

CAMBRIDGE UNIVERSITY PRESS  
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo, Delhi

Cambridge University Press  
32 Avenue of the Americas, New York, NY 10013-2473, USA

www.cambridge.org  
Information on this title: www.cambridge.org/9780521680035

© Cambridge University Press 2009

This publication is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published 2009

Printed in the United States of America

*A catalog record for this publication is available from the British Library.*

*Library of Congress Cataloging in Publication data*

A quantitative tour of the social sciences / [edited by] Andrew Gelman, Jeronimo Cortina.  
p. cm.

Includes bibliographical references and index.

ISBN 978-0-521-86198-4 (hardback) – ISBN 978-0-521-68003-5 (pbk.)

I. Social sciences – Methodology. 2. Social sciences. I. Gelman, Andrew.

II. Cortina, Jeronimo. III. Title.

H62.Q365 2009

300.72–dc22 2008054986

ISBN 978-0-521-86198-4 hardback

ISBN 978-0-521-68003-5 paperback

Cambridge University Press has no responsibility for the persistence or accuracy of URLs for external or third-party Internet Web sites referred to in this publication and does not guarantee that any content on such Web sites is, or will remain, accurate or appropriate. Information regarding prices, travel timetables, and other factual information given in this work are correct at the time of first printing, but Cambridge University Press does not guarantee the accuracy of such information thereafter.

# A QUANTITATIVE TOUR OF THE SOCIAL SCIENCES

Edited by

ANDREW GELMAN

*Columbia University*

JERONIMO CORTINA

*University of Houston*

 CAMBRIDGE  
UNIVERSITY PRESS

---

**PART V. POLITICAL SCIENCE**

---

*Charles Cameron*

## 16. The Politics of Supreme Court Nominations: The Critical Role of the Media Environment

In this chapter and the next, we'll work through an extended empirical political analysis involving the politics of Supreme Court nominations. I'll present this material in approximately the way the analysis took place, rather than in the cleaned-up and artificially neat way one presents it in a seminar or paper. I'll begin by reviewing an earlier study on how senators vote on Supreme Court nominations. The voting analysis raises a puzzle, apparently concerning the choices presidents make when they select nominees. But, as we shall see, the puzzle isn't really about presidents at all. Rather, the puzzle involves how the news media portray nominees and (ultimately) how hostile members of the Judiciary Committee manipulate the media environment in order to damage nominees. Or so I claim.

The point of these chapters is to show how a real analysis unfolds, which takes place in a way that is quite different from what one might think after reading journal articles or scholarly books. So, as we go, I'll discuss some personal background, the places where I was stumped, how luck entered in, where the ideas came from, and how I dealt with measurement issues. I'll also make comments about various statistical techniques and choices. The following chapter will consider the interplay of theory and data as well.

### THE SPATIAL MODEL OF VOTING

We start with the basic model that political scientists use to describe voting based on issue positions. This so-called spatial model is commonly used to characterize voters in an election, but here we will use it to describe U.S. senators deciding whether to vote to confirm nominees for the Supreme Court. The central idea comes out of mathematical psychology. (Remember when I said that political scientists often use methods developed in other social sciences? This is a case in point.) The original psychological question was: How can you tell if two objects, like a ball and a bat, are similar or different? Well, define some dimensions on which to measure the objects – weight, color, shape, and so on. The Cartesian product of these dimensions defines a space, and the characteristics of the ball and bat then yield two vectors in the hyperspace.

To put it another way: Suppose there are three continuous dimensions, which you can think of as the three dimensions of the room you're in: length, breadth,

and height. A specific item measured on these dimensions (length, breadth, and height) then becomes a point (a vector) in this three-dimensional space. Another item with somewhat different characteristics then becomes another point in the space. For example, you could have one object at (.25, .25, .5) and another at (.75, .75, .75). The triple  $(x, y, z)$  indicates the position of the items in terms of length, breadth, and height.

Now define a distance metric in the space, in other words, a yardstick for measuring the distance between the two points. One such measure is the Euclidean distance based on the Pythagorean formula you studied in school. In this case, the distance between object 1 and object 2 is

$$d_{12} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}.$$

In our example, this would be

$d_{12} = \sqrt{(.25 - .75)^2 + (.25 - .75)^2 + (.5 - .75)^2} = .75$ . In the psychological setup, you can use this metric to see how closely two items resemble each other.

You can play with the distance metric in various ways. For example, you can add weights to each dimension to indicate which ones are more important or less important (these are called “saliency weights” by political scientists). You can replace the squares with absolute values and dispense with the square root. You can dispense with the square root and just use the squared distance. And so on. You get the basic idea.

Political scientists stole this shiny piece of technology for their own purposes. Suppose the dimensions in the figure represent political dimensions of some sort. For example, we can represent a candidate’s general domestic policy liberalism as a location on one dimension, her general foreign policy position on another, and her position on some issue I particularly care about on the third. A point in the space then represents a candidate. And suppose I have preferences about policies – in particular, suppose I have a favorite policy point or location in the space, my “ideal point.” Then I can say which of two candidates is closer to my ideal position. If I am choosing between them, say in an election, I can vote for the closer of the two if I wish.

In fact, you can use the distance from the ideal point as a measure of the “utility” of one candidate or alternative. Then we imagine voters choosing the alternative with the greater utility, which is equivalent to saying that they choose the closer of two alternatives in the space.<sup>1</sup>

This set of tools has proven incredibly useful in political science, so much so that representing policies, laws, and candidates in spatial terms has become ubiquitous. For example, as an empirical matter, congressmembers appear to act as if they have a stable ideal point in a one-dimensional space, bills come with a location in the space (a number), and congressmembers vote for the closer bill

<sup>1</sup> In some circumstances, it might be better to vote for a more distant alternative rather than the closest one – for example, if you know that voting for the closest alternative is just “throwing your vote away,” while voting for a somewhat more distant choice can head off a really bad outcome. Any textbook on game theory and politics will explain situations like this in detail if you are interested (Ordeshook 1968). For more on spatial theory, see Enelow and Hinich (1984).

when they can (Poole and Rosenthal 1997). This policy space looks like a generic “liberal–conservative” dimension. At times during history, Congress appears more two-dimensional, with the second dimension typically related to race or geographically specific issues. But those times are relatively rare, and at present Congress looks very one-dimensional.

### VOTING ON SUPREME COURT NOMINEES

Define two variables:  $q_j$ , the “perceived qualifications” of nominee  $j$ , and  $d_{ij}$ , the ideological distance between senator  $i$  and nominee  $j$ .<sup>2</sup> We measured these using newspaper editorials: We had research assistants gather and code hundreds of editorials from liberal and conservative newspapers. To get  $q_j$ , we had them code the percentage of editorials that said that the nominee was “qualified” to be on the Court. To get the ideological position of the nominee, we coded the percentage of editorials that said that the nominee was “liberal.” (The variable was actually measured on a scale from  $-1$  to  $1$ .)

Coding of this sort is called “content analysis,” which is an almost absurdly fancy term for reading newspaper stories carefully. What separates content analysis from just perusing the newspaper is how systematic you are. We tried to be very systematic. For example, we wrote down rules for what counted as statements in the editorials. We employed several different coders and, using random samples, made sure that there was a high degree of intercoder reliability among them.<sup>3</sup> If you used the same rules and practiced a little, and then read and coded the stories yourself, I am confident that your results would strongly resemble ours – at least for ideology. The coders had some trouble getting consistent results on “quality” – they tended to pick up editorial endorsements and the general tone of the editorials. With practice, however, they seemed to get better at coding whether the editorialists said that the nominee was qualified to serve on the Supreme Court.

We wanted a measure of  $d_{ij}$ , the ideological distance between senator  $i$  and nominee  $j$ , so we also needed a measure of each senator’s ideology. To get that, we used ratings from the liberal interest group Americans for Democratic Action (ADA). Today, I would use the professional standard, the Poole-Rosenthal NOMINATE scores, which I mentioned in Chapter 15. You will recall that these are derived from scaling every roll call vote ever taken in Congress. The NOMINATE scores come in slightly different flavors, but the better ones to use here would be those that are comparable over time. The ADA scores fail in this regard, but I’ll pass over that point here. It turns out that, for our analysis, if we had used the NOMINATE scores, the substantive results would have been unchanged.

<sup>2</sup> Because every senator presumably finds better qualifications more attractive, perceived qualifications is a “valence dimension,” in the jargon of spatial theory. There is an interesting literature on valence dimensions in politics, but we don’t need to go into all the details.

<sup>3</sup> If you want to learn more about how social scientists use content analysis, you can read Weber (1990).

A tricky issue is that our measure of nominee ideology and our measure of senator ideology weren't measured on the same scale. You can think of the two scales as being something like Fahrenheit and Celsius scales for temperature. You can easily look up a formula for converting Fahrenheit into Celsius, but we had no such formula for these ideological measures. However, we needed one so that we could calculate distances between senators and nominees. We solved this problem with brute force.<sup>4</sup> That is, we assumed that the conversion formula between the two was linear (of the form  $y = ax + b$ ), calculated many such conversions, and ran the analysis over and over to see what conversion weights seemed to work best. A lot of subsequent work has gone into trying to find a common scale for congressmembers, presidents, and Supreme Court justices, and different methods have different strengths and weaknesses. Fortunately, these voting data are so strongly structured that the results are fairly insensitive to which method you use.

What did we find? To a surprising degree, senators appeared to vote as if they were using the following distance metric:

$$d_{ij} = b_1 + b_2(1 - q_j) + b_3(x_j - x_i)^2 + b_4(1 - q_j)(x_j - i)^2 \quad (1)$$

Then senators would vote "yes" if  $d_{ij}$  was below a critical threshold and "no" if it was above.

Let me explain equation (1) a little more. The first term is just a constant. The second term measures perceived lack of qualifications, since perceived qualifications are on a 0–1 scale. The coefficient  $b_2$  indicates the weight for lack of qualifications; this is a parameter you estimate from the data. The third term is the ideological distance between the senator and the nominee, with  $x_j$  being the measure for the nominee and  $x_i$  being the measure for the senator. I've explained how we measured those. The coefficient  $b_3$  is the weight on this term, again estimated from the data. The fourth term is an "interaction" between lack of quality and distance: For highly qualified nominees or very close candidates, the term is small. But for poorly qualified, distant ones, it is large. The coefficient  $b_4$  measures the importance of this term.

It turns out that this model provides an elegant and powerful description of senators' voting behavior on Supreme Court nominees, at least from 1937 to 1994. I emphasize description, since it's important to remember that all regression equations like this one are just summary descriptions of patterns in data. By themselves, they don't explain anything; supplying interpretation – in terms of causal mechanisms – is up to you. Anyway, I now know that the equation misses a few things; for example, after the *Brown v. Board of Education* Supreme Court decision, southern Democrats voted against a sequence of nominees they saw as racial liberals. This pattern ended shortly after the enactment of the Voting Rights Act. Then the equation kicked in again for southern Democrats. But improvements in the equation are not what I want to discuss.

Rather, I want to get back to the question, What does the equation mean substantively? Superficially, the equation says that a senator will vote in favor of a nominee if the senator likes the nominee enough – that is, if the senator sees the

<sup>4</sup> In the next chapter, I'll describe a method Jee-Kwang Park and I derived that I think is much better.

nominee as qualified and ideologically proximate. But in fact, the estimated weights in the equation say something much more intriguing than this. Consider the "quality" variable,  $q_j$ . Pick a high number, such as .8, so that the nominee is perceived as being of high quality. Then all the senators will vote for the nominee, regardless of their ideology. In other words, the nomination is consensual, and this is true even for ideologically extreme nominees. Now pick a low number for  $q_j$ , like .3. Senators who are close to the nominee lie on high contours, but those who are more distant from the nominee lie on low contours. So, the nominee's ideological soul mates vote for the nominee even if he or she is a "turkey," but those farther away vote against the nominee. The nomination becomes conflictual and ideologically polarized. This can happen only if the nominee is of low quality.

If you think about this for a moment, it should strike you as rather weird. It says that an extreme liberal like Ted Kennedy will vote for an extreme conservative, like Antonin Scalia, if Scalia is perceived as being of high quality by the editorialists (this actually happened, by the way). And an extreme conservative will vote for an extremely liberal nominee if she is perceived as being of high quality. In effect this occurred with the nomination of Arthur Goldberg and the initial nomination of Abe Fortas, in which conservatives acquiesced to a voice vote. But we haven't had a situation like this in a while.

Now you can see the first puzzle raised by the empirical results, which is: Why do senators behave this way? Why (as a senator) can you vote against someone you hate ideologically only if he or she is also a turkey on quality grounds? This is good puzzle indeed, but it's not one we will consider right now. So, let me continue laying out essentially the same results in a different way. This will help you see the magnitude of the effects we're talking about.

Let's switch from individual roll call votes to aggregate Yes–No votes for nonwithdrawn nominees between 1937 and 2005 (through the Roberts nomination). The data on individual votes allow us to connect senators' behavior to some well-developed theory, but if you are interested in the politics of Supreme Court nominations – especially outcomes – it's aggregate votes you really care about.

Figure 16.1 shows the distribution of Yeas and Nays in votes on Supreme Court nominations between 1937 and 2005. When there was a roll call vote on a nomination, I use the split in the actual roll call vote; for cases with a voice vote, I count the nomination as a unanimous Yes vote. So, the figure shows you the number of nominations with percent Yes votes in various ranges.

The key point to gain from Figure 16.1 is that most nominations are consensual, with all or almost all senators voting Yes. But there are quite a few nominations that are not consensual – there are substantial numbers of No votes. Incidentally, if we looked at all the data from Reconstruction period on, the pattern would be almost exactly the same.

If you put the logics of equation (1) and Figure 16.1 together, you can make a pretty sharp guess on which nominations are consensual and which are conflictual – in other words, where the No votes come from: No votes tend to occur when the nominee is perceived by the editorialists as of poor quality. Moreover, if you suspect that Democratic presidents typically nominate rather liberal justices

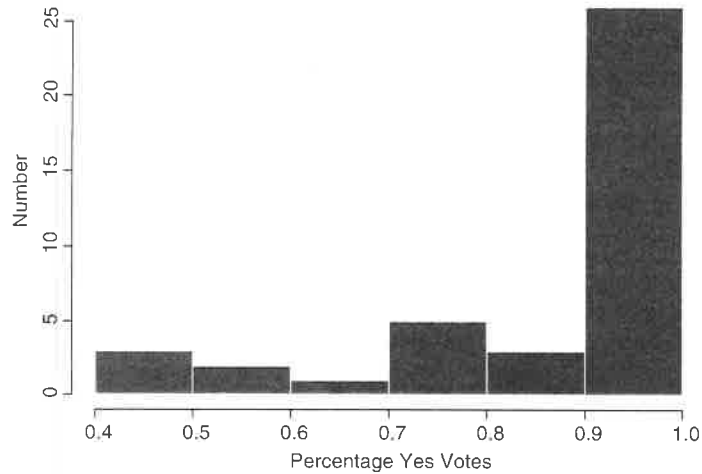


Figure 16.1. Distribution of the proportion of Yes votes in the Senate for all nonwithdrawn Supreme Court nominations, 1937–2005.

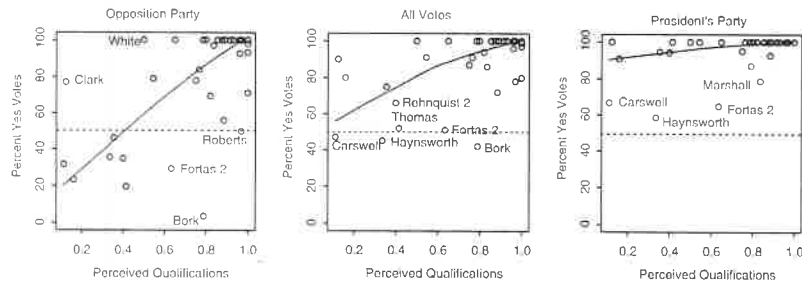


Figure 16.2. Percentage of Yes votes in the Senate as a function of perceived qualifications for Supreme Court nominations, 1937–2005.

and Republicans rather conservative ones (which is exactly what happens), then from equation (1) you might suspect that low qualifications provoke more No votes from members of the opposition party than from members of the president's party.

Figure 16.2 confirms these conjectures. What is shown is the percentage of Yes votes as a function of the nominee's perceived qualifications for each nomination from 1937 to 2005 (through Roberts) that went to the floor. (In other words, the data exclude the three nominations in that period that were withdrawn before a vote was taken: Homer Thornberry, Douglas Ginsburg, and Harriet Miers.) The middle panel displays votes from both parties; the panel on the left shows the relationship for members of the opposition party, while the panel on the right shows it for members of the president's party. The lines in the panels are the fits from highly flexible, nonparametric regression models, locally weighted

regressions, so-called lowess lines (Cleveland 1979).<sup>5</sup> I love using lowess regressions when exploring data. They let you see the main patterns in the data very clearly, including essential nonlinearities.

Let's start with the middle figure, which presents the overall relationship across both parties. As shown, the relationship seems clear and pretty strong. Particularly interesting, though, are those nominations that are unusually low. So, I've indicated which nominations those are. The percentages of Yes votes in the nominations of G. Harrold Carswell and William Rehnquist to be chief justice are low but seem in line with the overall pattern. However, the percentages of Yes votes in the nominations of Clement Haynsworth, Clarence Thomas, Abe Fortas for chief justice, and especially Robert Bork appear unusually low.

Now consider the left and right panels. One immediately perceives a dramatic difference in the behavior of members of the opposition party and the president's party. In the opposition party, support for the nominee plummets if his or her perceived qualifications are low. But support remains quite strong in the president's party, even for nominees whose qualifications are said to be poor in the newspaper editorials.<sup>6</sup> One can see in the figures that opposition Republicans displayed unusual mobilization against Abe Fortas (for chief justice), as did the Democrats against Robert Bork. In-party Republicans displayed unusually low support for Carswell and Haynsworth, and in-party Democrats showed unusually low support for Thurgood Marshall and Abe Fortas for chief justice. The Marshall defections reflect the opposition of southern Democrats to the first African American nominated to the Supreme Court.

Among the "safe" or consensual nominations, only 16% of the nominees were scored with low qualifications. Among the failed nominations, 100% received low qualifications ratings from the editorialists. Among the polarized nominations, two-thirds received low qualifications ratings from the editorialists.

### PROBABILITY OF CONFIRMATION

To look at the situation in one more way, consider the probability that nominees are confirmed simply as a function of their perceived qualifications. Essentially, if the president nominates someone with a perceived qualifications score above .8, the chance of approval is very high. But if he nominates someone whose score is below about .6, the chance of approval falls dramatically. Unsurprisingly, a simple logistic regression finds perceived qualifications to be a statistically significant predictor of approval.

The conclusion is clear: If the president nominates someone whom the editorialists score as highly qualified, the nomination will probably be consensual. Even if it isn't, it will probably avoid serious trouble. Conversely, if the

<sup>5</sup> If one runs simple linear regressions, the coefficients on perceived qualifications are highly statistically significant in all three cases, with intercepts and slopes (respectively) of 26 and 69 (opposition party), 58 and 39 (all votes), and 81 and 19 (president's party). I will present some regressions like this one shortly.

<sup>6</sup> At this level of aggregation, one cannot say whether the difference across the parties is due entirely to ideological distance or to some particular "party" factor, though regressions at the individual level suggest that any such party factor is small relative to ideological distance.

president nominates someone the editorialists perceive as low in quality, the nomination may fail. Even if it succeeds, there may well be a serious struggle in the Senate.

### THE PUZZLE . . . OR IS IT?

The puzzle should now be clear: If presidents can get whatever they want ideologically just by nominating the sort of person newspaper editorialists like, why do presidents ever nominate someone the editorialists see as a turkey?

I certainly didn't believe that presidents would deliberately choose poorly qualified nominees. (This analysis preceded Harriet Miers's nomination, which is hard to explain using any theory of rational presidential choice.) It was obvious that something big was going on with nominations beyond the narrow bounds of our spatial model. But what?

I didn't know, and that's where things sat for several years, until I joined the faculty at Columbia and (of course) had to teach several courses each year. The department at Columbia prided itself on offering small, intense, research-oriented seminars to seniors majoring in political science, and there was always a need for such courses. I didn't have a lot of big lecture-type courses ready, so I volunteered to teach a senior seminar on the politics of Supreme Court nominations. The course proved to be fairly popular, so I taught it for several years.

I had the good fortune to have a string of terrific students in the course. These students had little sense of modern social science research – perhaps this book will be a help here – but they were extremely bright and extraordinarily hard-working, and they needed topics for their seminar papers. So I dreamed up a series of questions that someone who didn't know any statistics or game theory would have a fair chance of answering through hard work alone. Most of these involved rather intense data collection, such as counting every question asked a nominee during the Judiciary Committee hearings or finding all the occasions when presidents “went public” on behalf of nominees.

Most of the students seemed to enjoy doing these papers. The idea that you can ask a clear social scientific question, collect systematic data, and then actually nail the question seemed to be a revelation to most of them. I was delighted that some graduated with honors as a result of their papers. One student even went on to graduate school in political science and now has a successful academic career (I claim no credit for her success).

A paper I wasn't enthusiastic about was one several students proposed, on scandals and Supreme Court nominations. The subject seemed rather frivolous. But once I got the students to formulate a clear research design (which – surprise – involved systematically collecting and coding newspaper stories), I was willing to let them try it.

### ENTER SCANDALS

The students worked hard, and after several months they brought back the data in Table 16.1. In retrospect, they didn't get things quite right but they came pretty close.

Table 16.1. *Scandals and results of Senate confirmations for U.S. Supreme Court nominations, 1937–1994*

	No Scandal	Scandal	
Confirmed	27	6	33
	1	4	5
Not Confirmed	28	10	38

A table like this ought to send shivers up the spine of any social scientist who likes data. It certainly did mine! What does it say? In 96% of nominations without scandals, the nominee was confirmed. In nominations with scandals, this figure was only 60%. Wow! But hold on. Perhaps this apparently striking difference is due to the smaller number of nominations with scandals – the supposed difference could be due to chance, given the smaller numbers. So, let's check with a chi-square test (easy to do with my handy computer). I calculate the chi square as 5.7 with 1 degree of freedom, which is statistically significant, indicating a lack of independence. In short, it is highly unlikely that the apparent difference in the confirmation rates between scandalous and non-scandalous nominations is due to chance alone.

Well, then, case solved! Presidents don't nominate turkeys. Rather, presidents nominate apparently good candidates. Some then turn out to be turkeys. Being a turkey tanks your perceived qualification score and provokes a conflictual vote, a withdrawal, or even a turn-down.

This may be true, but we haven't shown it yet. And (I will argue) this simple understanding may be a little naive about scandals. But let's take things one at a time.

### WHAT DRIVES “PERCEIVED QUALIFICATIONS”?

I am going to cheat a little at this point. What actually happened was that I got excited about scandals and went back and coded the newspapers extremely carefully for a much longer range of time (starting with the Reconstruction period that followed the Civil War). Then I started thinking about the strategic role of scandals in nominations. It was only later that I collected the data we are going to look at now. We'll take up strategy and scandals in the next chapter.

Next, I want to consider the question of what seems to drive the evaluations of the editorialists. What is being captured by our “perceived qualifications” measure? To answer the question, I am going to introduce three new variables:

*Objective qualifications:* A scale from 0 to 3, on which a nominee receives 1 point each if he or she has (1) served as a judge at either the state or federal level, (2) held a responsible position in the executive branch as a lawyer (for example, solicitor general or deputy attorney general), or (3) been a professor in a law school. This variable has a mean of 1.7, and half of the observations fall between 1 and 2.

*Negative tone of media coverage (“tone”):* The percentage of stories in *The New York Times* about the nominee that report “bad news,” including

improper behavior, poor qualifications, protests by groups and mobilization against the nominee, or extreme ideological positions. The mean for this variable is 22%, and half of the observations lie between 7% and 33%.<sup>7</sup>

*Ideological extremity:* Since ideology was measured on a -1 to 1 scale, extremity is just the absolute value of ideology. It has a mean of .59, and half of the observations fall between .40 and .81.

The connections between these variables and perceived qualifications, as indicated in editorials, are obvious. The variable "tone" provides a measure of overall media coverage, while "objective qualifications," "scandal," and "extremity" provide more detailed looks at the likely content of the media coverage.

Before undertaking any statistical analysis, it's always a good idea simply to look at the data. You'll be surprised how often you find data entry mistakes and miscodes or see a relationship you hadn't thought about. One useful device is a scatterplot matrix, which most statistical packages will produce. In doing this, we found perceived qualifications to be strongly (negatively) related to negative media tone, somewhat correlated with objective qualifications and scandals, and not particularly related to extremism.

Now let's do the thing most political scientists would find natural, which is to start running linear regressions. Let me warn you, though: The models I am going to show you next are terrible models of the data. They fundamentally misrepresent what's actually happening in the data. So, let's see if you can find out why, just from the results as they come out of the computer.

The estimated coefficients and standard errors for several different models are shown in Table 16.2. As you can see, they actually look pretty good – many of the coefficients are more than two standard errors away from zero, indicating statistical significance at the usual 5% level – except perhaps that objective qualifications don't seem to have any impact and extremism has the wrong sign. But you can easily rationalize the latter finding: It might be that presidents who want to nominate ideologically extreme nominees pick particularly attractive people to offset their extremity.

I've already suggested that we shouldn't be content with these results. The most important things to check in a linear regression model are nonlinearity and interactions. Basically, you have to worry about assuming that a variable comes in linearly when it doesn't.

In the example here, a key issue is interactions. How can you check quickly and easily for interactions? One approach is to use conditioning plots, so-called co-plots. Co-plots are an extremely elegant and extraordinarily powerful visual device for detecting interactions. Surprisingly, they are not widely used by political scientists. I suppose this is because they were invented fairly recently by

Table 16.2. *Estimated coefficients (and standard errors) for coefficients in regression models predicting Supreme Court nominees' perceived qualifications*

	Model 1	Model 2	Model 3	Model 4
Intercept	.96 (.05)	.97 (.05)	.89 (.09)	.82 (.09)
Tone	-.90 (.18)	-.67 (.20)	-.62 (.20)	-.76 (.20)
Scandal	—	-.19 (.08)	-.20 (.08)	-.19 (.08)
Objective qualifications	—	—	.04 (.04)	.02 (.04)
Extremism	—	—	—	.23 (.11)
Degrees of freedom	39	38	37	36
R <sup>2</sup>	.38	.47	.48	.54

*Note:* Cases are all nominations from 1937 to 2005. It would be better to display these estimates and uncertainties graphically (thus making comparisons clearer), but we show them as a table here to illustrate the common practice in social science research.

a statistician and don't appear in econometrics textbooks, which is where most political scientists learn statistics.<sup>8</sup> You can learn more about co-plots in Cleveland (1993), a wonderful book that I recommend highly to anyone who likes to interrogate numbers.

The basic idea in a co-plot is to look at a series of bivariate scatterplots (with lowess lines, of course) as you vary a third variable. If the fit in the bivariate scatterplot simply shifts upward or downward as the third variable moves from low to moderate to high, this is evidence of an additive effect from the third variable. But if the slope of the fit in the bivariate scatterplot changes in a major way, this is evidence of an interaction between the third variable and the *x*-axis variable. (You can vary a fourth variable simultaneously as well.) If an interaction is at all sizable, you will easily see it in a co-plot. In fact, the major danger when using powerful tools like co-plots is not missing something; it's overfitting the data.

So, let's take a look at the relationship between perceived qualifications and tone as we vary scandal. Since scandal only takes the value 0 or 1, this is about as simple a co-plot as you can get. It is shown in Figure 16.3, where the left panel shows the relationship between perceived qualifications and tone for the nominees without a scandal and the right panel shows the relationship for those with a scandal.

Figure 16.3 screams interaction. As you can see, if the nomination process failed to uncover a scandal, perceived qualifications were unrelated to negative media tone. In fact, media tone was mostly positive (less than .3) and perceived qualifications were favorable (above .8).<sup>9</sup> But if there was a scandal, perceived

<sup>7</sup> To identify the stories, I first searched the digital *New York Times* via Proquest, using the name of the nominee, from the time she was nominated until the date the nomination ended. Then I eliminated all stories that were not centrally about the nominee rather than a peripheral mention; these central stories are the denominator (but the statistical results don't change much if you just use the raw count of stories mentioning the nominee). Then I coded the central stories for "bad news" as defined previously.

<sup>8</sup> Sometimes political scientists call their statistics "econometrics" even though their work involves no economic theory. It's because they learned their statistics from econometrics books.

<sup>9</sup> The two obvious outliers are Sherman Minton and Harriet Miers, both of whom had low objective qualifications and had engaged in partisan political activity.



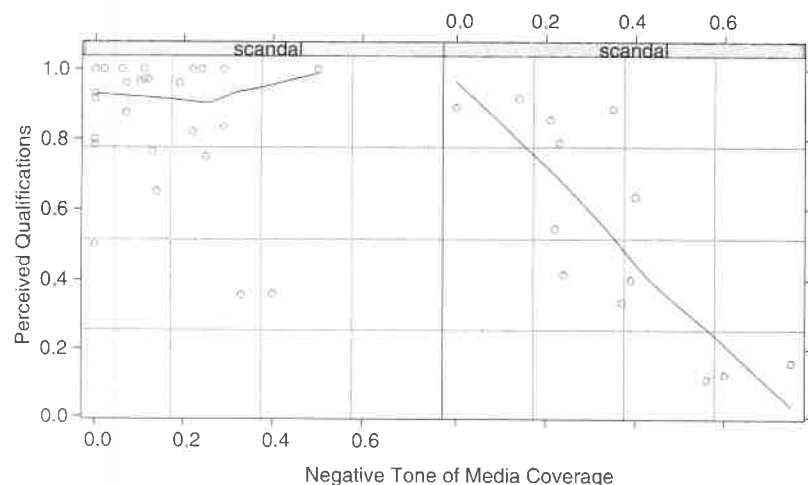


Figure 16.3 Conditioning plot of perceived qualifications, negative tone of media coverage, and scandal for Supreme Court nominations, 1937–2005.

qualifications were very strongly related to media tone. Here, most observations were greater than .3 for tone (that is, more negative in tone) and less than .8 for perceived qualifications.

Table 16.3 displays some simple linear regression fits exploring the interaction. The first two columns show models run separately for nominations with and without scandal. These strongly confirm the pattern in Figure 16.3. (Additional variables add little to these models, so I do not show them.) The third column shows the obvious model for all nominations: tone and scandal as main effects, plus the interaction between tone and scandal. The model suggests that only the interaction is doing any work. The final model forces the main effects to zero to focus on the interaction. Some analysts strongly object to dropping main effects in models with interactions, but little changes from the model with the direct effects.

What does this all tell us? One plausible interpretation is that negative media tone picks up a variety of information about the nominees, all sorts of bad news and possibly good news too. But a scandal is particularly bad news and it has a kind of double whammy on the editorialists, especially if it's a bad scandal.

Let's check this possibility by running some regressions on media tone and qualifications versus the content variables (scandal, objective qualifications, and extremism). These are shown in Table 16.4.

As you can see, the regressions suggest that perceived qualifications are largely driven by scandal: The presence of a scandal drops perceived qualifications by about 33 points. You will recall that a drop like this is usually enough to shift a nomination from consensual to conflictual. The coefficient for "Objective qualifications" has a plausible sign (although not statistically significant, being less than two standard errors from zero), so one might count this result as suggestive. "Extremism" seems to have the wrong sign, but its effect (if any) is measured very imprecisely.

Table 16.3. *Estimated coefficients (and standard errors) for coefficients in more regression models predicting Supreme Court nominees' perceived qualifications*

	Nominations without scandal	Nominations with scandal	All nominations (Model 1)	All nominations (Model 2)
Intercept	.90(.05)	.96(.10)	.90(.05)	.87(.03)
Tone	-.21(.24)	-1.23(.26)	-.21(.24)	—
Scandal	—	—	0.07(0.12)	—
Tone x Scandal	—	—	-1.01(.36)	-1.03(.15)
Deg of freedom	26	11	37	39
R <sup>2</sup>	.03	.66	.56	.54

Table 16.4. *Estimated coefficients (and standard errors) for coefficients in more regression models predicting two different outcomes regarding Supreme Court nominees*

	Perceived qualifications	Negative tone of media	Negative tone – model dropping one case (Clark)
Intercept	.70(.10)	.16(.07)	0.11(.07)
Scandal	-.33(.08)	.18(.05)	0.14(.05)
Objective qualifications	.065(.05)	-.06(.03)	-0.05(.03)
Extremism	.10(.12)	.17(.08)	0.24(.08)
Deg of freedom	37	37	36
R <sup>2</sup>	.36	.34	.36

*Note:* The regression attempts to estimate the effect of good and bad news on perceived qualifications and media tone.

In contrast, the second model suggests that negative media tone responds quite sensibly to good news and bad news about the nominee: Scandal raises negative media tone about 18 percentage points, as does extremism. Better objective qualifications are associated with a reduction in negative media tone, about 5 percentage points for every 1-point difference in qualifications. This is about the same magnitude uncovered in the previous regression, but it is estimated more precisely.

Examination of the residuals from the tone model uncovers little, except one moderately influential outlier (Clark again). The issue is that opposition support for Clark was unusually high, given the media tone. It might be worthwhile to recheck the Clark data to make sure that the coding was correct. However, if we drop the Clark observation and run the analysis again, the substantive results are very similar. So that you can see what I regard as little substantive change, I show this as the third model in the table. If the effects had been large and the data continued to look fine, I might consider using robust methods to down weigh the

observation without dropping it altogether. Fortunately, it doesn't appear to be very consequential. And this time, examination of co-plots finds little evidence of important interactions. Thus, a simple additive model seems to represent the structure in the tone data pretty well.

### VOTES RECONSIDERED

We now see that the variable perceived qualifications, coded from newspaper editorials, is largely a proxy for bad news in the form of scandals, as reported in news reports. So, suppose we go back to the original model that we devised for individual roll call votes and substitute the variable scandal for the variable perceived qualifications. What do you think happens? I leave this as an exercise for you to do. But perhaps you will believe me when I suggest that you will get substantively similar results.

### FITTING THE MODEL

Estimating parametric models for the data is not straightforward since the possibilities are numerous, once one admits the possibility of interactions and nonlinearities – and that's considering just three input variables! There is a tension between wanting to keep the models simple to avoid overfitting and wanting to capture evident patterns in the data. And many models will fit the data essentially equally well. But the different models will often have different substantive interpretations.

So, what should you do? In the absence of a theory strong enough to specify which variables to include and what functional form to employ – in other words, almost all the time in political science – we are well into “art.” Different people have different philosophies about this. In my view, you need to get to know the data well. This means using visualization aggressively and running a great many models but being extremely skeptical about each one. After a while, if you think hard about what you find, you will get a feel for robust patterns in the data. By “robust,” I mean patterns that almost certainly exist and that don't hinge on tweaking the specification in just the right way or including or dropping a critical outlier. In my view, you should report the models that highlight only these robust patterns, rather than the models with the best  $R^2$  or most impressive  $t$ -statistics. Then you can feel confident about the results, and you don't have to worry about everything falling apart the next time you get a new observation.

I should note a corollary to this “get to know the data” philosophy: It becomes almost impossible to really know the data once the number of predictor variables becomes large. The possible specifications proliferate beyond anyone's capacity to grasp. Political methodologist Chris Achen has formulated a “Rule of Three”: Don't believe regression results that include more than about three predictor variables. Because, almost certainly, the analyst doesn't really understand what's going on in the data (Achen 2002).<sup>10</sup>

<sup>10</sup> Here we are referring to predictors that you use to explain the outcome causally. It's certainly okay to include many predictors as control variables: for example, predictors such as age, sex, education, and ethnicity, which can capture some of the variability in a model of individual-level survey data.

To return to our example, let me suggest a few models for the aggregate voting data that satisfy Achen's Rule. (I urge you to try fitting some models yourself.) First, let's begin by fitting the opposition and copartisans separately, using tone and scandal. In other words, let's fit the top and bottom halves of Figure 16.4 separately. This will allow the intercepts and slopes to float freely, at the cost of smaller numbers of observations. Let's try the simplest additive model, and then include an interaction between tone and scandal. (I checked for nonlinearities in tone using a quadratic specification, but I won't show you the results since they don't suggest robust nonlinearities even in the lower right panel in Figure 16.3.)

These models are shown in Table 16.5. What are the take-away points? First, the intercepts in all four models look about the same, just as we saw in Figure 16.3. Second, in both populations, the main effects have the correct signs, seem to have plausible magnitudes (given what we know from Figure 16.4), and appear to differ across the two populations in the way we expect. They also are statistically significantly different from zero, except when the “Tone  $\times$  Scandal” interaction is included, in which case little is statistically significant. This is not unusual if the main effects and interaction display a high degree of multicollinearity. Especially in models with relatively small numbers of observations, it becomes difficult to measure effects precisely.

An obvious way to get more precise estimates is to pool the two populations while allowing some flexibility to capture important differences. An obvious strategy here is to force the same intercept across the two populations, estimate common main effects across the populations, and then estimate “add-on” main effects between the populations. This allows us to test whether the main effects actually do differ across the populations and to see how big the differences are. Because Figure 16.3 so strongly suggests an interaction, we can add the interaction to the basic additive model. But in the interest of avoiding overfitting, we can force the interaction to be constant across the populations. This is a middle path between overfitting and underfitting.

The results are shown in the rightmost two columns of Table 16.5. What are the main take-away points here? The tactic of pooling has increased the precision of the estimates without substantially changing their magnitudes from the earlier estimates. The coefficients on “Tone” and “Scandal” in Model 1 are essentially the same as those in Model 1a in Table 16.5. The coefficient on “Tone  $\times$  Copartisan” added to the “Tone” coefficient mimics that coefficient on “Tone” in Model 2a in Table 16.5, and similarly for the “Scandal” coefficients. As you can see, the magnitudes are about the same. We can test the difference between the two populations, captured by “Tone  $\times$  Copartisan” and “Scandal  $\times$  Copartisan.” As you can see, the differences appear to be real. The “Tone  $\times$  Scandal” interaction has the right sign and a plausible magnitude but is not statistically significant at the 95% level (the estimate is less than two standard errors from zero).

Are we done yet? No, because we need to examine residuals (always) and check for influential outliers. Doing so uncovers what one might expect: The two outliers we discussed in the lower right quadrant of Figure 16.3 are large and exert considerable leverage on the results. If we drop either or both of these

Table 16.5. Estimated coefficients (and standard errors) for coefficients in two regression models predicting total Senate votes (with the opposition party and the president's party considered separately, then pooled), given information on scandal and the tone of the media during the confirmation process

	Opposition		Copartisans	
	Model 1a	Model 1b	Model 2a	Model 2b
Intercept	100(5)	98(6)	101(2)	100(2)
Tone	-58(21)	-39(29)	-17(9)	-8(12)
Scandal	-27(8)	-18(13)	-7(3)	-2(6)
Tone x Scandal	—	-40(42)	—	-18(18)
Deg of freedom	37	36	37	36
R <sup>2</sup>	.50	.51	.28	.30

Table 16.6. The change in percentage of Yes votes in the Senate for Supreme Court nominees, comparing cases with and without negative media tone

	Copartisans		Opposition	
	No Scandal	Scandal	No Scandal	Scandal
Estimated separately, no interaction	-17	-17	-58	-58
Estimated jointly, no interaction	-16	-16	-60	-60
Estimated separately, with interaction	-8	-26	-39	-79
Estimated jointly, with interaction	-5	-34	-45	-74

observations, the size of the coefficients in Model 2b in Table 16.5 remain about the same (the shift effect of scandal becomes smaller and the “Scandal × Tone” interaction becomes somewhat larger). But the “Scandal × Tone” interaction becomes strongly statistically significant. A robust regression retaining the two outliers yields coefficients that differ little from those of Model 2b in Table 16.5.

I conclude that Model 2 is a pretty good representation of the aggregate voting data. In my opinion, we shouldn't be overly concerned about the lack of precision in the estimate of the Tone × Scandal interaction because the two outliers misleadingly raise the standard errors on this variable. Analysts who adopt a slavish attitude to *t*-tests would object and insist on the simple additive model. But if you have a good feel for the data, you'll believe that the simple additive model somewhat distorts what is the most likely pattern in the data.

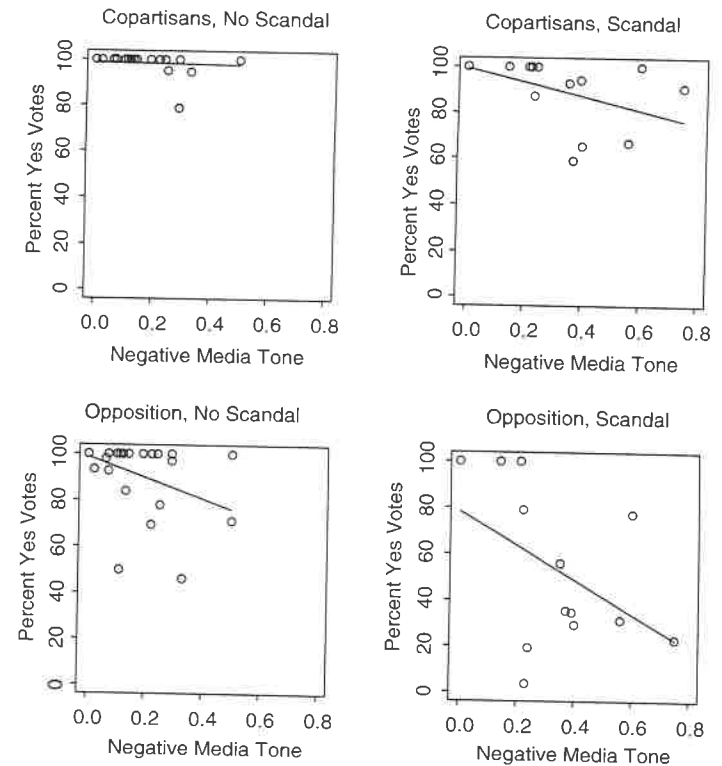


Figure 16.4 Predicting aggregate votes: the fit from the model in Table 16.5.

### INTERPRETING THE RESULTS

We still aren't done because we need to understand the substantive size of the estimated effects. This is somewhat complicated in the models with interactions. So, we'll do this in two different ways, first numerically, then visually.

Table 16.6 shows the marginal impact of tone on aggregate votes in the models we estimated in Table 16.5. To get these figures, I just took the partial derivative of aggregate votes with respect to tone. Equivalently, you can set the values of the indicators (scandal, copartisan) to 0 or 1 and add up the coefficients involving tone to get the corresponding marginal effects. This method will work in linear models like this.

As Table 16.6 shows, the effects of negative media tone appear to be substantial. Interestingly, the marginal effect of tone in the simple additive models is approximately an average of the separate “Scandal” and “No scandal” marginal effects in the models with the interaction.

Now consider a nomination without a scandal but with a terrible media tone. The estimated effects suggest that on average the nomination will nonetheless succeed, even when the opposition holds most of the seats in the Senate. But the nomination could be in serious trouble. On average, though, the size of the

coefficients suggests that it takes a scandal with adverse media tone to sink a nominee.

Before accepting these results about scandal at face value, though, we should add a condition. Accusations of financial improprieties, racism, or sexual misconduct mobilize the opposition and degrade support among the president's copartisans, especially when the accusations poison the coverage of the nominee in the media. That is the lesson from historical experience, as captured in the data and the models. But accusations of scandal are not the only thing that can mobilize the opposition. Some figures are so polarizing that the opposition behaves as if the nomination itself is a scandal. You can see this in the Bork nomination and also in that of L.Q.C. Lamar, the first southerner nominated to the Court after the Civil War. In both of these cases, the opposition acted as if it were in "scandal mode" despite the lack of any real evidence of misconduct by the nominee.<sup>11</sup> Nominating such figures can be problematic, at least during divided party government when the opposition holds a majority of seats in the Senate.

### CONCLUSION

Let me review what we have done so far. We started by reviewing earlier work on voting on Supreme Court nominees by individual senators. That work identified a key variable, perceived qualifications of the nominees, as indicated by newspaper editorialists. We then looked at aggregate voting patterns and found that, generally speaking, nominees get in trouble only when their perceived qualifications score is low. However, this happens surprisingly often.

We then took a closer look at the perceived qualification scores. We found that when the nominee is tarred by a scandal, his or her perceived qualifications score largely mirrors the tone of press coverage. Otherwise, the perceived qualification score is generally high. So, broadly speaking, the perceived qualifications score is mostly a proxy for scandal.

We then examined the tone of press coverage and found that it responds in a very sensible way to features of the nominee. In particular, better qualifications result in a more positive tone, accusations of scandal makes the tone of coverage more negative, and ideological extremism makes press coverage more negative. We finished by showing that the tone of press coverage, in tandem with accusations of scandal, works quite well in predicting aggregate voting on the nominees.

Where does this leave us? Again, what does it mean? The story that emerges emphasizes the media environment: This environment is critical for the success or failure of a nomination. If the environment is positive – composed mostly of good news – the nominee breezes through the Senate and receives a unanimous or near-unanimous vote. That's true regardless of the ideology of the nominee. But if the media environment turns nasty, the opposition party is likely to mobilize against the nominee. Support in the president's own party will weaken somewhat.

If the opposition is sufficiently numerous, the nominee can be in big trouble. The president may have to pull the nominee or risk defeat.

We used systematic data to paint a picture of what typically happens during Supreme Court nominations, but we haven't said much the calculations of political actors. However, our picture has strong implications for political strategy. In particular, management of the media environment is critical. If the president can craft a positive media environment, he can almost guarantee his nominee a seat on the Supreme Court. Conversely, if opponents of the nominee can poison the media environment, they have a real chance to derail the nomination.

How do you go about manipulating the media environment? This is what we will take up in the next chapter.

### Exercises

1. Devise a model predicting which nominees will have scandals.
2. Redo Figure 16.1, adding a kernel density smoother to the figure. Use a quantile-quantile plot to examine the distribution.
3. Download the updated roll call voting data from Lee Epstein's web page (<http://epstein.law.northwestern.edu>) and reestimate the original Cameron–Segal voting model. Can you improve it?

<sup>11</sup> That is, assuming one doesn't see as misconduct behavior like firing the Watergate special prosecutor, expressing a judicial philosophy far from the mainstream, or serving in the Confederate army. My point is, in both cases opposition senators acted as if they did consider such behavior scandalous.