

Judicial Positions on Political Reform

Designing common policy scores from judicial text

Benjamin G. Engst
University of Mannheim
Collaborative Research Center
engst@uni-mannheim.de

David M. Grundmanns
University of Mannheim
Collaborative Research Center
dgrundma@mail.uni-
mannheim.de

Thomas Gschwend
University of Mannheim
School of Social Sciences
gschwend@uni-mannheim.de

Abstract. What are the positions of highest courts towards political reforms? In order to answer this question scholarship on the US Supreme Court uses judicial votes but those are often not available in cross-country comparison. We present an approach scaling judicial decisions based on textual features other than votes. In particular, we argue that the legal outcome and tendencies field in briefs by scalable political actors can be used to compute a vote-matrix similar to roll-call data in legislative studies. With the obtained data we are able to estimate location scores employing a standard two-parameter item-response theory model using R `STAN`. Moreover, computing priors from the known manifesto positions of political actors allows as to meaningful anchor the court's locations in relation to the political positions in one common policy-space. The procedure applies independently of specific judicial systems but to outline the feasibility, we assess 100 senate decisions by the German Federal Constitutional Court. Our computed spatial measures allow for assessing the judiciary's willingness to align or diverge from the opinion of governing parties and their political reform. Moreover, the measure helps to validate common theoretical expectations which are based on the assumption that courts are strategic (political) actors.

We thank Linda Schill and Vanessa Müden for valuable research assistance and the Collaborative Research Center 884 at the University of Mannheim, funded by the German Research Foundation, for generous research opportunities. The article is developed as part of the project "Measuring a common space and the dynamics of reform positions" (C4) located at the aforementioned research center.

1 Judicial influence on political reform

What are the positions of highest courts towards political reform? An answer to this question is essential, as scholarship has come to accept that highest courts – in particular those with the power of judicial review – are able to excise political power (see Parcelle Jr., Curry and Marshall, 2011; Vanberg, 2015; Engst, 2018). In order to understand the judiciary's precise political influence it is necessary to relate political ideologies expressed in judicial decisions to political action. In short, we need to identify the political-ideological leaning of courts. One approach to do so – widely used in studies on the US Supreme Court – is to scale legal opinions based on individual judicial votes and to relate those to the known political ideologies of individual judges (e.g. Martin and Quinn, 2002; Epstein, Martin, Quinn and Segal, 2007; Epstein, Martin, Segal and Westerland, 2007; Hanretty, 2012^{b,a}, 2014, 2015; Engst et al., 2017). This allows to identify to what extent individual judges will hinder or support political reform.

Nevertheless, to understand political leanings of highest courts based on judicial votes has three major drawbacks. First, the publication of individual judicial votes is by no means a common feature in cross-country comparison (Raffaelli 2012, 30, Kelemen 2013, 1345); powerful courts like the German *Bundesverfassungsgericht*, the Austrian *Verfassungsgerichtshof* or the Italian *Corte costituzionale* publish votes either never or rarely (Kelemen, 2013, 1345). Second, the judicial vote alone does not carry substantial information on the (political) content of a judicial decision. Instead, this information is embedded in the text of the decision itself. Therefore, in order to understand a court's position towards political reform we need to scale the legal doctrine developed within the binding text of a decision. Finally, even if we are able to extract certain doctrinal leanings in court decisions it is not plausible to simply transfer those into a policy space with political actors. This is due to the fact that judicial text is different to text used to estimate policy-positions; such as party manifestos.

In order to overcome these short-comings we present an approach to scale judicial

decisions based on features that are commonly included in decisions and are independent of judicial votes. Moreover, we design our approach in a way so that judicial scores from a legal case-space (Clark and Lauderdale 2010; similar Lax 2011, Clark and Lauderdale 2012) transfer into a common policy-space. The resulting estimates enable us to compare the political-ideological leaning of highest courts to the manifesto positions of political actors.

In what follows we first outline in what way ideological scores in judicial decision-making are currently obtained to emphasize our contribution. Afterwards, we argue that briefs submitted by political actors and summarized in the written decisions allow us to compile data similar to data on roll-call votes in the U.S. Congress. Subsequently, our data-structure allows us to adopt the research design used by Jessee (2016), who jointly scale different groups of actors, as our scaling-approach for decisions of highest courts. Finally, we use 100 decisions on social-economic topics published by the German Federal Constitutional Court between 1990 to 2005 to outline the feasibility of our scaling-approach.

The estimated scores have implications for our capability to integrate judicial institutions fully into analyses of political systems. Moreover, the approach designed to obtain the political-ideological leaning of courts is a general one which can travel to research on similar institutions, that are “political players” but do not portray themselves as such; e.g. central banks.

2 Identifying judicial (political) ideology

In this section we briefly outline that the literature relies on two approaches to understand the (political-)ideological leaning of courts: the assessment of (1) individual judicial preferences or (2) features of judicial decisions. Moreover, we discuss the limitations of both approaches in understanding the political influence of courts.

Seminal work by Martin and Quinn (2002) uses individual judicial votes to understand judicial behavior and the ideological leaning of the US Supreme Court (see also Epstein, Martin, Quinn and Segal, 2007; Epstein, Martin, Segal and Westerland, 2007). Hanretty adopts the approaches to assess judicial behavior in Bulgaria (Hanretty, 2014), Estonia (Hanretty, 2015), Spain and Portugal (Hanretty, 2012a), as well as in the United Kingdom (Hanretty, 2012b). Moreover, we used separate opinions to understand judicial behavior in Germany (Engst et al., 2017). However, the challenge of analyses relying on votes is twofold. On the one hand, individual judicial votes are not always publicly available (Raffaelli, 2012; Kelemen, 2013) and on the other hand, highest courts often follow a norm of consent (Epstein, Segal and Spaeth, 2001), which is why published votes not necessarily vary. An alternative solution to overcome those limitation is to infer the individual preferences of judges from outside sources. To do so, scholars must anticipate judicial preferences before they are revealed in the court. This involves manually coding newspaper editorials on congressional hearings for judicial nominees (Segal and Cover, 1989) or process-tracing expert judgments, historical accounts, and “other existing and relevant documentary materials” to infer a preference position (Epstein, Knight and Shvetsova, 2001, 140). Although text is more informative than votes, this strategy depends on indirect sources that are external to the court and requires in-depth (“thick”) knowledge of every political environment. Hence, two challenges which come with this approach are that, (1) outside sources may not adequately capture judicial preferences and (2) the execution is labor-intensive.

In addition, the overarching challenges when using individual judicial preferences to understand the influence of courts on political reform is a conceptual one: Does the individual political leaning of a judge adequately reflect on the *substantial* political influence of a court? Scholars often implicitly assume this, as behavioral research on countries that do not publish individual votes illustrates. In respective studies, scholars assign a court a political position based on the ideological views of the

median judge, which are derived, for example, from the political views of the actor nominating the median judge (Hönnige, 2009; Carrubba et al., 2012; Sternberg et al., 2015; Brouard and Hönnige, 2017). This is an approach driven by strong assumptions. Even if this may uncover general patterns in a court's behavior this does not allow to understand the substantial impact of a certain judicial decision on a political reform. This information can be obtained only when placing the particular doctrine developed in judicial decisions in a policy space. Hence, content-specific features of judicial decisions are essential in identifying the political-ideological leaning of a court.

The body of research focusing on textual features of judicial decisions is growing in recent years (e.g. Clark and Lauderdale, 2010, 2012; Lauderdale and Clark, 2014) and law scholars have advocated in favor of interpreting decisions to classify them (Hall and Wright, 2008). Indeed, there are now a number of (automated) approaches that allow to scale decisions on some latent dimension. Citation networks across legal decisions as done by Fowler et al. (2007) provide for a form of scaling (also Lupu and Fowler, 2013). In particular when accounting for the fact that Clark and Lauderdale (2010) combine the assessment of citation patterns with individual judicial votes to uncover doctrinal spaces in search and seizure cases as well as decisions on the freedom of religion. Finally, supervised and unsupervised text-scaling approaches have been employed to locate opinions and briefs in common spaces (e.g. McGuire and Vanberg, 2005; Evans et al., 2007; Dyevre, 2015). While all these approaches are promising to the extent that judicial text is taken serious, the approaches do not claim to place decisions in a policy-space but rather in some form of a "doctrine space" (Clark and Lauderdale, 2010; Lax, 2011). Hence, scaling judicial decisions based on legal elements are promising to understand a court's doctrinal path but insufficient to transfer a court's action into a policy space. Therefore, additional steps are necessary to identify the political-ideological leaning of a court in a common policy-space.

In sum, in order to understand a court's political-ideological leaning with regard

to political reform we need to design an approach which overcomes two limitations: First, individual judicial behavior can explain a court's political influence with strong assumptions only and the feasibility of respective analyses is limited. Second, features of legal text help to understand a court's influence in the legal sphere, but this does not automatically transfer to the political arena. In what follows we outline a novel approach to place judicial decisions in a (common) policy space.

3 Research design

How should we go about generating a common policy space? When studying political reforms and the legislative process in particular, we need to compare the position of various political actors as well as the court in the same space. Comparing political actors is relative straightforward because they are composed of parties for which the literature developed common policy scales (e.g., Lowe et al., 2011; König, Marbach and Osnabrügge, 2013).

The challenge we face is that there is no measure to place individual decisions of the court in a common space short of simply mapping every decision to the position of the actor that nominated the median judge (Hönnige, 2009; Carrubba et al., 2012; Sternberg et al., 2015; Brouard and Hönnige, 2017). If we are interested in mapping and comparing different decisions we need a new strategy.

The strategy we propose is to leverage the information that several political actors submit briefs to the very same case the court will decide. Such legal briefs are written statements also field by political actors, e.g. the government presents a brief to a court that argues which side of a given case should prevail. All submitted briefs are summarized within the court's written decision on the same case. Thus, from analyzing the written decisions of the court we know, first, how the court decides a case and, second, how the political actors that submitted a brief would have decided the very

same legal question. We can therefore compare the court and the authors of the briefs because they clearly state on which side they are on the same case. If we now observe consistently that political actor A “votes” more often with the court than another political actor B then A should be placed closer to the court in a common space than B. An analysis across various cases therefore allows us to estimate where the court should be placed relative to the political actors that submit briefs. Finally, we can anchor this space using existing policy scales because the locations of the political actors on these policy scales are already known.

After providing the general intuition of our strategy how to map case outcomes on a common policy scale we introduce more formally how we go about doing this. We will employ a typical scaling model that allows to scale different types of actors in a common space using appropriate bridging observations (e.g., Bailey and Chang, 2001; Bailey, 2007; Jessee, 2009, 2016). This model allows to map the location of various actors, for instance of legislators and respondents (Jessee, 2009, 2016), on the same scale leveraging information on how those actors vote (legislators’ roll-call votes) or how would they have voted (respondents’ answers to a particular survey instrument) on a given bill proposal. This is functional equivalent to our situation where we have the court deciding a case and we have the case briefs of various political actors specifying how they would decide a particular case and, particularly, the governments’ decision to enact this law and, thus, ‘voted’ in favor of upholding it.

In order to be able to speak to prior work, we like to anchor the common space using existing policy scores. Therefore, and consistent with those models, we assume an unidimensional space. It is well known that the unidimensional spatial voting model is equivalent to the two-parameter item-response model (Clinton, Jackman and Rivers, 2004; Jackman, 2004). Thus, the model we estimate using R STAN is the following standard two-parameter item-response theory model with a probit link

$$P(Y_{ij} = 1) = \Phi[\beta_j(x_i - \gamma_j)], \quad (1)$$

where β_j is case j 's discrimination parameter, indicating the strength and direction of the relationship between actor i 's ideal point x_i and her likelihood agreeing to uphold case j , i.e., $P(Y_{ij} = 1)$, and γ_j represents the location of the cutpoint for case j .

Given that we have decisions of the court on a case to uphold it (and the government's decision to enact it in the first place) as well as the respective case briefs in which various other political actors take a stand on whether to uphold the very same law or not, we glue the different spaces together by assuming that β_j and γ_j is the same in each actor's utility function. This allows us to pool all outcomes together and estimate a common space while assuming that the particular policy space underlying the preferences of each scaled actor is structured in the same way. This assumption is justified at least to the extent that the court and the political actor respond to the exact same legal question and both actors need to present their political and judicial opinion in legal language. Subsequently, political and judicial actors are constrained by political and legal concerns.

In the next section we present our data and the strategies we use to code the available information in the written decisions of the court as well as the briefs of various political actors. Furthermore, we also describe how we identify our model in equation 1 and estimate it using R STAN.

4 Scaling the German Federal Constitutional Court

In this section we outline that the German Federal Constitutional Court (GFCC) and the decisions made by the court are particularly well suited to implement our scaling-approach. Moreover, we summarize how we identify the information necessary to estimate our model. Finally, we present findings on a use-case estimating the positions

of the two senates of the GFCC during the years 1990 to 2005. This allows us to illustrate to what extent the court and the four German governments during the same period align or hold different ideological views.

4.1 Case-selection & data

The German Federal Constitutional Court (GFCC) is a suitable case to apply our scaling-approach. The court is among the stronger courts with constitutional review powers in Europe (Engst, 2018, Ch. 3) and embodies decisive features of the typical Kelsenian constitutional court (Kelsen 2008 [1931]; Shapiro and Stone 1994). Moreover, the court rarely publishes votes and votes publicly available do not necessarily discriminate. In addition, the institution does allow judges to file separate opinions, but this is not often done (for an assessment see Wittig, 2016). Approaches currently used to systematically locate the court in a policy space either rely on a few dissenting opinions (Engst et al., 2017) or use indirect measures by assigning the median judge the policy manifesto score of the political actor nominating her (Hönnige, 2007, 2009; Brouard and Hönnige, 2017). This is why, research on the GFCC and other constitutional courts can benefit from our approach to locate courts in a policy space.

In order to illustrate our scaling approach we limit our analysis in two ways: First, we assess judicial decisions made between December 2nd, 1990 – Germany’s 12th legislative period – and November 22nd, 2005 – Germany’s 15th legislative period. The focus on this period has the major advantage that four federal governments composed of two different ideological blocs governed each for a period of seven respectively eight years. The conservative christian democrats (CDU/CSU) formed a coalition with the smaller liberal party (FDP) under Chancellor Kohl from 1990 to 1994 (Kohl I) and 1994 to 1998 (Kohl II). In addition, the social democrats (SPD) formed a coalition with the smaller green party (Die Grünen) under Chancellor Schröder from 1998 to 2002 (Schroeder I) and from 2002 to 2005 (Schroeder II). Subsequently, we are able to pool

the small number of decisions from 1990 to 1998 to estimate the position of the GFCC when a conservative center-right government is in power and we can pool decisions for seven years when a social-democratic center-left government is in power.

Second, we use data on the court decisions from the Constitutional Court Database (CCDB, Hönnige et al. 2015). This comprehensive database is designed to embed the decisions of the GFCC in the German legal, political and societal environment (see Wittig 2016; Engst 2018, Ch. 4). The CCDB also summarizes the legal outcome of each decision. This is an important variable enabling us to compute our data-matrix similar to roll-call data, which we required to estimate the position of the court. In total the GFCC has published 604 decisions made by the two senates between 1990 to 2005. However, the GFCC frequently publishes decisions that are not of interest to us. Provisional orders which temporarily regulate an issue until a main decision is made, reimbursement of expenses or requests to exclude a judge from a case do not carry any political information. Subsequently we exclude those decisions and a subset of 485 main decisions remains.

Finally, we assume that political actors may express different views towards legal decisions that address different (political) topics. In other words, it is not feasible to use all main decisions in one legislative period to scale the court on one common dimension. If we would do so, we would obtain an estimate for the court which can be anchored in the policy-space due to the known position of the political actors filing briefs, but we would be unable to assign a substantial meaning to the latent policy-space. This is why we restrict our analysis to decisions in one topic area. Moreover, in focusing on one policy-area to pool decisions is justified. The CCDB includes a variable which groups all decisions according to an adopted coding scheme from the Comparative Agendas Project (CAP, Bevan 2017).¹ Subsequently we choose to restrict our analysis

¹ The decisions were initially assigned to topic areas in a double-blind process by two trained coders.

to all 100 senate decisions published between 1990 to 2005 addressing economic and social policy issues.

The central feature to implement our approach are briefs filed by scalable political actors, which serve as bridging observations. In what follows we will outline how we extract briefs from decisions and code the necessary information to obtain a vote-matrix similar to roll-call data.

4.2 Automatization to extract information to scale courts

In this section we describe our approach to automatically extract and analyze written briefs by political actors to compute a vote-matrix similar to roll-call data. Roll-calls are known as recorded votes of legislators towards a particular issue. Roll-call data commonly summarizes all votes by different legislators on an issue in long format (cf. Clinton, Jackman and Rivers, 2004; Jessee, 2016). We argue that the legal question in a decision by the GFCC is an issue that the court openly votes on by presenting a verdict and political actors vote on the same issue by filing briefs using a language similar to the court. Subsequently, the first challenge is to extract information on briefs submitted by political actors from all the decisions made by the GFCC. In order to do so we manually converted the written senate decisions of the GFCC into `.txt`-files and processed the resulting corpus using regular expressions.

Briefs include opinions or provide for background information on a case argued before the GFCC and are filed by various - not exclusively political - actors. They are summarized in the published decisions by the GFCC. To automatically extract the briefs we need to account for the fact that different plaintiffs can initiate different proceedings before the GFCC; for example constitutional complaints by individual citizens, concrete judicial review by lower courts or abstract judicial review by certain political entities. Assessing decisions according to the different proceedings revealed the various senate decisions follow different structures. Hence, automatically extracting the briefs we

account for the different structures of the decisions. Subsequently, we first extracted sections that include briefs and second we extracted only those briefs filed by scaleable political actors.

An important feature of the summarized briefs are, that they reveal an actor's position towards the legal question at hand in the same language the court uses to decide a case. The GFCC assess whether a plaintiff's claim is admissible and justified on the merits (see Schlaich and Koriath, 2015, Ch.5). Admissibility is defined by procedural matters, e.g. that a plaintiff has used all available legal measures prior to referring to the GFCC. The decision on the merits is driven by the substantial assessment of a legal question at hand. A plaintiff's referral is successful only if a decision is (partially) admissible *and* (partially) justified. Subsequently, we aggregated the admissibility and the decision on the merits following the outlined rule to compute one judicial vote for our roll-call data. Moreover, briefs clearly state how a political actor assess a legal question at hand and we code briefs based on the opinion with regard to the justification on the merits but once those are not available we focus on the admissibility.² In sum, the judicial vote on the case and the political actors' opinions raised in briefs allow to compute roll-call votes for each decision.

In order to code the automatically extracted briefs, we first randomly selected a subset of 100 briefs filed by political actors between 1973 to 2010. Afterwards, two coders with a background in public law and political science, classified the briefs in a double-blind coding according to whether a brief supports or opposes a plaintiff's referral. If a political actor regards a plaintiff's claim as justified, then the respective brief is coded as 1 and 0 otherwise. The decision on the merits beats arguments on the decision's admissibility. Hence, the coding is based on the merits when available. The coders also coded neutral briefs but we do not account for those yet. The training

² Logically we assume that if a political actor regards a case as admissible but does not argue on the merits, then this actor still supports the plaintiff's referral. This scenario does not occur with regard to the court. Instead, the judges will always share their opinion on the merits once they regard a case as admissible.

data showed an inter-coder-reliability of 96 percent. Following the training the coders classified the 112 briefs filed by political actors in 80 of the 100 decisions analyzed here. In 20 decisions no political actor filed a brief. In order to increase the information of our currently scarce voting-matrix, we add information on the political actor who presented a law under consideration in a respective decision (e.g., the federal governments). The assumption is that this actor rejects a referral to the court but instead the courts should uphold the proposed law. This procedure increases the votes by political actors available to scale the decision to 167 votes in 85 decisions.

In sum, we compute a roll-call dataset which summarizes the votes by the court made in a decision and the votes by scalable political actors who filed briefs as well as the actors who initially presented a law under consideration. Subsequently, we can leverage a Bayesian ideal point estimation with strong prior information using the spatial positions of the political actors to estimate the ideal points of the two senates of the GFCC in a unidimensional policy-space. In the next section we present the results of our estimation

4.3 The GFCC and political parties from 1990 to 2005

In this section we present our strategy to estimate the model shown in equation 1 using court decisions including briefs filed by political actors during the legislative periods 12 and 13 (hereafter: Kohl I & Kohl II) as well as during the 14th and 15th seating of the first chamber, the German *Bundestag* (hereafter: Schroeder I & Schroeder II). The estimation also makes use of briefs filed by political actors during the 16th *Bundestag*, but due to the small sample size, we do not provide estimates on the senates from 2005 to 2009. Moreover, we expand our vote-matrix by adding information on the government that presented a law considered in a court decision. The government that initiated a given law is assumed to support to uphold the law.

4.3.1 Modeling approach

We use R STAN (see <http://mc-stan.org/>) to estimate our model. In order to define the priors of the political actors we make use of the MCSS posteriors estimated by König, Marbach and Osnabrügge (2013) for the German political parties based on the parties manifestos. The priors for the political actors who file briefs, such as the government Kohl I & II as well as Schroeder I & II are generated via bootstrapping from the posterior MCSS policy scores of each governing party weighted by the parties seat shares in the first legislative chamber (Bundestag). Each posterior is an estimate of the parties position in a unidimensional latent policy space. This approach is also adopted to estimate the position of the second legislative chamber (Bundesrat) or state governments who file briefs. The resulting distribution closely approximates a normal distribution of means, thus we use the mean and variance of the empirical distribution resulting from the bootstrap and approximate a normal distribution with these values to obtain our prior information used in the model. In short, we assume a prior distribution of x 's corresponding to the respective mean and standard deviation of the bootstrapped normal distribution.

Eventually we simulate the posterior distribution of a senate's ideal points using Markov Chain Monte Carlo. Our estimates are based on 40,000 iterations, of which the first 20,000 iterations are omitted as warm-up. While we have strong prior information on the position of the political actors, our assumption about a senate's ideal point is only weakly informative. Subsequently, we argue that a senate is a non-extremist actor and, thus, the ideal points of the senate are within the range of all ideal points of the political parties elected to the first chamber. In mathematical terms, the priors of each senate are drawn from a standard normal distribution (with mean 0 and standard deviation 1), which happens to represent centrist values on the MCSS scale. Extending the standard deviation to 1.5 does not impact the substantive interpretation of the estimated ideal points, but increases estimation uncertainty. In order to assess the

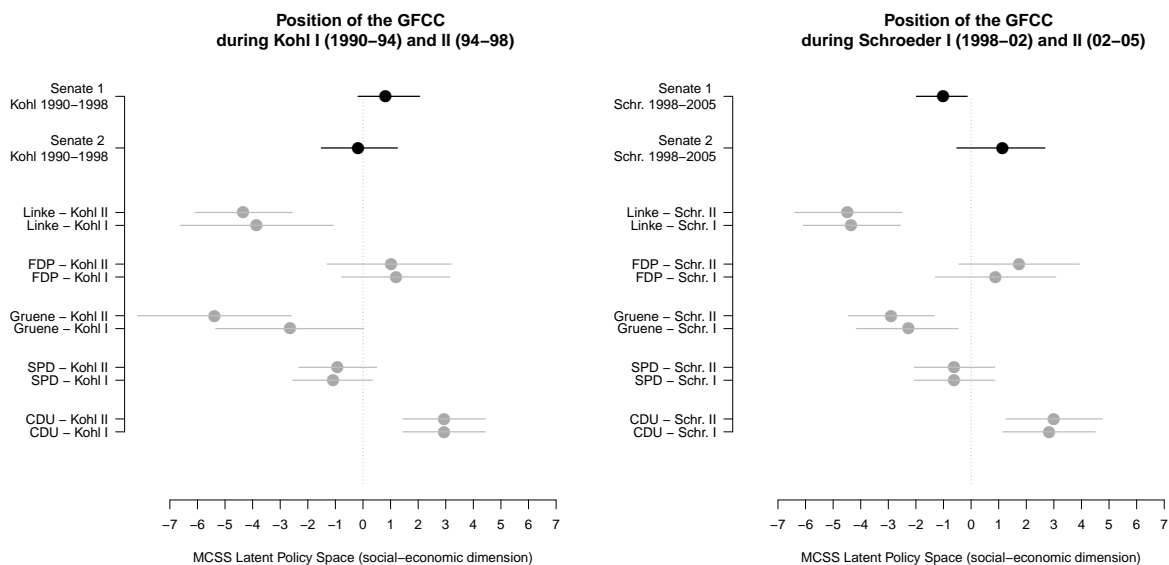
robustness of our approach we use the `logrile` posteriors by Lowe et al. (2011). In that case we constrain the standard deviation of the normal distribution to 0.5 to represent the narrower `logrile` scale.

The discrimination parameter β_j from our model can not be interpreted as the difficulty of the j^{th} item following classic item response theory. Instead, the factor moderates the effect of positional difference in the unidimensional space on the probability of a supportive sentiment. There may exist issues clearly dominated by ideological concerns - for example redistribution and social welfare - whereas the assessment of very specific claims may be influenced only weakly by an actor's position on the latent dimension. However, we have no prior information on the discrimination parameter. Thus we draw those from a standard normal distribution. Finally, the uninformative prior is assigned to the cutpoint parameter γ_j . The starting values correspond to the means of the posterior distributions of each political actor, and 0 is assigned for all senates and additional parameters.

In order to check the robustness of convergence, we simulated 1,000 chains for each senate. If there is any post-warm up divergences, the estimates would not be valid. To tune our model we have varied the resolution of the Gibbs sampler, that is, the step size between every iteration until no post-warm up divergences occurred in all 1,000 chains. Finally, we arrived at $\delta_{Senate1} = 0.85$ and $\delta_{Senate2} = 0.9$ as the corresponding maximum step size between every iteration for the respective models of the two senates. In addition, the estimated ideal points for both senates are substantially similar across the MCSS and `logrile` posteriors, albeit their corresponding estimation uncertainty differs. We randomly draw one chain to present our substantive findings below.³

³ The chain plots on that chain are summarized in Appendix 1

Figure 1: Estimated position of the two senates of the GFCC, 1990-2005



4.3.2 Estimation results

In order to scale the court and political actors on one common latent policy space we use our vote-matrix and the prior manifesto information on the various political actors. In essence, we identify the position of the court on a unidimensional latent policy space with a Bayesian ideal point estimation model. Afterwards, we compare the estimated position of the court to the position of the political parties in the German *Bundestag*. The analysis shows that there exist spatial differences between the two senates of the GFCC and differences between each senate and the political parties. The findings show that the second senate changes the own position as the ideological composition of the incoming government changes. This suggests that the second senate counter-balances governmental action albeit being composed of judges nominated by governing parties (cf. Hönnige 2009, 170).

Figure 1 summarizes our estimated ideal points (and the 95% BCI) for the first and second senate (black estimates) during the combined cabinets Kohl I & II (left panel) as well as Schroeder I & II (right panel). The average position of the first senate on the

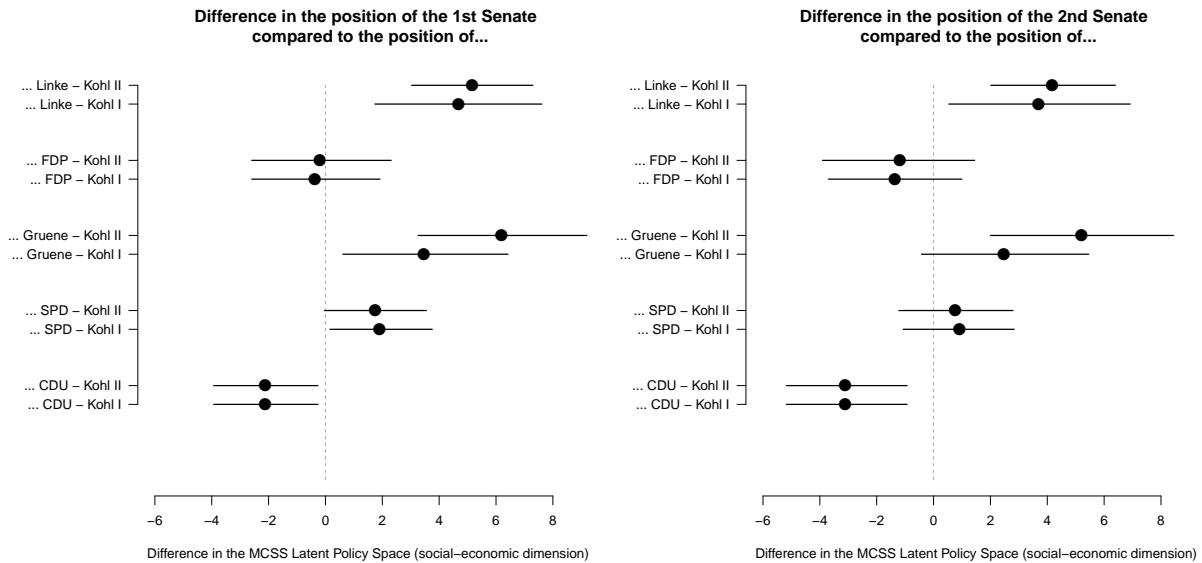
MCSS scale is 0.82 during Kohl I & II and the average position of the second senate is -0.17. During Schroeder I & II the first senate moves to the left on average with about -1.02, the second senate moves to the right with on average 1.14.

Surprisingly, the shifts in the ideal points of the 1st senate correspond to the change in the composition of the government from Kohl to Schroeder. In other words, as the governing parties shift to the left – as a consequence of the change from a center-right to a center-left government – the first senate changes to a center-left position as well (see Figure 1). Nevertheless, a similar change cannot be observed for the second senate. The senate initially holds a centrist position accounting for decisions on a social-economic dimension during Kohl I & II. However, the average ideal point during Schroeder I & II moves to the right on the MCSS scale.

We are only able to outline *possible* mechanism behind the shift in the court's ideal point due to changes in the government's composition. First, the government could file briefs in line with the rulings of the first senate. The senate mainly makes decisions on basic, fundamental rights and civil liberties. The possible political opposition towards those decisions – and especially towards the popular GFCC – could come with costs to the federal government. The public could regard the government as not supporting fundamental rights which could entail electoral costs. Subsequently, the government may seek to avoid to be portrait as opposing legal decisions on basic rights. Instead, the government is likely to build an image of adhering to those rights filing respective briefs. However, this is dependent on the government's ability to predict judicial decision. Governments file their briefs prior to the court voting on a plaintiff's claim. Hence, this argument violates the assumption of temporal causality. Therefore, another explanation – accounting for the temporal dimension – could follow along the lines that the court aligns with the briefs fielded by the federal government.

Figure 2 summarized the first differences between the first senate and the positions of the political parties during Kohl I & II (left panel) and the second senate (right

Figure 2: First difference comparing positions of the parties to the position of the GFCC, 1990-1998

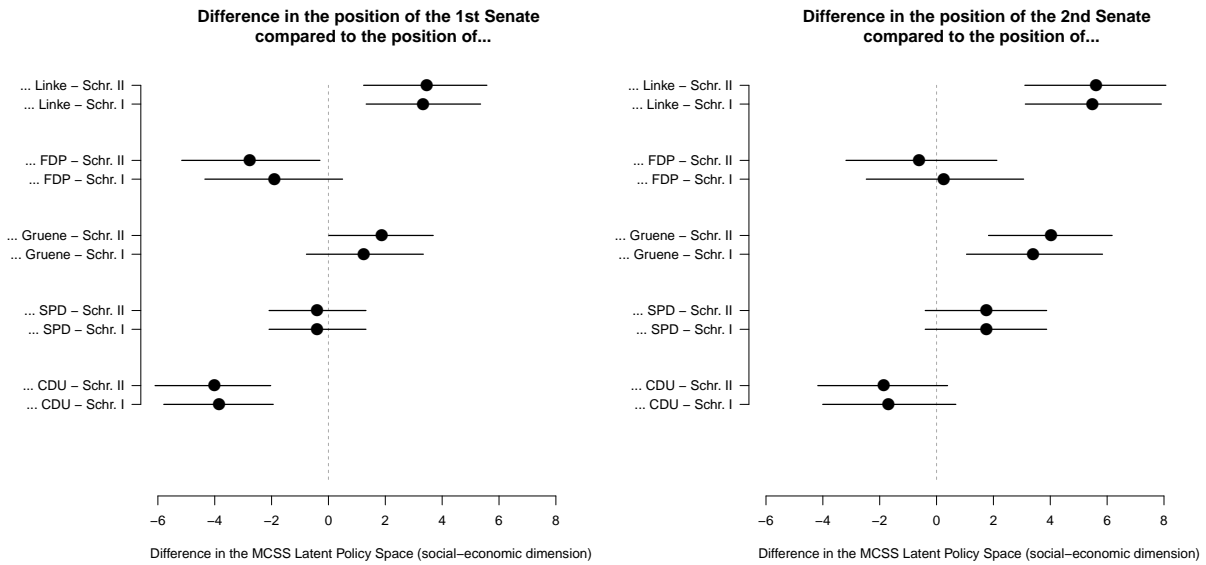


panel). The first senate is located significantly more to the left of the CDU as the major conservative governing party. The first senate holds a more social-liberal position on the social-economic dimension with a first different not significant different to the liberal party FDP and during Kohl II also not significant different to the SPD.

Moreover, the senates' position move to the left during the area of Schroeder I & II as shown in Figure 1. This can also be seen in the left panel of Figure 3. The first differences of the governing parties SPD and the Green party (Gruene) is not significant different compared to the position of the first senate. Moreover, the senate is positioned significantly more to the left compared to the position of the conservative CDU. Finally, the Green party moves further to the left following the election of 2002 and its position is significantly different compared to the position of the first senate. Subsequently, the first senate aligns with the social-democratic views while not adopting more extreme views as expressed by the party die Linke.

Comparing the right panels in Figure 2 and 3 shows that the second senate becomes more conservative over time. During the 12th *Bundestag* of the Kohl I government the difference between the social democrats (SPD), the green party, and the liberal FDP

Figure 3: First difference comparing positions of the parties to the position of the GFCC, 1998-2005



is not significant. However, this changes from Kohl II to Schroeder I & II. The Green party and the court drift apart while the average position of the FDP and the second senate move closer to one another. Finally, the significant difference between the CDU and the second senate diminishes during the area of Schroeder I & II.

How may we explain this movement? The second senate is in charge of decisions directed at the organization of the state. Given this jurisdiction it is plausible to regard the senate as a moderator to the government’s behavior compared to the first senate who protects fundamental rights. In other words, it is possible that the second senate is more often in a position where the court has to strongly oppose governmental claims.

In sum, our approach to estimate the ideal points of the first and second senate of the GFCC has shown that on a social-economic dimension the first senate holds more social-liberal views. The second senate appears to counter-balance governmental action which is why the senate becomes more conservative as the government changes from Kohl I & II to Schroeder I & II.

In order to assess the *robustness* of our scaling approach we computed our initial scores shown in Figure 1 also using the logrile scores estimated by Lowe et al. (2011).

The results in Appendix 2 show that the plotted ideal points of the first and second senate are substantially similar to the once in Figure 1. This speaks in favor of the robustness of our approach.

4.4 Validity test of our common judicial policy scores

In this section we evaluate the validity of our strategy to estimate judicial policy scores compared to the approach which assigns a court the position of the political party that nominates the median judge (Hönnige, 2009; Carrubba et al., 2012; Sternberg et al., 2015; Brouard and Hönnige, 2017). We compare the explanatory power of both approaches assessing a common theoretical believe in judicial politics. Namely, the hypothesis that *if the government is opposed to the court than the court will increase transparency surrounding a decision through oral hearings* (Vanberg 2001, Vanberg 2005, Krehbiel 2016, implicit also Staton 2010).

It is plausible to assume that the government increasingly opposes the court once the spatial distance between the court and the government increases. Hence, in line with the above hypothesis we expect a positive effect of a measure summarizing the absolute spatial distance between the court and the government on the occurrence of an oral hearing. In order to operationalize the spatial distance between the GFCC (remember that the court does not commonly publish individual judicial votes) and the German federal government we apply two strategies.

First, we use our *vote-matrix approach* and assign the court the position we estimated using the standard two-parameter item-response theory model outlined above. Moreover, we bootstrap the government's position drawing from the posterior distribution of the Manifesto Common Space Scores (MCSS; König, Marbach and Osnabrügge 2013) of each governing party as often as the respective party has seats in Germany's first chamber, the *Bundestag*. Finally, the position of the government is weighted by the sum of the legislative seats of the parties in government. In order to calculate the

difference between the government and the court we subtract the mean position of the government from the mean position of the court which we estimated above. The absolute value of the calculated difference is the distance between the government and the court.

Second, we use the previously established *party-label approach* and assign each judge the manifesto score of the party nominating her (Hönnige, 2007, 2009; Brouard and Hönnige, 2017). We once again use the MCS Scores for all the estimations and identify the position of the court as the manifesto score of the median judge. Next, we estimate the policy-score of the government weighting the manifesto scores of the parties in the governing coalition by the number of the respective representatives in the legislature. Finally, we subtract the position of the government from the position of the court and compute the absolute value. This value defines the distance between the court and the government.

In Table 1 we regress the two independent variables on an indicator for decisions with oral hearing (1) and without those hearings (0). In total we count 24 oral hearings compared to 85 decisions without oral hearings in our small sample. However, we need to omit four decisions as in those decisions the court has no power to schedule oral hearings. Subsequently, we have 22 decisions with oral hearings and 83 decisions without oral hearing. In the rare event logistic regressions (see King and Zeng, 2001) in the bivariate model 1.1 and the complete model 1.2 we use our measure of the absolute distance between the court and the government based on the vote-matrix approach as predictor. The estimates have the expected positive effect on decisions with oral hearings. In other words, the probability of an oral hearing significantly increases as the distances between the court and the government increases. This is well in line with expectations from the literature (Vanberg, 2001, 2005; Krehbiel, 2016). On the contrary, we find no significant effect of the distance between the court and the government on decisions with oral hearings in models 2.1 and 2.2. In these latter models we use the

Table 1: Rare event logit to predict oral hearings with spatial measures

	DVar: Oral hearing in judicial decision (=1)			
	Vote-matrix approach		Party-label approach	
	Model 1.1	Model 1.2	Model 2.1	Model 2.2
Abs. dist. btw. court & gov.	0.54* (0.27)	0.69* (0.31)	-0.36 (0.24)	-0.01 (0.28)

Length of the ruling (number of paragraphs)		0.03* (0.01)		0.02* (0.01)
1st Senate of the GFCC		0.11 (0.58)		0.03 (0.58)
(Intercept)	-1.67* (0.62)	-3.88* (1.08)	-0.02 (0.43)	-2.31* (0.91)
ePCP	0.819	0.848	0.790	0.819
BIC	112.39	99.51	114.49	106.26
Log Likelihood	-51.54	-40.45	-52.59	-43.82
Num. obs.	105	105	105	105
Observed hearings	22	22	22	22
Correctly predicted hearings	8	12	0	8

* $p < 0.05$

party-label approach to measure the distance between the court and the government. Moreover, the insignificant effects are on average negative in models 2.1 and 2.2. This implies that - if any - the likelihood of an oral hearing decreases as the distance between the court and the government increase. The insignificant finding contradicts common expectations from the literature. Subsequently, our measure produces valid (significant) findings well in line with expectations in judicial politics.

Nevertheless, one may claim that the theoretical expectations in the literature are false. However, Table 1 highlights that the models 1.1 and 1.2 provide for a better fit

compared with the models 2.1 and 2.2. The expected proportion correctly predicted (ePCP) highlight that the bivariate model 1.1 has already a fit similar to the complete model 2.2,⁴ while the complete model 1.2 provides for the best model fit over all models, with about 85 percent correctly predicted. In addition, the BIC confirms that the two models using the vote-matrix approach are slightly better than the two models using the party-label approach. Moreover, it is important to note that the more detailed comparison shown in the lower part of Table 1 reveals that the models 1.1 and 1.2 successfully correctly predict decisions with oral hearings (=1) while model 2.1 predicts none.

In sum, our validity test shows that our estimation strategy produces scores that seem to yield results the prior literature would expect us to find while the currently most often used alternative approach based on party-labels does not. We take this as strong evidence that our new vote-matrix approach is well suited to assess common theoretical expectations in judicial politics and provides for a better model-fit testing those expectations, when compared with the party-label approach. Subsequently, our novel approach to estimate common judicial policy scores is an improvement over established approaches.

5 Conclusion

What are the positions of highest courts towards political reform? In order to answer this question it is necessary to develop common spatial measures that allow us to locate highest courts on the same dimension as political actors who design reform processes. One common approach to scale judicial actors - mainly applied in the context of the US Supreme Court - is to analyze published, individual votes (e.g. Martin and Quinn, 2002; Epstein, Martin, Quinn and Segal, 2007; Epstein, Martin, Segal and Westerland,

⁴ The ePCP is similar in both models but a visual inspection shows that the two models do not predict the same decisions.

2007; Hanretty, 2012*b,a*, 2014, 2015; Engst et al., 2017). However, this approach does not travel to systems where courts rarely or never publish individual votes or where published vote do not discriminate while judges follow a norm of consensus. Moreover, the assessment of judicial votes becomes challenging where scholars need to fear imprisonment.⁵

In order to address these challenges we leverage information commonly embedded in the text of the decisions themselves, which allows us to anchor judicial decision in a common policy-space, together with positions of political actors. In particular, we assume that the overall outcome of a decision and briefs filed by political actors are suitable to compute a vote-matrix similar to roll-call data in legislative studies. Moreover, the knowledge on the manifesto position of the political actors (e.g., Lowe et al., 2011; König, Marbach and Osnabrügge, 2013) filing a brief can be used to bridge judicial decisions into a common policy-space. In order to estimate policy-scores we employ a standard two-parameter item-response theory model using R STAN.

What is the advantage of our approach over traditional approaches, for example assigning manifesto scores assuming that the scores are ‘inherited’ from the nominating party and, then, applying the scores to the court or the median judge? Intuitively, our approach is designed to leverage more information from the written decisions itself. In contrast to traditional work in judicial politics that do not use the content of those decisions at all, our approach is embedded in a recent trend in the literature that takes the content of judicial output seriously (e.g. Fowler et al., 2007; Clark and Lauderdale, 2010, 2012; Lauderdale and Clark, 2014). This helps to engage in broader and interdisciplinary scholarship to study the interaction of courts with other political actors in a particular system of governance. Our approach therefore has the potential to also bridge the gap between law scholars and political scientists because both groups

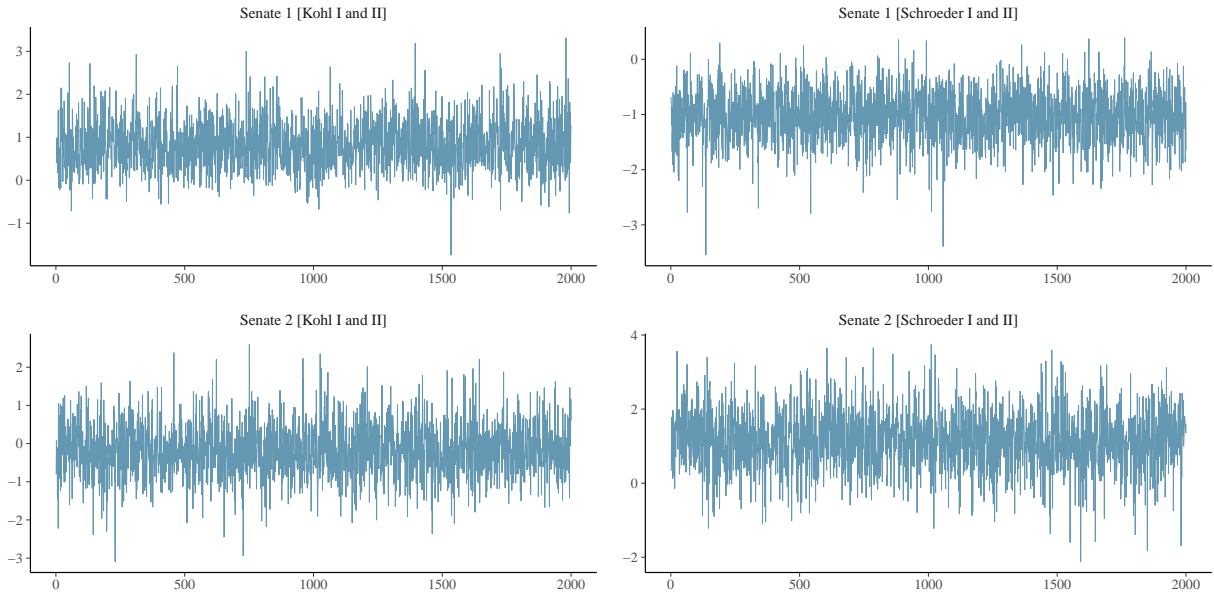
⁵ Artificiallawyer. 2019. “France Bans Judge Analytics, 5 Years In Prison For Rule Breakers.” <https://www.artificiallawyer.com/2019/06/04/france-bans-judge-analytics-5-years-in-prison-for-rule-breakers/>.

are interested in the same feature: analyzing the written text of judicial decisions. Moreover, assessing a common theoretical believe in judicial politics we showed that our approach is well equipped to generate valid and more plausible findings compared with the commonly applied party-label approach.

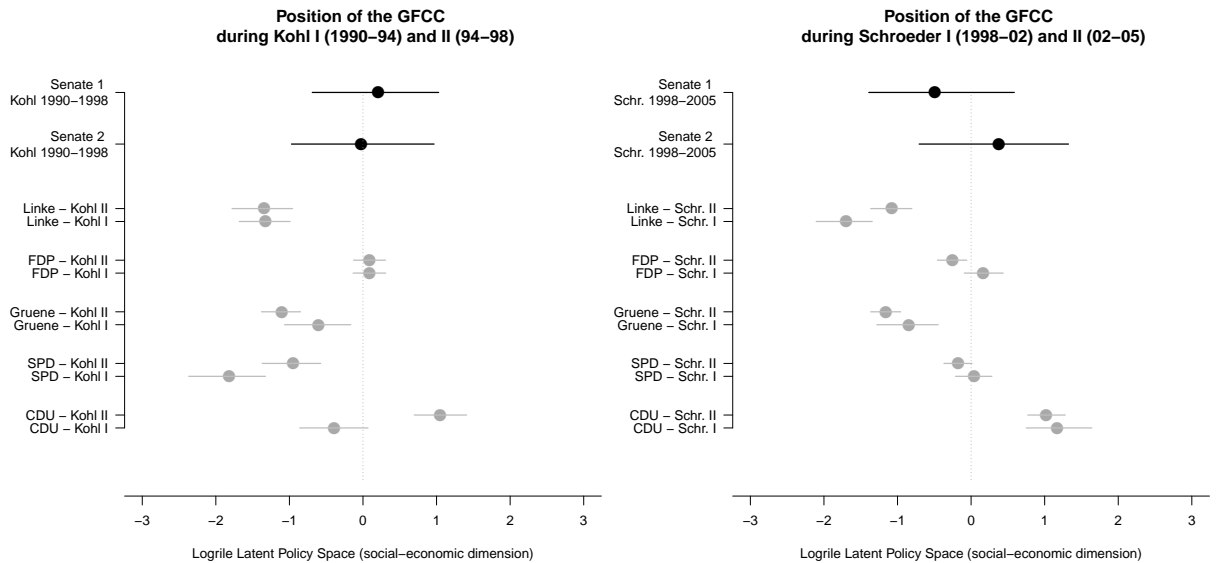
In moving the project forward we seek to collect more data inherent to decisions to enhance our vote-matrix and also weight decisive features of decisions. Moreover, the key when developing new measures such as judicial-policy scores is to validate the estimates. We have done so in our plausibility assessment analyzing the occurrence of oral hearing in judicial decisions but an additional attempt we undertake to validate our approach is to (crowd-source and) hand-code decisions and briefs by a group of experts with legal training.

The legal language has very particular characteristics and follows a special style – especially sentiments expressed in decisions are very different to sentiments used in common language. Finally, once it comes to legal arguments the text and the used language carries important weight. With our approach we take written legal decisions serious and encourage the community to keep following this trend in judicial politics.

Appendix 1: Trace plot of ideal point estimates of the GFCC



Appendix 2: Estimated position of the two senates of the GFCC on logrile scores, 1990-2005



References

- Bailey, Michael A. 2007. "Comparable Preference Estimates across Time and Institutions for the Court, Congress, and Presidency." *American Journal of Political Science* 51(3):433–448.
- Bailey, Michael A. and Kelly H. Chang. 2001. "Comparing Presidents, Senators, and Justices: Interinstitutional Preference Estimation." *The Journal of Law, Economics & Organization* 17(2):477–506.
- Bevan, Shaun. 2017. "Gone Fishing: The Creation of the Comparative Agendas Project Master Codebook." *Comparative Agendas Project (CAP) 0.9.1(Beta)*:08/26/2017.
- Brouard, Sylvain and Christoph Hönnige. 2017. "Constitutional courts as veto players: Lessons from the United States, France and Germany." *European Journal of Political Research* 56(3):529–552.
- Carrubba, Cliff, Barry Friedman, Andrew D. Martin and Georg Vanberg. 2012. "Who Controls the Content of Supreme Court Opinions?" *American Journal of Political Science* 56(2):400–412.
- Clark, Tom S. and Benjamin E. Lauderdale. 2010. "Locating Supreme Court Opinions in Doctrine Space." *American Journal of Political Science* 54(4):871–890.
- Clark, Tom S. and Benjamin E. Lauderdale. 2012. "The Genealogy of Law." *Political Analysis* 20(3):329–350.
- Clinton, Joshua D., Simon Jackman and Douglas Rivers. 2004. "The Statistical Analysis of Roll Call Data." *American Political Science Review* 98(2):355–370.
- Dyevre, Arthur. 2015. "The Promise and Pitfalls of Automated Text-Scaling Techniques for the Analysis of Judicial Opinions." *Working Paper* .
- Engst, Benjamin G. 2018. *The Two Faces of Judicial Power: The Dynamics of Judicial-Political Bargaining* PhD thesis University of Mannheim.
- Engst, Benjamin G., Thomas Gschwend, Nils Schaks, Sebastian Sternberg and Caroline Wittig. 2017. "Zum Einfluss der Parteinähe auf das Abstimmungsverhalten der Bundesverfassungsrichter – eine quantitative Untersuchung." *JuristenZeitung* 72(17):816–826.

- Epstein, Lee, Andrew D. Martin, Jeffrey A. Segal and Chad Westerland. 2007. "The Judicial Common Space." *Journal of Law, Economics, & Organization* 23(2):303–325.
- Epstein, Lee, Andrew D. Martin, Kevin M. Quinn and Jeffrey A. Segal. 2007. "Ideological Drift among Supreme Court Justices: Who, When, and How Important?" *Northwestern University Law Review* 101(4):1483–1541.
- Epstein, Lee, Jack Knight and Olga Shvetsova. 2001. "The Role of Constitutional Courts in the Establishment and Maintenance of Democratic Systems of Government." *Law & Society Review* 35(1):117–164.
- Epstein, Lee, Jeffrey A. Segal and Harold J. Spaeth. 2001. "The Norm of Consensus on the U.S. Supreme Court." *American Journal of Political Science* 45(2):362–377.
- Evans, Michael, Wayne McIntosh, Jimmy Lin and Cynthia Cates. 2007. "Recounting the courts? Applying automated content analysis to enhance empirical legal research." *Journal of Empirical Legal Studies* 4(4):1007–1039.
- Fowler, James H., Timothy R. Johnson, James F. Spriggs, Sangick Jeon and Paul J. Wahlbeck. 2007. "Network analysis and the law: Measuring the legal importance of precedents at the U.S. Supreme Court." *Political Analysis* 15(3):324–346.
- Hall, Mark A and Ronald F Wright. 2008. "Systematic Content Analysis of Judicial Opinions." *California Law Review* 96(1):63–122.
- Hanretty, Chris. 2012a. "Dissent in Iberia: The ideal points of justices on the Spanish and Portuguese Constitutional Tribunals." *European Journal of Political Research* 51(5):671–692.
- Hanretty, Chris. 2012b. "The Decisions and Ideal Points of British Law Lords." *British Journal of Political Science* 43(3):1–3.
- Hanretty, Chris. 2014. "The Bulgarian Constitutional Court as an Additional Legislative Chamber." *East European Politics & Societies* 28(3):540–558.
- Hanretty, Chris. 2015. "Judicial Disagreement need not be Political: Dissent on the Estonian Supreme Court." *Europe-Asia Studies* 67(6):970–988.
- Hönnige, Christoph. 2007. *Verfassungsgericht, Regierung und Opposition. Die vergleichende Analyse eines Spannungsdreiecks*. Wiesbaden: VS Verlag für Sozialwissenschaft.

- Hönnige, Christoph. 2009. "The Electoral Connection: How the Pivotal Judge Affects Oppositional Success at European Constitutional Courts." *West European Politics* 32(5):963–984.
- Hönnige, Christoph, Thomas Gschwend, Caroline Wittig and Benjamin G. Engst. 2015. *Constitutional Court Database (CCDB), V17.01 [Mar.]*.
URL: <http://ccdb.eu/>
- Jackman, Simon. 2004. "Bayesian Analysis for Political Research." *Annual Review of Political Science* 7:483–505.
- Jessee, Stephen. 2016. "(How) Can We Estimate the Ideology of Citizens and Political Elites on the Same Scale?" *American Journal of Political Science* 60(4):1108–1124.
- Jessee, Stephen A. 2009. "Spatial Voting in the 2004 Presidential Election." *American Political Science Review* 103(01):59–81.
URL: http://www.journals.cambridge.org/abstract_S000305540909008X
- Kelemen, Katalin. 2013. "Dissenting Opinions in Constitutional Courts." *German Law Journal* 14(8):1345–1371.
- Kelsen, Hans. 2008. [1931] *Wer soll der Hüter der Verfassung sein? Abhandlungen zur Theorie der Verfassungsgerichtsbarkeit in der pluralistischen, parlamentarischen Demokratie*. Tübingen: Mohr Siebeck.
- King, Gary and Langche Zeng. 2001. "Logistic Regression in Rare Events Data." *Political Analysis* 9(2):137–163.
URL: <http://pan.oxfordjournals.org/content/9/2/137.abstract>
- König, T., M. Marbach and M. Osnabrügge. 2013. "Estimating Party Positions across Countries and Time—A Dynamic Latent Variable Model for Manifesto Data." *Political Analysis* 21(4):468–491.
- Krehbiel, Jay N. 2016. "The Politics of Judicial Procedures: The Role of Public Oral Hearings in the German Constitutional Court." *American Journal of Political Science* 60(4):990–1005.
- Lauderdale, Benjamin E. and Tom S. Clark. 2014. "Scaling Politically Meaningful Dimensions Using Texts and Votes." *American Journal of Political Science* 58(3):754–771.
- Lax, Jeffrey R. 2011. "The New Judicial Politics of Legal Doctrine." *Annual Review of Political Science* 14(1):131–157.

- Lowe, Will, Kenneth Benoit, Slava Mikhaylov and Michael Laver. 2011. "Scaling Policy Preferences from Coded Political Texts." *Legislative Studies Quarterly* 36(1):123–155.
- Lupu, Yonatan and James H. Fowler. 2013. "Strategic Citations to Precedent on the U.S. Supreme Court." *Journal of Legal Studies* 42(1):151–186.
- Martin, Andrew D. and Kevin M. Quinn. 2002. "Dynamic Ideal Point Estimation via Markov Chain Monte Carlo for the U.S. Supreme Court, 1953-1999." *Political Analysis* 10(2):134–153.
- McGuire, Kevin T. and Georg Vanberg. 2005. "Mapping the Policies of the U.S. Supreme Court: Data, Opinions, and Constitutional Law." *Working Paper* .
- Parcelle Jr., Richard L., Brett W. Curry and Bryan W. Marshall. 2011. *Decision Making by the Modern Supreme Court*. Cambridge: Cambridge University Press.
- Raffaelli, Rosa. 2012. "Dissenting opinions in the Supreme Courts of the Member States." *European Parliament: Directorate General for Internal Policies* (PE 462.470).
- Schlaich, Klaus and Stefan Koriath. 2015. *Das Bundesverfassungsgericht. Stellung, Verfahren Entscheidungen*. 10th ed. München: C.H. Beck.
- Segal, Jeffrey A. and Albert D. Cover. 1989. "Ideological Values and the Votes of U.S. Supreme Court Justices." *American Political Science Review* 83(2):557–565.
- Shapiro, Martin and Alec Stone. 1994. "The New Constitutional Politics of Europe." *Comparative Political Studies* 26(4):397–420.
- Staton, Jeffrey K. 2010. *Judicial Power and Strategic Communication in Mexico*. Cambridge: Cambridge University Press.
- Sternberg, Sebastian, Thomas Gschwend, Caroline E. Wittig and Benjamin G. Engst. 2015. "Zum Einfluss der öffentlichen Meinung auf Entscheidungen des Bundesverfassungsgerichts: Eine Analyse von abstrakten Normenkontrollen sowie Bund-Länder-Streitigkeiten 1974 - 2010." *Politische Vierteljahresschrift* 56(4):570–598.
- Vanberg, Georg. 2001. "Legislative-Judicial Relations: A Game-Theoretic Approach to Constitutional Review." *American Journal of Political Science* 45(2):346–361.
- Vanberg, Georg. 2005. *The Politics of Constitutional Review in Germany*. Cambridge: Cambridge University Press.

Vanberg, Georg. 2015. "Constitutional Courts in Comparative Perspective: A Theoretical Assessment." *Annual Review of Political Science* 18(1):167–185.

Wittig, Caroline E. 2016. *The Occurrence of Separate Opinions at the Federal Constitutional Court. An Analysis with a Novel Database*. Berlin: Logos Verlag Berlin GmbH.