

# Measurement, model testing, and legislative influence in the European Union

Jonathan B Slapin  
University of Houston

## Abstract

Within the last several years, new data have become available to test the various theoretical models of EU decision-making, and, in doing so, to assess actor influence. This paper examines the extent to which the recent DEU and DEUII datasets provide sufficient information to distinguish between competing theoretical models of legislative decision-making, and accurately assess the power of the different branches of EU government. It argues that insufficient attention has been paid to measurement error in these data. Once measurement error is accounted for, it becomes clear that these data do not provide sufficient information to distinguish between most models of legislative politics. Moreover, empirical models that fail to account for measurement error are likely to lead researchers to erroneous conclusions about actors' legislative influence.

## Keywords

spatial models, bargaining models, ideal points, measurement, legislative politics,

## Introduction

Studies of separation-of-powers systems often wish to assess the relative power of the branches of government in the legislative process. In a bicameral system, is one chamber more influential in drafting and passing legislation than the other? What is the role of the executive branch? What are the sources of legislative influence — formal rules, preferences, or other institutional resources, such as budgets for staff and research? To answer these basic questions, we first must determine who “won” during the legislative process. Which actor in the political system is best able to change the outcome of negotiations in her favor? To assess actors’ relative success, scholars must measure both the outcome of the legislative process, as well as the preferences of actors involved. Doing so is not straightforward. This article explores a newly available and innovative dataset containing preferences and outcomes on 125 European Union (EU) legislative proposals — the *Decision-making in the European Union II* (DEUII) data (Thomson, Arregui, Leuffen, Costello, Cross, Hertz and Jensen, 2012). It examines the extent to which researchers can use these data to test theoretical decision-making models in the EU, and it assesses the power of the European Parliament (EP), Commission, and Council of Ministers in the EU legislative process.

The paper first discusses the problems associated with assessing the legislative influence of various branches of government in separation-of-powers systems, focusing in particular on the issue of measuring policy preferences and outcomes. It then discusses these issues specifically with respect to the EU and the DEU data.<sup>1</sup> Using Monte Carlo simulations, we demonstrate that a simple regression model accurately captures the relative influence of all actors, but, due to measurement error in the DEU data, the influence of the European Parliament and Commission is likely severely underestimated. After correcting for measurement error, the EP likely has significant influence over legislation, and its influence is greater under the ordinary legislative procedure (co-decision) than under consultation. In addition, measurement error may lead researchers to false conclusions when using the DEU data to test competing theoretical models of legislative decision-making. The results build on recent literature that use Monte Carlo simulations to explore how the DEU data can best be employed

to test spatial theories of legislative decision-making (Junge and König, 2007), as well as the literature that uses the DEU data to evaluate bargaining models and the relative influence of actors (Thomson, Stokman, Achen and König, 2006; Junge, 2010; Thomson, 2011; Aksoy, 2012; Golub, 2012). The findings have significant implications for how researchers test spatial models of legislative politics, and how they ought to collect preference data in the future.

## **Measuring preferences and outcomes in the legislative process**

Researchers wishing to assess the relative influence of various actors in the legislative process face a series of measurement and modeling problems. First, they must measure the policy preferences of the relevant actors in a common space, along with the position of the negotiated outcome, and the position of a reference point, reversion point, or status quo policy. Both the space and the positions are latent, meaning the researcher cannot directly observe the nature of the dimensions or the location of actors and outcomes on them. Instead, latent dimensions, as well as actor and bill positions, must be estimated using a measurement model (Jackman, 2008; Benoit and Laver, 2012).

Once estimates of latent positions are obtained, researchers often wish to use these data to determine the extent to which actors are powerful, i.e. influential in the policymaking process, simply lucky, perhaps due to their proximity to a truly powerful actor, or neither (Barry, 1980). And, finally, upon determining which actors have influence, researchers are often interested in what these results imply about the validity of theoretical models of legislative bargaining. Different bargaining models imply different levels of bargaining power for the relevant actors, and researchers wish to test these competing models against one another (e.g. Thomson et al., 2006; Schneider, Finke and Bailer, 2010; Thomson, 2011).

In the US Congress literature, there has been significant attention devoted to the problem of latent ideal point estimation within chambers, across chambers, across branches of government, and over time (e.g. Poole and Rosenthal, 1985; Clinton, Jackman and Rivers, 2004; McCarty and Shor, 2011). Scholars use ‘common space’ scores, which provide ideology es-

timates for the US House, Senate, and President in the same space, to examine gridlock (Binder, 1999), presidential veto power (Cameron, 2000, chap 6), and bureaucratic discretion (Shipan, 2004), among other issues.

Studies of legislative bargaining in the EU have paid less attention to issues of ideal point estimation in a common space. Given the lack of data, this oversight is not surprising. While there have been numerous studies of ideology in the EP (e.g. Hix, 2002; Hix, Noury and Roland, 2007; McElroy and Benoit, 2007; Proksch and Slapin, 2010; McElroy and Benoit, 2012), some work on policy preferences in the Council (e.g. Mattila and Lane, 2001), and a few studies of attitudes in the Commission (e.g. Hooghe, 2005), researchers have made few attempts to provide comparable estimates of ideology across branches. This is largely due to the poor quality of data. Roll call data is best in the EP, but even here the extent to which roll calls can be used to measure ideology is questionable (Carrubba, Gabel, Murrah, Clough, Montgomery and Schambach, 2006; Proksch and Slapin, 2010). Providing comparable policy positions across the branches of EU government is perhaps the primary contribution of the DEU project (Thomson et al., 2006; Thomson, 2011; Thomson et al., 2012).

Nevertheless, problems exist with how the DEU project collected these preference data. Unlike common space scores based on roll call votes in the US Congress, which are derived directly from spatial voting theory and estimated using an item response model, the DEU data are not derived from an underlying measurement model. Thus, they have no reported uncertainty associated with them. Clearly, though, these positions are not measured perfectly — no estimates of ideology are. All estimates of policy positions are subject some measurement error. When the positions of all actors are measured equally poorly or equally well, the lack of measurement error is less of an issue, but as we will see below, this is unlikely the case in the DEU. The position of the Council, as it is typically operationalized in DEU studies of bargaining models, is more accurately measured than the position of the EP and the Commission.

Measurement models assume that the concept being measured (in this case, ideology) is latent — it cannot be directly observed. Instead, researchers only get to observe noisy indicators of the latent variable (Carmines and Zeller, 1979, 10). Given that each indicator is

noisy, and only captures the ‘truth’ on average, the best means to assess the latent concept is to average over a number of indicators. Because the DEU data contain one position per issue for each member state, plus a position for the EP and the Commission, the Council’s position (generally taken as a weighted average of the member state positions) is better measured than the EP and Commission positions, which are simply point estimates. This discrepancy in the measurement quality of positions across branches creates problems both when attempting to assess the relative influence of these branches and when testing competing theoretical models of the bargaining process.

## **Model assessment and legislative influence**

There exists no shortage of sophisticated theoretical models of EU decision-making (e.g. Bueno de Mesquita and Stokman, 1994; Tsebelis, 1994; Crombez, 1996; Crombez, 1997; Tsebelis and Garrett, 2000). And these models often imply different power and influence for the various branches of EU government. For example, Tsebelis (1994) has famously argued that the EP had conditional agenda setting authority under the cooperation procedure, which it gave up in return for veto power with the move to codecision (Tsebelis and Garrett, 2000). Others have taken issue with his model, arguing about the relative merits of agenda-setting and veto powers (e.g. Moser, 1996; Crombez, 1997). Still others have suggested that EU decision-making is best modeled using a cooperative game theory concept, such as the Nash bargaining model or Shapley Shubik index (Achen, 2006; Schneider et al., 2010).

Testing these models has often led researchers to examine which branch of EU government has the most influence over legislation. Some studies examine the EP’s legislative success by examining the success of EP amendments (e.g. Kreppel, 1999; Tsebelis and Kalandrakis, 1999; Tsebelis, Jensen, Kalandrakis and Kreppel, 2001; Kasack, 2004; Häge and Kaeding, 2007). These studies tend to find evidence supporting the conditional agenda-setting power of the EP. But the authors are also very aware that their models assume that amendments are not offered strategically, an assumption that seems unlikely to hold, and which runs contrary to the assumptions of the rational decision-making models they seek to test. Others

have measured policy positions of all relevant actors on a few directives, and calculated the various theoretical predictions based on the observed data (e.g. König and Pöter, 2001). The König and Pöter study found that the EP is likely unable to use its veto power due to its pro-integrationist positions.

Lastly, numerous authors have used the DEU data to test various models legislative decision-making (Thomson et al., 2006; Thomson, 2011). The collaborators in the DEU project have collected policy positions for member states, the Commission, the EP, and the status quo (where one existed) on 125 controversial pieces of European legislation. The positions were collected through numerous in depth, semi-structured interviews with key informants, who were oftentimes participants in the legislative process from the Commission, member state permanent representations, or the EP. The informants were asked to identify individual issues within the legislation that were the most conflictual, define the endpoints of the issue scales, and identify actors' preferences, the status quo, and the outcome on a 0 to 100 scale representing the bargaining space.<sup>2</sup>

The DEU literature has tended to find that cooperative bargaining models, such as the Nash bargaining solution, better predict outcomes than procedural models that account for the rules of the decision-making process, such as Tsebelis' conditional agenda-setting model. While some using the DEU have found the EP is likely to have influence over decision-making, especially under codecision (Selck and Steunenberg, 2004), recent work by Thomson has suggested that both supranational actors — the EP and Commission — have relatively little influence over decision-making compared with the power suggested by theoretical decision-making models, such as the model proposed by Tsebelis and Garrett (2000) (Thomson and Hosli, 2006; Thomson, 2011, chap 8). Theoretically, the codecision (or ordinary) procedure ought to place to Council and Parliament on an equal footing, but Thomson finds that the supranational actors have, at best, influence equal to 30% of the Council's. Indeed, many of Thomson's models suggest significantly less power for both the EP and Commission, on the order of 3% to 15% of the Council's influence (Thomson, 2011, 207).

What accounts for the fact that, particularly in the DEU data, non-procedural weighted

means models so vastly outperform procedural models that take account of the structure of the legislative process? And why do the EP and Commission have relatively little influence over legislative outcomes in the DEU data when the formal rules suggest they ought to have more power? Using Monte Carlo simulations, as well as the DEU data itself, the remainder of the paper demonstrates that problems arising from measurement error can lead to wrong inferences about the accuracy of theoretical decision-making models and the influence of the EP. Correcting for measurement error, we find that we can no longer conclude that non-procedural models outperform procedural models in explaining legislative outcomes. Moreover, we find that the EP likely possesses more legislative influence than the previous results based on the DEU data suggest.

## Monte Carlo simulations

### Theory testing

We begin by conducting Monte Carlo simulations to assess the ability of data, such as those collected by the DEU project, to distinguish between competing theoretical models of legislative decision-making. We examine whether such data provide a researcher with sufficient statistical power to distinguish between a Romer-Rosenthal (R-R) agenda setter model (Romer and Rosenthal, 1978), chosen because it represents the simplest version of an institutional game, and the Nash bargaining solution (NBS), a cooperative game theory bargaining concept (Nash, 1950). Not only do these models generate simple institutional and non-institutional predictions, in a world with no measurement error they tend to make very different predictions. Thus, they should be relatively easy to distinguish from one another. If the data cannot distinguish between these models, they will not be able to distinguish between more nuanced models that make more similar predictions.

We set up a game in which two players,  $i \in \{A, B\}$ , bargain over 150 independent issues,  $j \in \{1, 150\}$ . On each issue,  $j$ , the players' have single-peaked, Euclidean preferences,  $x_{ij}$ , in a uni-dimensional space,  $S_j$ , consisting of 101 'locations' ranging from 0 to 100,  $S_j = \{0, 1, 2, 3, \dots, 100\}$ .<sup>3</sup> Given that we are discussing European integration, we refer to

these dimensions as representing players' integration preferences across the  $j$  issues, where the status quo level of integration,  $x_{SQ}$ , is 0, and 100 represents maximal change towards greater integration on the issue. But these labels are arbitrary and the dimensions could represent any conflictual issue. Players' ideal points on each issue are drawn from random uniform distributions so that player  $A$  is, on average, more Euroskeptic (has a preference closer to 0) than  $B$ .  $A$ 's ideal points,  $x_{Aj}$ , are drawn from  $S_{Aj} = \{0, 1, 2, 3, \dots, 60\}$ , while  $B$ 's ideal points,  $x_{Bj}$ , are drawn from  $S_{Bj} = \{40, 41, 42, 43, \dots, 100\}$ . We can think of  $A$  as representing the position of the Council, and  $B$  representing a supranational actor, either the Commission or EP. The data are generated in a manner as similar as possible to the DEU datasets.<sup>4</sup>

We will assume that R-R model represents the true data-generating process. Thus, the Monte Carlo exercise explores the circumstances under which we can correctly identify the R-R model as data generating process when tested against the NBS. We assume that player  $B$  is the agenda-setter and makes a closed-rule proposal to  $A$ , which she can either accept or reject. The equilibrium solution on an issue by issue basis is  $x_j^* = 2 * |x_{Aj} - x_{SQj}| + x_{SQj}$  if  $x_{Bj} > 2 * |x_{Aj} - x_{SQj}| + x_{SQj}$ , and  $x_j^* = x_{Bj}$  otherwise. In other words, on issues where  $B$  prefers significantly more integration than  $A$ ,  $B$  must offer to  $A$  the point that makes  $A$  indifferent to the status quo. Otherwise,  $B$  can realize her ideal point.<sup>5</sup> However, as researchers we do not observe true positions, only noisy estimates of the truth. The observed preferences are equal the true positions plus random noise:  $x_{obsAj} = x_{Aj} + \epsilon_{Aj}$ ,  $x_{obsBj} = x_{Bj} + \epsilon_{Bj}$ ,  $x_{obsj}^* = x_j^* + \epsilon_{x_j^*}$ , and  $x_{obsSQj}^* = x_{SQj}^* + \epsilon_{x_{SQj}^*}$ , where  $\epsilon_{ij} \sim N(0, \sigma)$ .<sup>6</sup> These observations are the data we, as researchers, get to observe. We can think of them as being the data produced by the DEU projects. As  $\sigma$  grows large, our measurements (observations) of the true positions and outcomes become worse, although they are still correct on average.

To test whether we are able to distinguish between the R-R model and the NBS with our observed data, we first must calculate the predictions of these two models for each issue on the basis of our observed preferences. We then examine how well these models explain the observed outcome. Thus, for each issue, we calculate the R-R prediction using the observed values rather than the true values:  $RR_{PREDj} = 2 * |x_{obs.Aj} - x_{obs.SQj}| + x_{obs.SQj}$  if



$x_{obs.Bj} > 2 * |x_{obs.Aj} - x_{obs.SQj}| + x_{obs.SQj}$ , and  $RR_{PREDj} = x_{obs.Bj}$  otherwise. The prediction for the NBS is calculated as:

$$NBS_{PREDj} = \max_{x_j^* \in S} [U_A(x_j^*) - U_A(x_{obs.SQj})] * [U_B(x_j^*) - U_B(x_{obs.SQj})].$$

In short, the NBS predicts that actors select the outcome in the bargaining space that maximizes the product of the differences of their individual utilities for the outcome and the status quo on each issue. We assume quadratic loss utility functions when calculating the NBS.

Having calculated the predictions of our two models for all  $j$  issues, we regress the observed outcome in our data on the predictions of the R-R and NBS models, calculated on the basis of our observed preferences and the observed status quo. We suppress the constant because, given the way the data are constructed, these two variables completely explain the variance in the dependent variable.<sup>7</sup> In fact, the R-R variable alone should completely explain the variance as it represents the true data generating mechanism.

In a world with no measurement error (i.e., a world in which the observed R-R variable precisely equals the true R-R issue-by-issue prediction, and the observed NBS variable precisely equals the true NBS issue-by-issue prediction) there should be a perfect 1:1 relationship between R-R variable and the outcome, and no relationship between the NBS variable and the outcome. As measurement error in one independent variable increases, though, the estimation of all coefficients in the model becomes biased. The coefficient on the poorly measured variable is biased towards zero, and the coefficient on the better measured variable may be biased in any direction. Moreover, the coefficient estimates are inconsistent, meaning that bias persists even as the sample size increases (Greene, 2000, 375–80). This is often referred to as the ‘error-in-variables’ problem, and despite its quite substantial consequences for estimation, it has received relatively little attention in the political science literature (but see Blackwell, Honaker and King, 2012). While we add the same amount of measurement error to all positions (both actors, the SQ, and the outcome), the measurement error does not affect the calculation of the R-R and the NBS variables in the same way. Because the

NBS is effectively a weighted average of the positions, some measurement error in the actors' positions tends to cancel out. On average, the measured NBS variable is, therefore, closer to the true NBS solution, than the measured R-R variable is to the true R-R solution. When  $\sigma = 8$ , the mean error for the NBS variable is approximately 9, while the mean error for the R-R variable is approximately 14. In other words, additional measurement error affects the calculation of the R-R variable to a greater extent than the NBS variable, and thus the R-R variable is more poorly measured. As measurement error increases, we would expect greater attenuation bias in the coefficient on R-R variable.

To start, we run four 1000-run simulations, varying  $\sigma$  from 2 to 12 to simulate increasing measurement error in the positions of the two actors. Thus the standard deviation of the error represents between 2% and 12% of the range of the scale, a relatively modest amount of error. The results are presented graphically using boxplots in Figure 1. Each simulation is represented by two boxplots, one displaying the distribution of the 1000 coefficient estimates for the R-R prediction variable, and the second displaying the distribution of coefficient estimates for the NBS variable.

[INSERT FIGURE 1 ABOUT HERE]

When  $\sigma$  is small (2), we are able to differentiate between the models quite easily. The regression model always correctly identifies the R-R variable as better explaining the outcome, with a regression coefficient of approximately 0.75, and always larger than the coefficient on the NBS variable. However, only slight increases in measurement error ( $\sigma = 4$  and above) lead to scenarios in which the two models are nearly impossible to distinguish, or the incorrect model (the NBS) better explains the outcome. When  $\sigma = 4$ , the model estimates the R-R coefficient to be greater than the NBS coefficient in 77% of simulations, when  $\sigma = 8$  the R-R coefficient is greater in less than 2% of simulations. As expected, additional measurement error leads the model to favor the better measured, but incorrect, NBS variable over the worse measured, but correct, R-R variable.

While the DEU provides us with no way to assess measurement error, given that positions are collected on a 0 – 100 scale, a standard deviation of 8 for any given position seems like

the least amount of measurement error that one might expect. It implies that the average position estimate in the data is off by only 8 points on the 101 point scale. In short, even relatively small amounts of measurement error make it difficult for researchers to distinguish between institutional models and non-institutional models, such as R-R and NBS. Even if the R-R model were the correct data generating process, when measurement error exists we could very easily draw the erroneous conclusion that the NBS better explains the bargaining process. Moreover, we have run simulations for the simplest of possible worlds — two actors playing the simplest of non-cooperative institutional games and employing the NBS. If we wished to use such data to distinguish between more nuanced theories that make more similar predictions, we would need much better, and much more data.

### **Legislative influence**

The above Monte Carlo simulations examined whether data resembling those collected in the DEU project are able to distinguish between competing theoretical decision making models. We found that they can only do so if we are willing to assume that positions of actors and outcomes are measured very exactly. We now turn to our attention to tests of legislative influence. Are we able to use the DEU data to assess the relative influence of different actors over the legislative process? To answer this question, we must account for how the positions of collective EU actors are measured in the DEU datasets. The DEU project has collected data on the positions of all member state governments, as well as the position of the EP, the Commission, and — where one could be identified — the legislative status quo. For a small number of issues in the DEUII project, positions for EP groups were collected, as well. However, the experts were not able to identify much variance in the group positions. Thus, for the most part, the Commission and EP preferences are represented by a single point estimate. In contrast, in most analyses, the estimate of the Council position takes into account numerous point estimates — namely the point estimates of each individual member state. It is often calculated as an average (often weighted, but sometimes not) of member state positions (Selck and Rhinard, 2005; Thomson and Hosli, 2006; Thomson, 2011) or using the Council median (Selck and Steunenbergh, 2004). Using Monte Carlo simulations, we will

show that, even when all positions (e.g. EP, Commission, member states, and outcomes) are measured with the same degree of error, as a result of averaging numerous point estimates, the aggregate Council position tends to be measured better than the positions of the EP and Commission. Due to the effects of the ‘error-in-variables’ problem discussed above, models of legislative influence, therefore, tend to overestimate the Council’s strength relative to the supranational actors.

In these Monte Carlo simulations, we begin by constructing true positions for a Council with seven members and a single position for the European Parliament for each of 150 issues. Thus, the simulation is, once again, designed to closely mirror the DEU data. The true positions of the seven Council members are drawn from a random uniform distribution ranging from 0 to 100. For each issue, we calculate the Council mean, which represents the Council’s ‘true’ bargaining position. The positions of the EP are drawn from a random uniform distribution ranging from 40 to 100. The EP, therefore tends to prefer more integration than the average Council member. The true outcome of negotiations on each issue is assumed to represent a negotiation between the EP and the Council in which each branch has equal bargaining power, plus some random error:

$$Outcome_{true} = 0.5 * EP_{true} + 0.5 * CouncilMean_{true} + \epsilon$$

where  $\epsilon \sim N(0, 5)$ . We next calculate our observed positions for each of the seven member states, the EP, and the outcome. They are equal to the true positions plus measurement error drawn from a normal distribution with  $\mu = 0$  and  $\sigma = 8$ . Thus, each position is measured with same degree of error. We obtain a single point estimate for the Council by taking the mean of these seven observed member state positions. The measurement of the Council position mirrors the operationalization found in Thomson and Hosli (2006) and Thomson (2011), and captures the position taken by Achen (2006) that the mean of Council members best captures the decision-making outcome.

Next, we regress the observed outcome on the average of the observed member state positions, i.e. the Council position, and EP position, suppressing the constant just as we

did above. We first regress the true outcome on the true mean Council position and the EP position to ensure that our regression model accurately captures the truth, i.e. the power we have, by construct, assigned to the Council and EP. We conduct 1000 runs of each simulation, and we again present the results using a series of boxplots, found in Figure 2, to represent the coefficients across all runs. The first two boxplots show that the regression model very accurately recaptures the truth, as we would expect, correctly assigning each actor equal power:  $\beta = 0.5$ . We next regress the observed outcome on the observed EP position and the observed Council position, measured as the mean of observed member state positions. Here, we see that the power of the EP is underestimated, while the power of the Council is significantly overestimated. In 92% of the runs the Council is estimated to have greater power than the EP even though we know their power is equal by design. These findings are a result of measurement error. When  $\sigma = 8$ , the average observed EP position is 6.4 away from the true EP position on our 0 to 100 scale, while the average Council mean is only 2.4 away from the true Council mean.<sup>8</sup>

By averaging over the positions of the member states to obtain the observed Council position, the measurement error in the observed Council position is decreased. In effect, the measurement error in individual state positions cancels out. Because the EP is only a point estimate, its measurement error remains. The measurement error in the EP variable leads to an attenuation in the estimate of EP strength, and upward bias on the effects of the better measured Council variable. As the number of positions averaged to obtain the observed Council position increases, this problem is becomes worse.<sup>9</sup>

[INSERT FIGURE 2 ABOUT HERE]

We next attempt to correct for the measurement error in the EP position using an error correction model, namely the SIMEX model developed by Cook and Stefanski (1994). SIMEX regression corrects for measurement error in an independent variable by adding additional error to that variable and using simulations to extrapolate back to a scenario in which no error is present. It requires that the researcher specify the level of measurement error he believes is present in the poorly measure variable. In our simulation, we have set

the level of error, so we can precisely specify it in the SIMEX model. We treat the EP position as the poorly measured variable and we define its measurement error as the average difference of the observed EP position from its true position — approximately 6.4 when  $\sigma = 8$ . The boxplots demonstrate that SIMEX model returns close to the correct coefficients on average, assigning almost equal power to the Council and EP. Moreover, the average difference between the two coefficients across all 1000 simulations is only 0.08.

These Monte Carlo simulations demonstrate that the existence of measurement error in data such as the position estimates collected in the DEU projects can lead researchers to incorrect conclusions about the validity of theoretical bargaining models and the power of actors in the decision-making process. Once we account for measurement error, we find that there is insufficient information in the DEU data to distinguish between even the simplest of competing bargaining models. Moreover, because of the way in which researchers tend to operationalize the Council position, the Council position is better measured than the EP, which could lead researchers to falsely conclude that the Council is more powerful than the EP. Both simulations demonstrate that whenever a researcher pits a variable operationalized as an average against a variable operationalized as a point estimate, the average will prevail even when the point estimate represents the correct model.

## Data and results

Until now, all of the results presented have been based on ‘made-up’ data — data constructed to simulate the data collection process of the DEU projects. Constructed data and Monte Carlo simulations are very helpful to understand what researchers can learn from certain types of data, but of course, at some point we wish to turn our attention to actual data. In this section, we use the DEUII data to assess the relative strength of the EP and the Council, using the knowledge gleaned from the Monte Carlo results presented above.

Using real data introduces additional issues. The DEU data contain missing observations, and the data are hierarchical in nature — multiple issues are nested within each legislative proposal. In addition, the data provide estimates of issue saliency, and many users of the data

have argued that accounting for issue saliency is important when testing bargaining models; actors likely bargain harder on issues most salient to them (e.g. Thomson, 2011; Golub, 2012). We attempt to take all these issues into account. First, to account for missing positions and saliency estimates, we use *Amelia II* to impute missing values (Honaker, King and Blackwell, 2011).<sup>10</sup> Accounting for the multilevel structure of the data is more difficult. Ideally, we would run a multilevel model. However, we already have a complicated error structure due to the presence of measurement error. Thus, it is even more difficult to properly account for the hierarchical nature of the data. There are no easy-to-implement error correction models that account for hierarchical data. While failing to account for missing data and measurement error can lead to biased coefficient estimates, failing to account for non-independence of observations in cross-sectional data only biases standard errors. As our primary focus is on correcting for potential bias in coefficients, we present SIMEX models rather than multilevel models. The OLS models, however, were run with robust standard errors, which have little impact on the magnitude of the standard errors. In addition, we ran random effects models with varying intercepts for legislative proposal. The difference between the random effects model and the simple OLS model is minimal, so we can be confident that ignoring the multilevel structure in the SIMEX models does not greatly affect our results.

To account for issue saliency, we first rescale the position data so the positions range from  $-10$  and  $10$ , and then weigh them so issues with low salience for an actor score near  $0$ , indicating indifference. Specifically, the rescaled positions are calculated as  $\frac{(pos_{orig}-50)*saliency}{500}$ . Outcome and reference variables are also rescaled to range from  $-10$  to  $10$  so they are on the same scale as the weighted position estimates. The Council position is calculated as the weighted average of member state preferences, where weighting is done by Council votes.

[INSERT TABLE 1 ABOUT HERE]

Table 1 presents five models. The first model regresses the outcome of negotiations on the positions of the EP, the Commission, and the Council, suppressing the constant, thus employing the same model used in the earlier Monte Carlo results. The second model reruns this model using a SIMEX regression. The third model includes a dummy for the decision-

making procedure (co-decision = 1) and interacts the dummy with position of the EP and the Commission to determine if the supranational actors are more influential under co-decision. The fourth and fifth models account for the position of the reference point by regressing the degree of change on each issue on the distance of the actors (Council, EP, and Commission) from the reference point. Model 4 presents an OLS regression and model 5 reruns model 4 as a SIMEX model.

The results of model 1 mirror the findings of others using the DEU data; the Council has significantly more influence than the EP or the Commission. The coefficient on the Council position is more than four times the magnitude of the EP and Commission coefficients. From the Monte Carlo results, though, we know that the model significantly underestimates the influence of the EP and Commission — the actors measured using a single point estimate only, rather than a weighted average. The second regression model presents the results of a SIMEX regression which corrects for measurement error in the Commission and EP variables. The model uses  $\sigma = 6$  as the standard deviation of measurement error, the approximate standard deviation of both the EP and Commission positions in the data. The results are robust to choosing different levels of measurement error. In this model, the coefficient on the Council variable remains identical to the previous model, but the Commission and EP coefficients both approximately double in magnitude. This model suggests that, taken together, the positions of the supranational actors explain a similar amount of variance in the outcome as the Council. The third model includes a control for issues decided under the co-decision procedure and interacts the co-decision variable with the EP and Commission preference variables. While theoretical models of EU legislative politics differ in how they discuss the various procedures and the influence they provide to actors, everyone agrees that the EP gained veto power under the co-decision procedure, and for the first time became a legislative co-equal with the Council. This model does not correct for measurement error as the error structure is more complex, involving both the position of the supranational actors and the interaction term. We can be confident, though, that the model underestimates the influence of the supranational actors. The aim of this model simply is to examine how co-decision impacts the influence of the supranational actors. The results show that



under co-decision, the EP has significantly more influence over legislative outcomes, and the Commission's influence remains virtually unchanged. Although the EP's influence remains one-third that of the Council's in this model, the earlier results have demonstrated that the model significantly understates its influence. Models 4 and 5 control for the position of the reference point. Here, we regress the distance of the outcome from the reference point on the distance of the actors' preferences from the reference point. Again, we see that, in model 4, which does not include any correction for measurement error, the Council distance better explains outcome change. However, once we correct for measurement error in model 5, the coefficient on the Council distance decreases in magnitude, while the coefficient on EP distance almost doubles.

## Conclusion

The DEU projects have undoubtedly added to our understanding of the EU legislative process, and they provide an invaluable source of information on actors' positions and salience on critical policy domains in the EU. However, researchers must carefully consider how best to use these data, taking into account the amount of information they contain. The data are unlikely to contain sufficient information to support many of claims in the literature regarding the strength of bargaining models in capturing EU decision-making, and which actors have greater bargaining power. Because of measurement error, theoretical models that involve taking a weighted average will always outperform models that do not when pitted against one another, regardless of which model actually represents the data generating process. Thus, it is unsurprising that many tests of competing bargaining models using the DEU data have found support for non-institutional models over institutional models, and that the Council is more influential. Given the simulations presented here, though, we would not want to conclude that the institutional models are wrong or that EP is powerless.

This is not to say that the DEU data are always ill-suited for testing bargaining models. Aksoy (2012) and Junge (2010) provide two recent examples of studies that use the DEU data to test the empirical implications of bargaining models without pitting a weighted

average variable against a point estimate. They both find support for institutional models of decision-making. Moreover, the data have been used in many other useful ways beyond tests of bargaining models. Thomson and Torenvlied (2010), for example, use the data as a measure of preferences to examine different perspectives on delegation to the Commission.

Future efforts to collect preference data in the EU must carefully consider measurement error and measurement models during the data collection process. Recent studies in the American literature that link responses from civil servants' internet survey responses to Congressional roll call votes provide a potentially fruitful guide to EU scholars (Clinton, Bertelli, Grose, Lewis and Nixon, 2012). These authors use a measurement model to generate truly comparable position estimates for parts of US government where roll call voting never occurs, namely executive branch agencies. Scholars of the EU face an identical problem — a lack of suitable data for measuring ideology across the branches of EU government. Until scholars are able to collect better measures, current analyses must carefully consider measurement error when presenting and interpreting results using the data we have.

### **Acknowledgements**

I would like to thank Sven-Oliver Proksch for helpful comments on an earlier version of this paper.

## Notes

<sup>1</sup>The DEUII data expand upon an earlier project — *Decision-making in the European Union* (DEU) (Thomson et al., 2006). The data collection process in the two projects was nearly identical, and the results presented in this paper apply to both datasets. Throughout the paper, the DEU abbreviation refers to both datasets. The data are publicly available and were downloaded from <http://www.robertthomson.info/research/resolving-controversy-in-the-eu>.

<sup>2</sup>See Thomson (2011) for a complete description of the data collection process, as well as a discussion of reliability and validity checks.

<sup>3</sup>On any single issue, players cannot occupy the same position. If random draws place both actors on the same location, +1 is added to  $B$ 's location.

<sup>4</sup>In fact, the Monte Carlo set up likely provides somewhat ‘cleaner’ data than those produced by the DEU project. In the DEU data, preferences tend to clump on certain points on the scale. Many of the issues are actually dichotomous, with actors’ preferences (along with the outcome and status quo) located at either 0 or 100. Assuming this clumping is an artifact of the data collection process, it makes testing procedural models even more difficult.

<sup>5</sup>Of course, this very simple game assumes that log-rolling across issues is not possible.

<sup>6</sup>In the simulations, these observed positions are rescaled back onto the 0-100 space, as the random draws may place some observed positions outside the 0–100 interval.

<sup>7</sup>Slapin (2008, 2011) employs a similar regression approach to testing competing models and actors’ bargaining power when examining negotiations at EU intergovernmental conferences.

<sup>8</sup>Indeed, because we consider each actors’ position as a normally distributed independent random variable, mathematically the Council’s position must be better measured than the EP’s. The variance of the mean of  $n$  normally distributed independent random variables is equal to  $Var(\frac{1}{n}X_1 + \frac{1}{n}X_2 + \dots + \frac{1}{n}X_n) = (\frac{1}{n})^2 Var(X_1) + (\frac{1}{n})^2 Var(X_2) + \dots + (\frac{1}{n})^2 Var(X_n)$ . As  $n$  increases, the

variance around the mean of these multiple distributions decreases. For example, imagine that the EP's position is normally distributed with a standard deviation equal to 8. Meanwhile, the Council is composed of two member states also with normally distributed independent positions, each with a standard deviation of 8 as well. If we measure the Council's position as the mean of the two states, its position will also be normally distributed, but its standard deviation will only be  $\sqrt{\left(\frac{1}{2}\right)^2 * 8^2 + \left(\frac{1}{2}\right)^2 * 8^2} = 5.66$ , significantly less than the standard deviation of the EP's position, even though the distributions of the EP and both member states had the same standard deviation.

<sup>9</sup>I wish to thank an anonymous reviewer for pointing out that the EP bargaining position may still be measured with greater accuracy than the Council bargaining position because experts were asked specifically to assess the EP bargaining position, whereas they were only asked to assess member state positions, and not a collective Council bargaining position. Instead, the Council bargaining position is estimated as a weighted average of the member state positions, thus making addition assumptions about Council bargaining works. Unfortunately, given the lack of uncertainty estimates, as well as the lack of positions for actors within the EP, there is no way to assess this statement. Moreover, given the number of member state positions being averaged to arrive at the Council position, the EP's position would need to be substantially better measured to overcome the math described above.

<sup>10</sup>Due to extreme missingness, we do not impute values for Bulgaria, Romania, or the EP groups.

## References

- Achen CH (2006) Evaluating political decision-making models. In: Thomson R, Stokman FN, Achen CH and König T (eds) *The European Union Decides*. Cambridge, UK: Cambridge University Press: 264–298.
- Aksoy D (2012) Institutional arrangements and logrolling: Evidence from the European Union. *American Journal of Political Science* 56(3): 538–552.
- Barry B (1980) Is it better to be powerful or lucky? Part 2. *Political Studies* 28(3): 338–352.
- Benoit K and Laver M (2012) The dimensionality of political space: Epistemological and methodological considerations. *European Union Politics* 13(2): 194–218.
- Binder S (1999) The dynamics of legislative gridlock, 1947–1996. *American Political Science Review* 93(3): 519–33.
- Blackwell M, Honaker J and King G (2012) Multiple overimputation: A unified approach to measurement error and missing data. Working Paper.
- Bueno de Mesquita B and Stokman F (eds) (1994) *European Community Decision Making: Models, Applications and Comparisons*. New Haven, CT: Yale University Press.
- Cameron CM (2000) *Veto Bargaining*. Cambridge, UK: Cambridge University Press.
- Carmines EG and Zeller R (1979) *Reliability and Validity Assessment*. Thousand Oaks, CA: Sage.
- Carrubba CJ, Gabel MJ, Murrah L, Clough R, Montgomery E and Schambach R (2006) Off the record: Unrecorded legislative votes, selection bias and roll-call vote analysis. *British Journal of Political Science* 36(4): 691–704.
- Clinton J, Bertelli A, Grose C, Lewis D and Nixon D (2012) Separated power in the United States: The ideology of agencies, presidents, and Congress. *American Journal of Political Science* 56(2): 341–354.

- Clinton J, Jackman SD and Rivers D (2004) The statistical analysis of roll call data. *American Political Science Review* 98(2): 355–70.
- Cook J and Stefanski L (1994) Simulation-extrapolation estimation in parametric measurement error models. *Journal of the American Statistical Association* 89(428): 1314–28.
- Crombez C (1996) Legislative procedures in the European Community. *British Journal of Political Science* 26(2): 199–228.
- Crombez C (1997) The co-decision procedure in the European Union. *Legislative Studies Quarterly* 22(1): 97–119.
- Golub J (2012) How the European Union does not work: National bargaining success in the Council of Ministers. *Journal of European Public Policy* 19(9): 1294–1315.
- Greene WH (2000) *Econometric Analysis*: Pearson Education, 4th edition.
- Häge FM and Kaeding M (2007) Reconsidering the European Parliament’s legislative influence: Formal vs. informal procedures. *Journal of European Integration* 29(3): 341–361.
- Hix S, Noury AG and Roland G (2007) *Democratic Politics in the European Parliament*. New York, NY: Cambridge University Press.
- Hix S (2002) Parliamentary behavior with two principals: Preferences, parties, and voting in the European Parliament. *American Journal of Political Science* 46(3): 688–98.
- Honaker J, King G and Blackwell M (2011) Amelia II: A program for missing data. *Journal of Statistical Software* 45(7): 1–47.
- Hooghe L (2005) Many roads lead to international norms, but few via international socialization. a case study of the European Commission. *International Organization* 59(4): 861–898.
- Jackman S (2008) Measurement. In: Box-Steffensmeier JM, Brady HE and Collier D (eds) *The Oxford Handbook of Political Methodology*: Oxford University Press: 119–151.

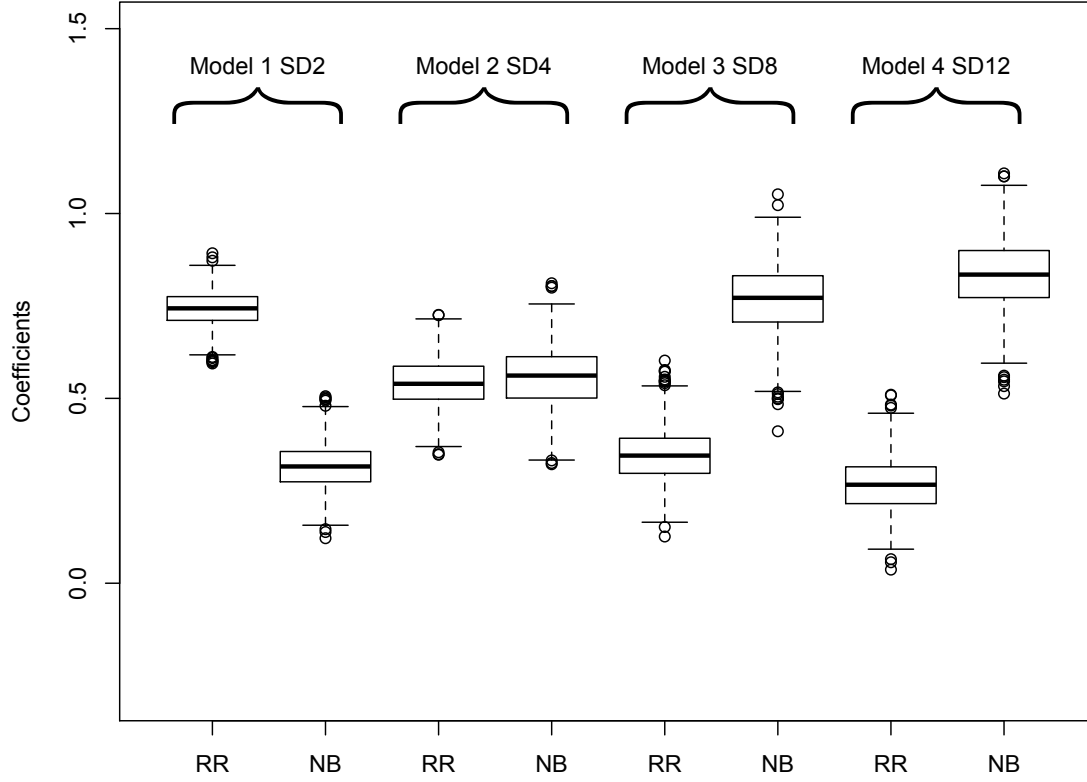
- Junge D (2010) Game theoretic models and the empirical analysis of EU policy making: Strategic interaction, collective decisions, and statistical inference. In: König T, Tsebelis G and Debus M (eds) *Reform Processes and Policy Change: Veto Players and Decision-Making in Modern Democracies*: Springer: 247–265.
- Junge D and König T (2007) What’s wrong with EU spatial analysis? The accuracy and robustness of empirical applications to the interpretation of the legislative process and the specification of preferences. *Journal of Theoretical Politics* 19(4): 465–487.
- Kasack C (2004) The legislative impact of the European Parliament under the revised co-decision procedure. *European Union Politics* 5(2).
- König T and Pöter M (2001) Examining the EU legislative process: The relative importance of agenda and veto power. *European Union Politics* 2(3): 329–351.
- Kreppel A (1999) What affects the European Parliament’s legislative influence? An analysis of the success of EP amendments. *Journal of Common Market Studies* 37(3): 521–38.
- Mattila M and Lane JE (2001) Why unanimity in the Council? A roll call analysis of Council voting. *European Union Politics* 2(1): 31–52.
- McCarty N and Shor B (2011) The ideological mapping of American legislatures. *American Political Science Review* 105(3): 530–551.
- McElroy G and Benoit K (2007) Party groups and policy positions in the European Parliament. *Party Politics* 13(1): 5–28.
- McElroy G and Benoit K (2012) Policy positioning in the European Parliament. *European Union Politics* 13(1): 150–167.
- Moser P (1996) The European Parliament as a conditional agenda setter: What are the conditions? a critique of Tsebelis (1994). *The American Political Science Review* 90(4): 834–838.
- Nash JF (1950) The bargaining problem. *Econometrica* 18(2): 155–62.

- Poole KT and Rosenthal H (1985) A spatial model for roll call analysis. *American Journal of Political Science* 29(2): 357–84.
- Proksch SO and Slapin JB (2010) Position taking in European Parliament speeches. *British Journal of Political Science* 40(3): 587–611.
- Romer T and Rosenthal H (1978) Political resource allocation, controlled agendas and the status quo. *Public Choice* 33(1): 27–43.
- Schneider G, Finke D and Bailer S (2010) Bargaining power in the European Union: An evaluation of competing game-theoretic models. *Political Studies* 58(1): 85–103.
- Selck TJ and Rhinard M (2005) *Pares Inter Pares* the bargaining success of the commission, the council, and the parliament in European Union legislative negotiations. *Swiss Political Science Review* 11(3): 123–140.
- Selck TJ and Steunenberg B (2004) Between power and luck: The European Parliament in the EU legislative process. *European Union Politics* 5(1).
- Shipan C (2004) Regulatory regimes, agency actions, and the conditional nature of political influence. *American Political Science Review* 98(2): 467–480.
- Slapin JB (2008) Bargaining power at Europe’s intergovernmental conferences: Testing institutional and intergovernmental theories. *International Organization* 62(1): 131–162.
- Slapin JB (2011) *Veto Power: Institutional Design in the European Union*. Ann Arbor, MI: University of Michigan Press.
- Thomson R (2011) *Resolving Controversy in the European Union: Legislative Decision-Making Before and After Enlargement*. Cambridge, UK: Cambridge University Press.
- Thomson R, Stokman F, Achen CH and König T (eds) (2006) *The European Union Decides*. Cambridge, UK: Cambridge University Press.



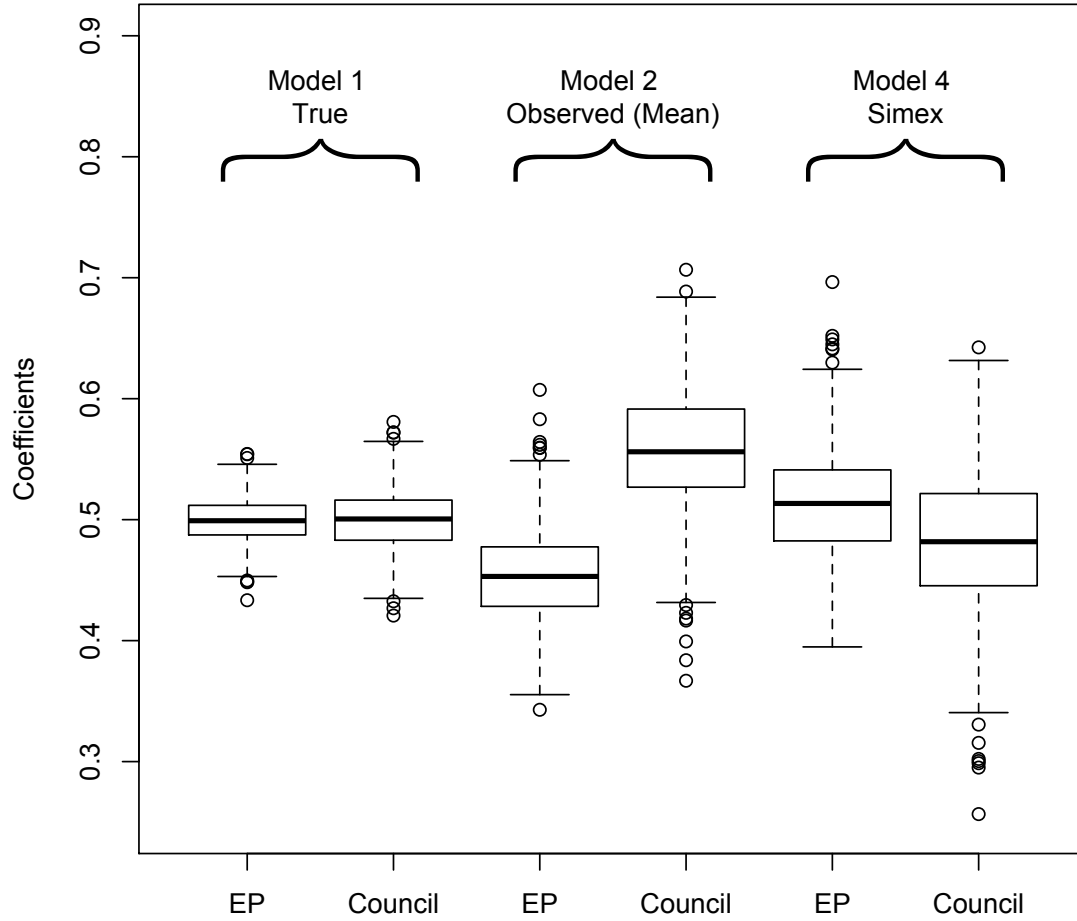
- Thomson R, Arregui J, Leuffen D, Costello R, Cross J, Hertz R and Jensen T (2012) A new dataset on decision-making in the European Union before and after the 2004 and 2007 enlargements (DEUII). *Journal of European Public Policy* 19(4): 604–622.
- Thomson R and Hosli M (2006) Who has power in the EU? The Commission, Council and Parliament in legislative decision-making. *Journal of Common Market Studies* 44(2).
- Thomson R and Torenvlied R (2010) Information, commitment and consensus: A comparison of three perspectives on delegation in the European Union. *British Journal of Political Science* 41(1): 139–159.
- Tsebelis G (1994) The power of the European Parliament as a conditional agenda setter. *The American Political Science Review* 88(1): 128–142.
- Tsebelis G and Garrett G (2000) Legislative politics in the European Union. *European Union Politics* 1(9): 9–36.
- Tsebelis G, Jensen CB, Kalandrakis A and Kreppel A (2001) Legislative procedures in the European Union: An empirical analysis. *British Journal of Political Science* 31.
- Tsebelis G and Kalandrakis A (1999) European Parliament and environmental legislation: The case of chemicals. *European Journal of Political Research* 36(1): 119–154.

Figure 1: Predictive Power of R-R Model and NBS with Varying Levels of Measurement Error.



*Note:* The boxplots are created from coefficient estimates of 1000 regressions. In each regression, the true bargaining outcome (constructed using the R-R model) is regressed on the prediction of the R-R model and the NBS model under various levels of measurement error in position estimates.

Figure 2: Power of EP and Council with Measurement Error.



*Note:* The boxplots are created from coefficient estimates of 1000 regressions. In each regression, the true bargaining outcome (constructed assuming equal power for the EP and Council) is regressed on the observed positions of the EP and Council, where the EP is measured using a single point estimate and the Council position is an average of member positions.

Table 1: Explaining DEUII Outcomes with Actor Preferences

	Model 1	Model 2	Model 3	Model 4	Model 5
	OLS – MI	SIMEX	OLS – MI	OLS – MI	SIMEX
EP	0.19*	0.44*	0.07		
	(0.06)	(0.08)	(0.11)		
Commission	0.17*	0.34*	0.17*		
	(0.05)	(0.06)	(0.08)		
Council	0.96*	0.99*	1.02*		
	(0.12)	(0.11)	(0.12)		
Codecision*EP			0.25*		
			(0.14)		
Codecision*Commission			0.03		
			(0.10)		
Codecision			-0.99*		
			(0.53)		
Distance Council				0.75*	0.65*
				(0.09)	(0.10)
Distance EP				0.15*	0.27*
				(0.06)	(0.09)
Distance Commission				0.14*	0.11*
				(0.06)	(0.05)
<i>N</i>	331	331	331	331	331

*Note:* Standard errors are reported in parentheses. \* indicates significance at  $p < 0.05$ . The dependent variable in models 1–3 is the outcome of negotiations. The dependent variable in models 4–5 is the absolute change in the outcome from the reference point.