

Dividing the Dollar With Formulas*

Gregory J. Martin[†]

March 6, 2017

Abstract

In advanced democracies, most government spending is allocated according to criteria approved by a legislature but implemented by the bureaucracy. I ask whether this fact imposes a binding constraint on the ability of legislators to engage in targeted redistribution, by constructing a model in which legislators are constrained to allocate spending by a formula of limited dimension - in contrast to benchmark models where proposers have the flexibility to manipulate the payoffs of individual members directly. The model predicts over-sized winning coalitions, positive distributions outside of the winning coalition, and the emergence of persistent voting blocs. I then apply the model to a sample of 31 US federal spending bills, using new data connecting spending outcomes to authorizing legislation. I find that most allocation formulas for spending programs involve 5 or fewer factors. Formulaic allocation imposes a tight constraint on targeting, eliminating more than 90% of Congressional proposers' degrees of freedom.

*The author acknowledges helpful comments from Steve Callander, Cliff Carrubba, Alan Wiseman, Dave Baron, Tasos Kalandrakis, and Claire Lim.

[†]Emory University, 1555 Dickey Dr., Atlanta, GA 30322. Send correspondence to: gregory.martin@emory.edu.

In advanced democracies, legislatures typically do not directly determine the distribution of government spending to the geographical sub-units their members represent. Instead, legislatures enact a set of criteria (a “formula”) that instruct the bureaucracy as to the way in which monies appropriated for a particular program ought to be spent. The elements of this set of criteria, rather than the ultimate distribution of funds, are the objects over which legislative bargaining and debate take place. While appropriations of funds directed to a particular district – “earmarks” – do occur, in practice they account for only a tiny fraction of government spending.

Canonical political economy models of legislative bargaining (Baron and Ferejohn 1989) and vote-buying (Snyder 1991) ignore this distinction between legislative proposal and ultimate outcome, effectively taking the earmark method of appropriations to be the rule rather than the exception. These models endow legislative proposers with the ability to directly manipulate the ultimate distribution to an individual legislator. In the US House of Representatives, for instance, this equates to a proposal with 435 dimensions, one for each legislator. In contrast, real-world legislative proposers manipulate the vector of spending outcomes only indirectly, by altering the relative weights of different factors in the criteria that bureaucrats use to award grants.

The objective of this paper is to show that this modeling abstraction misses an important feature of distributive politics: formulaic spending imposes a binding constraint on the ability of legislatures to redistribute. I present both theoretical and empirical evidence to support this claim. First, I extend the “divide the dollar” game due to Baron and Ferejohn (1989), and later generalized by Banks and Duggan (2000), to incorporate dimensionality limits on the bargaining space and extract testable predictions. Second, I take the model to a new data set of 31 U.S. spending bills and estimate its parameters. I recover an estimate of the degree to which formulaic allocation constrains distributive coalition formation, and show that this constraint is quite restrictive in

practice.

I model the formula constraint as imposing a limit on the dimension of legislative proposals. That is, proposers do not have complete freedom to directly specify the payoff to each district that will result if their proposal passes. They must condition spending allocations on a relatively low-dimensional set of observable attributes of the districts, rather than the identity of a particular district. This feature implies that changing the payoff to one member may require altering the payoffs of other members: buying one legislator often will bring along another from a district with similar characteristics that will also benefit from the formulaic appropriation. I extend the divide-the-dollar game to allow for these constraints and solve for the no-delay equilibrium proposals. Although the formulaic constraint complicates the legislative problem, both for the proposer and the rest of the chamber, the equilibrium exhibits three aggregate properties that are congruent with empirical observation but that are not shared by the standard model: oversized winning coalitions, positive distributions to members outside of the winning coalition, and persistent voting blocs of similar districts.

These properties have consequences for fundamental features of legislative politics. The endogenous generation of persistent voting blocs dramatically reduces the set of possible coalitions, and provides a natural focal point for the formation of legislative parties. The fact that benefits will often flow to members outside of the winning coalition provides flexibility for legislators to engage in position-taking. No votes without denying their constituents access to valuable programs. And the tendency for formulas to generate large coalitions produces more universalistic policies that benefit broader collections of districts.

Taking the model to the data, I construct a novel dataset of 31 US federal appropriations bills, which connects the program authorizing a particular federal grant or direct payment back to the

roll-call vote on the bill that appropriated funds for the program. Among programs in the sample with explicit statutory formulas, the median number of factors referenced in the formula is just 3; 90% have 5 or fewer factors. However, many programs do not have an explicit statutory formula, and appropriations bills often bundle together funds for multiple programs. I use the model to derive necessary conditions on the properties of winning coalitions and spending distributions in equilibrium, and apply these conditions to empirically recover an estimate of the effective dimension of the formula space in the U.S. House. The estimate is 42, rather than the possible 435, implying that formulaic spending procedures dramatically restrain legislators' ability to precisely target funds to marginal districts, eliminating more than 90% of proposers' degrees of freedom.

In this paper I take the existence of a dimensionality limit as given and exogenous, already determined by a prior stage of the policymaking game. The goal is to work out the implications of such a constraint for distributive politics in legislatures, and then use those implications to infer the degree of the constraint in the empirical data. I do not, however, impose any upper bound on the dimensionality parameter in the estimation; the fact that the estimated dimension is much less than the number of legislators suggests that the formula constraint is binding. Throughout the paper I adopt the interpretation that the constraint is due to the necessity of policies passing through the bureaucracy. In the discussion section I motivate this restriction more extensively, and sketch several extensions of the model that would deliver low-dimensional formulas in equilibrium. To provide better intuition for how these restrictions work in practice and why they might arise, consider the following classic example described extensively by Joskow and Schmalensee (1998).

A motivating example In 1990, members of Congress debated a set of amendments to the Clean Air Act that created a system of tradeable emissions permits for sulfur dioxide (SO_2) which could be bought and sold by electric utilities. A contentious issue surrounding the creation of this system

was that the allowances were to be given away free to utilities, meaning that Congress needed to determine the initial distribution of what was expected to be a valuable property right.

The rules that Congress ultimately adopted allocated allowances to generating units on the basis of the unit's characteristics - its fuel type, historical energy generation, and existing emissions-control devices, among other things. Though the final rules were lengthy and complex, the total distribution of allowances to each state did not differ much from what would have occurred under the very simple benchmark rule initially proposed by the EPA.¹ Where there were deviations, they often did not conform to expectations about the congressional bargaining process. Some states, like Utah, benefited from significant gains relative to the benchmark scenario despite both lacking any representation on the relevant committees or party leadership and voting No on final passage.

Utah's unexpected success in the bargaining process for SO₂ allowance allocations presents a puzzle for standard models. It clearly had nothing to do with holding a powerful committee position, and given that the Utah delegation voted No on a bill that passed by a wide margin, it cannot be that the majority "bought" the Utah vote by promising more allowances for Utah utilities. The possibility that I investigate in this paper is that Utah benefited from an unintended consequence of the formulaic allocation procedure: benefits that the Democratic leadership intended to target at swing states in Democratic but coal-heavy Appalachia also ended up flowing to similarly coal-heavy but Republican and anti-regulation Utah.

Related literature The classic models of legislative bargaining, on which this and many other papers build, are the "divide the dollar" game of Baron and Ferejohn (1989) and its extension to general social choice settings by Banks and Duggan (2000). In their setup, an exogenously

¹The EPA's "Basic Rule" would have determined the allocation by multiplying a unit's historical emissions of SO₂ by a constant. Joskow and Schmalensee's Table 2 shows that the Phase I allowance allocations by state generally fell within $\pm 5\%$ of the Basic Rule.

determined budget must be distributed amongst the members of an n -member legislature, which requires majority approval to pass an allocation. A fundamental result of this model is that in equilibrium all proposals garner the support of a minimal winning coalition (e.g., they pass by a bare majority), and the members of this coalition capture the entire budget.

Models of vote-buying, the classic example of which is Snyder (1991), share a similar structure and produce similar results. In these models, a legislative agenda-setter attempts to pass a bill by distributing side-payments to recalcitrant members in order to “buy” their votes. The vote-buying structure replaces the endogenous continuation value in bargaining games with an exogenous spatial component, but in other respects the predicted behavior is nearly identical. In equilibrium, the leader targets payments to the most marginal districts, spending only what is necessary to pass the bill with a minimal winning coalition; only members of the winning coalition receive positive payments.²

Empirical work on distributive politics has, however, consistently found these stark predictions to be violated in real data. To cite but one example, Knight (2005) examines a 1998 transportation bill in the US House on which 337 of the 435 representatives voted in favor, and 357 districts received funding for projects authorized by the bill. All available evidence indicates that situations like this one are the norm, rather than the exception.

This empirical regularity has been recognized for some time, and has spawned a variety of theoretical explanations. An early literature³ argued that a norm of universalism within Congress governed members’ behavior on distributive bills. While consonant with the empirical patterns, this approach was criticized for lacking a rational choice micro-foundation. As we will see, the

²While I pursue a direct extension of the Baron-Ferejohn bargaining model in the main text, the results can be equally well applied to vote buying, a connection which I elaborate in Appendix B.

³See, e.g. Niou and Ordeshook (1985), Stein and Bickers (1994), Weingast (1994).

present paper rationalizes several empirical patterns often attributed to norms of universalism in a legislative bargaining framework.

Groseclose and Snyder (1996) propose a variant of the basic vote-buying model where two competing leaders, one supporting and one opposing a bill, attempt to achieve their favored outcome on a vote by distributing side-payments. While this model provides a parsimonious explanation for supermajority coalitions, it retains the prediction of earlier work that side-payments will be targeted exclusively to the most marginal members of the ultimate winning coalition.

An alternative approach, due to Weingast and Marshall (1988), Carrubba and Volden (2000), and others, emphasizes the role of long-term agreements between members of Congress that are built into the structure of institutions such as the committee system. In this explanation, known as logrolling or vote-trading, members agree to cede control over one area of the budget in exchange for control over another.⁴ One problem with this approach is that, like all models of repeated games, there are many possible equilibria. The logrolling equilibrium relies on an unspecified mechanism for members to commit to the norm of deference to committees. Furthermore, there is no reason that distributive benefits should flow to members outside the winning coalition; if anything, members who vote against appropriations bills ought to be punished for their failure to go along with the cooperative strategy.

An equally large literature examines the participation and influence of the executive branch in distributive policymaking. Berry, Burden, and Howell (2010b) find that districts represented by members of the president's party receive higher levels of federal outlays. Levitt and Snyder (1997) compare the distribution of spending allocated by statutory formula to that of programs left to

⁴For example, rural representatives on the agriculture committee propose a generous farm-subsidies bill, urban representatives on the transportation committee propose generous federal funding for mass transit projects, and both bills pass with a broad urban-rural coalition.

the discretion of the executive branch, and show that the former are more responsive to changes in partisan control of Congress. And Stein and Bickers (1997) study the substantial durability of federal distributive programs, noting the long shadow of influence of “policy subsystems” within the bureaucracy on outlays.⁵

This article seeks to join the insights from these two strands of the literature on distributive politics, which at present are mostly distinct. Recent theories of legislative bargaining focus on the structure of institutions within the legislature to explain the tendency towards oversized coalitions on appropriations bills. I instead pursue the fact that policies, once passed, must be implemented. The basic, static Baron-Ferejohn and Snyder (1991) models get the incentives of legislators approximately right; what they miss is that legislatures do not directly control the distribution of government spending but instead rely on bureaucrats, who may have quite distinct motivations, to implement policies.⁶ The remainder of the article pursues the implications of this fact for coalition formation in the legislature.

The Model

The motivating example given in the introduction - bargaining over the distribution of SO₂ allowances in the Clean Air Act - demonstrated two salient features that are generally true of program and formula spending and that I aim to capture here. First, proposals took the form not of a list of payoffs for each member of Congress, but a much smaller list of formula elements: in the example, the relative weights placed on a unit’s fuel type, generating capacity, etc. in determining

⁵Though Berry, Burden, and Howell (2010a) offer the counterpoint that bureaucratic programs are not indefinitely lived, and are more likely to end when the current composition of Congress differs substantially from that of the program’s enacting coalition.

⁶See e.g. Prendergast (2007) and Besley and Ghatak (2005) on the incentives and motivation of individual bureaucrats, and the consequences for agency design.

the allowance allocation. Second, the mapping from formula to payoff was fixed and known: congresspeople were aware of the utilities operating within their districts and how manipulating the various formula elements was likely to help or hurt those utilities.

I capture these two features formally by postulating that bargaining occurs not directly over the ultimate distribution of funds, but indirectly over what I will call a *policy vector* x of length d , where d is less than the size of the legislature. The policy vector represents the set of “dials” that legislators can twist in determining how an appropriations bill’s budget is to be allocated. The number and identity of such “dials,” and hence the substantive interpretation of the policy vector, will vary depending on the context of a particular bill. For example, negotiations over a farm bill might take the form of alternative proposals for the relative subsidy levels granted to wheat versus corn versus soy production.

I model the process that translates the agreed-upon level of the policy vector x into ultimate distributive outcomes for every district by means of a *policy mapping matrix* Γ of dimension $(n \times d)$, which multiplies the policy vector x . Γ defines the (linear) mapping from policy choice to district payoffs.⁷ In reality, this translation from policy to outcome occurs via bureaucrats’ decisions to award funds on the basis of the criteria laid out in the authorizing legislation. Multiplication by Γ thus succinctly captures the process of implementation of a distributive bill’s award criteria.

The columns of Γ define the set of quantifiable measures or *attributes*, defined at the district level, that are available on which to condition the distributive consequences of the bill. Each column represents the vector of marginal effects on districts’ ultimate allocation of a change to a single formula element.⁸ The components of each column are multipliers, or weights, determining

⁷I assume that the mapping from policy to outcome is linear for analytical tractability. If the true mapping is non-linear, we can think of the linear version described here as a local approximation, valid for small changes of the elements of the policy vector.

⁸In the farm bill example, the column of Γ corresponding to the per-bushel corn production subsidy level would

the relative value of that particular formula criterion to a particular district. The total number of columns in the matrix, d , defines the dimensionality of the legislature's choice space. This number of benefits-determining factors is endogenous in the actual congressional process, but will be taken as fixed in the model; the Discussion section sketches some avenues for endogenizing this choice and argues that such an endogenously-chosen dimensionality constraint is likely to be low.

The mapping matrix for a given bill⁹ is assumed to be fixed, and known to legislators; hence, the effect of a tweak to any element of the policy vector on the ultimate distribution of funds is predictable. Appendix C presents some evidence for this assumption, showing that the cross-district allocation of funds for a given program is relatively stable from year to year. Note, however, that this structure by no means rules out uncertainty regarding the ultimate allocation. Under risk neutrality we can add forecast error - resulting, perhaps, from legislators' expectations regarding how the administering agency will interpret the statutory criteria being incorrect - with no change at all to the results. All that is required is that Γ gives the *expected* change in allocation to each district resulting from a change to the weight placed on each conditioning factor, and that these expectations are common across legislators.

The next sections lay this structure out in formal detail, and then derive several substantive conclusions. The main result is that $d < n$ (fewer policy levers than legislators) generates a partition of the legislature into voting blocs which always vote together. As d decreases, this partition becomes steadily coarser, leading to larger and possibly non-minimally sized winning coalitions. Furthermore, positive distributions to members outside the winning coalition are to be expected when the choice space defined by Γ is sufficiently constraining. For brevity, proofs are relegated to Appendix A.

be defined by measuring the expected bushels of corn produced by each district.

⁹As suggested by the farm bill and SO₂ allowance allocation examples, the set of relevant conditioning factors will vary depending on the policy area. Formally, this implies that there is potentially a different Γ for each distributive bill.

Primitives The primitives of the model are as follows. There is a set N of n legislators, who choose policy from a collection of choice sets $\{X_d\} : d \in \{\underline{d}, \bar{d}\}$. The index d indicates the dimension of the choice set. Each choice set satisfies (1) $X_d \subseteq \mathbb{R}^d$ and (2) $X_d \subset X_{d+1}$.

Each legislator is endowed with a utility function that takes a specific form. Namely, there exists, for every X_d , a fixed, exogenous matrix of characteristics Γ_d of dimension $(n \times d)$. Γ_d maps the “policy levers” in X_d into legislator utilities, such that if the chosen policy is $x \in X_d$, the vector of utilities for each district that results is:

$$u_d(x) = \Gamma_d x \tag{1}$$

The i th legislator’s utility from policy $x \in X_d$ is thus just the i th element of the vector $u_d(x)$. Adding an additional dimension does not change the values of the lower dimensions, e.g. $\Gamma_{d+1} = [\Gamma_d \mid \gamma_{d+1}]$ for some $(n \times 1)$ -dimensional vector γ_{d+1} . For technical reasons I make a final assumption, **(A1)**, on Γ_d that for every i there is some element of $(\Gamma_d)_i$ that is strictly positive.¹⁰

The choice sets X_d are bounded above by an exogenous budget constraint $B > 0$: for a policy x to be feasible ($x \in X_d$) it must satisfy $\beta(x) \leq B$, where $\beta(x)$ represents the cost of the policy. $\beta(x)$ takes the form:

$$\beta(x) = \sum_{i=1}^N [\Gamma_d x]_i$$

I.e., the cost of the policy equals the sum of the benefits it provides to all members. This is a purely distributive model, in the sense that the game being played is zero-sum. Legislators can choose only how to divide the “pie” amongst themselves; as the budget constraint B is fixed, there is no

¹⁰This requirement is trivially satisfied if we include a dimension d' in the choice set which has $(\Gamma_d)_{i,d'} = \frac{1}{n} \forall i$; e.g. it is possible to pay out the budget in equal shares to all legislators.

possibility of increasing the size of the pie.¹¹

The game form is the bargaining structure of Banks and Duggan (2000). Namely, in each round a single proposer is drawn from N ; legislator i is selected with probability ρ_i .¹² Whoever is recognized makes a proposal, which all members of N then vote up or down. If the proposal gains the support of a decisive coalition $D \in \mathcal{D}$, it passes, utilities are realized and the game ends; if not a new proposer is recognized and the game repeats. Legislators discount future payoffs at the common rate $\delta < 1$. I make two additional assumptions on the structure of the decisive coalition set \mathcal{D} : one, that it is proper, and two, that it is anonymous.¹³ Properness implies that if $D \in \mathcal{D}$, then $N - D \notin \mathcal{D}$; anonymity implies that only the number of legislators in the coalition matters for its success, or formally: if $D \in \mathcal{D}$ then $\{C : |C| = |D|\} \subset \mathcal{D}$. This structure subsumes both majority- and supermajority-rule voting procedures.

A strategy for an individual in the game where legislators bargain over X_d consists of an acceptance set $A_{i,d} \subseteq X_d$ and a proposal strategy $\pi_{i,d}$. $\pi_{i,d} \in \mathcal{P}(X_d)$, where $\mathcal{P}(X_d)$ is the set of all probability measures over X_d . Given a strategy profile for all individuals, the acceptance set of a coalition C is $A_{C,d} = \bigcap_{i \in C} A_{i,d}$. The social acceptance set is $A_d = \bigcup_{C \in \mathcal{D}} A_{C,d}$.

¹¹Note that the setup here includes the canonical bargaining model of Baron and Ferejohn (1989) as a special case; let $d = n$ and set Γ_d equal to the n -dimensional identity matrix. The model also nests the “naïve universalism” hypothesis - all legislators get equal shares of size $1/N$ - when $d = 1$ and Γ_d consists of a column of all ones.

¹²Note that the recognition probabilities are indexed by legislator because they need not be uniform across legislators; institutional features such as majority party status or committee positions can grant some legislators greater degrees of proposal power than others. Variation in recognition probabilities within the chamber thus allows the model to accommodate features like party influence or deference to committees.

¹³This second assumption rules out bicameralism, as a voting procedure requiring majorities in both houses for passage is not anonymous. To accommodate bicameral legislatures such as the US congress, the anonymity assumption can be dropped. Propositions 1 and 3 would go through unchanged in this setting. Versions of Propositions 2 and 4 can be proved without anonymity but both the proposition statements and the proofs are significantly more cumbersome. I focus on the case with anonymity for expositional clarity.

Solving the Model As in Banks and Duggan, I focus on no-delay stationary equilibria in which no legislator uses a weakly dominated voting strategy. The no-delay property implies that for any x proposed in equilibrium, $x \in A_d$. In such equilibria, we can define for each legislator a continuation value:

$$v_{i,d} = E[u_{i,d}] = \sum_{j=1}^n \rho_j \left[\int_{X_d} u_{i,d}(x) \pi_{j,d}(dx) \right]$$

Applying equation (1), we get:

$$v_d = \Gamma_d \sum_{j=1}^n \rho_j \left[\int_{X_d} x \pi_{j,d}(dx) \right] \quad (2)$$

In stationary equilibria, the individual acceptance sets $A_{i,d}$ are the upper contour set of $u_{i,d}$ at $u_{i,d}(v_{i,d})$. In other words, all legislators accept any proposal that gives them at least their continuation value from holding out for another round. For the proposer, the strategic question is how to assemble a coalition that satisfies enough legislators' continuation value constraints to get the bill to pass.

Existence of equilibrium For any dimension d , the model can be formulated as a special case of the general Banks and Duggan setup. Applying Theorem 1 of Banks-Duggan, we have immediately the following existence result:

Proposition 1. *There exists a no-delay stationary equilibrium. Furthermore, every stationary equilibrium is a no-delay equilibrium.*

Characterization of equilibrium Having established the existence of stationary equilibria, I proceed to the characterization of these equilibria, focusing on results relating to the dimension and rank of the matrix Γ_d . The first of these, Proposition 2, tells us that winning coalitions *cannot expand in size* if the dimension of the matrix Γ increases. No proposer who under dimension d proposed a policy to some coalition C would find it profitable, if an additional dimension became available, to propose a policy to a larger coalition $C^+ \supset C$.

Proposition 2. *Let $d^+ > d^-$. Let C^- be an arbitrary coalition of legislators that unanimously accepts a bill (i.e., is a winning coalition) with positive probability in some equilibrium under the smaller dimension, d^- . Let C^+ be any superset of C^- . If C^+ is not in the set of winning coalitions in any equilibrium under the smaller dimension d^- , then it is also not in the set of winning coalitions in any equilibrium under the larger dimension d^+ .*

The proof is in Appendix A, but the intuition is straightforward. As this is a zero-sum game, the proposer's goal is effectively to minimize the amount distributed to all other legislators, subject to her proposal being accepted. If it is possible to find a feasible policy vector x that satisfies all the members of C^- when there are d^- policy levers available, then x is still available and feasible when additional levers are added to the formula. The proof shows that all the members of C^- must still be satisfied with x when more dimensions become available, meaning that adding additional members would only increase the cost to the proposer while gaining her no advantage in getting the bill passed.

The second result concerns the structure of winning coalitions. If a set of districts are very similar along the dimensions available, the benefits they receive from any feasible policy will be correspondingly similar. In turn, these districts will have similar continuation values. This logic leads to the result of Proposition 3: any matrix of conditioning factors Γ with rank less than size

of the legislature generates a set of voting blocs whose members cannot be split by any feasible policy. In equilibrium, these voting blocs vote together in lockstep. As the dimension of the matrix Γ becomes smaller, the partition becomes more coarse, and the blocs become larger.

Proposition 3. *Suppose $d = \text{rank}(\Gamma_d) < n$. There exists a partition of N into d subsets $\{N_1, N_2, \dots, N_d\}$ such that all members of each subset always vote together in every equilibrium.*

This partitioning result implies that the task of the proposer becomes not assembling individual members (as in standard legislative bargaining formulations) but assembling *voting blocs*. Moreover, these blocs are generated endogenously by the observable characteristics of the districts that members represent. Formulaic allocation generates natural coalition partners and natural enemies, and may thus serve as an initial spark for the formation of legislative parties.

It is worth noting that this partitioned structure has two significant effects on the task of the proposer. First, low-dimensional formulas dramatically simplify the task's complexity, a point to which I return in the discussion section. To illustrate how dramatic this complexity reduction can be, consider the two extreme cases. If $d = 2$, the proposer has only two possible coalitions to choose from; finding the optimal proposal involves evaluating just two options to decide which can win at lowest cost. If $d = n$, on the other hand, under majority rule there are n choose $n/2$ possible winning coalitions, which for typical legislative sizes is an enormous number.

Second, the countervailing cost to this complexity reduction is that the value of proposal power declines as d decreases. The reasons for this are twofold. The first reason is that for small d it will generally not be possible to zero out the allocation to members outside of the winning coalition; in an attempt to win support of one voting bloc, another may get something as well, though not enough to convince its members to vote Yes. The second reason is that fitting together blocs rather than individuals will tend to lead to oversized coalitions. The final proposition states this formally.

Proposition 4 tells us that above a threshold value of d , the size of the voting blocs is small enough that there is always sufficient flexibility to achieve the proposer’s optimum (minimum-winning) coalition size of q . If d is below this threshold, the partition defined in Proposition 3 is too coarse. Oversized coalitions can be observed in equilibrium in this case, because the proposer does not have the flexibility to split up certain voting blocs.

Proposition 4. *Let q be the number of yes-votes required for a proposal to pass. If $\text{rank}(\Gamma_d) > n - \frac{q}{2}$, all coalitions are minimal winning. If $\text{rank}(\Gamma_d) \leq n - \frac{q}{2}$, there exist equilibria in which non-minimal winning coalitions occur with positive probability.*

Empirical implications: share of budget distributed to no-voters From the propositions just established, we know that a proposer, when recognized, minimizes the share of the budget paid to all other members, subject to the constraints that the bill has to expend the entire budget and has to pass.¹⁴ In real data, of course, it is not possible to observe the “proposer” in the Banks-Duggan sense.

Members’ votes on a bill, however, are observable, which allows us to write a weaker version of the minimization lemma (A3). Although the proposer’s identity remains unknown, the fact that the proposer must be a member of the winning coalition ensures that the following statement will hold regardless of the identity of the proposer: given an observed winning coalition C , the expected distribution $u(x^*)$ must minimize the aggregate payoff of all members in $N - C$, subject to the constraint that the full budget must be spent and all members of C must get at least $u(x^*)$. In other words, the equilibrium proposal x^* solves the linear program:

¹⁴See Lemma 3 in Appendix A for a formal statement of this property.

$$\begin{aligned}
& \min_{x \in X_d} \sum_{i \in N-C} [\Gamma_d x]_i & (3) \\
& \text{s.t.} \quad [\Gamma_d x]_j \geq u_j(x^*) \quad \forall j \in C \\
& \text{s.t.} \quad \sum_{k \in N} [\Gamma_d x]_k = B
\end{aligned}$$

The equilibrium distribution thus minimizes the share of the budget paid to no-voters, given the constraints enforced by the winning coalition's continuation values and the matrix Γ_d . As Γ_d becomes less restrictive - i.e. as $\text{rank}(\Gamma_d) \rightarrow n$ - the share of the budget captured by the winning coalition should approach 1, and no-voters should be increasingly cut out of the equilibrium distribution.

The result that the equilibrium distribution is the solution to a linear program means that given a hypothetical Γ_d , the winning coalition's optimal allocation to all members outside the coalition can be computed. Given a set of possible factors on which spending formulas could condition - e.g., district population, per-capita income, racial composition, etc. - it is thus possible to estimate the subset of factors involved in the formula. The estimation procedure works by choosing d to make the difference between 1) the winning-coalition-optimal allocation to no-voters when proposals are restricted to condition on no more than d factors, and 2) the actual observed allocation to no-voters, as small as possible. Details of an estimation routine which applies this principle to estimate d from appropriations data are reported in Appendix D.

In summary, the model predicts that allocation by formula has substantial consequences for legislative bargaining outcomes. The lower the dimension of the formula used to allocate appropriations, the more dramatically do the predictions depart from those of the classical model (and the

closer do they approach observed features of real-world distributive politics). Formulaic allocation breaks down the individualism of the legislature in a “state of nature.” It leads to the spontaneous generation of groups of districts who are aligned on some set of dimensions and share a common interest in moving policy in a particular direction. This change to within-legislature coalition politics, in turn, leads to changes outside the legislature as well: more universalistic policies that benefit broader collections of districts, and flows of distributive benefits even to constituents whose representatives opposed the authorizing legislation.

Results

The model just presented generates several testable predictions. The degree to which these predictions deviate from those of canonical models of bargaining depends entirely on the degree to which the complexity of the allocating formula, d , is smaller than the critical value $n - q/2$ defined in Proposition 4. Hence, the primary goal of the empirical analysis will be to estimate the effective degree of formula complexity d present in real appropriations legislation, and to assess the ability of the limited-complexity formula model to explain observed patterns in real spending data.

To that end I analyze a new dataset of US federal outlays covering 31 appropriations bills, the ultimate distribution of funds to each of the 435 congressional districts resulting from each bill, and the roll-call vote of each district on each bill. The data sources and collection methods, as well as summary tables and figures are presented in detail in Appendix E.

I first present some descriptive evidence demonstrating that the broad patterns predicted by the complexity-constrained model are in fact apparent in the data. Second, I measure the effective d of appropriations bills in the sample using two distinct approaches: a direct approach consisting of

reading the text of appropriations formulas included in the authorizing statute and determining the number of factors conditioned on by inspection, and an indirect approach which infers the effective d from the observed allocations to districts outside the winning coalition. The two approaches are complementary, each having relative strengths and weaknesses: the direct approach is more straightforward but applicable only at the program rather than bill level and only to the relatively small minority of programs with explicit statutory formulas. Third, I construct some model fit statistics and compare the model's performance with the baseline Baron-Ferejohn model that assumes that unconstrained redistribution is possible.

Empirical patterns

It is immediately evident from Table 5 that the typical coalition size on the spending bills in the sample is non-minimal-winning. The smallest margin of victory is the 109th Congress' Deficit Reduction Act of 2005, which passed by a margin of 6 votes, and there are a few near-universalistic coalitions. This fact alone does not constitute strong evidence for the theory, because there are a number of alternative explanations for large coalitions. Still, it is clear that the simple Baron-Ferejohn model provides a poor fit to the coalition-size data.¹⁵

Existing theories of legislative bargaining - including those that can account for oversized coalitions - predict that the winning coalition should capture all of the distributive benefits to be had in a given bill. Columns (1) - (3) of Table 1 examine this prediction in the data, presenting regression estimates of the share of the total bill budget received by a district as a function of the district's vote.¹⁶ All specifications include bill fixed effects, such that comparisons are relative to other dis-

¹⁵Knight (2005) draws the same conclusion in the context of estimating a structural model with earmark data from two transportation-spending bills.

¹⁶Table 6 in Appendix E presents bill-by-bill differences in mean receipts among Yes and No voters.

tricts on the same bill.¹⁷ As is clear from Table 6, the total budget expended and average per-district allocation varies widely by bill, a possibly confounding source of variation; transformation to share terms makes the levels of the dependent variables comparable across bills.¹⁸

The point estimates on voting No in this regression are consistently negative, and approach conventional levels of significance in some specifications.¹⁹ The difference is robust to inclusion of additional controls for partisanship, ideological distance from the chamber median, an indicator for being an at-large district, and Senate representation.²⁰ However, the power of the district's membership in the winning coalition to predict its share of distributive benefits in all specifications is very weak, with R^2 values on the order of 0.001. The substantive size of the estimate is moderate - compared to Yes voters on the *same bill*, No voters receive on average 0.061-0.066 percentage points less, which is approximately a quarter of the average share of 0.23%.²¹

The distribution of district-level expenditures for any given bill typically features a point mass at zero - that is, some subset of districts do not receive any grants authorized by the legislation.²² As it is possible that the credit-claiming benefits of voting for distributive legislation have more to do with the fact of a federal grant - and the attendant ribbon cutting ceremony photo opportunities and press releases to local media (Grimmer 2013) - than its absolute or relative dollar amount, I estimated the same model with a binary dummy for any positive grant amounts on the left-hand

¹⁷Table 8 in Appendix G presents the same specifications with district fixed effects.

¹⁸Table 9 in Appendix G presents analogous results in absolute-dollar terms.

¹⁹To allow for the fact that the variance in shares is correlated within bill, standard errors are clustered at the bill level in all regressions.

²⁰Party controls are included to account for theoretical expectations about of majority party agenda control (Cox and McCubbins 2007) or political influence by the executive branch in implementation decisions (Berry, Burden, and Howell 2010b). The at-large control matters because many programs distribute funds to state agencies with the stipulation that every state receive some minimum share. For at-large districts, all of this state-level funding, by definition, ends up in the district, whereas for multi-district states it will be split across districts.

²¹Shares always add to 1, so the average district share is 1/435, or 0.23%.

²²As Stein and Bickers (1997) note, this pattern is even more pronounced at the program level.

Table 1: Models of District Funding Receipts, 109-111 Congress

	Amount Received (% of Bill Total)			Any Grants Received		
	(1)	(2)	(3)	(4)	(5)	(6)
No Vote on Final Passage	-0.064*	-0.061	-0.066	-0.011	-0.009	-0.011
	(0.035)	(0.048)	(0.049)	(0.011)	(0.015)	(0.015)
Abstain on Final Passage	0.076	0.070	0.071	-0.006	-0.009	-0.009
	(0.065)	(0.064)	(0.065)	(0.018)	(0.018)	(0.018)
Majority Party		0.089	0.096		-0.001	0.004
		(0.071)	(0.067)		(0.030)	(0.029)
President's Party		-0.074**	-0.076**		-0.007	-0.012
		(0.037)	(0.036)		(0.021)	(0.020)
Ideological Extremity		0.094	0.088		-0.013	-0.009
		(0.120)	(0.119)		(0.033)	(0.033)
At-Large District		0.331**	0.334**		0.304***	0.306***
		(0.142)	(0.142)		(0.062)	(0.062)
Senate: Majority Party			-0.066**			-0.027*
			(0.034)			(0.015)
Senate: President's Party			0.022			0.033**
			(0.033)			(0.015)
Senate: Delegation Split			-0.038			0.008
			(0.036)			(0.014)
Fixed Effects:	Bill	Bill	Bill	Bill	Bill	Bill
Number of Bills	31	31	31	31	31	31
F-statistic	0.19	0.89	0.95	428.67	390.08	361.46
N	13,485	13,485	13,485	13,485	13,485	13,485
R ²	0.0004	0.002	0.003	0.505	0.511	0.512

*p < .1; **p < .05; ***p < .01

Cluster-robust standard errors in parentheses (clustered at the bill level). Bill-clustered standard errors are conservative relative to district-clustered or heteroskedasticity-robust versions. All right-hand-side variables are binary indicators except for Ideological Extremity, which is measured as the absolute deviation between the first-dimension DW-NOMINATE score of the district's representative and that of the chamber median. Dummies for party alignment of the district's Senate delegation are one only if both Senators are in the Senate majority or the president's party, respectively. The "Senate Delegation Split" dummy indicates that the state's Senate delegation consists of one Democratic and one Republican Senator.

side. The resulting linear probability models, which model the probability of receiving grant(s) of any size as a function of the same covariates, are presented in columns (4)-(6) of Table 1. The results here are similar: the no-vote effect is negative but small. Confidence intervals include zero in all specifications, and rule out substantively large effects - the lower bound of the confidence interval on the no-vote effect is negative 3-4%, compared to the baseline probability of about 50%.

In sum, the results here are difficult to reconcile with existing theories of bargaining under majority rule. While there is some evidence - in the form of negative point estimates - that winning coalitions attempt to reduce the amount of funds distributed to no-voters, the small substantive effects indicate that their ability to do so is very limited. Rather than being driven to zero, as predicted by existing theories, No-voters' average allocations are roughly three-quarters the size of their Yes-voting colleagues', and their probability of receiving a grant of any size is essentially the same. These patterns are, in contrast, entirely consistent with the theory of bargaining over formulas developed in the preceding section, provided that the effective formula dimension is sufficiently constraining. I next seek to measure this degree of constraint directly.

Estimating the dimension of proposals

A relatively small minority ($\sim 15\%$) of programs provided for by the appropriations bills in my sample have explicit statutory formulas. However, programs with associated statutory formulas tend to be larger than those without. In FY2011, for example, the average budget of a program with a statutory formula was \$2.6B, compared to an average of \$760M for programs without. They also tend to be more universalistic - in terms of the number of districts who receive direct benefits from the program - as shown in Appendix E, Figure 6.

For this minority of programs with a statutory formula, it is instructive to consider a direct,

textual approach to measurement of the formula dimension. For each program that the *Catalog of Federal Domestic Assistance* (CFDA) indicated had a statutory formula,²³ I located the text of the formula in either the U.S. Code or the Code of Federal Regulations and recorded the number of factors referenced in determining the allocation. For instance, one of the simplest allocation formulas applies to the HIV Emergency Relief Project Grants authorized by the Ryan White HIV/AIDS Treatment Extension Act of 2009 (abbreviated slightly for clarity):

(A) Subject to the extent of amounts made available in appropriations Acts, a grant made for purposes of this paragraph to an eligible area shall be made in an amount equal to the product of (i) an amount equal to the amount available for distribution for the fiscal year involved; and (ii) the percentage constituted by the ratio of the distribution factor for the eligible area to the sum of the respective distribution factors for all eligible areas (...)

(B) Distribution factor

For purposes of subparagraph (A)(ii), the term distribution factor means an amount equal to the living cases of HIV/AIDS (reported to and confirmed by the Director of the Centers for Disease Control and Prevention) in the eligible area involved.

This formula was coded as one-dimensional, as it references only a single factor: the number of living cases of HIV/AIDS in an area. The number of cases in some local area, relative to the number of cases in the US as a whole, entirely determines the allocation of HIV Emergency Relief grants to local health centers.

Quantiles of the distribution of formula dimension across all 80 programs in the sample with statutory formulas are presented in Table 2.²⁴ The typical formula is very simple: the median number of factors referenced is 3, and 90% of formula-grant programs have 5 or fewer factors in their allocation formulas. The outlier program is HUD's Public Housing Capital Fund (PHCF),

²³The "Formula and Matching Requirements" field in the CFDA indicates whether or not a program has a statutory formula and, typically, references the location in the U.S. Code or the Code of Federal Regulations where the formula can be found.

²⁴The distribution is shown in histogram form in Appendix E, Figure 5. The full coding of factors referenced in the formula for each program, with textual references, is available upon request from the author.

which has distinct sets of weights applicable to local housing agencies with less than 250 or more than 250 public housing units, each of which includes multiple characteristics of the public housing stock, such as average age, average number of bedrooms, and so on. HUD also exempts the New York City and Chicago Housing Authorities from the formula used for other large cities, using a separate process to determine their allocations. Taken together, there are a total of 40 distinct weighted factors appearing in the PHCF statutory formula.

10%	25%	50%	75%	90%	95%
1	2	3	4	5	8

Table 2: Quantiles of the distribution of the number of factors referenced in the allocation formula, among CFDA programs allocated by statutory formula.

With the exception of a few programs like the PHCF, statutory formulas are, by and large, quite low-dimensional. A typical formula guarantees some minimum percentage to every state or local agency applying for the program, and allocates the remaining appropriations according to total population.²⁵ They are also surprisingly durable and persistent over time: four distinct programs administered by the Department of Education and appropriated for by the 2009 American Recovery and Reinvestment Act (ARRA) all retain the same statutory formula originating in the Elementary and Secondary Education Act of 1965, one of the core elements of Lyndon Johnson’s “Great Society.”

What about the programs without explicit statutory formulas? The remaining 85% of programs are awarded competitively, according to criteria specified in the authorizing statute but which grant wide discretion in allocation decisions to the agency operating the program. For example, the

²⁵The Social Security Administration’s State Grants for Protection and Advocacy Services program, FEMA’s Emergency Management Performance Grants, and the Department of Human Health and Services’ Family Violence Prevention and Services grants all use this formula, coded as two-dimensional (a constant term plus total population).

Transportation Investment Generating Economic Recovery (TIGER) discretionary grant program, also funded by the ARRA, awarded grants according to the following criteria listed in CFDA:

The Primary Selection Criteria include (1) Long-Term Outcomes and (2) Jobs Creation & Economic Stimulus. The Secondary Selection Criteria include (1) Innovation and (2) Partnership. The Primary selection Criteria were intended to capture the primary objectives of the TIGER Discretionary Grants provision of the Recovery Act, which include near-term economic recovery and job creation, maximization of long-term economic benefits and impacts on the Nation, a region, or a metropolitan area, and assistance for those most affected by the current economic downturn. The Secondary Selection Criteria were intended to capture the benefits of new and/or innovative approaches to achieving programmatic objectives.

Direct measurement of the complexity of these non-statutory-formula programs is difficult. The text of the selection criteria omit legislators' implicit expectations about the outcomes of agency rulemaking and grant awarding processes, meaning that purely textual measures, such as counting the number of distinct criteria referenced, are likely to provide a poor approximation to the effective dimension of congressional proposals. And, as is evident in the TIGER example, criteria are often written vaguely enough that many different implementations of varying complexity are plausible, depending on the agency's interpretation of the statutory criteria.

An additional complication arises from the fact that the typical appropriations bill authorizes funding for multiple distinct programs, some of which are allocated by statutory formula and some of which are not. Figure 4 in Appendix E shows that some large appropriations bills authorize funds for double-digit numbers of programs; the median bill in the sample authorizes appropriations for 4 distinct programs.

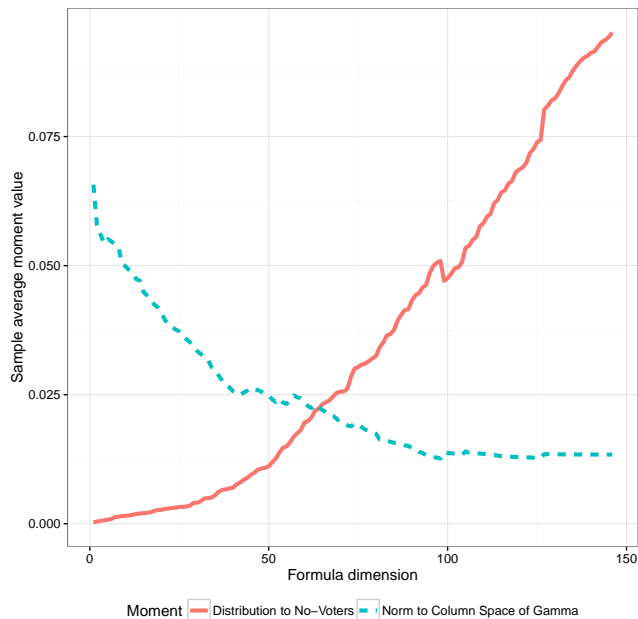
If non-statutory-formula programs look like statutory-formula programs in their complexity, and if there is no overlap in the sets of factors referenced by each program, then, given the preceding analysis of statutory formulas the typical appropriations bill allocation complexity would be on

the order of 12 dimensions. Of course, the differences in budget size and distribution of benefits between statutory-formula and non-statutory-formula programs are evidence that the two kinds of programs may not be comparable. And, within the subgroup of statutory-formula programs, some very common factors, such as total population or per-capita income, show up in the allocation formula for many different programs. Hence, this back-of-the-envelope estimate is unlikely to be the right one.

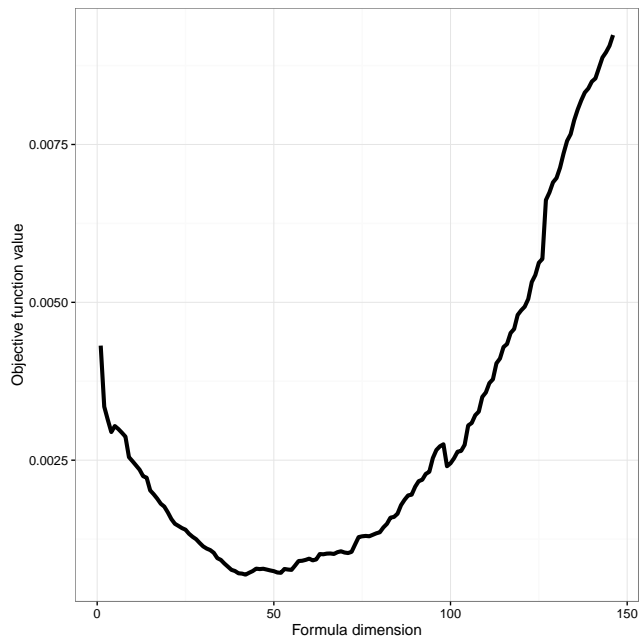
As a consequence of this measurement challenge, I adopt an alternative approach that relies only on data concerning distributive outcomes to *infer* the effective dimension of the choice space. The approach is to ask: supposing that the model presented above is the correct one, what value(s) of d are most likely to have generated a distribution of spending outcomes like the one we observe in the data? The overall fit of the model can then be assessed by evaluating how much additional variation in the data can be explained by the model at the best-fitting value of d , relative to the baseline (Baron-Ferejohn, $d = n$) variant.

The method required to implement this approach is described in detail in Appendix D. The essential component is that the observed distribution of spending under the model is the solution to the linear program (Equation 4) whose parameters are a function of the policy mapping matrix Γ_d . Using this property, it is possible to develop a simulation-based estimator for the effective dimension of the choice space, d . I can then apply the technique of McFadden (1989) to construct a GMM objective function and obtain parameter estimates and confidence intervals.

Specifically, there are two moments whose sum of squares the estimator seeks to minimize in the sample. Each moment is defined at the level of the bill; the optimal choice of d is that which minimizes the sample-average value of the objective function. The first moment is the difference between the observed distribution to no-voters and the proposer-optimal distribution, given the



(a) Sample average values of the two moments.



(b) Objective function values.

Figure 1: Sample average simulated moments and objective function values, as a function of the dimension of the choice space.

available set of factors on which to condition the spending formula. Figure 1a shows that this moment is generally increasing in d ; the more factors there are available, the easier it is to find an alternative policy that preserves the winning coalition’s benefit levels while reducing the distribution outside the winning coalition. When d is highly constrained, proposers have almost no flexibility and there are essentially no alternatives available that improve the position of the winners at the expense of the losers; hence this moment approaches zero at the smallest value of d tested, 2.²⁶ As additional factors are added, the moment rises, with an acceleration occurring around $d = 30$.

The second moment is the sum of squared residuals in a regression of the non-zero elements of

²⁶I skip the trivial case of $d = 1$, where there is only one feasible allocation and this moment value is exactly zero.

the observed funding distribution on the mapping matrix Γ .²⁷ The logic here is that in the model, the matrix Γ is the linear mapping from the equilibrium policy vector to the vector of district-level spending outcomes. That is, variation in outcomes is entirely “explained,” in the linear-regression sense, by variation in the columns of Γ - the R^2 in a regression of district receipts on Γ would be exactly 1. Of course, in reality a number of reasons - including measurement error, or unexpected agency implementation decisions that differ from legislators’ expectations - prevent this from holding exactly. The GMM approach thus tries to get this moment as small as possible, by choosing Γ that is complex enough to explain a significant fraction of variation in outcomes. Figure 1a shows that this moment declines as d increases, with most of the gains achieved by around $d = 40$.

The GMM objective - which weights both moments equally - trades off these two countervailing pressures. The resulting graph, shown in Figure 1b, thus displays an inverted-U shape with an interior minimum. The value of d that minimizes the objective function in the sample (the low point of Figure 1b) is 42. The simulation estimate is thus quite a bit larger than the extrapolation from the statutory-formula textual evidence presented earlier, suggesting that project grants and non-formula programs serve to increase congressional proposers’ flexibility relative to formula grants. Nonetheless, the estimate is still substantially restricted compared to the theoretical upper bound. The point estimate implies a 90% reduction in degrees of freedom relative to the Baron-Ferejohn benchmark.

A confidence interval on this estimate can be constructed by bootstrap resampling from the set of bills. I resample with replacement from the set of bills, and compute the minimizing value of d , exactly as just described, for each resample. The resulting distribution of estimates is presented in Appendix G, Figure 8. The central 95% interval of this distribution is [40, 62]; the 90% interval is

²⁷E.g., the Euclidean norm of the orthogonal projection of the vector of district spending outcomes onto Γ .

[40, 57].

Model fit and implications

In addition to the estimate of formula dimension just described, the bargaining simulation model yields some informative fit statistics and additional implications. First, the output of the model at the estimated dimension can be used to infer the share of funds outside the winning coalition that can be explained by formula constraints - that is, the fraction that the winning coalition could not have extracted, given the set of factors available. Second, the model yields the set of factors included in the simulated formula for each bill; by examining all bills in the sample and finding the factors that repeatedly appear, we can see which district-level attributes appear to be most important for determining the distribution of federal spending. Finally, the model outputs the proposer's optimal allocation to all districts for each bill. By comparing districts' allocations when they are in the winning coalition compared to when they are out of it, we can learn about the kinds of districts who benefit most from formulaic distribution.

Share of funds distributed to no-voters explained The first moment in the GMM objective used to fit the model is the difference between the observed distribution to no-voters and the proposer-optimal distribution for a given bill. The theoretical section provides a useful benchmark for this moment: if the proposer has complete flexibility ($d = N$, the Baron-Ferejohn case) the moment equals the total share of the bill budget distributed to members outside the winning coalition.²⁸ Comparing the moment value at any given d to this maximal case, then, provides a measure of model fit: the fraction of distributions outside the winning coalition that remain

²⁸In the benchmark case, the winning coalition takes everything; hence, all distributions to members outside the winning coalition are unexplained.

unexplained. Subtracting this ratio from one gives the fraction of distributions outside the winning coalition that are rationalized by the dimensionality restriction. Computing this ratio for each bill at the estimated value of d reveals that the fit is quite good: the fraction of distributions to no-voters “explained” ranges from 80.78% (P.L. 109-171) to 99.87% (P.L. 111-212). The average value across all 31 bills is 97.6%. Hence, even at $d = 42$ proposers have little available flexibility; by acting optimally they could reduce the share of funds going outside the winning coalition (and redistribute those funds to members of the winning coalition) by only about 2.5%.

Conditioning factors As detailed in Appendix D, the simulation selects columns to include in the mapping matrix Γ from the set of available district-level covariates on a bill-by-bill basis. By examining the factors included in the simulated formula for many bills, it is possible to get a sense of the covariates that appear to reliably predict levels of federal spending. Figure 9 in Appendix G presents the frequency of inclusion for the factors that appear in the simulated formula for at least half of the bills in the sample.

Examination of the resulting set reveals that measures of household and family size as well as the income and age distribution are among the most commonly included factors. This is likely related to the fact that many of the largest programs funded by bills in the sample are means-tested programs targeted to poor children and the elderly. Moving down the list, the geographic variables measuring land cover - the type of environments, human or natural, that make up a district’s surface area - appear frequently, likely related to farming and forestry subsidies that go only to districts with significant land area dedicated to cultivated crops or forest. Relatedly, the only occupational variable to make the list is that measuring the fraction of employed residents in farming and fishing occupations. The remainder of the set consists of measures of housing value, most likely due to these variables’ predictive value for the allocation of housing subsidies like the Section 8 voucher

program.

Position-taking and distributive benefits The model also generates some interesting predictions about those districts that benefit most from formulaic allocation. For each bill, the model outputs the proposer’s optimal distribution to all districts outside of the winning coalition. Using this information, it is possible to compare districts’ average performance in capturing appropriations when they are part of the winning coalition, to their average performance when they are out of the winning coalition and the proposer attempts to minimize their allocation given the limited set of conditioning factors available. The districts for whom this “yea-nay ratio” is highest are those which benefit most from formulaic allocation, as it is for these districts that the ability of a hostile majority to cut them out of a distributive deal is most limited.

Figure 2 plots this ratio for each district, against the first-dimension DW-NOMINATE score of the district’s representative. The denominator of the ratio is the average share of the total bill budget flowing to a district when that district is in the winning coalition; the numerator is the average share of the total bill budget that a proposer solving the optimization problem of equation (4) would have to send to the district when that district is outside of the winning coalition. There is a clear pattern that more conservative districts have higher ratios; most districts represented by Republicans have yea-nay ratios close to 1. At the liberal end of the NOMINATE distribution, however, the ratio drops off fairly steeply. Formulaic spending creates convenient cover for ideological conservatives to stake out positions opposed to various federal spending programs - and to back those positions up with an ideologically consistent roll-call voting record - without suffering the electorally dangerous consequence of actually denying their constituents access to those programs. Low-dimensional formulas allow legislators to pander to two different constituencies simultaneously, potentially engaging in pork-barrel politics while at the same time using their voting record to signal

ideological purity. The model endogenizes this kind of dual-pandering behavior, as the members for whom this strategy is feasible are those in districts with attributes that make them hard to cut out of distributive deals.

Discussion

It is a fact of distributive politics in the US federal context that all funds appropriated by Congress must pass through the federal bureaucracy before reaching their intended recipient. Before funds can be distributed, Congress must either assign one of the existing federal agencies to administer a program, or create a new one. Congress' instructions to the administering agency regarding the distribution of funds consist of a set of objective criteria on which bureaucrats may condition the award of grants, with weights attached indicating the relative importance of each criterion.

This paper has shown that this intermediation - treated as a detail that can be safely abstracted away in existing theories of distributive politics - is actually essential to understanding the structure, persistence and size of legislative coalitions on distributive issues. The less complex are the allocation criteria provided to the bureaucracy, in terms of the number of factors conditioned on, the more imprecise is the ability of legislative proposers to target funds to particular districts.

As a result, allocation by formulas of limited dimension provides a parsimonious explanation of multiple features of real distributive politics: oversized coalitions, distribution of funds to districts whose representatives voted against the authorizing legislation, and the emergence of persistent voting blocs of districts with common interests. Formulaic allocation also provides cover for legislators to engage in position-taking No votes without cutting themselves out of the deal, risking the ire of constituents. This feature is especially valuable for legislators, like those in the US House, who

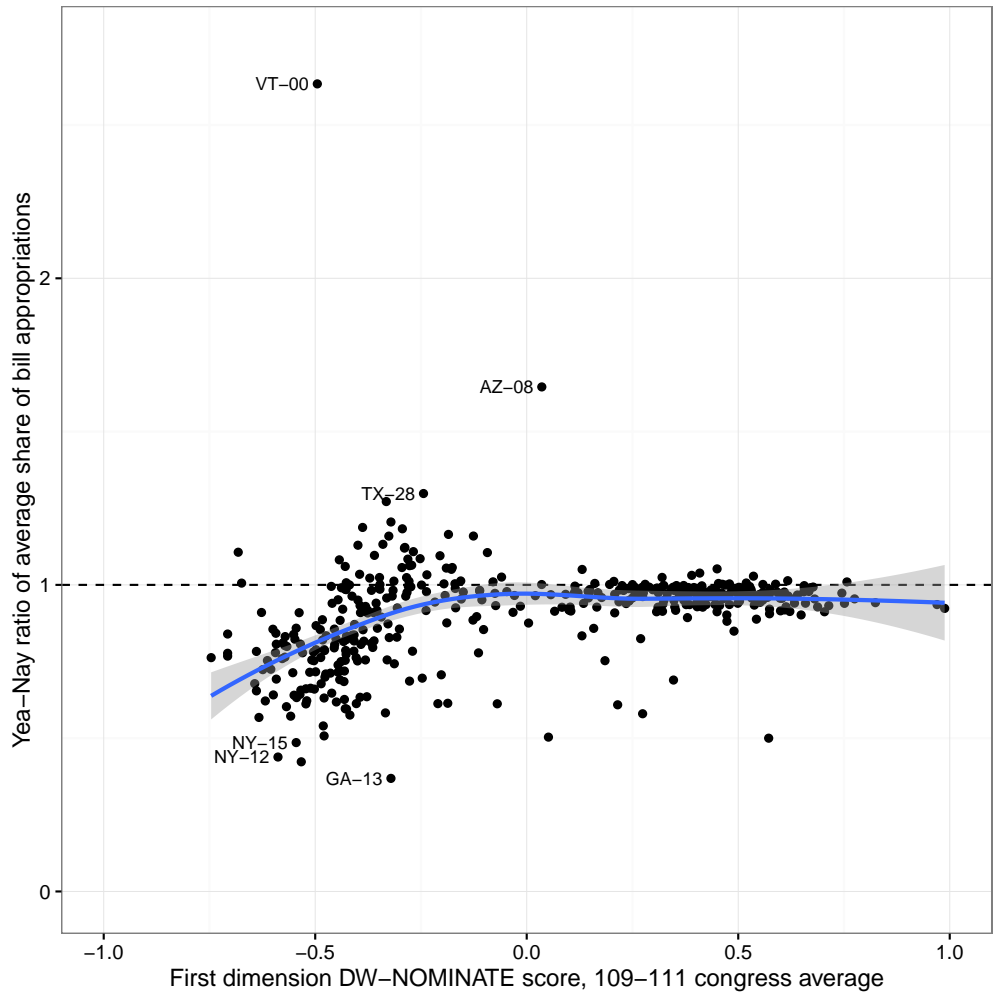


Figure 2: A comparison of districts’ relative performance in capturing appropriations when inside the winning coalition and when outside of it. Each point is a district; the Y-axis gives the ratio between the average share of the budget allocated to the district when voting Yea and the simulated proposer-optimal distribution to the district when voting Nay. Both quantities are averages across all 31 bills in the sample. The X-axis is the average first-dimension DW-NOMINATE score of the district’s representative in the 109-111th Congresses. The solid blue line is a local-linear smoother; the dashed black line intersects 1, e.g., the district does equally well when voting Yea as when voting Nay. The identities of a few outlier districts are labeled.

must face both ideologically-motivated primary electorates and more pragmatically-minded general electorates.

Empirically, we have seen that actual allocation formulas, where they can be directly observed, are in fact quite low-dimensional. In the remaining cases - where program criteria are too vague to permit explicit measurement from the text of the statute - I applied the model to data on distributive outcomes to recover an estimate of implicit formula complexity. Usefully, the model nests both the Baron-Ferejohn model of perfect targeting and the “naïve universalism” hypothesis of perfectly egalitarian distribution as special cases.²⁹ The data reject both of these polar cases, and support an intermediate level of the upper bound on formula complexity. The empirical estimate tilts decidedly towards the lower end of the spectrum of possibilities, however, granting congressional proposers less than 10% of the conceivable degrees of freedom.

Why would such low-dimensional formulas arise in equilibrium, when members of the legislative majority clearly have incentives to increase complexity and thereby improve their ability to precisely target funds? Again, one answer lies in the countervailing incentives of the bureaucrats with whom legislators must interact.

Bureaucratic agencies are staffed by agents with their own private motivations, which are unlikely to be purely mercenary (Prendergast 2007). Aligning the missions of federal agencies with the kinds of broad social missions likely to be present in the population of potential bureaucrats could allow Congress, a la Besley and Ghatak (2005), to limit the need for high-powered monetary incentives and may thus reduce the bureaucratic overhead costs involved in redistribution. The trade-off is that agencies with more universalistic missions are likely to use their budgets in ways that benefit districts in more universalistic fashion, reducing the opportunity for targeting funds to the majority.

²⁹ $d = N$ and $d = 1$, respectively.

Some degree of bureaucratic autonomy and professionalization is necessary for such considerations to bind. Hence, we can expect that in past eras with weaker civil service protections and more direct Congressional control over the bureaucracy, distributive outcomes would tilt more towards full capture by the winning legislative coalition. Conversely, a strengthening of the bargaining power of bureaucrats relative to legislators would be expected to shrink the effective formula dimension, leading to larger coalitions and more funds dispersed outside of the winning coalition.

A second answer is that there may be direct costs to complexity. Returning to the SO₂ allowance example given in the introduction, members of Congress probably had a general sense of what types of fuels powered the plants of their district's local electric utility, but it is highly improbable that they knew the exact operational characteristics of all the generating units operating in their districts. Moving from a formula involving only the former, general information to one involving the latter, specific information would have involved real costs in research staff time and in the potential for errors in prediction.

This complexity reduction benefit is especially pronounced for those legislators involved in drafting a proposal and whipping votes. When the proposal dimension grows large, the number of possible coalitions to consider quickly becomes astronomical. Limiting proposals to a small set of formula dimensions, while sacrificing some of the benefits of proposal power, keeps the proposer's optimization problem manageable. If legislators are boundedly rational a la Bendor (2010), this tradeoff may be well worthwhile.

References

- Banks, J. S. and J. Duggan (2000). A Bargaining Model of Collective Choice. *The American Political Science Review* 94(1), 73–88.
- Baron, D. P. and J. A. Ferejohn (1989). Bargaining in Legislatures. *The American Political*

- Science Review* 83(4), 1181–1206.
- Bendor, J. (2010). *Bounded Rationality and Politics*. Berkeley: University of California Press.
- Berry, C. R., B. C. Burden, and W. G. Howell (2010a). After enactment: The lives and deaths of federal programs. *American Journal of Political Science* 54(1), 1–17.
- Berry, C. R., B. C. Burden, and W. G. Howell (2010b). The president and the distribution of federal spending. *American Political Science Review* 104(04), 783–799.
- Besley, T. and M. Ghatak (2005). Competition and incentives with motivated agents. *The American Economic Review* 95(3), 616–636.
- Carrubba, C. J. and C. Volden (2000). Coalitional politics and logrolling in legislative institutions. *American Journal of Political Science*, 261–277.
- Cox, G. W. and M. D. McCubbins (2007). *Legislative Leviathan: Party Government in the House*. Cambridge University Press.
- Duggan, J. and T. Kalandrakis (2012). Dynamic legislative policy making. *Journal of Economic Theory* 147(5), 1653–1688.
- Grimmer, J. (2013). *Representational Style in Congress: What Legislators Say and Why It Matters*. Cambridge University Press.
- Groseclose, T. and J. M. Snyder (1996). Buying Supermajorities. *The American Political Science Review* 90(2), 303–315.
- Joskow, P. L. and R. Schmalensee (1998). The Political Economy of Market-Based Environmental Policy: The U. S. Acid Rain Program. *Journal of Law and Economics* 41(1), 37–83.
- Knight, B. (2005). Estimating the value of proposal power. *American Economic Review* 95(5), 1639–1652.
- Levitt, S. D. and J. M. Snyder (1997). The Impact of Federal Spending on House Election Outcomes. *Journal of Political Economy* 105(1), pp. 30–53.
- McFadden, D. (1989). A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica* 57(5), 995–1026.
- Minnesota Population Center (2011). National Historical Geographic Information System. Technical Report Version 2.0, University of Minnesota, Minneapolis, MN.
- Niou, E. M. and P. C. Ordeshook (1985). Universalism in congress. *American Journal of Political Science*, 246–258.
- Prendergast, C. (2007). The motivation and bias of bureaucrats. *The American Economic Review* 97(1), 180–196.
- Snyder, J. M. (1991). On Buying Legislatures. *Economics and Politics* 3(2), 93–109.

- Stein, R. M. and K. N. Bickers (1994). Universalism and the electoral connection: A test and some doubts. *Political Research Quarterly* 47(2), 295–317.
- Stein, R. M. and K. N. Bickers (1997). *Perpetuating the pork barrel: Policy subsystems and American democracy*. Cambridge University Press.
- Volden, C. and C. J. Carrubba (2004). The formation of oversized coalitions in parliamentary democracies. *American Journal of Political Science* 48(3), 521–537.
- Weingast, B. and W. Marshall (1988). The industrial organization of Congress; or, why legislatures, like firms, are not organized as markets. *The Journal of Political Economy*, 132–163.
- Weingast, B. R. (1994). Reflections on distributive politics and universalism. *Political Research Quarterly* 47(2), 319–327.

Appendices

A Proofs

Proof of Proposition 1

Proof. To apply Theorem 1 of Banks-Duggan, we first need to verify three technical conditions:³⁰

1. Explicitly specify the status quo, q as an element in X . I take the status quo to be the zero vector, which is by definition an element of every X_d .
2. Impose the requirement that the discount rate δ_i of every individual is identical, e.g. $\delta_i = \delta \forall i$.
3. Define individual i 's payoff from an outcome (x, t) as $(1 - \delta^{t-1})u_i(q) + \delta^{t-1}u_i(x)$ and normalize $u_i(q) = 0$ for every i . This condition is already satisfied given the definitions $q = 0$ and $u_{i,d}(x) = (\Gamma_d x)_i$.

And then check that the three main conditions of the theorem are met:

1. Each X_d is nonempty, convex, and compact.
2. The $u_{i,d}$'s are concave and continuous.
3. For every i there exists some $x_i \in X_d$ such that $u_{i,d}(x_i) > 0$.

1) $0 \in X_d$ for any B, Γ_d , so X_d is nonempty. Convexity and compactness follow immediately from the linear budget constraint and non-negativity constraints that define X_d . 2) Concavity and continuity are ensured by the linearity of the utility function. 3) The assumptions that $B > 0$ and

³⁰See the discussion on pp. 85 of Banks and Duggan (2000) for the logic underlying these conditions.

$\exists k_i$ s.t. $(\Gamma_d)_{i,k_i} > 0$ for all i imply that there is a feasible allocation $(0, 0, \dots, 0, B/(\sum_i (\Gamma_d)_{i,k_i}), 0, \dots, 0)$ - e.g., spending the entire budget on dimension k_i and zero on all other dimensions - that gives i strictly positive utility. \square

Proof of Proposition 2 To establish this claim, a few definitions are in order. First denote by $S(\pi_d^*)$ the support of the equilibrium proposal strategy profile π_d^* . Then, define for each X_d the set of decisive coalitions that unanimously accept a proposal in equilibrium with positive probability:

Definition 1. $W_d \equiv \{C \in \mathcal{D} : S(\pi_d^*) \cap A_{C,d} \neq \emptyset\}$

Next, a few lemmas prove useful in establishing the main result. Lemma 1 shows that all equilibrium proposals must expend the full budget:

Lemma 1. *For all $x \in S(\pi_d^*)$ and any d , $\beta(x) = B$.*

Proof. Suppose not, and let i be a member who proposes x with $\beta(x) < B$ in equilibrium. By **(A1)** there exists some basis vector e_k and some $\epsilon \in \mathbb{R}_{++}$ such that $x + \epsilon e_k$ yields strictly higher utility for i and weakly higher utility for all other members. This is inconsistent with the sequential rationality and no-weakly-dominated-voting assumptions. \square

Lemma 2 requires that the minimum cost of a policy acceptable to some coalition cannot decrease if additional members are added to the coalition. Furthermore, if there exists a policy such that all continuation value constraints for a given coalition can be exactly satisfied, the minimum-cost policy must strictly increase when additional members are added:

Lemma 2. *If $C' \subset C$, then for any d :*

1. $\min_{x \in A_{C,d}} \beta(x) \geq \min_{y \in A_{C',d}} \beta(y)$.

2. If $\exists y \in X_d$ with $\beta(y) = \sum_{i \in C'} \left(\sum_j \rho_j \int_{X_d} [\Gamma_d z]_i \pi_{j,d}^*(dz) \right)$ then the inequality is strict.

Proof. $C \supset C'$ directly implies that $A_{C,d} \subseteq A_{C',d}$. The first part of the lemma then follows immediately. To show the second part, note that in stationary equilibrium:

$$A_{C,d} = \{x \in X_d : [\Gamma_d x]_i \geq v_{i,d} \forall i \in C\}$$

Substituting into the definition of $\beta(\cdot)$ and v_i , we have that:

$$\begin{aligned} \beta(x) &\geq \sum_{i \in C} \left(\sum_j \rho_j \int_{X_d} [\Gamma_d z]_i \pi_{j,d}^*(dz) \right) \\ \beta(y) &\geq \sum_{i \in C'} \left(\sum_j \rho_j \int_{X_d} [\Gamma_d z]_i \pi_{j,d}^*(dz) \right) \end{aligned}$$

The sum defining the lower bound of $\beta(x)$ is strictly greater than the lower bound of $\beta(y)$, because each member's continuation value must be interior to the set of possible utility values and hence, if **(A1)** holds, must be strictly positive. Hence if there is a y that achieves the lower bound, $\min_{x \in A_{C,d}} \beta(x) > \min_{y \in A_{C',d}} \beta(y)$. \square

Finally, Lemma 3 shows that we can redefine the objective function of a proposer in terms of minimizing the amounts distributed to all other members. This property proves useful in generating necessary conditions for equilibrium.

Lemma 3. Let $Z_d \equiv \{x \in X_d : \beta(x) = B\} \cap A_d$, and define $\beta_{-i}(x) \equiv \sum_{j \neq i} [\Gamma_d x]_j$. When recognized, a member i proposes some $x_i \in \arg \min_{x \in Z_d} \beta_{-i}(x)$.

Proof. Given Lemma 1, we have that for $x \in S(\pi_{i,d}^*)$, $u_i(x) = B - \beta_{-i}(x)$. \square

With the above lemmas in hand, it is now possible to proceed to the proof of the main result in Proposition 2.

Proposition. *Let $d^+ > d^-$. If $C^+ \supset C^-$ where $C^- \in W_{d^-}$, then $C^+ \notin W_{d^-}$ implies $C^+ \notin W_{d^+}$.*

Proof. Suppose not. Let j be a member who is in C^+ but not in C^- . Lemma 2 implies $\min_{x \in A_{C^+, d^-}} \beta_{-i}(x) \geq \min_{y \in A_{C^-, d^-}} \beta_{-i}(y)$ for $i \in C^-$. By Lemma 3, the fact that $C^+ \notin W_{d^-}$ implies the inequality is strict. This implies that there exists some $x \in A_{C^-, d^-} \cap \{z \in X_{d^-} : \beta(z) = B\}$ which all members of C^- strictly prefer to any element of $A_{C^+, d^-} \cap \{z \in X_{d^-} : \beta(z) = B\}$, and therefore to any element of $A_{j, d^-} \cap \{z \in X_{d^-} : \beta(z) = B\}$.

Consider some $y \in A_{j, d^-} \cap \{z \in X_{d^-} : \beta(z) = B\}$. Such an element exists by **(A1)**. By the preceding argument, $\forall i \in C^-, x P_i y$. Applying the definition of the utility function, we have:

$$\begin{aligned} [\Gamma_{d^-}(x - y)]_i &> 0, \forall i \in C^- \\ \Rightarrow [\Gamma_{d^+}((x, 0) - (y, 0))]_i &> 0, \forall i \in C^- \end{aligned}$$

Furthermore, $\beta(x) = \beta(y) = B$, so $\beta(x - y) = 0$. Now consider any $z \in S(\pi_{d^+}^*) \cap A_{C^+, d^+}$, which exists by $C^+ \in W_{d^+}$. Given the definition of the choice spaces, $(y, 0)$ and $(x, 0)$ are in X_{d^+} . For any such z , $z + \epsilon((x, 0) - (y, 0))$ for some small scalar ϵ is both feasible and strictly preferred by all members of a decisive coalition, which is inconsistent with equilibrium. □

Proof of Proposition 3

Proposition. *Suppose $d = \text{rank}(\Gamma_d) < n$. There exists a partition of N into d subsets $\{N_1, N_2, \dots, N_d\}$*

such that all members of each subset always vote together in every equilibrium.

Proof. Let U be the n -dimensional space of utilities for all members of N , endowed with basis vectors (u_1, u_2, \dots, u_n) . Define V to be the projection $\Gamma_d x$, $x \in X_d$. V is thus the set of feasible utility vectors given Γ_d . By the fundamental theorem of linear algebra, V has dimension d , and therefore there exists a set of basis vectors (e_1, e_2, \dots, e_d) which span it. $V \subset U$, so (u_1, u_2, \dots, u_n) also span V . Any point $v \in V$ can be expressed as:

$$v = \sum_{k=1}^d a_k(v) e_k$$

or:

$$v = \sum_{k=1}^n b_k(v) u_k$$

This implies we can find a partition of N into d disjoint subsets $\{N_1, N_2, \dots, N_d\}$ (i.e., $N = \bigcup_{k=1}^d N_k$ and $N_k \cap N_l = \emptyset$ for any $k \neq l$) such that:

$$e_k = \sum_{l \in N_k} c_l u_l$$

For some constants $c_l > 0$. Then,

$$v = \sum_{k=1}^d a_k(v) e_k = \sum_{k=1}^d \left[a_k(v) \sum_{l \in N_k} c_l u_l \right]$$

This implies that for any $i, j \in N_k$:

$$u_i(v) = a_k(v)c_i$$

$$u_j(v) = a_k(v)c_j$$

In the bargaining model, the claim “i and j always vote together” is equivalent to the statement $u_i(v) \geq v_i \Leftrightarrow u_j(v) \geq v_j$. To prove this, note that:

$$u_i(v) \geq v_i \Rightarrow a_k(v)c_i \geq \sum_l \rho_l \int_V [a_k(z)e_k]_i \pi_d^*(dz)$$

Substituting for $a_k(v)$, we get:

$$\frac{c_i}{c_j} u_j(v) \geq \sum_l \rho_l \int_V [a_k(z)e_k]_i \pi_d^*(dz)$$

And multiplying through by $\frac{c_j}{c_i}$ gives the necessary equivalence. Reversing the indices proves the “only if” part of the claim. \square

Proof of Proposition 4

Proposition. *Define $q = \min_{D \in \mathcal{D}} |D|$. If $\text{rank}(\Gamma_d) > n - \frac{q}{2}$, all coalitions in W_d are minimal winning, i.e. $C \in W_d \Rightarrow |C| = q$. If $\text{rank}(\Gamma_d) \leq n - \frac{q}{2}$, there exist equilibria in which non-minimal winning coalitions occur with positive probability.*

Proof. Given Proposition 3, it suffices to consider possible combinations of the subsets in the partition $\{N_1, N_2, \dots, N_d\}$. A necessary condition for a non-minimal winning coalition is that the second smallest subset N_k included in the winning coalition must be at least size two; otherwise **all** subsets

in the winning coalition, if recognized, could do strictly better by dropping one of the size-one groups, and this bill would still have sufficient votes to pass. Constructing a minimal winning coalition from subsets of size 2 or larger is possible only if there are at least $\frac{q}{2}$ such subsets. This implies a maximum total number of subsets of $n - q + \frac{q}{2} = n - \frac{q}{2}$. \square

B Vote-buying variant

The results of the theoretical section apply equally well to a vote-buying model where a member with monopoly proposal power attempts to secure passage of a bill by distributing side-payments to members. Suppose that every member i 's utility value of passing some bill is given by v_i . Suppose, as in the basic model, that the proposer can select a vector $x \in X_d$ subject to the budget constraint B , and members' utilities from this side-payment are given by $\Gamma_d x$. A member's total utility is thus $v_i + [\Gamma_d x]_i$ if the package of bill and side-payments passes, and 0 otherwise.

This model is essentially identical to the bargaining model described in the main text, with the simplifying feature that the endogenous continuation value in the bargaining model is replaced with the constant value v_i . Propositions 2, 3 and 4 go through unchanged in this setting. With $v_{i,d} = v_i$ (a constant), we have the useful property that if $x \in A_{C,d}$ then $(x, \epsilon) \in A_{C,d'}$ for any $d' > d$, any coalition C , and any vector ϵ . Under the assumption that the proposer attempts to minimize total side-payments conditional on passing the bill, this property allows us to state the following additional comparative static on d :

Proposition 5. *Let $d^+ > d^-$, and choose any $x \in \pi_{d^-}^*$, $y \in \pi_{d^+}^*$. Then $\beta(y) \leq \beta(x)$.*

Proof. Given the structure of the choice sets, $(x, 0)$ is available in X_{d^+} . Further, the property that continuation values are constant implies that if $x \in A_{C,d^-}$ then $(x, 0) \in A_{C,d^+}$. So $(x, 0)$ is

both available, feasible (since $\beta((x, 0)) = \beta(x) \leq B$ by the fact that $x \in \pi_d^*$), and could pass with the support of a decisive coalition. Under the assumption that the proposer attempts to minimize total side-payments conditional on passing the bill, $y \neq (x, 0)$ can be chosen only if $\beta(y) \leq \beta((x, 0)) = \beta(x)$. \square

Proposition 5 gives us the intuitive result that the proposer’s cost of passing a bill decreases when side payments can be more accurately targeted. Cost is minimized when the full utility space is available, e.g. $d \geq n$.

C Predictability of spending program allocations

One possible interpretation of the model is that the elements of the proposal vector represent the budget allocated to various pre-existing distributive programs whose expected distribution of funds to districts is predictable and known in advance. This interpretation would be appropriate if the expenditure profile of each program were stable over time, that is, if the “recurrence rate” (Stein and Bickers 1997) - the probability that a district which receives benefits from program X in year Y also receives benefits from program X in year $Y + 1$ - were high. To assess the validity of this assumption, figures 3a and 3b show two measures of within-program stability.

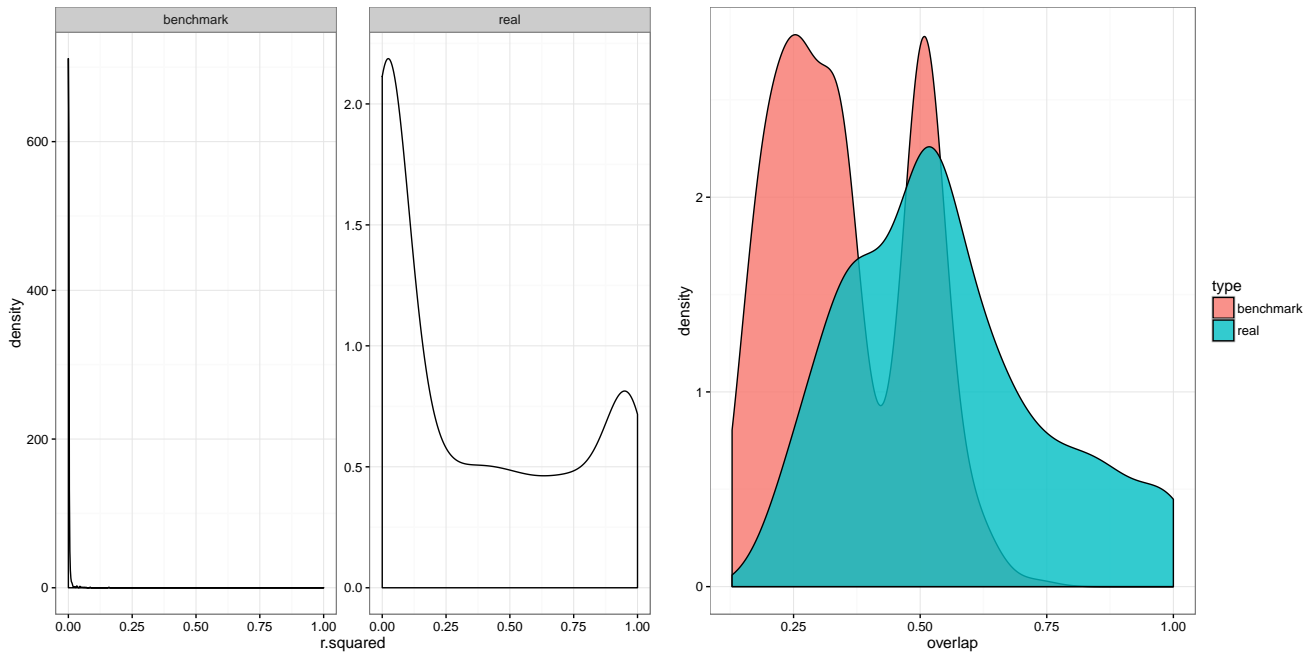
Figure 3a plots the distribution of R^2 values in regressions of each district’s allocation from a given program in year y on the same district’s allocation from the same program in year $y - 1$. There is one such regression for each program in the USAspending / CFDA data; the figure plots the kernel density estimate of the density of R^2 values over all programs. For comparison purposes, the right panel of the figure shows the result of the same computation when the actual program allocations are replaced with a random baseline in which each program’s allocations are drawn from

a Bernoulli-exponential mixture distribution with the same expected number of districts benefiting and same expected per-district benefit level as the actual program. Unsurprisingly, the baseline values are close to a point mass at 0, whereas a significant mass of the real distribution exceeds 0.75; in other words, there are many programs for which last year’s allocation is a very good predictor of this year’s.

Figure 3b uses a different measure of predictability that incorporates only the set of districts benefiting, not the benefit levels, analogous to the “recurrence rate” of Stein and Bickers (1997). For every program-year, I plot the fraction of districts which are ever (in any year) observed receiving benefits from that program, which receive benefits from that program in that year. By construction, this measure ranges from 0 to 1: it equals 1 if and only if the same set of districts benefit from the program in every year. I take the average of the yearly values within program across all years, and plot the resulting distribution of (program-level) overlap scores. Again, the actual distribution has a substantial mass close to 1. Because this measure will be nonzero even for completely random allocations, for comparison I plot the distribution of overlap scores resulting from the baseline Bernoulli-exponential mixture distribution with the same expected number of districts benefiting and same expected per-district benefit level as the actual program.

D Estimation Details

In this section I detail the procedure to estimate the dimension of the choice space, d . The procedure involves, for each bill in the sample, constructing the mapping matrix Γ for a given dimension, drawing a large number of possible values of the equilibrium policy vector x^* , and solving the linear program describing the proposer’s objective (given in equation 4) for each one. I describe each step



(a) Distribution of R^2 values in regressions of each district's allocation from a given program in year y on the same district's allocation from the same program in year $y - 1$. The distribution is compared to a baseline where each program's allocations are drawn at random from a Bernoulli-exponential mixture distribution with the same expected number of districts benefiting and same expected per-district benefit level as the actual program.

(b) Distribution of year-to-year overlap in program expenditure. Overlap is defined as the fraction of districts which received benefits from a program in any year which receive benefits in some particular year; overlap values are then averaged within program over all years in the sample. The distribution is compared to a baseline where each program's allocations are drawn at random from a Bernoulli-exponential mixture distribution with the same expected number of districts benefiting and same expected per-district benefit level as the actual program.

in turn.

D.1 Construction of Γ for each bill

The columns of the matrix Γ are a selection from a set of 160 district-level attributes described extensively in section E.3; the set includes a variety of demographic, geographic, and economic characteristics collected by the federal government. For each bill, I rank all 160 attributes according to the proportion of variance in spending outcomes among districts receiving positive allocations from that bill that is explained by the attribute. The matrix Γ_d for bill b , then, contains the $d - 1$ attributes with greatest explanatory power for bill b plus a column of ones (a constant term).³¹

I account for the fact that the distribution of outcomes involves a significant fraction of zeros by using the model presented in Table 1, column (5) to predict a probability p_{ij} that district i receives any grants as a result of bill j . This probability p_{ij} then multiplies the i -th row of the matrix Γ for bill j . I.e., the i -th row of Γ now gives the *expected* benefit to district i from changes in any of the formula elements.

D.2 The objective function

Equation (4) suggests a formulation of the objective as a function of the parameter d . At the true value of Γ_d , districts outside of the winning coalition get the minimum possible distribution given Γ_d and the constraint that all members of the winning coalition must get at least their observed distribution. To estimate d , then, we should select the value that, on average, yield optimal solutions that are close to the observed distribution.

³¹In other words, although the complete set of attributes is common to all bills, the attributes included in Γ_d for any given value of d will vary by bill.

I implement this idea by following the method of simulated moments (MSM) approach of McFadden (1989). This approach yields a GMM objective function which is somewhat time-consuming but straightforward to compute. It involves sampling from the feasible set of policy vectors x given Γ_d ,³² solving the linear program of equation (4) for each one, and computing an average deviation from the observed distribution.

To define the MSM objective function, first define for a given bill b the vector $\Delta^b(\Gamma)$ with N components indexed by i :

$$\Delta_i^b(\Gamma) = u_i^b - \tilde{u}_i^b(\Gamma) \quad (4)$$

Where \tilde{u}_i^b is the optimal allocation to district i in bill b given by the solution to the problem of equation (4); and u_i^b is the expected allocation to district i that would result from bill b given the bill's equilibrium policy vector x_b^* . Let l^b be a vector that takes value $l_i^b = 0$ if i is in the winning coalition on bill b , and $l_i^b = 1$ otherwise. Then, at the true value of Γ ,

$$E\left[\sum_i l_i^b \Delta_i^b\right] = 0 \quad (5)$$

Hence, we have a moment condition which is zero at the true value of Γ . Because the policy vector x_b^* is unobserved, neither the expected allocation u_i^b nor the proposer-optimal allocation \tilde{u}_i^b can be observed directly. I construct a simulator for this moment by drawing S values of x_b^* from the feasible set of policy vectors given Γ_d .³³

“Feasible” here means that the the vector expends the full budget of bill b and gives every

³²The feasible set is the set of values of x that, given a mapping matrix Γ_d , yields non-negative distributions for every member, and satisfies the overall budget constraint that the bill not expend more than the total observed amount.

³³In the results presented in the following sections, I will set $S = 5000$.

district a non-negative expected allocation. The sampling is accomplished by constructing a basis for the null space of the $(1 \times d)$ matrix consisting of the column sums of Γ , and sampling, at random, combinations of the vectors in the basis which are then added to a feasible starting point. I.e., given some feasible $x_{b,0}$, I construct a new sample $x_{b,s}$ as:

$$x_{b,s} = x_{b,0} + \Omega_{b,d} w_s \tag{6}$$

Where $\Omega_{b,d}$ is a $(d \times d - 1)$ matrix whose columns are a basis for the null space of the column sums matrix for bill b , and w_s is a vector of $d - 1$ weights drawn from a uniform distribution. This works because the basis vectors in the null space of the column sum matrix are, by construction, zero-net-cost, and the ranges for the uniform draws can be chosen such that the result respects the non-negativity constraints for every district. Hence, if $x_{b,0}$ is feasible, then $x_{b,s}$ is also feasible. This process is repeated 5000 times to generate 5000 samples of $x_{b,s}$.

The simulated moment \hat{m}_1^b for each bill b is thus defined by:

$$\hat{m}_1^b = \sum_i \frac{1}{S} \sum_{s=1}^S l_i^b \Delta_{i,s}^b(\Gamma) \tag{7}$$

A second moment arises from a condition imposed by the fact that Γ defines a linear mapping from X to U , which is that the equilibrium distribution must lie in the column space of Γ . In real data with any amount of measurement error, this condition will never be exactly satisfied unless $d = N$. Similar logic to that above implies that we should try to get the distance from the observed allocation to the column space as small as possible. In other words, we should search for values of d that could have produced distributions of outcomes similar to the ones we observe in the data. To achieve this I construct another moment equal to the average difference between the observed

outcome for district i on bill b and the projection of the vector of outcomes onto the column space of Γ :

$$\hat{m}_2^b = \sum_i u_i^b - (\Gamma(\Gamma'\Gamma)^{-1}\Gamma'u^b)_i \quad (8)$$

Putting the two components together as $\hat{m}^b = \begin{bmatrix} \hat{m}_1^b \\ \hat{m}_2^b \end{bmatrix}$, we have a total of 2 moments. Since there is only one parameter (d), the system is overidentified, and estimation must be by GMM. The GMM objective function is:

$$Q_B(d) = \left[\frac{1}{B} \sum_{b=1}^B \hat{m}^b \right]' W_B \left[\frac{1}{B} \sum_{b=1}^B \hat{m}^b \right] \quad (9)$$

Where b indexes bills, B is the total number of bills in the sample, and W_B is a 2×2 positive definite weighting matrix. Because the two moments are similarly scaled and I have no *a priori* reason to prefer one over the other, I choose W_B to be the two-dimensional identity matrix.³⁴ The objective function is thus just the sum of squares of the sample average moment vector $\frac{1}{B} \sum_{b=1}^B \hat{m}^b$.

One complication to the GMM approach is that the parameter d (the dimension of the choice space) takes only integer values. Hence, it is inappropriate to use standard gradient- or simplex-based methods to find the minimum of the objective function. I circumvent this problem by noting that the possible values of d fall into a finite and known range defined by Proposition 4, namely: $d \in \{2, 3, \dots, n - \frac{q}{2}\}$, where q is the minimum coalition size needed to pass a bill. It is, therefore, possible to simply compute values of the moments, and thus objective function values, for each candidate value of d . Comparing the function values across candidate values of d and selecting the

³⁴Other reasonable choices for W_B , such as using an estimate of the inverse covariance matrix of the sample moments, yield similar results.

minimum of these yields my estimate of d .

D.3 Estimation of d

As described above, computing a point estimate for d simply requires finding the minimum of the function value over all values of d computed in the previous section. To construct confidence intervals, I employed a simple nonparametric bootstrap technique, resampling bills with replacement and computing the minimizing value of d as described above for each bootstrapped resample. I repeated this procedure for all 1000 bootstrap resamples, and report the central 95% interval of the resulting (discrete) distribution as the confidence interval for d .

E Data

Estimating the effective degree of formula complexity d present in real appropriations legislation requires three distinct data sources. First, **spending** data on funds appropriated by a legislature that allows me to identify both the district in which funds are spent and the authorizing legislation. Second, **voting** data that reveals each district’s position on each spending bill. And third, observable **characteristics** of the districts that form the basis of the policy mapping matrix Γ . I focus on the US House of Representatives and use spending, voting, and characteristic data defined at the congressional district level. Each of the three categories of data are described in detail below.

E.1 Spending

Spending data comes from the US Department of Treasury’s USAspending.gov, a website established by the Federal Funding Accountability and Transparency Act of 2006 in order to “give the American

public access to information on how their tax dollars are being spent.”³⁵ USAspending data covers federal contract, grant, loan, and other financial assistance awards of more than \$25,000, identifying the amount of the award, the location of the recipient and the place of performance. I downloaded USAspending data covering fiscal years 2005 through 2011, covering the sessions of the 109th to 111th Congresses.

USAspending records domestic recipients of federal assistance (including loans and cash grants to individuals and companies as well as assistance to state and local governments) from 195 federal agencies, covering a total of 639 distinct assistance programs. Critically for my purposes, USAspending provides geographic identifying information, allowing me to determine the congressional district of each recipient. I use the recipient’s congressional district except where USAspending identifies the recipient as a state government or state-controlled institution of higher education, in which case I use the “place of performance” congressional district. This is to ensure that, for instance, an NSF grant to researchers at the University of Alabama in Huntsville counts for the district containing Huntsville (AL-05) and not that containing the headquarters of the Alabama university system in Tuscaloosa (AL-07).³⁶

USAspending also provides two additional identifiers on each grant which prove important: the federal agency that made the grant, and the specific program under which the grant was made. The program numbers correspond to programs listed in the Catalog of Federal Domestic Assistance (CFDA), produced by the US Census department.³⁷ The CFDA contains a field that indicates,

³⁵<https://www.usaspending.gov/about/usaspending/Pages/default.aspx>

³⁶In cases where USAspending did not provide congressional district information directly, I used the ZIP code or county of recipient or place of performance to identify the congressional district. As some counties / ZIP codes cross district boundaries, I used the Census Department’s ZIP- or county-to-district relationship files to allocate the amounts for these entries according to the fraction of the total population of the ZIP / county living in each congressional district.

³⁷The CFDA has been used to track the histories of federal spending programs by, among others, Stein and Bickers (1997) and Berry, Burden, and Howell (2010a).

for each program, the legislation that authorized the creation of the program. For example, the CFDA allows me to determine that funding for program 10.025, “Plant and Animal Disease, Pest Control, and Animal Care,” was authorized by the Plant Protection Act, Public Law 106-224. Grants made by agencies under program 10.025 subsequent to the passage of 106-224 that show up in USAspending can therefore be traced back to Public Law 106-224.³⁸ In cases where multiple bills authorize appropriations for the same program - as occurs when a reauthorization appropriates new funding for a continuing program established in a previous bill - I attribute all amounts distributed after the passage of the bill and before the passage of the *next* bill appropriating for the same program to the prior bill.

Finally, USAspending classifies expenditures as one of several assistance types. The distribution of expenditures by assistance type is shown Table 3. The largest category, making up approximately half of the total expenditures in USAspending, are direct payments to individuals, corporations and institutions, a category that includes agricultural subsidies like the Dairy Product Price Support program as well as federal student aid such as the Pell Grant program. The second largest category are formula grants, which are allocated according to statutory formula and typically, though not always, directed to state or local governments or non-profit institutions. Block grants, similarly, are directed to states according to formula but with fewer restrictions on how they may be used or allocated within the state. Project grants and cooperative agreements are grants for specific, localized projects - such as investments in new public transportation infrastructure - either allocated through a competitive application process or “earmarked” for a specific project specified in the statute. Loans, loan guarantees and insurance make up the remainder.³⁹

³⁸Public Law numbers are sequential; 106-224 indicates the 224th law passed by the 106th Congress.

³⁹As Stein and Bickers (1997) note, there can be huge differences between the potential liability incurred by a federal loan guarantee and the amount that is actually disbursed; a billion-dollar loan guarantee may ultimately require no expenditure at all if the debtor does not default. I follow Stein and Bickers (1997) in using the amounts

Type	Total.Allocation (\$B)	# Programs	# Bills
Direct Payment	638.89	69	14
Formula Grant	389.25	118	24
Project Grant	185.64	367	44
Cooperative Agreement	42.59	234	32
Block Grant	33.84	69	19
Other	13.06	39	8
Loans and Loan Guarantees	2.74	20	5
Insurance	0.15	4	3

Table 3: The distribution of federal outlays in the 51 public laws in the sample. Dollar amounts are in billions. The number of programs and number of bills columns count the number of programs or bills, respectively, that provide any assistance of the listed type.

I aggregate the spending data by district and authorizing public law, such that each entry in my final spending dataset contains the total amount spent in a single district during FY2005-2011 that was authorized by a single law. I include only laws passed by the 109th through 111th Congresses, for two reasons. One, there was no significant redrawing of districts during this period, such that the district boundaries in these three Congresses are the same as the districts identified in USAspending. Two, my spending data covers only 2005-2011 and as such is only likely to be a good measure of the short-term consequences of recently passed bills.

This initial screen yielded a total of 58 separate bills in this period cited as authorizing spending on CFDA-listed programs. Of these, I excluded from the sample seven bills that show up in the USAspending data but that are inappropriate to include. I dropped, for example, defense authorization bills, whose primary distributive consequences do not show up in CFDA-listed programs. I also dropped several bills with strong ideological content and minor distributive consequences, including the Emergency Economic Stabilization Act of 2008, the Dodd-Frank Wall Street Reform Act, and the Children’s Supplemental Health Insurance Reauthorization Act.

Of the remaining 51 bills, I retained those which authorized positive amounts in either the appropriated rather than the nominal value of the loan, with the result that many loan guarantee programs have zero expenditure.

formula grant, direct payment, or block grant categories. The model is most applicable to these types of spending, which are allocated either by explicit statutory formula or according to well-defined, objective criteria.⁴⁰ As Table 4 shows, bills authorizing some of at least one of these types of spending account for the vast majority of all spending in the data; bills authorizing only project grants or loans are several orders of magnitude smaller in terms of budgetary impact. However, I include *all* spending types (not just formula grants, block grants, and direct payments) in the calculation of total district-level spending authorized by each bill, for two reasons. One, members vote on the full package of appropriations authorized by a bill, not separately on the formula and project grant components; a yes vote on final passage can be interpreted as approval of the bundle but not necessarily of any of its components individually. Two, to the extent that congressional leaders can use project grants or other forms of spending to target important swing districts, this flexibility should appropriately be counted in any estimate of the dimensionality of congressional proposals. Including only the formula components would tend to bias such estimates downward, relative to the actual menu of proposals available to congressional leaders.

The final dataset thus defined contains 31 bills, with 435 rows for each bill corresponding to the total district-level spending that the bill authorized. The bills included in the sample are given in Table 5. This table shows the number of districts voting yes on each bill, the number of districts receiving positive spending allocations, as well as the amount of funding received by the district with the median and highest spending, respectively, that resulted from the bill.

⁴⁰For example, Pell Grants are available to full-time undergraduate students meeting family income limits, and milk price support payments are available to owners of dairy cattle.

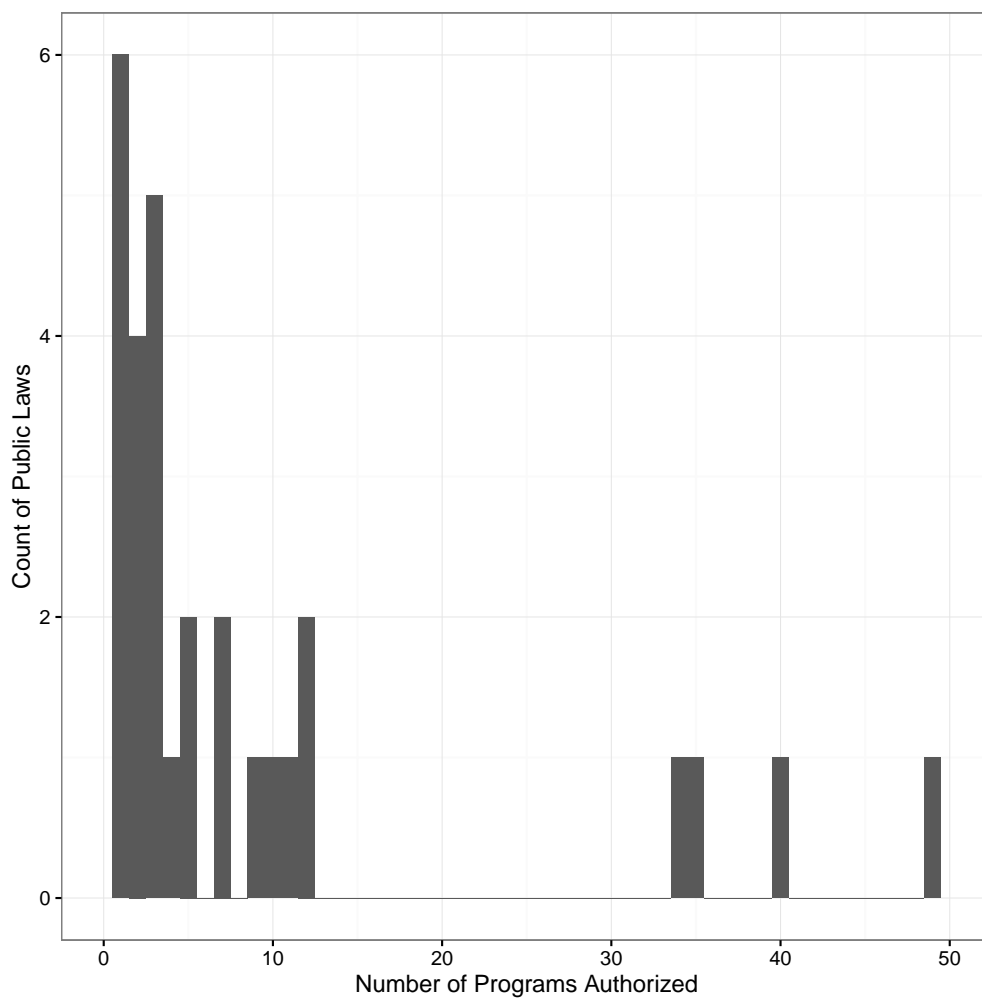


Figure 4: The distribution of the number of distinct (CFDA-recognized) programs for which funds are authorized by bills in the sample.

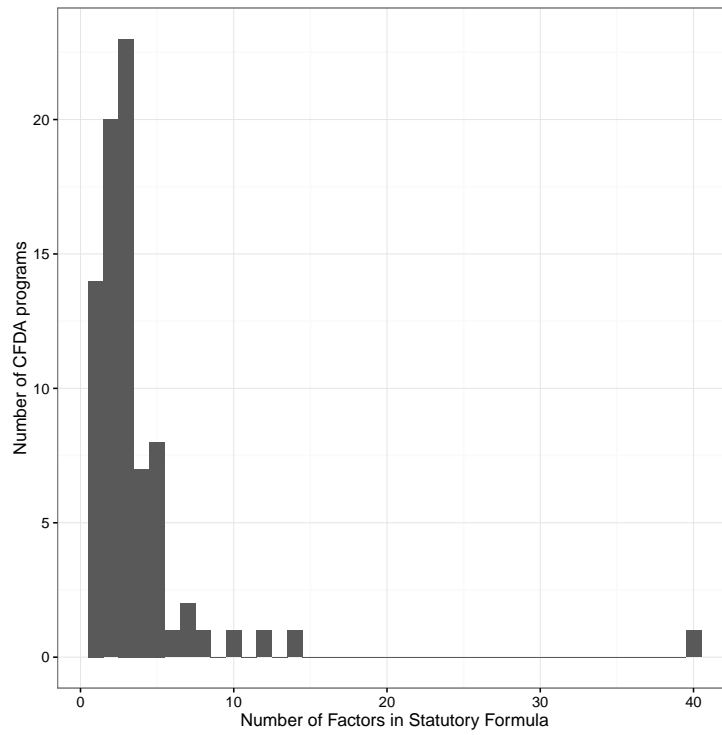


Figure 5: Distribution of CFDA programs allocated by statutory formula, by the number of factors referenced in the allocation formula.

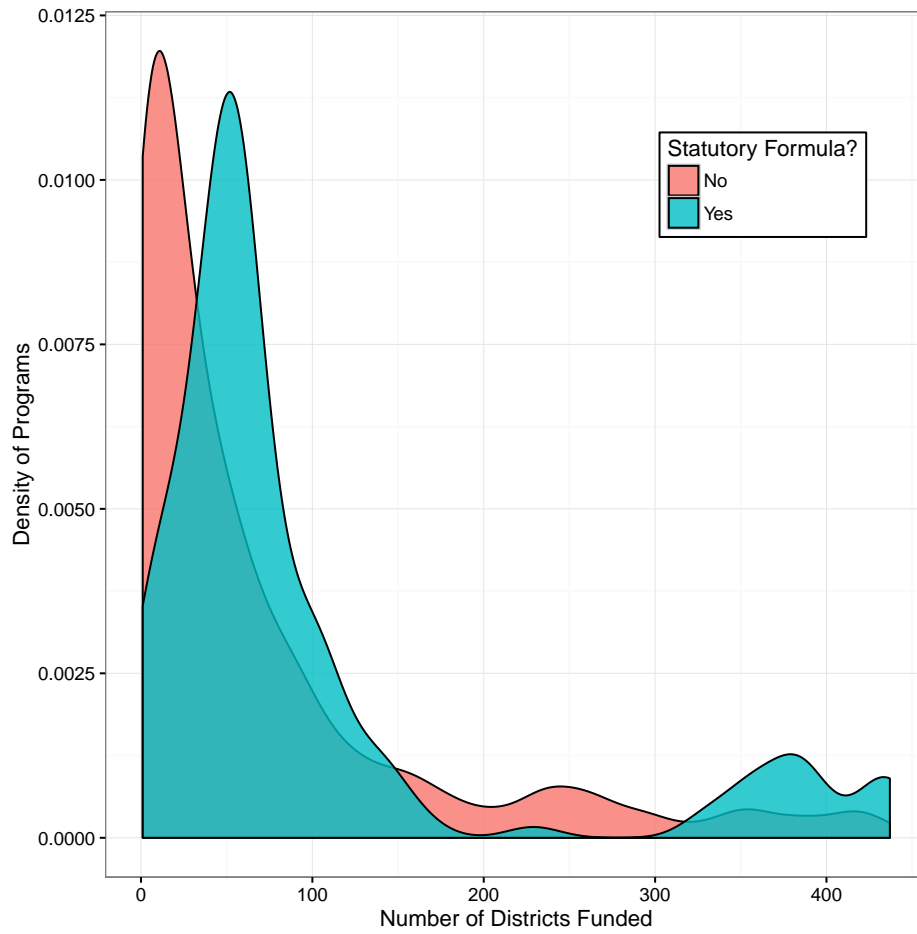


Figure 6: Kernel density estimate of the number of districts receiving funds from CFDA programs, among programs with and without statutory formulas.

Sample	N	Only Specified Categories		All Spending	
		Median	Total	Median	Total
Formula	21	0.30	389.26	3.62	1246.94
Formula + Direct	26	0.44	1028.15	1.91	1249.00
Formula + Direct + Block	31	0.69	1061.99	2.39	1301.83
Others	20	0.08	4.34	0.08	4.33

Table 4: Median and total spending levels among different sample definitions. Dollar amounts are in millions. 'Formula' is the set of bills which authorize positive levels of Formula Grants; Formula + Direct is the set of bills which authorize positive amounts in either the Formula Grant or Direct Payment category; Formula + Direct + Block is the set of bills which authorize positive amounts in either Formula Grant, Direct Payment or Block Grant categories. 'Others' is the set of bills which do not meet any of these criteria. The third and fourth columns are the amounts authorized in the categories used to define the subsets, e.g. only Formula Grants for the Formula Grants sample. The fifth and sixth columns include all spending types.

E.2 Votes

Roll-call vote data for the 109th, 110th, and 111th Congresses was downloaded from Keith Poole's voteview.com. To extract only the votes on the bills in the sample defined in table 5, I consulted the Library of Congress' THOMAS database.⁴¹ THOMAS' Bills and Resolutions database provided a list of all House bills and the last major action on each - including, when applicable, the public law that the bill ultimately became. Using this database I was able to match, for instance, the 109th Congress' H.R.1270 into Public Law 109-6.

I then searched for the house resolution number in the roll-call vote database, and extracted the appropriate *final passage* vote. This vector of votes was matched to the corresponding law in the sample. I coded all yea-voters on a given bill as the winning coalition for that bill, and all non-yea-voters (i.e. those who voted nay or abstained) as not in the winning coalition. Table 5 provides summary measures of the coalition sizes for each bill.

⁴¹<http://thomas.loc.gov>

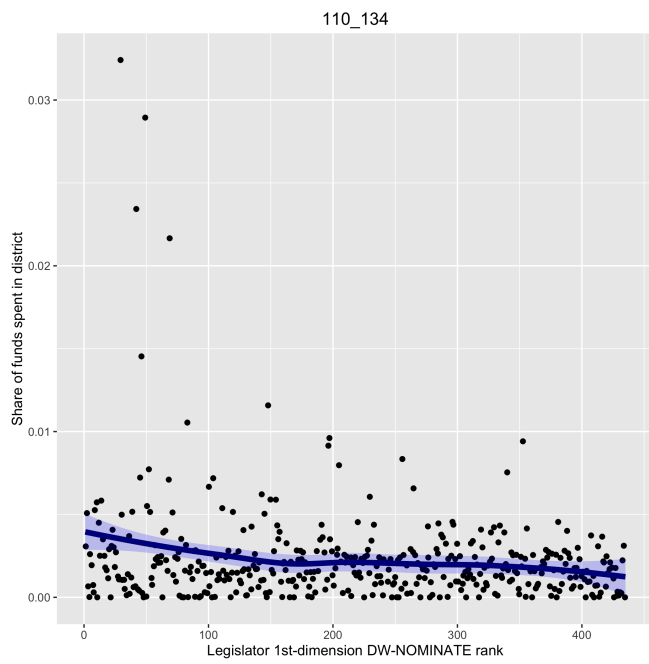
A visual inspection of the data shows that the inverted-U-shaped distribution predicted by standard vote buying models is difficult to discern, as even very extreme members get significant distributions of funds in most bills in the sample. Figures 7a and 7b show the allocation of funds in two example bills, the 110th Congress' Improving Head Start for School Readiness Act of 2007 (Public Law 110-134) and the 111th Congress' American Recovery and Reinvestment Act (ARRA, also known as the Recovery Act, Public Law 111-5). The y-axis measures the share of the bill's total expenditures spent in each district; points are individual districts, arranged by DW-NOMINATE rank from most liberal to most conservative.

The Democratic congressional majority responsible for both 110-134 and 111-5 appears to have achieved a slight Democratic tilt to the appropriations, with particularly conservative Republican districts getting below-average shares in both bills. But marginal Democratic districts - the kind of swing-district votes which the leadership would need to secure to ensure passage of the bill - do not appear to do any better than the most liberal members of the caucus; if anything, they do worse. This pattern of results is consistent with the vote-buying variant of the model presented here: proposers (in this case, the Democratic party leadership) attempt to target funds to marginal districts, but due to formula restrictions end up buying off inframarginal districts as well.

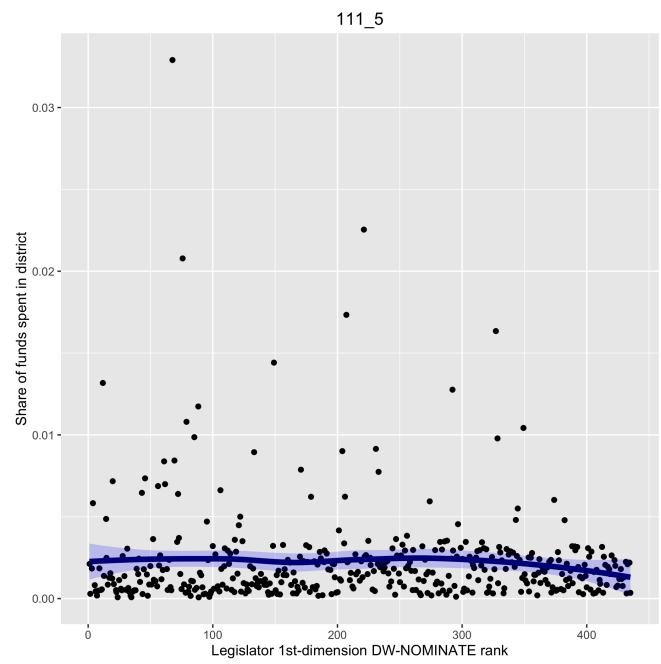
E.3 Characteristics

To construct the policy mapping matrix Γ , I used data from two primary sources: demographic data from the US Census, and geographic data from the US Geological Survey (USGS) and the National Oceanic and Atmospheric Administration (NOAA).

Demographic variables are derived from the the Census Department's 2010 Census of Population and Housing 110-112th Congressional District Summary File, as provided by the National



(a) Public Law 110-134



(b) Public Law 111-5

Figure 7: Distribution of funds by ideology in two example bills.

Historical Geographic Information System (NHGIS) (Minnesota Population Center 2011). The census variables include information on each district’s urban/rural makeup, age distribution, income distribution, poverty rates, household composition, occupational and employment distribution, education levels, housing stock, and numerous other factors. The complete set of variables is listed in Table 7.

Geographic variables came from the GTOPO30 Digital Elevation Model⁴² and National Land Cover Database 2006⁴³ datasets produced by the USGS, as well as the $.25 \times .25$ Unified Precipitation⁴⁴ dataset produced by NOAA. As these datasets are quite high-resolution, I generated summary statistics at the congressional district level according to a process described in detail in Appendix F.

The geographic variables included in the vector of observables for each district are: long-term average monthly precipitation in each month of the year, percentage of the land area covered by each of the land cover types defined by the USGS⁴⁵, percentage of the land area that is impervious (e.g., developed), percentage of the land area covered by tree canopy, and the mean and standard deviation of elevation (in meters). Combining these variables with the Census-derived population and economic variables yields a final set of 160 distinct variables.

F Construction of district geographic characteristics

Estimation of the structural model relies on a set of fixed, known district-level characteristics. The US government releases a huge amount of geographical data (covering features such as topography,

⁴²http://eros.usgs.gov/#/Find_Data/Products_and_Data_Available/gtopo30_info

⁴³http://www.mrlc.gov/nlcd06_data.php

⁴⁴<http://www.esrl.noaa.gov/psd/data/gridded/data.unified.html>

⁴⁵For a list of the land cover types, see http://www.mrlc.gov/nlcd06_leg.php

land cover, and climate), much of which is likely to provide useful information regarding the relative benefits of different types of public spending to different congressional districts. However, the observational unit of the geographical data is not a congressional district; it is typically a grid cell or point, of which a congressional district may contain very many, or none. As a result it is necessary to perform some aggregation and/or interpolation to get the data into the required district level of observation. I describe the process employed for each of the datasets in what follows; all code used to implement the technique is available from the author's website. The source datasets are available freely on the web from the publishing agencies.

F.1 National Land Cover Database 2006

The USGS' National Land Cover Database (NLCD) is a high-resolution raster dataset which splits the US up into a very fine grid of cells (pixels) and assigns each pixel to one of 20 land cover categories, such as urbanized area, farmland, forest, etc.⁴⁶ Each pixel is 30 meters square, implying that a large congressional district (such as the single at-large district of Wyoming) might contain on the order of 10 billion pixels. As the raster datasets are huge image files for which read/write access is slow, it is computationally prohibitive to read billions of pixels for each congressional district. Instead, I estimate district-level means with a simple geographically stratified sampling technique.

The method employed is to first construct the rectangular bounding box of each district, using the 110th Congressional district shapefiles provided by the US Census Bureau. I then uniformly sample 10,000 latitude/longitude points from within this bounding box, access the pixel at that location, and add its value to the vector of observations for the district in which the point falls.⁴⁷

⁴⁶For a full list of the categories, see http://www.mrlc.gov/nlcd06_leg.php

⁴⁷This may not be the same as the original district from which the bounding box was constructed, as Congressional districts are rarely rectangular.

After repeating this process for each of the 435 Congressional districts (leading to a total sample of 43,500,000 points), I compute the proportion of observations in each district that fall into each of the 20 land cover types. The result is a 435×20 matrix of estimated proportions, each row of which sums to 1.

I repeat the identical procedure for the NLCD's Percent Developed Impervious and Percent Tree Canopy datasets, again using 10,000 sample points per district. The results are estimates of the percentage of land in each district that is covered by impervious materials and tree canopy, respectively.

F.2 GTOPO30

The USGS's GTOPO30 dataset is a raster dataset containing measurements of elevation, covering the globe at a resolution of 30 arc-seconds (approximately 900m at the equator). While not as extreme as the NLCD, the problem of excessive data remains. I use the same technique described in the previous section to estimate mean elevation for each district. Along with the absolute elevation, the hilliness of terrain (e.g., the variability of elevation) may also influence what types of public investments are feasible for a given district. I therefore compute sample standard deviations of elevation for each district along with the sample mean.

F.3 Unified Precipitation Dataset

NOAA's Unified Precipitation Dataset contains monthly long-term average rainfall totals (in inches) for a grid of points covering the US at .25 degrees of latitude / longitude intervals (approximately 27.5 km). This grid spacing is wide enough that the total number of points is manageable to directly compute district means; however, some small urban districts may not contain any grid points. I

deal with this problem by overlaying the grid points onto the district shapefile; for any district that contains one or more grid points, I take the average value of the monthly precipitation variables for each point that falls within the district. For districts that do not contain any grid points, I take the value from the grid point that is closest to the district's geographical centroid.

G Additional tables and model output

This section includes some additional regressions and model output omitted from the main text for brevity. Table 8 presents models analogous to those in Table 1 but with fixed effects at the district level. The comparison here is within-district, asking how the same district's share varies when it is in the winning coalition compared to when it is out. Table 9 presents models analogous to those in Tables 1 and 8, but where the left-hand side is the absolute dollar amount of spending authorized by the bill. Table 10 presents the same analysis as Table 1, but for the Senate: the unit of observation is the bill-state, and the independent variables of interest are the state's Senate delegation's vote on the authorizing legislation. Results are qualitatively similar to those in the main text for the House, although the sample size is smaller because the Senate did not record a roll-call vote on final passage of some of the bills in the sample.

Figure 8 reports the distribution of formula dimension estimates resulting from a bootstrapping exercise described in the results section. And Figure 9 presents a histogram of the frequency of inclusion in simulated formulas for the most common district-level factors.

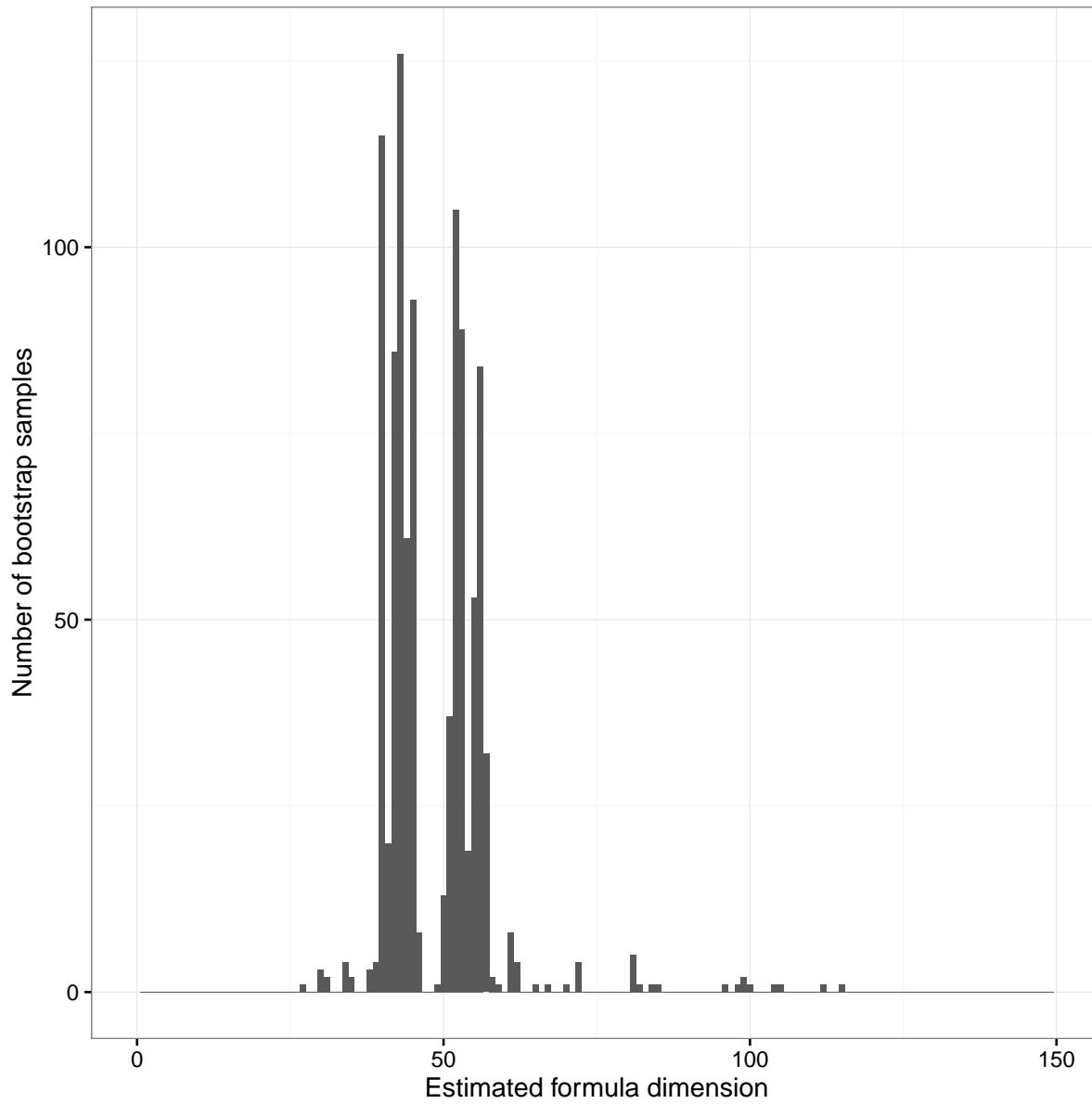


Figure 8: The bootstrapped distribution of formula dimension estimates. The distribution is constructed by resampling with replacement from the set of bills, and computing the minimizing value of d for each resample.

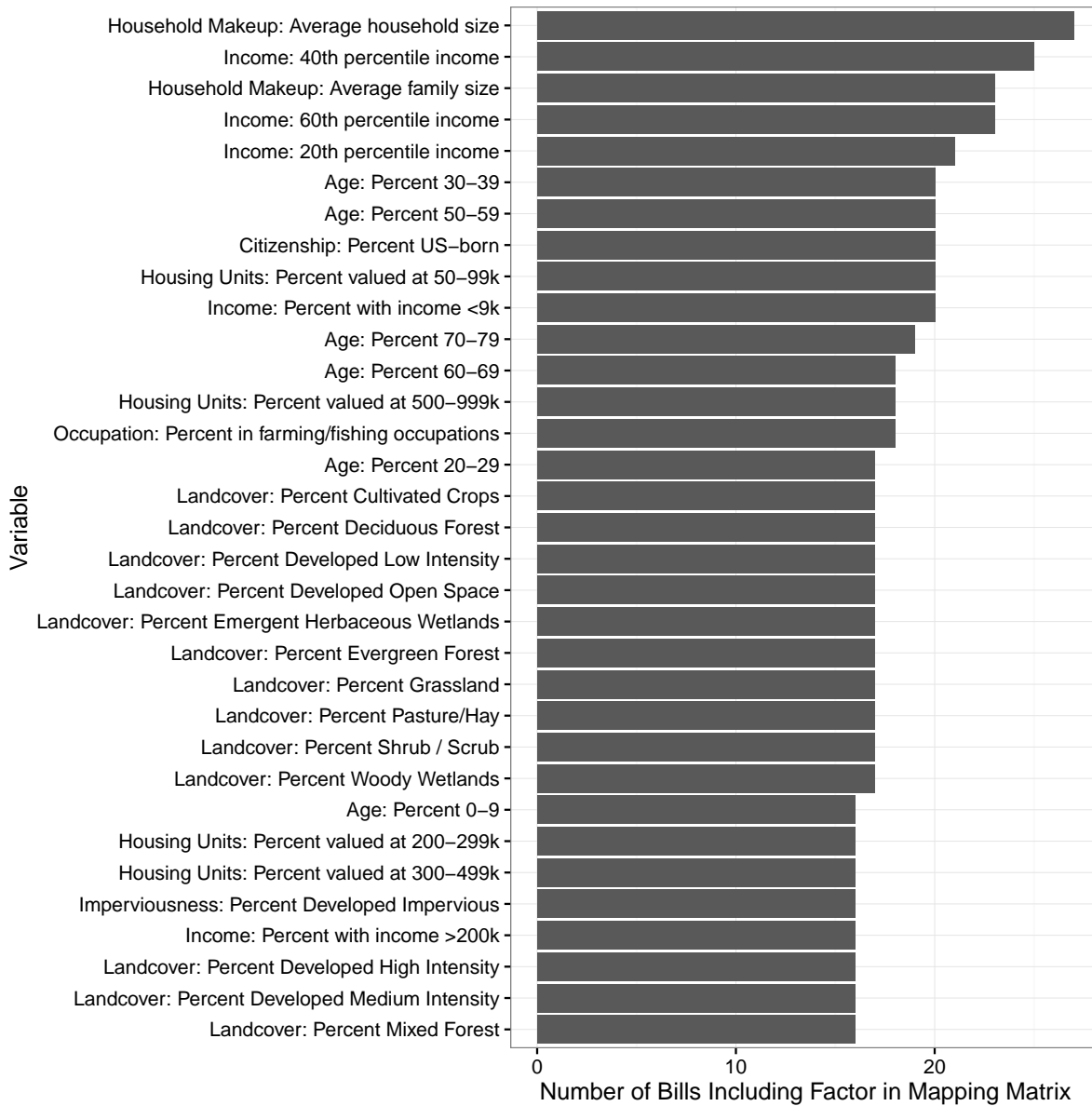


Figure 9: The district-level covariates most commonly included in simulated formulas over the sample of appropriations bills. The horizontal axis gives the number of bills in the sample, out of the set of 31, in which the indicated factor is included in the formula generated by the simulation.

Congress	Public Law	Law Name	Yea Voters	Districts Receiving Funding	Median Distric
109	58	Energy Policy Act of 2005	275	402	
109	59	Safe, Accountable, Flexible, Efficient Transportation Equity Act: A Legacy for Users	412	388	
109	171	Deficit Reduction Act of 2005	212	208	
109	307	Children's Hospital GME Support Reauthorization Act of 2006	421	54	
109	347	SAFE Port Act	409	46	
109	415	Ryan White HIV/AIDS Treatment Modernization Act of 2006	325	259	
109	482	National Institutes of Health Reform Act of 2006	414	265	
110	5	Revised Continuing Appropriations Resolution, 2007	286	374	
110	53	Implementing Recommendations of the 9/11 Commission Act of 2007	371	176	
110	134	Improving Head Start for School Readiness Act of 2007	381	417	
110	140	Energy Independence and Security Act of 2007	314	411	
110	161	Consolidated Appropriations Act, 2008	272	434	
110	206	Traumatic Brain Injury Act of 2008	392	57	
110	246	Food, Conservation, and Energy Act of 2008	317	434	
110	252	Supplemental Appropriations Act, 2008	416	265	
110	275	Medicare Improvements for Patients and Providers Act of 2008	383	100	
110	293	Tom Lantos and Henry J. Hyde United States Global Leadership Against HIV/AIDS, Tuberculosis, and Malaria Reauthorization Act of 2008	303	44	
110	329	Consolidated Security, Disaster Assistance, and Continuing Appropriations Act, 2009	370	180	
110	377	Poison Center Support, Enhancement, and Awareness Act of 2008	403	58	
110	411	Native American Housing Assistance and Self-Determination Reauthorization Act of 2008	333	107	
111	5	American Recovery and Reinvestment Act of 2009	246	435	
111	8	Omnibus Appropriations Act, 2009	237	425	
111	13	Serve America Act	419	128	
111	80	Agriculture, Rural Development, Food and Drug Administration, and Related Agencies Appropriations Act, 2010	263	45	
111	87	Ryan White HIV/AIDS Treatment Extension Act of 2009	408	283	
111	88	Department of the Interior, Environment, and Related Agencies Appropriations Act, 2010	247	64	
111	117	Consolidated Appropriations Act, 2010	221	427	
111	152	Health Care and Education Reconciliation Act of 2010	220	100	
111	212	Supplemental Appropriations Act, 2010	308	6	
111	296	Healthy, Hunger-Free Kids Act of 2010	264	75	
111	312	Tax Relief, Unemployment Insurance Reauthorization, and Job Creation Act of 2010	277	44	

Table 5: The sample of spending bills

Congress	Public Law	Yea	Nay	Abstain
109	58	\$ 29M	\$ 48M	\$ 26M
109	59	\$ 9M	\$ 2.3M	\$ 4.4M
109	171	\$ 869.4K	\$ 1.3M	\$ 3M
109	307	\$ 1.5M	\$ 0	\$ 1.6M
109	347	\$ 131K	\$ 0	\$ 50K
109	415	\$ 3.5M	\$ 12.1M	\$ 6.6M
109	482	\$ 9.8M	\$ 5.6M	\$ 9.1M
110	5	\$ 2M	\$ 1.1M	\$ 1.2M
110	53	\$ 13.6M	\$ 3.2M	\$ 61.2M
110	134	\$ 105M	\$ 59M	\$ 144M
110	140	\$ 32.5M	\$ 9.1M	\$ 14.2M
110	161	\$ 12M	\$ 24M	\$ 30M
110	206	\$ 25K	\$ 0	\$ 28K
110	246	\$ 13.7M	\$ 8.9M	\$ 10.4M
110	252	\$ 2.8M	\$ 1.9M	\$ 2.4M
110	275	\$ 420K	\$ 60K	\$ 146K
110	293	\$ 4.2M	\$ 414.6K	\$ 0
110	329	\$ 816K	\$ 163K	\$ 252K
110	377	\$ 152K	\$ 65K	\$ 202K
110	411	\$ 8.6M	\$ 2.5M	\$ 22.1M
111	5	\$ 2.7B	\$ 2.3B	\$ 2.2B
111	8	\$ 197M	\$ 76M	\$ 133M
111	13	\$ 542K		\$ 145K
111	80	\$ 950K	\$ 372K	\$ 0
111	87	\$ 8.5M	\$ 1.1M	\$ 8.4M
111	88	\$ 187K	\$ 124K	\$ 0
111	117	\$ 7.1M	\$ 4.3M	\$ 15.8M
111	152	\$ 7.3M	\$ 4.9M	\$ 4.8M
111	212	\$ 1.9K	\$ 0	\$ 14.5K
111	296	\$ 28M	\$ 19M	\$ 12M
111	312	\$ 22K	\$ 58K	\$ 0

Table 6: USAspending distributions by vote. The columns labeled Yea, Nay and Abstain give the average USAspending funds received by districts voting Yea, Nay, or abstaining, respectively, on the final passage vote of the indicated public law.

Category	Variables
Urbanity	Percent urban; Percent suburban; Percent rural
Race	Percent white; Percent black; Percent indian; Percent asian; Percent other; Percent mixed; Percent hispanic
Gender	Percent male
Age	Percent 0-9; Percent 10-19; Percent 20-29; Percent 30-39; Percent 40-49; Percent 50-59; Percent 60-69; Percent 70-79; Percent 80-89
Household Makeup	Percent married family households; Percent unmarried family households; Percent single non family households; Percent joint non family households; Average household size; Average family size
Group Quarters Population	Percent of pop. in prison; Percent of pop. in juvenile group housing; Percent of pop. in nursing homes; Percent of pop. in college dormitories; Percent of pop. in military base housing
Housing Tenure	Percent owners; Percent renters
Citizenship	Percent US-born; Percent naturalized citizens; Percent noncitizens
Commutes	Percent driving to work; Percent carpooling to work; Percent taking transit to work; Percent walking to work; Percent telecommuting; Percent commutes <15min; Percent commutes 15-30min; Percent commutes 30-45min; Percent commutes 45-90min
Education	Percent with no high school; Percent high school dropout; Percent high school graduate, no college; Percent with some college, no degree; Percent with bachelors degree only; Percent with postgraduate degree
Poverty	Percent below poverty line
Employment	Percent employed; Percent unemployed
Income	Percent with income <9k; Percent with income 10-14k; Percent with income 15-19k; Percent with income 20-24k; Percent with income 25-29k; Percent with income 30-34k; Percent with income 35-39k; Percent with income 40-44k; Percent with income 45-49k; Percent with income 50-59k; Percent with income 60-74k; Percent with income 75-99k; Percent with income 100-124k; Percent with income 125-149k; Percent with income 150-199k; Percent with income >200k; Percent with social security income; Percent with supplemental security income; Percent receiving public assistance; Percent receiving food stamps; Percent with retirement plan income; 20th percentile income; 40th percentile income; 60th percentile income; 80th percentile income; 95th percentile income
Veterans	Percent veterans
Occupation	Percent in management occupations; Percent in science/engineering occupations; Percent in education occupations; Percent in health occupations; Percent in health support occupations; Percent in protective occupations; Percent in food service occupations; Percent in customer service occupations; Percent in caregiving occupations; Percent in sales occupations; Percent in administrative occupations; Percent in farming/fishing occupations; Percent in construction occupations; Percent in maintenance occupations; Percent in production occupations; Percent in transportation occupations; Percent in moving occupations
Industries	Percent in agriculture industry; Percent in mining industry; Percent in construction industry; Percent in manufacturing industry; Percent in wholesale industry; Percent in retail industry; Percent in transportation industry; Percent in utilities industry; Percent in information industry; Percent in finance industry; Percent in real estate industry; Percent in professional services industry; Percent in management industry; Percent in administrative support industry; Percent in education industry; Percent in health industry; Percent in entertainment industry; Percent in hospitality industry; Percent in other service industry; Percent in public administration industry
Housing Units	Percent in single family homes; Percent in small multi-unit buildings; Percent in large multi-unit buildings; Percent in mobile homes; Median year built; Percent with 0 bedrooms; Percent with 1 bedrooms; Percent with 2 bedrooms; Percent with 3 bedrooms; Percent with 4 bedrooms; Percent with 5 or more bedrooms; Percent valued at 0-50k; Percent valued at 50-99k; Percent valued at 100-149k; Percent valued at 150-199k; Percent valued at 200-299k; Percent valued at 300-499k; Percent valued at 500-999k; Percent valued at >1m; Percent vacant housing units

Table 7: District-level census variables included in the set of district attributes in the simulation.

Table 8: Models of District Funding Receipts, 109-111 Congress

	Amount Received (% of Bill Total)			Any Grants Received		
	(1)	(2)	(3)	(4)	(5)	(6)
No Vote on Final Passage	-0.009 (0.029)	-0.017 (0.040)	-0.020 (0.039)	0.087 (0.072)	0.087 (0.085)	0.088 (0.084)
Abstain on Final Passage	0.108 (0.070)	0.105 (0.068)	0.105 (0.068)	-0.019 (0.036)	-0.020 (0.037)	-0.020 (0.037)
Majority Party		0.080 (0.119)	0.101 (0.119)		-0.034 (0.054)	-0.038 (0.057)
President's Party		0.009 (0.043)	0.004 (0.043)		0.028 (0.030)	0.032 (0.025)
Ideological Extremity		0.253 (0.253)	0.269 (0.249)		-0.023 (0.102)	-0.040 (0.101)
At-Large District			-0.087** (0.035)			-0.012 (0.034)
Senate: Majority Party			0.031 (0.036)			-0.021 (0.040)
Senate: President's Party			-0.065 (0.054)			-0.028 (0.075)
Fixed Effects:	District	District	District	District	District	District
Number of Bills	31	31	31	31	31	31
F-statistic	2.42	2.41	2.41	2.27	2.28	2.28
N	13,485	13,485	13,485	13,485	13,485	13,485
R ²	0.075	0.075	0.075	0.070	0.071	0.072

*p < .1; **p < .05; ***p < .01

Cluster-robust standard errors in parentheses (clustered at the bill level). Bill-clustered standard errors are conservative relative to district-clustered or heteroskedasticity-robust versions. All right-hand-side variables are binary indicators except for Ideological Extremity, which is measured as the absolute deviation between the first-dimension DW-NOMINATE score of the district's representative and that of the chamber median. Dummies for party alignment of the district's Senate delegation are one only if both Senators are in the Senate majority or the president's party, respectively. The "Senate Delegation Split" dummy indicates that the state's Senate delegation consists of one Democratic and one Republican Senator.

Table 9: Models of District Funding Receipts, 109-111 Congress

	Amount Received (\$MM)					
	(1)	(2)	(3)	(4)	(5)	(6)
No Vote on Final Passage	-28.918 (21.202)	-18.006 (11.448)	-19.976 (12.860)	100.936 (101.956)	130.454 (131.927)	130.291 (131.822)
Abstain on Final Passage	-6.761 (9.922)	-4.603 (7.498)	-4.694 (7.543)	-27.248 (34.178)	-16.575 (24.290)	-16.747 (24.494)
Majority Party		6.432 (10.433)	8.801 (8.988)		49.211 (52.547)	50.067 (52.341)
President's Party		8.581 (10.056)	12.916 (14.053)		67.728 (66.092)	66.141 (65.062)
Ideological Extremity		-16.808 (40.760)	-21.922 (45.151)		33.568 (50.959)	46.230 (60.044)
At-Large District		99.592 (83.530)	100.156 (84.292)			
Senate: Majority Party			-25.718 (24.536)			17.439 (19.179)
Senate: President's Party			-23.299 (23.716)			8.960 (8.085)
Senate: Delegation Split			-24.164 (21.093)			6.028 (12.597)
Fixed Effects:	Bill	Bill	Bill	District	District	District
Number of Bills	31	31	31	31	31	31
F-statistic	225.59	200.8	185.66	0.89	0.96	0.95
N	13,485	13,485	13,485	13,485	13,485	13,485
R ²	0.349	0.350	0.350	0.029	0.031	0.031

*p < .1; **p < .05; ***p < .01

Cluster-robust standard errors in parentheses (clustered at the bill level). Bill-clustered standard errors are conservative relative to district-clustered or heteroskedasticity-robust versions. All right-hand-side variables are binary indicators except for Ideological Extremity, which is measured as the absolute deviation between the first-dimension DW-NOMINATE score of the district's representative and that of the chamber median. Dummies for party alignment of the district's Senate delegation are one only if both Senators are in the Senate majority or the president's party, respectively. The "Senate Delegation Split" dummy indicates that the state's Senate delegation consists of one Democratic and one Republican Senator.

Table 10: Models of District Funding Receipts, 109-111 Congress (Senate)

	Amount Received (% of Bill Total)		Any Grants Received	
	(1)	(2)	(3)	(4)
No Vote on Final Passage	-0.003 (0.004)	-0.004 (0.004)	0.037 (0.035)	0.054* (0.032)
Split on Final Passage	-0.001 (0.006)	-0.003 (0.006)	0.025 (0.030)	0.021 (0.027)
Majority Party		0.003 (0.005)		0.018 (0.022)
President's Party		-0.004 (0.004)		0.022 (0.020)
Split-Party Delegation		-0.004 (0.005)		0.031 (0.026)
Fixed Effects:	Bill	Bill	Bill	Bill
Number of Bills	23	23	23	23
F-statistic	0.01	0.11	46.13	41.06
N	1,150	1,150	1,150	1,150
R ²	0.0002	0.003	0.496	0.497

*p < .1; **p < .05; ***p < .01

Cluster-robust standard errors in parentheses (clustered at the bill level). Bill-clustered standard errors are conservative relative to district-clustered or heteroskedasticity-robust versions. All right-hand-side variables are binary indicators. Dummies for party alignment of the district's Senate delegation are one only if both Senators are in the Senate majority or the president's party, respectively.