

Measuring Foreign Policy Positions of Members of the US Congress*

GYUNG-HO JEONG

Studies have shown that a foreign policy position of a member of Congress is often distinct from a domestic one. Despite this, measures commonly used to determine the foreign policy positions of members of Congress are based on congressional votes on domestic as well as foreign policy matters. As foreign policy votes take up only a small portion of all congressional votes, these measures conflate a member's foreign policy position with his or her domestic policy position. While there are other measures based exclusively on foreign policy votes, these are also problematic because they tend to use a small number of controversial votes and thus inflate extremism. To address these shortcomings, I present a new measure by applying a Bayesian item response theory model to all foreign policy votes. This paper demonstrates the similarities, differences, and advantages of this measure by comparing it with the existing measures in a series of analyses of foreign policy positions of political parties and individual legislators.

In explaining how members of the US Congress make foreign policy decisions, a large number of studies have found that ideology¹ of individual members is one of the most important factors. Individual members' ideology has been found to be as important as, or more important than, constituent interests and political party affiliation in affecting members' decisions on foreign policy matters. For instance, Bernstein and Anthony (1974) found that a senator's ideology was more important than the senator's party affiliation and the economic interests of the senator's state in explaining the senator's vote on an anti-ballistic missile system. More recently, Meernik and Oldmixon (2008) found that a senator's support for the President on foreign policy is significantly affected by the similarity of ideology between the senator and the President.

In addition, a number of studies have documented that a member's position on foreign policy matters is often distinct from their domestic policy position. As summarized by Senator Arthur Vandenberg's (R-MI) statement "politics stops at the water's edge," the politics of foreign policy has often been viewed differently from that of domestic policy. While this statement sounds less and less relevant, as congressional foreign policymaking becomes increasingly influenced by party politics, the literature on the two presidencies thesis has shown that, until

* Gyung-Ho Jeong, Assistant Professor at the University of British Columbia, C425 - 1866 Main Mall, Vancouver, B.C. V6T 1Z1, Canada (gyung-ho.jeong@ubc.ca). An earlier version of this paper was presented at the 2013 Annual Meeting of the Midwest Political Science Association. The author would like to thank Ralph Carter, James McCormick, and three anonymous reviewers for their comments. This research was supported by the Social Sciences and Humanities Research Council of Canada. To view supplementary material for this article, please visit <http://dx.doi.org/10.1017/psrm.2016.3>

¹ In this paper, I use "ideology" and "position" interchangeably, despite their conceptual differences. As discussed extensively by existing studies (see Kingdon 1981), congressional voting is a product of multiple factors, such as a member's constituency, party influence, and the President, as well as the member's ideology. Therefore, any measures of ideology based on congressional votes are subject to this criticism of measurement error. Here, I do not claim that my measure is immune from this criticism. Instead, my goal is to improve on the existing measures of foreign policy ideology.

recently, Congress has tended to defer to presidential leadership on foreign policy matters (Canes-Wrone, Howell and Lewis 2007).

The history of congressional foreign policymaking also shows that foreign policy ideology is distinct from that of domestic policy. While domestic policy conservatives supported isolationist foreign policy until the 1960s, they now support interventionist foreign policy (except for some domestic policy conservatives, such as Rand Paul). Conversely, domestic policy liberals supported interventionist foreign policy until the 1960s. They are now less supportive of intervention in general, except for certain humanitarian interventions. These changes in what it means to be conservative or liberal in foreign policy are also documented on specific foreign policy issues. For instance, domestic policy conservatives supported protectionist trade policy and opposed increased defense spending before the 1970s, but have since adopted the opposite positions (Keech and Pak 1995; Trubowitz 1998; Cronin and Fordham 1999; Irwin and Kroszner 1999; Shoch 2001; Fordham 2007; Karol 2009).

Therefore, for a better understanding of congressional foreign policymaking, we need a distinct measure of foreign policy positions based on members' foreign policy voting behavior. While NOMINATE measures and interest group ratings, such as Americans for Democratic Action (ADA) scores, are commonly used to measure a member's foreign policy position, they are based on members' votes on domestic as well as foreign policy issues. As foreign policy votes tend to take up only a small portion of all congressional votes, these measures conflate a member's foreign policy position with his or her domestic policy position. Therefore, we need to build a measure of foreign policy position using foreign policy votes. For this reason, *National Journal* vote ratings are useful because they are based purely on foreign policy votes. However, they tend to utilize only a small number of selective foreign policy votes. As will be demonstrated later, these measures tend to exaggerate ideological differences among legislators and are less informative.

To address these problems, I present a measure of foreign policy positions of members of Congress by applying a Bayesian item response theory (IRT) model to all foreign policy votes. By utilizing only foreign policy votes, this measure avoids conflating a member's foreign policy position and his or her domestic policy position and thus captures important changes in members' foreign policy positions. By utilizing all foreign policy votes, this approach also produces a finer and more informative measure than the measures based on a small number of selective foreign policy votes. Another key advantage of my approach is that it can be implemented through free software to estimate foreign policy positions that are comparable over time.

In the next section, I discuss the distinctiveness of foreign policy ideology from that of domestic policy, and the problems of the existing measures. I then present my approach to address these problems. To illustrate the usefulness of the new measure, I first compare it with existing measures, using the foreign policy positions of members in the 103rd House of Representatives. The advantages of this new measure are then demonstrated with a series of analyses of the foreign policy positions of political parties and key senators since World War II.

EXISTING MEASURES OF FOREIGN POLICY POSITIONS OF MEMBERS

Commonly used measures of ideology utilize members' votes on the floor of Congress. For instance, NOMINATE measures developed by Poole and Rosenthal (1997) utilize all congressional votes to estimate ideology of members, while interest/advocacy groups—like ADA—select a small number of votes to measure members' political liberalism or conservativeness.

While these measures have served researchers by allowing them to have a proxy measure of ideology, there are some problems in using them to analyze foreign policy decisions of legislators. First, NOMINATE measures and ADA scores are based on members' votes on

domestic as well as foreign policy issues.² The problem is that the number of foreign policy votes is quite small compared with that of domestic policy votes. On average, the percentage of foreign policy votes in each Congress is only 18.8 percent in the House and 21.3 percent in the Senate. We can thus safely assume—and demonstrate later—that these measures are capturing mainly domestic policy positions of members.

This predominance of domestic policy votes will not be a problem if there is no difference between domestic and foreign policy ideologies. However, this assumption does not square with the findings of existing studies that foreign policy ideology is distinct from that of domestic policy. According to Cronin and Fordham (1999), the mid-20th century saw a dramatic change in the meaning of being conservative or liberal on foreign policy. In particular, they show that conservatives had supported nationalist (isolationist) foreign policy until the early 1960s, after which they supported more internationalist policy. Conversely, liberals had supported internationalist foreign policy until the early 1960s, after which they became more nationalistic. This distinctiveness of foreign and domestic policy ideologies is confirmed by other studies. For instance, Karol (2009) documents changes in the meaning of being conservative and liberal on foreign trade and defense spending. According to him, until the 1960s, Democrats and liberals supported free trade and defense spending, while Republicans and conservatives opposed free trade and defense spending; however, during the 1960s, liberals and conservatives switched their positions. Thus, since the 1960s, liberals have supported protectionist trade policy and dovish defense policy, while conservatives have supported pro-trade policy and hawkish defense policy. Other studies also document such ideological position switches on foreign policy issues (Keech and Pak 1995; Trubowitz 1998; Irwin and Kroszner 1999; Shoch 2001; Fordham 2007).

In addition, the literature of the two presidencies thesis has long debated whether there are systematic differences between domestic and foreign policy in the way that the President and Congress interact with each other. While the original thesis formulated by Wildavsky (1966) has been empirically challenged, the thesis has received at least limited or conditional support, making it difficult to ignore the possible differences between domestic and foreign policy votes. For instance, in their analysis of presidents' success rates between domestic and foreign policy votes in Congress, Fleisher and Bond (1988) found that Republican presidents were able to enjoy deference from members on foreign policy matters. Later, Fleisher et al. (2000) found that the thesis still applied to minority presidents, although the degree of presidential success on foreign policy votes had declined. More recently, Canes-Wrone, Howell and Lewis (2007), using appropriations votes and congressional choices of bureaucratic structures, found that the thesis is still valid. Thus, the practice of ignoring the difference between foreign and domestic policy votes in Congress is cause for concern.

In this regard, the national security voting index compiled by the American Security Council (ASC) and the *National Journal* vote ratings on foreign policy are useful, as these measures are based purely on foreign policy votes. However, these measures are problematic for two reasons. First, they are available only for a limited number of years. The *National Journal* ratings are available from 1993. The ASC voting indices are available from 1969, with some years missing after that.

² To be clear, my issue with NOMINATE measures is not with its method or procedure. Ideally, NOMINATE procedure can be applied to foreign policy votes, producing a measure of foreign policy ideology. However, there is no publicly available software to implement DW-NOMINATE procedure, which allows for cross-time comparison. Although software for W-NOMINATE is available, W-NOMINATE score is not suitable for cross-time comparison. For this reason, my discussion of NOMINATE focuses on its measures, not on its method. Also, the debate over the methodological differences between NOMINATE and Bayesian IRT estimation (Carroll et al. 2009a; Clinton and Jackman 2009) suggests that the discrepancy between the measures produced by the two methods will be insignificant if the same data are used.

The more serious problem of these measures is that they are based on selective votes on foreign policy. For instance, according to the *National Journal* website, their analysts select votes that “show ideological distinctions between members;” thus, only controversial votes are selected, while votes that involve less controversial issues are excluded. As the comparison of my measure with these measures will demonstrate, this approach causes two problems. First, this approach tends to overemphasize ideological distinction among members, producing a more polarized distribution of ideology in Congress. As consensus votes and less controversial votes are excluded, the ideological differences among members will be estimated to be much larger than they are, making liberals more liberal and conservatives more conservative than they really are (Snyder 1992; Poole and Rosenthal 1997). Another problem of this approach is that the measures are unable to distinguish among non-moderate legislators, producing less informative measures of ideology. As only a small number of controversial votes are used to measure members’ ideology, there will be few votes that distinguish among non-moderate legislators. As a result, many non-moderate legislators will be estimated to have the same or similar positions, even when there are ideological differences among them.

Given these problems of the existing measures, we need a new approach that allows for the utilization of all foreign policy votes, at the same time taking into account differences across votes in terms of importance and relevance to foreign policy.

A NEW MEASURE OF FOREIGN POLICY POSITIONS OF MEMBERS

This section discusses how applying a Bayesian IRT model to all foreign policy votes improves the measurement of foreign policy positions of members.

Data

To measure foreign policy positions of members, I use all congressional votes on foreign and defense policy since 1945. The data include all votes on foreign relations, national defense, foreign trade, and immigration.³ To collect foreign policy votes, I searched the Poole and Rosenthal (1997) data set of all congressional votes for foreign policy votes in the House and the Senate, from 1945 through 2010. Poole and Rosenthal have collected and coded all congressional votes starting with the First Congress. Therefore, an efficient way to identify foreign policy votes is to do keyword searches in their codebooks. The keywords used in my search include, but are not limited to, *security, defense, treaty, weapons, foreign, trade, immigration, refugees, asylum, General Agreement on Tariffs and Trade (GATT), World Trade Organization, World Bank, International Monetary Fund, diplomacy, aircraft, military, international, nuclear, arms, Army, Navy, Air Forces, Marines, United Nations, intelligence, Soviet, China, and Vietnam*. This produces around 50 to 200 foreign policy votes per chamber in each Congress.

For verification, I compared the list of foreign policy votes with that of key votes by *Congressional Quarterly*. Since 1947, *Congressional Quarterly* has selected votes in each Congress that are of great public interest. As *Congressional Quarterly* lists key votes on various issues, checking this source is a useful step in evaluating whether my keyword search of the

³ On national defense, I did not include votes on veterans’ benefits. On immigration, I included votes regarding refugees and asylum, whereas votes on the legalization of undocumented immigrants are excluded. Also, I estimated foreign policy positions using only foreign relations and national defense votes, excluding trade and immigration votes, because some studies find that political parties are more polarized on foreign economic issues (Crabb and Holt 1992; Prins and Marshall 2001). The exclusion of trade and immigration votes does not change the measure significantly. See supplemental files.

Poole and Rosenthal data set has missed any important foreign policy votes. My keyword search has captured all of the key foreign policy votes selected by *Congressional Quarterly*.

Model

After creating the data set of foreign policy votes, the next step is to estimate the ideologies of members using these votes. While some measures, like the *National Journal* vote ratings, count the number or percentage of members’ “yea” votes to measure their ideologies, this method ignores the possibility that not all congressional votes affect foreign policy equally. Some votes are more important than others. Even among controversial votes, there is no guarantee that all controversial votes are equally important. Therefore, in estimating foreign policy ideologies, we need a model that can differentiate congressional votes by their importance. To deal with this issue, I employ a Bayesian IRT model.

In this model, the probability of legislator *i* voting “yea” on vote *j* is modeled using three parameters: the policy position or ideal point of legislator *i* (θ_i), a discrimination parameter of vote *j* (a_j), and a difficulty parameter of vote *j* (b_j). Then, if we assume the probit link function, the probability is written as

$$\Pr(Y_{ij} = 1 | \theta_i, a_j, b_j) = \Phi(a_j\theta_i - b_j), \tag{1}$$

where Y_{ij} is 1 if legislator *i* voted “yea” on vote *j* and 0 otherwise. Consequently, the likelihood is

$$L(\theta_i, a_j, b_j | Y) = \prod_{i=1}^n \prod_{j=1}^m [\Phi(a_j\theta_i - b_j)]^{Y_{ij}} \times [1 - \Phi(a_j\theta_i - b_j)]^{1 - Y_{ij}}, \tag{2}$$

where Y is a $n \times m$ matrix of roll call votes, n the number of legislators, m the number of votes.

In a logit or probit model, the characteristics of votes are ignored because only characteristics of legislators— X_i —are assumed to affect the probability of voting yes. That is, the structure of a probit model is $P(Y_{ij} = 1|\beta) = \Phi(X_i \beta)$. In a probit model, there is no parameter that captures the characteristics of vote *j*. In an IRT model, the vote-specific characteristics are captured by difficulty and discrimination parameters.

First, the difficulty parameter (b_j) is useful in capturing the degree to which a vote is extreme or moderate by measuring the location of the cut-off point (b_j/a_j) for each vote.⁴ That is, the difficulty parameter is useful in defining the location on a space that divides those who are likely to vote for and those who are likely to vote against a bill.

Second, a discrimination parameter (a_j) measures the extent to which a legislator’s ideology affects his or her voting decision. When a vote has a large discrimination parameter, the voting behavior of conservative and liberal legislators will differ significantly. On the other hand, when a vote has a discrimination parameter of 0, there will be little difference in voting between conservative and liberal legislators. In other words, we can think of the discrimination parameter as a weight of ideology in the voting decision. Therefore, a vote is likely to have a low discrimination parameter when the vote is unrelated to the issue at hand.

In analyzing foreign policy votes that include votes on defense spending, this is a particularly useful feature of an IRT model. In my data, the proportion of votes on defense appropriations bills in each Congress ranges from 5 to 32 percent. A major problem in dealing with defense spending votes—and foreign trade votes, for that matter—is that non-ideological factors, such as constituent interests, can have significant influence on how legislators vote. Although studies

⁴ The right-hand side of Equation (1) can be rewritten as $\Phi(a_j (\theta_i - c_j))$ if we set $c_j = b_j/a_j$, where c_j is a cut-point. For more details on the IRT model, see Clinton, Jackman and Rivers (2004).

have found that the ideology of the member of Congress is the most important factor in explaining defense spending votes, constituent interests are also found to have a certain degree of influence (Bernstein and Anthony 1974; Carter 1989; Lindsay 1990). Therefore, it is likely that some votes are more related to constituent interests than to the ideology of the legislator. If this is the case, an IRT model will help us deal with such votes, because these votes are likely to have a small or 0 discrimination parameter.

In summary, an IRT model allows us to utilize multiple votes to estimate the ideologies of legislators and vote-specific parameters. By taking into account such differences between votes, this model minimizes the likelihood of an idiosyncratic vote skewing the results. This model can be estimated using the Markov chain Monte Carlo (MCMC) method. We can use WinBUGS or MCMCpack package in R for implementation.⁵ For implementation, we need to specify the prior distributions. For bill parameters, I used diffuse priors. For ideal points, I used standard normal distributions for model identification. Specifically, the following prior distributions are used:

$$\theta_i \sim N(0, 1); a_i \sim N(0, 10); b_j \sim N(0, 10). \quad (3)$$

While the use of standard normal distributions for the prior distributions of ideal points solves the scale invariance problem, we still need to constrain some ideal points or bill parameters to determine which direction is conservative or liberal.⁶ In most cases, I constrained some extreme legislators to be positive or negative to identify the model. For instance, Senators Jesse Helms (R-NC), and Strom Thurmond (R-SC) are constrained to be positive, while Senators Edward Kennedy (D-MA) and John Kerry (D-MA) are constrained to be negative.

COMPARISON 1: 103RD US HOUSE OF REPRESENTATIVES

To demonstrate the differences and similarities between the Bayesian IRT measure and the existing measures, I use the estimates from the 103rd House of Representatives. As *National Journal* vote ratings are available from this Congress, we can compare all four existing measures—NOMINATE score, ADA score, ASC national security voting index, and *National Journal* foreign policy vote rating—with the IRT measure. For the NOMINATE score, I use the dimension 1 score of W-NOMINATE.⁷ As ADA score, ASC index, and *National Journal* vote rating are available each year, I took the average of the scores in 1993 and 1994 to come up with the scores for the 103rd House.

My measure of foreign policy positions comes from fitting the one-dimensional IRT model to the data from the 103rd House. Out of 131 votes on foreign policy, only four votes are estimated to have a discrimination parameter that is not significantly different from 0, which indicates that

⁵ The Geweke test scores for most MCMC chains were greater than -1.96 and <1.96 , indicating that the chains converged to a stationary series. See supplemental files for a sample of traceplots and density plots.

⁶ See Rivers (2003) for the identification issue in IRT models.

⁷ While other NOMINATE measures are also available, I chose the dimension 1 score of W-NOMINATE because it is the most commonly used non-dynamic measure of ideal points. In addition, other measures provide very similar results. For instance, A-NOMINATE, which is a new NOMINATE measure that relaxes the assumption about the utility function (Carroll et al. 2013), produces ideal point estimates that are very similar to the dimension 1 score of W-NOMINATE (see supplemental files). There are two reasons for the decision to use the dimension 1 of W-NOMINATE instead of the dimension 2. First, foreign policy votes tend to split members on the first dimension rather than the second dimension of W-NOMINATE, which is indicated by the fact that most cutting line angles are between 45 and 135° (see supplemental files). Second, the dimension 2 score of NOMINATE does not perform better than the dimension 1 score of NOMINATE in predicting foreign policy votes (see Figure 2 in the next section).

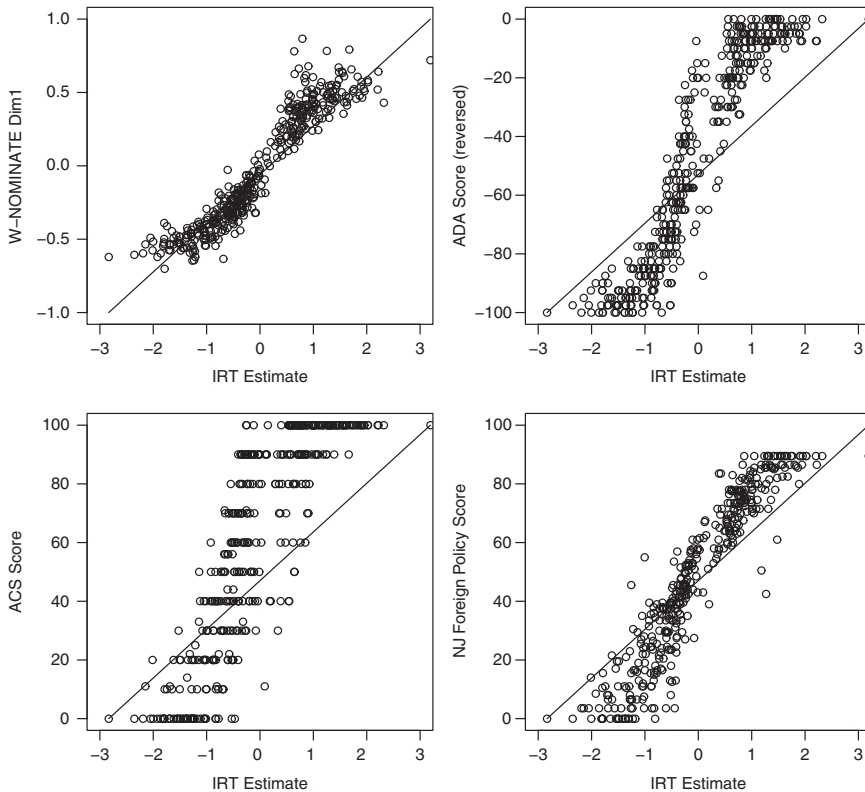


Fig. 1. Comparison of item response theory (IRT) estimate, W-NOMINATE, Americans for Democratic Action (ADA) score, American Security Council (ASC) score, and National Journal (NJ) foreign policy vote ratings (103rd House)

Note: The location of each dot on the horizontal dimension (Dim) represents the IRT estimate of a legislator. The location of the vertical dimension represents the legislator’s score in W-NOMINATE, ADA score, ASC score, and NJ vote rating, respectively.

the one-dimensional model captures the data effectively. For comparison, Figure 1 plots the Bayesian IRT estimates against each of the four existing measures. To put these measures in a comparable scale, the 45° line was created by connecting the possible minimum and maximum scores in each of the horizontal and vertical dimensions.⁸ The figure shows that these measures are highly correlated. The coefficients of correlation between the IRT measure and each of the existing measures are 0.94 (W-NOMINATE), 0.92 (ADA score), 0.86 (ASC score), and 0.93 (*National Journal* rating).

The figure shows that, while there is no significant difference between the dimension 1 score of W-NOMINATE and the IRT measure, there are notable differences between the IRT measure and the three interest/advocacy group measures—ADA, ASC, and *National Journal* scores. First, all three measures have a tendency to overemphasize the degree of polarization, compared with the IRT measure. That is, each of these measures tends to have higher scores (above the

⁸ NOMINATE measures are constrained to be between -1 and 1, while ADA, ASC, and *National Journal* scores range from 0 to 100. The estimates from an IRT model are not constrained. However, most estimates tend to be between -3 and 3.

45° line) when the IRT measure is positive, and lower scores (below the 45° line) when the IRT measure is negative. We can attribute this inflation of the ideological difference among legislators to the selectiveness of their votes. As *National Journal's* selection criterion makes clear, the *National Journal* scores tend to focus on controversial votes and exclude less controversial votes. Accordingly, the measures based only on controversial votes are likely to inflate the differences among legislators and downplay the similarities among them.⁹

There is another difference between the IRT measure and the three existing measures that use selective votes. The IRT measure is better at distinguishing ideological differences among non-moderate legislators than the ADA score, the ASC score, and *National Journal* vote ratings are. For instance, in the bottom left panel of Figure 1, there are a group of legislators with a 100 percent score on the ASC score, while their positions are well sorted out on the horizontal dimension. This is also the case with the ADA score and *National Journal* vote ratings, indicating that the use of selective votes produces less informative measures of ideology, especially among non-moderate legislators. By using less controversial or lopsided votes as well as controversial votes, my approach is able to capture subtle differences among non-moderate legislators.

While the differences between the IRT measure and the three measures are noticeable, there are few systematic differences between the IRT measure and the W-NOMINATE score. However, as I will demonstrate in the next section, the measures based on domestic as well as foreign policy votes have serious problems when one engages in a dynamic analysis of foreign policymaking in Congress.

COMPARISON 2: DYNAMIC ANALYSIS OF POLITICAL PARTIES AND SENATORS

In this section, I demonstrate the advantages of my measure and the problems of the existing measures by analyzing foreign policy positions of political parties and individual legislators over time. For cross-Congress comparison, I use a dynamic version of the IRT model developed by Martin and Quinn (2002). The difficulty in comparing the measures across Congresses is that the agenda content of each Congress is likely to differ, creating different scales across different Congresses. For instance, policy positions can be estimated to be more polarized in one Congress just because there were more controversial votes in that particular Congress, even though there were few changes in members' policy positions. For this reason, we should be careful when we compare estimates from one Congress with those from another Congress. Martin and Quinn (2002) address this issue by introducing the estimates of policy position in the previous Congress as the prior distribution for the policy position of the next Congress in Bayesian estimation.¹⁰ By combining information from previous Congresses in estimating policy positions in the current Congress, this approach allows us to directly compare measures from one Congress to another. I fit a one-dimensional dynamic IRT model to all foreign policy votes from 1945 to 2010 for the Senate.¹¹

⁹ For details, see Snyder (1992) and Poole and Rosenthal (1997). Also, see supplemental files for Monte Carlo experiments that illustrate this problem.

¹⁰ This model can be estimated using MCMCdynamicIRT1d in the MCMCpack package. See supplemental files for the sample R code.

¹¹ One possible objection to fitting a one-dimensional model is that foreign policy issues might be better captured by a multidimensional model. I do not argue that foreign policy debates are always one dimensional. Instead, I suggest and find that congressional debates over foreign policy issues for a given time can be captured by a one-dimensional model. Usually, one issue dimension tends to be dominant for a given time in Congress. For instance, the isolationist versus interventionist dimension was dominant before the 1950s. However, after the 1950s, the unilateralist versus multilateralist dimension became dominant. Additionally, the small percentage of

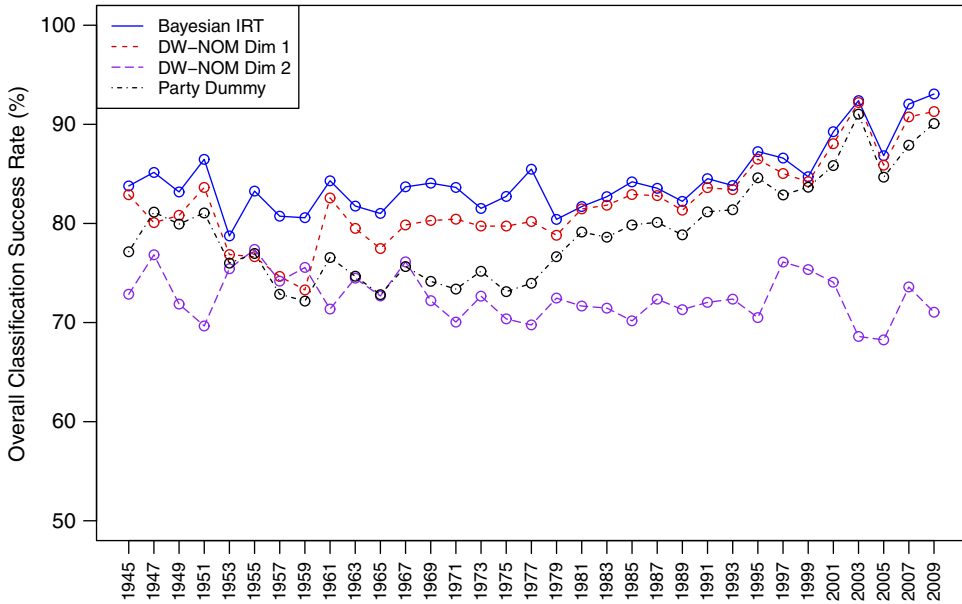


Fig. 2. Overall classification rates of logit models fit to foreign policy votes (Senate)

Note: The classification rates represent the overall classification rates of logit models fit to each of the foreign policy votes in each Congress. Each logit model includes each one of the measures—the Bayesian item response theory (IRT) measure, the dimension (Dim) 1 score of DW-NOMINATE (DW-NOM), the Dim 2 score of DW-NOM, and a party indicator variable (GOP = 1)—as the sole predictor.

For the purpose of illustration, I compare the performance of my measure with that of DW-NOMINATE, which is a dynamic version of W-NOMINATE. As ADA score, ASC score, and *National Journal* vote ratings do not address the issue of agenda change across Congresses, they are not suitable for a dynamic analysis. In addition, as ASC score and *National Journal* vote ratings cover a limited number of years, they are not useful for a dynamic analysis. DW-NOMINATE addresses the issue of agenda changes across Congresses by modeling policy position changes of a legislator with a polynomial process. (See Carroll et al. 2009b for more details.)

Before I examine how the Bayesian IRT measure and the dimension 1 score of DW-NOMINATE measure up to the historical accounts of congressional foreign policymaking, I check the performance of the measures by comparing the classification rates of the logit models that include each of the measures as the sole predictor. That is, I fit the following model to each of the foreign policy votes: $\Pr(Y_i = 1) = \Lambda(\alpha + \beta X_i)$, where X_i is each of the IRT measure, the dimension 1 score of DW-NOMINATE, the dimension 2 score of DW-NOMINATE, and party indicator variable (GOP = 1).¹² I then compute the classification success rate using the model estimate. Figure 2

(Footnote continued)

votes with 0 discrimination parameters suggests that the one-dimensional model is sufficient and appropriate. See supplemental files.

¹² Alternatively, one might compare classification rates of the IRT and DW-NOMINATE models directly. However, this approach is more suitable when we compare the performance of the models rather than that of the measures. As these measures are more likely to be used in a logit or probit model, comparing the classification rates of the logit models using each of the measures as the only predictor will be more useful for potential users of the measures. See supplemental files for the comparison of classification rates of the IRT and DW-NOMINATE models.

reports the average classification rates by Congress. The logit model with a party indicator variable serves as the baseline. The dimension 2 score of DW-NOMINATE was used to check the possibility that foreign policy votes might have loaded on the second dimension. The figure, however, shows that foreign policy votes are more related to the first dimension of DW-NOMINATE than the second dimension, as the classification rates of the dimension 2 score of DW-NOMINATE are almost always lower than the dimension 1 score of DW-NOMINATE.

For our purposes, two things are noteworthy in the figure. First, the IRT measure performs better than the dimension 1 score of DW-NOMINATE in most Congresses. However, there is variation across time. While the IRT measure clearly performs better than DW-NOMINATE before 1979, its performance is only marginally better after 1980. Interestingly, this coincides with the time when the performance of the model with the party indicator variable improves significantly. This indicates that, when political parties are polarized (i.e., when knowing a legislator's party identification is sufficient to predict his or her votes >80 percent of the time), the IRT measure and the dimension 1 score of DW-NOMINATE do not make much difference in predicting a legislator's foreign policy voting. This suggests that increased party polarization in Congress has reduced the difference between members' domestic and foreign policy positions. However, when political parties were not polarized over foreign policy, as in the 1950s and 1960s, the IRT measure clearly predicts members' voting on foreign policy better than the dimension 1 score of DW-NOMINATE. In the following sections, I examine how the Bayesian IRT estimate and the dimension 1 score of DW-NOMINATE measure up to historical accounts of foreign policy positions of political parties and individual legislators since 1945.

Foreign Policy Positions of Political Parties Since 1945

A simple way to trace the policy position of a political party is to examine the location of the median of the party members' policy positions. Figure 3 shows the location of each party's median foreign policy position since 1945.¹³ The Bayesian IRT estimate in the top panel of the figure illustrates that partisan difference was significant throughout the postwar period, with the exception of the 1950s. Also, it shows that the ideological difference between the two parties on foreign policy has been increasing since the 1960s.

For the purpose of comparison, the bottom panel of Figure 3 plots the median of each party using the dimension 1 score of DW-NOMINATE. While there are some similar findings, such as increasing partisan polarization from the 1960s, the biggest difference between the IRT measure and the DW-NOMINATE score is that the DW-NOMINATE score does not provide any evidence of partisan position changes. Throughout the entire postwar period, Republicans are found to have been conservative, whereas Democrats are estimated to be liberal.

However, the top panel of Figure 3 documents that political parties switched their foreign policy positions in the late 1950s. Until the late 1950s, the center of the Democratic Party was estimated to be more hawkish than that of the Republican Party. However, since the early 1960s, the Republican Party has become more hawkish. As discussed earlier, this partisan position switch on foreign policy is consistent with the findings of Cronin and Fordham (1999). While Democrats were liberal on domestic policy throughout the period, their foreign policy positions changed from being hawkish in the 1940s and 1950s to being dovish since then. Likewise, Republicans' foreign policy positions changed from being dovish before the 1960s to being hawkish since then, even though their domestic policy positions can be characterized to

¹³ In order to check whether the findings are affected by the choice of votes, I plotted Senate party medians with congressional quarterly key votes. The results are similar. See supplemental files.

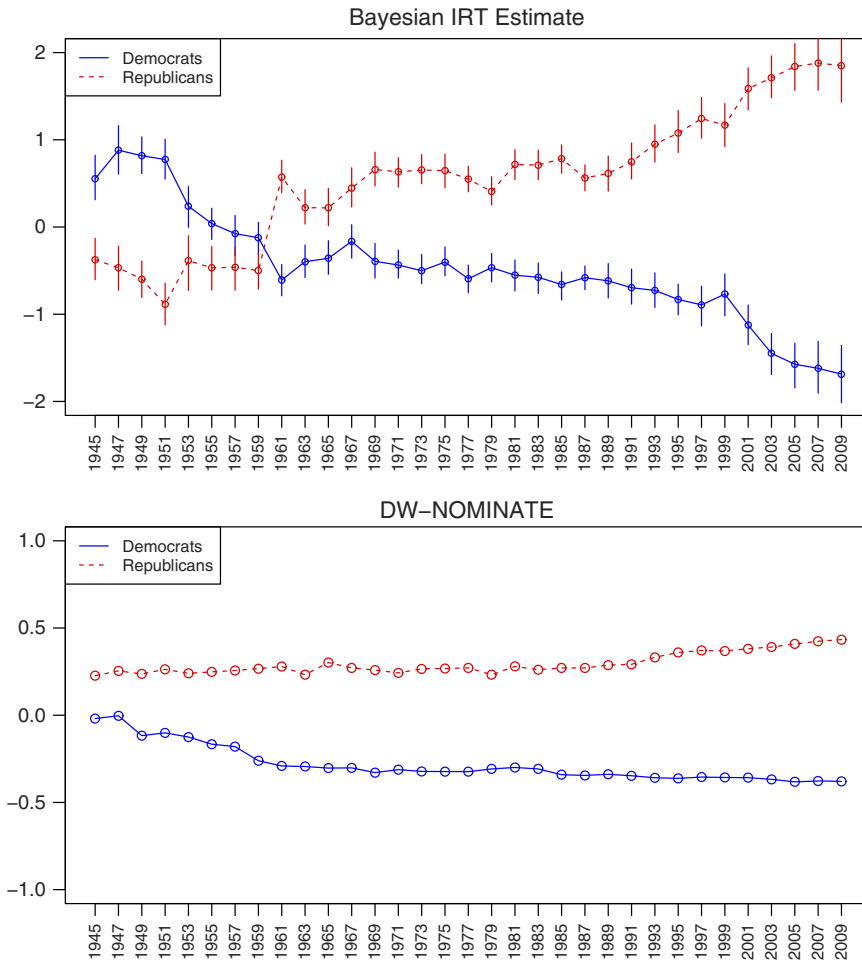


Fig. 3. Median foreign policy positions of political parties (Senate)
 Note: Each party's median represents the location of each party's median member. In the top panel, the bars around the dots represent the 95 percent credible intervals. IRT = item response theory.

be conservative throughout the period. This justifies the need to have a distinctive measure of foreign policy positions.

In my measure of foreign policy positions, the meaning of hawkishness has also changed around the 1960s. Before the 1960s, being hawkish represented supporting multilateral interventions (such as support for the UN and the NATO), defense spending, and free trade. However, since the 1960s, hawkishness has been more closely associated with supporting unilateral interventions, while support for defense spending and free trade is still related to being hawkish. Likewise, the meaning of dovishness has changed around the 1960s. Before the 1960s, dovishness meant opposition to multilateral interventions, defense spending, and free trade. Since the 1960s, being dovish has been associated with supporting international organizations, although it still represents the opposition to defense spending and free trade. Therefore, while a unidimensional model is sufficient to capture tensions over foreign policy in Congress, we need to recognize that the meaning of being hawkish before and after the 1960s is not exactly the same.

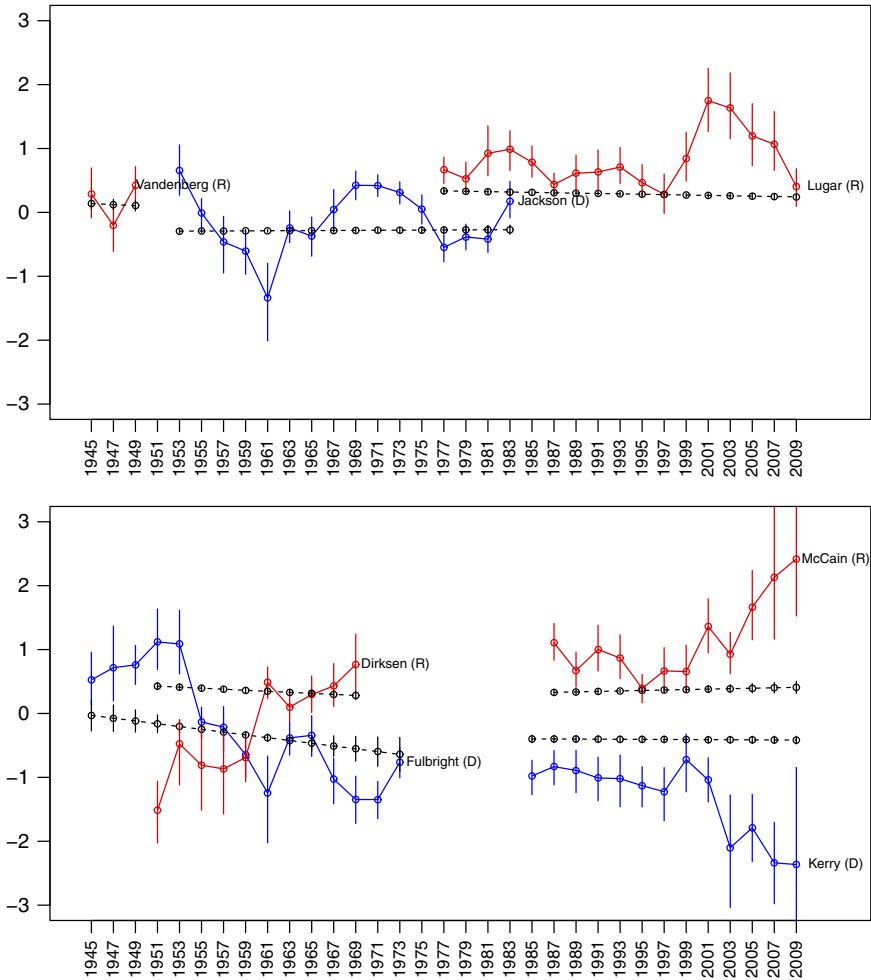


Fig. 4. Foreign policy positions of key senators

Note: The solid lines trace the mean estimates of a dynamic item response theory (IRT) model. The dotted lines trace the point estimates of DW-NOMINATE. Vertical bars represent 95 percent credible and confidence intervals for the IRT estimates and the DW-NOMINATE scores, respectively.

Foreign Policy Positions of Key Senators

In this section, I focus on individual senators' foreign policy positions to illustrate the differences between my measure and the DW-NOMINATE score at the micro level. Figure 4 presents the estimated foreign policy positions of some of the senators who were influential on foreign policy matters. Dotted lines represent the senators' DW-NOMINATE score, while the solid lines represent the IRT measure. The vertical bars around the point estimates represent 95 percent credible and confidence intervals for the IRT estimates and the DW-NOMINATE scores, respectively.¹⁴ As the DW-NOMINATE score and my measure are in different scales,

¹⁴ The 95 percent credible intervals for the IRT estimates tend to be larger than the 95 percent confidence intervals for the DW-NOMINATE scores for two reasons. First, the DW-NOMINATE measure is based on all

there is no point in comparing the absolute values. Instead, I focus on comparing how the two measures capture over-time *changes* in individual senators' foreign policy positions.

In the top panel of the figure, I plot some of the moderate senators who had significant influences on foreign policy. Arthur Vandenberg (R-MI), who was the chair of the Senate Foreign Relations Committee (SFRC), cooperated with the Truman Administration in setting the basis of Cold War foreign policy. Henry "Scoop" Jackson (D-WA) was one of the hawkish Democrats who had a reputation for anti-communist and pro-defense spending positions. Richard Lugar (R-IN) was a moderate Republican who was deeply involved in foreign policymaking as the chair and ranking member of the SFRC until he was defeated in the recent Republican primary.

For all three senators, if we rely on the DW-NOMINATE score to measure their foreign policy positions, we will conclude that there were no or few changes in these senators' foreign policy positions. However, the IRT measure reveals that their foreign policy positions changed rather significantly over the years. For instance, the IRT measure reveals that Senator Vandenberg moved slightly toward a more hawkish direction during the 81st Congress (1949–1950). In the case of Senator Jackson, the IRT estimate shows that his reputation as an anti-communist and pro-defense hawk was the result of his distinct foreign policy positions between 1967 and 1976. During this period, his foreign policy positions moved in the opposite direction of Senate Democrats. As the movement of the Democratic median (the top panel of Figure 3) and Senator J. William Fulbright (the bottom panel of Figure 4) indicate, many Democrats during this period changed their foreign policy stance from hawkish to dovish positions. However, Senator Jackson continued to support the Vietnam War and cooperated with the Nixon Administration on foreign policy issues (Kaufmann 2000, 242–5). While he modified his hawkish foreign policy positions during the Carter Administration, the election of a Republican President seems to have set him free to express his hawkish positions in the early 1980s.¹⁵ In the case of Senator Lugar, the IRT estimate shows that he maintained his moderate stance throughout his tenure in the Senate, with the exception of the period right after the 9/11 terrorist attacks in 2001. As the chair of the SFRC (2003–2006), he supported President Bush's foreign policy until he openly criticized Bush's Iraq policy in 2007 (Zeleny 2007).

While the top panel of Figure 4 shows that the IRT measure captures subtle but significant changes in foreign policy positions of even moderate senators, the bottom panel provides convincing evidence against using the DW-NOMINATE score as a measure of foreign policy positions, by analyzing the senators whose position changes are relatively well known. First, the IRT measure shows that Senator Everett Dirksen (R-IL) had a dovish position in the early 1950s and then transformed himself into a pro-defense hawk through the late 1950s and early 1960s. This is consistent with the fact that, as a Midwestern Republican, he was a member of the party's Old Guard, led by isolationist Robert Taft of Ohio. However, during the Eisenhower Administration, Dirksen gradually changed his foreign policy position in support of interventionism and strong defense (Dueck 2010, 125–6). While the IRT

(Footnote continued)

congressional votes, while my measure is based only on foreign policy votes, which add up only to 20 percent of total votes in each Congress on average. Another reason that the uncertainty measure is larger for the IRT estimates is the difference in scale between DW-NOMINATE and the IRT estimates. While DW-NOMINATE scores are constrained to be between -1 and 1 , IRT estimates tend to range from -3 to 3 .

¹⁵ Similarly, Senator Jackson's position during the Kennedy Administration can be explained by his deference to President Kennedy. For instance, he endorsed the Test Ban Treaty with the Soviet Union out of deference to the President (Kaufmann 2000, 148–53). This type of deference to co-partisan presidents on foreign policy is also observed in Senators Lugar (George W. Bush on the Iraq War), Dirksen (Dwight Eisenhower on interventionist policy), Fulbright (Lyndon B. Johnson on the Vietnam War), and McCain (George W. Bush on the Iraq War).

measure captures this change in Senator Dirksen's foreign policy positions, the DW-NOMINATE score does not show such changes at all.

Senator Fulbright is famous for his withdrawal of support for President Lyndon Johnson's escalation of the Vietnam War. As the chair of the SFRC during the Johnson Administration, Senator Fulbright had supported President Johnson's escalation of the war in Vietnam by sponsoring the Gulf of Tonkin Resolution. However, in the late 1960s, he changed his position and opposed the American involvement in Vietnam by holding a series of congressional hearings on the Vietnam War (Johnson 2006, 115–8). The IRT measure captures such changes in his position during the 1960s. The IRT measure also captures Senator Fulbright's support for international involvements—such as joining the UN, the creation of the Fulbright Scholarship program, and the endorsement of US intervention in Korea—in the 1940s and 1950s (Woods 1995, 166). Simply put, the IRT measure closely traces Senator Fulbright's foreign policy positions as being full of ups and downs. In contrast, the DW-NOMINATE score does not capture Senator Fulbright's tumultuous position changes. Rather, his DW-NOMINATE score suggests that his move toward liberalism was consistent.

Finally, Senators John McCain (R-AZ) and John Kerry (D-MA) show that the IRT measure captures senators' foreign policy positions effectively. In the bottom panel of Figure 4, the IRT measure shows that Senator McCain became more hawkish after the US invasion of Iraq in 2003 and in the period leading up to the Republican primaries in 2008 when he became the Republican nominee for the President. As he missed many votes in the Senate during his election campaigns, the 95 percent credible intervals are quite large for his estimates for the 110th and 111th Senates. Nevertheless, the IRT measure makes it clear that he became more hawkish relative to his position before 2001. The DW-NOMINATE score does not reflect such changes in his foreign policy positions. On the other hand, according to the IRT measure, Senator Kerry moved in the opposite direction during the same period. He became more dovish around his primary campaigns for the Democratic nomination for President in 2004. While the IRT measure captures such changes in his positions, his DW-NOMINATE score does not show any change.

CONCLUSION

In this paper, I have proposed an alternative measure of foreign policy positions of members of Congress. In doing so, I have demonstrated the similarities, differences, and advantages of my measure by comparing it with commonly used measures. While my analysis of the 103rd House of Representatives suggests that my measure is quite similar to the existing measures, I noted important differences. First, I demonstrated that the existing measures based on selective votes tend to exaggerate ideological differences among members, because their data consist of controversial foreign policy votes (ADA, ASC, and *National Journal* scores). It was also noted that ADA, ASC, and *National Journal* measures provide less informative estimates for non-moderate legislators because they use a small number of selective votes.

By analyzing the historical evolution of the foreign policy positions of political parties and certain influential senators, I demonstrated the problems with using all congressional votes to measure foreign policy positions. Given the predominance of domestic policy votes (about 80 percent of all votes), DW-NOMINATE scores conflate a member's foreign policy position with his or her domestic policy position and thus were not able to identify important changes in foreign policy positions of legislators. In a way, this is consistent with the fact that DW-NOMINATE uses bridge legislators and in general shows that legislators tend to be static in their ideology (Poole 2007). We can thus take this as demonstrating that, while overall ideology

of a legislator is stable, their position on a specific policy, like foreign policy, can show interesting changes. It is therefore better to determine the foreign policy position of a member from his or her foreign policy votes in order to produce a measure consistent with historical accounts of congressional foreign policymaking.

REFERENCES

- Bernstein, Robert A., and William W. Anthony. 1974. 'The ABM Issue in the Senate, 1968–1970: The Importance of Ideology'. *American Political Science Review* 68(3):1198–206.
- Canes-Wrone, Brandice, William G. Howell, and David E. Lewis. 2007. 'Toward a Broader Understanding of Presidential Power: A Reevaluation of the Two Presidencies Thesis'. *Journal of Politics* 69:1–16.
- Carroll, Royce, Jeffrey Lewis, James Lo, Keith Poole, and Howard Rosenthal. 2009a. Comparing NOMINATE and IDEAL: Points of Difference and Monte Carlo Tests'. *Legislative Studies Quarterly* 34:555–91.
- Carroll, Royce, Jeffrey B. Lewis, James Lo, Keith T. Poole, and Howard Rosenthal. 2009b. Measuring Bias and Uncertainty in DW-NOMINATE Ideal Point Estimates via the Parametric Bootstrap'. *Political Analysis* 17:261–75.
- Carroll, Royce, Jeffrey B. Lewis, James Lo, Keith T. Poole, and Howard Rosenthal. 2013. 'The Structure of Utility in Spatial Voting Models of Voting'. *American Journal of Political Science* 57:1008–28.
- Carter, Ralph G. 1989. 'Senate Defense Budgeting, 1981–1988: The Impact of Ideology, Party, and Constituency Benefits on the Decision to Support the President'. *American Politics Quarterly* 17(3): 332–47.
- Clinton, Joshua, and Simon Jackman. 2009. 'To Simulate or NOMINATE?'. *Legislative Studies Quarterly* 34:593–621.
- Clinton, Joshua, Simon Jackman, and Douglas Rivers. 2004. 'The Statistical Analysis of Roll Call Data'. *American Political Science Review* 98:355–70.
- Crabb, Cecil V., and Pat M. Holt. 1992. *Invitation to Struggle: Congress, the President, and Foreign Policy*, 4th ed. Washington, DC: Congressional Quarterly Press.
- Cronin, Patrick, and Benjamin O. Fordham. 1999. 'Timeless Principles or Today's Fashion? Testing the Stability of the Linkage Between Ideology and Foreign Policy in the Senate'. *The Journal of Politics* 61(4):967–98.
- Dueck, Colin. 2010. *Hard Line: The Republican Party and U.S. Foreign Policy Since World War II*. Princeton, NJ: Princeton University Press.
- Fordham, Benjamin O. 2007. 'The Evolution of Republican and Democratic Positions on Cold War Military Spending'. *Social Science History* 31:603–36.
- Fleisher, Richard, and John R. Bond. 1988. 'Are There Two Presidencies? Yes, But Only for Republicans'. *Journal of Politics* 50:746–67.
- Fleisher, Richard, Jon R. Bond, Glen S. Krutz, and Stephen Hanna. 2000. 'The Demise of the Two Presidencies'. *American Politics Quarterly* 28(1):3–25.
- Irwin, Douglas A., and Randall S. Kroszner. 1999. 'Interests, Institutions, and Ideology in Securing Policy Change: The Republican Conversion to the Trade Liberalization After Smoot-Hawley'. *Journal of Law and Economics* 42(2):643–73.
- Johnson, Robert D. 2006. *Congress and the Cold War*. Cambridge: Cambridge University Press.
- Karol, David. 2009. *Party Position Change in American Politics: Coalition Management*. Cambridge: Cambridge University Press.
- Kaufmann, Robert G. 2000. *Henry M. Jackson: A Life in Politics*. Seattle, WA: University of Washington Press.
- Keech, William R., and Kyoungsan Pak. 1995. 'Partisanship, Institutions, and Change in American Trade Politics'. *Journal of Politics* 57(4):1130–42.
- Kingdon, John W. 1981. *Congressmen's Voting Decisions*. Ann Arbor: University of Michigan Press.
- Lindsay, James M. 1990. 'Parochialism, Policy, and Constituency Constraints: Congressional Voting on Strategic Weapons Systems'. *American Journal of Political Science* 34(4):936–60.

- 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* 10:134–53.
- Meernik, James, and Elizabeth Oldmixon. 2008. 'The President, the Senate, and the Costs of Internationalism'. *Foreign Policy Analysis* 4:187–206.
- Poole, Keith T. 2007. 'Changing Minds? Not in Congress!'. *Public Choice* 131:435–51.
- Poole, Keith T., and Howard Rosenthal. 1997. *Congress: A Political Economic History of Roll Call Voting*. New York: Oxford University Press.
- Prins, Brandon C., and Bryan W. Marshall. 2001. 'Congressional Support of the President: A Comparison of Foreign, Defense, and Domestic Policy Decision-Making During and After the Cold War'. *Presidential Studies Quarterly* 31:660–79.
- Rivers, Douglas. 2003. 'Identification of Multidimensional Item-Response Models'. Typescript, Stanford University, Stanford, California.
- Shoch, James. 2001. *Trading Blows: Party Competition and U.S. Trade Policy in a Globalizing Era*. Chapel Hill, NC: University of North Carolina Press.
- Snyder, James. 1992. 'Artificial Extremism in Interest Group Ratings'. *Legislative Studies Quarterly* 17:319–45.
- Trubowitz, Peter. 1998. *Defining the National Interest: Conflict and Change in American Foreign Policy*. Chicago, IL: University of Chicago Press.
- Wildavsky, Aaron. 1966. 'The Two Presidencies'. *Transaction* 4:7–14.
- Woods, Randall B. 1995. *Fulbright: A Biography*. Cambridge: Cambridge University Press.
- Zeleny, Jeffrey. 2007. 'Lugar Urges Quick Shift in Iraq War Strategy'. *New York Times*, 27 June.