

The Politics of Constitutional Review: Evidence from the European Court of Justice^{*}

Michael Malecki[†]

2009-09-15

^{*}A version of this paper was prepared for the 2009 annual meeting of the American Political Science Association, Toronto, Ontario, Canada, 2–6 September 2009.

[†]Ph.D. Candidate, Washington University in St. Louis

Abstract

The European Court of Justice (ECJ) has fostered the development of a common European legal order, and in doing so, has asserted itself and its supremacy more, and more successfully, than any other international court. It has maintained features of international courts such as its composition of one judge per member state, while employing other tools of national high courts such as en banc decisions and organization into chambers, that together hide internal dissent and shield the ECJ from direct monitoring or curbing by the member states. The same shield has frustrated efforts to quantify the court's responsiveness to member states, with limited evidence that the ECJ yields to some member-state interest some of the time, but nonetheless has advanced integration beyond national governments' wishes. This equivocation arises at least in part from a failure to include relevant information about the court's composition and organization.

In fact, the six-year renewable terms of judges, their previous qualifications and affiliations, and the internal organization into chambers all provide prior information that can and should be incorporated into a more complete model of judicial behavior. I develop an extension of the well-studied item-response model to infer judges' preferences, using the structured ecological data from the cases they heard and relevant prior information about judges and the national governments that appoint them as well as information about cases. I offer new, more rigorous tests for existing theoretical hypotheses about the ECJ's deference to certain actors and preference for integration. The model is applicable to other settings of structured ecological data. Many other national and international courts hear cases in subset chambers, and relevant prior information should be included rather than ignored in models of judicial behavior.

1 Introduction

Among the EU institutions, the Court of Justice is probably least often the subject of calls for reform or for greater democratic accountability. The twenty-seven judges – one per member state – are entrusted with the power to compel member states to uphold their commitments codified in the treaties. Its role is often to balance the interests of member states against each other or against other EU institutions. It seems an incredible amount of trust granted to an extremely small group of men (and even fewer women). If we can dismiss the possibility that a robe and ascot confer superhuman reasoning, why would member states do this?

Member states have reasonably strong incentives to delegate to the court much of the tough work of completing the contracts of the treaties. They suffer from time-horizon problems, where considerable short-term benefits of renegeing are detrimental in the long run; the relative salience of issue areas varies widely from one member state to another; and they can use the court like the EU generally as a straw man of domestic politics. The Commission and Brussels bureaucracy are easily tarred as zealous Eurocrats, the EP and MEPs as irrelevant, unaccountable, and out of touch, and the Council as the bastion of shortsighted national interest, curtailed only slightly by weighted voting rules. The Court of Justice, by contrast, is burdened with no expectation of democratic accountability or responsiveness.

The ECJ has been theorized primarily as a unitary strategic actor with the European Union institutions and member states. In effect it has been a black box that may evince some preference or deference. Indeed scholars have learned a good deal about the other institutions by this analytical strategy. Treating the court as a single voice has clarified how the European Parliament interacts with the Commission, for example (Jupille, 2004).

Most importantly, it is clear that subnational courts in member states are empowered, through

the ECJ, with a form of judicial review apart from whatever may or may not be intended by their national constitutions (Burley and Mattli, 1993). The several national modes of judicial review (and implied legislative responses) described in Stone Sweet (2000) are potentially subverted by ECJ review.

Apart from a formal framework of judicial review, the European Union comprises systematically different “legal cultures” (Gibson and Caldeira, 1996), some more litigious than others, viewing the national courts – much less the European court – in quite different terms. Gibson and Caldeira (1995, 1998), concerned with the ECJ’s reservoir of popular legitimacy, find that it compares unfavorably to national high courts, which themselves are not wellsprings of legitimacy. The ECJ’s assertion that “European” legal questions should be directed to it via preliminary references thus reach quite different audiences, with implications for the role not only of the ECJ itself but of the national courts.

In this paper I address three conditionally related questions. First, is it possible to recover information about individual dispositions from the collective decisions of the ECJ? Second, are these individual positions predictable based on other attributes known about the judges? And third, do individual judges at the court evince responsiveness to their own member state governments? The short answer to all of these questions is yes: ECJ judges have heterogeneous preferences, which correlate weakly with the left–right positions of their appointing governments, and they are slightly more likely to decide cases on the side their home government prefers.

In the next section, I discuss the organization of the ECJ and how its rules of procedure make it possible to answer these questions. Then, I describe a novel statistical model that permits inferences about individuals from aggregate decisions like those of the ECJ. Next, I present estimates from the model of judges’ preferences, as well as predictors at the case, judge, and case-judge levels.

1.1 Collective Decisions by Diverse Membership

Any theory about the Court's behavior implies either a set of preferences or constraints that affect the individual members of the court. Mostly, they assume that the judges have a uniform preference for greater supranational governance, even though they may be constrained by the threat of override or noncompliance.

The basis for thinking about the ECJ as a strategic actor in European Union politics is that the court has preferences worth measuring: preferences over integration and the balance of Commission, EP, and national power. Moreover, the preferences of the court are endogenous because the court comprises judges appointed by member states. The incentives provided to the judges by their appointing principal government vary considerably, so much that some judges may appear to drift far from their government's wishes. Even so, thinking about the court as a collection of national judges is an advance over treating it as either an oracle of the true law or as a sneaky partner with the Commission for integration.

In general, institutions that perform constitutional review are designed to balance two goals: *representativeness* and *independence*. Many such courts – among them the ECJ – derive representative legitimacy through diversity of appointment procedures. The judges of the Constitutional Court of Spain, for just one example, are appointed by the legislature (four from each chamber), the executive (two), and the regular judiciary (two). Independence of constitutional judges is usually considered a function of appointment length (though lifetime appointments are rare) and difficulty of removal. Helmke (2005) has considered other potential limits on independence as well, such as judges' salaries or resources available to the court. In addition, many courts (constitutional and otherwise) tip the balance toward independence by shielding individual judge-decisions from scrutiny. The feature of collective decisions taken by

diverse membership is a common one in institutions that balance independence and representativeness.

Chambers or subsets of collegial judges – those that do not make public their internal disagreements through either votes or published dissents from decisions – enable a body of judges to divide work among its members. This increases efficiency, and may bolster the reputation and legitimacy of the court as a whole by shielding individual judges from public scrutiny. However, this “shield” rests on an assumption of uniformity among chambers, and more importantly, among judges. The ECJ in particular uses a nonrandom method of assignment (cases are assigned by the President of the Court) and stable (not randomly constituted) chambers. Both of these facts should be troubling to scholars, and the public, who view the court as a monolithic body.

The system of chambers at the Court of Justice has evolved over time from an innovation in the Rules of Procedure and some codification in the Treaties back to greater latitude for the Court to self-regulate. Although the membership was ostensibly fixed for two to three years at a time, enlargements, retirements, and replacements create more variation even among older cases than the two-year duration would suggest. In fact, though the court’s rules would suggest about between 50 and 70 different combinations should exist in the 1549 decisions analyzed in this paper, decisions were rendered by 296 distinct aggregations of judges.

Every instance where judges names are included in a judgment is parsed to a participation table. Older Reports of the Court do not consistently report, for each case, the hearing judges. Many cases still have to be coded by hand from signatures listed in the print Reports of the Court prior to 1986. Newer cases always indicate which judges heard the case, reporting “the Court, composed of...”

Listing the hearing judges became more important with a rule change that coincided with the Nice treaty in 2004. Chambers became pools with rotating judges, introducing more variation in

hearing judges. With enlargement, the Court first added a chamber of three and then moved entirely to the rotation-in-pools system. Most cases are heard by five-judge chambers including the Chamber President, the Judge-Rapporteur of the case, and three members drawn from the chamber's pool.

The fact that many or all ECJ judges have decided with multiple subsets of their colleagues gives us hope that a model can tease out individual behavior from these combinations. The model would not work if the chamber membership were less variable than it is, or if we did not know the exact composition of judges hearing the case.¹ The chamber subsets of judges at some other courts are often randomly constituted from the pool of judges. Indeed the ECJ appears to be moving slowly toward a model with greater randomization. Though Chambers of three judges are rarely, if ever, used anymore, the greater variation in composition of Chambers of Five (and the greater volume of cases) should increase, or at least not decrease, the precision of our estimates of their ideal points.

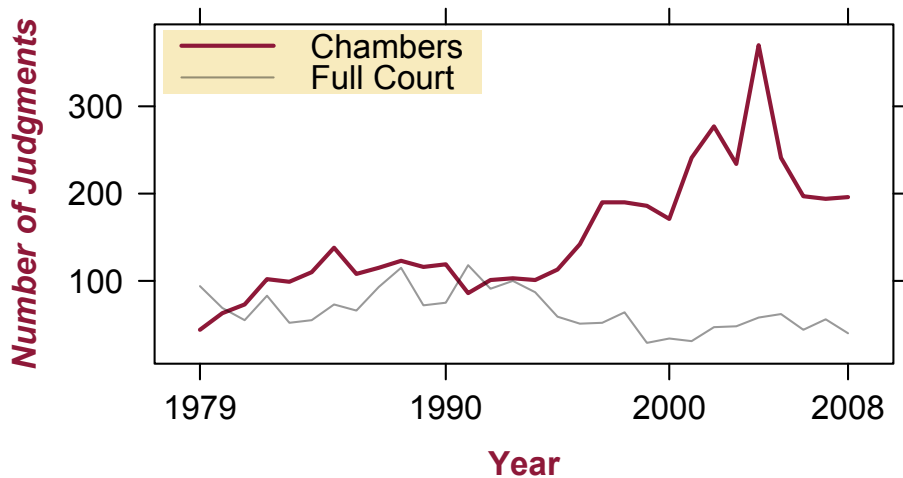
2 Ecological Item-Response Model

2.1 Motivation and Intuition

The item response model is a standard tool for analyzing legislative voting histories and estimating legislators' preferences (Jackman, 2001; Clinton, Jackman and Rivers, 2004). Packages such as Jackman's "pscl" and Martin, Quinn, and Park's "MCMCpack" (Martin, Quinn and Park, 2003-) have made item-response models usable, and political scientists have brought such methods to bear on an ever-widening variety of data. The standard specification is consistent with a

¹In fact this is true for some early cases when the chamber number is indicated but the judges are not. Work is underway to examine the printed records of the Reports for Hearing and final Judgments to recover individual judge participation where the signature portion of Judgments has been removed from electronic records.

Figure 1. The total number of decisions by chambers of judges or the full court. After 2004 in the graph, the Grand Chamber is counted with the full court. In recent years the ECJ has heard fewer than 10 cases as a full court.



quadratic-loss utility function around a so-called ideal point. The outcome modeled here is whether the Court agreed with the Commission. Agreement with the Commission is assumed to be a *pro-integration* decision (Stone Sweet and Brunell, 1998).

The Bayesian framework naturally considers the ideal points as distributions rather than point estimates. Applications in political science have often focused on the ready availability of standard errors for ideal points (Han, 2007) and the strength of Bayesian methods with small datasets (Hagemann, 2007). Modeled subject parameters have been suggested as a way to identify the model, but not generally as parameters of interest themselves (Bafumi et al., 2005). Because we view the item-response framework as a model of voting behavior, rather than a scaling exercise, the advantages of including relevant prior information are obvious. It would be foolish to ignore or discard directly relevant information, such as possible loyalty or prospects career advancement in the national setting, from a model of judicial behavior. Although some earlier critiques of

Bayesian methods were skeptical of the explicit inclusion of prior knowledge for having too much influence over parameter estimates and model fit, today it is more widely recognized that many of the assumptions underlying classical techniques can be thought of as “special cases” of Bayesian priors, and often ones that researchers would not adopt if forced to specify them in a Bayesian framework. In addition, hierarchical modeling is much more straightforward in the Bayesian setup, where parameter distributions are modeled, than in the classical one, where attention focuses on correlation structures of “error” terms.

Judicial scholarship focusing on the behavior – the voting behavior – of individual judges has been developed mostly in the context of the United States Supreme Court. It is well understood and widely accepted that justices have preferences and vote on case outcomes consistent with those preferences. Other work has investigated individual judges’ behavior at lower courts, mostly the courts of appeals but also federal district courts. We rarely think about “the Supreme Court” speaking with one voice, bargaining over policy with Congress and the Executive. Certainly this is due in large part to published dissents and on-the-record individual justice-votes. But in their absence, should we assume its decisions are always unanimous? What are the implications of a divided Court of Justice behind the shield of *en banc* decisions? What would be optimal strategies for national governments and the European Commission if they considered individual judges as strategic actors?

Judges at other courts, most notably the European Court of Human Rights and the International Court of Justice, have been treated as individual strategic actors. Whether this is due to their disclosure of votes or the courts’ relatively ambiguous strategic environment is unclear. In comparison to these others, the ECJ’s environment, with the Commission, Council, member states as parties, interveners, or observers in cases, is far more complicated.

The findings of Posner and de Figueiredo (2005) that the International Criminal Court judges

exhibit bias with regard to their home countries and others similarly situated gives hope for how member states might see their interests reflected. Voeten (2007) finds similar patterns at the European Court of Human Rights (ECtHR), with some interesting implications for the ECJ. Four current ECJ justices, all from post-Communist states, previously served at the ECtHR, where their individual decisions were observable. This result is curious in light of Voeten's finding that such judges at the ECtHR were systematically more likely to find against other post-Communist states.

The fact that after the enlargement of 2004 judges who had previously served at the ECtHR were the first appointees to the European Court of Justice is evidence of a new opportunity for advancement for international judges from those countries.

Some scholars consider features of judicial appointment as evidence of courts' independence or subordination to other institutions. The variation in pre- and post-ECJ service of judges from different member states highlights the difficulty of considering longer terms or lifetime appointments as evidence of independence. In effect, most judges from Belgium and the Netherlands are renewed until they die or retire – renewals have occurred as often as not by caretaker governments or delicately balanced coalitions.

By contrast most judges from Germany and the United Kingdom are replaced if the government has changed hands since their appointment. France has tried both perpetual renewal and prompt replacement. The judges appointed from the 10-state 2004 enlargement are up for renewal or replacement in the coming year.

The model's intuition is a subtle generalization of a standard probit model. An outcome of "1" results from a function of latent case and judge parameters. Thus we note the similarities to the case-level probit analysis of Gabel, Carrubba, and Hankla (2008) where only case-level (rather than judge- or case-judge-level) effects can be estimated.

The data observed comprise a single outcome, attributed to a chamber of judges, per legal issue.

A complete “case” before the Court may encompass several legal issues, with a ruling on each of them. Furthermore, observations tendered by the Commission or member states may pertain to one or several legal issues. The fact that cases may contain several legal issues is irrelevant and potentially confusing, since the data are observed at the level of legal issue. I will therefore use the terms “case” and “legal issue” interchangeably to mean “the object upon which observations are submitted, decisions are made, and outcomes revealed.”

2.2 Model

The parameters of interest in the model are judge ideal points and case location parameters.

Define the set of judges $j = 1, 2, \dots, J$. A judge’s ideal point is denoted θ_j .

The usual “two-parameter” IRT model is so called because each item k has both a location and relevance parameter (a and b respectively), and is fully described by $2 \times K + J$ parameters estimated from $J \times K$ data points. The relevance parameter is analogous to a factor loading. In practice, researchers often examine which bills have high relevance to bolster their claims about the measured latent ideal points. (High relevance of items relating to government involvement in the economy, for example, support the claim that measured ideal points pertain to left–right ideology.) The outcome I model here, pro-Commission decisions, lets me assume constant relevance b and fit a “one-parameter” or Rasch model, estimating only location.

Formally, let $k = 1, 2, \dots, K$ be the vector of cases (legal issues) and $y_k \in Y$ be the observed decisions:

$$y_k = \begin{cases} 1 & \text{if pro-EU} \\ 0 & \text{if anti-EU} \end{cases} \quad (1)$$

The model deals with the probability that an outcome y_k is observed to be 1; in English, a

decision on a legal issue is in favor of the Commission. Formally we model

$$\Pr(y_k = 1) = \Phi(f(\theta^g, a_k)), \quad (2)$$

which is a mapping into a probability of case parameters and judge parameters. Now we turn to a discussion of f .

Denote the judges on a given chamber j^g and all others j^{-g} . Thus each case is informative about θ^g and not about θ^{-g} . It is worth noting that in contrast to the typical item-response framework, we observe a single outcome per case, netting K total data points, rather than the $J \times K$ for individual justice-votes. To identify the model we must take care to restrict the number of parameters estimated below this upper bound. The reason we have any traction on this model at all is through the organization of judges into chambers and the rotation of members on chambers. Each case is heard by one and only one chamber comprising a subset of judges. In a hypothetical one-judge chamber $f(\cdot)$ would simply be the probability that judge j sided with the Commission. But the smallest chamber is 3 judges, so label for each $j \in g$ this probability as ω_{jk} :

$$\omega_{jk} = \Phi(a_k + \theta_j). \quad (3)$$

.

To complete the definition of f , aggregate Ω over the j^g judges. The simplest chamber-judge allocation function is the mean of ω ,

$$\frac{\sum_{j=1}^{\#j^g} \omega_{jk} | \theta_j \in \theta^g}{\#j^g}. \quad (4)$$

The quantity in (4) could also also be thought of as the expectation of a multinomial outcome,

where the number of possible outcomes is the number of judges in the chamber who vote pro-Commission. That is, on a chamber of three judges, this could be 1, 2, or 3 votes in any combination. The expectation of each outcome in a multinomial distribution is the expectation of an independent binomial distribution. Another reasonable alternative, but not a smooth function, is for f to return the probability that the median judge voted pro-EU.

The chamber-judge weighting function –mean, median, or other – is not a purely technical point; it relates to procedures at the Court itself. For example, if the judge-rapporteur is more strongly influential on case outcomes than her peers, the model should make a judge’s “own” cases contribute more heavily to the estimate of preference. Alternative specifications of the chamber-judge weighting function will be explored in future research. Using the chamber mean, we nonetheless are able to control for some judge-specific effects by looking for case-judge-specific variations.

In terms of implementation, we use a “participation matrix,” where entries are 1 for judges hearing a case and 0 for all others. The contribution of each case for those judges is the result of a typical data-augmentation draw from a truncated normal distribution (truncated at 0 from below when $Y = 1$ and at 0 from above when $Y = 0$), with the mean from equation (3). All other judges on the given case are considered missing and the data are uninformative; conversely all other cases for a given judge are missing and uninformative. Missing data result in sampler draws from a standard normal distribution.

Typically, both item-response models and other techniques such as NOMINATE require for identification that at least one subject be either fixed at a point (most liberal, most conservative) or constrained to positive or negative values. That is, normally all θ are sampled from the same $\mathcal{N}(0, \sigma_\theta^2)$ distribution, except for some, which are fixed at a point or constrained to be positive or negative. Such a strategy incorporates a small amount of prior knowledge, enough to identify the

model – but not in the case of grouped data. In the typical setup we learn what the latent θ are, but practically nothing about them other than their distribution. The hierarchical setting enables us to incorporate directly knowledge that we have about all the judges and also learn about the relationship between these characteristics and their observed behavior.

Using a hierarchical model specification, we simultaneously fit the data more closely while making weaker assumptions. Thus, we will assume that ideal points θ have a person-specific mean and common variance:

$$\theta_j \sim \mathcal{N}(X_j\beta, \sigma_\theta^2) \quad (5)$$

$$a_k \sim \mathcal{N}(C_k\gamma, \sigma_\gamma^2) \quad (6)$$

$$\begin{aligned} \beta &\sim \mathcal{N}(b_o, B_o), & \gamma &\sim \mathcal{N}(g_o, G_o), \\ \sigma_\theta^2, \sigma_\gamma^2 &\sim \mathcal{IG}(v/2, v/2). \end{aligned} \quad (7)$$

where x is a vector judge covariates and c a vector of case covariates. Under this model the ideal point is effectively constrained to an interval subset of θ , and β tells us the effect of each judge attribute. The hierarchical model more closely resembles the typically identified IRT model when we fix values $\sigma^2 = 1$, and as few as one j to have $X_j\beta = \{-1, 1\}$ and for all others $x = 0$. Even then these are weaker assumptions than the strong constraining identification strategy, which actually fixes θ for some j . By varying the location and precision of the parameters β and γ , we can detect the sensitivity of the model to the prior.

The intuition and interpretation of the second-level parameters is straightforward and powerful:

- Ideal points are estimated using both person-level predictors and a vector of outcomes. Variance σ_θ^2 decreases as the effect of $X_j\beta$ explains vote choices. Put simply, to the extent that β explains vote choices, the lower the uncertainty contained in σ^2 .

- The model reduces to a standard unidentified ideal-point model as $X\beta \rightarrow 0$ (that is, all ideal points would be draws from the same normal distribution with mean μ_θ).
- Predictions about ideal points (and behavior) can be made given only the elements of x , judge characteristics. In other words, given absolutely no voting history, the ideal point of a new member can be estimated, with appropriate uncertainty, knowing only those attributes that make up x . Predictions of voting behavior could be made by specifying the location of a number of cases.

3 Data and Results

The model has a huge number of “moving parts,” or sampled posterior distributions. They are needed, though, to answer the central questions of this research. First, do the data reveal heterogeneous judge preferences? This is answered by considering the marginal distributions of judge parameters. Second, can we predict judge positions with other covariates? We consider two separate judge-level variables that enter the model as linear predictors of judge positions. And third, do judges show signs of greater responsiveness to their own governments’ observations? We extend the model described above to account for attributes of cases to affect some but not all judges in the chamber.

The data comprise 1549 Article 234 cases – preliminary references – from 1970 to 1997, in which the Commission announced a position via a written observation and argued its position at the hearing. These cases were heard by 296 unique combinations of judges; the median number of cases per chamber is 2 and the mean 5; Judges Zuleeg, Montinho de Almedia, and Grévisse, sat as a Chamber of Three in 95 cases.

Larger configurations of seven, nine, eleven, or thirteen judges have also existed but were used

far less frequently and not at all after the 2005 reform of the Rules of Procedure. The model, of course, is agnostic of formal membership in chambers or the order of rotation – for all (and only) the judges hearing a given case, it contributes to the estimate. I also present one model that includes infringement cases brought by the Commission against a member state or member states against each other.

The Commission makes and supports more observations than any other party, but it is frequently joined or opposed by other governments. The presence of other member state governments is used to predict the location of the case-specific parameter of the model. Future work considers a more sophisticated coding that takes into account whether the Commission is alone or joined by the others.

Overall, the Court sides with the Commission in Preliminary References about 75% of the time. The model depends on the composition of the chambers deciding against the Commission. Accounting for the participation matrix, judges side with the Commission in these cases from 72% to 85% of the time.

3.1 Hypotheses

Judges appointed by national governments must be minimally acceptable to the government making the appointment. The left–right ideology of mainstream political parties in Europe has been demonstrated to correlate strongly with positions and policies regarding European integration. Although extreme parties of both the left and the right often strongly oppose integration, within the mainstream – those parties that govern or join governing coalitions – left parties are consistently more supportive of supranational governance than right parties. I expect that this relationship holds in the judges they would appoint to the ECJ. As a measure of the left–right orientation of the government at the time of first appointment, I use cabinet-share

weighted manifesto scores from Kim and Fording (2002).

Hypothesis 1: Judges Judges appointed by right or anti-EU governments will be more likely to decide cases against the Commission. *Positive* values β increase the “distance” in the latent utility space, resulting in a lower propensity to vote in favor of the Commission.

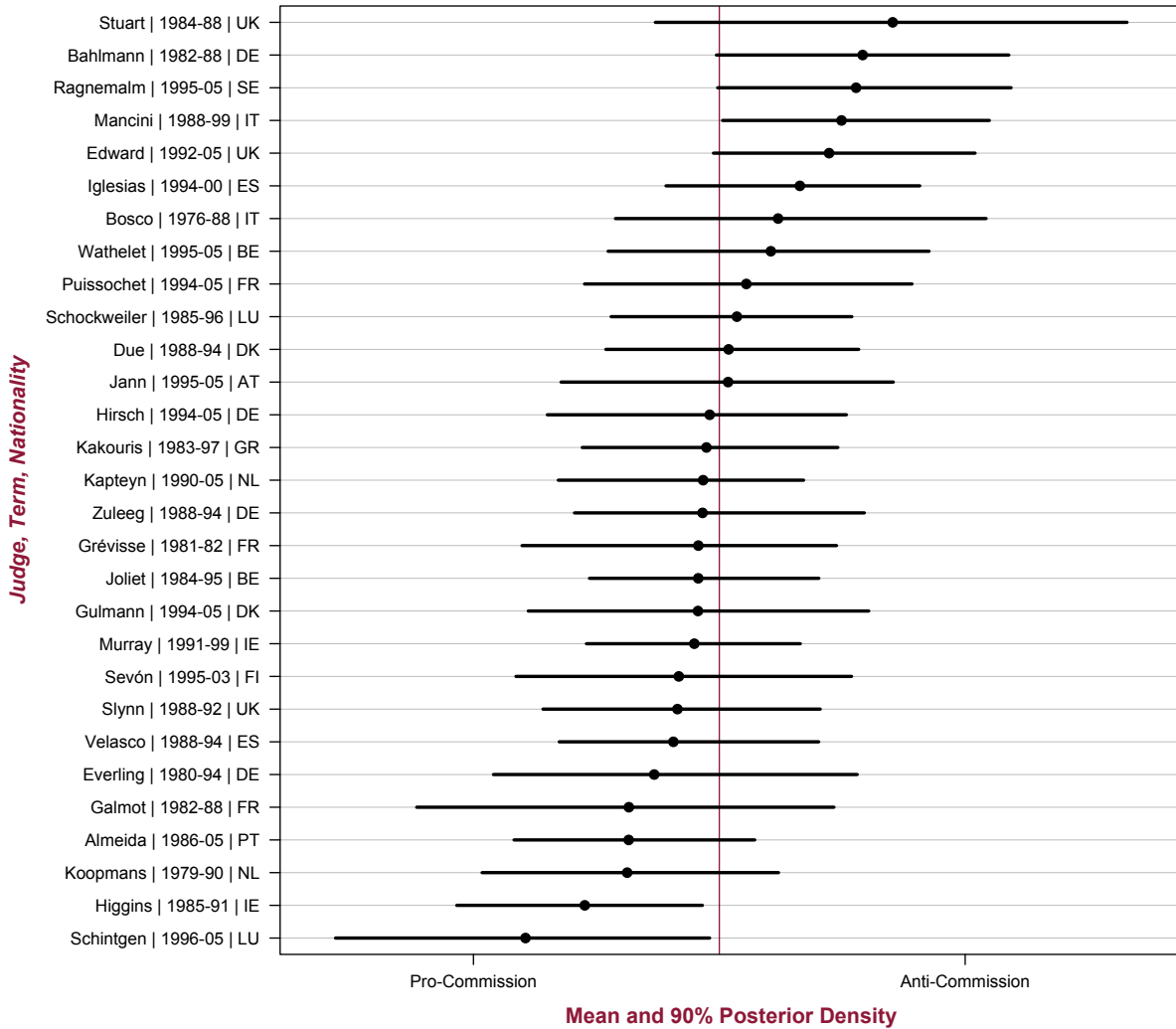
Hypothesis 2: Cases Some attributes of cases apply to all the judges on the panel. Carrubba, Gabel and Hankla (2008) suggested that member state observations may signal the threat of noncompliance and showed that the balance of observations is a strong predictor of case outcome. The net number of observations with respect to the Commission (opposing minus supporting) thus predicts the location of the case parameter in the latent space. *Negative* values of γ would be consistent with the findings in Carrubba, Gabel and Hankla (2008).

3.2 Results

Figure 2 shows the estimated ideal points of judges from 1549 preliminary references and the predictors discussed above for judge and case positions. The presence of government observations is a meaningful predictor of case position (how much a case is pro- or anti-commission). The appointing government left–right position is also a weak but generally positive predictor of ideal points.

The prior distribution of judge positions, predicted by the right–left position of their national governments at the time of appointment, is quickly overwhelmed by the data. This result increases our confidence in the measurement model. At the same time, we have tested a theory and now have evidence that the left-right composition of appointing governments is a very weak predictor of judicial behavior at the ECJ. Most importantly, this result is robust to the precision of the prior

Figure 2. Estimated ECJ judge positions (θ) from ecological item response (EIRT) model.



and alternative starting values.² Even when we begin with relatively strong or weak beliefs about judge parameters, the data imply a posterior distribution like the one in Figure 2.

²All the models in the paper have attained conventional measures of convergence, often relatively quickly and with very low autocorrelation thanks to a parameter-expansion sampling scheme. All starting values are random draws at the time of initialization and multiple chains yielded similar posterior draws; only draws from one chain are presented. The variance of hyperpriors – the spread of β and γ – was set at 1, 10, and 100 with similar results. The models presented have a relatively “flat” prior variance of 10 for both β and γ .

The variation in chamber composition has enabled us to separate the likely effects of the different groupings, controlling for the attributes of the case that would affect any chamber equally. However, because we do not observe every possible combination of judges, we learn more about some judges than others. (As I noted earlier, the chamber of Zuleeg–Almeida–Grévisse heard more cases together than any other grouping and is thus poorly separated.)

Figure 3. Other correlates of judges’ preferences: the percentage of cases they heard on any panel in which the outcome was in favor of the Commission; and the right–left ideology of their national government at the time of their appointment. The latter is used as a predictor in the hierarchical model. The scale is reversed so that the theoretical prediction that right or conservative governments appoint more Euroskeptical judges. The ordering below is the same as in Figure 2.

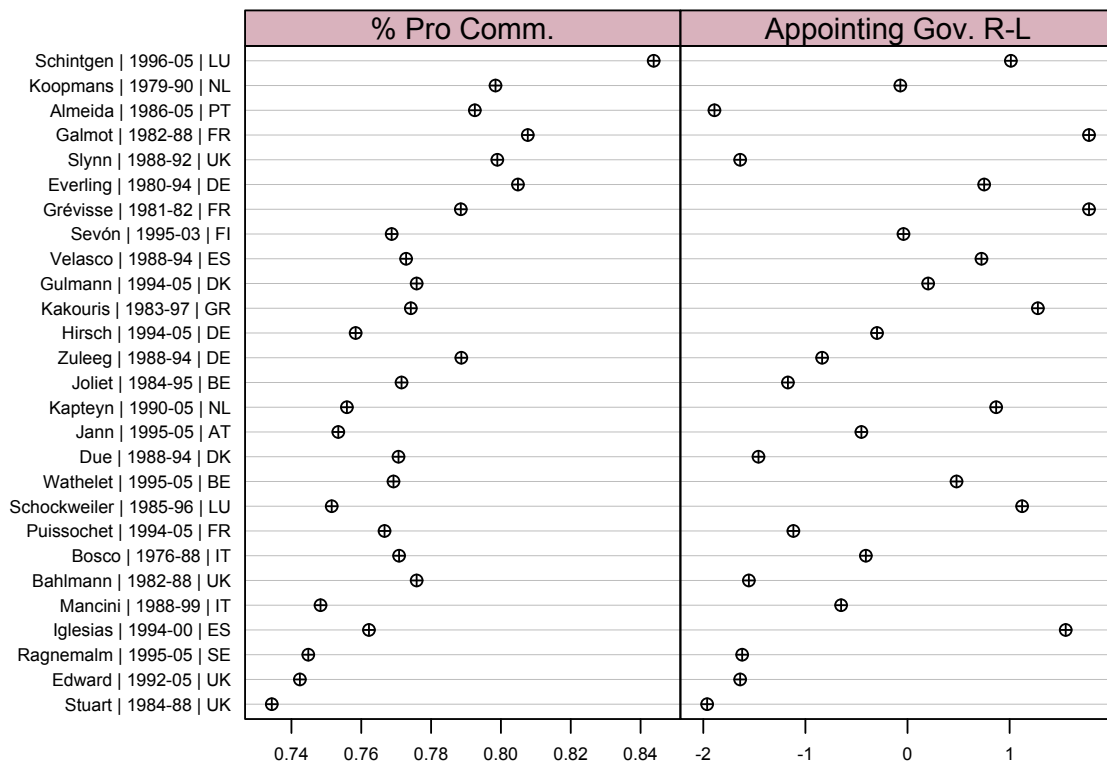
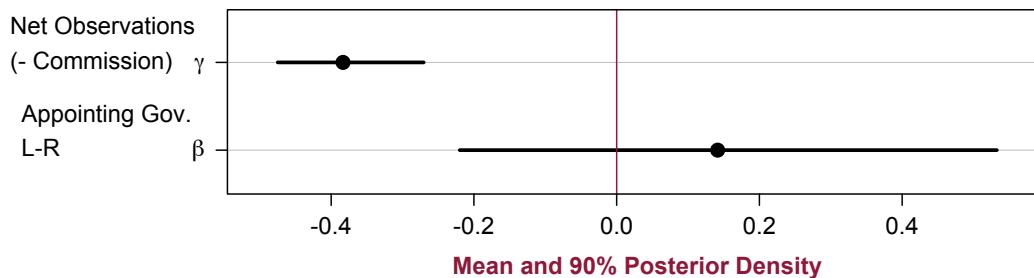


Figure 3 plots the two other measures of judge positions (percentage of cases siding with the

Commission and appointing government R–L score from Kim and Fording (2002)). Again, judges are ordered according to the EIRT measure. Reassuringly, the EIRT measure tracks this raw pro-Commission measure, but it is not identical. At the same time, it makes sense that the judges whose chambers most frequently ruled against (for) the Commission have the most extreme ideal point estimates. The judges with the longest terms are estimated with greater precision than those with shorter histories.

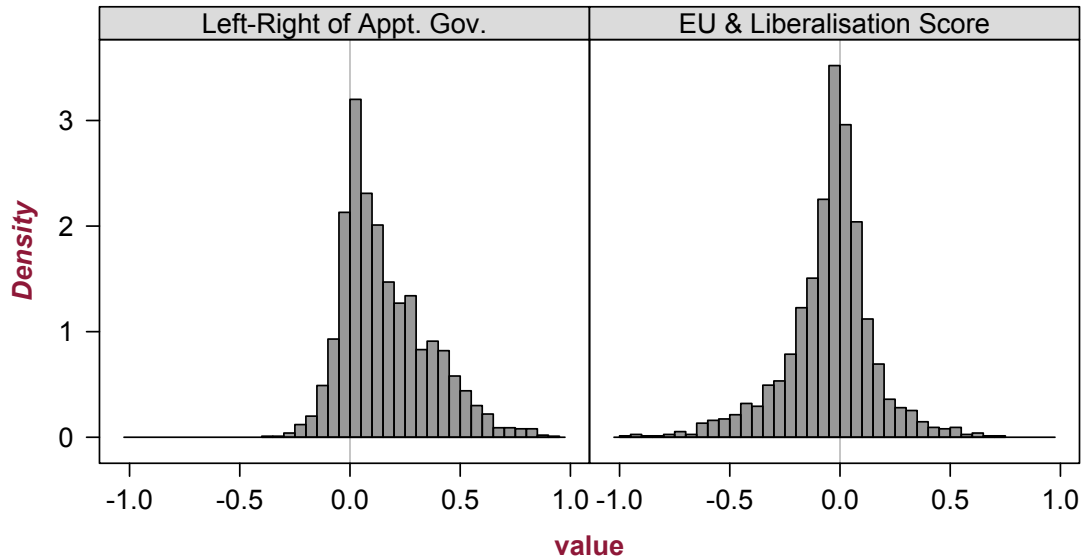
Figure 4. Prior predictors β and γ for judges and cases, respectively.



Estimates of the regressions of case and judge covariates on the hierarchical priors are shown in Figure 4. Above, I noted that the posterior distribution of second-level prior predictors, in particular those pertaining to judges, could be used to test theories about the influences on judicial behavior. The posterior distribution of β , the vector of coefficients in the regression of predictors on ideal point estimates, allows us to do exactly that. Figure 5 shows two alternative covariates to predict judge behavior: the left–right cabinet median discussed earlier, and a similar cabinet median of a similar normalized index of two other items coded in the Comparative Manifesto Project dataset: position (positive or negative) on the European Communities / European Union, and position on liberalization versus protectionism.³

³This alternative “integration index” is available at the same level of precision (all countries in all years) as the

Figure 5. Posterior density plots of two prior predictors of judges' preferences.



The posterior distributions of these hierarchical prior predictors produce two notable findings:

- 1) the distribution of β_{LR} is heavy at 0, with another peak around 0.4. It is 80% on the positive side of zero, indicating a weak prediction that judges appointed by right governments are more likely to hand down judgments against the European Commission and judges appointed by left governments are more favorably inclined toward the Commission's position. The shape of the posterior distribution refutes the claim that some left governments would appoint protectionist judges and some right governments free-market crusaders: these claims would together imply negative β . Instead it is apparent that for some judges the appointing government has no predictive power and for others it is stronger.⁴
- 2) The left–right positions of parties in government

Manifesto-based left–right index. It correlates weakly with the L–R measure and identifies the same countries as being “anti-integration” as expert surveys that only classify parties/governments trichotomously. A trichotomous coding (pro, anti, neutral) produces almost identical results to those presented in Figure 5 with the continuous Manifesto-based covariate.

⁴Indeed oddly-shaped posterior distributions like the one for β_{LR} above is a major advantage of Bayesian modeling: we do not have to assume that the effect is normally distributed about some value of β . A mixture makes sense, given the

is a better indicator of the kinds of judges governments will appoint than is a comparable Manifesto-based indicator of their positions on the European Union and market liberalization. Other expert assessments of left–right or pro–anti-integration correlate strongly with the measures used here, and do not do a better job predicting judges’ observed behavior of siding with or against the European Commission, taken as an indicator of support for integration or regulation at the European rather than national level, which is centered at -0.03 and almost perfectly symmetric about zero.

We also tested whether the number of observations of more national governments predicted case positions. The quantity γ is our estimate of the influence of each additional government on the “location” of the case. It is the unit effect that applies to all the judges in the chamber. Consistent with previous findings, observations from national governments are a strong predictor of case outcomes.

3.3 Judge-Specific Observation Effects

The next model estimates the effect of a judge’s own country’s observations on outcomes of preliminary references. A $(k \times j)$ matrix is constructed where entries D are 1 when a judge’s (columns j) member state submits an observation on the same side as the Commission in case k , -1 when a judge’s member state submits an observation against the Commission’s position. This type of model which estimates an effect conditionally for some subjects on some items is typically called “differential item functioning” and implies the following model:

$$\omega_{jk} = \Phi \left(a_k - \frac{\sum_{j=1}^{\#g} \theta_j | \theta_j \in \theta^g}{\#j^g} - \delta_{jk} D \right) \quad (8)$$

variations in appointment patterns described earlier. It is likely that countries where ECJ appointments are politicized and contentious are the ones contributing higher positive values of β here. Coding and controlling for these country-level variations is the subject of future research.

where a is still predicted by γ times case covariates c that affect the whole panel, and θ by judge covariates x that affect the judge on all cases with strength β .

The number of nonzero entries in D ranges (by judge) from 0 to 95, with a median and mean both around 25. It assumes that observations for and against the Commission have the same weight. Not surprisingly, controlling for these observations increases slightly the precision of the estimates of judges' positions, shown in Figure 6. The order of a few is changed, but in broad strokes the results are the same.

Hypothesis 3 Observations from a judge's own government sway opinion independently and more strongly than those from other governments. A *negative* value of δ is evidence of responsiveness in the expected direction given the coding of D .

The purpose of estimating the case-judge-specific effect δ was to test the theory that national governments influence their judges through the observations they submit. Indeed, this is the case, as the distribution of δ and β in Figure 7 shows.

The wide, overlapping bands in Figure 2 understate just how much information the model has actually recovered from the permutations of chambers. Although we have not observed a single within-chamber judge-vote, the ordering and position of judges relative to one another is described by the pairwise comparison of their marginal distributions. The matrix of these pairwise comparisons is shown in Figure 8.

Figure 6. Estimated ECJ judge positions (θ) controlling for the presence and position of their own governments' observations in cases.

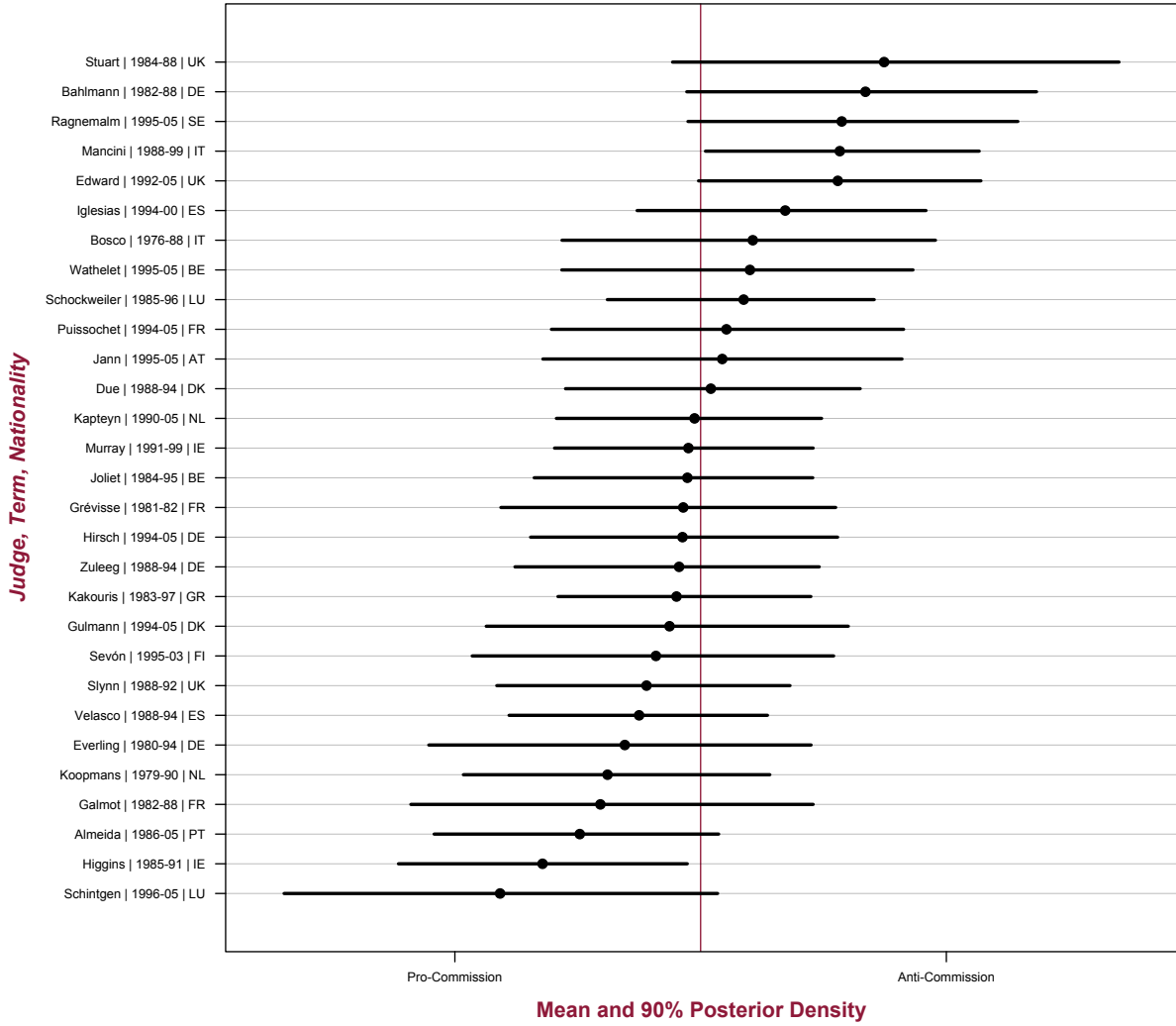


Figure 7. Prior predictors β , δ , and γ : judge positions, the effect of home-government observations, and chamber effects of all observations.

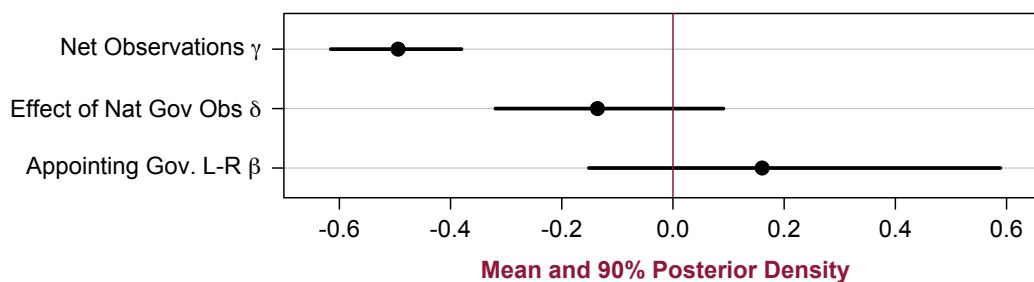


Figure 8. The matrix of pairwise comparisons of judges' preference estimates. Numbers indicate the degree of certainty of the ranking of the row relative to the column. For example, we are 63% certain that Schintgen lies to the left of Higgins and 80% certain he is left of Everling. Even though both ends of the joint distributions in Figure 6 – Judges Schintgen and Stuart – overlap each other, we are over 99% certain that Stuart is to the right of Schintgen.

[image removed for size: see malecki.wustl.edu/ecoirt.pdf]

4 Conclusion

This paper set out to address three questions about the politics of decisionmaking at the European Court of Justice, to shed light on the class of political institutions called constitutional courts.

Do judges reveal observable patterns of heterogeneous preferences? Yes, though as in most item-response models, the positions of adjacent judges are sometimes ambivalent. Two types of null result were possible: the distributions could exhibit greater, or even total, overlap; or they could be overly sensitive to the prior information. Instead, the preference estimates of the judges is determined by the data rather than the prior distributions.

Second, can we predict the dispositions of the judges based on other information known about them? Yes. Consistent with expectations, the median left–right position of the government that initially appointed each judge is a weak predictor of judges’ preference estimates. An alternative measure of government positions based on the median party’s position on the EU and trade liberalization, did not predict the judges’ positions at all.

Third, do judges respond more strongly to observations submitted by their own home governments? Again, evidence supports the claim that judges favor the position of the observation of their own governments, whether supporting or opposing the Commission, over and above other governments’ observations.

Most normative and positive theories about constitutional review are based on the incentives presented to individual judges. Institutions are designed to safeguard independence, but constraints such as the threat of override and the threat of removal (or non-reappointment) also encourage more responsive behavior. Theories about the ECJ have been stymied by the ecological problem of collective decisions, which is a common attribute of “independence” especially for constitutional courts.

The model developed here uses the organization of the court into chambers of various compositions to infer individual judges' preferences. To date, little research has conceptualized the court as a collection of individuals each with interests and preferences. The EIRT model developed here offers new, more rigorous tests for existing theoretical hypotheses about the ECJ's deference to certain actors and preference for integration. The complex interactions of national politics, integration, and enlargement offer a rich vein for future research.

Finally, and most broadly, I would speculate that some of the ECJ's design features – above all the multiple sources of appointment – that it shares with other constitutional courts probably has similar effects in other settings. Whoever makes appointments can be expected to appoint like-minded individuals as judges; judges, in turn, appear to behave in both representative and responsive ways.

References

- Bafumi, Joseph, Andrew Gelman, David K. Park, and Noah Kaplan. 2005. "Practical Issues in Implementing and Understanding Bayesian Ideal Point Estimation." *Political Analysis* 13:171–187.
- Burley, Anne-Marie, and Walter Mattli. 1993. "Europe before the Court: A Political Theory of Legal Integration." *International Organization* 47:41–76.
- Carrubba, Clifford J., Matthew Gabel, and Charles Hankla. 2008. "Judicial Behavior under Political Constraints: Evidence from the European Court of Justice." *American Political Science Review* 102:435–452.
- Clinton, Joshua, Simon Jackman, and Douglas Rivers. 2004. "The Statistical Analysis of Roll Call Data." *American Political Science Review* 98:355–370.
- Gibson, James L., and Gregory A. Caldeira. 1995. "The legitimacy of transnational legal institutions: Compliance, support, and the European Court of Justice." *American Journal of Political Science* 39:459–89.
- Gibson, James L., and Gregory A. Caldeira. 1996. "The Legal Cultures of Europe." *Law & society review* 30:55.
- Gibson, James L., and Gregory A. Caldeira. 1998. "Changes in the Legitimacy of the European Court of Justice: A Post-Maastricht Analysis." *British Journal of Political Science* 28:63–91.
- Hagemann, Sara. 2007. "Applying Ideal Point Estimation Methods to the Council of Ministers." *European Union Politics* 8:279–296.
- Han, Jeong-Hun. 2007. "Analysing Roll Calls of the European Parliament: A Bayesian Application." *European Union Politics* 8:479–507.
- Helmke, Gretchen. 2005. *Courts under constraints: judges, generals, and presidents in Argentina*. Cambridge University Press.
- Jackman, Simon. 2001. "Multidimensional Analysis of Roll Call Data via Bayesian Simulation: Identification, Estimation, Inference, and Model Checking." *Political Analysis* 9 (June):227–241.
- Jupille, Joseph. 2004. *Procedural Politics: Issues, Influence, and Institutional Choice in the European Union*. Cambridge University Press.
- Kim, HeeMin, and Richard C. Fording. 2002. "Government partisanship in Western democracies, 1945-1998." *European Journal of Political Research* 41:187–206.

- Martin, Andrew D., Kevin M. Quinn, and Jong Hee Park. 2003-. "MCMCpack: Markov chain Monte Carlo (MCMC) Package." *R package version 0.9-5*, URL <http://mcmcpack.wustl.edu> .
- Posner, Eric A., and Miguel F. P. de Figueiredo. 2005. "Is the International Court of Justice Biased?" *Journal of Legal Studies* 34:599-630.
- Stone Sweet, Alec. 2000. *Governing with judges*. New York: Oxford University.
- Stone Sweet, Alec, and Thomas L. Brunell. 1998. "Constructing a supranational constitution: Dispute resolution and governance in the European Community." *American Political Science Review* 92:63-81.
- Voeten, Erik. 2007. "What Motivates International Judges? Evidence from the European Court of Human Rights." 2007 American Political Science Association Annual Meetings.