Advance Access publication February 12, 2008 *Political Analysis* (2008) 16:235–249
doi:10.1093/pan/mpm038

Pass the Pork: Measuring Legislator Shares in Congress

Benjamin E. Lauderdale
Department of Politics, Princeton University, Corwin Hall, Princeton, NJ 08544
e-mail: blauderd@princeton.edu

Linear regression models are frequently used to analyze distributive politics in the U.S. Congress; however, authors have used a variety of specifications with different implicit assumptions about how bicameralism shapes legislative bargaining. I derive a model that describes district or state spending authorizations as the aggregation of spending secured by multiple legislators working on behalf of overlapping constituencies. This bicameral shares model allows the disaggregation of House and Senate influence through simultaneous estimation of the relative bargaining power of the two chambers and the advantages that accrue to legislators holding partisan, committee, and other relevant affiliations. In the 2005 transportation bill, the model better predicts the functional form of small state advantage than recently employed specifications in the literature.

1 Introduction

The empirical study of Congress's distributive bargaining is made more difficult by its bicameral structure. Even where it is possible to measure spending authorizations accurately at the level of congressional districts, we still must consider the influence of the district's two senators if we are to know which attributes of representatives yield greater bargaining power. In the more typical case of state-level data, the ecological inference problem of estimating the influence of individual representatives is even more difficult. In either case, two hurdles must be overcome simultaneously: modeling the bargaining power of individual representatives and modeling the process of aggregation over multiple representatives that leads to the observable outcomes.

Empirical work has usually recognized that bicameralism is important; however, the tools employed have generally been limited to either comparing House and Senate versions of the same legislation or adjusting for state size through the inclusion of various transformations of population in linear regressions. The first approach uses the different stages of bill development to simplify the inferential problem. Lee follows this approach in examining district-level data for the 1998 transportation bill, assigning each of the earmarks in the House version of the bill to a congressional district or districts (Lee 2003).

Author's note: Coding the earmark data by place and county was completed by Taxpayers for Common Sense, which released it to me for this project. I am especially indebted to Chris Laumann for his assistance in programming the algorithm for matching location names with congressional districts and to Frances Lee for sharing her transportation earmark data. Chris Achen, Doug Arnold, Brandice Canes-Wrone, Kosuke Imai, Nolan McCarty, Jasjeet Sekhon, Aaron Strauss, and several anonymous reviewers provided helpful suggestions.

She finds that committee membership correlates to higher funding receipt, whereas merit variables such as highway mileage have little influence on earmark spending. Knight (2004) follows Lee in eliminating the Senate problem by using only the House's initial draft of the 1991 and 1998 transportation earmarks to test a formal model similar to that of Baron and Ferejohn (1987). Hauk and Wacziarg (2007) compare the coefficient on their log(population) variable in the House and final versions of a bill to assess the consequences of Senate involvement in the legislative process. The central problem with this approach is that it focuses attention on an intermediate stage of a bargaining game, which may not reliably reveal the mechanisms driving final outcomes.

The second approach—modeling the final bargaining outcome—presents a different inferential problem. The overlapping representation of the House and Senate make it necessary to account for possible variation in distribution as a function of state size. Lee and Oppenheimer (1999) employ an index of per capita Senate representation (inverse population) to assess whether large states receive less distributive spending per capita. Using a similar functional form, Lee (2000) analyzes the change in the distribution of formula authorizations at the state level between the transportation bills of 1991 and 1998. She employs the reciprocal of state size as a model variable, but the change in authorization levels as the dependent variable instead of the raw total. Larcinese, Rizzo, and Testa (2006) remove the state size variation in their data by using state fixed effects in a time-series cross-sectional analysis. Hauk and Wacziarg (2007) directly test for a state size effect in the 2005 transportation earmarks aggregated at the state level, following the common approach of logarithmic transformation for modeling strictly positive quantities. Although all these specifications are at least descriptive of the direction and magnitude of trends, they implicitly assume different aggregation processes for individual legislator influence. The authors diverge from each other in their basic assumptions about how representation maps onto outcomes without discussion or justification.

This paper takes up one approach to explicitly deriving the functional form of a statistical specification for distributive data. I address the problem of aggregating legislator influence by treating each legislator as an independent actor attempting to get as much funding for his or her constituents as possible and the total observed authorizations for a state or district as simply the sum of the authorizations that each of its representatives was able to secure. These legislators each has bargaining power that depends on their position in the legislature: party, committee, and other relevant attributes. Although these assumptions are surely a simplification of legislators' motivations, they are similar to those used in some formal models of bicameral bargaining (Diermeier and Myerson 1999; McCarty 2000). Thus, although the bicameral share model introduced here does not test any particular formal model, it draws on similar assumptions to those used in legislative bargaining models. These assumptions turn out to reproduce the state size dependence assumed by Lee and Oppenheimer (1999), but with a different legislator covariate dependence.

After deriving the model, I use a case study of the earmarks in the 2005 transportation legislation to show that the model fits distributive spending data well and to contrast the results with the specification of Hauk and Wacziarg (2007) on the same data. I then expand the analysis to the overall authorizations in that legislation, showing that there is less small state advantage in the bill overall than in the earmarks. I use the contrast between these two classes of authorizations to show that the previously observed generosity to the Democratic minority in earmarks (Hauk and Wacziarg 2007) obscured the fact that overall spending authorized by the 2005 transportation bill preferentially rewarded areas of the country with more Republican representatives in the House.

2 Modeling Congressional Spending

The standard approach to dealing with a known predictor of outcome variation is to include that variable—in this case state size—in the regression. Lee and Oppenheimer (1999) do this by employing an index of per capita Senate representation to assess which states are being short changed in distributive spending because of their relative lack of Senate representation. Their model for state-level spending per capita G_i as a function of inverse state population $1/P_i$ and covariates $\mathbf{X_i}$ takes the form

$$G_i = \alpha_0 + \alpha_1 \frac{1}{P_i} + \beta \mathbf{X_i} + \varepsilon_i. \tag{1}$$

Rewritten as a model for total state spending, $S_i = G_i P_i$, the deterministic part of the model becomes a linear model in state size:

$$S_i = \alpha_0 P_i + \alpha_1 + \beta \mathbf{X_i} P_i. \tag{2}$$

This model implies a particular functional form associated with the small state advantage. A positive value of α_1 describes the funding level of a hypothetical state with zero population (two senators and zero representatives). Because it is linear in state size, the returns to additional House representatives are constant.[1] Unfortunately, the specification of Lee and Oppenheimer suffers from heteroskedastic errors. The authors note persistent unit-specific heteroskedasticity in their time-series cross-section and use panel-corrected standard errors (PCSEs) for their hypothesis testing. However, PCSEs do not change the ordinary least squares coefficients, and thus predictable, systematic variation in error magnitude as a function of state size is ignored in estimation.

Hauk and Wacziarg (2007) test for a state size effect in the 2005 transportation earmarks aggregated at the state level using the common approach of logarithmic transformation for modeling strictly positive quantities:

$$\log(G_i) = \alpha_0 + \alpha_1 \log(P_i) + \beta \mathbf{X_i} + \varepsilon_i. \tag{3}$$

$$S_i = e^{\alpha_0} P_i^{1+\alpha_1} e^{\beta \mathbf{X_i}}. \tag{4}$$

This functional form imposes the constraint that spending goes to zero for states with zero population as well as a declining marginal return to House delegation size if $\alpha_1 < 0$, the range of parameter values that correspond to a per capita small state advantage. Hauk and Wacziarg observe that a state size effect appears when the total amount of earmarks doubled from the House version to the final version of the 2005 transportation bill. The authors argue that it was Senate involvement that led to this per capita inequality. However, to induce a decline in the marginal return on population, the added spending after Senate involvement in the legislative process would have to be not merely biased toward small states in per capita terms but also in *absolute dollars*. Since the Senate has equal representation by state, there is no reason to expect such a dependence and there is certainly no reason to mandate it by functional form. Although taking logarithmic transformations solves the heteroskedasticity problem, it yields a problematic functional form. A better

[1]Using a similar functional form, Lee (2000) analyzes the change in the distribution of formula authorizations at the state level between the transportation bills of 1991 and 1998, employing the reciprocal of state size as a model variable, but the change in authorization levels as the dependent variable. Larcinese, Rizzo, and Testa (2006) use population untransformed in their regression analysis of total federal spending in each state as well as state fixed effects that make the population variable nearly irrelevant.

solution is to leave regression models behind and design an estimator appropriate for the particular features of bargaining in the U.S. Congress.

2.1 *A Specification for Bicameral Shares*

Formal models tend to predict that—proposal, veto, and voting power being equal—legislators in the same chamber should have an equal propensity (in expectation) to bring home federal dollars to their constituents, regardless of how many constituents they might have. Committee membership or chairmanship, partisanship, and other potential sources of additional bargaining power will modify this baseline of intra-chamber equality.

We can write down a simple model for how we might expect district shares to vary given that both representatives and senators will be working on behalf of the district, under a baseline assumption that intra-chamber bargaining power is equally distributed. Where D_i is the spending authorized to district i and N_i is the number of House representatives in that district's state, we can decompose spending outcomes in terms of a parameter for the relative bargaining power of the House and Senate, θ:

$$\frac{D_i}{\sum_{i'=1}^{435} D_i'} = \left[\theta \cdot \frac{1}{435} + (1-\theta) \cdot \frac{1}{N_i} \cdot \frac{2}{100}\right]. \tag{5}$$

Equation (5) simply says that the share of the total spending received by district i is equal to the sum of House and Senate contributions. The ith district's share is equal to their representative's contribution $\theta \cdot \frac{1}{435}$ plus a fraction of the two senators contribution $(1-\theta) \cdot \frac{2}{100}$ after it is divided across the N_i districts in the ith district's state. The relative size of the House and Senate contributions is determined by θ. If $\theta = 1$, the Senate contribution is zero and total spending is per capita equal across states. Smaller values of θ yield increasing small state advantage per capita.

As is, there is no variation among legislators within a chamber, so all that equation (5) does is infer the relative power of the two chambers from the distribution of spending as a function of state size. It is straightforward to introduce predictors of variation in intra-chamber bargaining power. We simply replace the fractions 1/435 and 2/100 with expressions for the fraction of spending received by individual legislators. In equation (6), both inter- and intra-chamber bargaining are modeled, with each member receiving a share of the chamber total that is defined by their personal covariates ($X_{i\mathrm{H}}$ for House members, $X_{i\mathrm{S}1}$ and $X_{i\mathrm{S}2}$ for each district's two senators, the corresponding β_H and β_S are the estimated coefficients associated with those covariates).

$$\frac{D_i}{\sum_{i'=1}^{435} D_i{}'} = \left[\theta \cdot \left(\frac{1 + X_{i\mathrm{H}}^\top \beta_\mathrm{H}}{\sum_{j=1}^{435}(1 + X_{j\mathrm{H}}^\top \beta_\mathrm{H})}\right)\right.$$
$$\left. +(1-\theta) \cdot \frac{1}{N_i} \cdot \left(\frac{2 + X_{i\mathrm{S}1}^\top \beta_\mathrm{S} + X_{i\mathrm{S}2}^\top \beta_\mathrm{S}}{\sum_{k=1}^{50}(2 + X_{k\mathrm{S}1}^\top \beta_\mathrm{S} + X_{k\mathrm{S}2}^\top \beta_\mathrm{S})}\right)\right]. \tag{6}$$

I consider share prediction models where the covariates are additive and multiplicative (additive on a log-scale). Equation (6) is for the additive relationship; the multiplicative model is generated by replacing each $1 + X^\top \beta$ with $\exp(X^\top \beta)$. Either model yields coefficients with straightforward interpretations. If a coefficient β_1 has the value 0.2, the interpretation in the additive case is that a legislator receives an additional 0.2 of the baseline share per unit of X_1, whereas in the multiplicative case it is that receives

$e^{0.2} = 1.22$ times as much as the baseline. So long as the substantive covariate effects are small compared to the baseline shares, the two specifications will yield similar coefficients and predictions. Only when the coefficients grow large will they diverge. In the data considered in this paper, the substantive results are nearly identical, so I present only the additive models.

In a linear regression model for spending or log(spending), coefficients are interpreted as marginal effects on the unit outcome: state or district spending authorizations. In the bicameral shares model, the legislator coefficients describe the influence of legislator attributes on a legislator's share of the total bill. By modeling a share for each legislator that can be predicted by their individual characteristics, it is possible to vary the particular covariates considered from case to case depending on data availability and the question of interest.[2] The model implies that any advantage that accrues to senators will be of the same magnitude regardless of their state's population.[3]

With shares as the dependent variable, the disturbances cannot be independent. The errors must add to zero exactly, not merely in expectation. Fortunately, with 50 or 435 units, the resulting interdependence is negligible. Several specifications of the disturbance term yield the same substantive results. A lognormal specification has the virtue of being strictly positive; however, because its mode and mean are different, it yields biased estimates in MLE. Alternatively, one can treat the individual legislator shares as normally distributed. We can then use the rules for addition and division of normally distributed random variables to create a heteroskedastic error function that depends on the number of representatives in a district's state and allows for different levels of legislator share variation in the two chambers:

$$\sigma^2_{\text{District}} = \sigma^2_{\text{H}} + 2\left(\frac{\sigma_{\text{S}}}{N_i}\right)^2. \tag{7}$$

Although this specification yields nonzero likelihood for shares below zero, in the data this is not a problem because the truncation covers little of the distribution. This approach addresses potential heteroskedasticity and the relative values of these parameters provide a diagnostic for whether the model is a good fit to the data. I have adopted the normal sum error term for the analysis in this paper.[4]

Both the model for inter-chamber shares in equation (5) and for inter- and intra-chamber shares in equation (6) can be aggregated up to the state level by summing over the districts in each state. This yields specifications for state-level data:

$$\frac{S_i}{\sum_{i'=1}^{50} S_{i'}} = \left[\theta \cdot \frac{N_i}{435} + (1-\theta)\cdot\frac{1}{50}\right]. \tag{8}$$

$$\frac{S_i}{\sum_{i'=1}^{50} S_i{}'} = \left[\theta\cdot\left(\frac{\sum^{N_i} 1 + X_{i\text{H}}^{\top}\beta_{\text{H}}}{\sum_{j=1}^{435} 1 + X_{j\text{H}}^{\top}\beta_{\text{H}}}\right) + (1-\theta)\cdot\left(\frac{2 + X_{i\text{S}1}^{\top}\beta_{\text{S}} + X_{i\text{S}2}^{\top}\beta_{\text{S}}}{\sum_{k=1}^{50} 2 + X_{k\text{S}1}^{\top}\beta_{\text{S}} + X_{k\text{S}2}^{\top}\beta_{\text{S}}}\right)\right]. \tag{9}$$

$$\sigma^2_{\text{State}} = N_i\sigma^2_{\text{H}} + 2\sigma^2_{\text{S}}. \tag{10}$$

[2]Using multiple bills, one could also model θ. Previous research suggests that different legislation will have different relative House and Senate influence (Lee and Oppenheimer 1999).

[3]Consequently, the return on committee membership (if positive) will be much more electorally potent for senators from small states. This in turn implies the empirical regularity that senators from small states preferentially seek out committees that distribute federal spending (Lee and Oppenheimer 1999).

[4]Using the lognormal specification does not change the substantive conclusions of the analysis in later sections.

These models must be applied with more care than the district-level models. Making claims about individual-level parameters for House representatives based on data aggregated up to the state level is subject to the same risk of aggregation bias that afflicts other ecological inference problems (Achen and Shively 1995). Moreover, with 50 data points instead of 435, any analysis with more than a few covariates is at risk of overfitting the data. I present this state-level specification for two reasons. The first is that the results of analyzing the earmark data in the 2005 transportation bill suggest that understanding the legislation requires examination of the total spending allocated in the legislation, which cannot be disaggregated to the district level. Similarly, if the model is to be useful for analyzing distributive spending generally—most of which is measurable only at the state level—it must work on state-level data.[5]

The second reason for examining the state-level version of the bicameral shares model is that it provides theoretical foundations for preferring linear regressions that follow the form of equation (1) rather than equation (3). The simpler form of the bicameral shares model in equation (5) predicts a peculiar set of equations for district-level spending as a function of state size that are positive linear combinations of a constant and the reciprocal of the number of congressional districts in the district's state. Aggregated up to the state level (equation (8)), these simply become positive linear combinations of a constant and the number of congressional districts in the state, which is approximately P_i, the population of the state. This provides microfoundations for the functional form used by Lee and Oppenheimer (1999) to describe the state size dependence of distributive spending. I consider a case study in detail here to provide evidence that this class of functional forms are better fits to the data than those used by Hauk and Wacziarg (2007).

3 Case Study: The 2005 Transportation Bill

I consider two lists of earmarked projects, those in the House version of the 2005 transportation bill "SAFETEA-LU"[6] (H.R. 3) as it was passed in March 2005 and those in the conference report ultimately passed by Congress and signed by President Bush in August 2005. Most transportation spending is not earmarked, so after considering the subset of the spending for which district-level analysis is possible, I analyze the state-level authorization levels for the bill as a whole.

Although Hauk and Wacziarg (2007) analyze the 2005 earmarks at the state level, it is possible to disaggregate transportation projects to the level of congressional districts, eliminating some of the risks associated with ecological inference. The earmarked projects listed in both the House bill (H.R. 3) and the final conference report were coded at the level of city, county, or state by Taxpayers for Common Sense, depending on the detail available in the text of the bill.[7] I use this information to link the data to U.S. Census lists that give the congressional districts of all incorporated places and counties. Where there was more than one congressional district for a place or county, or where only the state was listed, all congressional districts were credited with having received equal shares of the funding for

[5]Because the bicameral shares model simply describes a linear decomposition of spending into per capita and per-state equal components, the functional form does not change if one aggregates spending across multiple bills. Generally, if θ_i is the chamber split on the ith bill and the T_i total spending distributed on that bill, the total spending allocated per capita equally (the total House share) will be $\sum_{i=1}^{n} \theta_i T_i$ and the total money distributed $\sum_{i=1}^{n} T_i$. Thus, the value of θ for the total spending authorized across a collection of n bills will simply be the weighted sum $((\sum_i^n \theta_i T_i)/(\sum_i^n T_i)$.

[6]"Safe, Accountable, Flexible, Efficient Transportation Equity Act: A Legacy for Users."

[7]As is usual practice in Congress, the Senate bill listed no earmarks, so these are the only two earmark lists available.

Table 1 Success rates for matching of projects to multiple congressional districts (CDs)

Classification	*Place*	*County*	*State*	*1 CD*	*2*	*3–5*	*6–10*	*11+*	*Total*
H.R. 3	1808	664	1623	1633	517	543	415	987	4095
	0.44	0.16	0.40	0.40	0.13	0.13	0.10	0.24	
Final	2656	1012	2678	2712	834	863	682	1255	6346
	0.42	0.16	0.42	0.43	0.13	0.14	0.11	0.20	

Note. The left side of the table lists the matching specificity, whether the project was matched to an incorporated place's name, a county name, or only to a state. The right side of the table indicates the extent to which projects were assigned to multiple districts. The projects that were spread out over multiple districts will lead to underestimation of within-state variation.

that earmark.[8] Approximately two-fifths of all projects were classified as single-district projects and two-thirds were classified to five or fewer districts (Table 1). The classification success rates vary little across the two versions of the bill. District-level variation in spending will be underestimated because of the projects that are split evenly among a large number of districts.[9] Projects for the District of Columbia and for territories were excluded from the analysis. In addition to the spending data, I code the votes on the final version of the bill, the party, committee and conference affiliations and the tenure in office for the representative and two senators of each district.[10] Three House leaders' districts are given dummy variables because they received massive projects and risk biasing other coefficients: Don Young (R-AK; House Transportation Chairman), Dennis Hastert (R-IL; Speaker of the House), and Bill Thomas[11] (R-CA; House Appropriations Chairman).

In contrast to some previous studies, I have not included any variables related to the merit of the projects, such as physical size of states, total highway mileage, or mass transit systems. This exclusion of merit variables does not constitute a claim that Congress ignores merit. However, inclusion of both legislative influence variables and merit variables will generally lead to endogeneity bias.[12] The observed distributive outcomes in the legislation are the result of a bargaining process. This bargaining process determines the mapping of many possible merit variables into a final distribution. This mapping is politically malleable even if its components are fixed variables like highway miles, mass transit use, and weather. If merit variables were included in the analysis, they would necessarily have effects

[8]This is the procedure used by Lee (2003). Knight (2004) finds no difference in his results if he divides projects across districts as done here or omits them entirely. Since he is only using House projects—I am interested in Senate projects as well—dividing projects is more appropriate here. If I were to eliminate such projects, it would bias downward the total spending level in states with large cities, which include all the large states. This would bias all model parameters of interest.

[9]Although 20% of the projects were placed in 11 or more districts, this is partially due to the projects in cities and counties with a large number of districts. In these locations, projects can be expected to benefit multiple districts, so representatives may have stronger incentives to cooperate. Others are transit projects that benefit many districts and so no place or county information was coded.

[10]Christopher Cox (R-CA) was appointed Securities and Exchange Commission Chairman on June 2, 2005 and consequently did not vote on the reconciled bill. He is included in the data set, was appointed to the conference committee, and is considered to have abstained from the final vote.

[11]Both CA 20 and CA 22 are assigned to this dummy variable because Bakersfield, CA, is split between the two districts. Whereas Young's infamous "Bridge to Nowhere" was subsequently defunded, Thomas's district can still look forward to a new state route ($100 million), a widening of an existing highway ($60 million), a "Bakersfield Beltway System" ($140 million), and the most expensive project in the entire bill, a "Centennial Corridor Loop" ($330 million).

[12]I thank multiple reviewers for encouraging me to clarify my position on this point. Lee's (2003) research on the House version of the 1998 transportation earmarks finds no connection between earmark spending and merit variables; however, this result is not the basis for my omission of merit variables.

because the spending distribution is itself the result of a formula using these same variables. But these effects would be due to the endogenous relationship between the set of merit variables and bargaining power: the relative weights on the possible merit variables are themselves determined by bargaining. This is why the relative formula allocations to public transit and highways shifted toward the latter after the Republican takeover of Congress in 1994 (Lee 2000). Determining the extent to which Congress distributes federal resources meritoriously is an important research question, but it is not one that can be answered by putting merit variables and bargaining power variables into a single equation model.

3.1 *Inter-Chamber Earmark Bargaining*

The total earmark spending (excluding DC and territories) increased from \$12.5 billion in H.R. 3 to \$24.2 billion in the final legislation. As Fig. 1 indicates, there is a striking difference in the spending distribution by state size between the initial House bill and the final legislation.[13] In the first pair of columns in Table 2, I provide the results of fitting the bicameral shares model on H.R. 3 and the conference report versions of the legislation. In H.R. 3, $\hat{\theta} = 0.94$ (SE = 0.03), whereas in the conference report, $\hat{\theta} = 0.56$ (0.06).[14] For this analysis, I use dummy variables for the three influential House Republican leaders because two of the three are at extreme values of state size and risk biasing the results.[15]

The Hauk and Wacziarg model also recovers a shift in distribution between the two versions of the legislation, finding in their model without covariates that the coefficient α_1 on $\log(P_i)$ shifted from an insignificant -0.112 (0.119) to an "economically large" effect of -0.585 (0.081). Using the bicameral shares model, we can add some context to the shift. The House distributed a little over half the earmark total, dividing it approximately evenly across districts. The Senate distributed the remaining earmark dollars, dividing them approximately evenly across states.

Figure 1 suggests that the bicameral shares model is a better fit to the data than Hauk and Wacziarg model. I plot total earmarks by state and by district for both versions of the legislation. In the state plots, I show the state-level fit for the bicameral shares model without any dummy variables (equation (8)). In H.R. 3, the fitted state size curves for the Hauk and Wacziarg model and the bicameral shares model are similar; however, in the conference report, the two models differ noticeably. At both extremes of state size, the bicameral shares model predicts higher levels of earmarks. For small states, the bicameral shares model is closer to the data and is a more plausible trend line: state spending is not going to zero as state size goes to zero.[16] For large states, the comparison is more ambiguous, bicameral shares is closer to the New York (29) and California (53) points and further from Florida (25) and Texas (32). The case for the bicameral shares model is

[13]The shift in distribution and the approximate doubling ($12.5/24.2 = 0.52$) of the total earmark quantity between the two versions of the legislation might suggest a roughly equal split in earmark proposal rights between the House and Senate. However, H.R. 3 represents an intermediate stage in a bargaining game, so although the proximity of the apparent proposal shares to the estimated values of θ is intriguing, it does not by itself tell us much about the performance of the model.

[14]Unfortunately, the 1998 transportation earmark data compiled by Frances Lee is only for the initial House version of that legislation. Consequently it is useful only for showing that the H.R. 3 legislation was not historically exceptional. If a similar analysis is performed on the earmarks in the House version of the 1998 legislation, the estimate $\hat{\theta} = 0.95$ (0.02).

[15]Removing all three dummies and running the model described in equation (5) yields estimates of 0.85 (0.04) and 0.55 (0.08) with log-likelihoods of -7810.1 and -8201.1 on H.R. 3 and the conference report, respectively. Because Alaska is a one-representative state, the very large share received by Don Young in H.R. 3 leads to omitted variable bias in the estimated Senate influence.

[16]Fitting the bivariate regression model, $Si = \beta 0 + \beta 1Pi$ yields similar results. The fitted intercept $\hat{\beta}_0 =$ \$188.1 million (42.8) is significantly different from zero ($p = 6 \times 10^{-5}$).

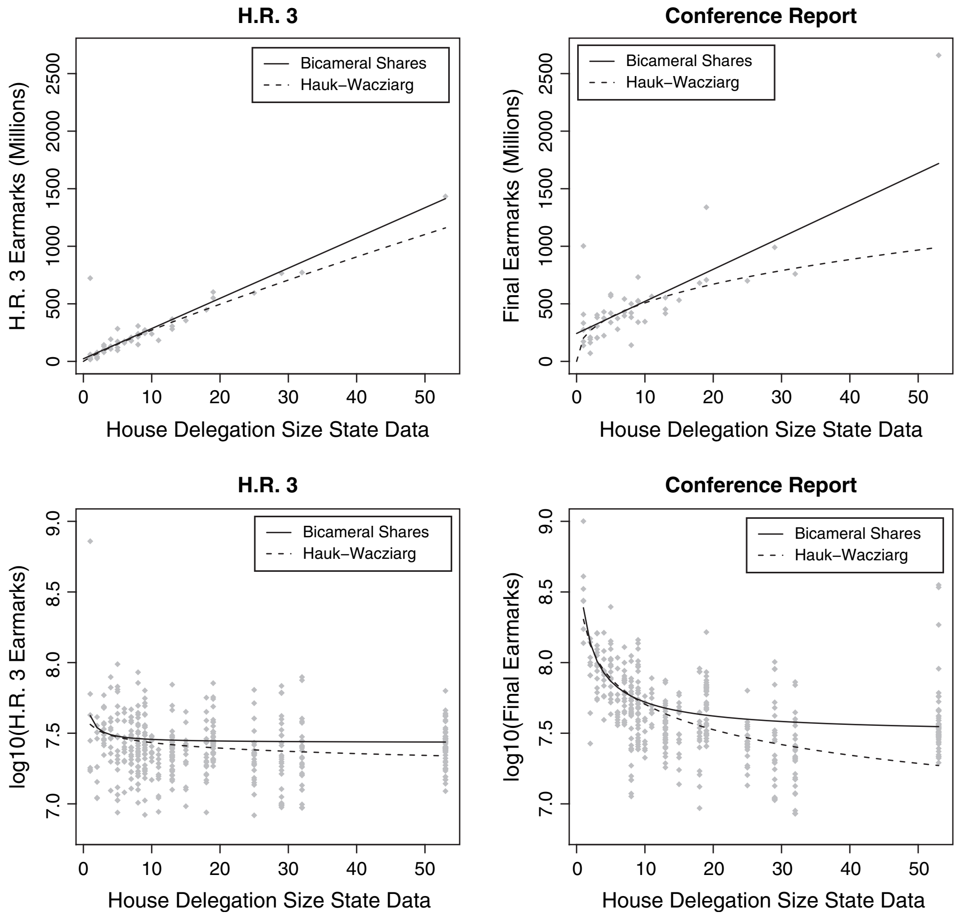

Fig. 1 Funding levels as a function of state size. The top pair of plots show that the relationship between state-level spending and state size is close to linear in state size. The initial House legislation shows little evidence of small state advantage (positive intercept), whereas there is a clearly positive intercept for the final conference report. The bottom pair of plots show district-level spending, again showing that the small state advantage appears only in the final legislation. Solid lines are predictions based on the bicameral share fit for the variable θ without legislator covariates. The bicameral shares specification is far more successful than the Hauk and Wacziarg specification in predicting California because it does not assume a declining marginal return on increasing House representation. The major outliers toward higher spending are states which had influential House members (AK, IL, CA). Outliers toward lower spending are states which had representatives and senators vote against the final version of the legislation (NH, AZ, TX).

stronger because the analysis of intra-chamber bargaining to follow shows that several of the Texas legislators may have opted out of the bargaining process because they were anti-pork. The residual on the Hauk-Wacziarg fit for California is \$1.67 billion, or about 7% of the total earmarks in the bill. The authors do not list California as one of their outliers (noting only Alaska and Arizona) because their model is on a log scale. Since trade-offs in distribution between states all occur on a linear scale, logarithmic specifications for studying Congressional spending may be generally problematic. The residual for the bicameral shares model is still quite large, \$0.85 billion. However, this is nearly accounted for by the projects in Bill Thomas district and the other district covering Bakersfield, CA, which total \$0.70 billion. Thus, in the district-level plots for the conference report, the bicameral shares model fits most California districts well, whereas the Hauk-Wacziarg model underpredicts all 53 districts.

3.2 *Intra-Chamber Earmark Bargaining*

The final two pairs of columns in Table 2 show bicameral shares model fits on the earmark data using district data with legislator covariates (equation (6)). In the first of these two models, I add covariates for being in the majority, the transportation committee, and the conference committee. In the analysis of H.R. 3, I add only the House covariates; in the conference report, I add both House and Senate covariates. The Senate covariates, by virtue of the large value of θ, can only explain a small amount of variation in the spending in H.R. 3. Therefore, they tend to take on apparently large but substantively and statistically insignificant values if included in the model.

In the second model, I add a variable for whether each legislator voted against the final conference report and a variable for tenure in office. The voting variable is a debatable inclusion. Chronologically, it appears that a low share of the earmarked spending must lead to a no vote and not vice versa. I suspect that this is not the causal mechanism though. Most of the eight representatives[17] and four senators[18] who voted against the final bill are legislators who campaign against government waste. Therefore, the primary causal mechanism driving the relationship between their earmark receipts and their votes is likely to be that these legislators did not request earmarks because they disapprove of earmark spending and voted against the bill for the same reason. Thus, the vote variable is acting as a proxy for legislators who opted out of the distributive game. Because all these legislators are Republicans and few were on the relevant committees, not including this variable may lead to omitted variable bias in the estimation of committee and party effects. The author prefers the specification with the vote variable; however, readers who believe that potential endogeneity is a greater concern should discount the last two columns of the table as they deem appropriate.

Aside from the individual legislator dummy variables, there are two variables that are significant and substantively large across all specifications. As Hauk and Wacziarg observe, Republicans did worse than Democrats in both chambers, significantly so in the House. Even accounting for the fact that all the anti-pork legislators proxied by final vote were Republicans, this effect is significant in the House. The lack of a pro-majority effect is consistent with the general finding that pork is bipartisan (Evans 2004), but the prominority bias in this bill is more difficult to explain (Hauk and Wacziarg 2007). In the next section, I apply the bicameral shares model to the overall spending in the bill to offer an explanation of why the majority might have secured fewer earmark dollars. The other consistent effect is that Senate conferees did extremely well, receiving 54% more earmarks than the baseline Senator. Senate earmarks have traditionally been omitted from bills voted on the floor and are instead added in conference, which appears to give the conference committee members extra influence.

Some of the other significant effects vary across versions of the bill or specifications of the model. The magnitude of the House committee membership advantage is much smaller once the Senate becomes involved in the legislative process. It is likely that the House

[17] Shadegg (R-AZ), Flake (R-AZ), Royce (R-CA), Jones (R-NC), Boehner (R-OH), Thornberry (R-TX), Hensarling (R-TX), and Sensenbrenner (R-WI).

[18] Kyl (R-AZ), McCain (R-AZ), Gregg (R-NH), and Cornyn (R-TX).

Table 2 Comparison of bicameral share model fit for H.R. 3 with the final legislation

	H.R. 3		*Final*		*H.R. 3*		*Final*		*H.R. 3*		*Final*	
Bill Version	*Estimate*	*SE*	*Estimate*	*SE*	*Estimate*	*SE*	*Estimate*	*SE*	*Estimate*	*SE*	*Estimate*	*SE*
House Majority					**−0.16**	0.04	**−0.25**	0.07	**−0.16**	0.05	**−0.26**	0.09
House Committee					**0.50**	0.07	0.12	0.12	**0.62**	0.11	0.21	0.16
House Conference					**0.23**	0.08	**0.27**	0.13	**0.21**	0.09	0.23	0.16
House Voted No.									−0.29	0.20	−0.33	0.35
House log(Tenure)									**0.07**	0.03	0.12	0.07
Senate Majority							−0.16	0.13			−0.01	0.13
Senate Committee							−0.17	0.23			0.12	0.22
Senate Conference							**0.83**	0.31			**0.54**	0.25
Senate Voted No.											**−1.19**	0.45
Senate log(Tenure)											−0.08	0.05
Don Young	**26.75**	1.10	**27.37**	4.81	**26.69**	0.98	**22.01**	4.49	**30.88**	2.39	**27.46**	6.27
Bill Thomas	−0.07	0.36	**11.46**	0.96	−0.08	0.34	**10.60**	0.93	−0.04	0.39	**13.59**	2.21
Dennis Hastert	**1.77**	0.52	**4.54**	0.87	**1.98**	0.48	**4.31**	0.78	**2.22**	0.57	**5.36**	1.19
θ (House Share)	**0.94**	0.03	**0.56**	0.06	**0.95**	0.02	**0.56**	0.07	**0.95**	0.02	**0.55**	0.06
$\sigma_H/10^6$	13.0		21.0		11.9		20.1		11.7		20.1	
$\sigma_S/10^6$	13.0		88.6		5.5		83.2		6.7		73.7	
log(L)	−7755.3		−8039.9		−7706.0		−8019.0		−7702.0		−8006.7	
n	435		435		435		435		435		435	

Note. Senate conferees did extremely well in the final legislation, whereas members of the House Republican majority did unexpectedly poorly. Even after eliminating anti-pork Republicans (proxied by a no vote on the final legislation), the majority received less earmarked money than the minority. Entries significant at the 95% level or higher are in bold.

Table 3 Total state-level spending fits under the bicameral shares model

	Estimate	*SE*	*Estimate*	*SE*
House—Majority			**0.70**	0.30
House—Committee			−0.08	0.37
House—Conference			0.38	0.35
Senate—Majority			0.59	0.63
Senate—Committee			0.47	0.81
Senate—Conference			1.22	0.96
House Share	**0.79**	0.03	**0.81**	0.03
$\sigma_H/10^6$	37.8		31.1	
$\sigma_S/10^6$	46.6		36.5	
Log-likelihood	−999.4		−988.7	
n	50		50	

Note. In contrast to the distribution of earmarked spending, states with more members of the House Republican majority received far more transportation spending overall. Entries significant at the 95% level or higher are in bold.

conference effect is significant because it proxies for more influential members of the House and the transportation committee. As anticipated, the vote proxy for anti-pork legislators weakens many of the committee and conference effects. The size of the coefficients on the vote variables are reasonable if one views that variable as a proxy for dropping out of the earmark bargaining game. In the Senate, the cost is about one senator share, whereas in the House the cost is about a third of a representative's share. The House cost to opting out is underestimated because the coding of projects underestimates cross-district variation in multidistrict states. The effect of opting out of the legislation is visible in the raw data in Fig. 1, especially in the three states that had senators oppose the legislation. Texas—which had one senator and two representatives vote against the bill—is visibly below the trend at 32 House seats. New Hampshire had one senator opposed to the bill, and received the least earmarked dollars of any state despite having two representatives rather than one. Arizona had *both* Senators and two of its eight representatives opposed to the bill, and is clearly visible as an outlier in the data plots.[19]

3.3 *Bargaining over Total Spending*

Although the earmark spending data has the advantage of allowing analysis at the level of congressional districts, over 90% of the spending in the transportation bill was allotted by formula to state governments rather than earmarked for particular projects. In Table 3, I fit the state-level version of the bicameral shares model on the total spending in the transportation bill.[20] The House share is much higher, 0.79 or 0.81 depending on covariate specification, indicating a more egalitarian per capita apportionment in overall spending

[19]Hauk and Wacziarg (2007) observe that Arizona is an outlier, but give no explanation for its status. Similarly, they observe that Alaska is an outlier toward higher levels of spending. Although the authors call attention to the fact that all three legislators from the state were on the conference committee, 70% of Alaska's earmarks were already in place in H.R. 3. In fact, Alaska is a more severe outlier in that version of the legislation because there is so little variation across other districts. Thus, Don Young's position as chairman of the originating committee seems a more likely explanation for Alaska's windfall.

[20]Unfortunately, a comparison of the state-by-state distribution of overall expenditures with that proposed by H.R. 3 before Senate consideration was impossible. Estimates of the state-level expected funding for the intermediate version of the bill are not available from the Department of Transportation, which supplied the estimates for the final bill.

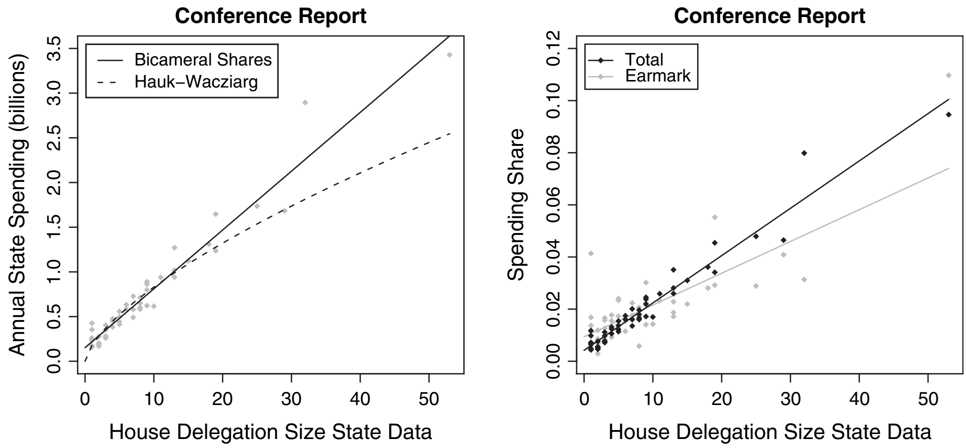

Fig. 2 At left, the Hauk-Wacziarg model underpredicts the spending levels of the smallest and largest states in overall spending as it does in earmarks. At right, the different state-level spending distributions of the earmark and overall spending are shown. Overall spending is closer to per capita egalitarian than earmarked spending, but still exhibits a small state advantage.

than in earmarks. The only strongly significant predictor for overall spending is being in the House majority, which is associated with an increase in legislator share of 80%. Since Republicans disproportionately represent areas that benefit from highway spending, this is consistent with the previously studied shift from public transit toward highways in U.S. transportation spending after 1994 (Lee 2000).

The Republican advantage in formula spending overwhelms the minor Democratic advantage in the earmarks. Why would the majority grant such favorable treatment on the most visible portion of the legislation? Perhaps the goal was to make it difficult for the minority to use the strongly pro-Republican slant in overall spending as a campaign issue. Balla et al. (2002) argue that the reason why minority parties are included in pork distribution is that the appearance of nonpartisanship inoculates the majority against criticism for taking the most valuable awards. Here, the majority was willing to forgo the visibility of earmarked projects for the far greater monetary rewards available in the formula spending. Less charitably put, the minority appears to have been bought off with earmarks.

The left panel of Fig. 2 shows again that the bicameral shares model is a better fit than the specification used by Hauk and Wacziarg. As before, the constraint that the latter model must pass through zero leads it to underestimate small states' shares.[21] The difference is clearer in the largest states: the assumption of declining marginal return on additional House representatives leads to an underestimate of the annual transportation allocation to Texas by $1.08 billion and to California by $0.88 billion.

The final remaining question is why the division between the House and Senate is so different (right panel of Fig. 2) in earmark and overall spending. Why did the Senate have more proposal power in the earmarked spending than in the bill as a whole? This result contradicts the prediction that House members have more to gain from projects than senators because they provide a rare opportunity to visibly bring home benefits to their district (Lee 2003). More broadly though, it is consistent with previous work that finds

[21]Following the same analysis as in footnote 16, the intercept of a bivariate regression fit of spending on House delegation size is significantly nonzero and positive: $154.3 million (28.4).

differences in the size of small state advantage in discretionary and nondiscretionary distributive spending (Lee and Oppenheimer 1999). Nondiscretionary programs are observed to have a greater small state advantage because Congress directly controls spending distribution, whereas in discretionary programs it yields some control to the executive branch. This logic might hold within the transportation bill: there is more fine grained control over the distribution of earmarks than over the formula.

4 Conclusion

In this paper, I provide a statistical model that allows estimation of the relative bargaining power between and within the chambers of the U.S. Congress. This allows us to tell a more complete story of how the structure of bicameral representation generates the observable data of distributive politics. The set of models developed here can be used to describe distributive outcomes in any bills where it is possible to measure funding at either the state or the district level. When compared to the linear models that have been widely used in analysis of congressional spending authorizations, the bicameral shares model provide a more theoretically motivated specification, transparent parameter interpretation, and a better fit to the data. That the model predicts two of the basic features of the small state advantage—positive intercept and constant returns to increasing population—suggests that the cross-legislator independence assumption and the distinction between inter- and intra-chamber bargaining are useful ways of approximating the Congressional bargaining process.

The introduction of the parameter θ to describe the relative influence of the House and Senate provides a convenient and comparable measure for characterizing the degree of per capita malapportionment. In SAFETEA-LU, expected earmarks for single-district states increased from about \$55 million over 5 years under per capita equality ($\theta = 1$) to \$250 million for the observed $\theta \approx 0.55$. Overall spending for the smallest states increased from \$0.42 billion over 5 years to \$1.06 billion ($\theta \approx 0.80$). The existence of the Senate was worth over \$200 per capita per year to the citizens of the smallest states, *on one bill*.

Some authors may still wish to use linear models to analyze distribution data for initial exploratory data analysis or for other reasons. The preceding analysis offers some guidance about how to set up such a model. Legislative influence variables and merit variables are endogenous and should not be included in the same model. One can then model total spending as a linear function of state size as in Lee and Oppenheimer (1999), which will get the functional form of the state size effect right and get the disturbance distribution wrong. Alternatively, one can get the functional form wrong and fix heteroskedasticity of the errors by taking logarithmic transformations as in Hauk and Wacziarg (2007). Since incorrectly specified disturbances do not necessarily preclude consistent parameter estimates, getting the disturbance wrong will almost always be preferred. Thus, models for total state-level spending should include state size linearly. Models for state-level per capita spending and for district-level spending should include the reciprocal of state size.

Moving in the direction of better theorized statistical specifications for distributive politics requires making clear assumptions about how legislator influence is aggregated. This paper has described one model that results from applying some very simple additivity and normality assumptions. At best, these assumptions are approximations. They are unlikely to be the best ones. Other specifications are possible and might be indicated by either empirical or formal research. Other statistical approaches may also be helpful, such as Markov chain Monte Carlo sampling approaches for simulating individual legislator shares conditional on the observed distribution of authorizations across units. Whichever of these methodological innovations prove to be useful, they will need to be responsive to the substance of distributive politics. Application of standard regression solutions such as logarithmic transformations may be able to shed light on general trends; however, without an assessment of the substantive constraints that are being imposed, such modeling may obscure rather than illuminate the processes that generate distributive outcomes.

References

Achen, Christopher H., and W. Phillips Shively. 1995. *Cross-level inference.* Chicago, IL: University of Chicago Press.

Balla, Steven J., Eric D. Lawrence, Forrest Maltzman, and Lee Sigelman. 2002. Partisanship, blame avoidance, and the distribution of legislative pork. *American Journal of Political Science* 46:515–25.

Baron, David P., and John F. Ferejohn. 1987. Bargaining and agenda formation in legislatures. *The American Economic Review* 77:303–9.

Diermeier, Daniel, and Roger B. Myerson. 1999. Bicameralism and its consequences for the internal organization of legislatures. *The American Economic Review* 89:1182–96.

Evans, Diana. 2004. *Greasing the wheels: Using pork barrel projects to build majority coalitions in Congress.* Cambridge: Cambridge University Press.

Hauk, William R., and Romain Wacziarg. 2007. Small states, big pork. *Quarterly Journal of Political Science* 2:95–106.

Knight, Brian G. 2004. Parochial interests and the centralized provision of local public goods: Evidence from congressional voting on transportation projects. *Journal of Public Economics* 88:845–66.

Larcinese, Valentino, Leonzio Rizzo, and Cecilia Testa. 2006. Allocating the US federal budget to the states: The impact of the president. *Journal of Politics* 68:447.

Lee, Frances E. 2000. Senate representation and coalition building in distributional politics. *American Political Science Review* 94:59–72.

Lee, Frances E. 2003. Geographic politics in the U.S. house of representatives: Coalition building and distribution of benefits. *American Journal of Political Science*. 47:714–28.

Lee, Frances E., and Bruce I. Oppenheimer. 1999. *Sizing up the Senate: The unequal consequences of equal representation.* Chicago, IL: University of Chicago Press.

McCarty, Nolan. 2000. Proposal rights, veto rights, and political bargaining. *American Journal of Political Science* 44:506–22.