NBER WORKING PAPER SERIES

ENDOGENOUS NETWORK FORMATION IN CONGRESS

Nathan Canen
Francesco Trebbi

Working Paper 22756
http://www.nber.org/papers/w22756

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2016

Kareem Carr provided excellent assistance in the data collection. The authors would like to thank Matilde Bombardini, Thorsten Rogall, and seminar participants at various institutions for their comments and suggestions. Francesco Trebbi gratefully acknowledges support by the Canadian Institute For Advanced Research and the Social Sciences and Humanities Research Council of Canada. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Endogenous Network Formation in Congress
Nathan Canen and Francesco Trebbi
NBER Working Paper No. 22756
October 2016
JEL No. P16,P48

ABSTRACT

This paper presents and structurally estimates a model of endogenous network formation and legislative activity of career-motivated politicians. Employing data on socialization and legislative effort of members of the 105th-110th U.S. Congresses, our model reconciles a set of empirical regularities, including: recent trends in Congressional productivity; the complementarity of socialization processes and legislative activities in the House of Representatives; substantial heterogeneity across legislators in terms of effort and success rate in passing specific legislation. We avoid taking the social structure of Congress as exogenously given and instead embed it in a model of endogenous network formation useful for developing relevant counterfactuals, including some pertinent to the congressional emergency response to the 2008-09 financial crisis. Our counterfactual analysis further demonstrates how to empirically identify the specific equilibrium at play within each Congress among the multiple equilibria typically present in this class of games.

Nathan Canen
Vancouver School of Economics
6000 Iona Drive
Vancouver, BC, V6T 1L4
Canada
njcanen@alumni.ubc.ca

Francesco Trebbi
University of British Columbia
6000 Iona Drive
Vancouver, BC V6T 1L4
Canada
and CIFAR
and also NBER
ftrebbi@mail.ubc.ca

A data appendix is available at http://www.nber.org/data-appendix/w22756

1. Introduction

Deliberative bodies of non-trivial size typically rely for their productive functioning on various modes of informal organization, the most prominent of which may be social networks. Information percolates through the networks' edges and individuals form links aimed at guaranteeing reciprocal support and success. Possibly due to its perceived relevance, the study of the role of interpersonal relationship in the legislative process dates at least back to the 1930s (Routt (1938)), but only in the last fifteen years this area of research has come to full prominence (Lazer (2011)). The complexity of simultaneously modeling relational dependence and political decision-making worthy of attention (congressional voting in primis) may be a reason for this delayed uptake.

It is not trivial to operate within theoretical environments where strategic socialization choices and considerations of effort aimed at fulfilling individual political goals of legislators may be jointly determined. This is the core motivation for the construction of the model presented in this paper.

Our theoretical framework, which is specifically designed with empirical estimation in mind, aims at capturing the role of social interactions in fostering legislative activity (we will focus on passage of legislation in the US House of Representatives). We avoid taking the social structure of Congress as exogenously given and instead embed it in a model of network formation useful for developing relevant counterfactuals.

The process of socialization in our setup is instrumental to the legislative success of bills presented by individual sponsors. Legislative success, in return, is instrumental to a politician's own electoral success and reelection. The socialization effort of each member works in a complementary fashion with the effort the politician exerts on legislative activity. Importantly, the socialization and legislative effort decisions of every politician depend on the equilibrium choices of everybody else in the social network[1] that forms endogenously. As typical of the literature on games played on endogenous networks, there are multiple equilibria, that in our setup can be fully characterized.

We structurally estimate the model employing data on cosponsorship and legislative effort of members of the 105th-110th US Congresses. Our estimation approach does not rely on the characteristics of the specific equilibrium at play, but through a careful use of counterfactual analysis we are able to identify which equilibrium is at play nonetheless. Our set of counterfactuals further includes exercises pertinent to the congressional emergency response to the 2008-09 financial crisis.

Our model reconciles a set of empirical regularities: recent trends in Congressional productivity across all bills; the complementarity of socialization processes and legislative activities in the House of Representatives; substantial heterogeneity across legislators in terms of effort and success rate in passing legislation.

[1]This is a relevant dimension of games played on networks. See Bramoullé et al. (2014) on the role of relational links in shaping strategic interactions in equilibrium for a large class of games.

From the theoretical perspective, this paper benefits from the advances in endogenous network modeling within Economics and Mathematical Sociology upon which we build, in particular Cabrales et al. (2011) and Jackson and Zenou (2015). The literature of games on networks includes other important contributions, such as Jackson and Watts (2002), Hojman and Szeidl (2006), Galeotti et al. (2006). Jackson (2008) offers an exhaustive survey.

This paper contributes to an established literature in Political Economy and Political Science on the role of social networks in legislative environments. In an early contribution, Fowler (2006) studies networks in Congress, as defined by cosponsorship links. Using a connectedness measure, he shows that more connected members of Congress are able to get more amendments approved and have more success on roll call votes on their sponsored bills.[2] Again using cosponsorship links, Cho and Fowler (2010) show that Congress can be understood as a small-world network, as it appears subdivided in multiple dense components with few intermediaries tying them in between. The small-world effects seem significant in explaining legislative productivity over time (number of important laws passed, as defined by Mayhew (2005)). Importantly, these papers take networks as exogenous. In this work, the networks will be formed endogenously, and we also do not assume that the observed cosponsorship relationships are the true underlying social network.

Kirkland (2011) focuses on the relationship between bill survival and weak ties (measured as sparsity in cosponsorship) of the sponsor, showing a positive correlation for both eight state legislatures and the US House of Representatives. Cohen and Malloy (2014) study the role of network proximity in Senate voting, focusing on alumni networks and seat assignments, in a reduced form setting. The authors employ identification restrictions aimed at ascertaining causal effects of networks on voting behavior (e.g. using the quasi-at-random seating arrangements of Freshman Senators) and emphasize that these effects are particularly important for bills that are "irrelevant" to the Senator (in which the network carries more sway relative to individual incentives).[3]

Alemán and Calvo (2013), Koger (2003), and Bratton and Rouse (2011) study the incentives for cosponsoring in different settings (focusing on ideological similarity, tenure, etc.). Other works on social networks in political environments, but not necessarily looking at legislative productivity exist, such as Desmarais et al. (2015) who look at the diffusion of legislation/laws across states. Methodologically and in terms of theoretical formalization, we complement and add to these papers.

Beyond their role in social networks, Wilson and Young (1997) study the signaling content of cosponsorships. Coming from a literature initiated by Mayhew (1974), they note that cosponsorship is a reasonably costless way of signaling to the median voter in the district about one's congressional activity. They identify three different explanations for cosponsorships and their possible signaling impact: (i) bandwagoning, (ii) ideology, and (iii) expertise. They find

[2]A similar study is Zhang et al. (2008).

[3]As further motivating evidence, in Appendix we also offer an alternative reduced form analysis of the role of networks on legislative outcomes, where we employ as source of identification the variation in relational links of sponsors of the same bill across House and Senate.

a null to moderate effect of cosponsorship on bill success, as measured as successive progress of the bills through Congress hurdles. Kessler and Krehbiel (1996) instead point out that the timing of cosponsorships would indicate that it is not as much a signaling to voters, as to other politicians (for example, they show that extremists seem to cosponsor earlier). Still in the context of bill sponsorships, Anderson et al. (2003) find correlations with legislative productivity (i.e. the bill passing through different stages in Congress) for congressmember who sponsor more bills and use more floor time (albeit at a declining marginal rate).

The rest of the paper is organized as follows. Section 2 derives the empirical model and discusses identification. Section 3 presents the data. Section 4 illustrates the estimation and the moment conditions. The estimation results and the assessent of the model's fit are reported in Section 5. Section 6 contains the main counterfactuals and discusses the equilibria of the model from an empirical perspective. Section 7 concludes.

2. Model

This section presents a model of endogenous network formation in Congress. The model centers around career incentives. Politicians care about being reelected and they can affect the probability of being reelected by exerting effort in Congress and by building connections instrumental to having specific legislation passed (we think of it as policy favorable to the politician's constituents).

Connections are built through a dedicated costly socialization effort. Through connections, a politician's actions affect his/her whole personal network's outcomes via an intuitive reciprocity mechanism: a politician's legislative effort increases the chances of success of policy bills favorable to each of his/her connections, and vice versa. Each politician anticipates these effects on his/her reelection chances and chooses strategically effort levels in a game on the endogenous network.

The model implies an utility function similar to those in Cabrales et al. (2011) and Jackson and Zenou (2015). Based on extant analysis for this class of games, we can characterize the equilibria of the congressional game. Given data on cosponsorships and legislative activity, which we will discuss below, we further prove identification of the model. Our counterfactual analysis will also demonstrate how it is possible to empirically identify the specific equilibrium at play among the multiple equilibria typically generated by this class of games.

2.1. Setup.

Congress is composed by N members and, for simplicity, we will focus on one chamber (e.g. the House). Let $\mathcal{N} = \{1, 2, ..., N\}$ represent the set of politicians in Congress. We assume that each congressional cycle has two periods, 1 and 2.

2.1.1. *Preferences.*

Politicians are assumed to be career motivated and to exert costly effort with the aim of increasing their chances of being reelected. In period 1, each congressman can present a policy proposal, which for brevity we refer to as a "bill". The bill consists of a policy goal the congressmember intends to fulfill, for instance passing a statute targeted to his or her constituency, landing a subsidy, or obtaining an earmark beneficial to firms in the home district. We will describe below how getting i's policy goal fulfilled will map into an increase in i's chances of being reelected.

Politician i is assumed to choose: i) overall legislative effort x_i, including but not limited to activity devoted to i's constituency-specific bill; ii) a socialization effort s_i, in order to increase the chance of getting his bill approved and his electoral support. Both forms of effort are costly. The cost of congressional effort is given by $\frac{c}{2}x_i^2$, with $c > 0$, and the total cost of socialization is given by $\frac{1}{2}s_i^2$. The parameter c governs the relative cost of congressional effort to socialization effort.

The politician will choose (s_i, x_i) in order to maximize his utility. i's objective is (myopically) to get reelected in the next period (an approximation to a fully intertemporal optimization decision). Specifically, a politician's utility is given by:

$$u_i = Pr(reelected) - \frac{c}{2}x_i^2 - \frac{1}{2}s_i^2. \tag{2.1}$$

We will now proceed to unpack $Pr(reelected)$ explicitly and clarify the nature of x_i and s_i.

2.1.2. *Socialization.* The overall congressional network $\mathbf{G} = \{g_{i,j}\}_{i,j\in\mathcal{N}}$ is endogenous and not observable to the econometrician. $\mathbf{G}$ arises from the equilibrium choices $\mathbf{s}$ of socialization by the N agents, as in Cabrales et al. (2011), and each $i, j \in \mathcal{N}$ link's likelihood is given by:

$$\begin{aligned} g_{i,j}(\mathbf{s}) &= \frac{s_i s_j}{\sum_{k\in\mathcal{N}} s_k} \quad if \quad \sum_{k\in\mathcal{N}} s_k > 0 \quad and \quad i \neq j \\ &= 0, \quad otherwise. \end{aligned}$$

The choices of s_i are generic: increasing s_i means i is more likely to connect with all other agents. They are also complementary, if s_j is higher, it is more likely to be connected to j (all else constant).[4]

2.1.3. *Reelection Probability.* The choice of x_i, the level of overall congressional activity exerted by i, affects the overall legislative support for i's own bill, Y_i, through a Cobb-Douglas support function:

$$Y_i = \varepsilon_i x_i \left(\sum_{j\in\mathcal{N}} g_{i,j}(\mathbf{s}) x_j \right).$$

Both i's own congressional effort, x_i, and that of his network, $\sum_{j\in\mathcal{N}} g_{i,j}(\mathbf{s})x_j$, matter for the ultimate support received by i's bill. If i puts no effort in legislative activity, then the bill will not be approved ($x_i = 0$ implies $Y_i = 0$). The same thing happens if everybody in i's network

[4]Notice that, although we will use cosponsorships of Congressional bills as proxies for socialization efforts, we choose not model the directed targeting by i of cosponsorships toward any specific congressmember j (or other i, j directed linkages). Rather we posit a generic socialization effort. Individual linkages between $i, j \in \mathcal{N}$ could be potentially studied and have been investigated in a literature largely led by Fowler (2006), but each cosponsorship action appears a very noisy signal of any actual link. We do not pursue this approach here, but rather estimate the network links that i has with any other politician j, $\{g_{i,j}\}_{j\in\mathcal{N}}$.

exerts no effort. We assume x_i to be observable by the agents in our model (to all politicians and voters).

Y_i is stochastic and depends also on a random shock ε_i, assumed to be standard Pareto distributed with scale parameter $\gamma > 0$ and i.i.d. across politicians. We assume that ε_i is realized after the choice of $\mathbf{x}$, the vector of x_j across all politicians $j \in \mathcal{N}$. Because ε_i is a shock following the realized legislative support, i must take expectations over its value when choosing (x_i, s_i). Also notice that each link between politician i and j is an endogenous function of the socialization behavior of everybody else, hence the dependency $g_{i,j}(\mathbf{s})$ on $\mathbf{s}$, the vector of s_j across politicians $j \in \mathcal{N}$.

The bill is approved if $Y_i > m$, where $m > 0$ is an institutional threshold.[5] The probability of having the bill approved is thus given by:

$$\begin{aligned} Pr(Y_i > m) &= \Pr\left(\varepsilon_i > \frac{m}{x_i\left(\sum_{j\in\mathcal{N}} g_{i,j}(\mathbf{s})x_j\right)}\right) \\ &= \left(\frac{\gamma}{m}\right)\left(\sum_{j\in\mathcal{N}} g_{i,j}(\mathbf{s})x_j\right)x_i, \end{aligned} \tag{2.2}$$

where we use the distributional assumption on ε.[6] Actual passage of the bill sponsored by i is represented by the indicator function $I_{[Y_i > m]}$.

We interpret x_i as the observable congressional effort by i, instrumental to the approval of i's bill, and we postulate that voters will approve of a politician exerting high congressional effort x_i. We also allow for voters to care about whether in fact the bill passes conditional on effort. That is, we allow for the political principals (the voters) to reward their agent i for effort x_i, networking s_i, and ultimately luck ε_i. To get reelected, the politician must have an approval rate in the electorate that is sufficiently large. Similarly to Bartels (1993), the electoral approval rate of i is modeled as a variable V_i:

$$V_i = \rho V_{i,0} + \zeta I_{[Y_i > m]} + \alpha_i x_i + \eta_i$$

where η_i is assumed to be a mean zero electoral shock, uniformly distributed on $[-0.5, 0.5]$, and where $V_{i,0}$ stands for the baseline approval rate before the start of the term (i.e. before period 1 in the model). Hence, this set-up allows for approval rates to be persistent, but also to react when a politician is capable of getting a bill approved $I_{[Y_i > m]}$ (with a gain ζ) or when i exerts high congressional effort x_i. $\rho > 0$ measures persistence in approval rates, which may be due to the politician's characteristics (such as incumbency advantage, committee membership, majority party affiliation). The parameter ζ, which could be equal to zero empirically, governs the relative importance of a bill actually passing vis-à-vis congressional effort. The direct effect

[5] Naturally, m can be function of a simple majority requirement or even supermajority restrictions.

[6] More generally, one can take Y_i to represent the average approval rate of i's multiple bills. In this case, each b is a separate bill by a politician i. The conditions for our model are unchanged, as long as bills are not strategically introduced (i.e. specifically shocks ε_b are still i.i.d. within i).

of x_i, captured by the parameter α_i, is i-specific and may depend on party affiliation, majority status, congressional delegation, etc. Finally, notice that, while there is no direct value to the voters of the politician having more socialization, the value of s_i matters implicitly, being instrumental in getting legislation approved.

In period 2, i is reelected if his/her electoral approval level, V_i is larger than an electoral threshold w. So, the probability of being reelected is given by:

$$\begin{aligned} & Pr(\rho V_{i,0} + \zeta I_{(Y_i > m)} + \alpha_i x_i + \eta_i > w) \\ = \quad & (1-w) + \rho V_{i,0} + \zeta I_{(Y_i > m)} + \alpha_i x_i \end{aligned}$$

where we use the distribution assumption on η. Note that, in period 1 when making his effort decisions, i does not know the value of ε_i. So taking the expectation over ε of the above implies an expected probability of reelection, when choosing (s_i, x_i), given by:

$$\begin{aligned} Pr(reelected) &= \mathbb{E}^{\varepsilon} Pr(\rho V_{i,0} + \zeta I_{(Y_i > m)} + \alpha_i x_i + \eta_i > w) = \\ &= (1-w) + \rho V_{i,0} + \zeta \frac{\gamma}{m} \sum_{j \in \mathcal{N}} g_{i,j}(\mathbf{s}) x_i x_j + \alpha_i x_i \end{aligned} \tag{2.3}$$

where $\mathbb{E}^{\varepsilon} I_{(Y_i > m)} = Pr(Y_i > m)$, as given by (2.2).

Replacing (2.3) into the utility function (2.1) yields:

$$u_i(x_i, x_{-i}) = (1-w) + \rho V_{i,0} + \zeta \frac{\gamma}{m} \sum_{j \in \mathcal{N}} g_{i,j}(\mathbf{s}) x_i x_j + \alpha_i x_i - \frac{c}{2} x_i^2 - \frac{1}{2} s_i^2$$

Since the terms $(1-w)$ and $\rho V_{i,0}$ do not affect the maximization problem, we can rewrite an equivalent utility for the problem:

$$\tilde{u}_i(x_i, x_{-i}, s_i, s_{-i}) = \alpha_i x_i + \phi \sum_{j \in \mathcal{N}, j \neq i} \frac{s_i s_j}{\sum_{k \in \mathcal{N}} s_k} x_i x_j - \frac{c}{2} x_i^2 - \frac{1}{2} s_i^2 \tag{2.4}$$

where $\phi = \zeta \frac{\gamma}{m}$.

(2.4) is a utility function with correspondence to a class of games on endogenous social networks studied by Cabrales et al. (2011) and Jackson and Zenou (2015). These papers characterize the equilibria of the game, which we now present. This characterization will be useful, because it will provide the basis for identification.

Let us first define:

$$\phi(\alpha) = \phi \psi \tag{2.5}$$

where $\alpha = \{\alpha_i\}_{i \in \mathcal{N}}$ and $\psi = \frac{\sum_{i \in \mathcal{N}} \alpha_i^2}{\sum_{i \in \mathcal{N}} \alpha_i}$ and let us posit non-trivial costs of effort so to avoid explosive (infinite) socialization:

Assumption 1 : $2(c/3)^{3/2} > \phi(\alpha) > 0$.

We will verify ex post this assumption's validity, based on the actual parameters' estimates we obtain.[7] We can now state:

Theorem 2.1 (Cabrales et al. (2011), Theorem 1)**.** Suppose that **Assumption 1** holds. Consider q replica games of the game described above.[8] Then there exists q^* such that for all q-replica games with $q \geq q^*$, there are exactly two interior pure strategy Nash equilibria[9]. These pure strategy Nash equilibria are such that, for all players i the strategies (s_i, x_i) converge to (s_i^*, x_i^*) as $q \to \infty$, where:

$$\begin{aligned} s_i^* &= \alpha_i s \\ x_i^* &= \alpha_i x \end{aligned}$$

for all $i = 1, ..., N$, with (s, x) given by the positive solutions to:

$$\begin{aligned} s &= \phi(\alpha) x^2 \\ 1 &= x[c - \phi(\alpha) s] \end{aligned}$$

The theorem above gives a characterization of the interior pure strategy Nash equilibria of the game, defined by the simultaneous choice (s_i^*, x_i^*) of all agents when maximizing their utility defined by (2.4). Proposition 1 in Cabrales et al. (2011) also proves that these two equilibria are stable.

Note some interesting characteristics of this result, as Jackson and Zenou (2015) point out. The marginal rate of substitution of socialization effort to congressional effort is uniform across politicians; as $\frac{s_i^*}{x_i^*} = \frac{s_j^*}{x_j^*}$. Furthermore, differences in any of the actions (socialization or congressional effort) reflect directly the differences in individual returns α_i, as $\frac{x_j^*}{x_i^*} = \frac{s_j^*}{s_i^*} = \frac{\alpha_j}{\alpha_i}$. Finally, the difference in the probability of approval across agents is also only a function of differences in individual traits, as the ratio between two agents from (2.2) is a function of $g_{i,j}(\mathbf{s}), x_i^*, x_j^*$ for $j = 1, ..., N$ which are, themselves, only functions of $\{\alpha_j\}_{j \in \mathcal{N}}$.

One can further discuss whether in the case of the contemporary Congress, we are in a situation in which q^* is sufficiently large (i.e. that there is sufficient stability to be in the converged stable pure strategy Nash equilibria). Given the amount of repeated interaction among members occurring on Capitol Hill, this does not appear unreasonable.

[7]Notice that this assumption is also prima facie empirically valid. Explosive socialization is not observed in Congressional data, if one for instance focuses on cosponsorship behavior.

[8]According to the authors, "this replica game allows us to take limits as the population becomes large without having to specify the types of the new individuals that are added". Cabrales et al. (2011) allow for agents to have common types, i.e. N agents of $T < N$ types. Here, we are assuming $T = N$ and each agent is a different type α_i. Our empirical approach therefore naturally encompasses also the case with multiple politicians of the same type (e.g. by party, congressional delegation, tenure, etc.).

[9]Lemma 2 in Cabrales et al. (2011) also characterises a semi-corner Nash equilibrium, given by $(s_i^*, x_i^*) = (0, \alpha_i/c)\ \forall i = 1, ..., N$. This equilibrium does not require Assumption 1.

2.2. **Preliminaries to Estimation.** Let us assume that the econometrician observes equilibrium effort s_i^*, x_i^* with noise, as neither of these dimensions may be perfectly observable by the researcher. For instance, bill cosponsorships capture a lot of the socialization patterns and are often used in the political science literature as the basis for the construction of political networks. However, there are further parts of process socialization we cannot fully observe (e.g. close-doors meetings, fund-raisers, and so on). Similarly, although we can partially observe congressional effort through standard proxies (e.g. times the Congressmember was present on the floor for speeches, number of bills written[10], presence in roll call voting), we cannot perfectly observe the effort they put in their activities in Congress.

Denote $\mathcal{N}_\tau$ to be the agents present in Congress τ. This is a set which can vary across different τ.[11] Introducing classical measurement error, for agent i in Congress τ, we observe:

$$s_{i,\tau} = s_{i,\tau}^* + \upsilon_{i,\tau} \tag{2.6}$$

$$x_{i,\tau} = x_{i,\tau}^* + \lambda_{i,\tau} \tag{2.7}$$

with $\upsilon_{i,\tau}, \lambda_{i,\tau}$ i.i.d. across agents (and time) and mutually independent. We posit standard assumptions on the measurement errors. They are mean zero, conditional on the observed data, and have constant variance: $\mathbb{E}(\upsilon_{i,\tau} \mid \{\alpha_j\}_{j\in\mathcal{N}_\tau}) = \mathbb{E}(\lambda_{i,\tau} \mid \{\alpha_j\}_{j\in\mathcal{N}_\tau}) = 0$, and $\mathbb{E}(\lambda_{i,\tau}^2 \mid \{\alpha_j\}_{j\in\mathcal{N}_\tau}) = \sigma^2$.[12]

The information we employ is the following.Let us indicate with $\{y_{i,\tau} = I_{(Y_{i,\tau}>m)}\}$ whether each bill was approved or not, where $i \in \mathcal{N}_\tau$ and τ is a given Congressional cycle.[13] $\{s_{i,\tau}\}$ indicates the cosponsorship decisions per politician $i \in \mathcal{N}_\tau$. This is our proxy for the equilibrium socialization choice $\left\{s_{i,\tau}^*\right\}$. $\{x_{i,\tau}\}$ indicates a vector of observable proxies for congressional effort $\left\{x_{i,\tau}^*\right\}$. This is constructed using data on floor speeches (word counts per politician during a term) and on roll call presence/votes. We employ a procedure (Non-Negative Matrix Factorization, see Trebbi and Weese (2016)), to reduce the dimensionality of this set of proxies to a single dimension.[14]

[10]Both highlighted as important for legislative success in Anderson et al. (2003).

[11]The data is observed for multiple Congresses and we provide identification results for parameters specific to each Congress. This means we allow our parameters to differ across different Congresses and we can construct time-series estimates of the parameters.

[12]Note that, once one conditions on the set $\{\alpha_j\}_{j\in\mathcal{N}_\tau}$, one is conditioning on all individual observations $(s_{i,\tau}, x_{i,\tau})$.
Indeed, from Theorem 2.1; (s^*, x^*) are completely determined by the parameters that govern the system. Then, all individual choices are just functions of parameters and of the set of types $\{\alpha_j\}_{j\in\mathcal{N}_\tau}$. So we can rewrite the assumptions as: $\mathbb{E}(\upsilon_{i,\tau} \mid s_{i,\tau}, x_{i,\tau}) = \mathbb{E}(\lambda_{i,\tau} \mid s_{i,\tau}, x_{i,\tau}) = 0$, and $\mathbb{E}(\lambda_{i,\tau}^2 \mid s_{i,\tau}, x_{i,\tau}) = \sigma^2$.
We are assuming that s_i^* is what is chosen, and that it is randomly and independently decomposed between the observable s_i, and an error function. s_i, x_i are observed, and an arbitrary measurement error, conditional on this observation (and on the data we have), is mean zero, with variance σ^2 (for the observed effort) and independent of all the other measurement errors across individuals and time. We do not need to impose that the measurement errors in both types of effort have the same distribution.

[13]A complete data description section follows below.

[14]Sponsorship of bills is already included, as we will use the separate bills independently. Further details on this, the data and procedure to lead to the effort proxy are in the next Section.

As we will be performing our analysis within a Congress, we suppress the notation τ. We assume that a single pure strategy Nash equilibrium, as defined in Theorem 2.1, is played in each Congress. We do not impose, however, that the same equilibrium is played across different Congresses, rather we identify which equilibrium is played empirically in Section 6. More precisely, we show that, given the observed data, one can uniquely pin down the equilibrium which is actually being played. In fact, although there are two interior pure strategy Nash equilibria, conditional on observing $\{s_i, x_i\}_{i\in\mathcal{N}}$, there is only one set of values consistent with it. For identification of the parameters of our model it is not necessary to identify the nature of the equilibrium at play.

Our aim is, given $\{y_i, s_i, x_i\}_{i\in\mathcal{N}}$, to estimate the parameters $(\gamma, c, \phi, \psi, \zeta, \{\alpha_i\}_{i\in\mathcal{N}}, \sigma^2)$. The basis for identification will be Theorem 2.1 and the systems of equations it implies. Formal identification of our model, given the information available to the econometrician is proved, in Appendix.

3. Data

We use cosponsorship data from Fowler (2006), compiled from the Library of Congress, covering the 105th to the 110th United States Congress (from 1997 to 2008). This data contains cosponsorship decisions by politician, and within that data, who sponsors and who cosponsors each bill. It also contains information on whether the each bill was approved in Congress or not (we will focus on passage in the House of Representatives). Figure 1 shows that measures of inter-connectedness of Congress, for example the total number of cosponsorship links in legislative acts across members of the House (Fowler (2006)), have been steadily increasing.

Per Congressional cycle, we compute how many bills each politician cosponsors. Cosponsorships act as empirical proxies for the socialization effort $\{s_{i,\tau}\}_{i\in\mathcal{N}_\tau}$. The individual bill success outcome (i.e. if the bill passes or not) maps into $\{y_{i,\tau}\}_{i\in\mathcal{N}}$. We then use the sponsorship information to link the outcome of the bill to the network characteristics and individual decisions.

To compute our proxies for congressional effort, $\{x_{i,\tau}\}_{i\in\mathcal{N}_\tau}$, we first collect data on Roll Call voting and floor speeches in Congress. Data for Roll Call voting comes from VoteView. We compute an index, for each politician and for each term in Congress, as the times the congressmember voted as a proportion of total Roll Call votes. This measure, which we call *Roll Call Effort*, is defined as $1-$ (number of times i was "Not Voting"/ total Number of Roll Call votes in a Congress).

Following Anderson et al. (2003), we also use data on floor speeches as a measure of congressional effort. To do so, we compile the amount of words that each Congressmember used in his/her floor speeches across the duration of one term (we call this variable $Words$). Our *Floor Speeches* variable is $log(1+Words_{i,\tau})$. Data comes from Gentzkow and Shapiro (2015), available on ICPSR.[15] That these measures of social interaction and legislative activity may

[15] As there are changes in the composition of Congress within a term, for instance due to death or resignation among other reasons, we have some observations whose co-sponsorship numbers and word counts do not correspond to a full term. To mend this, we scale up values proportionally to the recorded behavior while in Congress. In other words, if a politician leaves halfway through his term, we double the values of these observations.

be germane to one another is evident from the significant and positive raw correlation of link formation and proxies of legislative activity and effort, for instance floor speeches in Figure 2 and Roll Call Effort in Figure 3. This complementarity between effort choices is fully consistent with our theoretical setup.

We proceed to construct $\{x_{i,\tau}\}_{i\in\mathcal{N}_\tau}$, by using both Roll Call Effort and Floor Speeches. An appropriate combination of these variables can be obtained through dimensionality reduction methods. Since effort should be non-negative, we employ a procedure that guarantees positive values (i.e. we cannot use methodologies like principal components analysis, as this involves a centering of data and negative values).[16] We employ Non-Negative Matrix Factorization (NNMF). Non-Negative Matrix Factorization is a dimensionality reduction procedure which imposes constraints so that the resulting elements are all non-negative. A survey of this methodology can be found in Wang and Zhang (2013). NNMF works by factorizing a matrix, call it A, into two positive matrices W, H, under a quadratic loss function. The product WH is an approximation to A of smaller dimension, as there are less columns in W than rows in A. We then apply the transformation proposed in Tsuge et al. (2001).

Summary statistics for all our variables can be found in Table 1. Figures 4-6 complement these, by showing graphically the different distributions of the raw variables (Cosponsorship decisions, Words in Floor Speeches, Roll Call Effort) over Congress cycles.

We restrict the data to Congresses 105th-110th for multiple reasons. Firstly, the data we employ to compute effort from floor speeches is only available from the 104th Congress onwards. Secondly, the 104th Congress (corresponding to the Republican Revolution) provides a structural break in the analysis of Congressional behavior. With multiple changes to Congressional composition and structure during the 104th, it becomes hard to compare the costs and socialization of this specific Congress to others, preceding or following, without having to further delve into the exceptionality of this congressional cycle, which is not the aim of this work.[17]

Identification of $\{\alpha_i\}_{i\in\mathcal{N}}$ requires setting a normalization $\alpha_1 = 1$. The choice of which politician to use as the normalizer does not impact our estimates of an individual Congress, but can affect the comparison across Congresses if the normalizer changes. For this reason, we choose Representative Ed Markey as a common normalizer $i = 1$, as he is member of all Congresses in our sample. Rep. Markey also exhibits a stable behavior and was member of prior and subsequent Congresses, excluding fluctuations due to rookie or retirement effects. Available upon request are also results obtained using as alternative normalizers Representatives Neil Abercrombie, Rick Boucher, Dan Burton, or Fred Upton. The results appear stable and qualitatively identical.

[16]Our qualitative results still hold if we use either of these variables individually. However, the magnitudes of the estimates change due to the different scales of Roll Call Effort (between 0 and 1) and the floor speech data (in hundreds of words).

[17]In addition, without ad-hoc modifications to the estimating model specifically designed to accommodate the idiosyncrasies of the 104th Congress, this lack of stability would also likely undermine any effort of structural estimation.

Finally, we perform an additional trimming of the data: across all Congresses in the sample we drop 7 out of 2681 observations of politicians that have cosponsorship figures less than 3 bills over a full term. From our identification equations, we have to use the inverse value of α for all individuals. For those with less than 3 cosponsorships (and given that our normalizer and most politicians cosponsor in the hundreds), the inversion becomes close to infinite and this harms our inference. We also remove a set of 19 observations, that have the number of words in Floor Speeches set to 0 in the data of Gentzkow and Shapiro (2015). These observations relate almost exclusively to some Congressmen who either resigned or died during that term[18]. Since the data is zero, the rescaling above does not prove to be adequate, so we drop these observations.

4. Estimation

The moment conditions necessary to identify and estimate the model's parameters are[19]:

$$\alpha_i = \frac{\mathbb{E}(s_i)}{\mathbb{E}(s_1)} \tag{4.1}$$

$$\psi = \frac{\sum_{j\in\mathcal{N}}\left(\frac{\mathbb{E}(s_j)}{\mathbb{E}(s_1)}\right)^2}{\sum_{j\in\mathcal{N}}\left(\frac{\mathbb{E}(s_j)}{\mathbb{E}(s_1)}\right)} \tag{4.2}$$

$$\mathbb{E}s_i\alpha_i = \phi\psi(\mathbb{E}x_i^2 + \sigma^2) \tag{4.3}$$

$$\mathbb{E}\left(\alpha_i - x_i\left(c - \phi\psi\frac{s_i}{\alpha_i}\right)\right) = 0 \tag{4.4}$$

$$\sigma^2 = \frac{\mathbb{E}x_k x_j \alpha_i^2}{\alpha_k\alpha_j} - \mathbb{E}x_i^2 \tag{4.5}$$

$$\mathbb{E}\left(y_i - \gamma\sum_{j\in\mathcal{N}}\frac{s_i s_j}{\sum_{k\in\mathcal{N}} s_k}x_j x_i\right) = 0. \tag{4.6}$$

We proceed by providing estimates using a two step procedure. In the first step, we compute estimates for the set of types $\{\alpha_i\}_{i\in\mathcal{N}}$ and ψ using the empirical analogues to 4.1 and 4.2. Given these estimates, we proceed by Generalized Method of Moments to estimate the parameter vector $(c, \phi, \sigma^2, \gamma)$ through the identified system given by (4.3), (4.4), (4.5) and (4.6).

It is important to note that because we are estimating $N-1$ parameters for $\{\alpha_i\}_{i\in\mathcal{N}}$ for each Congress, while we only observe one realization of variables for each individual, we cannot make robust inference on the set of parameters $\{\alpha_i\}_{i\in\mathcal{N}}$ (and, hence, on ψ) in the first step. We can, however, obtain conditional confidence intervals. These confidence intervals are conditional on our approximation of α_i and have proved to be reliable in Montecarlo simulations. Furthermore, the small significance (and size[20]) of the estimates of σ^2 presented in the

[18]Such as Representatives Jo Ann Davis in the 110th Congress, Sony Bono in the 105th, or resignations as Representative Bobby Jindal in the 110th.

[19]See Appendix for a full derivation and a description of the moment conditions.

[20]If we compare it to the average number of cosponsorships, used to estimate each α_i

next section provide evidence that these approximations are reasonable in terms of magnitude (that is, notwithstanding the presence of measurement error in observed congressional and socialization effort).

Concerning the information of whether a bill passed or not $\{y_{i,\tau}\}_{i\in\mathcal{N}}$, the model assumes each politician presents only one bill. Because a good fraction of members of Congress sponsor multiple bills, however, we work with $L > N$ bills in the actual data. This is easily accommodated in the estimation. Recall that ε are i.i.d. across time and bills. For each politician i, all i's bills have the same associated network $g_{i,j}$, as it comes from the same politician and his same network and effort choices (as well as those of his network). The different ε realizations, however, represent different bill qualities or institutional arrangements within politician, meaning that the same politician may have one bill approved and not another. The dimensionality of the problem can be decreased by simply averaging out each bill's success by politician. This is made possible by the fact that equation (4.6) holds for all bills, implying that it must hold for all politicians as well. Similarly, we average out equation (4.5) over k, j for estimation.

5. Results

Table 2 shows the average estimates for the set of individual types α_i, $i \in \mathcal{N}_\tau$. Figures 7-8 show graphically the distributions of the estimated α_i (Figure 7) over time and by party (Figure 8). These distributions appear stable across Congresses. However, splitting the samples by party, we can observe important differences in the estimated distributions of Republicans and Democrats. Democrats have a higher average and dispersion, while Republicans have tighter distributions. This implies different socialization patterns across parties, as Democrats socialize more and (by our socialization function) more often with other Democrats.[21]

Figure 9 shows the estimates of ψ, which also appear quite stable over time. This is consistent with Figures 7-8, given the relationship between ψ and the moments of the distribution of α, which are shown to be stable.

Table 3 presents our estimates for the parameters $(\psi, \phi, c, \sigma^2, \gamma)$, together with their conditional confidence intervals. These results are also shown graphically in Figures 9-12.

We can see that ϕ is estimated very precisely and with similar magnitude over time, between 1.05 and 1.25. This provides evidence that the returns to socialization are quantitatively sizeable (above the median estimated α_i for most Congresses) and have not changed much over time. In the context of the model, these returns represent the (expected) electoral gain to the politician of having a bill approved relative to the direct electoral return to congressional effort. ϕ being stable suggests that the returns of having a bill approved have not significantly changed over the period of time studied, and that the electorate seems to continue to value significantly the returns of having projects approved, possibly more than congressional effort per se.

[21] This may not necessarily hold for Republicans. While this is a shortcoming, a modification of the socialization function $g_{i,j}(\mathbf{s})$ to become party-specific appears unfeasible for tractability reasons. Ultimately, the quality of the the fit of our model (discussed below and shown in Table 5) is sufficiently high to assuage any concern pertinent to this specific matter.

The relative cost of congressional effort c is estimated to be increasing over time. One can note that these changes in c track increases in cosponsorship decisions over time, as reported in Figure 1. As recent Congresses exhibit more cosponsorship/socialization, the model specifically attributes this trend in cosponsorships to a decrease in the relative cost of socialization (an increase in the cost of congressional effort or a drop in the cost of socialization). Intuitively, as socialization becomes relatively cheaper, politicians increase their socialization effort.

The estimates of γ are noisy. Still, we can obtain estimated probabilities of bill approval, as shown in Figure 13. By comparing Figure 13 with the average bill passage rates in the Summary Statistics (Table 1), we can see that the model can generate a good match at the mean approval rate (which we observe), while our structural assumptions allow us to represent the whole distribution of expected probabilities of having a bill approved across different politicians. These indicate some variation over time. Later Congresses (108th, 109th and 110th), as well as the 106th Congress, show a higher predicted approval rate for most politicians. A higher c leads to a reoptimization which benefits all politicians, as they are able to approve more bills. Table 4 shows that the changes in the estimated c tracks the changes in the mean probability of approval. In the next section, we will discuss how this positive correlation is consistent with our model.

Concerning the sample fit of the model we perform two different exercises. First, we predict the specific cosponsorship i, j links at the level of each Congressmember based on what predicted by our $g_{i,j}(\mathbf{s})$ function separately for each Congress in our sample. The correlations between the estimated $g_{i,j}(\mathbf{s})$ and any i, j individual cosponsorships are reported in Table 5. The correlations are very high, showing an excellent fit of our parsimonious model to more complex individual socialization choices. Second, we fit the empirical success rate Y_i for each politician in Congress τ to the estimated one $\mathbb{E}^{\varepsilon} I_{(Y_i > m)} = \frac{\gamma}{m} \sum_{j \in \mathcal{N}} g_{i,j}(\mathbf{s}) x_i x_j$ at τ. For this exercise, we focus on politicians who have sponsored a sufficiently large amount of bills (over 40 House Bills sponsored), so to have a reasonable empirical approximation to an average success rate. We show a positive correlation of 0.297 between estimated and empirical approval rates. This is illustrated in Figure 14 and shows that our model captures part of the empirically observed relationships.

6. Counterfactuals

As in Cabrales et al. (2011), the theoretical analysis above identifies multiple equilibria in pure strategies. Our objective here will be to identify which is being played in each Congress and derive suitable implications. It turns out that identification of the empirically extant equilibrium across multiple potential equilibria is possible through counterfactuals.

For convenience let us remind the reader which multiple equilibria exist. First of all, there is a semi-corner equilibrium which exists even without Assumption 1 holding. In this equilibrium congressional effort is exerted and is chosen at level $x_i^* = \alpha_i / c$, but there is no socialization, i.e. $s_i^* = 0$ for all i. This occurs because socialization is too costly (relative to congressional effort) whenever c is low enough. Effort is still provided in the model, because there are incentives for reelection given by (2.3). This equilibrium is obviously counterfactual, as typically $s_i^* > 0$, and not of interest to us.

Theorem 2.1 characterizes two stable interior pure-strategy Nash equilibria existing under the parametric restriction of Assumption 1: $2(c/3)^{1.5} > \psi\phi$. Based on the empirical estimates in the previous section, this restriction is verified for all Congresses and reasonably slack (c is estimated in the hundreds, while ϕ and ψ are estimated in the range $[1, 2]$).

As described in Cabrales et al. (2011), the two equilibria have different welfare properties, as one (the high action equilibrium) is Pareto-superior to the other (the low action equilibrium). Cabrales et al. (2011) also notice that the socially efficient outcome lies in between the two equilibria. As such, the equilibria are to be considered "too high" socialization and congressional effort (in high action) and "too low" socialization and congressional effort (in low action). The inefficiency arises from the fact that the agents fail to internalize the effect of their choices on other agents' behavior through the network when maximizing their individual utility.

While Assumption 1 guarantees two interior stable equilibria, we do not know which equilibrium is at play without further inspection.

We can isolate through counterfactuals which equilibrium is at play in each Congress in an intuitive way.[22] The approach is straightforward once one realizes that we unambiguously know which equilibrium is being played in the data by looking at appropriate comparative statics for the estimated curves defined in Theorem 2.1. One obtains comparative statics of opposite sign depending on which equilibrium is at play. Indeed, regarding comparative statics, Cabrales et al. (2011) state: *"It turns out that, when the returns increase, all equilibrium actions decrease at the Pareto-superior equilibrium, while they increase at the Pareto-inferior equilibrium"*. Figure 2 in their work illustrates the relevant comparative statics.

Operationally, this requires using the pair of equations in Theorem 2.1 based on the estimated parameters for each Congress and producing a form of comparative statics by modifying the appropriate parameters. For instance, a decrease in the relative cost of congressional effort c will imply moving towards a counterfactual equilibrium with strictly higher congressional effort and higher socialization effort when starting from the low action equilibrium, while the exact opposite (i.e. lower congressional effort and lower socialization) will happen to agents' counterfactual equilibrium actions when instead the high action equilibrium is being currently played.

While assessing empirically multiple equilibria constitutes important motivation for performing counterfactual analysis, the reader may be also interested in quantitatively assessing the model. For this reason we perform quantitative exercises involving either γ or c. Notice also that when we change γ, we automatically change either ϕ or ζ, by definition. This means there are two possible counterfactuals for γ: changing γ by keeping ζ constant, or changing γ by keeping ϕ constant. Since changing ϕ or changing c are similar counterfactuals, as both change the returns to socialize (the comparative statics go in the same direction), we present

[22] Also notice that each Congress could be potentially playing a different type of equilibrium.

counterfactuals in c as they are straightforward and lend themselves to a clearer interpretation. We focus first on changing γ keeping ϕ constant (i.e. changing the bill quality shock). Subsequently, we explore changes in c. Finally we conclude with a counterfactual analysis of the Congressional response to the 2008-09 financial crisis.

6.1. **Change in γ with ϕ constant.** For each of the 105th-110th Congresses, the subfigures in Figure 15 show counterfactuals reducing γ while keeping ϕ constant (i.e. ζ adjusts inversely). Recall that the shock to the bill passage ε is assumed to be standard Pareto distributed with scale parameter $\gamma > 0$, hence lower γ determines a lower median draw of the positive shock ε and lower chances of legislative success. As the system of equations in Theorem 2.1 does not change (the equations depend on ϕ directly), only the probability of approval will be affected as per equation (2.2). Hence, γ only changes the shape of the bill approval function. Quantitatively a decrease in γ by 10% leads to a sizeable shift of the probability of approval curve to the left (which in Table 1 are shown to vary from 9.57% to 12.85%). This is shown in Figure 15. As we only change the values of γ in (4.6), the percentage change in the expected probability of bill passage is linear, dropping by 10% as well.

6.2. **Change in c.** We now perform counterfactuals with respect to c. By lowering c we lower the cost of congressional effort relative to the socialization effort cost. By inspection of Assumption 1, it is clear that the magnitude of the change in c may trigger a violation of this assumption. Hence, the magnitude of the change in c will matter for the analysis below. We discuss two cases.

As a first case, consider taking the limit $c \to 0$. This change will immediately violate Assumption 1, inducing the semi-corner equilibrium without socialization effort. In this equilibrium, congressional effort will still be provided, even absent social networks, because of positive reelection returns to effort. This is because in equation (2.3), which governs the chance of reelection, there are positive returns to a higher x_i^* even in case the bill does not get passed. Voters are assumed to care about effort per se when deciding whether to reappoint a politician and politicians care about getting reelected. Politicians will not be able, however, to pass legislation. Finally, let us note that this semi-corner equilibrium is unstable in the sense that, were one politician to change to a positive s_j, so would all the other players.

As a second case, suppose the reduction in c does not trigger a violation of Assumption 1. The counterfactual results presented in Table 6 change c without violating this restriction. This modifies the incentives for socialization and effort, as it changes their relative costs. A reduction in c affects the system governed by (A.1)-(A.4). While two possible interior equilibria exist, the returns to socialization increase when lowering c and this is proven to decrease all equilibrium actions at the Pareto superior equilibrium and increase them at the Pareto inferior equilibrium. Figure 16 shows the estimated system of equations for each Congress. We can see that the only equilibrium (where the curves cross) consistent with the observed value of the proxies for effort is the Pareto superior equilibrium for every Congress. Under the estimated parameters, Congress appears to be systematically in the high action equilibrium with over-investment in socialization and effort.

Let us now assess the magnitude of these effects. We decrease the estimated value of c by 1% and by 10%, and assess the changes in the probability of bill approval at the high action equilibrium. We find, expectably, that there is a negative effect of decreasing c on the likelihood of bill passage. We, however, find that this effect is more than linear. There is a decrease of bill success probability of almost 17% on average across all Congresses, vis-à-vis a decrease of 10% in c. This is because in the high action equilibrium, a decrease in c leads to a decreased choice of congressional effort and socialization of an individual, which reduces the returns to socialization of other individuals, which then further lower their own congressional effort until a new equilibrium is reached. This feedback effect compounds and leads to a larger negative effect, pointing out the importance of socialization in the approval of bills.

While we are able to conclude that Congress is in the high action equilibrium case, the alternative Pareto inferior equilibrium can be also analyzed. Based on our parameter estimates, the low action equilibrium appears essentially a "government shutdown" case. This is shown more closely in Figure 17, where we have a closer look at the Pareto inferior equilibrium for the 110th Congress. We can see that this is very close to an equilibrium where agents produce no legislative effort and no socialization. Inaction in Congress or shutdown are intuitive: when no one is making any effort, there is no incentive to either make effort (outside of the direct reelection incentives) or socialize, even while Assumption 1 holds.

This multiplicity of equilibria requires some discussion, possibly because of its empirical relevance. In recent congressional history we have indeed observed multiple shutdowns consistent with jumping to the Pareto inferior equilibrium. The United States federal government shut down in October 2013 during the 113th Congress. Previous events include the November 1995 shutdown and a subsequent episode between December 1995 and January 1996, both during the 104th Congress. Why and how these transitions to different equilibria may manifest is unfortunately beyond the scope of this paper (and a common limitations of many models with multiple equilibria). Yet it is an important feature that our theoretical framework may accommodate such empirical phenomena.

6.3. **Counterfactual of the Democratic Party Takeover in the 110th Congress.** We conclude this section by proposing a counterfactual of congressional behavior during the 110th cycle. Elected in November 2006, the House of Representative turned Democratic majority after twelve years of consecutive Republican control. This revealed to be a particularly consequential election, as it was the 110th Congress that voted between the Summer and the Fall of 2008 a host of emergency economic measures in response to the 2008-09 financial crisis (for an analysis of these Congressional votes, see Mian et al. (2010)). Some of this legislative activity happened to be extremely momentous, including the vote of the Emergency Economic Stabilization Act of 2008 (EESA, also known as the "TARP" from the Troubled Asset Relief Program), which initially failed passage in the House , inducing one of the largest intraday losses in NYSE's history.

An important counterfactual is to asseess how relevant the role of congressional networks was in eventually guaranteeing a responsive legislative intervention to the financial crisis. How much different would have legislative activity looked absent the Democratic Party takeover in 2006? Would have Congress behaved differently under Republican control?

Within our framework this counterfactual corresponds to replacing the α_i estimated parameter of every i Democrat elected to the 110th Congress with the α_j of the j Republican i had replaced in office, and then generating new equilibrium actions and a counterfactual social network. Figure 18 reports the distribution in the data and in the counterfactual. The counterfactual distributions appears only mildly less right skewed than what is estimated for the 110th Congress.

However, even small differences in the vector of α could potentially produce substantial effects through the network in terms of bill likelihood of success. In Table 7 we inspect the magnitude of the counterfactual reductions in the likelihood of bill passage for some of the most important emergency response legislation during the Fall of 2008. This set includes in addition to the EESA, the Housing and Economic Recovery Act of 2008 (aimed at foreclosure prevention, also studied by Mian et al. (2010)), the Economic Stimulus Act of 2008 and the Supplementary Appropriations Act, both large bills precursor of the fiscal intervention of 2009. Table 7 reports the relative differences in bill passage probabilities between the counterfactual and the estimated model. It appears all differences are around 1/2 a percent reduction in likelihood of success, a quantitatively small effect. This allows to reject the claim that, absent the Democratic takeover of the House in 2006, the financial crisis response would have been substantially different, with a more restrained government intervention under a Republican Congress (obviously, conditionally on the same set of emergency legislation being pushed forth by Treasury Secretary Hank Paulson and Federal Reserve Chairman Ben Bernanke).

7. Conclusions

This paper develops and estimates a structural model of legislative activity for the US Congress in which endogenous social interactions play a role in promoting bill passage. We show that socialization effort matters quantitatively for legislative activity.

By endogenizing both legislative effort and socialization activity, we are able to accommodate the complementarity in these two sets of equilibrium actions, which we indeed observe in the raw data. The model further fits recent trends in Congressional productivity and accommodates substantial heterogeneity across legislators in terms of effort and success rate in passing sponsored legislation. By employing a parsimonious model of socialization, we are also able to develop relevant counterfactuals, including some pertinent to the congressional response to the 2008-09 financial crisis. With more recent waves of data, further counterfactuals (for example, related to the role of the Tea Party movement) could be performed.

Multiple equilibria arise naturally within our theoretical setting (as it is typical of models of endogenous network formation). By careful use of theory and counterfactual exercises, we are able to identify which equilibrium is in fact at play in the data. Congress appears systematically in a high action equilibrium, where both congressional effort and socialization are inefficiently high. A latent low action equilibrium is also possible and it amounts to a government shutdown, not dissimilar from the actual episodes such as October 2013 and November through January 1995.

References

Alemán, E. and Calvo, E. (2013). Explaining policy ties in presidential congresses: A network analysis of bill initiation data. *Political Studies*, 61(2):356–377. 3

Anderson, W. D., Box-Steffensmeier, J. M., and Sinclair-Chapman, V. (2003). The keys to legislative success in the us house of representatives. *Legislative Studies Quarterly*, 28(3):357–386. 4, 9, 10

Bartels, L. M. (1993). Messages received: the political impact of media exposure. *American Political Science Review*, 87(02):267–285. 6

Bramoullé, Y., Kranton, R., and D'Amours, M. (2014). Strategic interaction and networks. *American Economic Review*, 104(3):898–930. 2

Bratton, K. A. and Rouse, S. M. (2011). Networks in the legislative arena: How group dynamics affect cosponsorship. *Legislative Studies Quarterly*, 36(3):423–460. 3

Cabrales, A., Calvó-Armengol, A., and Zenou, Y. (2011). Social interactions and spillovers. *Games and Economic Behavior*, 72(2):339–360. 3, 4, 5, 7, 8, 14, 15

Cho, W. K. T. and Fowler, J. H. (2010). Legislative success in a small world: Social network analysis and the dynamics of congressional legislation. *The Journal of Politics*, 72(01):124–135. 3

Cohen, L. and Malloy, C. J. (2014). Friends in high places. *American Economic Journal: Economic Policy*, 6(3):63–91. 3

Desmarais, B. A., La Raja, R. J., and Kowal, M. S. (2015). The fates of challengers in us house elections: The role of extended party networks in supporting candidates and shaping electoral outcomes. *American Journal of Political Science*, 59(1):194–211. 3

Fowler, J. H. (2006). Connecting the congress: A study of cosponsorship networks. *Political Analysis*, 14(4):456–487. 3, 5, 10

Galeotti, A., Goyal, S., and Kamphorst, J. (2006). Network formation with heterogeneous players. *Games and Economic Behavior*, 54(2):353–372. 3

Gentzkow, M. and Shapiro, J. (2015). Congressional record for 104th-110th congresses: Text and phrase counts. ICPSR33501-v4. 10, 12

Hojman, D. A. and Szeidl, A. (2006). Endogenous networks, social games, and evolution. *Games and Economic Behavior*, 55(1):112–130. 3

Jackson, M. O. (2008). *Social and Economic Networks*. Princeton University Press, Princeton, NJ, USA. 3

Jackson, M. O. and Watts, A. (2002). The evolution of social and economic networks. *Journal of Economic Theory*, 106(2):265–295. 3

Jackson, M. O. and Zenou, Y. (2015). Games on networks. volume 4 of *Handbook of Game Theory with Economic Applications*, pages 95–163. Elsevier. 3, 4, 7, 8

Kessler, D. and Krehbiel, K. (1996). Dynamics of cosponsorship. *American Political Science Review*, 90(03):555–566. 4

Kirkland, J. H. (2011). The relational determinants of legislative outcomes: Strong and weak ties between legislators. *The Journal of Politics*, 73(03):887–898. 3

Koger, G. (2003). Position taking and cosponsorship in the us house. *Legislative Studies Quarterly*, 28(2):225–246. 3

Lazer, D. (2011). Networks in political science: Back to the future. *PS: Political Science & Politics*, 44(01):61–68. 2

Mayhew, D. R. (1974). *Congress: The electoral connection*. Yale University Press. 3

Mayhew, D. R. (2005). *Divided we govern: Party control, lawmaking and investigations, 1946-2002*. Yale University Press. 3

Mian, A., Sufi, A., and Trebbi, F. (2010). The political economy of the us mortgage default crisis. *The American Economic Review*, 100(5):1967–1998. 17, 18

Routt, G. C. (1938). Interpersonal relationships and the legislative process. *The Annals of the American Academy of Political and Social Science*, 195(1):129–136. 2

Trebbi, F. and Weese, E. (2016). Insurgency and small wars: Estimation of unobserved coalition structures. *Working Paper*. 9

Tsuge, S., Shishibori, M., Kuroiwa, S., and Kita, K. (2001). Dimensionality reduction using non-negative matrix factorization for information retrieval. In *Systems, Man, and Cybernetics, 2001 IEEE International Conference on*, volume 2, pages 960–965. 11

Wang, Y. X. and Zhang, Y. J. (2013). Nonnegative matrix factorization: A comprehensive review. *IEEE Transactions on Knowledge and Data Engineering*, 25(6):1336–1353. 11

Wilson, R. K. and Young, C. D. (1997). Cosponsorship in the u. s. congress. *Legislative Studies Quarterly*, 22(1):25–43. 3

Zhang, Y., Friend, A., Traud, A. L., Porter, M. A., Fowler, J. H., and Mucha, P. J. (2008). Community structure in congressional cosponsorship networks. *Physica A: Statistical Mechanics and its Applications*, 387(7):1705–1712. 3

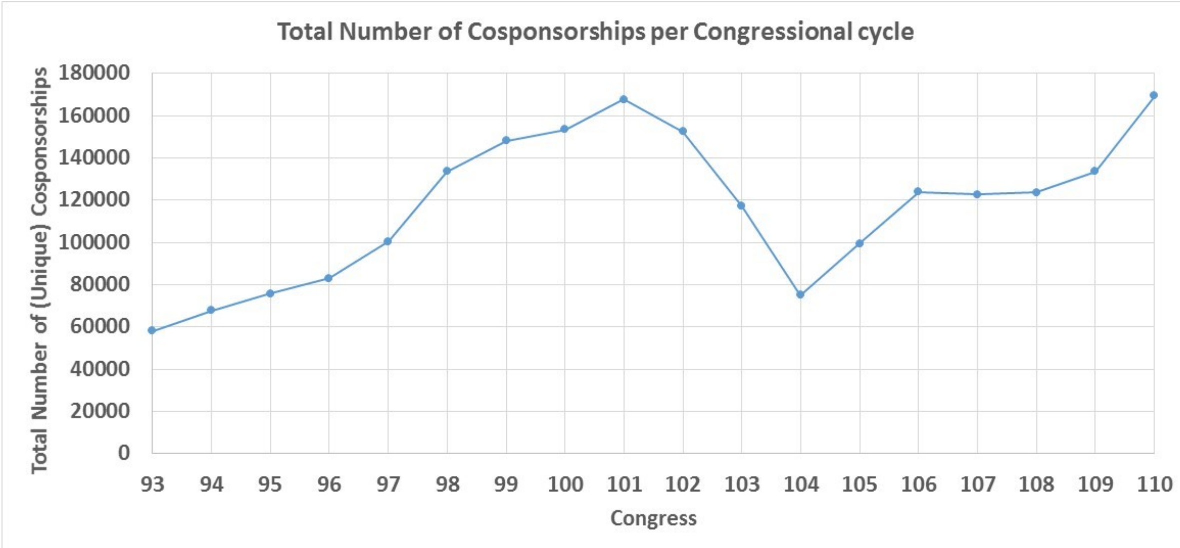

Figure 1. Total Number of Cosponsorships per Congressional cycle

The figure shows the evolution of the total number of (unique) cosponsorships during a congressional cycle (i.e. anytime a politician has cosponsored another in a directed way) over time.

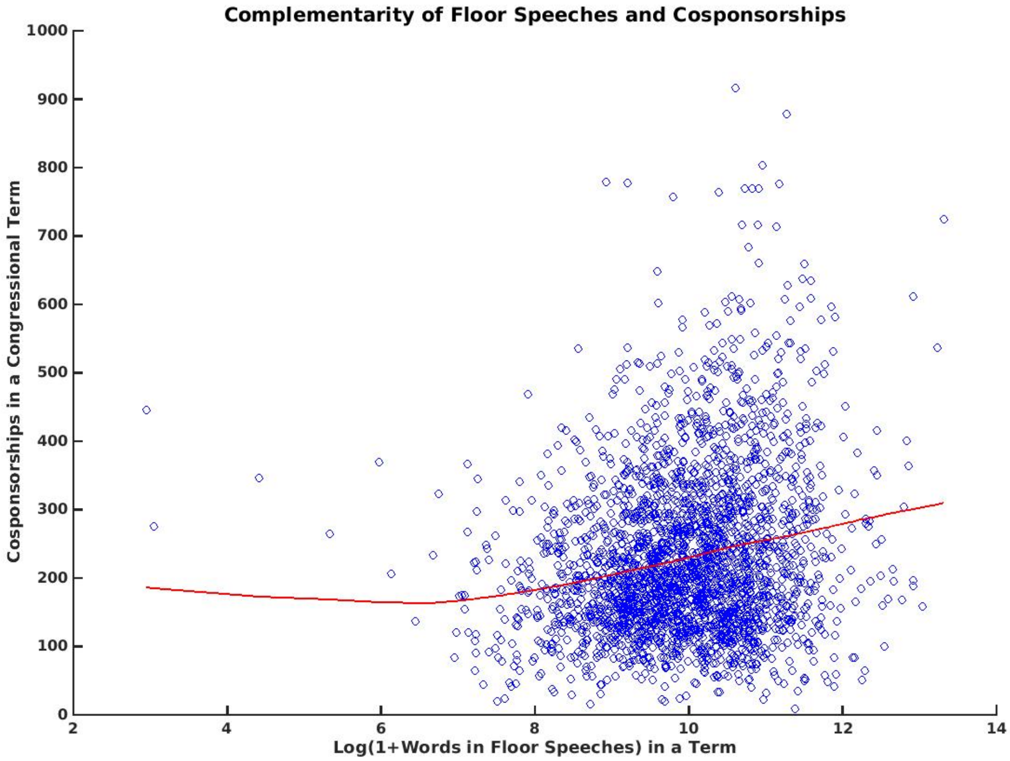

FIGURE 2. Correlation between the raw data of log(1+Words) in Floor Speeches and Cosponsorship decisions.

The figure shows the positive correlation between proxies for socialization (cosponsorships) and legislative effort (number of words in floor speeches). The graph presents the variables in raw form, without rescaling or removal of members with low cosponsorship. We present a LOWESS (locally weighted scatterplot smoothing) fit, with bandwidth (span) equal to 0.9, fitting the relationship between the variables. We do remove, as described in the Data section, observations that have total words equal to zero, which are mostly due to death/resignations in that term.

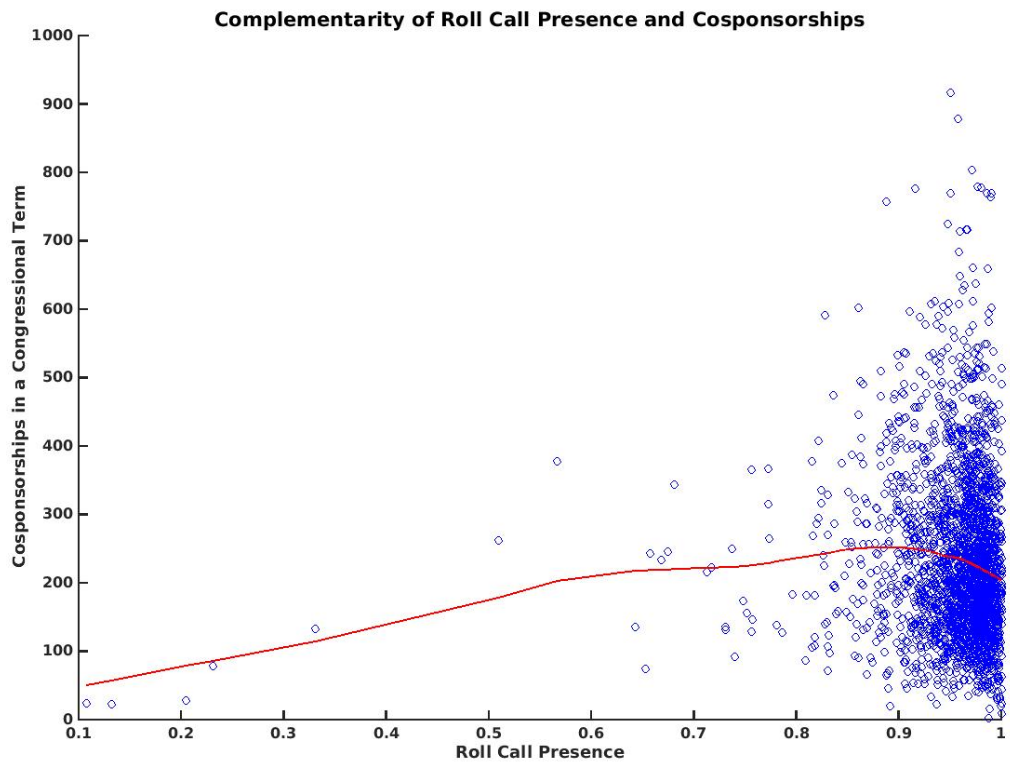

FIGURE 3. Correlation between the raw data of Roll Call Presence and Cosponsorship decisions.

The figure shows the positive correlation between proxies for socialization (cosponsorships) and legislative effort (times the politician is Voting in Roll Call). The graph presents the variables in raw form, without rescaling or removal of members with low cosponsorship. We present a LOWESS (locally weighted scatterplot smoothing) fit, with bandwidth (span) equal to 0.9, fitting the relationship between the variables.

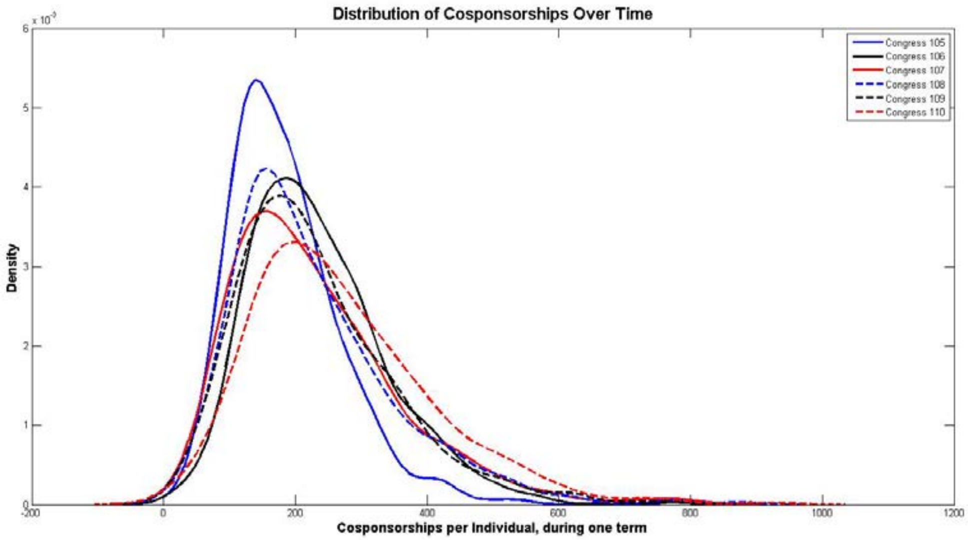

FIGURE 4. Distribution of Number of Cosponsorships of Politicians Across Congresses

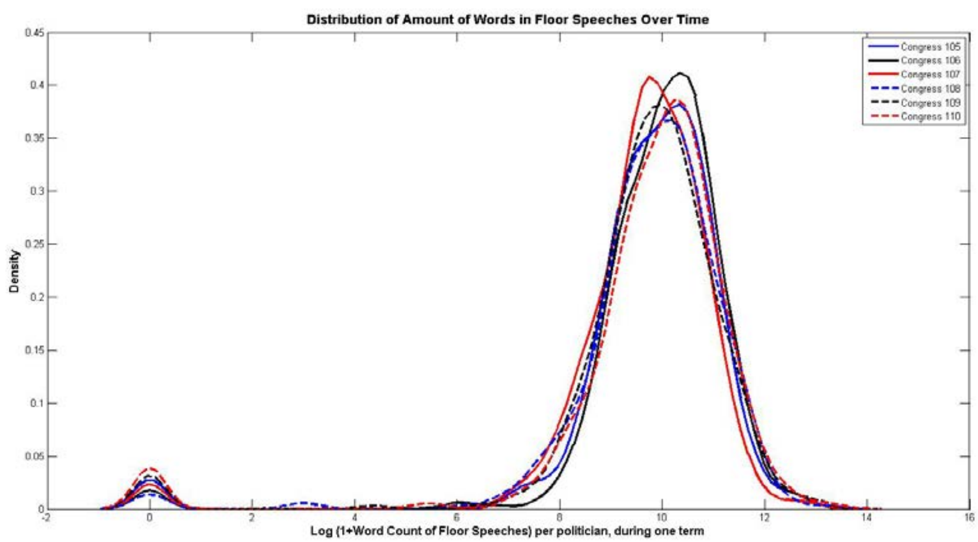

FIGURE 5. Distribution of log(1+Words) in Floor Speeches of Politicians Across Congresses

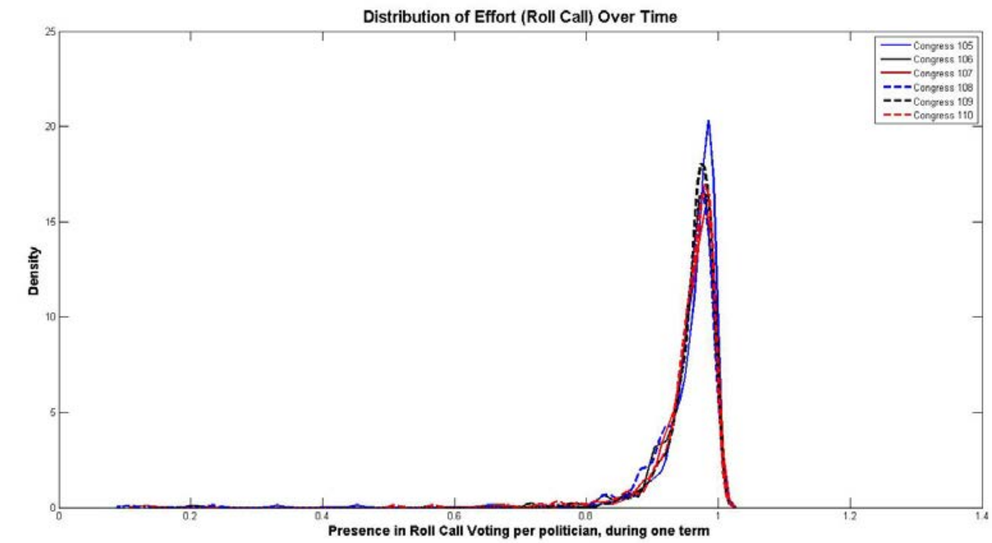

FIGURE 6. Distribution of Roll Call Presence of Politicians Across Congresses

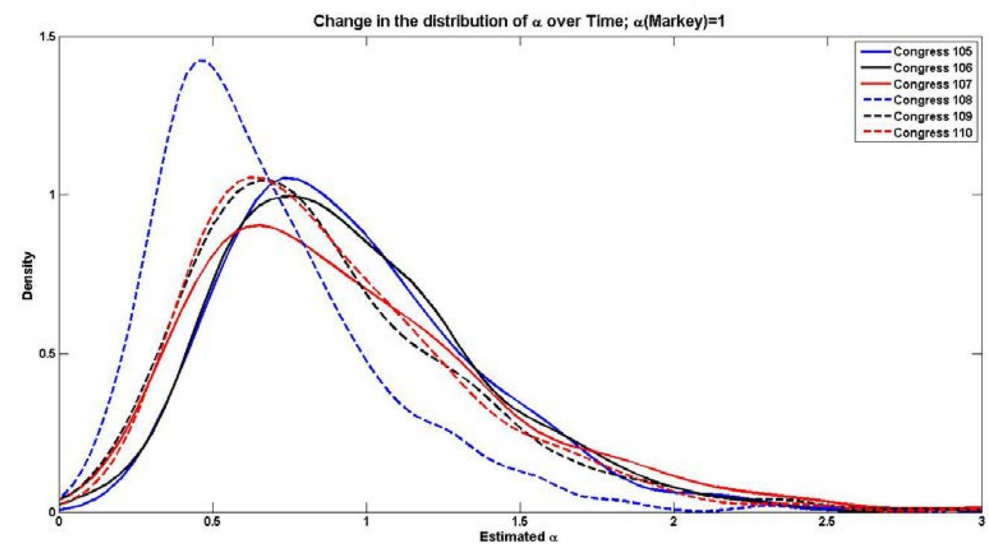

FIGURE 7. Distribution of Estimated α over Time

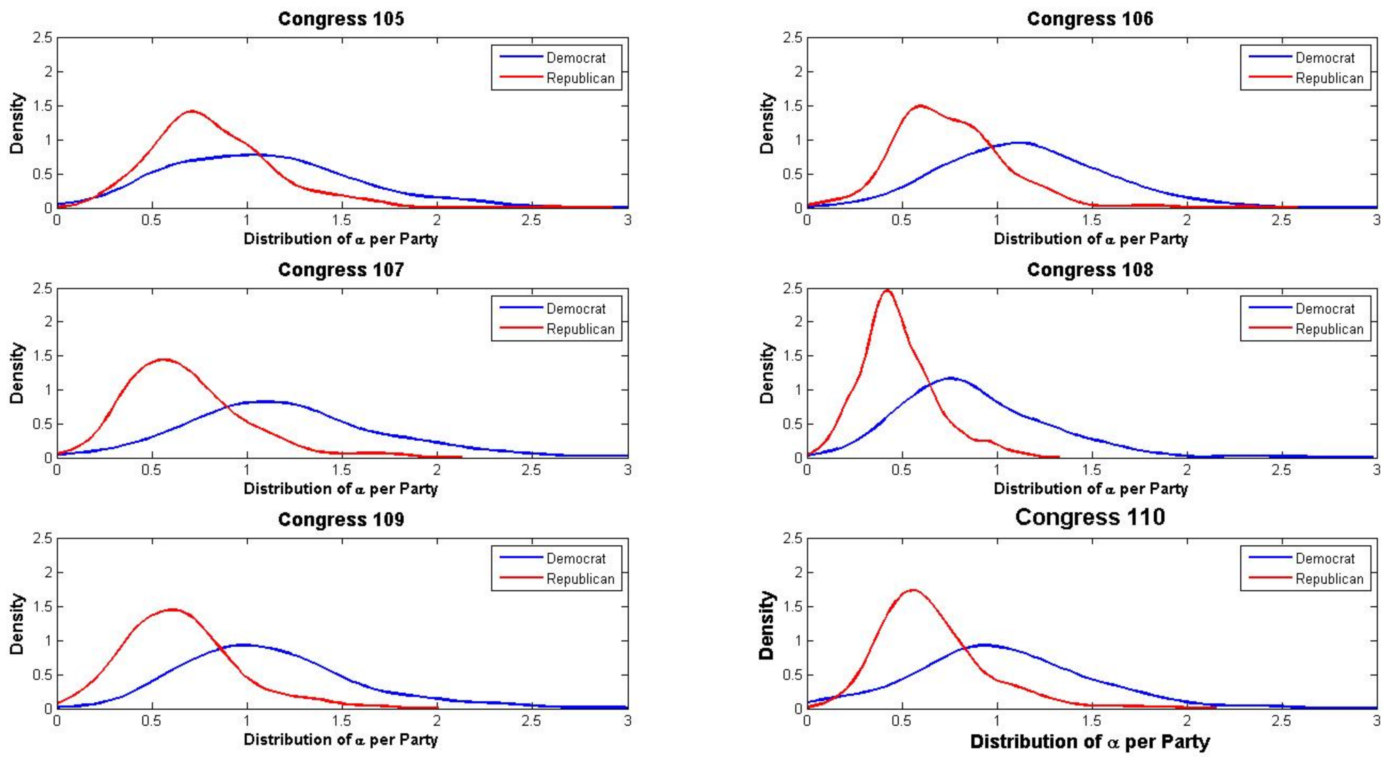

FIGURE 8. Distribution of Estimated α Across Parties over Time

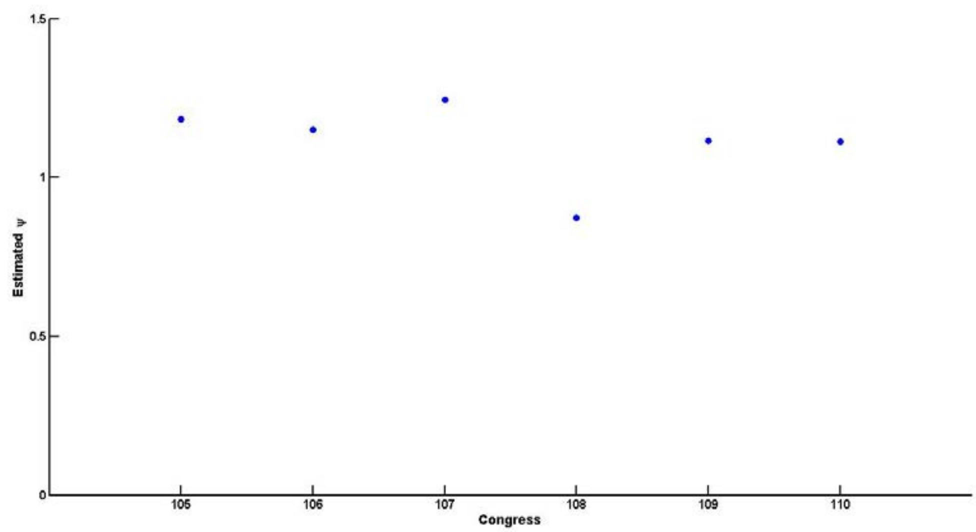

FIGURE 9. Estimated ψ over Time

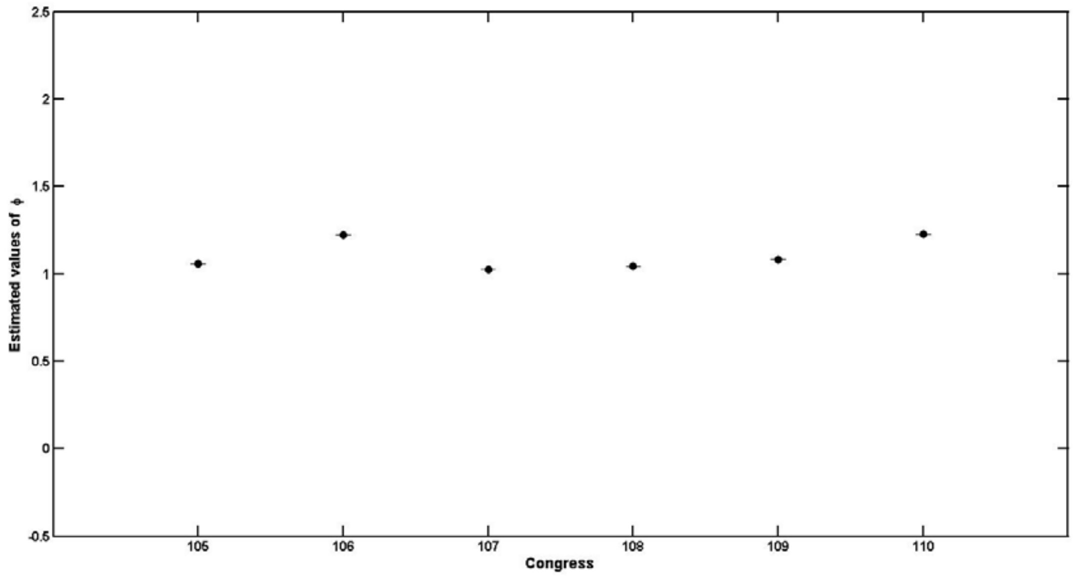

FIGURE 10. Estimated ϕ over Time

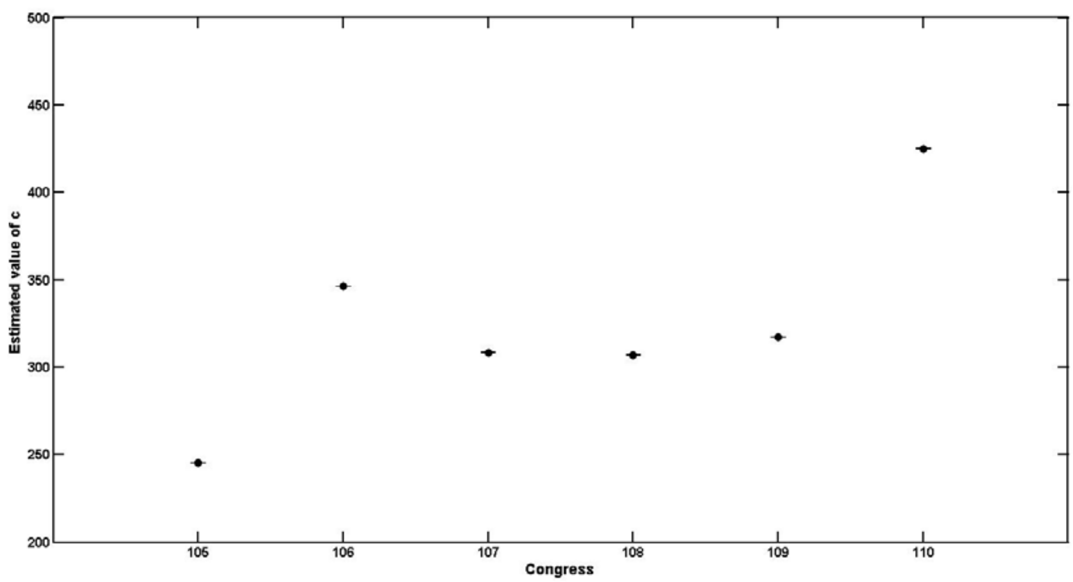

FIGURE 11. Estimated c over Time

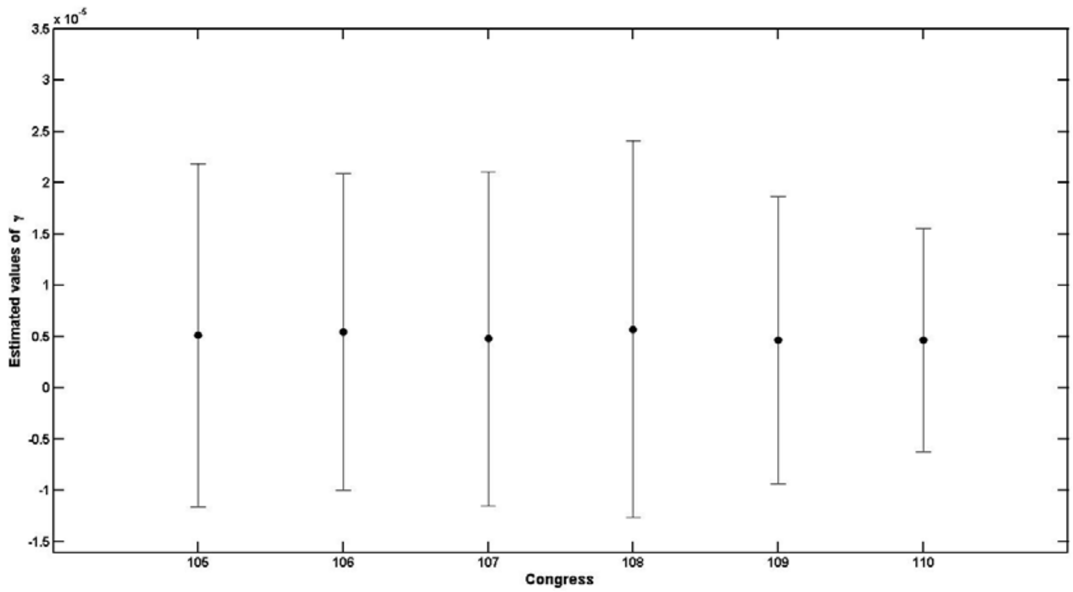

FIGURE 12. Estimated γ over Time

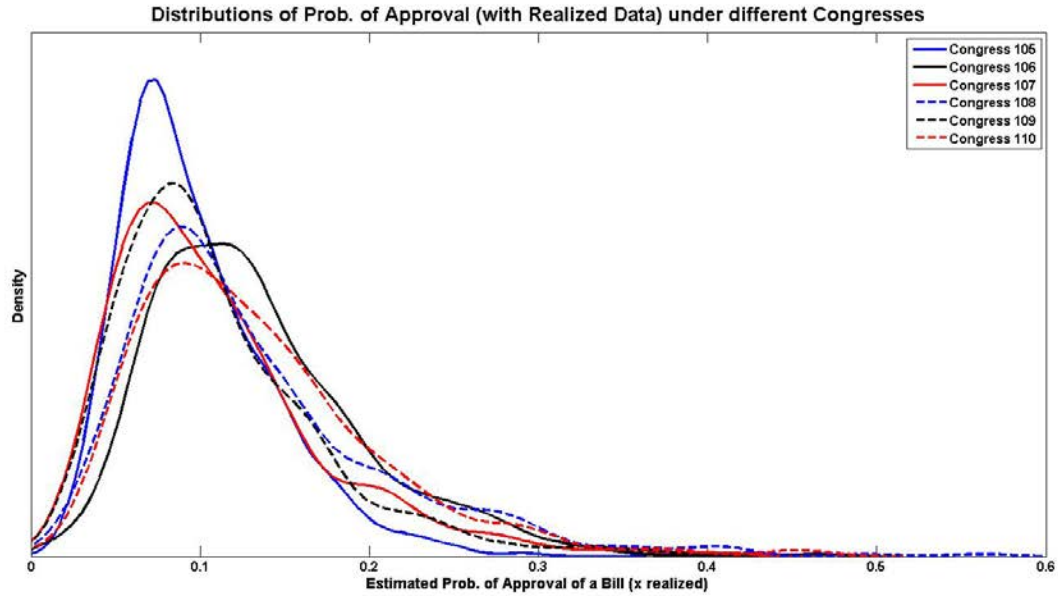

FIGURE 13. Estimated Probability of Approval, Congresses 105-110

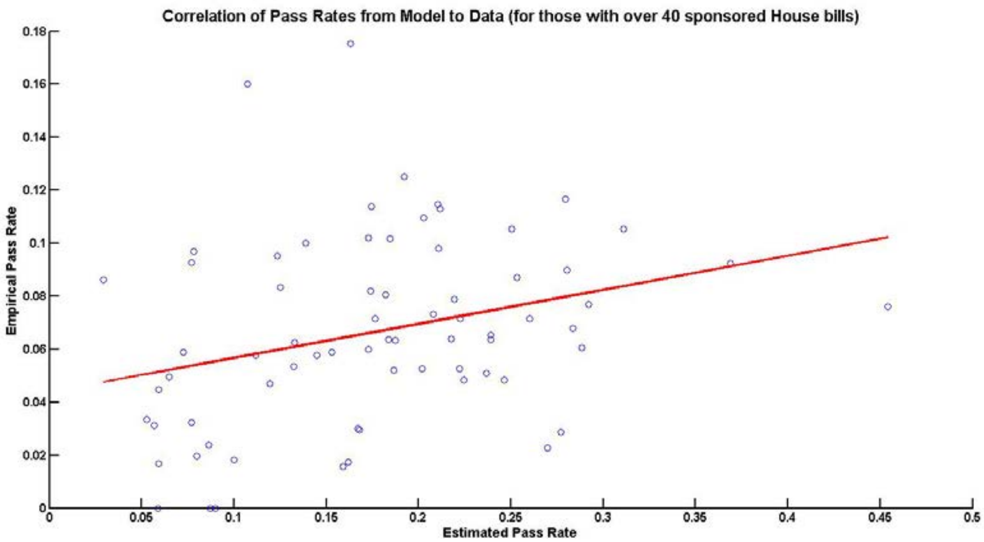

Figure 14. Fit of the Model: Estimated Probability of Approval to Empirical Probability per Politician

The figure shows the fit of the model when looking at the model's estimated probability of a politician's bill being approved, given by $\frac{\hat{\gamma}}{m} x_i \sum_{j \in \mathcal{N}} g_{i,j}(\mathbf{s}) x_j$, as compared to the empirical one (computed by the number of bills a politician sponsored and that passed the House, including amendments and resolutions, divided by the total number of those he sponsored in that Congressional term). We present the figure for the observations that had more than 40 House bills sponsored in a Congressional term. This restriction is needed to make sure we have a consistent empirical estimate, as a small number of bills implies that predicted averages can be far from the realized ones. A linear fit is shown in red.

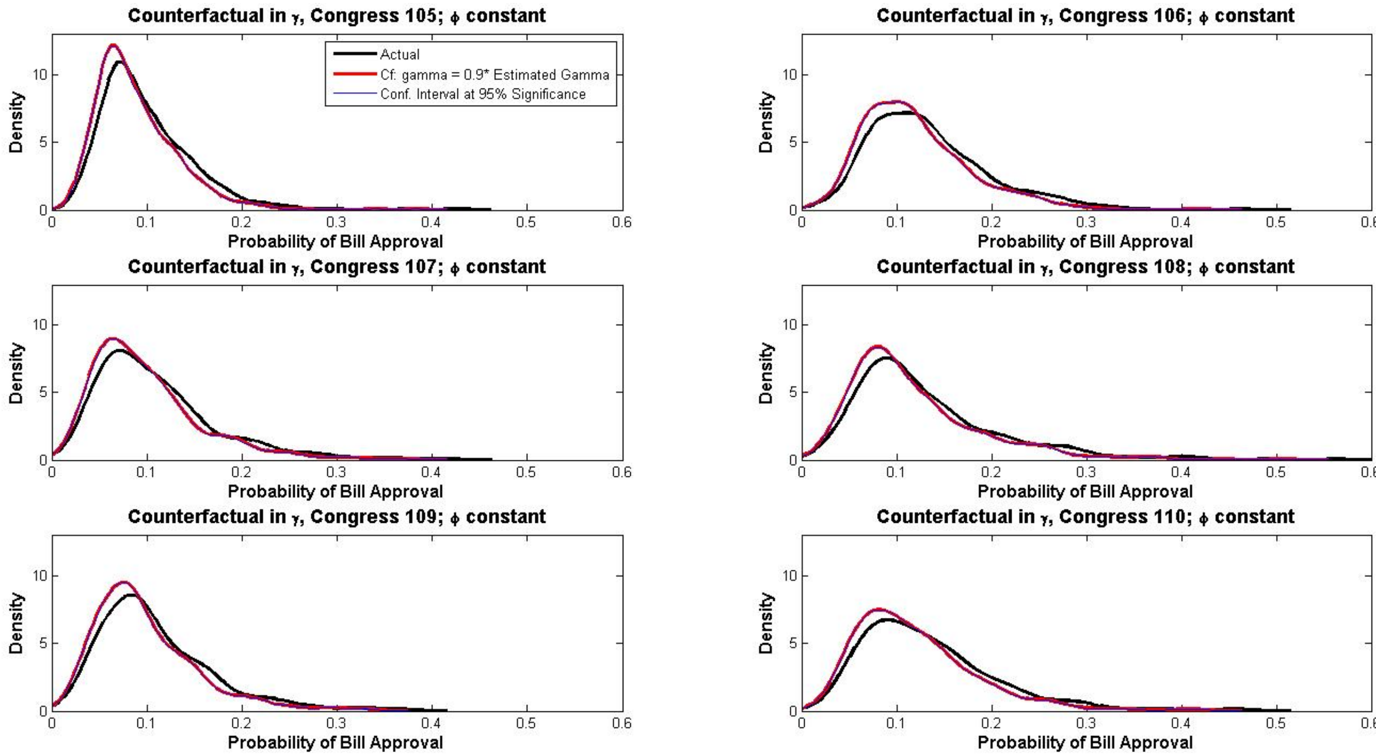

FIGURE 15. Counterfactuals: Decreasing $\hat{\gamma}$ by 10%, ϕ constant

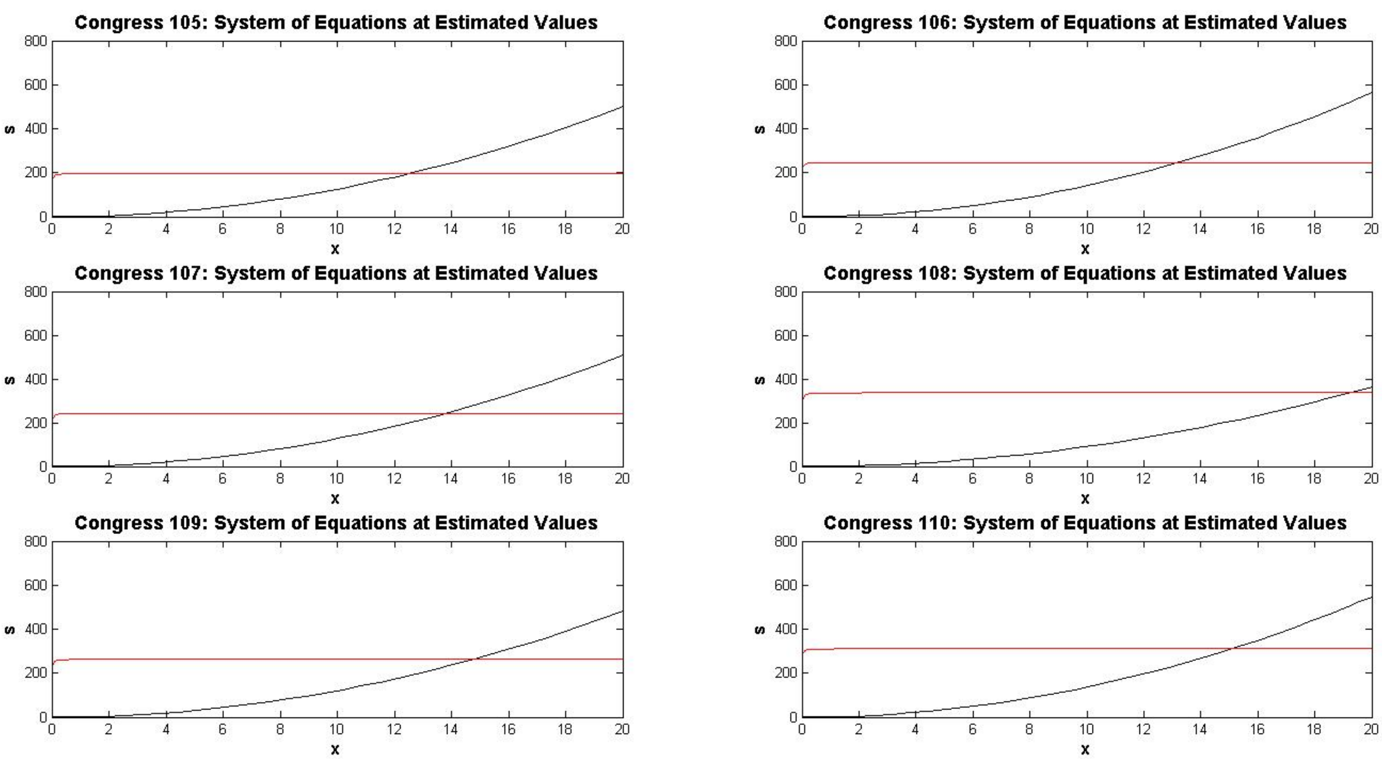

FIGURE 16. Estimated System of Equations from Theorem 2.1 across Congresses

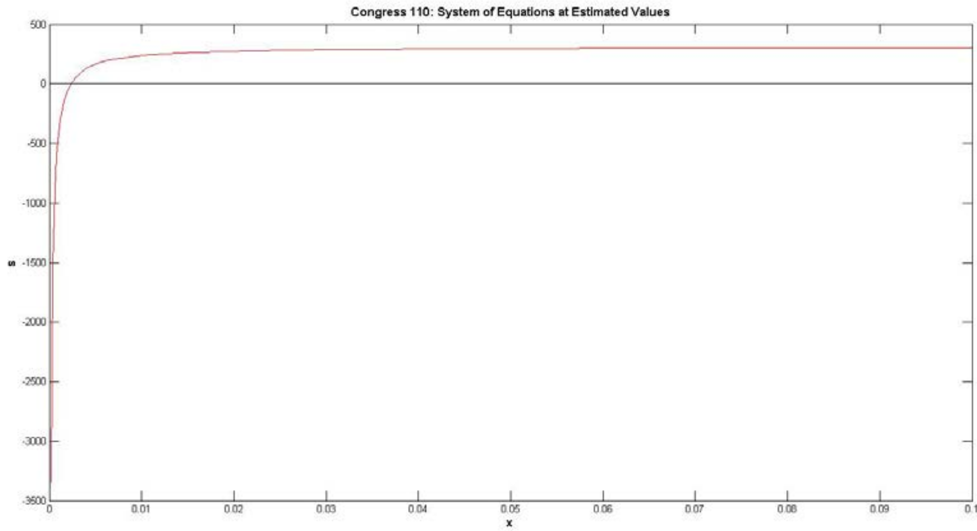

FIGURE 17. Estimated System of Equations from Theorem 2.1, Congress 110, enlarged detail close to $(0, 0)$

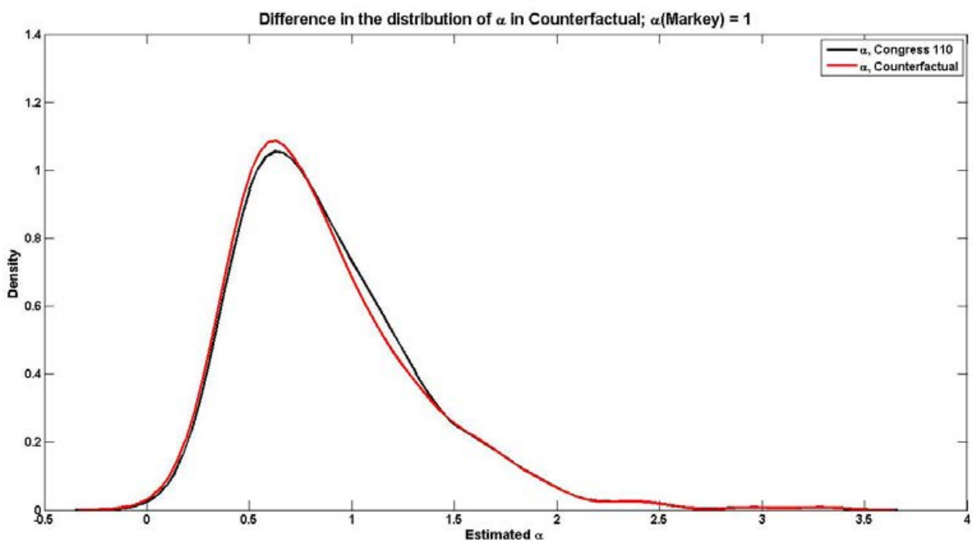

FIGURE 18. Distribution of α in Congress 110 and Counterfactual (Democrats who won seats in 110 that had been Republican are replaced by those Republicans from Congress 109)

Table 1. Summary Statistics

Congress	**105**	**106**	**107**	**108**	**109**	**110**
Cosponsorships						
Mean	185.74	234.57	229.79	226.75	230.74	269.65
Standard Deviation	85.79	102.91	127.03	124.08	119.48	135.90
Minimum	2	28	22	20	15	9
Maximum	683	769	804	878	779	916
Words in Floor Speeches						
Mean	32938.633	36282.23	27906.61	33490.47	33985.21	37416.96
Standard Deviation	38503.19	39234.14	34421.74	42334.30	45922.73	51212.574
Minimum	0	0	0	18	0	0
Maximum	360967	384094	376449	556633	405859	609928
Presence in Roll Call Votes						
Mean	0.9620	0.9524	0.9556	0.9505	0.9605	0.9551
Standard Deviation	0.0514	0.0574	0.0579	0.0665	0.0380	0.0497
Minimum	0.3310	0.2051	0.1323	0.1076	0.5669	0.5094
Maximum	1	1	1	1	1	1
N	442	435	440	439	438	445
Approval of House Bills						
Mean	0.1087	0.1246	0.0981	0.1138	0.0957	0.1285
Standard Deviation	0.3758	0.3782	0.3092	0.3439	0.3690	0.3687
Minimum	0	0	0	0	0	0
Maximum	1	1	1	1	1	1
Number of Bills	4874	5681	5767	5431	6436	7340

The table presents summary statistics for the variables used in the structural estimation, across Congresses. Summary Statistics for individual bills and politicians can be found in the Online Appendix, relevant for the Reduced Form section. Presence in Roll Call is defined as the proportion of Roll Call votes that the politician does not appear as "Not Voting". Number of words said in floor speeches aggregates the number of words said by a politician across all his speeches in a term. Cosponsorships and number of words are scaled to full term length (i.e. if a politician leaves mid-office and is replaced mid-office; then both him and the replacement have those variables multiplied by 2.). For estimation, we remove the observations (bills and politicians) we do not have or cannot match to identifying numbers, and those with less than 3 Cosponsorships (see the Data Section). These are mostly Congressmen who substitute others mid-term. Data used for bills is House bills (H.R.).

Table 2. Estimated Mean α and Dispersion of its Estimates

Congress	**105**	**106**	**107**	**108**	**109**	**110**
Average $\hat{\alpha}$	0.9730	0.9613	0.9636	0.6784	0.8860	0.8896
	(0.4516)	(0.4273)	(0.5208)	(0.3641)	(0.4521)	(0.4449)

The table presents the average estimated α for each year, $\alpha_{Markey} = 1$. Standard deviation of this mean estimate is presented in brackets.

Table 3. Parameter Estimates

Congress	**105**	**106**	**107**	**108**	**109**	**110**
$\hat{\psi}$	1.1826	1.1512	1.2451	0.8738	1.1167	1.1122
$\hat{\phi}$	1.0586***	1.2228***	1.0238***	1.0428***	1.0800***	1.2292***
	(0.706E-5)	(1.246E-5)	(0.684E-5)	(1.043E-5)	(0.865E-5)	(1.532E-5)
$\hat{c}$	245.48***	346.39***	308.59***	307.14***	317.29***	425.23***
	(0.0461)	(0.0450)	(0.0533)	(0.0381)	(0.0468)	(0.0475)
$\hat{\sigma}^2$	79.20	90.34	130.39	119.88	115.97	122.68
	(203.51)	(184.31)	(267.40)	(263.73)	(234.75)	(251.13)
$\hat{\gamma}$	5.142E-6	5.491E-6	4.792E-6	5.729e-6	4.680E-6	4.683E-6
	(8.510E-6)	(7.869E-6)	(8.305E-6)	(9.372E-6)	(7.137E-6)	(5.556E-7)
N	437	434	438	439	434	437

Standard Errors are in Parentheses. Standard errors for $(\hat{c}, \hat{\phi}, \hat{\sigma}^2)$ are conditional on the approximation of α by $\hat{\alpha}$. *** represents significance at 1%, ** 5%, * 10%. γ estimated by GMM, from our identified set of equations. Effort variable is the Non-Negative Matrix Factorization of the variables $[log(1 + Words_{i,\tau}), RollCallEffort_{i,\tau}]$, where the second variable is the presence in Roll Call Votes. Socialization proxy is the number of Cosponsorships. $E - 7$ denotes $\times 10^{-7}$. The confidence intervals are conditional on the estimates of α and ψ.

Table 4. Consistency of the Model: c is positively correlated to *realized* approval rates

Congress	**105**	**106**	**107**	**108**	**109**	**110**
Estimate of c	245.48	346.39	308.59	307.14	317.29	425.23
Mean Approval of House Bills	0.1087	0.1246	0.0981	0.1138	0.0957	0.1285

Our model predicts that c and the probability of bill approval should be positively correlated (see Table 6). In this table, we show that our model is also consistent with the data: the correlation between our estimated value of c and the *realized* mean approval rate is 0.63.

Table 5. Model Fit: Correlation of Network from the Model to Observed Individual Cosponsorship decisions between politicians

Congress	**105**	**106**	**107**	**108**	**109**	**110**
Correlation between Model links and observed cosponsorship links	0.6492	0.7653	0.8230	0.8048	0.7779	0.7555

The table presents the correlation between the network predicted by the model, given by $g_{i,j}(\mathbf{s})$ under the estimated parameters, to the observed individual cosponsorship decisions. The later is measured by the undirected network, with an i, j entry defined as the number of times i cosponsors j, j cosponsors i or i and j both cosponsor the same bill.

Table 6. Counterfactuals in c: Predicted (Proportional) Change in the (Mean) Probability of Bill Approval

Congress	**105**	**106**	**107**	**108**	**109**	**110**
Decrease in 1% in $\hat{c}$	-0.0188	-0.0181	-0.0164	-0.0156	-0.0169	-0.0162
Decrease in 10% in $\hat{c}$	-0.1809	-0.1775	-0.1623	-0.1573	-0.1677	-0.1632

The table presents the change (here, they are all *decreases*) in the average probability of bill approval under the counterfactual, where the estimated cost c, is reduced by either 1% or 10% (of the estimated value $\hat{c}$). We do this by calculating the the implied optimal $\{x_i^*, s_i^*\}$ under the estimated parameters in Table 3 (and replacing $\hat{c}$ by the appropriate counterfactual) using the system in Theorem 2.1, and calculate the probability of approval defined as $\gamma \sum_{j \in N} g_{i,j} x_i x_j$. We then find the percentual change over the predicted values under Table 3. Note that the estimated probability of approval here differs from the curves in Figures 13 as we are using the projected $\{x_i^*, s_i^*\}$ and not the observed one. This is to make a fair comparison across these two worlds, (where we do not observe values realized with measurement error in one of them). Finally, note that the percentage change in the other counterfactual, by decreasing γ but keeping ϕ constant is always 0.9, as that only moves the probability of approval proportionally (so we omit from the table, only showing the figures in Figure 15).

Table 7. Counterfactuals in α: Looking at the Changes in (Ex-Ante) predicted probability of Emergency Crisis bills in the 110th Congress, if the Republicans who lost their seats remained

Act	Proportional Change	Baseline Probability of Success
Emergency Economic Stabilization Act of 2008. (H.R. 1424) Sponsor: Patrick Kennedy, Democrat - RI	-0.0054	0.5256
Housing and Economic Recovery Act of 2008 (H.R. 3221) Sponsor: Nancy Pelosi, Democrat - CA	-0.0053	0.0003
Economic Stimulus Act of 2008 (H.R. 5140) Sponsor: Nancy Pelosi, Democrat - CA	-0.0053	0.0003
Supplementary Appropriations Act, 2008 (H.R. 2642) Sponsor: Chet Edwards, Democrat - TX	-0.0053	0.0887

The table presents the proportional change (Counterfactual/Predicted - 1) of the probability of each bill passing under our counterfactual scenario (where we replace the newly elected Democrats who took over Republican seats, by the Republicans they replaced). This is done by replacing the estimated α's of these new members by the estimated ones of those Republicans in Congress 109. We then calculate the projected probability of bill approval using our estimated parameters, for the estimated α and for the counterfactual distribution, as in Table 6. The baseline (predicted) probability is shown in the second column.

Appendix A. Identification

To proceed, we normalize $\alpha_1 = 1$, where agent $i = 1$ is the normalizer. Who this person is is immaterial, but, for comparison purposes, it is convenient that $i = 1$ is present in all τ. Hence, all α_j, $j \in \mathcal{N}$ will be relative values to the normalizer. As we cannot separately identify m from the scale parameter γ, we normalize $m = 1$.

Using Theorem 2.1 with $\alpha_1 = 1$, we have that $s^* = s_1^*$ and $x^* = x_1^*$. As $s_i^* = s_i - \upsilon_i$ and $x_i^* = x_i - \lambda_i$, we can rewrite the system in Theorem 2.1 as:

$$s_i - \upsilon_i = \alpha_i(s_1 - \upsilon_1) \tag{A.1}$$

$$x_i - \lambda_i = \alpha_i(x_1 - \lambda_1) \tag{A.2}$$

for all $i \in \mathcal{N}$, with:

$$(s_1 - \upsilon_1) = \phi\psi(x_1 - \lambda_1)^2 \tag{A.3}$$

$$1 = (x_1 - \lambda_1)[c - \phi\psi(s_1 - \upsilon_1)] \tag{A.4}$$

where $\psi = \frac{\sum_{i \in \mathcal{N}} \alpha_i^2}{\sum_{i \in \mathcal{N}} \alpha_i}$.

It follows from (A.1):

$$\mathbb{E}(s_1 - \upsilon_1)\alpha_j = \mathbb{E}(s_j - \upsilon_j) \quad \forall j \in \mathcal{N} \tag{A.5}$$

or

$$\alpha_j = \frac{\mathbb{E}(s_j)}{\mathbb{E}(s_1)} \tag{A.6}$$

where we have used the mean zero assumption on υ_j, $\quad \forall j \in \mathcal{N}$.

Hence, α_j is identified, since $\{s_i\}_{i \in \mathcal{N}}$ is observed. As this is true for any $j \in \mathcal{N}$, $\{\alpha_i\}_{i \in \mathcal{N}}$ is identified. It follows that ψ is now identified, as it is a function of $\{\alpha_i\}_{i \in \mathcal{N}}$. In particular,

$$\psi = \frac{\sum_{j \in \mathcal{N}} \left(\frac{\mathbb{E}(s_j)}{\mathbb{E}(s_1)}\right)^2}{\sum_{j \in \mathcal{N}} \left(\frac{\mathbb{E}(s_j)}{\mathbb{E}(s_1)}\right)}. \tag{A.7}$$

(A.3) implies that:

$$(s_1 - \upsilon_1) = \phi\psi(x_1 - \lambda_1)^2$$

so that

$$\left(\frac{s_i - \upsilon_i}{\alpha_i}\right) = \phi\psi\left(\frac{x_i - \lambda_i}{\alpha_i}\right)^2$$

and

$$(s_i - \upsilon_i)\,\alpha_i = \phi\psi(x_i^2 - 2\lambda_i x_i + \lambda_i^2),$$

where the second line uses that $x_i - \lambda_i = \alpha_i(x_1 - \lambda_1)$ and that $s_i - \upsilon_i = \alpha_i(s_1 - \upsilon_1)$.

Applying the Expectation operator to the above implies that:

$$\mathbb{E}s_i\alpha_i = \phi\psi\left(\mathbb{E}x_i^2 - 2\mathbb{E}\lambda_i x_i + \mathbb{E}\lambda_i^2\right),$$

and hence, that:

$$\mathbb{E}s_i\alpha_i = \phi\psi(\mathbb{E}x_i^2 + \sigma^2) \quad \forall i \in \mathcal{N} \tag{A.8}$$

In particular, note that for $i = 1$, $\mathbb{E}s_1 = \phi\psi(\mathbb{E}x_1^2 + \sigma^2)$. This captures that measurement error leads to an additional term: the variance of the measurement error is added on to the observable counterpart of equation $s^* = \phi\psi x^{*2}$.

Now, we can rewrite (A.4) as:

$$1 = \frac{x_i - \lambda_i}{\alpha_i}\left(c - \phi\psi\frac{s_i - \upsilon_i}{\alpha_i}\right)$$

or

$$\alpha_i = (x_i - \lambda_i)\left(c - \phi\psi\frac{s_i - \upsilon_i}{\alpha_i}\right),$$

where the first line uses that $x_i^* = \alpha_i x_1^*$ and $s_i^* = \alpha_i s_1^*$.

Applying the expectation operator to the above yields:

$$\begin{aligned}
\mathbb{E}\alpha_i &= \mathbb{E}\left((x_i - \lambda_i)[c - \phi\psi\left(\frac{s_i - \upsilon_i}{\alpha_i}\right)]\right) \\
&= \mathbb{E}\left(x_i c - \phi\psi x_i\frac{s_i}{\alpha_i} + \phi\psi\upsilon_i\frac{x_i}{\alpha_i} - \lambda_i c + \phi\psi\lambda_i\frac{s_i}{\alpha_i} - \phi\psi\lambda_i\frac{\upsilon_i}{\alpha_i})\right) \\
&= \mathbb{E}\left(x_i[c - \phi\psi\frac{s_i}{\alpha_i}]\right),
\end{aligned}$$

where the third step uses that $\mathbb{E}(\upsilon_i \mid x_i, \alpha_i) = \mathbb{E}(\lambda_i \mid s_i, \alpha_i) = 0$, and the independence of λ_i and υ_i guarantees that the last term is 0 in expected value. Hence, we have that:

$$0 = \mathbb{E}\left(\alpha_i - x_i\left(c - \phi\psi\frac{s_i}{\alpha_i}\right)\right) \tag{A.9}$$

In particular, we note that for $i = 1$, this equation is $0 = 1 - \mathbb{E}x_1(c - \phi\psi s_1)$, which is the counterpart to the original one in Theorem 2.1. Since the measurement error is additive, it cancels out to leave the same structure. The original equation no longer holds exactly, but in expectation.

Consider that (A.2) is valid for every individual. Then let, $i, j \in \mathcal{N}$ be such that $i \neq k, j \neq k$, but arbitrary. Multiplying (A.2) for i and j implies that:

$$(x_k - \lambda_k)(x_j - \lambda_j) = \alpha_k \alpha_j \left(\frac{x_i - \lambda_i}{\alpha_i} \right)^2$$

or

$$x_k x_j + \lambda_k \lambda_j - \lambda_k x_j - \lambda_j x_k = \frac{\alpha_k \alpha_j}{\alpha_i^2}(x_i^2 - 2x_i\lambda_i + \lambda_i^2)$$

Applying expectation on both sides, using that $\mathbb{E}\lambda_k\lambda_j = 0$ by independence and mean zero measurement error, and $\mathbb{E}\lambda_k x_j = \mathbb{E}\lambda_j x_k = 0$ yields:

$$\mathbb{E}x_k x_j = \frac{\alpha_k \alpha_j}{\alpha_i^2}(\mathbb{E}x_i^2 + \mathbb{E}\lambda_i^2) \tag{A.10}$$

$$\frac{\mathbb{E}x_k x_j \alpha_i^2}{\alpha_k \alpha_j} = \mathbb{E}x_i^2 + \sigma^2 \tag{A.11}$$

$$\sigma^2 = \frac{\mathbb{E}x_k x_j \alpha_i^2}{\alpha_k \alpha_j} - \mathbb{E}x_i^2 \tag{A.12}$$

Hence, σ^2 is identified. Intuitively, this expression for σ^2 captures that the measurement error induces variance in excess of what predicted by the model. The model predicts that x_k, x_j will be as correlated as "α_m/α_i" multiples of x_i, with $m = k, j$. However, in the data we might observe correlation in excess of that.

Since this equation is valid for every i, j, k such that $i \neq j$, $i \neq k$, our equivalent moment condition is done by averaging first over $j \in \mathcal{N}$, then by averaging over $k \in \mathcal{N}$, giving us enhanced stability in our procedure. The moment is then written, as the other equations, in terms of i.

Since (A.8) is a function of only the parameters ψ, σ^2, ϕ, and we now have identified ψ, σ^2, we can replace (A.12) into (A.8) and recover ϕ uniquely (as it is linear in ϕ). Finally, with ϕ and σ^2, equation (A.9) recovers c uniquely, as it is linear in c.

To conclude we derive the identification of γ. If we knew the measurement error, we would have from (2.2), setting $m = 1$:

$$
\begin{aligned}
Pr(y_i = 1 \mid \{\lambda_j, \upsilon_j\}_{j\in\mathcal{N}}) &= Pr(Y_i > m \mid \{\lambda_j\}_{j\in\mathcal{N}}) \\
&= \left(\frac{\gamma}{m}\right)\left(\sum_{j\in\mathcal{N}} g_{i,j}(\mathbf{s}^*)x_j^*\right)x_i^* \\
&= \gamma(\sum_{j\in\mathcal{N}} g_{i,j}(\mathbf{s}^*)(x_j-\lambda_j)(x_i-\lambda_i)) \\
&= \gamma\sum_{j\in\mathcal{N}}\frac{s_i^*s_j^*}{\sum_{k\in\mathcal{N}} s_k^*}(x_jx_i+\lambda_j\lambda_i-\lambda_jx_i-\lambda_ix_j) \\
&= \gamma\sum_{j\in\mathcal{N}}\frac{(s_i-\upsilon_i)(s_j-\upsilon_j)}{\sum_{k\in\mathcal{N}}(s_k-\upsilon_k)}(x_jx_i+\lambda_j\lambda_i-\lambda_jx_i-\lambda_ix_j)
\end{aligned}
$$

However, we do not observe the measurement error. We first make an approximation to remove the term $\sum_{k\in\mathcal{N}}\upsilon_k$ from the denominator. We note that, for a large N (as we have in the data), that term must be negligible as the measurement error is i.i.d. across agents and mean zero. Set:

$$
\begin{aligned}
&\frac{(s_i-\upsilon_i)(s_j-\upsilon_j)}{\sum_{k\in\mathcal{N}}(s_k-\upsilon_k)} \\
=\ &\frac{\frac{1}{N}(s_i-\upsilon_i)(s_j-\upsilon_j)}{\frac{1}{N}\sum_{k\in\mathcal{N}}(s_k-\upsilon_k)} \\
=\ &\frac{\frac{1}{N}(s_i-\upsilon_i)(s_j-\upsilon_j)}{\frac{1}{N}\sum_{k\in\mathcal{N}}s_k-\frac{1}{N}\sum_{k\in\mathcal{N}}\upsilon_k} \\
=\ &\frac{\frac{1}{N}(s_i-\upsilon_i)(s_j-\upsilon_j)}{\frac{1}{N}\sum_{k\in\mathcal{N}}s_k-o_P(1)} \\
=\ &\frac{\frac{1}{N}(s_i-\upsilon_i)(s_j-\upsilon_j)}{\frac{1}{N}\sum_{k\in\mathcal{N}}s_k}-o_P(1) \\
=\ &\frac{(s_i-\upsilon_i)(s_j-\upsilon_j)}{\sum_{k\in\mathcal{N}}s_k}-o_P(1)
\end{aligned}
$$

where we are taking N to be large (as we see in the data), such that the correlation between υ_i and $\sum_{k\in\mathcal{N}}\upsilon_k/N$ is negligible.

With this, we can substitute back to get:

$$
\begin{aligned}
&Pr(y_i = 1 \mid \{\lambda_j, \upsilon_j\}_{j\in\mathcal{N}}) = \\
&\gamma\sum_{j\in\mathcal{N}}\frac{(s_i-\upsilon_i)(s_j-\upsilon_j)}{\sum_{k\in\mathcal{N}}s_k}(x_jx_i+\lambda_j\lambda_i-\lambda_jx_i-\lambda_ix_j)-o_P(1)
\end{aligned}
$$

Integrating out the measurement errors, which are independent of each other and i.i.d. for all agents, we get, for large N almost exactly:

$$Pr(y_i = 1) = \gamma \mathbb{E}\left(\sum_{j \in \mathcal{N}} \frac{s_i s_j}{\sum_{k \in \mathcal{N}} s_k} x_j x_i\right), \tag{A.13}$$

which implies that γ is identified.

Note that we can rewrite (A.13) as a moment condition, using that $\mathbb{E}(y_i) = Pr(y_i = 1)$ since y_i is an indicator function and that ε_i is i.i.d. across bills:

$$\mathbb{E}\left(y_i - \gamma \sum_{j \in \mathcal{N}} \frac{s_i s_j}{\sum_{k \in \mathcal{N}} s_k} x_j x_i\right) = 0 \tag{A.14}$$

Multiplying both sides by ζ, and using that $\phi = \zeta\gamma$, we can recover ζ.

Hence, the parameters $(\{\alpha_i\}_{i \in \mathcal{N}}, \phi, c, \psi, \gamma, \zeta, \sigma^2)$ are identified.