporated into *Clarify*, a statistical suite that is available free on the Internet.

1 Introduction

Katz and King (1999) propose a new model to explain and predict electoral results in multiparty democracies. The authors argue that ordinary least-squares (OLS) regression is inappropriate when several parties contest an election and the dependent variable measures the share of the vote going to each party. First, OLS assumes that the dependent variable is

Authors' note: All data and relevant unpublished materials are available at the *Political Analysis* Web site. We thank Jonathan Katz and Gary King for helpful comments and access to their data and computer code. We are also grateful to John Gibson for sharing his data, to Scott Golenbock for research assistance, and to Justin Burchett, Simon Jackman, Gary King, Dan Posner, Ken Scheve, David Simon, Robert Van Houweling, Steven Webb, and David Weimer for useful suggestions. Part of this research was conducted while Wittenberg was a visiting scholar at the Helen Kellogg Institute for International Studies at Notre Dame.

theoretically unbounded and could fall anywhere on the real number line, whereas in reality each party's share of the electoral pie must lie between 0 and 1. Second, OLS treats the vote for each party independently, but they are not independent because the proportions for all parties must sum to 1. Given these problems, Katz and King suggest a method analogous to OLS but tailored to the unique features of multiparty systems. Their approach, which involves transforming the data and maximizing the likelihood of a multivariate t distribution, overcomes the limitations of OLS.

Their methodological innovation is overdue, since the vast majority of democracies around the globe feature elections with more than two parties. In the postwar period, the United States stands alone as the only industrialized country with a consistent two-party system; in other industrialized nations, three or more significant parties regularly compete for power. A wider sample, including not only advanced but also developing countries, displays a similarly diverse electoral landscape. Cox (1997) calculated the effective number of electoral parties for 54 democracies at various stages of development and found that 44 had significantly more than two effective parties. Of the 10 that did not, only 1 (the United States) had a population above 2.5 million people. Clearly, most people in democratic countries vote in elections with more than two relevant parties. To analyze these elections properly, researchers need a model that reflects the special characteristics of multiparty data.

Regrettably, the methodological advance of Katz and King (KK) comes at considerable cost. First, the model requires a high level of statistical expertise: users must know how to program and maximize likelihood functions, evaluate high-dimensional integrals, and understand t-based regression, tasks beyond the ken of most researchers. Second, as Katz and King themselves note, the model is computationally demanding for more than three political parties.

We offer a sophisticated yet practical alternative, based on Zellner's (1962) seemingly unrelated regression (SUR) model. Our approach is nearly as easy to use as OLS yet performs as well as KK on a wide range of criteria, including the ability to predict the distribution of votes and winners of elections. SUR also scales easily to accommodate a large number of parties. The procedure is implemented in most statistics packages and has been integrated into *Clarify* (King et al. 2000; Tomz et al. 2001), a suite of programs available for free on the World Wide Web. Using *Clarify*, researchers can run SUR and interpret the results with only four simple commands, which we summarize at the *Political Analysis* Web site.

We are not alone in seeking a practical alternative to KK. Jackson (2002) and Mikhailov, Niemi, and Weimer (MNW) (2002) also recommend versions of SUR for multiparty election data.[1] Following Theil (1970), Jackson and MNW note that aggregate voting data may be heteroskedastic both within and across electoral districts. As a solution, the authors propose innovative weighting schemes and iterative techniques. Their contributions go a long way toward making SUR a flexible research tool for students of electoral politics. Our approach is even more streamlined. Although we applaud any effort to model possible sources of error, the extra variability that Jackson and MNW capture varies inversely with the number of voters in a district, so in districts of moderate size the errors should be relatively small. Our procedure is more straightforward and accessible, since it rests on feasible generalized least squares (FGLS) as programmed in many statistics packages.

In the sections that follow we compare our approach with KK, widely considered the "state of the art" in multiparty voting models. Our tests show that SUR performs about as well as—and sometimes better than—KK, without resorting to t-based regression, iterated

[1]Both papers extend the pioneering work of Jackson et al. (1998).

estimation algorithms, or elaborate weighting procedures. We begin by introducing our variant of SUR, showing how it differs from KK, and describing algorithms for interpreting the results. We then evaluate the performance of SUR versus KK, first via a Monte Carlo experiment and then using the same British election data that Katz and King selected to highlight features of their model. In both tests, KK and SUR yield virtually identical results. To illustrate how SUR accommodates a large number of parties, we examine the Polish parliamentary election of 1993 and the fates of seven parties that contested for seats. With SUR we draw several new and interesting conclusions about the effects of economic conditions on the composition of the Polish parliament. We conclude by noting applications beyond the sphere of multiparty elections.

2 SUR

Aggregate voting data exhibit two important features that we should build into statistical models. Each party's proportion of the total vote must fall between 0 and 1, and the shares of all parties must sum to 1. Adopting the notation of Katz and King, the proportion of the vote ($\mathbf{V}$) for each party j $(j = 1, \ldots, J)$ in electoral district i $(i = 1, \ldots, I)$ must meet the following criteria: $\mathbf{V}_{ij} \in [0, 1]$ for all i and j, and $\sum_{j=1}^{J} \mathbf{V}_{ij} = 1$ for all i.

To analyze such data we recommend SUR, a multiequation version of OLS. As noted previously, OLS requires a potentially unbounded dependent variable. Following Katz and King, we convert the votes to an unbounded scale by applying the multivariate logistic transformation. Specifically, we calculate the natural log of each party's share of the vote, relative to that of a reference party, J. We denote the vector of $J - 1$ log ratios for electoral district i as $\mathbf{Y}_i = [\ln(\mathbf{V}_{i1}/\mathbf{V}_{iJ}), \ln(\mathbf{V}_{i2}/\mathbf{V}_{iJ}), \ldots, \ln(\mathbf{V}_{i(J-1)}/\mathbf{V}_{iJ})]$ and assume it is multivariate Normal with mean vector $\boldsymbol{\mu}_i$ and variance matrix $\boldsymbol{\Sigma}$. We then model $\boldsymbol{\mu}_i$ as a linear function of explanatory variables ($\mathbf{x}$) and effect coefficients ($\boldsymbol{\beta}$), such that $\boldsymbol{\mu}_i = [\mathbf{x}_{i1}\boldsymbol{\beta}_1, \mathbf{x}_{i2}\boldsymbol{\beta}_2, \ldots, \mathbf{x}_{i(J-1)}\boldsymbol{\beta}_{(J-1)}]$.

Researchers can estimate $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ via seemingly unrelated regression, a method for studying regression equations that appear to be unrelated but in fact have correlated errors. $\mathbf{Y}_i \sim N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma})$ is a stack of $J - 1$ regressions. Moreover, the error terms are correlated across these equations because the dependent variable is constructed from vote shares, such that a higher log ratio for one party means a lower log ratio for the others. SUR is, therefore, an appropriate estimation technique. Our variant of SUR employs the FGLS algorithm as described by Greene (2000).

Of course, one could estimate the $\boldsymbol{\beta}$'s via $J - 1$ separate linear regressions, but SUR is more convenient, since in most statistics packages users can estimate all the equations with one simple command. SUR is potentially more efficient as well: if the explanatory variables differ from one equation to the next, SUR will take advantage of interesting information about covariance among the equations. At a minimum, though, SUR will be more convenient and no less efficient than equation-by-equation OLS.

Having run SUR, researchers would like to report quantities of direct interest to political scientists. For instance, they might say how certain variables affect the prospects of an important party or they might make an electoral forecast: a prediction about how votes are likely to be distributed given some constellation of values for the explanatory variables. Researchers cannot obtain these quantities simply by inspecting the parameter estimates and their standard errors, however. The $\boldsymbol{\beta}$'s tell how a 1-unit change in some explanatory variable would alter the *log ratio* for a particular political party, but students of the democratic process want to know about votes, not log ratios. Fortunately, the solution is relatively mechanical. To interpret the results, choose hypothetical or real values for the $\mathbf{x}$'s and then calculate a new set of **Y**'s based on the chosen **x**'s and the parameter estimates. Finally, reverse the logit transformation, thereby converting the **Y**'s back into votes.[2]

3 How SUR Differs from KK

SUR and KK differ in their approach to partially contested districts and in assumptions about the distribution of the errors. We outline these differences below.

3.1 *Partially Contested Districts*

Political parties sometimes choose not to run everywhere. A party may compete only in the southeast, for example, or only in urban areas. When one or more parties opt not to compete in a particular district, researchers regard that district as partially contested. In England during the 1950s and 1960s, for example, the Conservative and Labour Parties battled in nearly every electoral district but the Liberals abstained in some areas.

The most common method for dealing with partially contested districts is to exclude them from analysis. Indeed, most statistics packages routinely drop districts in which the outcome or explanatory variables contain missing values. This practice, often called listwise deletion, not only discards potentially useful information but also might result in a nonrepresentative sample (King et al. 2001).

Katz and King address the problem of partial contestation by estimating the *effective* rather than the *actual* vote. The effective vote for a party is the share it would have received if all relevant parties had competed. Where all parties run the effective vote is identical to the actual vote. In partially contested districts these two quantities differ, however: the statistical insertion of a party into a race awards that party more votes than it actually earned, at the expense of parties that truly competed.

The KK procedure is appropriate for research on the effective vote, but as Katz and King stress, users should pay close attention to their assumptions. In their analysis of the British case, Katz and King presumed that a party's decision to compete in one district would not affect how well it performed in others. We caution that electoral competition is costly but party resources are finite, so what a party spends in one district must subtract from human and financial resources available elsewhere. Consequently, adding a party to one district may reduce the vote it wins in others, though the precise impact will depend on political and institutional conditions. Treating elections as "costless" may lead to overly optimistic estimates of the nationwide vote for the party that did not contest all districts and overly pessimistic estimates of the nationwide vote for parties that mustered the resources to run in all districts. These topics are worthy of future research.

We suggest a different approach to partially contested districts: run a separate analysis for each pattern of contestation. This may seem awkward, but in practice it is quite easy. To study areas of full contestation, estimate a model based exclusively on those districts. In most statistics packages this involves no more than running the model on the full data set, since the partially contested districts would be excluded automatically as missing. For inferences about partially contested districts, conduct a separate analysis on that subset of observations. Later in the paper, we show how to do this for British elections in the 1950s and the 1960s, when the Liberal Party failed to compete everywhere.

Our approach differs from KK in two respects. First, we follow most students of comparative politics by focusing on actual rather than hypothetical election results. Second, our

[2]This procedure has now been automated in *Clarify*.

approach allows the parameter estimates (and thus any quantities derived from them) to vary between fully contested and partially contested districts, whereas KK involves one set of parameters for all districts. The inclusion of partially contested districts may contaminate the estimates for fully contested districts. We present some evidence for this in our analysis of British elections.

Our approach does have certain limitations, however. Most obviously, it cannot tell what party j would have received if, contrary to fact, it had run in district i. Second, our method does not work well for complex patterns of partial contestation. In the three-party case, for instance, possible combinations of contestation include $\{1, 2, 3\}$, $\{1, 3\}$, $\{2, 3\}$, $\{1, 2\}$, $\{1\}$, $\{2\}$, $\{3\}$, and even $\{\ \}$. More generally, up to 2^J forms of partial contestation could arise (where J is the number of parties), making it inconvenient or even impossible to run a separate model for each set of outcomes. Moreover, the two-stage approach requires the researcher to estimate more parameters (one set for fully contested districts, and one or more sets for partially contested areas), thereby imposing more demands on the data.

With highly variegated patterns of contestation, researchers may prefer KK to SUR. We offer two qualifications, however. First, KK could address such cases in theory, but it would not be trivial to implement the technique in practice; researchers would need to program a likelihood function with as many as 2^J pieces. Second, our experience suggests that patterns of partial contestation among significant parties are fairly limited in most countries. In postwar England, for example, the Conservative and Labour Parties ran everywhere, while prior to the 1970s the Liberal Party ran only in selected districts. Thus, only two patterns of partial contestation emerged—{Conservative, Labour, Liberal} and {Conservative, Labour}—even though eight were theoretically possible. The situation changed dramatically in the 1970s, when all three parties fielded candidates in every district, a strategy that has continued for the last three decades.

Even in the new democracies of Eastern Europe and the former Soviet Union, full contestation is the norm for the most important political parties. For example, there were 13 national parliamentary elections in the Czech Republic, Hungary, Poland, Russia, and Slovakia between 1992 and 1999. In the national party list component of each of the elections, at least the top seven vote-getters contested all electoral districts.[3] Moreover, parties that ran everywhere captured at least 99% of the vote in 10 of the elections and more than 90% in the others. An even stronger pattern emerges when we consider presidential rather than parliamentary elections: all candidates contested all districts in the presidential elections that took place in Poland, Russia, and Slovakia during the same period (presidents are elected by parliament in the Czech Republic and Hungary). Students of minor parties may want to address issues of partial contestation, but overall the phenomenon seems less pervasive than one might suspect.

3.2 *Distributional Assumptions*

The models also use different statistical distributions: SUR employs a multivariate Normal, whereas KK relies on a multivariate t, which has fatter tails and narrower peaks. Katz and King claim that the t fits British electoral data better than the Normal. This should come as no surprise, since the Normal is the limiting case of the t with infinite degrees of freedom. However, we show in Sections 4 and 5 that the improved fit does not translate into greater

[3]The number of electoral regions per country is as follows: Russia, 79; Poland, 49; Hungary, 20; Slovakia, 38; and the Czech Republic, 76.

accuracy for several quantities of interest, such as vote shares and seat distributions in the legislature. By adopting the Normal we lose little of substantive interest but gain quite a bit in ease of use.[4]

4 Monte Carlo Comparison of SUR and KK

We designed a Monte Carlo experiment to investigate how well SUR would perform in a world tailor-made for KK. In particular, we simulated electoral data from a multivariate t, ran SUR and KK on those data, and measured how accurately each model recovered key quantities of interest. In theory, KK should have won this Monte Carlo contest, since the data were generated under the assumptions of that model. In practice, we found only minor differences in performance across the two techniques.

We constructed a three-party system in which the log ratios for the larger parties (relative to the smallest party) were linear functions of two explanatory variables and a constant term. The system contained 500 voting districts, roughly the number that exist in England. Rather than make arbitrary assumptions about the pattern of partial contestation, we assumed that all three parties ran in every district. (We defer analysis of partial contestation to Section 5, where we discuss patterns that were generated by the real world.) We sampled the log ratios from a t distribution with only five degrees of freedom, a clear deviation from normality. We then ran both SUR and KK on the simulated log ratios and calculated the expected vote shares ($\mathbf{V}$) for each district, conditional on the parameter estimates. To ensure maximum comparability with KK, we used Gauss code provided by Katz and King to estimate a multivariate t model and compute the expected votes. Details of the Monte Carlo procedure are given in the Appendix.

Although the data were t distributed by construction, SUR recovered the true $\boldsymbol{\beta}$'s with less bias than KK.[5] This may have been due to shortcomings of the programming language that Katz and King used or to the necessity, under KK, of estimating not only the effect parameters but also the degrees of freedom, ν. As Honaker et al. (2002) point out, ν is not globally concave. The challenge of estimating ν may have prevented KK from recovering all the $\boldsymbol{\beta}$'s accurately, even though the data were simulated conditional on the assumptions of that model.

Ultimately, though, we were more interested in votes than in raw statistical output. Thus, we compared the bias, variance, and mean squared error (MSE) of the vote shares that each model generated. To quantify the bias, we computed the mean of $\mathbf{V} - \hat{\mathbf{V}}$ for each district and party, where $\mathbf{V}$ represents the true vote shares. After converting these measures of bias into absolute values, we averaged across the 500 districts to obtain the mean absolute bias for each model and party. Following a similar procedure we computed the variance as $\mathrm{var}(\hat{\mathbf{V}})$ and the MSE as $(\mathbf{V} - \hat{\mathbf{V}})^2 + \mathrm{var}(\hat{\mathbf{V}})$.

Table 1A reports results for the largest and smallest parties (results for the intermediate party were like those for the large party). The table shows that both models fared well, with mean absolute biases of less than 0.25 percentage points. To our surprise, the estimated

[4]Breusch et al. (1997) reach a similar conclusion. After comparing the Normal and t regression models, they find that "although mathematically the two models are different, for purposes of statistical inference they are indistinguishable" (p. 269), such that "there is nothing gained by considering the technically more complex alternative" (p. 285).

[5]The bias with KK was 2–10 times larger than that with SUR. We list the biases for KK and SUR, respectively, for each parameter: $\boldsymbol{\beta}_{01}$, 0.018 vs. 0.003; $\boldsymbol{\beta}_{11}$, -0.007 vs. -0.002; $\boldsymbol{\beta}_{21}$, -0.006 vs. -0.003; $\boldsymbol{\beta}_{02}$, 0.008 vs. 0.001; $\boldsymbol{\beta}_{12}$, 0.008 vs. -0.0003; $\boldsymbol{\beta}_{22}$, -0.004 vs. -0.0006. Note, however, that even KK's estimates differ from the true values by no more than a few percent.

Table 1 Monte Carlo comparison of KK and SUR

(A) Vote shares

	Large party		*Small party*	
	KK	*SUR*	*KK*	*SUR*
Mean absolute bias (%)	0.24	0.14	0.13	0.04
Variance	8.58	11.16	1.91	2.40
Mean squared error	8.67	11.19	1.94	2.41

(B) Probabilities

	KK	*SUR*	*Difference*
Predict the district winner	0.95	0.94	0.01
Identify competitive districts	0.55	0.52	0.03
Identify uncompetitive districts	0.99	0.99	0.00
Identify small parties that got <5%	0.82	0.81	0.01
Identify small parties that got >5%	0.99	0.99	0.00

vote shares were slightly more biased under KK. Moreover, both models more accurately recovered the true vote shares of the small party than of the large one. The reason is simple: the votes for the small party clustered more tightly around their mean, making it easier to approximate the truth. Although the SUR estimates were slightly less biased, KK outperformed SUR in variance and, in the end, struck a somewhat better balance between bias and efficiency, as suggested by its smaller MSE. This was as expected, since the data were generated to fit the assumptions of KK.

The MSE is a common measure of dispersion around a known value, but its magnitude is hard to interpret. Are the MSE differentials in Table 1A large enough to affect substantive conclusions? To shed light on this question, we calculated a more intuitive measure: the probability that the expected vote falls within a given distance from the true value. To obtain the probability for, say, KK at the 1% range, we took 1000 KK estimates for a district and determined what percentage of them fell within plus or minus 1% of the true value. We repeated this for each of the 500 districts and averaged the results.

Figure 1 shows that SUR performed nearly as well as KK across a wide range of error tolerances for large and small parties. The dashed line (KK) typically appears above the

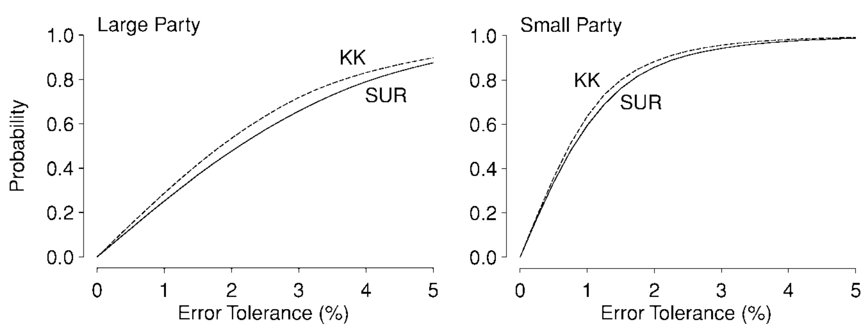

Fig. 1 Probability that the expected vote falls within the tolerable range.

solid line (SUR), indicating that the KK estimates were more likely to fall within some permissible distance from the truth. Judging from the narrow vertical gap between the lines, however, the differences were minute. In the worst case, which pertained to the large party and an error tolerance of 2.75 percentage points, the vertical distance between the lines was only 0.06, meaning that KK would outperform in only 6 of 100 random samples. For the small party, the maximum gap was only 0.04, implying that KK would beat SUR only 4 of 100 times, even if the error tolerance had been chosen to accentuate the difference in performance. It seems safe to conclude, therefore, that the larger MSE of SUR is not of great substantive importance for the t distributed data we simulated.

Figure 1 exhibits other interesting patterns. Where the error tolerance is close to 0 the models perform almost identically, since the smaller bias of SUR compensates for its larger variance. As we move toward error tolerances of a half-point or more, the smaller variance of KK means that more of the sampling distribution is closer to the true value. For even larger tolerances the two models converge once more. Convergence is especially evident for the small party, where the curves overlap once the permissible error is approximately 4%.

In Table 1B we present other quantities for which a t-based model could conceivably outperform SUR. The first row shows the probability of predicting the correct district winner. For each district, we counted the proportion of times (of 1000 expected vote shares from the Monte Carlo experiment) that each model predicted a plurality for the party that actually won. We then averaged across the 500 districts to yield the value in the table. We also computed the probability of correctly identifying a competitive district, where each party received between 30 and 40% of the true vote, and we did the same for uncompetitive districts, where at least one party fell outside the 30–40% range. Finally, we asked how often the models correctly classified cases where the small party got less than 5% of the vote, which in England would have caused it to lose its electoral deposit, as well as cases where the small party crossed the 5% threshold.

SUR performed nearly as well as KK for all five quantities, and even where the disparity was largest—in identifying competitive districts—KK did better than SUR only 3 times out of 100.[6] Table 1B also shows where the models tended to go wrong. Both models faltered when identifying competitive districts, probably because so few (only 3 of 500) existed in the simulated data. Likewise, the models had trouble classifying the nine cases in which the small party got less than 5% of the vote. SUR and KK would need to be modified to account more accurately for such rare events.

5 British Electoral Data

The previous section used simulation to compare KK and SUR. We now consider the British electoral data that Katz and King used to illustrate strengths of their model. The authors studied district-level electoral results covering 10 parliamentary elections in England between 1959 and 1992. Their data set contained 4829 observations, each representing a district in a given election year. For each election there were two equations, one modeling the log of Conservative to Liberal votes, and a second modeling the log of Labour to Liberal votes. Each equation had six explanatory variables: a constant term, lagged log ratios for the Conservative and Labour parties, and three dummy variables indicating which party (Conservative, Labour, or Liberal), if any, ran an incumbent candidate. After reproducing

[6]We thank Gary King for suggesting these quantities. We presume that there must be some cases where the t affects substantive conclusions, but we were not able to find them.

the published results of Katz and King, we applied SUR to the same data, including the same explanatory variables, and evaluated output from the two models.

5.1 *The Normal Versus the* t

We began by considering how the distributional assumptions of SUR and KK accorded with English data. Katz and King showed that, in England, the transformed votes followed a multivariate t *before* accounting for the explanatory variables. To make a stronger case for KK, one would need to prove that the *residuals* $(\mathbf{Y} - \hat{\mathbf{Y}})$, as distinct from the $\mathbf{Y}$'s, better approximate a t than a Normal. A footnote in the original article acknowledges this point: "the fit of the model could be closer to a normal after conditioning on explanatory variables" (Katz and King 1999, p. 22). Thus, we investigated whether the residuals followed a t and whether the Normal provided a reasonable approximation.

For our test, we ran SUR and calculated the residuals $\mathbf{e} = (\mathbf{e}_1, \mathbf{e}_2) = (\mathbf{Y}_1 - \hat{\mathbf{Y}}_1, \mathbf{Y}_2 - \hat{\mathbf{Y}}_2)$. Note that $\mathbf{Y}_1$ and $\mathbf{Y}_2$ were known with certainty only for fully contested districts, so our calculations pertained only to that subset of the data. Under SUR, the residuals should have followed a bivariate Normal with mean (0, 0) and variance $\hat{\boldsymbol{\Sigma}}$. Many techniques exist to evaluate the assumption of bivariate Normality; we followed Katz and King by reporting confidence region coverage. If the residuals were truly Normal, 95% of the pairs should have fallen within the 95% confidence region, 90% should have resided within the 90% region, etc. By comparing the expected versus the actual proportion of points in a given region, we judged how closely the residuals approximated the hypothesized distribution.

In the British data, the residuals were approximately Normal in some ranges but not in others. Bivariate Normal contours are ellipses of the form $\{\mathbf{e} : \mathbf{e}\boldsymbol{\Sigma}^{-1}\mathbf{e}'\}$, where the radius r is a χ^2 with 2 degrees of freedom.[7] To determine the share of points within a given confidence region, we set r equal to, say, the 95% quantile of χ^2_2 and calculated the percentage of residuals that satisfied $\mathbf{e}\boldsymbol{\Sigma}^{-1}\mathbf{e}' < r$. Across all election years, approximately 94% of the residuals fell within the 95% region and 90% landed within the 90% contours. The fit deteriorated as we moved toward smaller confidence regions. For instance, nearly 61% of the residuals fell within the 50% ellipse. These findings, which fully conditioned on the values of the explanatory variables, confirm that many English constituencies clustered more heavily around the mean than the bivariate Normal would allow.

To see whether the t fit better than the Normal, we ran KK and calculated $\mathbf{e}$. By the assumptions of the model, these residuals should have followed a bivariate t with mean (0, 0), scatter matrix $\hat{\boldsymbol{\Sigma}}$, and $\hat{\nu}$ degrees of freedom. Like their Normal counterparts, contours of the bivariate t are elliptical with the form $\{\mathbf{e} : \mathbf{e}\boldsymbol{\Sigma}^{-1}\mathbf{e}'\}$, but r is distributed as $2F_{2,\hat{\nu}}$, which is double the value of an F distribution with 2,$\hat{\nu}$ degrees of freedom (Lange et al. 1989, p. 881). Thus, we evaluated the fit by asking what percentage of the residuals satisfied $\mathbf{e}\boldsymbol{\Sigma}^{-1}\mathbf{e}' < 2F_{2,\hat{\nu}}$, the desired quantile of the F distribution. We found that the t fit the data somewhat better than the Normal, particularly for the inner region, where approximately 50% of the residuals fell within the 50% confidence region.

In summary, the t distribution was more appropriate for the data Katz and King used, even after conditioning on covariates. Nevertheless, the following sections show that KK did not perform noticeably better for quantities such as the expected vote and the percentage

[7]The value $\{\mathbf{e} : \mathbf{e}\boldsymbol{\Sigma}^{-1}\mathbf{e}'\}$, often called the Mahalanobis distance, measures the distance from the origin in standard deviations and, therefore, takes into account the variability of the residuals in all dimensions. Johnson and Wichern (1998, Chap. 4) prove that under the assumption of bivariate Normality, $\mathbf{e}\boldsymbol{\Sigma}^{-1}\mathbf{e}'$ is χ^2_2 with 2 df. Tong (1990, Chap. 4) provides a more technical treatment of elliptically contoured distributions.

of districts correctly predicted. Perhaps there are applications in which Normal and t-based models would point to substantially different conclusions (e.g., Jackman 2000), but apparently that is not the case for the data and quantities in this paper.

5.2 *Expected Vote Shares*

We next compared the estimated distribution of votes from the two models. The panels in Fig. 2 display the expected value of the vote as generated by KK (vertical axis) and SUR (horizontal axis). Each point represents one English district. To generate these panels we ran KK and SUR for each election and computed the expected vote for each party in each district, conditional on the true values of the explanatory variables. The first three panels present the vote shares for fully contested districts in all 10 elections. Diagonal lines emanate from the origins at a 45° angle. If KK and SUR had produced *identical* expected values, all points would have fallen exactly on the 45° lines. Though there is some scatter, the models made strikingly similar predictions. The correlation between the expected votes was stronger than 0.99 for all three parties. Neither a politician nor a social scientist could choose between KK and SUR based solely on the expected vote shares.

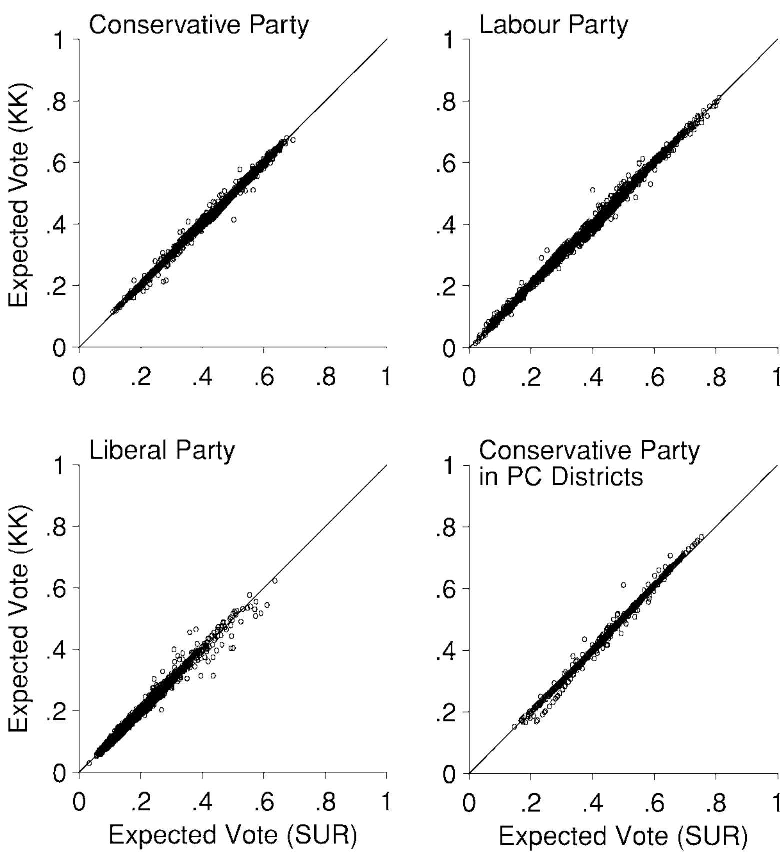

Fig. 2 Expected vote using SUR and KK. Most points fall very close to the 45° line, indicating that SUR and KK produced nearly identical estimates of the vote for every party and electoral district in England.

Katz and King gave special attention to partially contested districts, so it seemed important to measure performance in those critical cases. Our sample contained 986 partially contested districts, almost exclusively from the period before October 1974. All but two of those districts involved a bilateral contest between Labour and the Conservatives instead of the three-way race that has typified English politics since the mid-1970s. To cope with those districts, we estimated a two-party version of SUR with Labour as the reference party. Our setup amounted to a simple, single-equation linear model in which the dependent variable was the natural log of the ratio of Conservative-to-Labour votes. We did not predict any votes for the Liberals since they abstained from competition. In contrast, KK produced a hypothetical vote for the Liberals, who did not run, and compensated by taking votes from Labour and the Conservatives. Statistically speaking, their model added a third party and imputed votes for it.

To make the outputs comparable, we calculated the Conservative percentage of the two-party vote for each model and plotted the results in the lower-right panel in Fig. 2. This panel reveals that, even in partially contested districts, both models produced similar expectations about the division of support between the Conservatives and Labour. With few exceptions, the points fall near the 45° line, suggesting virtually indistinguishable results.

To the extent that the models differed, SUR's results were as accurate as KK's. We calculated the absolute error $|\mathbf{V}_i - \hat{\mathbf{V}}_i|$ for all parties, districts, and elections, and then averaged the results. Neither model erred by more than 3 percentage points for any party. In fully contested districts, for example, KK misestimated the Conservative vote by 2.10 percentage points on average, while SUR erred by 2.11 percentage points.[8] More importantly, the difference in errors between the two models was negligible, never exceeding 0.02 percentage points for any party in a fully contested district. Note that between 1955 and 1992 no English districts were decided by less than 0.02% of the vote.

For partially contested districts, SUR had a slightly smaller mean absolute error (1.69) than KK (1.80). The performance gap probably arose from KK's approach to partially contested districts. In such areas, the effective vote for all three parties deviated from the actual vote, and this may have affected the ratios as well; there was no guarantee that the Conservative share of the effective two-party vote would match the Conservative share of the actual two-party vote. Some may regard this as a virtue of KK, but our approach is probably better suited for research on the actual vote.

5.3 *Percentage Correctly Predicted*

We found little reason to prefer KK over SUR for research on the expected vote, but politicians and scholars are interested in votes mainly because they get translated into seats in the legislature and, therefore, control over public policy. Thus, we examined which model more accurately predicted who won a seat.

With British electoral data, SUR predicted the correct winner no less often than KK. For each model and election year, we calculated the expected percentage of districts in which the model correctly anticipated the winner. Our algorithm, detailed at the *Political Analysis* Web site, built upon work by Herron (1999), who proposed the concept of "expected percentage correctly predicted" (ePCP) in the context of logit models.

Figure 3 displays the ePCP for both models. The dots represent our estimates for SUR and the X's mark our estimates for KK. The vertical lines, which span the 95% confidence

[8]The errors, in percentage points, were 2.37 (KK) versus 2.35 (SUR) for Labour and 2.63 (KK) versus 2.62 (SUR) for the Liberals.

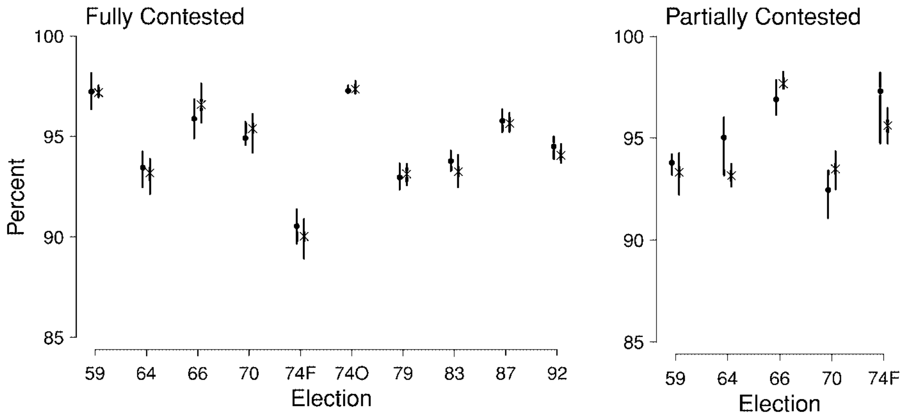

Fig. 3 Expected percentage correctly predicted. For each election, the dot represents our estimate for SUR, and the X is our estimate for KK. The vertical bars, which span the 95% confidence intervals, overlap so thoroughly that we cannot favor one model over the other with a high degree of confidence.

intervals, overlap enough in each election that we cannot favor one model over the other with a high degree of confidence. This result holds not only for fully contested districts, but also for partially contested ones, which exist in the data throughout the 1950s and 1960s but virtually disappear after February 1974.[9] Judging from the evidence in Fig. 3, researchers who want to predict the winner of the election and the composition of the parliament could choose SUR without compromising their results.

5.4 *Incumbency Advantage*

Katz and King used their model to address an ongoing scholarly debate: Do incumbents have an advantage when they run for election? They concluded that incumbency advantage in Britain is "small but meaningful, varies across parties, and is not growing" (Katz and King 1999, p. 15). In a final effort to uncover substantive differences between KK and SUR, we estimated the incumbency effect under each model. Katz and King defined incumbency advantage as the extra votes a party could expect to receive by running an incumbent instead of nominating a nonincumbent. To compute this advantage, we employed the same algorithm they did. For the Conservatives, we calculated the expected vote when a Conservative incumbent ran (other variables held at their means) and subtracted the expected vote given an open seat. We repeated this procedure for all parties and elections. Figure 4 displays the results from both models. The dots (SUR) and X's (KK) represent estimates of the incumbency advantage, while the length of the bars—spanning a 95% confidence interval—signals our uncertainty.[10] For instance, the top-left panel of Fig. 4 shows that, according to both SUR and KK, the incumbency advantage for the Conservative Party was less than 1% in 1979.

KK and SUR often differed, however, on the question of incumbency advantage. KK's estimates had relatively tight confidence intervals and hovered closer to 0, whereas SUR's

[9]There was one partially contested district in October 1974 and 10 such districts in 1979. These samples are obviously too small to allow a meaningful measure of ePCP. After 1979, there were no partially contested districts in the data.

[10]We omitted the estimate of incumbency advantage for the Liberals in 1959, since only one Liberal ran as an incumbent in that election.

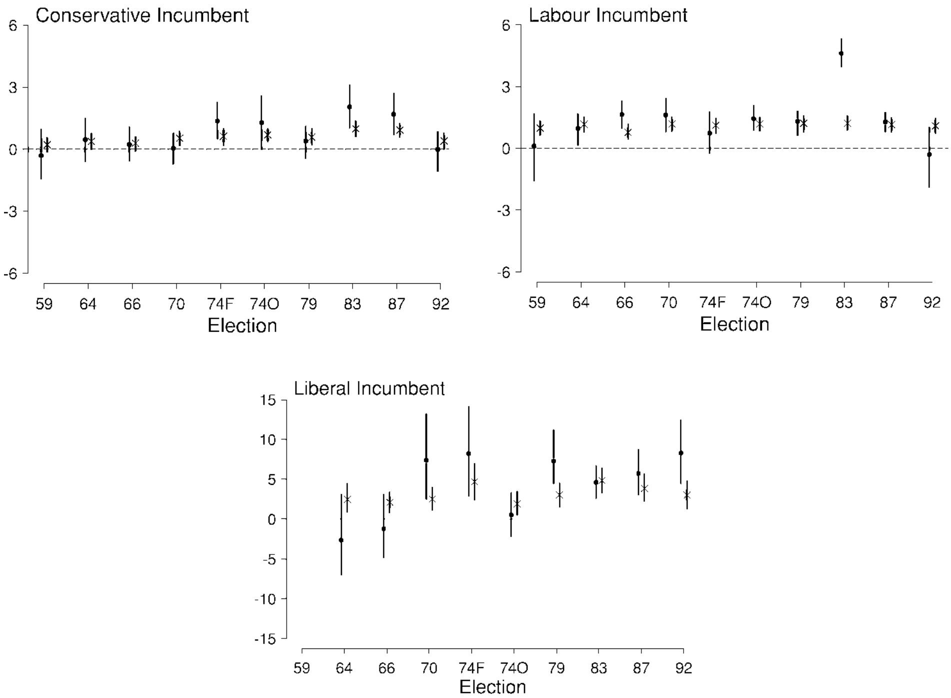

Fig. 4 Estimated incumbency advantage (%). For each election, the dot marks the incumbency advantage as estimated by SUR; the X represents the KK estimate. The vertical bars represent the 95% confidence intervals.

exhibited more variability within and across elections. The difference was most apparent for the Labour Party in 1983, where SUR found an incumbency advantage of more than 4 percentage points, a value considerably higher than in the previous or the following election. No such jump was evident in the KK estimates, which varied smoothly over time. This sharp increase in incumbency advantage deserves further investigation, given the peculiar events of 1983: the Labour party split before the election, leaving a socialist rump party to contest against the defectors (social democrats) and the Conservatives.

We traced these differences to the use of empirical Bayesian priors, which shrank the KK estimates toward the pooled average and reduced their variability. Katz and King applied priors to the incumbency dummies but not to the other explanatory variables. Since incumbency advantage had a relatively small effect on the vote, these priors did not cause the models to produce noticeably different estimates of the expected vote or the likely winner of an election. On the issue of incumbency, though, the priors exerted a clear effect. To verify that the differences between KK and SUR arose mainly from priors, rather than the t distribution or other features of the model, we reran KK without priors and found that the results matched SUR quite closely (evidence is available at the *Political Analysis* Web site). Thus, priors did indeed help Katz and King reduce the variability within and across elections. We leave the appropriateness of empirical Bayesian priors as a topic for future research.

6 The Polish Parliamentary Election of 1993

The previous sections compared SUR and KK with both simulated and real data and showed in each case that they produced virtually identical results for the most common quantities of interest to scholars and politicians. For this section we applied SUR to address a substantive problem in comparative politics: How do economic factors influence support for proreform political parties in East–Central Europe? To answer this question, we replicated and extended the work of John Gibson and Anna Cielecka (1995), who asked whether more favorable trends in unemployment, economic growth, and privatization would have increased support for proreform parties in the 1993 Polish parliamentary election. Poland was at the forefront of economic reform in the postcommunist world, and it was also among the first countries to experience the return of the former communists to power. This return provoked widespread fear that postcommunist societies had little toleration for the hardships of reform and that the former ruling parties would make a comeback everywhere. The authors asked whether a stronger economy might have prevented the former communists from regaining power. Through a reanalysis of their data, we show that the postcommunist successor party would have emerged victorious even if economic conditions had been markedly better, but it would have been in a weaker bargaining position vis-à-vis its coalition partner.

Gibson and Cielecka hypothesized that lower unemployment, faster economic growth, and more comprehensive privatization would have increased support for reformist parties to the detriment of the former communists, the Democratic Left Alliance (SLD). To test their hypothesis, they estimated seven separate OLS equations, one for each of seven major parties. The vote shares for each party served as the dependent variables and were regressed on a battery of social and economic indicators. The authors then imagined a hypothetical scenario: How would reformist parties have fared if unemployment had not increased by 2.3% shortly before the election, if yearly industrial growth had been 5 percentage points higher, and/or if 10% of the workforce had moved from the public to the private nonfarm sector (which would have occurred if more of the Polish economy had been privatized)?

Gibson and Cielecka found that if all three counterfactual changes had taken place simultaneously, the electoral gap between the SLD and the leading reform party, the Democratic Union (UD), would have shrunk from 10 to 6.8% in favor of the SLD. Thus the SLD would have beaten the UD even in the most optimistic case. The results were more favorable when reformists were considered as a bloc, however. In that case, they estimated that the vote for the reformist UD, the Non-Party Reform Bloc (BBWR), and the Congress of Liberal Democrats (KLD) taken together would have exceeded that of the SLD by 4.7%.

By focusing exclusively on predicted vote shares, the paper left an equally interesting question unasked: How would the composition of parliament have looked if economic conditions had been better? Would the reformist parties have gained the upper hand? Table 2 reports our estimates of the number of seats that each party would have won if unemployment, economic growth, and privatization had all changed according to the same "what if" scenario. To generate these estimates, we first reran Gibson and Cielecka's analysis using SUR instead of OLS.[11] We then obtained expected vote shares for each party in each district, conditional on the counterfactual scenario. Finally, we converted the votes into seats, according to Poland's electoral law.[12] Since our estimate of each party's vote was uncertain, so too was our estimate of their seats in the postelection parliament. Thus, we calculated 95% confidence intervals around our estimated seat totals, which are also listed in Table 2.

[11] We augmented the data from Gibson and Cielecka by including seat totals from Florczyka (1993). We also corrected a few minor errors.

[12] Poland used the d'Hont method with thresholds for distributing seats at both the regional and the national level in 1993. The algorithm we used to convert votes into seats is available at the *Political Analysis* Web site.

Table 2 Estimated composition of the Polish parliament following the 1993 elections with shifts in unemployment, growth, and labor market composition

	Actual outcome		*Counterfactual*	
	Vote (%)	*Seats*	*Seats*	*95% CI*
Parties				
SLD	20.4	171	161	145, 178
PSL	15.4	132	138	135, 142
UD	10.6	74	87	77, 94
UP	7.3	41	41	33, 44
KPN	5.8	22	0	0, 0
BBWR	5.4	16	28	22, 43
KLD	4.0	0	0	0, 0
Gaps				
SLD–UD	9.8	97	74	51, 101
SLD–Reformists	0.4	81	45	9, 78
Possible coalitions				
SLD + PSL	35.8	303	299	282, 314
Reformists	20.0	90	115	100, 137
Reformists + PSL	35.4	222	254	236, 276

Note. Reformists = UD, BBWR, and KLD. Source for percentage of vote and seat totals: Florczyka (1993).

Table 2 confirms the authors' first conclusion: shifting economic conditions in the way they posit would have decreased the popularity of the SLD and increased the vote for the UD but in no way would have closed the gap between the two. Even considerably better economic circumstances would have left the UD trailing the SLD by 74 seats in the postelection parliament. This would have represented an improvement over the actual gap of 97 seats, but it still would have left the SLD with a commanding lead. Thus, even a sizable increase in growth, a reduction in unemployment, and more widespread privatization would not have propelled the UD past the SLD.

Surprisingly, Table 2 points to a similar conclusion about competition between the SLD and the reformist parties (UD, BBWR, and KLD) as a group. Gibson and Cielecka predicted that the reformists would have captured 4.7% more of the vote than the SLD under their counterfactual economic scenario. Based on their analysis, we would have expected the reformists to receive more seats than the SLD, thereby putting them in a position to form a government. But when votes are translated into seats, we see that this is not the case: the SLD still would have possessed 45 more seats than the reformist bloc. A rosier economy would have benefited the reformists, to be sure, but even together they would not have overtaken the SLD, nor could they have formed a reformist government alone.

Our most counterintuitive finding, however, is that the real "winner" from an improvement in economic conditions would have been the Polish Peasant Party (PSL). The party drew most of its support from farmers and is generally regarded as a successor to a former satellite party from the Communist period, so it would appear to be an unlikely beneficiary of postreform improvements in economic conditions. Indirectly, though, an economic upturn would have converted the PSL into the pivotal party in Polish politics. Following the actual election, the SLD formed a coalition with the PSL. Together they held 303 seats, well above the 231 necessary to form a majority. Equally importantly, the reformists and the PSL held only 222 seats and, thus, could not have formed a government. Under the hypothetical scenario, in contrast, the reformists could have gained a majority by persuading the PSL to join their coalition. (A similar coalition had actually been in power 3 years earlier). Together the UD, the BBWR, and the PSL would have controlled 254 seats in the legislature, 32 more than they obtained in the real election. Ironically, then, the PSL, which itself would have picked up only six seats under the hypothesized scenario, could nonetheless have become a "kingmaker" in postelection coalition negotiations had the economic scenario come to fruition: both the SLD and the reform bloc could have formed a ruling coalition with the PSL, and neither could have governed without it!

7 Conclusion

In this paper, we offer a model of multiparty electoral data that performs as well as KK across a broad range of indicators and is nearly as easy to use as OLS. We show that SUR matches KK in producing accurate estimates of the expected vote, the composition of parliament, and several other quantities of substantive interest to politicians and scholars. Our results hold not only in the pristine world of Monte Carlo studies, but also in the messy reality of real election data. Given how easy it is to estimate SUR with widely available software, and how readily the model scales to an arbitrarily large number of parties, the attractions of SUR are clear. Of course, KK has its own attractions, particularly for work on the effective vote. Researchers are now in the fortunate position of being able to choose among several models, each with its own strengths and weaknesses.

In this issue Honaker et al. present another technique (HKK) that researchers may want to consider. They propose a t-based regression that can be approximated by running weighted least squares on multiply imputed data sets. Like our method, HKK is faster than the original KK procedure and handles an arbitrarily large number of political parties. Nevertheless, the approach requires researchers to generate multiply imputed data sets, understand t-based regression models and implement weighting procedures to calculate them, and then combine and interpret the statistical output from several imputed data sets. For certain types of research, especially concerning the effective vote, KK and HKK may be optimal. For other types of inquiries, researchers could do as well with far less effort by running SUR.[13]

Although this paper has concentrated on multiparty elections, SUR can help researchers study other kinds of compositional data. Indeed, the model is appropriate for any substantive problem in which the outcome variables, like proportions of a pie, are nonnegative and sum to unity. Budget allocations, trade flows, and the partisan makeup of legislatures all exhibit compositional characteristics. Researchers now have a convenient tool to shed light on these important topics.

Appendix

Details of the Monte Carlo Procedure

Our Monte Carlo experiment involved two steps. First, we generated true vote shares for each of 500 electoral districts. To do this we drew two explanatory variables ($\mathbf{x}_1$ and $\mathbf{x}_2$) from a standard Normal distribution and then calculated $E(\mathbf{Y}_1) = \boldsymbol{\mu}_1 = \boldsymbol{\beta}_{01} + \mathbf{x}_1\boldsymbol{\beta}_{11} + \mathbf{x}_2\boldsymbol{\beta}_{21}$ and $E(\mathbf{Y}_2) = \boldsymbol{\mu}_2 = \boldsymbol{\beta}_{02} + \mathbf{x}_1\boldsymbol{\beta}_{12} + \mathbf{x}_2\boldsymbol{\beta}_{22}$ using $\mathbf{x}$'s unique to each district and $\boldsymbol{\beta}$'s

[13] It would also be interesting to compare KK with other variants of SUR. We leave that for future research.

whose values remained constant across districts. We set values of the $\boldsymbol{\beta}$'s that, after taking the inverse logit transformation of the $\boldsymbol{\mu}$'s, would yield reasonable proportions for the votes. In particular, we set $\beta_{01} = 0.2$, $\beta_{11} = 0.2$, $\beta_{21} = 0.5$, $\beta_{02} = 0.8$, $\beta_{12} = -0.7$, and $\beta_{22} = 0.5$. In our case, the largest party received 45% of the vote on average, with a minimum showing of 10% and a maximum of 82%. The intermediate party captured 36% of the vote (with 5 and 82% as lower and upper bounds), while the minor party claimed only 19% of the vote, though its showing in any particular district ranged from 4 to 49%. Each district had a true pair of $\boldsymbol{\mu}$'s and a corresponding three-tuple of true vote shares, $\mathbf{V} = (\mathbf{V}_1, \mathbf{V}_2, \mathbf{V}_3)$. The goal of the experiment was to see how well SUR and KK recovered the truth.

Second, we sampled the $\mathbf{Y}$'s from a multivariate t with mean $\boldsymbol{\mu} = (\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)$. To take these draws, we fixed the degrees of freedom (ν) and the 2×2 symmetric variance matrix at values close to what KK estimated, based on the British data. Specifically, we set $\nu = 5$, the variance for party 1 equal to 0.06, the variance for party 2 equal to 0.08, and the covariance at 0.06. With $\nu = 5$, the data were clearly t rather than Normal. We then ran both SUR and KK on the log ratios.

By repeating the second step 1000 times, we built up sampling distributions of the point estimates $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\mu}}$, and $\hat{\boldsymbol{V}}$ for both models. We then compared the estimates with their true values to see how well the models fared in a world designed to favor KK.

References

Breusch, T. S., J. C. Robertson, and A. H. Welsh. 1997. "The Emperor's New Clothes: A Critique of the Multivariate *t* Regression Model." *Statistica Neerlandica* 51(3):269–286.

Cox, Gary W. 1997. *Making Votes Count: Strategic Coordination in the World's Electoral Systems*. Cambridge: Cambridge University Press.

Florczyka, Andrzej (ed). 1993. *Wybory Do Sejmu Rzeczpospolitej Polskiej*. Warsaw: Panstowa Komisja Wyborcza.

Gibson, John, and Anna Cielecka. 1995. "Economic Influences on the Political Support for Market Reforms in Post-Communist Transitions: Some Evidence from the 1993 Polish Parliamentary Elections." *Europe-Asia Studies* 47(5):765–785.

Greene, William H. 2000. *Econometric Analysis*, 4th ed. Upper Saddle River, NJ: Prentice Hall.

Herron, Michael C. 1999. "Postestimation Uncertainty in Limited Dependent Variable Models." *Political Analysis* 8:83–98.

Honaker, James, Jonathan N. Katz, and Gary King. 2002. "A Fast, Easy, and Efficient Estimator for Multiparty Electoral Data." *Political Analysis* 10:84–100.

Jackman, Simon. 2000. "Estimation and Inference are Missing Data Problems: Unifying Social Science Statistics via Bayesian Simulation." *Political Analysis* 8:307–332.

Jackson, John E. 2002. "A Seemingly Unrelated Regression Model for Analyzing Multiparty Elections." *Political Analysis* 10:49–65.

Jackson, John E., Jacek Klich, and Krystyna Poznánska. 1998. "Democratic Institutions and Economic Reform: The Polish Case," Working paper, Ann Arbor: University of Michigan.

Johnson, Richard A., and Dean W. Wichern. 1998. *Applied Multivariate Statistical Analysis*, 4th ed. Upper Saddle River, NJ: Prentice Hall.

Katz, Jonathan, and Gary King. 1999. "A Statistical Model for Multiparty Electoral Data." *American Political Science Review* 93(1):15–32.

King, Gary, Michael Tomz, and Jason Wittenberg. 2000. "Making the Most of Statistical Analyses: Improving Interpretation and Presentation." *American Journal of Political Science* 44(2):347–361.

King, Gary, James Honaker, Anne Joseph, and Kenneth Scheve. 2001. "Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation." *American Political Science Review* 95(1): 49–69.

Lange, Kenneth L., Roderick J. A. Little, and Jeremy M. G. Taylor. 1989. "Robust Statistical Modeling Using the *t* Distribution." *Journal of the American Statistical Association* 84(408):881–896.

Mikhailov, Nikolai, Richard G. Niemi, and David W. Weimer. 2002. "Application of Theil Group Logit Methods to District-Level Vote Shares: Tests of Prospective and Retrospective Voting in the 1991, 1993, and 1997 Polish Elections." *Electoral Studies* (in press).

Theil, Henri. 1970. "On the Estimation of Relationships Involving Qualitative Variables." *American Journal of Sociology* 76(1):103–154.

Tomz, Michael, Jason Wittenberg, and Gary King. 2001. *CLARIFY: Software for Interpreting and Presenting Statistical Results, Version 2.0*. Cambridge, MA: Harvard University, June 1 (http://gking.harvard.edu/).

Tong, Y. L. 1990. *The Multivariate Normal Distribution*. New York: Springer-Verlag.

Zellner, Arnold. 1962. "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias." *Journal of the American Statistical Association* 57(298):348–368.