State Politics & Policy Quarterly
11(4) 371–389

DOI: 10.1177/1532440011408929
http://sppq.sagepub.com
SAGE

Linear Models with Outliers: Choosing between Conditional-Mean and Conditional-Median Methods

Jeffrey J. Harden[1] and Bruce A. Desmarais[2]

Abstract

State politics researchers commonly employ ordinary least squares (OLS) regression or one of its variants to test linear hypotheses. However, OLS is easily influenced by outliers and thus can produce misleading results when the error term distribution has heavy tails. Here we demonstrate that median regression (MR), an alternative to OLS that conditions the median of the dependent variable (rather than the mean) on the independent variables, can be a solution to this problem. Then we propose and validate a hypothesis test that applied researchers can use to select between OLS and MR in a given sample of data. Finally, we present two examples from state politics research in which (1) the test selects MR over OLS and (2) differences in results between the two methods could lead to different substantive inferences. We conclude that MR and the test we propose can improve linear models in state politics research.

Keywords

outliers, linear models, ordinary least squares, median regression, cross-validation

Research in state politics, and political science in general, has relied on the linear regression model to test key hypotheses for several decades. By estimating the linear additive relationship between several independent variables and a measure of the

[1]University of North Carolina at Chapel Hill, USA
[2]University of Massachusetts–Amherst, USA

Corresponding Author:
Jeffrey J. Harden, University of North Carolina at Chapel Hill, Department of Political Science, 312 Hamilton Hall, CB #3265, Chapel Hill, NC 27599
Email: jjharden@unc.edu

center of the distribution of a dependent variable, the linear model is substantively intuitive and permits a straightforward interpretation of results. Although careful theoretical specification of the variables to include in the analysis is the most critical component in model construction, another important step is to select an appropriate method to estimate the parameters of the model. Linear regression by ordinary least squares (OLS), which conditions the mean of the dependent variable on the independent variables, is the most common choice at this step in the process in political science (Krueger and Lewis-Beck 2008). Analysts are often justified in this choice because OLS holds many desirable statistical properties; for instance, it is the best linear unbiased estimator (BLUE) when the Gauss–Markov assumptions are met and the minimum variance unbiased estimator (MVUE) when the error term is distributed normally (Gujarati 2003, 65–68). However, heavy-tailed error distributions can reduce OLS efficiency by masking the true systematic behavior with extreme and atypical random variation, or outliers. This is problematic because inefficiency can produce misleading estimates in a single sample of data, even when the estimator is unbiased. Thus, researchers using OLS with a heavy-tailed error term are more likely to draw inferences that are not warranted by the data.

In this article, we make two contributions in response to this problem. First, we demonstrate that another estimator, median regression (MR), which has been popular in several disciplines since the late 1970s, is a more efficient choice when the error term has heavy tails. Second, we provide a model comparison test that greatly enhances the utility of MR in applied research. Though MR often gives different results than OLS, there is currently no method for determining which set of estimates should be considered more reliable. Below we provide a detailed review of MR as a robust alternative to OLS and introduce a new test that applied researchers can use to select between OLS and MR in a given sample of data.

As any statistics textbook explains, the sample median serves as a more robust measure of central tendency than the mean when outlying observations are present. Similarly, MR, which conditions the median of the dependent variable on the independent variables, can serve as an alternative to the conditional-mean framework of OLS under these conditions.[1] By minimizing the sum of absolute residuals rather than squared residuals like OLS, MR is not disproportionately influenced by outliers.[2] As a result, MR is more efficient than OLS and allows the analyst to handle heavy-tailed data without giving special treatment to outliers.[3] Instead of setting arbitrary criteria from which to define and delete observations (e.g., Studentized residuals or Cook's D; Cook 1977), including indicator variables for outlying observations, or transforming variables to accommodate the assumptions of OLS, MR permits the model to be informed entirely by theory.[4]

We present three main points as evidence of the utility of MR in state politics research. First, we explain that MR is a more efficient estimator of the linear model than OLS when the error term is drawn from a heavy-tailed distribution, and that this gain in relative efficiency holds when analyzing both small and large samples of data.[5] Second, we propose the use of a new hypothesis test, called the cross-validated difference in means (CVDM) test, as a method for choosing between OLS and MR in a sample of data.[6] Finally, we show that heavy-tailed data are common in state politics research and that differences between OLS and MR can lead to different substantive conclusions. Because MR offers an improvement to OLS under these common conditions while still modeling a similar quantity, we conclude that the method and the CVDM test can be useful tools for scholars of state politics.

Robust Regression Methods

The search for approaches to data analysis that are robust to the presence of outliers, including linear regression, dates back over two centuries (Western 1995). The major hindrance to their widespread use was the nonlinearity of most robust estimators, which makes their implementation more computationally demanding than OLS.[7] However, modern computing power eliminates this concern, and so researchers can benefit from the fact that estimators such as MR can be more efficient than OLS even when OLS is BLUE.[8] This leads to the question of which robust estimator a researcher should use in an applied analysis.

Why MR?

In many cases where the Gauss–Markov assumptions are met, several more efficient (albeit nonlinear) unbiased estimators of the linear model can be identified. These methods include *L*p regression, of which OLS and MR are special cases (Lai and Lee 2005), regression with Student's *t* distributed errors (Lange, Little, and Taylor 1989), regression with Cauchy distributed errors (Morgenthaler 1994), and M-estimation (P. Huber 1964; Western 1995), of which OLS and MR are also special cases.[9] All of these methods constitute nonlinear estimators that are designed to be more efficient than OLS when the distribution of the error term is heavy tailed.

Though we cannot claim it is the "best" option for all cases in which the error term has heavy tails, we focus exclusively on MR for three reasons. First, the median, like the mean, is an intuitive measure of central tendency, which leads to a straightforward interpretation of the conditioning effect of independent variables on a dependent variable with MR. None of the other robust methods mentioned above are derived as estimators of relationships between independent variables and common measures of central tendency of the dependent variable.[10] This means that MR is more similar to OLS with respect to the quantity it models. Thus, virtually any set of hypotheses testable with OLS is also testable with MR. Second, software routines for performing MR, exactly as we describe it, are available in most statistical analysis packages used by political scientists. Some of the other methods described above are more difficult to implement. Third, like OLS with a normally distributed error term, MR is the maximum likelihood (ML) estimator when the residuals are drawn from a Laplace distribution. The fact that MR is firmly within the likelihood framework allows the analyst to use the familiar information-theoretic and likelihood-based methods of model comparison and selection (e.g., Akaike information criterion, Bayesian information criterion, or likelihood ratio tests). This contrasts with some of the other methods listed above, which have fewer and less intuitive options for assessing goodness of fit and conducting model comparisons.

Regression through the Conditional Median

Our first main objective is to briefly summarize the intuition behind the MR method of model estimation. For more detailed discussions of MR, see Koenker (2005) or Hao and Naiman (2007). Because the two techniques are similar in many respects, MR can be best understood in comparison to OLS. In an OLS model, the conditional mean of the dependent variable (y) is theorized to be a linear combination of k independent variables (x). Similar logic applies in a MR model, except that the median of the dependent variable is conditioned on x. Like OLS, a MR model takes on the familiar regression form, in which y is equal to a weighted combination of the independent variables plus an error term $y_i = x_i'\beta + \varepsilon_i$. In OLS ε_i is the deviation from the mean, and in MR it is the deviation from the median. This produces residuals as the predicted values subtracted from the actual values, or $y_i - \hat{y}_i$.

The key difference between the two methods is the way in which each solves for β. OLS minimizes the sum of the squared residuals: $\min \Sigma(y_i - \hat{y}_i)^2$. In contrast, MR minimizes the sum of the *absolute deviations*. Instead of squaring the residuals, the method minimizes the sum of their absolute values: $\min \Sigma |y_i - \hat{y}_i|$. Through one of two methods, this approach fits the regression line such that the number of points above the line is equal to the number below the line.[11] Although this may seem intuitive, the primary reason for the dominance of the squared error criterion is that the OLS solution is available in closed (linear) form whereas the least absolute deviations criterion must be optimized iteratively. In fact, the least absolute deviations criterion actually appeared in published work in 1755 (Maire and Boscovich 1755), a half century before the first published work on OLS (Legendre 1805).

OLS and MR can both be defined as solutions to a minimization problem. However, they differ markedly in how they are affected by deviations from central tendency. The OLS estimator responds exponentially to deviations from the conditional mean whereas MR responds linearly to deviations from the conditional median. In other words, OLS is heavily influenced by outlying observations, but MR is not. Thus, MR is a better, more efficient estimator if large, infrequent deviations from central tendency (outliers) are a result of stochastic variation (and thus should be treated as random errors) and not systematic variation that should draw the regression line away from regularities in the data. Formally, as Bassett and Koenker (1978, 621) prove, "[MR is] more efficient than [OLS] for any error distribution for which the median is more efficient than the mean." Of course, a researcher never knows the true distribution of the error term in an applied example and therefore cannot determine a priori whether OLS or MR is appropriate. However, we show below that it is simple to test whether MR or OLS is more appropriate for the data on hand.

A Test for Choosing between OLS and MR

Though MR is similar in approach to OLS and easy to implement, it has failed to become popular in political science likely because there is little guidance for researchers as to when to use a method other than OLS to estimate a linear model. In this section we present the CVDM test as a simple method for making this choice. We provide a brief description of the intuition behind the test, give its formal definition, and assess its performance in a simulation study. In the online appendix (available at http://sppq.sagepub.com) we provide more details on the test's statistical properties.

OLS and MR with ML

Before describing the test, note that it requires OLS and MR to be defined as ML estimators.[12] OLS and MR estimates can be obtained through ML with a normal and Laplace likelihood function, respectively. These functions are maximized to arrive at the coefficient estimates. Even if the error term is neither normally nor Laplace distributed, the fact that both can be expressed as ML estimators plays an important role in determining whether the conditional-mean or conditional-median model is closer to the true data generating process.

A Brief Description of the CVDM Test

A likelihood function can be used to compute the likelihood of having observed the data on hand, given that the data were drawn from the model corresponding to that likelihood function and vector of parameters. Moreover, when the observations are assumed to be independent, the likelihood function can be used to compute the likelihood of observing any particular data point (an individual likelihood). One approach to selecting between two models—which we justify in the online appendix—is to consider the model corresponding to the higher average value of the individual likelihood (or log likelihood) function to be closer to the data generating process since it is more likely to have generated the data. The CVDM test takes this one step further by using *cross-validated* log likelihood (CVLL) values. Specifically, the likelihood for a single observation is calculated from a model estimated on all of the other observations in the data set. This "leave-one-out" procedure is done for each observation, making the test an out-of-sample measure of fit. This protects the analyst against overfitting the model, which can otherwise produce bias in the test's selection (Desmarais and Harden n.d.; Smyth 2000).

More formally, let $\delta^{(cv)}$ be an N-length vector of differences in individual OLS CVLLs and individual MR CVLLs and $L_{\mathrm{OLS}}(.)$ and $L_{\mathrm{MR}}(.)$ be the OLS and MR likelihood functions. Then,

$$\delta_i^{(cv)} = L_{\mathrm{OLS}}(y_i | x_i, \hat{\beta}_{-i}) - L_{\mathrm{MR}}(y_i | x_i, \tilde{\beta}_{-i}), \tag{1}$$

where the $-i$ subscripts on $\hat{\beta}_{-i}$ and $\tilde{\beta}_{-i}$ indicate that the ith observation is excluded from the sample used to estimate the regression coefficients. The CVDM test is a test of the null hypothesis that $E[\delta^{(cv)}] = 0$. This amounts to a difference in means test. If the mean CVLL for OLS is statistically significantly larger than the mean CVLL for MR, then the data are more likely to have come from the normal likelihood model, and OLS is selected as the better fitting estimator. The reverse is true if the average MR CVLL is significantly larger than the average OLS CVLL. The starting point for the CVDM test is a conventional t-test applied to $\delta^{(cv)}$. Because it is possible that there is significant skew in the distribution of $\delta^{(cv)}$, the t-statistic is adjusted for skewness using the procedure suggested by N. Johnson (1978). Specifically, let $\bar{\delta}^{(cv)}$ be the sample mean of $\delta^{(cv)}$. Then the unbiased estimator of the skewness of $\delta^{(cv)}$ is $\hat{\mu}^3 = n(n-1)^{-1}(n-2)^{-1}\sum_{i=1}^{n}(\delta_i^{(cv)} - \bar{\delta}^{(cv)})^3$. The CVDM test statistic is,

$$\text{CVDM} = \left[\ \bar{\delta}^{(cv)} + \frac{\hat{\mu}^3}{6s^2 n} + \frac{\hat{\mu}^3}{3s^4}\ (\bar{\delta}^{(cv)})^2\ \right]\frac{s}{\sqrt{n}}\,, \tag{A.3}$$

where s is the conventional estimator of the standard deviation of $\delta^{(cv)}$. To derive p values, the CVDM statistic is evaluated with respect to a Student's t distribution with $n - 1$ degrees of freedom. See the online appendix for more details on the CVDM test.

Simulation Study

Before implementing the CVDM test in applied research, it is necessary to assess whether the better fitting estimator as selected by the CVDM test is also the estimator that provides the more efficient estimates. To this end, we present results from a simulation study in which we created a large number of data sets favoring either OLS or MR (to varying degrees) and recorded the test's selection each time. Technically speaking, the CVDM test is designed to make an unbiased selection of the model with the lowest Kullback–Leibler divergence from the true model (see the online appendix). However, to facilitate ease of interpretation, in this simulation study we measure estimator performance with mean squared error (MSE).[13] If the CVDM test selects the estimator with lower MSE in simulated data, then we can be more confident in its ability to select the estimator that produces coefficients closer to the true parameters in real data.

Specifically, we generated a variable y dependent on a single predictor variable x distributed standard normal. We manipulated the MSE of the two estimators by drawing the error term from an exponential power (EP) distribution (see Mineo and Ruggieri 2005). With the EP distribution, the parameter p determines the relative MSE of OLS to MR. When $p = 1$, the EP distribution is equivalent to the Laplace, and when $p = 2$ it is equivalent to the normal. This means that MR is the better choice (lower MSE) when $p = 1$, and, as p increases, MR MSE increases while OLS MSE decreases. In addition, this design allows us to assess the performance of the CVDM test even when the distribution is neither normal nor Laplace (i.e., when $1 < p < 2$; see Mineo and Ruggieri 2005). This is important because applied researchers never know the true distribution of their data. Thus, it is crucial that the test is not adversely affected by distributional misspecification.[14]

We conducted the simulations for 50 values of p ranging from 1 to 2. At each value of p, we created 1,000 data sets, then estimated OLS and MR models and conducted the CVDM test in each one. Then we calculated the OLS and MR MSE as well as the number of times the test selected OLS across those 1,000 data sets. We did this for four different sample sizes, $N = 50$, $N = 100$, $N = 500$, and $N = 1{,}000$, to assess whether CVDM test performance improves with more observations.[15] This produced a total of 50 values of $p \times 1{,}000$ data sets $\times$ 4 sample sizes = 200,000 simulations.

Figure 1 displays results from the simulations, with panels a through d showing output from each sample size. Each point on the graph represents one set of 1,000 simulated data sets. The x-axes plot the relative MSE of OLS to MR. When this is equal to 1, OLS and MR efficiency is equal. When it is less than 1, OLS has smaller MSE than MR, and when it is larger than 1, MR has smaller MSE. The y-axes plot the percentage of times in those 1,000 data sets that the CVDM test selected OLS. If the test performed perfectly, the points should fall at 100% on the y-axis when relative MSE is less than 1 (i.e., always select OLS when OLS MSE is smaller than MR MSE) and 0% when relative MSE is larger than 1 (i.e., never select OLS when OLS MSE is larger). Gray sections of the graphs indicate areas in which the test makes an incorrect selection more than 50% of the time.

Three main results emerge from these simulations. First, overall test performance is quite strong. In general, the test selects the estimator with lower MSE much more often than not.[16] Second, test performance improves as the sample size increases. Although it still performs reasonably well at $N = 50$, there are fewer points in the gray regions in each subsequent graph. Finally, note that most of the cases in which the test chooses the incorrect estimator more than 50% of the time are also the cases in which relative MSE is close to one. When MSE is close to 1, OLS and MR estimates are very similar, and thus less likely to show meaningful substantive differences. Thus, the CVDM test is most likely to make correct selections at the most critical situation for applied researchers—when differences between the two estimators are large. In sum, these simulations show that the CVDM test is effective in choosing between OLS and MR, and improves with increases in sample size and as differences between the two estimators increase. Thus, we recommend its use to applied researchers. However, as with any statistical test, we recommend caution when using the CVDM test with a small sample size.[17]

Applying MR to State Politics Research

Having shown both the intuition behind MR as an alternative to OLS and presented the CVDM test as an effective method for choosing the more efficient of the two estimators, our next objective is to consider the applicability of MR to state politics research. MR can be a useful tool in state politics because several studies in the literature utilize data with heavy tails.[18] Examples include interest group population

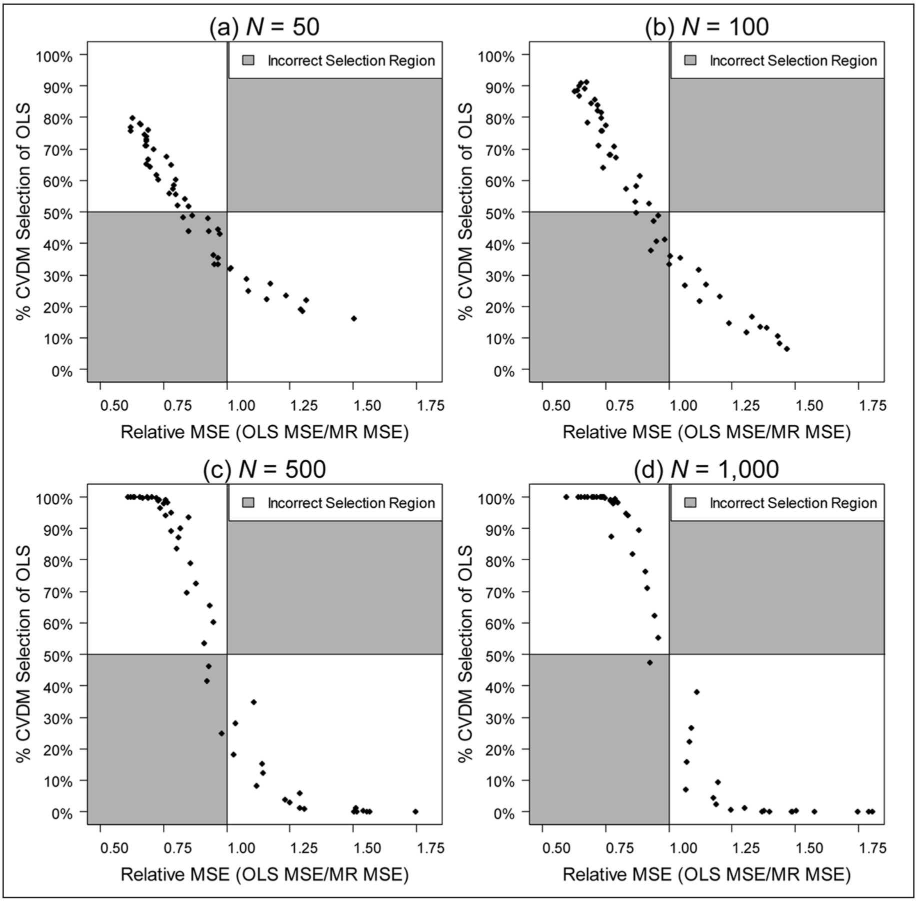

Figure 1. CVDM test performance in the simulations
Note: CVDM = cross-validated difference in means; MSE = mean squared error; OLS = ordinary least squares; MR = median regression. The graphs report simulation results at four different sample sizes, with each point summarizing 1,000 simulated data sets. The x-axes plot the relative MSE of OLS to MR. When this is equal to 1, OLS and MR performance is equal. When it is less than 1, OLS has smaller MSE than MR, and when it is larger than 1, MR has smaller MSE. The y-axes plot the percentage of times the CVDM test selected OLS. Note that (1) the CVDM test performs well, selecting the estimator with lower MSE most of the time, and (2) that performance improves as *N* increases.

density (Lowery and Gray 1995; Wolak, Lowery, and Gray 2001), state legislative control of bureaucracy (J. Huber, Shipan, and Pfahler 2001), margin of victory in U.S. congressional elections (Leighton and Lopez 2002), and reorganizations of U.S. state Executive branches (Berkman and Reenock 2004). In addition, as we show below, MR is important for state politics researchers to consider because differences between OLS and MR may lead to different conclusions. Thus, selecting the more appropriate estimator is crucial for securing accurate inferences from model results.

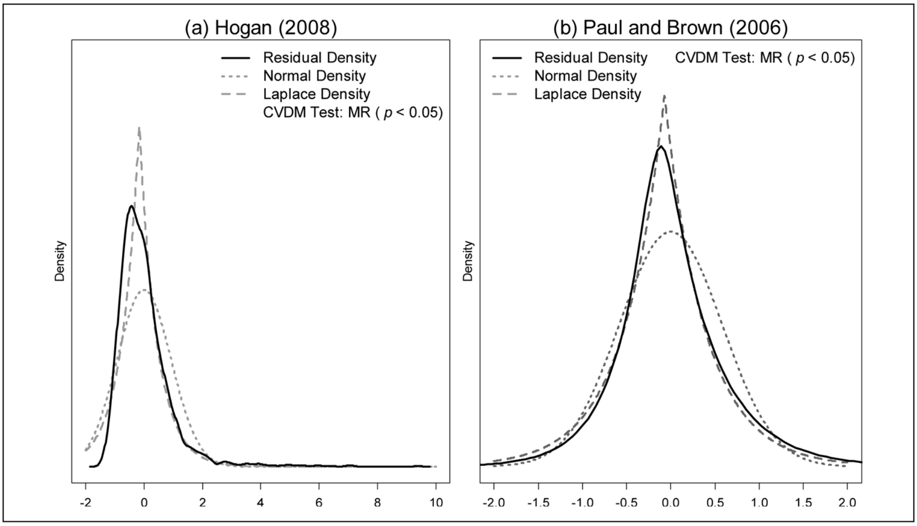

Figure 2. OLS residual densities and parameter-matched normal and Laplace densities for the replication examples

Note: OLS = ordinary least squares; MR = median regression. Panels (a) and (b), respectively, plot density estimates of the Hogan (2008) and Paul and Brown (2006) OLS residuals (solid lines), along with parameter-matched normal densities (dotted lines) and Laplace densities (dashed lines). Note that in each case the residual density looks more like the Laplace than the normal. This is informal evidence that MR is likely a better estimator for each model.

As evidence of this claim, we report replication results of two recent articles from the state politics literature that use OLS to test linear hypotheses. To illustrate the applicability of MR at different sample sizes, we show one example with a large sample size and one with a small sample. Hogan (2008) looks at the influence of state legislative incumbent partisan voting behavior on 1,816 challengers' fund-raising success, and Paul and Brown (2006) examine the role of elites in 41 referendum votes on building sports stadiums. The dependent variables are, respectively, challenger spending as a percentage of incumbent spending in 1996 and 1998 and the vote percentage in favor of stadium referendums from 1984 to 2001.[19] In each case, the authors' original estimation procedure is OLS, but the CVDM test selects MR as the more efficient estimator at statistically significant levels ($p < .05$).[20] Furthermore, these particular models show that the differences between OLS and MR results—in coefficient magnitude, standard error size, or both—could potentially lead to different inferences.[21]

Before detailing these key differences, we begin our replication analysis with a graphical illustration of why MR is a likely better choice for these data sets. Figure 2 plots a density estimate of each model's OLS residuals (solid lines) along with a normal density (dotted lines) and a Laplace density (dashed lines) with the same parameterizations. This is done to provide informal evidence that the error terms have heavy tails. Note that in each case the density estimate looks more like the Laplace density than the normal density. To be sure, these plots are not substitutes for the CVDM test. Nonetheless, they provide a more intuitive, visual depiction in favor of MR over OLS.

Partisan Voting and Challenger Fund-Raising in State Legislatures

Hogan (2008) examines how the voting behavior of state legislators can influence their chances of reelection. He looks at this process in three areas: the decision of challengers to run against incumbents, challenger fund-raising success, and votes received by challengers and incumbents. The goal is to test whether the classic spatial proximity model (e.g., Downs 1957) holds in the low-information environment of state legislative elections. While the Downsian model predicts that an extreme incumbent should suffer from voting away from the district median, the fact that few citizens are knowledgeable about state legislative candidates may curtail that relationship. In such a case, partisan voting may actually serve to stimulate the incumbent's core supporters, resulting in electoral benefits.

The data come from incumbents in both the lower and upper houses of 14 state legislatures in 1996 and 1998. Here we focus on the OLS model predicting challenger spending as a percentage of incumbent spending. The main independent variable, *partisan policy position*, is a measure of incumbent divergence from expected district preferences on economic and regulatory policy, with large values representing legislators whose voting records are strongly divergent from district preferences. Building from the Downsian framework, Hogan hypothesizes that greater divergence will correspond to increased spending by challengers—an incumbent out of touch with his or her district will draw more significant opposition with respect to fund-raising. This leads to the expectation of a positive coefficient on partisan policy position.

Panel a of Figure 2 shows that the peak of the normal density falls slightly to the right, while the Laplace density peak does not. In addition, the shoulders and tails of the residual density map more closely to the Laplace than the normal. In short, there is good informal evidence that MR is preferred over OLS. This is further validated by the CVDM test, which selects MR at a statistically significant level ($p < .05$). Figure 3 shows the extent to which differences in results between OLS and MR affect substantive conclusions. Coefficients are represented by points and 95% confidence intervals are represented by lines.[22]

Notice first that Hogan's main hypothesis is supported with the OLS model; the coefficient on partisan policy position is positive and statistically significant at the 95% level. As the Downsian framework would predict, challengers spend (and thus, raise) more money when facing incumbents with extreme partisan voting records. However, the effect weakens with MR. Though still positive and significant, the MR coefficient on partisan policy position is about 45% of the magnitude of the OLS estimate. While the OLS results indicate that a standard deviation increase in partisan policy position corresponds to a 5.4% increase in challenger spending as a percentage of incumbent spending, the MR model estimates that effect to be only a 2.4% increase.

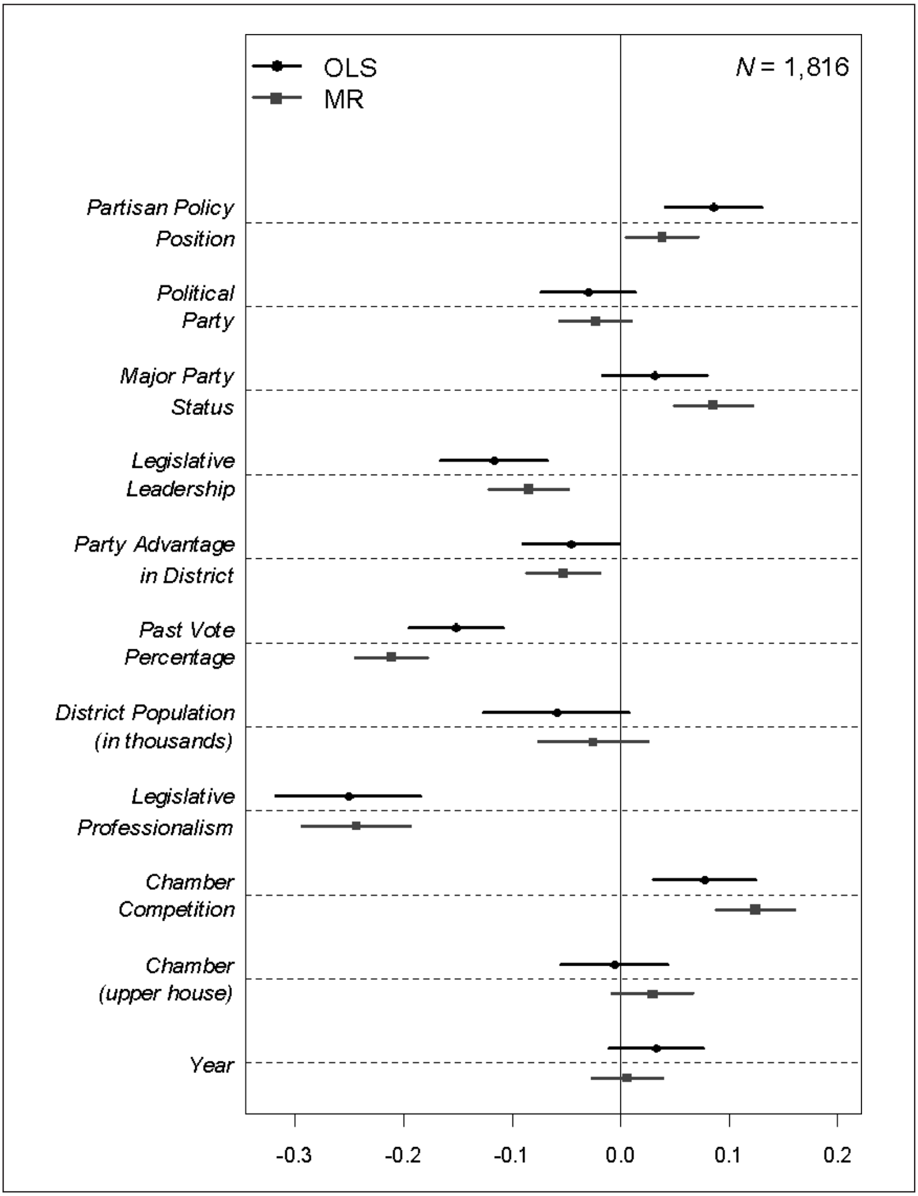

Figure 3. Reanalysis of factors affecting challenger spending as a percentage of incumbent spending

Note: OLS = ordinary least squares; MR = median regression. The graph shows OLS and MR results for the model of challenger spending as a percentage of incumbent spending. The OLS results show support for the expectations described in Hogan (2008). The coefficient on partisan policy position is positive and statistically significant at the 95% level. However, though still positive and significant, the MR coefficient on partisan policy position is about 45% of the magnitude of the OLS estimate.

Source: Hogan (2008, Table 2).

Thus, the MR model may show weaker support for Hogan's main hypothesis.[23] However, it is also important to note that the MR model produces stronger and more precise estimates for other variables. For instance, district partisanship in favor of the incumbent (party advantage in district) only weakly reduces challenger spending according to the OLS results, but with MR that effect is stronger in magnitude and statistically significant.

Elite Influence on Referendum Outcomes

Paul and Brown (2006) study the effects of elites on the outcome of voter referendums, looking specifically at votes on building sports stadiums in major cities. While past work shows that elites can drive public opinion in several contexts, their study is unique in that it provides a comparative look at the influence of different types of elites, such as elected officials and unelected business, minority, and community leaders. They expect that all of these types of elites affect referendum outcomes but that political elites have more influence because their desire for reelection makes them more credible to voters (Paul and Brown 2006, 277). They also expect that there is less support for stadiums that would impose a bigger tax burden on the public.

The authors test these hypotheses with data from 41 stadium referendum votes between 1984 and 2001. Here we examine a model of the final vote percentage in favor of the referendum. Key independent variables include indicators for whether there is Republican support or opposition and Democratic support or opposition. In particular, in the model we replicate, these factors are measured as instances in which a Republican or Democrat was the leading supporter or opponent (*Republican leading proponent*/*Democratic leading proponent* and *Republican leading opponent*/*Democratic leading opponent*). Other variables include support or opposition from minority and citizen groups (*minority support*/*citizen group support* and *minority opposition*/*citizen group opposition*) and a measure of the taxpayer burden of the proposed stadium (*public share*). The authors hypothesize that the proponent variables exert positive effects on the final outcome while the opponent variables and public share have negative effects.

Panel b of Figure 2 shows that the OLS residual density estimate has heavier tails than the normal and is a better visual match with the equivalent Laplace density than with the normal. Thus, it is not surprising that the CVDM test selects MR over OLS at a statistically significant level ($p < .05$). Figure 4 shows the differences between the two estimators. Again, coefficients are represented by points and 95% confidence intervals are represented by lines.

First note that the original OLS model provides support for the authors' hypotheses. In particular, all of the elite opinion coefficients are signed in the expected directions, and three are statistically significant (Republican leading proponent, Republican

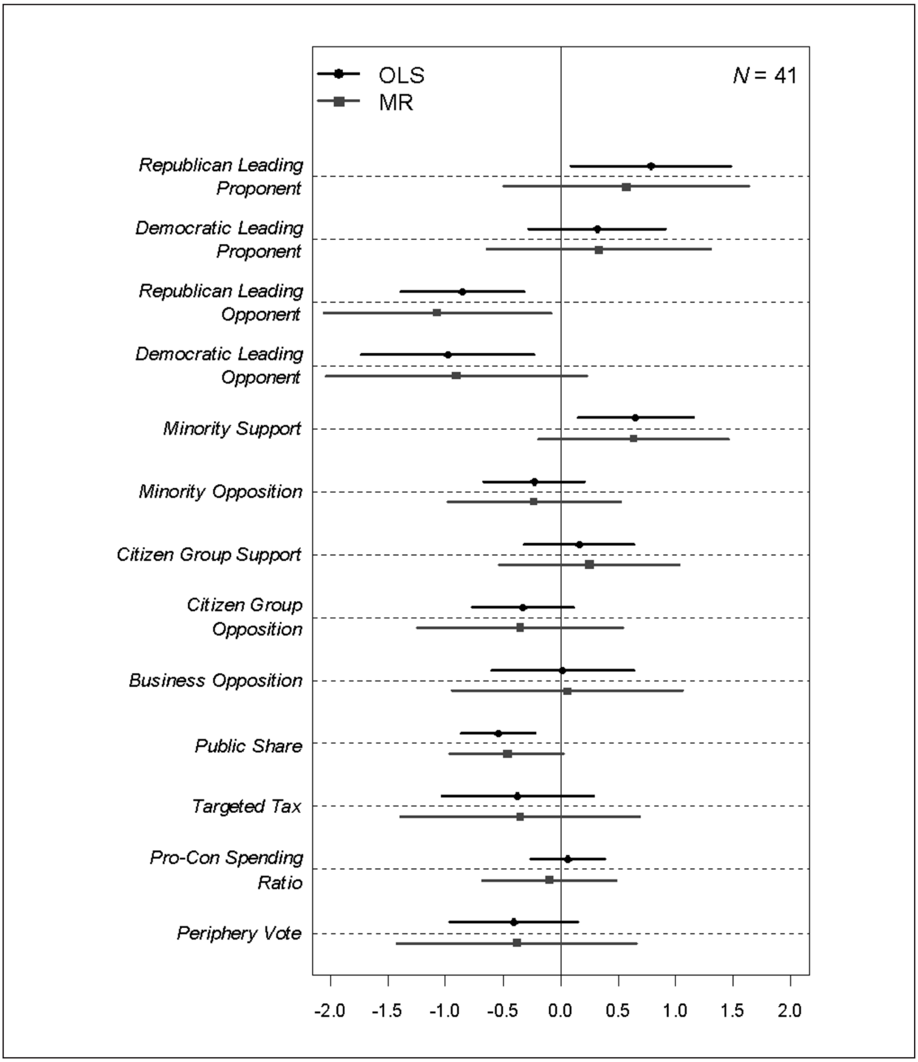

Figure 4. Reanalysis of sports facility referendum support

Note: OLS = ordinary least squares; MR = median regression. The graph shows OLS and MR results for the model of elite influence on referendums. The OLS results show support for the expectations described in Paul and Brown (2006); all four of the political elite estimates are signed in the expected direction and three are statistically significant (Republican leading proponent, Republican leading opponent, and Democratic leading opponent). In addition, minority support produces a statistically significant estimate, and public share is negative, as expected, and statistically significant. In contrast, with MR only one of the elite support variables produces a significant estimate (Republican leading opponent). The coefficient on public share also drops in magnitude and becomes nonsignificant.

Source: Paul and Brown (2006, Table 2).

leading opponent, and Democratic leading opponent). Minority support also produces a statistically significant estimate, and public share is negative, as expected, and statistically significant. Thus, with OLS there is evidence in favor of the hypotheses that elites can influence referendum outcomes and that support declines as the cost to the public increases.

However, with the better fitting MR model, the magnitude of several variables—including Republican leading opponent and public share—weakens slightly, and only one of the elite support variables produces a significant coefficient (Republican leading opponent). In addition, the coefficient on public share becomes statistically nonsignificant. Thus, at least with respect to statistical significance, the MR results show less support for the authors' hypotheses regarding the influence of elites and public burden on voter support for stadium referendums.

Conclusions

The linear model is a cornerstone of empirical political science, defining the basis on which all regression models are built, and linear hypotheses are quite prevalent in many subfields, including state politics. However, the discipline is oriented such that the conditional-mean framework of OLS is virtually the only technique for estimating the parameters of the linear model, despite the fact that many types of data present suboptimal conditions for the method. Several outcome variables important in state politics research exhibit heavy tails, and MR is often a more appropriate estimator for these data. In addition, though a number of alternatives to OLS in heavy-tailed data exist, MR holds several advantages for the applied researcher. In particular, interpretation of MR results is straightforward and similar to that of OLS; the MR linear predictor gives the expected median (rather than the mean) of the dependent variable conditional on a vector of independent variables. Thus, the analyst is not forced to alter his or her underlying theory to use MR but rather can simply gain from its more efficient estimates.

It is, of course, nearly always the case that the distribution of the error term is unknown to the researcher. We describe a simple testing procedure that can assess the relative fit of OLS to MR, even when the models are misspecified. These tests can be applied to a sample of data to choose the best estimator of the coefficients of the linear model and their standard errors. Then the researcher can present the favored estimator, or at least both sets of estimates and the results of the CVDM test. The test is easy to use and provides clear evidence as to which model should be favored in a particular applied setting. Deletion of cases, the creation of indicators for problematic outliers, and variable transformation are not needed.

Finally, we show here that the choice between OLS and MR can have important implications for interpreting results of a linear model. We describe two instances in which statistical evaluation of hypotheses changes—at least to some degree—with the use of MR. These examples are both recent publications in two prominent journals in state politics research: *American Journal of Political Science* and *State Politics &*

Policy Quarterly. In sum, we conclude that MR is a relevant and valuable tool to scholars of state politics. We recommend that researchers testing linear hypotheses carefully consider the choice between OLS and MR and use the CVDM test to determine which estimator is more appropriate for their data.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Notes

1. Median regression (MR) is quantile regression with the median as the conditional quantile of interest (Koenker and Bassett 1978).
2. Although ordinary least squares (OLS) is known for its robustness to violations of basic assumptions, we show that improvement is feasible when the tails of the error term are heavier than a normal distribution and that such improvements can make important differences in results. Indeed, while outliers can have a substantial effect on OLS estimates, their impact on those of MR is negligible.
3. We explicitly address outliers only on the dependent variable in this research.
4. As we show below, our method for choosing between OLS and MR still relies on the arbitrary nature of hypothesis testing. Nonetheless, its arbitrary *p* value provides an answer to a more specific and useful question—whether one estimation method is preferred over another—than whether a particular observation is an outlier.
5. In fact, the problems for OLS that are caused by heavy-tailed disturbances are just as important in large samples as they are in small samples.
6. The test is different from a normality assessment test such as the Kolmogorov–Smirnov test. Instead of testing whether the residuals are normally distributed, our test assesses whether OLS or MR fits the data significantly better.
7. However, MR can still be estimated just as easily as OLS on any computer with standard statistical software (e.g., R or Stata).
8. A key element of the result that MR can be more efficient than OLS even when OLS is BLUE is the important distinction between "best linear unbiasedness" and "minimum variance unbiasedness" (MVUE). In the online appendix, we discuss that distinction and the concepts of efficiency, BLUE, and MVUE in detail. The main point from that discussion is that the BLUE property of OLS means that OLS is optimal among estimators that use a linear algorithm to obtain coefficient estimates and standard errors. It has nothing to do with the theoretical specification of a linear additive relationship between a dependent and independent variables. While MR estimates a model in which there is such a linear additive relationship, it does not use a linear algorithm to obtain parameter estimates like OLS.

Thus, in a technical sense MR is a nonlinear estimator. This means that even when OLS is BLUE, MR can still be more efficient.

9. Still another approach involves the representation of the mean of the dependent variable as a complex and flexible function of the independent variables, noting that outliers can arise through misspecification of the regression function. Examples of these methods include multivariate fractional polynomial regression (Royston and Altman 1997), kernel regression (Nadaraya 1964), and generalized additive models (Hastie and Tibshirani 1986).
10. Even in regression with *t*-distributed errors, the analyst is modeling a conditional quantity that is a mixture of the mean and median dependent on the degrees of freedom (*df*) parameter. When $df \to \infty$ the location parameter is equal to the mean, and when $df \to 0$, the location parameter is equal to the median.
11. These methods are the simplex algorithm (Koenker 2005) or maximum likelihood (ML) with a Laplace likelihood function.
12. One concern when using ML is its performance in small samples. However, a small sample is not primarily a problem for the coefficient estimates, but for statistical inference. Indeed, MR defined as ML with a Laplace likelihood has been shown to be unbiased in finite samples even under certain departures from iid Laplace errors (see Jung 1996). Like all ML estimators, asymptotic (i.e., large sample) theory is used to justify inverting the negative Hessian to get estimates of the MR standard errors. In contrast, the sampling distribution of OLS is known exactly in a finite sample (assuming the Gauss–Markov assumptions are valid). If the analyst is uncomfortable using the asymptotic standard errors for MR in a small sample, several nonparametric and semiparametric routines can be used to construct inference statistics. For instance, in Stata MR standard errors can be bootstrapped or jackknifed. The R implementation of MR makes possible bootstrapping, jackknifing, a kernel density estimator, and an inverted rank test. Thus, we do not see small samples as reason to avoid MR, but simply as a common statistical issue that can be dealt with through acknowledging caution and through using alternative means of conducting statistical inference.
13. For an estimate of a parameter θ, mean squared error (MSE) is defined as the mean of the squared difference between the estimate ($\hat{\theta}$) and the true parameter, or $E[(\hat{\theta} - \theta)^2]$.
14. We also simulated data with a *t* distribution and varied the *df* parameter. Results were similar to the exponential power results we report here.
15. We also tried $N = 2{,}000$ and found the same pattern of results as shown here.
16. Note that we do not expect a direct correspondence between MSE and the cross-validated difference in means (CVDM) statistic. Though MSE is a common and intuitive measure of estimator performance, it is not the quantity that the CVDM statistic measures. The CVDM statistic measures the relative Kullback–Leibler divergence of OLS and MR, which is a more direct measure of the relative appropriateness of the parametric assumptions underlying OLS and MR but is less intuitive than relative MSE.
17. We do not have a specific sample size at which we can say researchers should or should not use the test. It is an unbiased estimator of the difference in Kullback–Leibler divergences at any sample size, though it becomes more efficient with more data points. We show the simulation results simply to point out that at 50 observations—an important number for state politics research—there is evidence that the test performs reasonably well.

18. However, its usefulness is not limited to state politics. Indeed, when we began this project, we collected many more models than just these three. Table A1 in the online appendix summarizes our replication findings from 15 different models, including the two reported here. That table shows (1) that the CVDM test can be used in several instances across political science, (2) that the test chooses both OLS and MR across the different examples, and (3) that MR could potentially lead to less support, more support, no change, or mixed results compared to the original OLS model with respect to either statistical or substantive significance.
19. Although these examples use OLS, both dependent variables may not be appropriate for OLS or MR because they are constructed as percentages.
20. The *t*-statistics are −10.73 (Hogan 2008) and −1.87 (Paul and Brown 2006).
21. However, we do not wish to make strong claims about the conclusions of either analysis. We show these examples only to illustrate the potential usefulness of MR. Interpretation of the change in support for hypotheses can be somewhat subjective. Most importantly, researchers should be mindful of both statistical and substantive significance when comparing the results.
22. Coefficients are standardized for ease of presentation only, not to make comparisons between effects (see King 1986).
23. To be sure, some may consider a 2.4% increase to be support for Hogan's hypothesis.

References

Bassett, Gilbert, and Roger Koenker. 1978. "Asymptotic Theory of Least Absolute Error Regression." *Journal of the American Statistical Association* 73:618–22.

Berkman, Michael B., and Christopher Reenock. 2004. "Incremental Consolidation and Comprehensive Reorganization of American State Executive Branches." *American Journal of Political Science* 48:796–812.

Cook, R. Dennis. 1977. "Detection of Influential Observation in Linear Regression." *Technometrics* 19:15–18.

Desmarais, Bruce A., and Jeffrey J. Harden. N.d. "An Unbiased Model Comparison Test Using Cross-Validation." http://ssrn.com/abstract=1765845.

Downs, Anthony. 1957. *An Economic Theory of Democracy*. New York: Harper & Row.

Gujarati, Damodar N. 2003. *Basic Econometrics*. International ed. New York: McGraw-Hill.

Hao, Lingxin, and Daniel Q. Naiman. 2007. *Quantile Regression. Quantitative Applications in the Social Sciences*. Thousand Oaks, CA: Sage.

Hastie, Trevor, and Robert Tibshirani. 1986. "Generalized Additive Models." *Statistical Science* 1:297–310.

Hogan, Robert E. 2008. "Policy Responsiveness and Incumbent Reelection in State Legislatures." *American Journal of Political Science* 52:858–73.

Huber, John D., Charles R. Shipan, and Madelaine Pfahler. 2001. "Legislatures and Statutory Control of Bureaucracy." *American Journal of Political Science* 45:330–45.

Huber, Peter J. 1964. "Robust Estimation of a Location Parameter." *Annals of Mathematical Statistics* 35:73–101.

Johnson, Norman J. 1978. “Modified *t* Tests and Confidence Intervals for Asymmetrical Populations.” *Journal of the American Statistical Association* 73:536–44.

Jung, Sin-Ho. 1996. “Quasi-Likelihood for Median Regression Models.” *Journal of the American Statistical Association* 91:251–57.

King, Gary. 1986. “How Not to Lie with Statistics: Avoiding Common Mistakes in Quantitative Political Science.” *American Journal of Political Science* 30:666–87.

Koenker, Roger. 2005. *Quantile Regression*. New York: Cambridge University Press.

Koenker, Roger, and Gilbert Bassett. 1978. “Regression Quantiles.” *Econometrica* 46:33–50.

Krueger, James S., and Michael S. Lewis-Beck. 2008. “Is OLS Dead?” The *Political Methodologist* 15(2): 2–4.

Lai, P. Y., and Stephen M. S. Lee. 2005. “An Overview of Asymptotic Properties of *L*p Regression under General Classes of Error Distributions.” *Journal of the American Statistical Association* 100:446–58.

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: 881–96.

Legendre, Adrien M. 1805. *Nouvelles Metermination des Orbites des Comethodes pour la Detes*. Paris: F. Didot.

Leighton, Wayne A., and Edward J. Lopez. 2002. “Committee Assignments and the Cost of Party Loyalty.” *Political Research Quarterly* 55:59–90.

Lowery, David, and Virginia Gray. 1995. “The Population Ecology of Gucci Gulch, or the Natural Regulation of Interest Group Numbers in the American States.” *American Journal of Political Science* 39:1–29.

Maire, Christopher, and Ruggero Giuseppe Boscovich. 1755. *De Litteraria Expeditione per Pontificiam Ditionem ad Dimetiendos duos Meridiani Gradus: et Corrigendam Mappam Geographicam, iussu et auspiciis Benedicti XIV Pont. Max*. In Typographio Palladis excudebant N. et M. Palearini.

Mineo, Angelo M., and Mariantonietta Ruggieri. 2005. “A Software Tool for the Exponential Power Distribution: The normalp Package.” *Journal of Statistical Software* 12(4): 1–24.

Morgenthaler, Stephan. 1994. “Small Sample Efficiency and Exact Fit for Cauchy Regression Models.” *Statistics and Probability Letters* 19:381–85.

Nadaraya, E. A. 1964. “On Estimating Regression.” *Theory of Probability and Its Applications* 9:141–42.

Paul, David M., and Clyde Brown. 2006. “The Dynamics of Elite Endorsements in Professional Sports Facility Referendums.” *State Politics & Policy Quarterly* 6:272–99.

Royston, Patrick, and Douglas Altman. 1997. “Approximating Statistical Functions by Using Fractional Polynomial Regression.” *Statistician* 46:411–22.

Smyth, Padhraic. 2000. “Model Selection for Probabilistic Clustering Using Cross-Validated Likelihood.” *Statistics and Computing* 10:63–72.

Western, Bruce. 1995. “Concepts and Suggestions for Robust Regression Analysis.” *American Journal of Political Science* 39:786–817.

Wolak, Jennifer, David Lowery, and Virginia Gray. 2001. “California Dreaming: Replicating the ESA Model, Unusual Cases, and Comparative State Political Analysis.” *State Politics & Policy Quarterly* 1:255–72.

Bios

Jeffrey J. Harden is a PhD candidate in political science at the University of North Carolina at Chapel Hill. His research focuses primarily on U.S. state politics with a particular emphasis in public opinion, representation, and methodological issues.

Bruce A. Desmarais is an assistant professor in the Department of Political Science and core faculty member of the Computational Social Science Initiative at the University of Massachusetts-Amherst. His primary areas of research are in political network analysis and out-of-sample evaluation of models of political processes.