Political Science

Lecture 3: Congressional Hearings and Supreme Court Nominations

1. Introduction

In the last lecture, we concluded that manipulating the media environment is a central concern for strategic players in the "game" of Supreme Court nominations. From this perspective, the President and his allies must try to establish a favorable media environment for the nominee. Individuals opposed to the nominee and President must do the reverse: establish a hostile environment. Of course, a favorable environment may include "flying under the radar" – no attention in the media, or very little.

How can strategic actors manipulate the information environment? A useful distinction is between *direct* and *indirect* manipulation of the media environment. By direct manipulation I mean, for example, the president "going public" (making speeches) on behalf of the nominee, interest groups staging protests and demonstrations, and members of the Judiciary Committee shaking their finger at the nominee and asking "gotcha" questions about judicial doctrine. These are direct manipulations because the media may deem them newsworthy and broadcast them or report on them.

Indirect manipulation is somewhat more subtle. The idea is that you engage in actions that make direct manipulations (your own or other's) more or less effective, or encourage reporters to write the "right" kind of stories on their own. Consider, for example, the president's choice of a moderate versus an extreme nominee, or a highly qualified one versus a marginally qualified one. These choices are indirect manipulations of the media because they affect how the media portray the nominee. These choices also make the president's statements about the nominee more or less credible (at least arguably). As we saw last week, selection of a moderate, able nominee will encourage favorable press reports while selection of an extremist hack will tend to generate negative publicity. As another example of indirect manipulation, interest groups compile incriminating statements – e.g., "The Book of Bork" assembled by anti-Bork activists – and turn them over to the press. Thus, the interest group does the work investigating the nominee for the press, which leads to increased press coverage of the kind the group wants.

What about the Judiciary Committee? How can the Chair or the majority indirectly manipulate the press? Suppose the Chair is hostile to the nominee. Then an obvious way to indirectly manipulate the media is to do a thorough investigation, versus a superficial job, and uncover some really damaging evidence of improper behavior. (We saw last time the huge impact of scandals on media coverage.) Then, during the hearing itself, the Chair and like-minded members of the Committee can beat the nominee over the head with his or her transgressions. Thus, the indirect manipulation – searching for a scandal – enhances the direct manipulation, the public hearings.

One might suppose a friendly Chair should do the same, investigate thoroughly to reveal how clean the nominee is, then publicize it during the hearings. But, as we saw, when it comes to Supreme Court nominations no news is good news: absent a scandal and a hugely hostile Senate, a nominee is practically sure to be confirmed. So a better bet for a friendly Chair is to move the nominee through the process in such a way that he avoids accusations of scandal. In practice, this means, hustle him through as fast as possible.

In this lecture, I'd like to explore these ideas in more detail. We'll take a close look at the duration and intensity of public hearings during Supreme Court nominations. I'll try to convince you the public hearings are longer and more intense when 1) the Chair of the Judiciary Committee is hostile to the nominee, 2) the Chair has some red meat to work with, that is, when the prior search or good luck uncovered a scandal or other bad conduct, and 3) staff resources are more abundant, so prolonged hearings are less costly of effort to senators. We'll also find some evidence suggesting hearings became longer and more intense beginning in the late 1960s, in the wake of the controversial Warren Court.

You may be willing to believe all this already, without looking at any data or models. So at some level the findings may not be terribly surprising. But my point isn't to unfold a shocking revelation about American politics, though I do want to explore this material carefully. Instead, I want to return to a theme I introduced in the first lecture: the interplay of theory and data. I'm going to try to show you, in a simple way, how this works.

The analysis proceeds in six steps. First, we'll probe the logic of the situation. To do so, we'll develop a simple formal – that is, mathematical – model of the Chair's behavior. (I claim the model is a more general representation of congressional investigations of the executive.) We'll interrogate the model deductively to generate specific empirical predictions. Thus, we'll know in a very precise way how a set of theoretical assumptions map into empirical hypotheses.

Second, we'll try to measure the key variables indicated by the theory. This will involve a fair amount of work (all too common, I'm afraid). Third, we'll search the data for structure, looking for the predicted patterns but also for unexpected ones. As before, we'll rely on artful visualizations and highly flexible, non-parametric modeling.

Fourth, we'll fit parametric models to the data, "testing" the formal model's predictions. We'll use models appropriate for count data and briefly discuss some issues that arise there. Fifth, we'll ask what could go wrong with the analysis – how hard should I believe the theory?

Finally, we'll address some normative questions. Given what we've found, how should we evaluate the kind of partisan behavior emphasized by the theory? We'll just gesture in this direction but, I hope, provide some food for thought.

2. A New Theory of Congressional Hearings

What do political scientists know about the length and intensity of congressional hearings? Is there a standard model for the process? In fact, there is an empirical literature that examines how frequently Congress holds hearings and what the hearings are about (Aberbach 1991 and 2002, Jones and Baumgartner 2002). There is a small theoretical literature that treats hearings as signals from a congressional committee to an agency about the committee's interest in a matter (Ferejohn and Shipan 1989, Cameron and Rosendorff 1993), or as a signal from a committee to floor members of Congress about the committee's knowledge of an issue (Diermeirer and Feddersen 2000). None of the theoretical works focus on hearings as an opportunity for congressmen to gain publicity or to damage the investigation target. So let's build a new model of public hearings from scratch, bearing in mind the case of Supreme Court nominees. Our theory will be based on the idea of publicity-seeking partisans who use hearings to generate news stories.

We'll begin by assuming that 1) public hearings generate negative messages about a nominee, 2) negative messages bring a reward to the Chair of the committee, at least when he is ideologically estranged from the nominee, and 3) public hearings are costly to the Chair to conduct, because of time and effort costs and forgone leisure. We will assume that the objective of the Chair is to maximize his net rewards, that is, maximize rewards less effort costs. So, the pieces of our model are *negative messages*, a *production function* for creating negative messages, a *reward function* from generating negative messages, and a *cost function* for producing negative messages. We'll think a little harder about each of these pieces, then put them all together and consider the Chair's behavior in the model.

Let's denote the number of negative messages as n. And, we'll denote the negative message production function as $n = f(x;\theta)$, where x indicates the duration of the hearing and θ the stock of "bad stuff" known about the nominee. We'll take θ as fixed – it reflects the prior work by the committee staff and interest groups in digging up dirt on the nominee. The Chair "builds" negative messages by extending the public hearings. So, negative messages increase with the length of the hearings. A seemingly sensible assumption is declining marginal productivity – longer hearings mean more bad messages, but at a declining rate from each additional day of hearings. Another sensible assumption is the more ammunition the Chair has to use against the nominee (the bigger is θ), the greater the number of bad messages, for the same level of effort. In addition – and this is critical – we assume "increasing differences" in duration and dirt. A figure is useful at this point, especially to explain increasing differences.

Examine Figure 1, which illustrates the type of production function I have in mind. The lower curve shows the effect of longer hearings (*x*), at a lower level of dirt (θ), the higher curve the effect at a higher level. You can see the declining marginal effect of hearing duration with each curve, in that they tend to flatten out. And, you can see increasing differences at work: as duration increases, the difference between the high dirt curve and the low dirt curve increases. At any duration, the marginal effect of duration is greater when θ is greater.[1] An example of a function like this is $n = \theta \log(x)$, which is the function in the figure.

OK, now let's consider the return to the Chair from the production of negative messages. We'll assume that the chair's return is a function of negative messages – and that the effect of negative messages depends on the ideological distance between the Chair and the nominee. Let's call this ideological distance δ. Then we can denote the return to the Chair from negative messages as $r(n;\delta)$. How do we imagine this return function works? If the Chair is close enough to the nominee, he doesn't want any negative messages about the nominee: $r(n;\delta)$ is negative for

[1] If the production functions are differentiable, the increasing differences assumption is just

$\frac{\partial^2}{\partial x \partial \theta} f(x,\theta) > 0$: more dirt raises the marginal increase in negative messages from longer hearings.

positive *n* if δ is "small" enough. But beyond a certain degree of ideological estrangement, $r(n;\delta)$ will be positive for positive *n*. Plausibly the greater the ideological distance between the Chair and nominee, δ, the more the value to him of negative messages.

Again, it may be helpful to look at a picture, so examine Figure 2 which illustrates a return function. The lower curve is the return from negative messages to a Chair who is ideologically proximate to the nominee. As shown, the return to him is negative so he wants no negative messages about the nominee. The middle line is the return function to a Chair who is somewhat estranged from the nominee, say, a moderate Democrat facing a somewhat conservative Republican nominee. The top curve is the return function for a very liberal Chair who faces a very conservative nominee. An example of a return function like this is: $r(n;\delta)=(\delta-\bar{\delta})n$, where $\bar{\delta}$ represents the critical distance at which the return function switches from negative to positive.

Our third building block is the cost of holding public hearings. Here we'll assume a cost function $c(x;\kappa)$ in which costs rise with the duration of the hearing but decline with, say, the abundance of staff resources κ. Presumably as well, the cost of zero hearings is zero, that is, $c(0)$ = 0. An example of such a function is x/κ.

Alright, we now have our building blocks. Let's assemble them to describe the Chair's objective. Pursuing the idea that the Chair is interested in the net returns from hearings, we have

$$\Pi(x,\theta,\delta,\kappa)=r\big(f(x,\theta),\delta\big)-c(x,\kappa) \qquad (1)$$

This is just the return function minus the cost function. We imagine the Chair wishes to maximize this function by choosing a hearing duration *x*. In Figure 3, I've used the component equations I mentioned above to construct a specific net return function at two different levels of θ. As you can see, at least in this example, there is a clear optimal length for each level of dirt. Call this length x^*.

An obvious question is, when will equation (1) have a well-defined maximum? This is not a trivial matter and has been the subject of much thought by mathematicians. You can probably see that if the net return function is continuous and the allowable hearing duration is a continuous closed interval, like 0-6, the net return function must have a maximum value in the interval. In such a case, if the net return function is differentiable, you can find the (or at least a) maximum using simple calculus. Things can get more complicated though if, for example, the net return function has jumps or the allowed duration of hearings takes only integer values, and so on.

At this point we could go off on a long digression on the theory of optimization but instead let's keep our eye on the theory-empirics link. To do so, we need to think harder about the optimal hearing duration *x**. Presumably *x** is the length of public hearings the Chair will choose, at least that's our assumption. We can observe hearing lengths across different nominations, which will become the dependent variable of interest. Now, the optimal hearing length will be a function of the exogenous variables in equation (1), specifically, 1) the amount of dirt θ, 2) the ideological distance δ between the Chair and the nominee, and 3) the size of staff resources κ. In other words, optimal length is itself a function $x^*(\theta, \delta, \kappa)$. Over a series of nominations, we may be able to observe the values of these variables, which will become our independent variables. So if we can coax our theoretical model into making predictions about the behavior of the function $x^*(\theta, \delta, \kappa)$ as θ, δ, and κ change, we will have some testable predictions relating our dependent and independent variables.

Let me introduce a little jargon. Assuming that $x^*(\theta, \delta, \kappa)$ is differentiable (which we don't actually have to assume), the predictions we want concern $\frac{\partial x^*}{\partial \theta}$ (the change in x^* as θ increases), $\frac{\partial x^*}{\partial \delta}$ (the change in *x** as δ increases), and $\frac{\partial x^*}{\partial \kappa}$ (the change in *x** as κ increases). These three partial derivatives are called "comparative static multipliers," which is probably not completely transparent terminology. But the terminology is too well established to change, so let's go ahead and use it.

From a theoretical point of view, the comparative static multipliers summarize the model's empirical content – for most practical purposes, they *are* its empirical content. My good friend Rebecca Morton, in her useful book *Methods and Models*, discusses some other ways to test theoretical models in the social sciences (Morton 1999). But in general comparative statics is the meat and potatoes of theoretically driven empirics. In fact, as a broad rule: no comparative statics, no empirical content.

So, our problem is to derive the comparative static multipliers for our model. I should note, typically we will not be able to establish exact numerical values for the comparative static multipliers. For example, we will not be able to say, "The theory predicts that if the stock of bad news about a nominee goes up by one unit, the hearing duration will increase by 5 hours." Predictions like this demand too much from theory. But we may be able to get predictions about the *sign* of the multipliers – e.g., whether x^* gets larger or smaller or stays the same (or, we can't tell) as the stock of bad news about a nominee increases.

Well, how do we derive the comparative static multipliers in our model? Basically, there are three ways to do it. The first is to derive a closed form solution. This is the easiest method to understand. The idea is to assume specific functional forms for the production function, return function, and cost function, and then actually solve out for the function $x^*(\theta, \delta, \kappa)$. If we are careful about specifying the underlying functions, this may be possible. Then, given an explicit $x^*(\theta, \delta, \kappa)$ function, we can see how x^* changes as the exogenous variables change, either through inspection, calculus techniques, or direct calculation.

I will work out a closed form solution in a minute, but the obvious difficulty with this approach is that the results are hostage to the specific equations we assume. Consequently, there may be little or no generality to the results. Perhaps if you used a somewhat different but equally plausible return function (for instance), you might get completely different predictions. And, any test of the model based on a closed form solution will really be a test of the full set of assumptions, i.e., not just the general equation (1) but all the specific forms for each of the components.

A second way to go is to use some powerful results in the theory of robust comparative statics. In recent years, economic theorists have worked out general results concerning the behavior of equations like (1).[2] If you feel comfortable making some rather broad assumptions about the component parts of (1) – the kind of assumptions we made earlier, like "increases in x and θ with increasing differences between x and θ" – you can apply powerful theorems directly to the problem and immediately get the comparative static results. I was careful to make the right kinds of assumptions as we went along, so we can do this. I won't go into too much detail about this, but a nice implication of this approach is that, because we know the results are perfectly general so long as we maintain the underlying assumptions about the component equations, we can use any tractable functional forms we want and derive closed form solutions whose behavior is (in a qualitative sense) perfectly general. These in turn can supply specific estimating equations. (You may want to think about the logic of this for a minute.) So combining the two methods is very attractive when taking theory to data.

A third method is somewhat intermediate between the specificity of the first approach and the generality of the second. Here, you make some general assumptions about the shapes of the component equations but not their specific form, and also some strong assumptions about their smoothness and continuity. Then you can use calculus-based methods to work out the comparative statics. This approach typically involves quite a bit of algebra so we won't do any more with it.

---

[2] For example, see Milgrom and Shannon (1998). An accessible introduction written for political scientists is Ashworth and Bueno de Mesquita (2006). An unpublished monograph, Athey, Milgrom, and Roberts (1998), available at http://post.economics.harvard.edu/faculty/athey/athey.html is a lovely primer.

Alright then, let's assume the specific functional forms I used earlier, since we know the results we'll find in this case are actually quite general. Thus, equation (1) becomes

$$\left(\delta - \bar{\delta}\right)\theta \log(x) - x/\kappa \tag{2}$$

This function is what is shown in Figure 3, for the values $\delta = 3$, $\bar{\delta} = .7$, $\kappa = 1/2$, and $\theta = 3$ and $\theta = 4$. (I picked these values through a little trial and error, to get pretty shapes for the curves.) As you can see, there is a clear optimal hearing length, the point at which the net return is highest. And, this optimum length increases as the amount of available "dirt" increases. Make sure you understand this point: the optimal length (the point where the curve is highest) shifts to the right as θ increases. Simply from inspection of the curves, we can see our first comparative static in action.

I could re-do the picture varying each of the parameters in equation (2) to give a pictorial version of the comparative statics. Instead, let's go ahead and derive an expression for the optimum length. Using a little algebra,[3] I find this to be

$$x^*\left(\delta, \bar{\delta}, \theta, \kappa\right) = \left(\delta - \bar{\delta}\right)\kappa\theta \tag{3}$$

Now you have to admit that's pretty neat! The optimal length is simply the multiplicand of the three relevant parameters (counting $\left(\delta - \bar{\delta}\right)$ as a single parameter.)

An obvious point that jumps out of equation (3) is that if the Chair is close enough to the nominee (if $\delta \leq \bar{\delta}$) the chair will not want to hold any hearings at all: x^* is zero (obviously, hearing length can't be negative). Or, if zero hearings aren't feasible, the Chair will hold the shortest feasible hearings. Before the late 1940s or so, the Senate Judiciary Committee sometimes allowed Supreme Court nominees to go forward to floor votes without holding public hearings.

---

[3] Take the first derivative of (2) with respect to *x* and set this equal to zero (i.e., assume an interior maximum), then solve for *x* at the interior maximum.

After the late 1940s the Committee stopped doing this – seemingly, having no public hearings at all appeared unacceptably cavalier. But even then some hearings were very brief, pro forma affairs. We'll look at some data in a minute.

Using equation (3), we can get our comparative static predictions directly.[4] When $\delta > \bar{\delta}$ and $\theta, \kappa > 0$:

$$\frac{\partial}{\partial \delta} x^*(.) = \kappa\theta > 0$$

$$\frac{\partial}{\partial \theta} x^*(.) = (\delta - \bar{\delta})\kappa > 0 \qquad (4)$$

$$\frac{\partial}{\partial \kappa} x^*(.) = (\delta - \bar{\delta})\theta > 0$$

In words, when the Chair and nominee are estranged ideologically, hearing duration lengthens as the degree of estrangement deepens, the stock of dirt increases, and the committee staff swells. If we just used the robust approach, we would get similar results, except in general each of the equalities would be weak (greater than or zero, rather than greater than). Note that if $\delta \leq \bar{\delta}$ (the nominee is close to the Chair), these all become (weakly) negative, implying a minimal hearing length.

We'll take the model to data in a few minute. We'll proceed in several steps. First, we'll visualize the relationships non-parametrically using extremely flexible functional forms. This is a good way to listen to the data – we'll let the data talk to us without forcing it into any procrustean beds. This approach fits neatly with the robust comparative statics philosophy: we eschew specific functional forms and look for the broad patterns. But we'll also estimate parametric equations inspired by our closed form solution. This is attractive since (3) is so simple, and the parametrics

---

[4] Just take the partial derivative of (3) with respect to the parameter of interest.

allow us to do standard diagnostics on this regression, e.g., making sure the results are robust to influential outliers, and so on.

Should I summarize where we are? We have developed a new theory of congressional hearings, with an eye toward Supreme Court nominations. Our theory is based on the idea that congressmen are ideological partisans. According to the theory, the Chair of the Judiciary Committee drags out the hearings in order to publicize nasty bits of information about an ideological enemy. In contrast, the Chair hustles an ideological soul mate through the hearings in order to limit any damage to him. Thus, public hearings on Supreme Court nominees are not a seminar in a philosophy class or law school. Rather, they are a kind of political warfare, at least when the Chair and the nominee are ideologically estranged. Our key independent variable is the ideological distance between the Chair and the nominee. Also important is the availability of dirt on the nominee. Under some reasonable and fairly general assumptions, we worked out the empirical implications of our theory. Because we did this mathematically, we can be sure our logic is tight: our empirical hypotheses are nailed down quite nicely, thank you.

OK, then, let's turn to the data.

3. Data: Definitions and Overview

The data come from all nominations from Hugo Black (1937) to Samuel Alito (2005). Two variables, *length of hearing* (duration in days) and *volume of the hearing report* (length in pages) represent the intensity of hearings and are our dependent variables. Our theory suggests three variables as independent variables: 1) a measure of ideological distance between the nominee and the Chair of the Judiciary Committee, 2) a dummy variable for scandal, measuring the "dirt" or bad news about the nominee at the time of the hearing, and 3) the number of Judiciary Committee staff. In addition, a little knowledge of history suggests a fourth: a dummy variable for Senate Courtesy (explained shortly).

Length of Hearing

The first measure of intensity of public hearings is the number of hearing days of hearings for each Supreme Court nominee, as reported in Rutkus and Bearden (2006). This measure is the actual number of days on which public hearings were held. For example, in the case of the Powell nomination, the public hearing started on 11/03/1971 and ended on 11/10/1971, a total of eight days. However, hearings were not held on 11/05, 11/06, and 11/07. Thus the recorded hearing days are five days rather than eight days.[5]

// Insert Figure 4 about here //

Using this measure, the mean duration of hearings was 3.3 days and the median 2 days. The minimum length was zero days, the maximum 12 days (the Bork nomination), and the variance was a little over 10 days. The left hand panel of Figure 4 shows the distribution of hearing lengths in days; the right hand panel shows the distribution in logged days (actually, days plus one, so that a hearing of zero days results in a logged duration of zero). The figures are histograms adjusted to read as a probability density (in other words, the sum of the bar heights times the bar widths equals one). In addition, I have plotted a kernel density smoother over the bins of the histograms, providing another measure of the shape of the data as a probability density. This is the thick, curving line.

Histograms and kernel density smoothers are a great way to visualize the shape of a distribution of data. But they require a bit of experimentation – how many bins to use, how smooth to make the estimated empirical density, and so on. In this case, with a little experimentation I set

[5] Because the nominations of Douglas Ginsburg and Harriet Meiers were withdrawn before hearings, they are not in our data set. In addition, the hearing report for the Alito hearings has not been published at the time we write, so this observation is missing for the analysis of page counts.

the bin widths and the degree of smoothing to what seem to me to be intermediate levels – not too smooth, but fairly smooth.

If you focus on the left panel, using the raw data, you can see that the data do not look normally distributed at all. Instead, they look like they were generated by some sort of exponential distribution, perhaps a Poisson distribution with a bit of over-dispersion (more on this later).[6] It would take a bold (unreflective?) analyst to fit an OLS regression to data that look like this.

Taking logs – shown in the right hand panel – makes the data look more normal. Actually, though, the kernel density smoother hints that there might actually be two or even three distinct populations in the data, each more or less Poisson distributed. Rather looks that way, doesn't it? Any ideas on what those populations might be? We'll come back to this in a while.

Volume of the Hearing Report

The Senate Judiciary Committee publishes an official report on each hearing. The report contains transcripts of the witness's testimony, senators' questions, and the nominees' answers, plus various items the senators want to put on the record. It seems natural to view the volume of the report as proportional to the intensity of the public hearing.

The average number of pages in the committee reports is 454 while the median length is 128. The minimum length is zero and the maximum is 3350 (again, the Bork nomination). The variance is huge.

Figure 5 is just like the previous figure, but shows the distribution of the pages data. Again, the left-hand panel shows the raw data, the right-hand panel the data logged. At this point it should

[6] "Over-dispersion" means the data is spread out more than one should see in the reference distribution, here the Poisson distribution. An interesting feature of a Poisson distribution is that the mean and variance have the same value. So you can see that this data does appear somewhat over-dispersed, but not hugely so. We'll do a test in a while.

be no surprise that the raw data look very non-normal. In this case, they look very over-dispersed. Taking logs eliminates the long right hand tail and again hints that there may be two distinct populations.

// Insert Figure 5 about here //

Ideological Distance

The theory indicates that the more distant the nominee's ideology is from that of the Judiciary Chair, the longer the hearing will be (ceteris paribus). To estimate the distance from the nominee's ideology to the Judiciary committee chair, we need to measure the ideology of the Chair and the nominee and put them on the same scale. This takes some work!

Assigning a score to the Judiciary Chair is quite straightforward. We first identified the Judiciary Committee chairs at the time of each nomination and then located their DW-NOMINATE scores in Rosenthal and Poole's dataset.[7] I mentioned these scores in the first lecture. Poole and Rosenthal's DW-NOMINATE scores, based on all recorded roll call votes in Congress, are widely used by political scientists as measures of congressmen's ideologies and are generally considered highly reliable. These are shown in Table 1.

// Insert Table 1 about here //

The process is much more complicated for the nominees and is going to take a little explaining. First, given our use of a roll call based measure for senators, it might seem natural to do the same thing for Supreme Court justices. Then, given something like a DW-NOMINATE score

---

[7] Available at www.voteview.com

for a justice, we could try to put the nominee measure and the senator measure in sinc somehow. In fact, several scholars have tried to do this (Bailey and Chang 2001, Epstein et al 2005).

I believe voting scores for Supreme Justices have their uses, but they are not well-suited for studying nomination politics. First, there are no voting scores for failed nominees since they never made it onto the Court (obviously). So that creates a problem right off the bat. Second, even for the successful nominees, we are more interested in the way they were perceived at the time of the hearings then how they actually voted later on the Court. The two are surely related, but they are not the same thing. Finally, there are good theoretical and historical reasons to believe that the voting behavior of Supreme Court justices may not be a simple expression of their preferences. For one thing, their votes surely reflect the changing docket of cases, which they themselves choose.

To see why this is a problem, consider the voting of an ideologically moderate justice when a conservative majority controls the docket. This justice, who will often split from the conservative majority and vote with liberals on the Court, is apt to look rather liberal himself. But if a block of liberals controlled the docket, this same moderate justice would often vote with conservatives, thus making him look rather conservative. So, you get an artificial extremism that fluctuates with the docket. But this is hardly the only problem with voting scores.

In addition to docket effects, the voting behavior of justices may reflect the political environment facing the Court. History suggests that the justices often moderate in the face of political opposition from the other branches (Rosenberg 1992). In my opinion, given a largely or completely endogenous docket and the possibility of strategic voting by the justices, it takes a lot of faith to see the justices' voting scores as a pure measure of their ideologies. In fairness, similar criticisms have been directed at the congressional DW- NOMINATE scores, but despite some rather intense work on the subject, no one has been able to show that docket effects or strategic behavior contaminate those scores in a profound way.

The bottom line is, for the nominees we need something other than their subsequent voting scores as a measure of their ideology, and we want this measure to be scaled like DW-NOMINATE scores. What to do?

Here is out solution. For the nominees who served in Congress, we start with their DW-NOMINATE scores. More specifically, five nominees (Black, Byrnes, Burton, Vinson, and Minton) served as legislators before they were nominated, so we can directly use their DW-NOMINATE scores as a measure of their ideology.[8] For those who did not, we start with the NOMINATE score of the president who nominated them.[9] This is hardly perfect but it seems reasonable to believe that the ideology of the nominating president contains valuable information about the nominee. To complement this measure, we also employ the nominee ideology scores calculated by Jeffrey Segal and Albert Cover (1989). I mentioned these in the second lecture. These scores are based on contemporaneous editorial evaluations in four major newspapers (two liberal, two conservative), *The New York Times*, the *Washington Post*, the *Chicago Tribune*, and the *Los Angeles Times*. Roughly speaking, they indicate the percentage of editorial evaluations that score the nominee as a liberal.

We believe both sets of scores tap into the nominee's true ideology, though probably with error. To recover the posited common factor underlying the two measures, we performed a Principal-Component Analysis (PCA) on the two measures. Then, to put the PCA scores into the DW-NOMINATE space, we simply regressed them on the DW-NOMINATE measure associated with each nominee (their score or the president's score). Using the estimated coefficient, we then converted each PCA score into a DW-NOMINATE score. Essentially, we take a DW-NOMINATE

---

[8] Aficionados of ideology scores will note that we converted all of the scores into the Senate space.

[9] Because presidents announce their position on many bills, they have been scaled just like senators or House members.

score for each nominee (either his or her own, or that their nominating president), then adjust it based on contemporaneous perceptions of the nominee.

The recovered DW-NOMINATE scores are shown in Table 2. Negative scores connote liberals while positive scores connote conservatives, and the scores lie between -1 and +1 (so do DW-NOMINATE scores). Generally speaking, the scores have a great deal of face validity, that is, justices we know were liberals show up that way, and similarly for conservatives and moderates. As a side note, these scores do a pretty good job predicting the future voting behavior of the nominees who make it onto the Court.

// Insert Table 2 about here //

Finally, given the Chair's score and the imputed score of the nominee, we calculate the simple distance between them (the absolute value). Given the range of observed DW-NOMINATE scores, the distances range between zero and one. The mean distance is .34 and median distance .31. The minimum distance is virtually zero and the maximum almost one. The variance is .06. We show the distribution of the ideological distances in Figure 6. The data is single peaked, with a long right tail. Taking logs pulls in the tail somewhat.

// Insert Figure 6 about here //

Scandal

The theory indicates the importance of some "dirt" for the Chair to work with. We will use the presence of a scandal as our measure of the stock of dirt. According to our data, coded from stories in *The New York Times*, thirteen nominees experienced some sort of scandal or scandals. Since we are interested in the effect of scandal on the length of hearing and the volume of the hearing report, the actual timing of the scandal is critical. We classified the nominees with scandal into three categories according to this timing (See Table 2).

// Insert Table 2 about here //

We do not score late-breaking scandals – those that happened after the public hearings had concluded – as an effective scandal for our purposes, since we seek a measure of dirt existing at the time of the hearings. We do count the scandal involving Black, as it emerged between the time Roosevelt nominated him but before he was confirmed, which occurred without a hearing. A trickier case involves the scandals that emerged during the hearings. If one suspects that the Chair engineered their emergence, or that they had an effect on the subsequent duration or intensity of the hearings, one should count them as dirt. But if one views them as non-engineered and not liable to affect the choreography of the hearings, one might not count them. So, one might count all four of the "hearing scandals" as scandals, or none of them, or just the two that occurred prior to the last day (indicated in the table). In our data analysis we opt for the first of these choices, but we re-ran all the results using the other coding choices, and it doesn't make much difference to the results.

// Insert Table 2 about here //

Staff Size

For the size of the Senate Judiciary committee staff, we began with Keith W. Smith's compilation of data on congressional committee staff (1947-1998).[10] Smith compiled the data from standard sources. We have updated and checked these figures, using staff directories published by Congressional Quarterly. For pre-1947 nominations, for which we've had trouble locating exact numbers, we use the number for 1947. Figure 7 displays the data over time. As you can see, the

---

[10] Available on his personal web site, http://trc.ucdavis.edu/kwsmith .

size of the committee staff has varied a lot over time. Basically, staff increased steadily until the Stevens nomination in 1975, and then fell to a moderate and fairly stable level. So this is not a story of simple growth.

Senate Courtesy

Supreme Court nominations are usually referred to the Senate Judiciary Committee, which typically schedules a public hearing on the nominee. However, by unanimous consent the Senate can directly consider and confirm a nomination without referral to the Judiciary committee. This fast-track nomination process, called Senate Courtesy, has been invoked for a few nominees, invariably former or current senators. For example, Byrnes did not go through the hearing due to Senate Courtesy. Of course, Senate Courtesy is hardly an inviolable rule (we find it hard to imagine it would be invoked today). In fact, some former or current senators were nonetheless sent to the Judiciary committee for hearings. However, in these cases the nominee usually had a nominal hearing and was quickly sent to the full Senate for a vote. We identify Black, Byrnes, Burton, Vinson, and Minton as filling the historic requirements of Senate Courtesy.

4. Visualizing the Data

As we've emphasized repeatedly, visualizing your data is an essential part of any analysis and a cornerstone of modern applied statistics (Cleveland 1993). This is still true when you are testing a formal model, because you want to be fair with the data. By this we mean you want to be confident the predicted patterns are really there. In addition, you want to know if the data contains interesting patterns that the theory doesn't predict. Surprises are great, because they are food for thought. They may even stimulate you to create a new theory. And there's nothing like clever visualizations for uncovering surprising patterns (also for finding coding errors or other mistakes, which happen all the time).[11]

So, where do we start? I like a scatter-plot matrix with non-parametric smoothers imposed on each panel: simple to do but often very effective. This is shown in Figure 8.

// Insert Figure 8 about here //

Recall that each row has the indicated variable on the *y*-axis and each of the other variables on the *x*-axis. For example, look at the bottom row in the figure. The *y*-variable here is days of hearings (logged). Log-days is arrayed against each of the other variables in turn. It's nice to see log-days and log-pages track each other pretty tightly, suggesting both tap into the same thing, presumably hearing intensity. In fact, you will observe that all the key independent variables (taken singly) appear to have a strong impact on the duration of the hearings and, in the next to bottom row, log-pages. In the next to bottom row, you can see what looks like a threshold effect from staff on log-pages. Looking again in the bottom row, you may see something similar for duration, now that you look harder. In other words, once you get beyond a certain level of staff (130 or so), additional staff doesn't seem to have much effect. Now that we're a little sensitized to threshold effects, look again at the effect of distance on days and pages. This is interesting since the distance-duration relationship is central to the theory. Again, it looks like we're seeing a threshold effect. This is gratifying, because our theoretical model predicted it (remember minimal hearings for proximate nominees?). It's too early to break out the champagne, but the patterns in the figure look promising.

---

[11] If you use Excel spreadsheets to manage your data and use the cut and paste commands, it is very easy to make a mistake. Also if you are careless when sorting on a variable, its easy to scramble your data. It's also extremely easy to mess up when merging data sets or creating new variables. *Whenever you do this, you should look at your data!* Otherwise, you are asking for trouble.

Examining the last column, you will observe that most of the variables increase over time, though staff turns down again (we saw that earlier). This co-movement over time is a little worrisome because it suggests we may have trouble untangling all the relationships. I will return to this point later. Skimming the matrix for other strong relationships, I see that staff increases with the distance between the Chair and the nominee. This is hard to take literally, but it suggests the Committee has more staff when it is estranged from the executive. It might be worth thinking harder about potential endogeneity, but we won't spend more time on this here.

So, despite a few worrisome issues, our initial cut at the data looks promising. What would be a potentially illuminating visualization, given our theory? Personally, I want to see the relationship between duration and ideological distance, for scandal ridden nominees vs. non-scandalous ones, and for staff-poor vs. staff-rich nominations. This is a natural application of conditioning plots, which I introduced in the last lecture. Let's use this nifty tool again.[12]

// Insert Figure 9 about here //

Figure 9 shows log-days as a function of the ideological distance between the Chair and the nominee. The left-hand panel shows the relationship for nominations in which no scandal emerged before or during the hearings. As shown, when the chair and the nominee are ideological soul mates (ideological distances under about .3 or .4 in DW-NOMINATE units) hearing durations are usually very short. Nor, in these circumstances, are durations particularly responsive to distance. But when the Chair and the nominee are estranged (ideological distances greater than .4 or so) the durations get noticeably longer as ideological distance increases.

---

[12] We will be using W.S. Cleveland's wonderful suite of graphics programs, which he calls trellis graphics. This is available in S-Plus or R. You can coerce other programs to do something similar, if you work at it.

The right hand panel in Figure 9 shows data from nominations with scandals, and there the situation is even more dramatic. The hearing duration is much longer even for ideological soul mates and goes up from there. In fact, the duration is longer for the closest scandal-ridden nominee than it is for the more distant scandal-free ones. Beyond the .4 mark, the slopes of the two lines don't really look that different to me, suggesting a merely additive effect of scandal, rather than an interaction. One caveat worth noting: as you can see in the panels, most of the scandal-free nominations involve ideologically proximate nominees. The scandal-ridden nominations span the ideological spectrum.

// Insert Figure 10 about here //

Now let's check out the effect of staff, as shown in Figure 10. Again the figure shows the duration-distance relationship, but now the left hand panel shows nominations in which the Committee staffing was relatively skimpy (below 133, with a minimum value of 27). The right-hand panel shows the relationship for nominations in which staffing was relatively abundant (above 133, with a maximum of 251). In the left-hand panel you will notice that all the ideological distances are rather small. So, the figure says: when distances are small and staffing tight, durations were typically short. It is tempting to add: "and, they don't vary with distance." But we have to careful here since because of the limited range of the data – for really big distances the durations might tick up, but we can't tell since we have no observations in that range.[13] The right-hand panel

[13] Analysts are often pretty sloppy about projecting relationships outside the range of their data. Relying exclusively on tables of regression coefficients encourages this often risky practice. The defaults in the "trellis graphics" package I'm using won't project the smoothers outside the data, which is nice. More generally, when showing the fit from a regression, it is a good idea also to show the actual data points if possible or, barring that, a "rug" at the bottom of the

has observations spanning the ideological spectrum. We see both a shift upward in the curve, and a notably positive slope. Again, all this looks quite consistent with our theory.

An obvious question now is: what happens if we vary staffing and scandal at the same time? If we had enough observations, we could use conditioning plots to examine both relationships simultaneously. Unfortunately, we don't have enough scandal-ridden nominations to do much with. However, we can take a closer look at the scandal-free nominations. Figure 11 shows them at low and high levels of staffing.

// Insert Figure 11 about here //

If you compare this figure with the preceding one, you'll see that the curves in the left hand panels appear virtually identical: low staffing nominations typically involve ideologically proximate nominees, and distance seems not to matter to duration at such low levels. The curve in the right hand panel of Figure 11 has about the same shape as that in the right hand panel of the previous figure, except it's shifted down slightly. So, it looks scandal has a simple additive effect, at least at high levels of staffing.

5. Fitting Parametric Models

We've learned a lot about the data's structure just from looking at it in artful ways. So let's go ahead and fit some parametric models. However, before we do, we should ask ourselves: what class of models should we fit? There are at least three possibilities.

---

figure with a dash for each data point, so the reader can see the distribution of the data. Simple practices like this would contribute to more honesty (and self-awareness) in quantitative political analysis.

First, based on equation (3), we could fit logged days or logged pages using ordinary least squares. This is certainly the simplest option and, in light of the what we saw above, we know it is likely to work well (there might still be heteroskedasticity with the pages data, though). However, there are several problems here. First, we have to decide what to do with the zero duration nominations, since log[0] is undefined. For the exploratory analysis in the previous section I used log[days + 1] but this is ad hoc. Also, it is a pain to constantly have to convert back from log-days to days.

The second option is to treat the data as count data and model them appropriately. This means fitting the data with Poisson or perhaps negative binomial regressions. This approach automatically takes care of the zero duration nominations and we won't have to worry about converting back from log-days to days. Running count models is simple these days, so this is an attractive choice.

Third, some people will view the hearing data as duration data. Accordingly, they would want to estimate survival curves and hazard rates. This is a perfectly sensible choice. But note, the theory does *not* say, the Chair holds a hearing for a day, decides whether to stop or continue depending on the realization of some random variable, holds it for another day if he continues, decides to stop or not, etcetera. That would truly suggest a survival analysis. Rather, the theory says the Chair determines the length of the hearings ex ante, based on the information he has in hand. This really suggests a count model.[14]

---

[14] The data are not quite duration data in the sense that the dependent variable is a count of "hearing days" on which the hearing was actually conducted. If we were to pursue the idea of duration model, we would want a measure of the onset of each spell of hearings, noting each day a spell was in the risk set of termination. Be that as it may, we checked the statistical results of the Poisson and negative binomial regressions with a duration analysis, a Weibull parametric regression. The results of the Weibull analysis are essentially the same as with the count models.

A Quick Review of Poisson Regression

We opt for the count approach since it seems the closest to the theory. So let me say a few words about regression models of count data. My discussion follows Wooldridge 2003 (pp. 573-578), which is very basic. An advanced and extremely thorough reference is Cameron and Trivedi 1998.

We want to model the expected value of days or pages, conditional on the value of the independent variables. In other words, we want to model $\mathrm{E}(y \mid \Delta, \theta, \kappa)$ (in this notation, $\Delta = \delta - \bar{\delta}$ and y is either days or pages). We have seen that a reasonable first cut uses the logarithm of y and relates log-y to the independent variables. This suggests a model something like

$$\ln(\mathrm{E}(y \mid \Delta, \theta, \kappa)) = \beta_0 + \beta_1 \Delta + \beta_2 \theta + \beta_3 \kappa$$

But now we want to avoid logs and model $\mathrm{E}(y \mid \Delta, \theta, \kappa)$ directly. So, let's take exponents of both sides to get

$$\mathrm{E}(y \mid \Delta, \theta, \kappa) = e^{\beta_0 + \beta_1 \Delta + \beta_2 \theta + \beta_3 \kappa}$$

We could model this with non-linear least squares but that won't take into account the special statistical distributions associated with counts. In fact, we've observed that days and pages are definitely not normally distributed. The archetypal distribution for count data is the Poisson distribution. Relatively rare events often display a Poisson distribution, for example, arrivals in hospital emergency rooms, people arriving at a check out counter or (its first application) deaths from mule kicks in the Prussian Army.

Let's take a closer look at the Poisson distribution. First, it takes only non-negative integer values. Its density looks like this: $\frac{e^{-\lambda} \lambda^{y}}{y!}$, where y is the count. So the density gives the probability of drawing a particular number, e.g., the probability of a three or a four. The key parameter governing the density, λ , is both the mean and the variance of the distribution. (The Poisson distribution has many amazing properties and this is just one.) You will recall that the mean duration of the public hearings in days is 3.3. Figure 12 shows the density of a Poisson distribution with λ = 3.3. If you compare this with the kernel density smoother in the left-hand panel of Figure 4, you'll see they actually do resemble each other.

//Insert Figure 12 about here //

Now, the typical regression model based on the Poisson distribution holds that the conditional expectation of an observation (conditional on the values of the independent variables) equals lambda, that is $\mathrm{E}[y \mid \Delta, \theta, k] = \lambda$. I hope this makes sense to you. Now combine this with the previous equation to get

$$E[y \mid \Delta, \theta, k] = \lambda \ = \ e^{\beta_0 + \beta_1\Delta + \beta_2\theta + \beta_3\kappa}$$

This expression can be estimated with maximum likelihood methods and, as I mentioned, is programmed into many statistical packages. You'll often find it in the generalized linear model (GLM) component of statistical packages.

But – how do you interpret the coefficients from estimating an equation like this? As usual, to get the marginal effect of an increase in a variable take the partial derivative of the expression with respect to the variable of interest, say Δ, to get

$$\frac{\partial}{\partial \Delta} E[y \mid \Delta, \theta, \kappa] = e^{\beta_0 + \beta_1\Delta + \beta_2\theta + \beta_3\kappa}\beta_2 = \ E[y \mid \Delta, \theta, k]\beta_2$$

In other words, the effect of an increase in distance is a *proportional* increase in the conditional mean.

To illustrate, let's cheat a bit and look ahead to Table 4. Glance at Model 1a in the table. You can see the estimated coefficients in a very simple model. The mean values of Distance and Staff are .34 and 114, respectively. If we set those variables as their mean values and set scandal to zero, we calculate $E[y \mid \overline{\Delta}, \theta = 0, \overline{\kappa}) = e^{-.36+1.14(.34)+.61(0)+.01(114)}$ = 3.2 days. This is the predicted duration of a scandal free nomination, when staff and distance take their mean values. A similar calculation shows that the equivalent value for a scandal-wracked nomination is 5.9 days. You can see that the switch from no scandal to scandal increases the duration by (5.9 – 3.2)/3.2 = .84. In other words, an 84% increase. An approximation that is often used is that roughly the percentage change in $\mathrm{E}[y \mid x_1,...x_n]$ = (100 times the coefficient) times the size of the change. So, for example, if ideological distance increases from zero to one, duration will go up 114%, that is, more than double (using the coefficients in Model 1a). This approximation suggests about a 60% increase with the switch from no-scandal to scandal.

We are almost ready to run some models. But we should discuss one more issue. The Poisson model restricts the conditional variance so that it equals the conditional mean, that is, it forces $Var(y \mid x_1,...x_n) = E(y \mid x_1,...x_n)$. However, it's often the case that observed count data are too spread out to exactly fit a Poisson distribution (over-dispersion). Much less often, you'll see court data that are too tightly distributed. Now, as Cameron and Trivedi note, the Poisson maximum likelihood estimator is robust to distributional misspecification, in the same way OLS is robust to non-normality in the linear regression context. So if the specification is correct, the coefficient estimates are consistent though not efficient. In other words, we need to adjust the standard errors and *t*-statistics (the standard errors will be too small, if there is over-dispersion). This is similar to having to adjust standard errors in OLS models in the face of heteroskedasticity.

How can we do this? The most direct way is to use quasi-maximum likelihood estimation (QMLE) to estimate a Poisson-like model with greater flexibility in the variance. Basically, the Poisson QMLE model will adjust the standard errors to take account of the over-dispersion. Wooldridge discusses this a little more, if you want a few more details. This is quite easy to do in some statistical packages.

An alternative approach is to make some additional assumptions about the cause of the over-dispersion. For example, one might assume that there is unobserved heterogeneity for each observation (nominee in our cases), that acts like a random error term in the exponential regression model. If the rest of the model is Poisson distributed and this error term is Gamma distributed, you get a negative binomial distribution, which has a larger variance than a Poisson distribution. This is also easy to estimate in many statistical packages. Various other assumptions can also get you to a negative binomial distribution for count data – there are more details in Cameron and Trivedi (1998).

Standard statistical packages will automatically give you a statistical test for over-dispersion, which is just a chi-square test. If the data are over-dispersed, you should either use Poisson QMLE (Wooldridge's preferred option) or negative binomial regression, if you don't mind making the additional assumptions.

Two Simple Models

Ok, what would be a reasonable first-cut model, in light of the theory and our knowledge of the data's structure? The most straightforward answer is to model hearing length (days) and intensity (pages) as an additive function of Ideological Distance, Scandal, and Staff Size. In addition, we should investigate Senatorial Courtesy.

Take a look at the models of the hearing length (days) data in Table 4. Model 1 is a plain vanilla model of the data. It contains just the three key variables, Ideological Distance, Scandal, and Staff Size. Model 2 then adds Senatorial Courtesy. We estimate the models with Poisson regression and (mostly for pedagogical reasons) negative binominal regression. The three key independent variables are significant at the .01 level and have the positive coefficient predicted by the formal model's comparative statics. Senate Courtesy is also significant at the .10 level (P-value = .06). The chi-square test suggests that the data is not over-dispersed, though the test is close. We don't bother to show the QMLE estimates but the dispersion parameter (the factor by which one inflates the standard errors) is taken to be 1.15, which does not change the substantive interpretation of the results in any way.

Now glance at the same models estimated with the data on the volume of the hearing report, the pages data. These are shown in Table 5. Here the chi-square test definitely indicates over-dispersion. Table 6 shows QMLE Poisson estimates of the same two models. In the QMLE estimates, the extreme inflation of the standard errors makes the staff and courtesy variables statistically insignificantly different from zero. In contrast, those variables remain statistically significant in the negative binomial models, shown in Table 5. The over-dispersion seems so severe in the Poisson framework that it seems reasonable to consider the negative binomial as more plausible.

Adding a Threshold

Our earlier discussion emphasized a threshold in distance as a distinctive theoretical prediction, one that received some support when we visualized the data. Let's fit a parametric model with a threshold to the days data and see how it compares to the earlier models. I leave a similar exercise with the pages data as an exercise for you.

How do we go about doing this? Here is a simple way.[15] Let's say we think the threshold should be .4, on the distance scale. First, define a dummy variable that takes the value zero if Distance is less than .4 and one if greater. Then, interact this variable with Distance - .4. This procedure creates an "elbow" shaped threshold variable: the variable takes the value of zero at and below the threshold, then increases linearly (from zero) as Distance increases above .4.

---

[15] Another method is to estimate a generalized additive model (a "gam"), with distance entered in a flexible way, say via smoothing splines (see Hastie and Tibshirani 1990). This gives similar results.

Let’s estimate this model using the elbow version of Distance and compare it with our simple model. Table 7 shows the results. The first column repeats Model 2a, for purposes of comparison (that is, the Distance measure is the simple absolute value used earlier). The second column shows the same model, but using the threshold version of Distance. As you can see, the coefficients on the other variables have about the same magnitude (the intercept is different of course). However, the new Distance measure is steeper (and more statistically significant). The overall fit of the model is modestly better. In sum, the data is equally compatible, and perhaps favors, the threshold version of ideological distance.

What Could Go Wrong?

At this point, we could declare victory and go home. After all, we crafted a new theory of committee hearings, which we formalized in a mathematical model; we interrogated the formal model to uncover its empirical implications; we collected appropriate data to test the empirical predictions; we explored the data for structure using non-parametric models; we fit parametric models to the data; and, both the parametric and non-parametric models behaved like the formal model predicted. Given this, we can now interpret the politics of public hearings on Supreme Court nominations in a way that is both theoretically informed and historically grounded. Admittedly, understanding the duration and intensity of public hearings on Supreme Court nominations is a small question, not like World Peace or ending poverty. But still, in a modest way we might call this a good day at the office.

But wait! We should ask one more question: what could go wrong with our pretty picture? Is there some way in which we haven’t been completely fair with the data, something we overlooked perhaps because it didn’t mesh with the theory?

Well, when we were visualizing we found two troubling patterns: 1) log-days looks like it might have been generated by two or possibly three different data generating processes (the multiple peaks in the distribution of log-days, Figure 4), and 2) the independent variables mostly take low values in the early years though things are more variable in the later years.[16] Might these things be related?

Return to Figure 4. If we identify the separate populations of cases and map their distributions on the figure with a kernel density smoother, the peaks of the distinct populations should correspond to the identified peaks in the overall distribution. This is what we have done in Figure 13. The thick line is again the overall distribution, showing three peaks, which occur at about 2.1 days, 5.75 days, and 12 days. The three thin lines are kernel density smoothers from three distinct sub-populations. The distribution with a tall peak to the left, at about 2 days, is nominations that occurred prior to 1967 (only one of these involved scandals prior to or during the hearing). The distribution whose peak corresponds to the middle peak is scandal-free nominations after 1967.[17] The distribution with two large peaks, one in the middle and one at the far right, is post-1967 scandal-ridden nominations.[18]

I will leave the following as an exercise for you. Re-run the model, partitioning the data into early (pre-1967) and late (1967 and forward) groups. Since there are no "courtesy" nominees in the later period, you will have to drop this variable for the latter. You will find that the model works well in the later period (though staff is no longer significant). Even the coefficients look similar.

---

[16] Refer back to the scatter plot matrix, Figure 8. Look at the last column, which shows each of the variables plotted against time. Compare the second half of each scatter plot with the first half. You'll see that Distance takes high and low values in the second half of the period, but the highest values tend to come late; scandals come only in the second half of the period; and Staff takes high and low values in the second half of the period but only low values in the first half.

[17] This distribution has one very long duration, that of Homer Thornberry. Thornberry's hearings were joint with Fortas's, whose nomination was scandal-plagued.

[18] A few scandal-ridden nominations had short durations, e.g., Kennedy with 3 days and Breyer with 4 days.

But in the earlier period, the model seems to fall apart, in that only the courtesy variable remains statistically significant.

What are we to make of this? There are two obvious interpretations. One interpretation is that the relationships elaborated in the theory hold throughout the period under study, but the effect of ideological distance is muted in the early period by the low values Distance always took in that period. Then, in the later period, we can detect it because Distance takes on a wide range of values, all the way from zero to one. The second interpretation goes: the world suddenly changed in 1967 (or thereabouts). Prior to that, the theory simply didn't apply because ideology was unimportant. Then suddenly, ideological conflict became critical in Supreme Court nominations. So, the first interpretation says: the evidence powerfully favors the theory post-1967, and we can understand why it would be hard to detect in the earlier period even if it held (the circumscribed range of the variables). So we should assume the theory holds throughout. The second says: the theory holds only in the last part of data set, due to an (unexplained) structural break.

// Insert Table 7 about here //

Let's see if we can get a little leverage on the two interpretations by examining residuals. Consider the top row of Figure 14. It displays the residuals from Model 3 in Table 7, our best model so far. The left hand panels show the residuals for the early observations, arrayed against the (raw) Distance measure. The right hand panel shows the residuals for the late observations. A loess smoother shows the pattern in the data. As you can see, the model tends to over-estimate durations in the early period, especially if Distance is greater than the threshold, .4. And, the model tends to under-estimate durations in the late nominations. This pattern indeed suggests a pre/post-1967 effect, but perhaps only a shift rather than a radical break.

// Insert Figure 14 about here //

So, let's estimate that model. This is Model 4 in Table 7. There you can see that a pre1967 dummy variable is statistically significant. In this model, the staff variable is no longer statistically significant but this is not terribly surprising, given the time pattern in staffing. The bottom rows in Figure 14 display the residuals from this model. As you can see, it fits the early observations much better. In the later period, the model no longer consistently under-estimates the durations. However, it tends to slightly over-estimate durations for nominations with low distance and under-estimate durations for those with large ideological distances. So, there remains some unexplained structure in the data, but Model 4 clearly out-performs the earlier ones.

The pre/post-1967 effect is completely ad hoc – it emerged from the data rather than the theory. Of course, we can imagine some plausible reasons why 1967 might denote a watershed in the politics of Supreme Court nominations. We might point to the judicial activism of the Warren Court and the proliferation of new interest groups that occurred at that time. Controversial Supreme Court decisions coupled with burgeoning groups may have boosted the political salience of Supreme Court nominations, leading to greater publicity-seeking behavior on the Judiciary Committee, beyond the simple effect of increased ideological distance.

So, it looks like we might want to tweak the theory, to account for the varying political salience of the Court. In turn, this could lead us to collect systematic data on salience so that we can unpack the pre/post-1967 effect and see if our conjecture about salience and groups holds water.

I'm going to stop here, but I hope you are getting a feel for the interplay between theory and data, a sort of back-and-forth tennis game between deduction and induction. Both halves of the court are fun and worthwhile but what I enjoy most is volleying back and forth.

5. Discussion: The Normative Issues

In the first lecture, I said it was important to ask normative, evaluative questions, once we had a solid foundation of theory and data upon which to stand. Well, we have a theory that works quite well, according to the data. What are the normative questions here? The model assumes the Chair tries to make hay by exposing the scandalous short-comings of ideologically distant nominees. What are the implications? Is this process a good thing?

Clearly, the process has some positive features. For one thing, “turkeys” get a good airing, at least when the nominee is ideologically far from the Judiciary Chair. So far as it goes, this is surely a good thing, as it lets senators know what they are voting for or against. It also makes it easier for citizens to evaluate their senators’ votes and the president’s performance as a chooser of nominees.

But, I don’t want to exaggerate. Even at the simplest level, the asymmetries in the process are disturbing. First, ideologically proximate turkeys tend to get a free pass in the Judiciary Committee, or at least a less costly passage. Hugo Black is an extreme case in point. So if you think exposing turkeys is a good thing, then the manifest ideological bias in the process is a bad feature. If you feel this way, you might favor giving greater scope to the minority on the Judiciary Committee, who will be inclined to be tougher on nominees the majority favors.

There is a second kind of asymmetry that many people find disturbing, I mean the asymmetry between “scandal” and “ideology.” The model says the real engine driving the process is ideology. But the hearings aren’t forthrightly about ideology – instead, a hostile Chair makes hay by publicizing the nominee’s unsavory or shady behavior. Some people may see this behavior or trivializing or even mis-directing the process. The example of Black may be disturbing in this regard, as many observers believe he was an excellent justice despite his early sojourn as a KKK member. But he surely would have had a tough time in hearings before a Republican controlled Judiciary Committee, and might well have been rejected by a Republican Senate. So, the emphasis on scandal may do a poor job of screening future justices.

The nasty way the hearings unfold during periods of ideological polarization may also contribute to public alienation or disgust with government. Obviously, we haven’t confirmed this conjecture with systematic data, and it might be hard to do so. But the point does suggest itself.

Still, even if this is true, it is probably an inevitable consequence of ideological polarization: historically in America, polarized politics is nasty politics.

The process clearly creates incentives for the president, who can anticipate the action in the Judiciary Committee. First, it creates an incentive for the president to vet nominees and weed out the turkeys before they enter the hearing room, especially when the Judiciary Committee is hostile. Surely this is a good thing. It would be interesting if archival evidence showed this happening. Second, because of the way scrutiny is tied to ideology, the process *may* create an incentive for the president to avoid ideological extremists, at least when the Chair is far from the president. By doing so, the president can take some of the steam out the hearings. If so, one might see this as a good feature of the process, especially in an age when most voters are moderates and most politicians are extremists. But we would need to investigate the president's choices in depth to see if this is really true before we start lauding the process for encouraging moderation.

Ultimately we are concerned with a much larger normative question: Does the nomination process keep the Supreme Court within the ideological mainstream of American politics, and staffed with judges of intellectual competence and high integrity? For if the Court wanders outside the mainstream, it is apt to provoke potentially disastrous assaults on judicial independence. And a judiciary filled with corrupt, venal hacks will inevitably undermine the rule of law. Of course, the theoretical and empirical work in this chapter is only a very small step toward answering the big question we really care about. But the only way to build a cathedral (as it were) is by laying one brick on top of another, one at a time.

This concludes my section of the course. I hope I've conveyed some sense of the fun of political science. Perhaps I've challenged your notions of how to engage in political argument. If so, I'll count the lecture as a success. In any event, good luck with your future encounters with political data, political argument, and politics.

Problem Set for Chapter 3

1. Theory. The formal model in Section 2 has implications about how many negative stories about the nominee we should see during the period of public hearings. To derive these implications, substitute equation (3) back into equation (2) and solve for the number of negative stories, *n*. Using this expression, derive comparative static predictions about *n*. Indicate what data you would need to test these predictions and indicate what statistical methods you would use.

2. Co-plots. Form a conditioning plot similar to Figure 9-11, but conditioning on senatorial courtesy. Use the data shown in the scatter plot matrix to make two more conditioning plots of your choice.

3. Senatorial courtesy. Devise a logit or probit model that predicts when we should see senatorial courtesy. Think about the possible explanatory variables such as ideological distance between the two parties. Do you believe your model?

4. Analysis of the pages data. Complete an analysis of the pages data similar to that in the chapter on the days data. In particular, run Negative Binomial and QMLE regressions after including the pre-1976 variable, and test whether a threshold version of distance is comparable with the data (Refer to Table 7).

References

Aberbach, Joel D. 1991. *Keeping a Watchful Eye: the Politics of Congressional Oversight*. Washington, D.C.: The Brooking Institution.

Aberbach, Joel D. 2002. "What's Happened to the Watchful Eye?" *Congress and the Presidency* 29:3-23.

Ashworth, Scott and Bueno de Mesquita. 2006. "Monotone Comparative Statistics for Models of Politics," *American Journal of Political Science* forthcoming.

Athey, Susan, Paul Milgrom, and John Roberts. 1998. *Robust Comparative Statics.* unpublished working paper.

Bailey, Michael and Kelly Chang. 2001. "Comparing Presidents, Senators, and Justices: Interinstitutional Preference Estimation." *Journal of Law, Economics, and Organization* 17:477-506.

Baumgartner, Frank and Bryan Jones (eds). 2002. *Policy Dynamics*. Chicago: University of Chicago Press.

Cameron, A Colin and Pravin K. Trivedi. 1998. *Regression Analysis of Count Data*. New York: Cambridge University Press.

Cameron, Charles M. and Peter B. Resendorff. 1993. "A Signaling Theory of Congressional Oversight." *Games and Economic Behavior* (special issue on game theoretic models of politics) 5:44-70.

Cleveland, William S. 1993. *Visualizing Data.* Hobart Press.

Diermeier, Daniel and Timothy J. Fedderson. 2000. "Information and Congressional Hearings." *American Journal of Political Science* 44:51-65.

Epstein, Lee, Andrew D. Martin, Jeffrey A. Segal, and Chad Westerland, 2005, *Judicial Common Space*, working paper Department of Political Science Washington University.

Ferejohn and Charles Shipan. 1989. "Congressional Influence on Administrative Agencies: A Case Study of Telecommunications Policy." *Congress Reconsidered* (4th edition) Larry Dodd and Bruce Oppenheimer (eds.). Congressional Quarterly Press.

Milgrom, Paul and C. Shannon. 1998. "Monotone Comparative Statics," *Econometrica* 62:157-180.

Morton, Rebecca. 1999. *Methods and Models: A Guide to the Empirical Analysis of Formal Models in Political science.* New York: Cambridge University Press.

Rosenberg, Gerald. 1992. "Judicial Independence and the Reality of Political Power," *The Review of Politics* 54:369-398.

Rutkus, Denis Steven and Maureen Bearden. 2006. *Supreme Court Nominations, 1789-2005: Actions by the Senate, the Judiciary Committee, and the President*. Congressional Research Service Report for Congress RL33225.

Segal, Jeffrey and Albert Cover. 1989. "Ideological Values and the Votes of U.S. Supreme Court Justices." *American Political Science Review* 83:557-565.

Tibshirani, Rob J. and Trevor Hastie. 1990. *Generalized Additive Models.* Chapman & Hall/CRC.

Wooldridge, Jeffrey M. 2003. *Introductory Econometrics: A Modern Approach*. Mason, Ohio: Thomson-Southwestern.

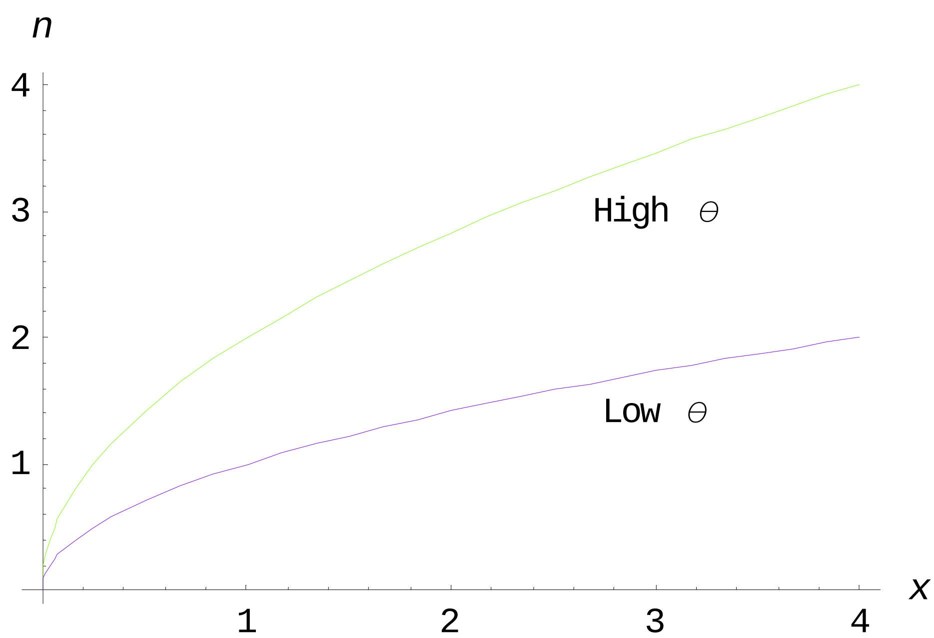

Figure 1. Production functions for negative messages.

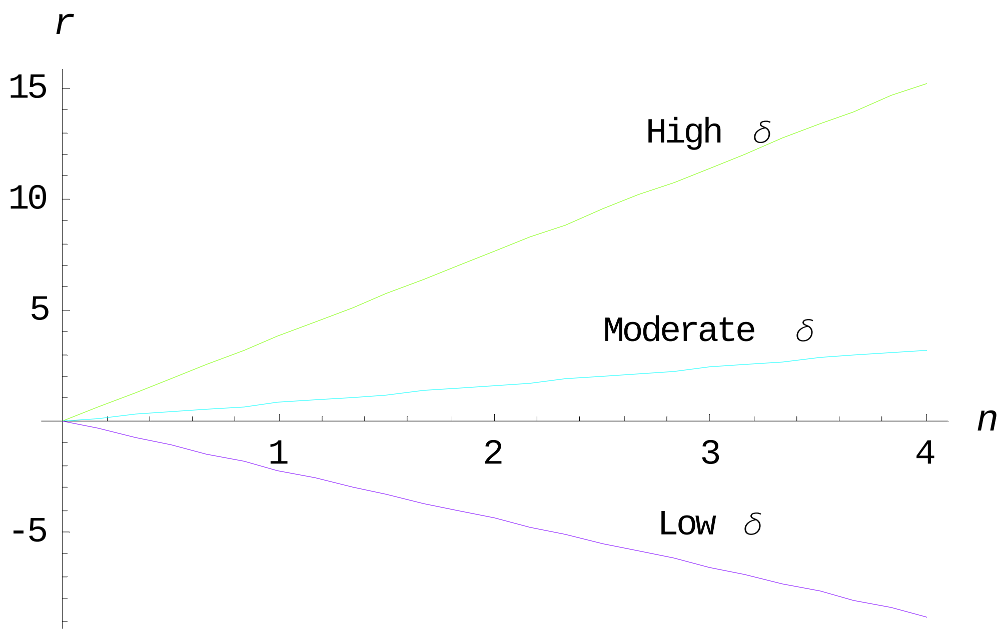

Figure 2. Return functions from negative messages.

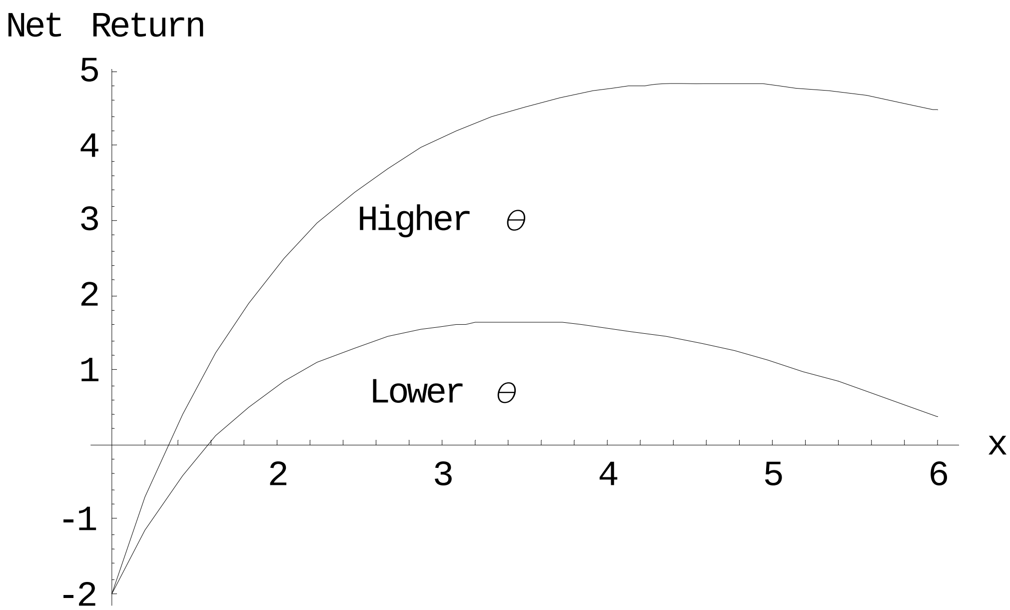

Figure 3. Net returns from hearing lengths, at high and low levels of “dirt.”

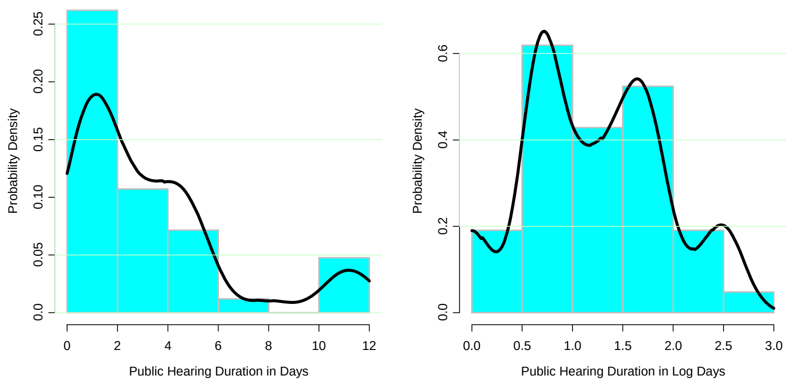

Figure 4. Distribution of duration of public hearings, in days.

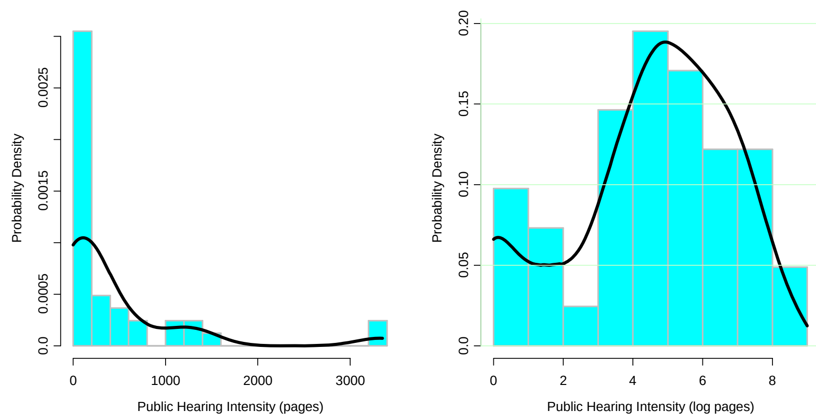

Figure 5. Distribution of intensity of public hearings, as measured by number of pages in committee reports.

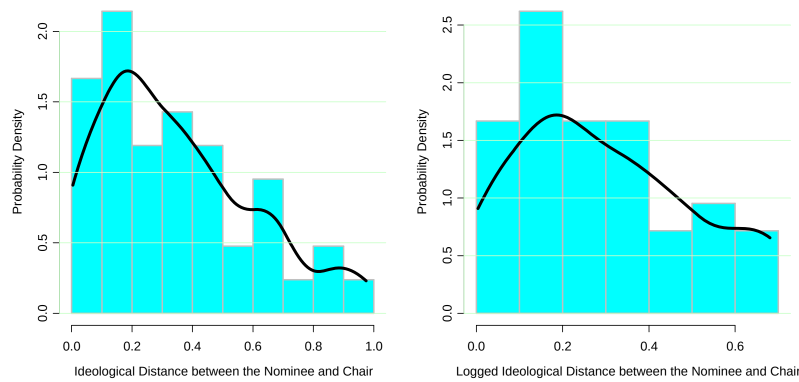

Figure 6. Distribution of ideological distances between nominees and the chairs of the Judiciary Committee (absolute values, NOMINATE scale)

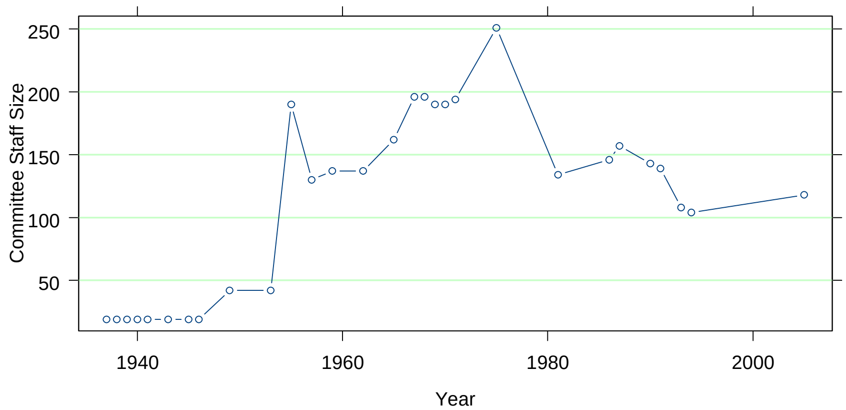

Figure 7. Size of the Senate Judiciary Committee staff at the time of Supreme Court nominations, 1937-2005

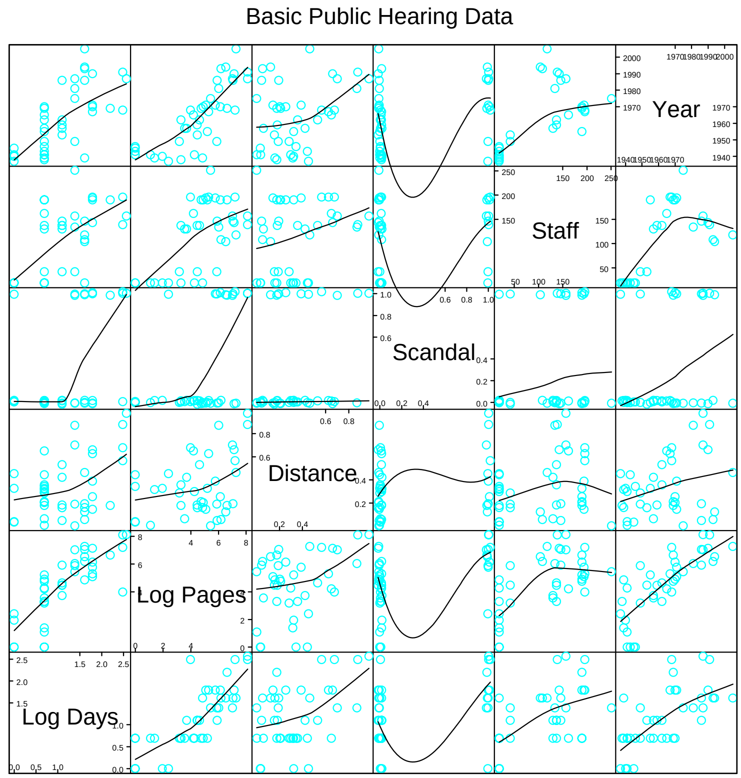

Figure 8. Scatter plot matrix of the public hearing data.

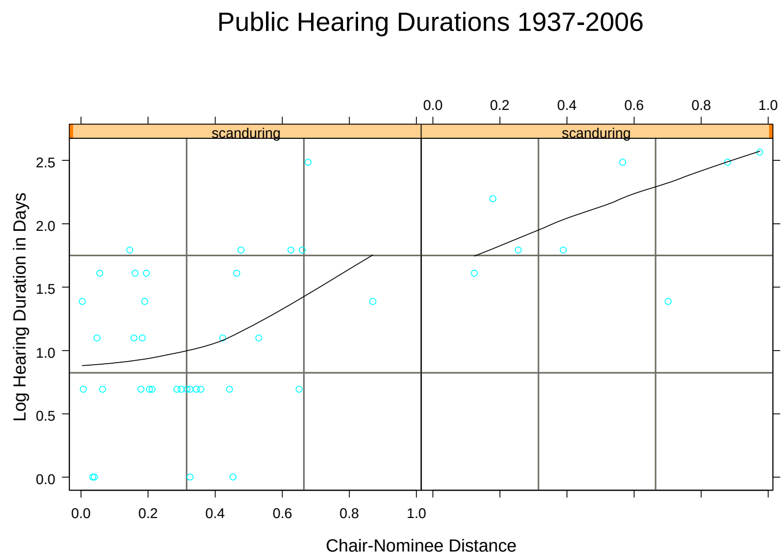

Figure 9. Hearing duration and ideological distance between the Chair and the nominee, without a scandal (left panel) and with one (right panel).

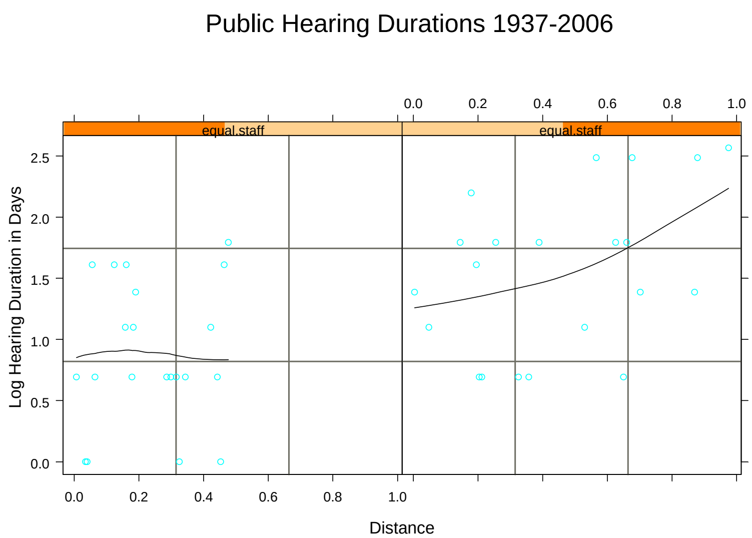

Figure 10. Hearing duration and ideological distance between the Chair and the nominee, at low levels of staffing (left panel) and relatively high levels (right panel)

Public Hearing Durations 1937-2006: Scandal-Free Only

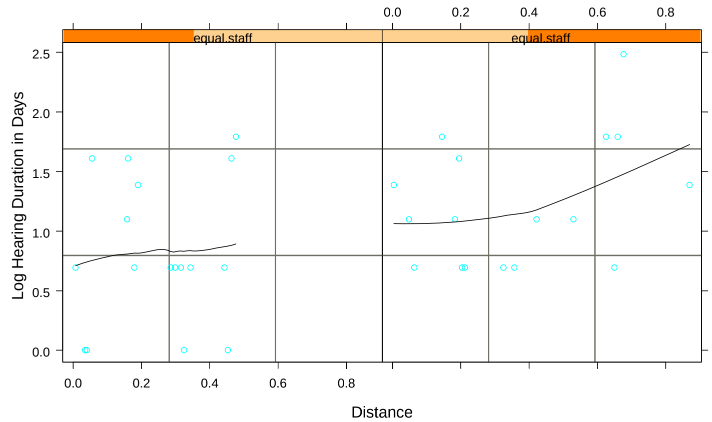

Figure 11. Scandal-free nominations at low staffing levels (left panel) and high staffing levels (right panel).

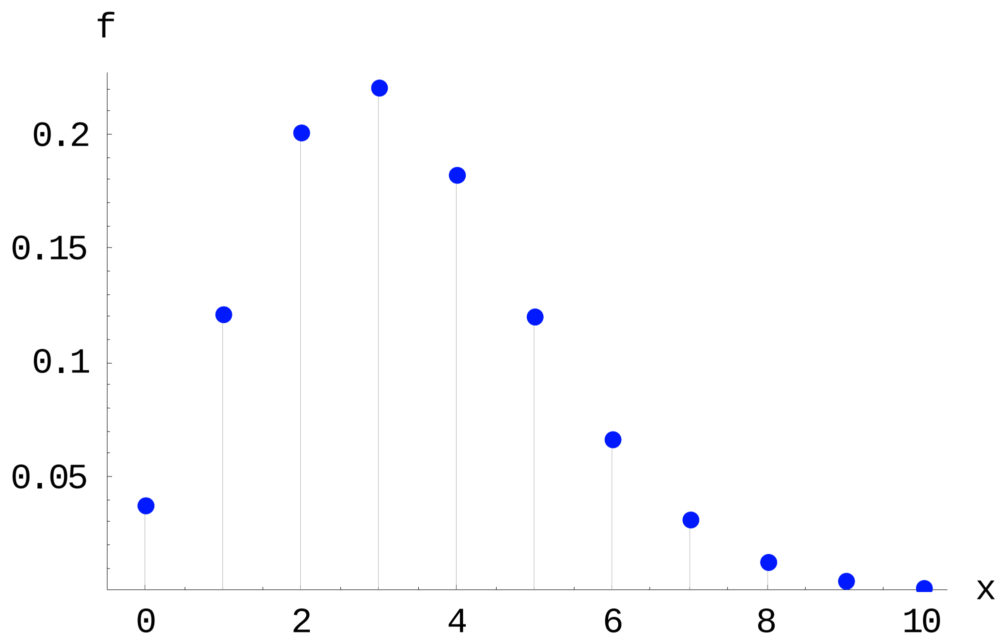

Figure 12. Probability mass function of a Poisson random variable with mean 3.3.

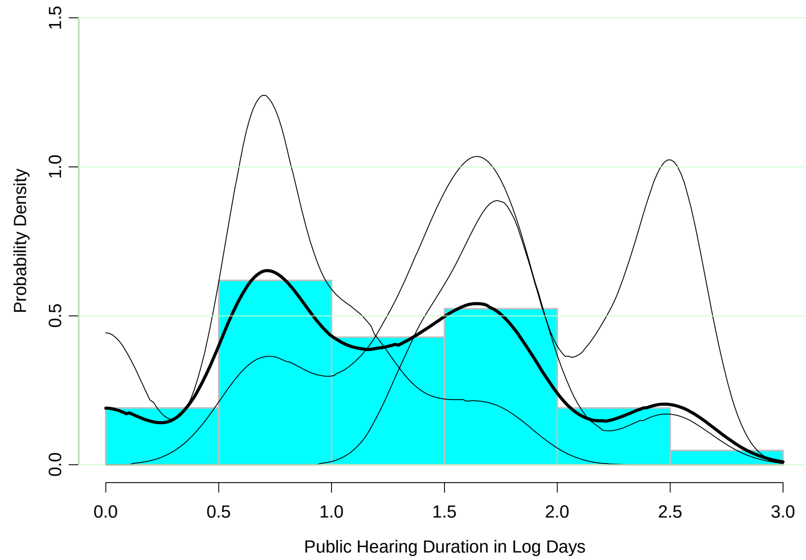

Figure 13. Three populations of nominations: early nominations, late scandal-free nominations, and late scandal-ridden nominations.

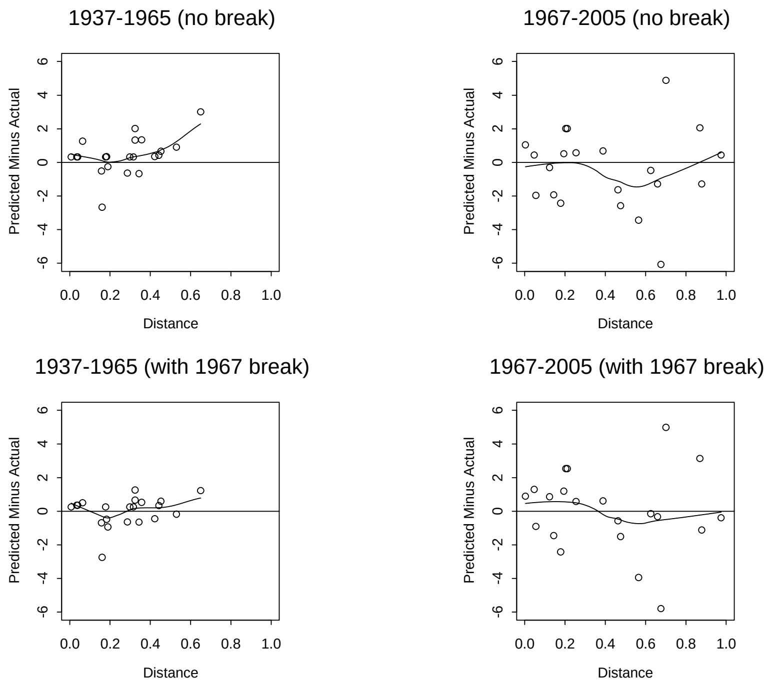

Figure 14. Early and late residuals vs. ideological distance, without (top row) and with (bottom row) a 1967 break

Chair	**Year**	**Party**	**State**	**DW-NOMINATE 1**	**DW-NOMINATE 2**
Henry F. Ashurst	1933 - 1941	Demo	Arizona	-.220	-.004
Frederick Van Nuys	1941 - 1945	Demo	Indiana	-.136	.029
Pat McCarran	1945 - 1947	Demo	Nevada	.086	.799
Alexander Wiley	1947 - 1949	Rep	Wisconsin	.174	-.451
Pat McCarran	1949 - 1953	Demo	Nevada	.086	.799
William Langer	1953 - 1955	Rep	N. Carolina	.066	.650
Harley M. Kilgore	1955 - 1956	Demo	W. Virginia	-.502	.036
James O. Eastland	1956 - 1978	Demo	Mississippi	.111	.965
Edward M. Kennedy	1979 - 1981	Demo	Massachusetts	-.480	-.403
Strom Thurmond	1981 - 1987	Rep	S. Carolina	.425	.382
Joseph R. Biden, Jr.	1987 - 1995	Demo	Delaware	-.371	.007
Orrin G. Hatch	1995 - 2001	Rep	Utah	.359	.016
Patrick J. Leahy	Jan 3 - Jan 20, 2001	Demo	Vermont	-.452	-.074
Orrin G. Hatch	Jan 20 – June 5, 2001	Rep	Utah	.359	.016
Patrick J. Leahy	June 6 – Jan 15, 2003	Demo	Vermont	-.452	-.074
Orrin G. Hatch	2003 - 2005	Rep	Utah	.359	.016
Arlen Specter	2005 -	Repn	Pennsylvania	.024	-.648

Table 1. The Chair of the Judiciary Committee, 1937-2005

Nominee	Ideology Score (estimated)
Black	-.642
Reed	-.368
Frankfurter	-.372
Douglas	-.527
Murphy	-.536
Stone2	-.065
Byrnes	-.112
R. Jackson	-.371
Rutledge	-.493
Burton	.122
Vinson	-.257
Clark	-.101
Minton	-.197
Warren	-.154
Harlan	-.125
Brennan	-.314
Whittaker	.044
Stewart	-.072
B.White	-.243
Goldberg	-.416
Fortas1	-.529
Marshall	-.502
Fortas2	-.442
Thornberry	-.553
Burger	.339
Haynsworth	.306
Carswell	.382
Blackmun	.331
Powell	.275
Rehnquist1	.519
Stevens	.132
OConnor	.320
Rehnquist2	.631
Scalia	.484
Bork	.582
D.Ginsburg	.563
Kennedy	.309
Souter	.275
Thomas	.502
R.B.Ginsburg	-.312
Breyer	-.244
Roberts	.525
Miers	.430
Alito	.538

Table 2. Estimated ideology scores for the nominees

Nominees with Hearings			No Hearing
Scandal Prior to Hearing	**Scandal During Hearing**	**Scandal After Hearing**	
Haynsworth	Clark (last day)	Jackson	Black
Carswell	Fortas 2	Warren	
Rehnquiest1	Rehnquist2 (last day)		
Bork	Thomas		
Kennedy			
Breyer			

Table 3. Nominees with Scandals

	Poisson		Negative Binomial	
Variables	**Model 1a**	**Model 2a**	**Model 1b**	**Model 2b**
Ideological Distance	1.14 (.31)***	1.08 (.31)***	1.12 (.34)***	1.06 (.32)***
Scandal	.61 (.18)***	.64 (.18)***	.61(.19)***	.64 (.18)***
Staff Size	.01 (.00)***	.01 (.00)***	.01(.00)***	.01 (.00)***
Senate Courtesy	---	-1.45** (.75)	---	-1.45 (.75)
Constant	-.36 (.27)	-.04 (.29)	-.35 (.28)	-.04 (.30)
N	42	42	40	42
Pseudo R^2	.30	.32	.18	.21
Log Likelihood	-78.9	-76.1	-78.8	-76.1

Dependent variable: number of days of public hearings. For Model 1a: Goodness-of-fit chi2 = 48.6, Prob > chi2(35) = 0.12. For Model 2a: Goodness-of-fit chi2 = 43.0, Prob > chi2(35) = .23. * p < .10, ** p < .05, *** p < .01

Table 4. Two simple models of hearing duration (days).

Variables	Poisson		Negative Binomial	
	Model 1c	**Model 2c**	**Model 1d**	**Model 2d**
Ideological Distance	2.17 (.03)***	2.08 (.03)***	1.59 (.82)**	1.69 (.74)**
Scandal	1.32 (.02)***	1.39 (.02)***	1.28 (.49)***	1.24 (.44)***
Staff Size	.005 (.00)***	.001(.000)***	.011 (.004)***	.007 (.004)*
Senate Courtesy	---	-3.76 (.18)***	---	-2.97 (.69)***
Constant	3.89(.03)***	4.44 (.03)***	3.28 (.51)***	3.83 (.55)***
N	41	41	41	41
Pseudo R^2	.66	.71	.04	.06
Log Likelihood	-6146.0	-5199	-260.3	-254

Dependent variable: Pages of committee report. Model 1c: Goodness-of-fit chi2 = 12038, Prob > chi2(35) = 0.00. Model 2c: Goodness-of-fit chi2= 10145, Prob > chi2(35)= .00. * p < .10, ** p < .05, *** p < .01

Table 5. Two simple models of the volume of hearings (pages).

Variables	QMLE Poisson	
	Model 1e	**Model 2e**
Ideological Distance	2.17 (.54)***	2.08 (.49)***
Scandal	1.32 (.35)***	1.39 (.34)***
Staff Size	.005 (.003)	.001(.003)
Senate Courtesy	---	-3.76 (3.36)
Constant	3.89(.54)***	4.44 (.52)***
N	41	41
Null Deviance	35461	35461
Residual Deviance	12039	10145

Dispersion parameter taken to be 425 in Model 1e and 366 in Model 2e.
p < .10, ** p < .05, *** p < .01

Table 6. QMLE Poisson estimation of the volume of hearings data (pages).

Variables	**Model 2a**	**Model 3**	**Model 4**
Ideological Distance	1.08 (.31)***	1.67 (.44)***	1.37 (.45)***
Scandal	.64 (.18)***	.61 (.18)***	.46 (.19)**
Staff Size	.005 (.002)***	.005 (.002)***	.002 (.002)
Senate Courtesy	-1.45** (.75)	-1.40 (.74)*	-1.28 (.75)*
Pre1967	---	---	-.76 (.31)**
Constant	-.04 (.29)	.19 (.27)	.96 (.40)**
N	42	42	42
Residual Deviance	43.0	41.6	35.1

Dependent variable is duration of hearings in days. In Models 3 and 4 Ideological Distance is a threshold variable (see text for discussion). Null deviance is 115.1.

Table 7. Poisson regressions of hearing durations in days, with a distance threshold and a mid-1960s break.