# Estimating Candidate Valence*

Kei Kawai†  Takeaki Sunada‡
U.C. Berkeley  University of Rochester

December 3, 2024

**Abstract**

We estimate valence measures of candidates running in U.S. House elections from data on vote shares. Our identification and estimation strategy builds on ideas developed for estimating production functions, allowing us to control for possible endogeneity of campaign spending and sample selection of candidates due to endogenous entry. We find that incumbents have substantially higher valence measures than challengers running against them, resulting in about 3.8 percentage-point differences in the vote share, on average. Eliminating differences in the valence of challengers and incumbents results in an increase in the winning probability of a challenger from 6.2% to 12.2%. Our measure of candidate valence can be used to study various substantive questions of political economy. We illustrate its usefulness by studying the source of incumbency advantage in U.S. House elections.

Key words: Candidate Valence, Production Function, Dynamic Game, Incumbency Advantage
JEL classification: D72, D24

*We thank Daniel Diermeier, Igal Hendel, Aviv Nevo, Walter Stone, and Michael Whinston, for insightful comments.
†Department of Economics, U.C. Berkeley Email: kei@berkeley.edu
‡Simon Business School, University of Rochester Email: tsunada@simon.rochester.edu

# 1 Introduction

In many models of political economy, political candidates are not only horizontally differentiated through ideology but vertically differentiated through valence. Vertical differences among candidates affect various aspects of political competition such as policy convergence between candidates, alignment of policy with the preferences of the voters, and campaign finance decisions of the candidates. While the development of various measures of horizontal differentiation, such as the DW-NOMINATE scores (Poole and Rosenthal 1985) and donation-based ideology measures (Bonica, 2023), has made empirical studies of political ideology possible, there has not been as much progress on the development of measures of candidate valence. This has made it challenging to empirically study models of political economy with vertical differentiation among candidates.

In this study, we identify and estimate a metric of candidate valence from data on vote shares. Although the vote share of each candidate should, in principle, be informative about candidate valence, we need to isolate the effect of valence from other factors such as campaign activities of the candidates. We model the vote shares in each election as a random variable that is determined, in part, by candidates' endogenous campaign spending, their policy positions and their valence. The valence of the candidate corresponds to a candidate specific constant term (unobservable to the researcher) in the equation that determines votes. Our measure of valence is in units of vote share, capturing the differences in expected vote shares across candidates holding everything else constant. Our measure contrasts with those based on an index of observable candidate characteristics, which may miss aspects of candidate valence that are unobservable to researchers, or survey-based measures that do not allow for a straightforward interpretation of their magnitude.

In order to isolate the effect of valence on the vote share, we need to account for the fact that campaign activities of the candidates are endogenous and that candidate entry and exit induce selection in the valence of the candidates that compete. To overcome these challenges, we exploit a natural parallel between recovering candidate valence from vote shares and recovering unobserved firm-level productivity from firm output. In the context of production function estimation, the input decisions and entry/exit decisions of the firms depend on the unobserved (to the researcher) productivity of the firms. This dependence gives rise to endogeneity and sample selection bias similar to the one considered here. The approaches developed for estimating production functions enable one to recover the production function and firm productivity measures even in the presence of endogeneity and sample selection.

We adapt the control function approach developed in Olley and Pakes (1996) to our setup. Specifically, we embed the model of vote shares into a dynamic game with spending, fund-raising, savings, challenger entry and incumbent retirement and show that the incumbents' policy functions are one-to-one between the valence and the observed actions. The one-to-one property allows us to recover the valence of incumbents as a function of their actions. We also use the model of challengers' entry decision to derive a sufficient statistic for the valence of the challengers that choose to enter. The structure of the dynamic game allows us to link the vote shares to the valence of the candidates while controlling for endogeneity of spending and selection of candidates. The dynamic game also provides optimality conditions of the candidates' problem that serve as extra moment conditions for identification.

We use our approach to estimate the valence of each candidate running for U.S. House elections between 1984 and 2008. Our estimates suggest that there are substantial differences in the valence measures between incumbents and challengers. We find that the average valence measure among incumbents is about 3.8 percentage points higher in terms of vote share than among challengers who run against them. We also find larger dispersion of valence measures among challengers than among incumbents. The inter-quartile range of valence measures among incumbents is about 3.1 percentage points, while that among challengers is about 7.7 percentage points. Our findings are consistent with the fact that incumbents are selected partly by valence. Eliminating the differences in the valence between incumbents and challengers increases the challenger's winning probability from 6.2% to 12.2%.

Regarding open-seat candidates, we find that the upper tail of the valence distribution resembles that of the incumbents. However, unlike the valence distribution of the incumbents, there is a substantial fraction of low valence candidates that increases the dispersion of the distribution. The mean valence measure among open-seat candidates is comparable to that among challengers who run against incumbents. The inter-quartile range is about 8.1 percentage points.

Our measure of candidate valence can be used to study various substantive topics in political economy. To illustrate its usefulness, we study incumbency advantage in U.S. House elections. In particular, using the valence measures of candidates as outcome variables, we build on the regression discontinuity design used in Lee (2008) to identify the incumbency effect that can be attributed to differences in candidate valence. We also use the regression discontinuity design to identify the incumbency effect that is attributable to differential spending and policy positions between incumbents and challengers. We hence offer a decomposition of the incumbency advantage identified in Lee (2008). Our results imply that about 35 percent of the incumbency advantage is explained by differences in valence. We also find that differences in spending account for 45 percent of the incumbency advantage, and differences in policy positions account for about 20 percent. Differences in policy positions explain part of the incumbency advantage because challengers adopt more extreme policy positions than incumbents, on average. Our finding that spending accounts for less than half of the incumbency advantage suggests that policy interventions designed to eliminate incumbency advantage through the spending channel (e.g., subsidizing challengers' campaigns) will be only partially effective.

**Literature** Candidate valence plays an important role in many models of political competition. Differences in the valence of candidates affect convergence of platforms between candidates and alignment of policy with the preferences of the voters (Aragones and Palfrey 2004, Carter and Patty 2015, Buisseret and Van Weelden 2022). Candidate valence also plays a significant role in models of political selection (e.g., Snyder and Ting 2011, Serra 2011, Adams and Merrill III 2008). Despite their importance, existing measures of candidate valence for Congressional candidates have largely relied on candidate characteristics such as candidates' occupation, political experience, legislative accomplishment, etc. (e.g., Green and Krasno 1988, Maestas and Rugeley 2008), or on survey responses (e.g., Stone et. al. 2010, and Stone and Simas 2010). However, valence measures based on candidate characteristics miss potentially important components of valence that are unobservable to the researcher, and those based on surveys typically do not allow for a straightforward interpretation of their magnitude. Because our measures of valence correspond to candidate fixed effects in the model of vote shares, they capture any observable and unobservable candidate specific characteristics that affect votes. In addition, our measures are defined in units of vote share, allowing for a straightforward interpretation of their magnitude.[1]

An important part of our empirical exercise is to separately identify the effect of campaign spending, candidate valence and other district characteristics on the vote share. In

[1] Other papers that estimate candidate valence as latent variables include Rekkas (2007), for the Canadian Parliament, Gordon and Harmann (2013), for the U.S. President, Kendall, Nannicini, and Trebbi (2015), for an Italian mayor, Sieg and Yoon (2017), for U.S. Governors, Montero (2023), for Mexican Chamber of Deputies, Frey et. al. (2021), for Mexican mayors, Iaryczower et. al. (2023), for Brazilian Chamber of Deputies.

this regard, our paper is related to the extensive literature that estimates the causal effect of candidate spending on the vote share, including Jacobson (1978), Green and Krasno (1988), Levitt (1994), Gerber (1998), Erikson and Palfrey (2000), and da Silveira and de Mello (2011).[2] The key difference between our paper and previous work is that we focus on identifying candidate valence. Much of the previous work has treated candidate valence as nuisance parameters, for example, by differencing them out.

Finally, our paper contributes to the study of incumbency advantage. Starting from the early work of Erikson (1971), various approaches have been used to identify the incumbency advantage in U.S. Congressional elections. Gelman and King (1990) and Levitt and Wolfram (1997) use panel data methods. Gowrisankaran et. al. (2008) exploit reelection probabilities as a function of the number of terms since an open seat election and the incumbent's tenure.[3] Ansolabehere et. al. (2000) use legislative redistricting and Lee (2008) uses regression discontinuity around a vote share of 50%.[4] Our analysis of incumbency advantage in Section 6 is based on Lee (2008).

# 2 Model

**Overview** We embed a model of vote shares in a dynamic model of U.S. House elections with endogenous spending, saving, policy positioning, entry and retirement decisions. In each period $t$ ($t = 1, 2, ...\infty$), there is a stage game which is either an election with an incumbent seeking re-election or an open-seat election. In an election with an incumbent, potential challengers from the out-party (i.e., not the incumbent's party) decide whether or not to enter, and conditional on challenger entry, the incumbent and the challenger simultaneously make spending, saving, and fund-raising decisions. We model the vote share as a function of the spending, policy position and the valence of the candidates, state variables (such as district characteristics) and a random shock. The winner becomes the incumbent next period. An open-seat election is the same as an election with an incumbent except that challengers from both parties make entry decisions. The time between the periods is two years, because Congressional elections take place every two years.

[2] See Stratmann (2005) for a survey.

[3] Gowrisankaran et. al. (2008) estimate the distribution of candidate valence as in our paper. They impose parametric functional form restrictions on the distribution.

[4] Levitt and Wolfram (1997) decompose the incumbency advantage into direct officeholder benefits, the ability of incumbents to scare off high quality challengers, and higher average quality of incumbents vis-à-vis the typical open-seat candidate. Ansolabehere et. al. (2000) separates the electoral benefits of "homestyle" from other sources of incumbency advantage.

**Sequence of Events within the Stage Game** In an election with an incumbent, events occur in the following order:

1. Nature draws $N \in \{0, 1, 2, \cdots\}$, the number of potential challengers from the out-party according to a distribution $F_N$. The valence (*q*uality) and the policy position of the potential candidates, $\{\{q_{C,1}, p_{C,1}\}$, $\{q_{C,2}, p_{C,2}\}$, ..., $\{q_{C,N}, p_{C,N}\}\}$, are drawn independently across candidates according to $F_{q_C, p_C}$. We do not consider entry for the incumbent's party.[5] Each potential challenger observes the current state, own and incumbent valence as well as own and incumbent policy position. Potential challengers simultaneously make entry decisions by comparing the value of entering and the cost of entry, $\kappa$.

2(a). If exactly one challenger enters, that challenger becomes the party nominee. In the general election, the incumbent and the nominee of the out-party simultaneously decide how much to spend, raise and save, taking as given own and opponent's valence and policy position. The vote shares are determined as a function of the spending, the valence, the policy position of the candidates, state variables and a random shock.

If $M$ ($1 < M \leq N$) potential challengers enter, there is a Primary election. We do not explicitly model the Primary, but we assume that an entrant with valence $q_{C,m}$ and policy position $p_{C,m}$ is selected to be the party nominee with probability $\pi(q_{C,m}, p_{C,m}, \mathbf{q}_{C,-m}, \mathbf{p}_{C,-m})$, where $\mathbf{q}_{C,-m} = (q_{C,1}, \ldots, q_{C,m-1}, q_{C,m+1}, \ldots, q_{C,M})$ and $\mathbf{p}_{C,-m}$ is defined analogously for policy positions. The party nominee competes against the incumbent in the general election.

2(b). If no potential challenger enters, the incumbent decides how much to spend, raise and save, and the incumbent becomes the winner with probability one.

3. The winner of the election receives utility $B$. State variables such as the incumbent's war chest and district characteristics evolve from current values to the next. Before the start of the next period, the winner chooses to retire or run for reelection. She also chooses the next period policy position. Conditional on running for reelection, the winner becomes the incumbent next period with war chest determined by the amount of money she saved in the previous period. If the incumbent retires, the stage game of the next period becomes an open-seat election.

[5] Almost all incumbents in our sample become the party nominee, barring a major scandal. See Online Appendix 10.7 (Kawai and Sunada, 2024) for the set of elections we drop due to scandals.

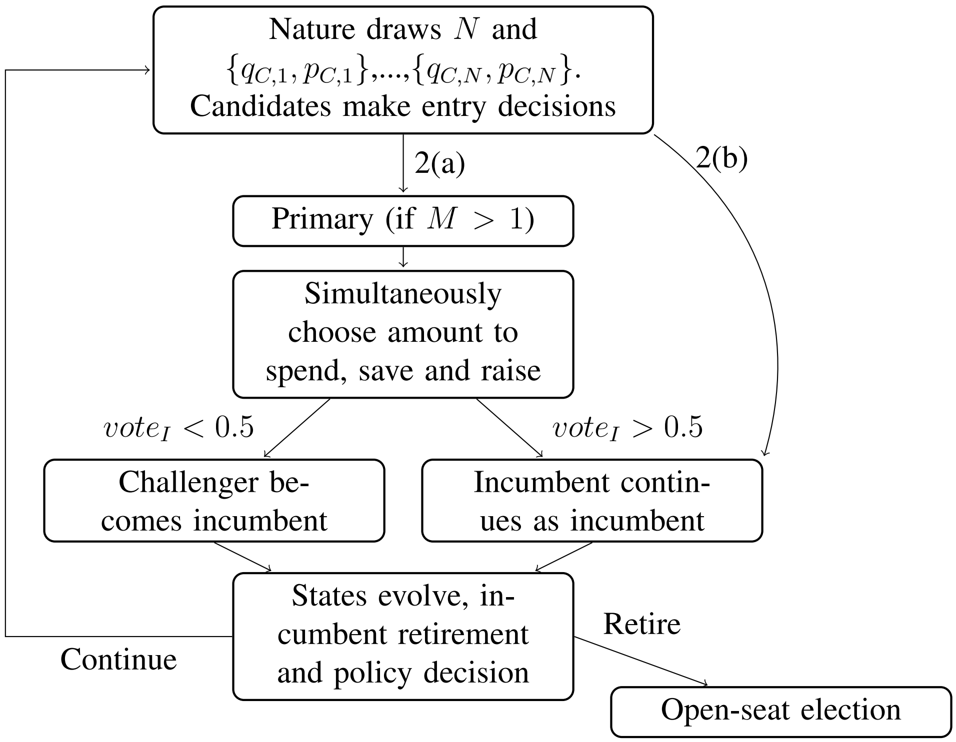



Figure 1: Timeline of the Stage Game for an Election with an Incumbent

Figure 1 illustrates the timeline of an election with an incumbent. $vote_I$ denotes the incumbent's vote share. We assume that the policy position of the incumbent is determined prior to the beginning of the stage game as a result of her voting record while in office. Policy positions of the potential challengers are drawn exogenously at the beginning of the stage game, but both the entry decision of the challengers and the Primary make the policy position of the general election challenger endogenous.

In open-seat elections, potential challengers from both parties make simultaneous entry decisions and the candidate selection process described in steps 1 and 2 applies to both parties. Once a single candidate is selected from each party, they compete in the general election in a way analogous to elections with an incumbent. Because the model of open-seat elections is similar to that of elections with an incumbent, we focus on the case of elections with an incumbent in the following discussion. A full description of the model of open-seat elections is given in Online Appendix 10.1 (Kawai and Sunada, 2024).

**The Vote Share Equation** We begin by describing how vote shares are determined in the general election (last step of case 2(a)). We specify the vote share equation as follows:

$$\begin{aligned} vote_I = &\beta_I \ln d_I + \beta_C \ln d_C + \beta_P (p_I - p^*)^2 - \beta_P (p_C - p^*)^2 \\ &+ \beta_{ten} ten_I + \beta_X X + q_I - q_C + \varepsilon, \end{aligned} \tag{1}$$

where $vote_I$ is the incumbent's vote share, $\ln d_I$ is the log spending (*d*isbursement) of the incumbent and $\ln d_C$ is the log spending of the challenger. We let $p_I$ and $p_C$ denote the policy positions of the incumbent and the challenger, and $p^*$ denote the ideal policy position of the voters in the district. We assume that the vote share depends on the distance between the policy position of the candidates and the ideal position of the district. The term $ten_I$ is the tenure of the incumbent (i.e., the number of consecutive terms served in office), $X$ is a vector of exogenous control variables such as district characteristics, $q_I$ and $q_C$ are the valence of the incumbent and the challenger, and $\varepsilon$ is a random shock.

The valence terms, $q_I$ and $q_C$, capture the candidate's ability to attract votes.[6] They may include candidates' personal traits such as name recognition, perceived leadership skills, public speaking skills, etc. They are unobserved to the researcher but observed by the candidates. Because candidates observe $q_I$ and $q_C$ when making their decisions, $\ln d_I$ and $\ln d_C$ are potentially correlated with $q_I$ and $q_C$. On the other hand, $\varepsilon$ is not observed to the candidates, and hence orthogonal to their decisions.

We assume that the error term, $\varepsilon$, follows a Normal distribution with mean $0.5$ and variance $\sigma_\varepsilon^2$. The probability that the incumbent wins, i.e. the probability that the vote share of the incumbent exceeds 0.5, can be written as follows:

$$\begin{aligned} &\Pr(vote_I > 0.5) = \\ &\Phi\left(\frac{1}{\sigma_\varepsilon}(\beta_I \ln d_I + \beta_C \ln d_C + \beta_P (p_I - p^*)^2 - \beta_P (p_C - p^*)^2 + \beta_{ten} ten_I + \beta_X X + q_I - q_C)\right), \end{aligned} \tag{2}$$

where $\Phi$ is the c.d.f. of a standard Normal distribution.

[6] We assume that candidate valence is a scalar unobservable. In this regard, our setting is similar to the one considered in Olley and Pakes (1996). In that paper, firms are heterogeneous with regard to a scalar Hicks-neutral productivity term. For models with multidimensional firm heterogeneity, see e.g., Doraszelski and Jaumandreu (2018) and Demirer (2022).

**Incumbent's Problem** Consider the problem of the incumbent when the incumbent faces a challenger (after the Primary in case 2(a)). For a given strategy of the challenger, a contested incumbent solves the following dynamic programming problem:

$$v_I(\mathbf{s}, q_C, p_C) = \max_{d_I \geq 0, w_I' \geq 0} u_I + \delta \Pr(vote_I > 0.5) \mathbb{E}_{\mathbf{s}'|\mathbf{s}}[(1 - \lambda(\mathbf{s}')) V_I(\mathbf{s}')],$$
$$\text{where } u_I = B \cdot \Pr(vote_I > 0.5) - C_I(w_I' + d_I - w_I; q_I) + H_I(d_I). \tag{3}$$

The incumbent chooses the amount of spending, $d_I$, and the amount of savings, $w_I'$, given the challenger's valence $q_C$, challenger's policy position $p_C$, and the state s, where $\mathrm{s} = \{q_I, w_I, ten_I, p_I, p^*, X\}$. The term $w_I$ is the war chest of the incumbent at the beginning of the period. Note that s includes the incumbent's valence $q_I$, and policy position, $p_I$. The flow payoff, $u_I$, consists of four terms. $B$ is the utility from winning and it is multiplied by the probability of winning, given by Expression (2). The term $C_I(\cdot)$ captures the costs that the incumbent incurs from raising money. The amount raised by the incumbent is the sum of future savings, $w_I'$, and spending, $d_I$, less the war chest, $w_I$. We let $C_I(\cdot)$ depend on $q_I$ and assume that the marginal cost of raising money is strictly decreasing in $q_I$, so that candidates with higher valence have lower marginal cost of raising money. The term $H_I(\cdot)$ represents the consumption value of spending. It captures the fact that candidates can and do spend money on things that seem only tangentially related to increasing his or her vote share. For instance, candidates can spend money on charitable donations, buy gifts and meals, etc.[7]

The second term of the value function is the continuation value which is a product of the discount factor $\delta$, the probability of winning, and the continuation payoff $\mathbb{E}_{\mathbf{s}'|\mathbf{s}}[(1 - \lambda(\mathrm{s}'))V_I(\mathrm{s}')]$. The term, $1 - \lambda(\mathrm{s}')$, denotes the probability that the incumbent runs for reelection and $V_I(\mathrm{s}')$ denotes the next period's (ex-ante) value function, both of which are defined below. The expectation of the next period's value function is taken with respect to $X'$, $p_I'$, and $p^{*\prime}$.[8]

When no challenger enters (case 2(b)), the election is uncontested and the incumbent wins with probability one. The problem of the incumbent in an uncontested election is as

[7] In the estimation, we allow for the possibility that $H_I(\cdot) = 0$. The reason for including $H_I(\cdot)$ is to account for the fact that there is incumbent spending even in periods when the incumbent seems almost certain to win.

[8] We assume that if an incumbent retires, she receives zero payoff thereafter.

follows:

$$\tilde{v}_I(\mathbf{s}) = \max_{d_I \geq 0, w'_I \geq 0} \tilde{u}_I + \delta \mathbb{E}_{\mathbf{s}'|\mathbf{s}}[(1-\lambda(\mathbf{s}'))V_I(\mathbf{s}')],$$
$$\text{where } \tilde{u}_I = B - \tilde{C}_I(w'_I + d_I - w_I; q_I) + \tilde{H}_I(d_I). \tag{4}$$

The term $\tilde{u}_I$ is the period utility of the incumbent when she is uncontested, and the expression is obtained by replacing $\Pr(vote_I > 0.5)$ with 1, and by replacing $C_I(\cdot)$ and $H_I(\cdot)$ with $\tilde{C}_I(\cdot)$ and $\tilde{H}_I(\cdot)$ in expression (3). $\tilde{C}_I(\cdot)$ and $\tilde{H}_I(\cdot)$ are the costs of raising money and the consumption value from spending in uncontested periods, respectively. We assume that the marginal cost of raising money is strictly decreasing in $q_I$.

After each election, state variables evolve. We assume that $X$ and the ideal policy position of the electorate, $p^*$, follow an exogenous Markov process. We assume a deterministic transition for $w_I$, $ten_I$ and $q_I$. The incumbent war chest in the next period equals the amount that the incumbent saves in the current period plus $10\%$ interest.[9] The tenure of the incumbent increases by 1 as $ten'_I = ten_I + 1$. We assume that $q_I$ is constant over time. While this is restrictive, we can account for deterministic trends in electoral strength through $ten_I$. Allowing for $q_I$ to evolve stochastically is conceptually straightforward, but the estimation of such a model becomes data-intensive. We discuss this point in detail in Online Appendix 10.8.

While the incumbent is in office, and prior to the next election, the incumbent decides what her policy position will be in the next election, and whether or not to retire. We let $p_I(\cdot)$ denote the policy function that determines the incumbent's policy position in the next period, and $\lambda(\cdot)$ denote the policy function for retirement. We let $p_I(\cdot)$ and $\lambda(\cdot)$ be stochastic. These policy functions may result from dynamic optimization problems, but we do not explicitly model them, instead treating them as probability distributions over outcomes to be estimated directly from the data.[10] Because these choices are endogenous, we let these distributions depend on the full set of state variables, $q_I, w'_I$, $ten'_I, p^{*\prime}$, $X'$, and her previous policy position, $p_I$.[11]

[9]Since the time between periods is two years, a 10% interest implies an annual interest of about 5%.

[10]Choice of policy may depend on electoral considerations as well as those outside of our model, such as within-party dynamics in Congress. We opt for a reduced-form representation of the policy choice, but the policy function, $p_I(\cdot)$, may result from these strategic considerations. Similarly, the retirement choice $\lambda(\cdot)$, may also be the result of an optimal exit decision. For an explicit model of exit, see Diermeier et. al. (2005). In that paper, the authors estimate a career model in which House members can choose to run for reelection, run for Senate, or exit, where exit can mean working for the private sector, public sector, or retiring.

[11]In expression (3) and in what follows, the exit probability is expressed as $\lambda(\mathbf{s}')$. However, $\lambda(\cdot)$ depends on the previous policy position, $p_I$, and not on the current policy position $p'_I$. Although this is an abuse

The incumbent's ex-ante value function at the beginning of the stage game is as follows:

$$V_I(\mathbf{s}) = (1 - P_e(\mathbf{s}))\tilde{v}_I(\mathbf{s}) + P_e(\mathbf{s}) \int_{q_C,p_C} v_I(\mathbf{s}, q_C, p_C) dG_{q_C,p_C}(q_C, p_C|\mathbf{s}), \tag{5}$$

where $P_e(\mathbf{s})$ is the probability that a challenger enters. The value function consists of two terms: the first is the value of the incumbent when she is uncontested and the second is the value when she is contested. Because the challenger's valence and policy position, $\{q_C, p_C\}$, are uncertain at the beginning of the period, we take the expectation of $v_I(\mathbf{s}, q_C, p_C)$ with respect to $q_C$ and $p_C$. We denote their joint distribution by $G_{q_C,p_C}(\cdot|\mathbf{s})$.

Note that $P_e(\mathbf{s})$ and $G_{q_C,p_C}(\cdot|\mathbf{s})$ are equilibrium objects that are endogenously determined by the challengers' entry decisions. We describe how $P_e(\mathbf{s})$ and $G_{q_C,p_C}(\cdot|\mathbf{s})$ are determined by the problem of the challenger below.

**Challenger's Problem** The problem of a challenger who competes in the general election is given as follows:

$$\begin{aligned} v_C(\mathbf{s}, q_C, p_C) = \max_{d_C \geq 0, w'_C \geq 0} & B \cdot \Pr(vote_I < 0.5) - C_C(w'_C + d_C, q_C) \\ & + H_C(d_C) + \delta \Pr(vote_I < 0.5)\mathbb{E}_{\mathbf{s}'|\mathbf{s}}[(1 - \lambda(\mathbf{s}'))V_I(\mathbf{s}')]. \end{aligned} \tag{6}$$

The challenger chooses the amount of spending, $d_C$, and the amount of savings, $w'_C$, given her own valence $q_C$, policy position $p_C$, and s, where $\mathbf{s} = \{q_I,\, w_I,\, ten_I,\, p_I,\, p^*,\, X\}$. The terms $C_C(\cdot)$ and $H_C(\cdot)$ capture the challenger's cost of fund-raising and personal benefit from spending, respectively. Challengers start out with no war chest. The challenger's next-period value function is the same as that of the incumbent, $V_I$, because the challenger becomes the incumbent if she wins. The next-period value, $V_I$, depends on the valence of the challenger, savings from period $t$ (plus 10% interest), tenure ($= 1$), the policy position in the next period (chosen by the winning challenger prior to the next election), and the vector of exogenous variables, $p^{*\prime}$ and $X'$, so that $\mathbf{s}' = \{q_C,\, 1.1w'_C,\, 1,\, p'_C,\, p^{*\prime},\, X'\}$.[12] The next period policy position $p'_C$ is determined by the policy function $p_I(\cdot)$.

We now consider the challenger's entry decision. We assume that challengers make simultaneous entry decisions by comparing the value of entering with the cost of entry, $\kappa$. The value of entering is the product of the challenger's value function, $v_C(\mathbf{s}, q_C, p_C)$, and

of notation, we opted to express the retirement probability as $\lambda(\mathbf{s}')$ instead of $\lambda(q_I, w'_I, ten'_I, p_I, p^{*\prime}, X')$ in order to keep the expressions simple.

[12]We define the tenure of a candidate running for the first time as an incumbent to be one.

the probability of winning the primary, which we denote by $p(\mathbf{s}, q_C, p_C)$. The probability of winning the Primary is as follows:

$$p(\mathbf{s}, q_{C,m}, p_{C,m}) = \mathbb{E}_M \left[ \int \pi(q_{C,m}, p_{C,m}, \mathbf{q}_{C,-m}, \mathbf{p}_{C,-m}) dF_{\mathbf{q}_{C,-m},\mathbf{p}_{C,-m}}(\mathbf{q}_{C,-m}, \mathbf{p}_{C,-m}|M, \mathbf{s}) \middle| \mathbf{s} \right], \tag{7}$$

where $\pi(q_{C,m}, p_{C,m}, \mathbf{q}_{C,-m}, \mathbf{p}_{C,-m})$ is the probability that entrant $m$ wins the primary if her valence and policy position are $\{q_{C,m}, p_{C,m}\}$ and those of her opponents are $\{\mathbf{q}_{C,-m}, \mathbf{p}_{C,-m}\}$. We do not explicitly model the Primary but instead let $\pi(\cdot)$ represent the candidate selection process in a reduced-form way. We let $\pi(\cdot)$ be flexible, only requiring $\pi(\cdot)$ to be symmetric across candidates and increasing in own valence.[13] To obtain the ex-ante probability of winning the Primary, $p(\mathbf{s}, q_C, p_C)$, we integrate $\pi(q_{C,m}, p_{C,m}, \mathbf{q}_{C,-m}, \mathbf{p}_{C,-m})$ with respect to the valence measures and the policy positions of the opponents, $\{\mathbf{q}_{C,-m}, \mathbf{p}_{C,-m}\}$, and the number of total entrants, $M$. We denote by $F_{\mathbf{q}_{C,-m},\mathbf{p}_{C,-m}}(\mathbf{q}_{C,-m}, \mathbf{p}_{C,-m}|M, \mathbf{s})$ the joint distribution of the opponents' valence measures and policy positions conditional on $M$ and the state $\mathbf{s}$. The distribution of $\mathbf{q}_{C,-m}$, $\mathbf{p}_{C,-m}$ and $M$ are both endogenous.

Each potential challenger chooses to enter if the value of entry, $p(\mathbf{s}, q_C, p_C)v_C(\mathbf{s}, q_C, p_C)$, is higher than the entry cost, $\kappa$. This implies that, as long as $p(\mathbf{s}, q_C, p_C)v_C(\mathbf{s}, q_C, p_C)$ is increasing in $q_C$, the entry decision of a challenger can be expressed by the following cutoff rule in $q_C$:[14]

$$\chi(\mathbf{s}, q_C, p_C) = \begin{cases} 1\text{: if } q_C > \bar{q}_C(\mathbf{s}, p_C) \\ [0,1]\text{: if } q_C = \bar{q}_C(\mathbf{s}, p_C) \\ 0\text{: if } q_C < \bar{q}_C(\mathbf{s}, p_C) \end{cases},$$

where $\bar{q}_C(\mathbf{s}, p_C)$ is defined implicitly as the solution to $p(\mathbf{s}, \cdot, p_C)v_C(\mathbf{s}, \cdot, p_C) - \kappa = 0$. $\bar{q}_C$ is the type of challenger that is indifferent between entering and not entering.

[13]When exactly one potential challenger enters ($M = 1$), the entrant becomes the party nominee with probability 1. One example of $\pi(\cdot)$ that satisfies our assumptions is the Tullock contest function.

[14]Because we assume that $\pi(q_{C,m}, p_{C,m}, \mathbf{q}_{C,-m}, \mathbf{p}_{C,-m})$ is increasing in $q_{C,m}$, $p(\mathbf{s}, q_C, p_C)$ is increasing in $q_C$ by assumption. Given that $p(\mathbf{s}, q_C, p_C)$ and $v_C(\mathbf{s}, q_C, p_C)$ are non-negative, $p(\mathbf{s}, q_C, p_C)v_C(\mathbf{s}, q_C, p_C)$ is increasing in $q_C$ when $v_C(\mathbf{s}, q_C, p_C)$ is not too decreasing in $q_C$. Because higher $q_C$ lowers fund-raising costs and increases the vote share (absent equilibrium response), $v_C(\mathbf{s}, q_C, p_C)$ should intuitively be increasing in $q_C$. In Online Appendix 10.2, we show that, at the estimated parameter values and evaluated at observed states, $v_C(\mathbf{s}, q_C, p_C)$ is increasing in $q_C$ for about 85.8% of the challengers in our data.

**Equilibrium** We now define the equilibrium of the game. Formally, the players of the game are the incumbent and an infinite sequence of potential challengers. The strategies of the game are how much to spend, save, and raise for both the incumbent and the general election challenger, as well as the entry decisions of the potential challengers. We assume that $p_I(\cdot)$ and $\lambda(\cdot)$ are given. The solution concept we use is Markov Perfect Equilibria (Maskin and Tirole 1988).

It is intuitive to think of the equilibrium of the game as a fixed point in the potential challenger's entry threshold, $\bar{q}_C(\cdot)$ $(\{\mathbf{s}, p_C\} \mapsto \mathbb{R})$. To see this, consider fixing a threshold $\bar{q}_C(\cdot)$. As we explain below, fixing a threshold determines the probability that a challenger enters, $P_e(\mathbf{s})$, the equilibrium distribution of the challengers' valence and policy, $G_{q_C,p_C}(\cdot|\mathbf{s})$, and the probability that a challenger wins the Primary, $p(\mathbf{s}, q_C, p_C)$.

Note that $P_e(\mathbf{s})$ and $G_{q_C,p_C}(\cdot|\mathbf{s})$ are endogenous objects that enter into the incumbent's value function (Expression (5)). Once $P_e(\mathbf{s})$ and $G_{q_C,p_C}(\cdot|\mathbf{s})$ are determined, expressions (3) through (6) define a dynamic game of spending, saving and fund-raising indexed by $\bar{q}_C(\cdot)$ (taking as given $p_I(\cdot)$ and $\lambda(\cdot)$). Now, consider the value function of the challenger $v_C$ associated with the solution of this dynamic game. The value function $v_C$, along with $p(\mathbf{s}, q_C, p_C)$, defines a threshold for challenger entry given by $p(\mathbf{s}, \cdot, p_C)v_C(\mathbf{s}, \cdot, p_C) - \kappa = 0$. In equilibrium, the threshold for entry that solves this expression must coincide with the entry threshold that we fixed at the outset.

We now show that $P_e(\mathbf{s})$, $G_{q_C,p_C}(\cdot|\mathbf{s})$ and $p(\mathbf{s}, q_C, p_C)$ can be expressed as a function of $\bar{q}_C(\cdot)$. We start with the equilibrium entry probability, $P_e(\mathbf{s})$. Given that the entry probability is equal to one minus the probability of no entry, it can be expressed using $\bar{q}_C(\mathbf{s}, p_C)$ as follows:

$$P_e(\mathbf{s}) = \mathbb{E}_{N|\mathbf{s}} \left[ 1 - \mathbb{E}_{p_{C,1},\ldots,p_{C,N}} \left[ \prod_{n=1}^{N} F_{q_C|p_{C,n}}(\bar{q}_C(\mathbf{s}, p_{C,n})) \right] \right], \tag{8}$$

where $F_{q_C|p_{C,n}}(\bar{q}_C(\mathbf{s}, p_{C,n}))$ is the probability that the realization of $q_{C,n}$ is too low for candidate $n$ with policy $p_{C,n}$ to enter.

The equilibrium valence distribution of the challengers, $G_{q_C,p_C}(\cdot|\mathbf{s})$, can also be expressed as a function of $\bar{q}_C(\mathbf{s}, p_C)$. Consider the case for $N = 2$:

$$\begin{aligned}
&G_{q_C,p_C}(t_q, t_p|\mathbf{s}, N=2) = \Pr(\text{Valence of Challenger} \leq t_q \cap \text{Policy} \leq t_p|\mathbf{s}, N=2) \\
&= \frac{1}{P_e(\mathbf{s})} \times (\Pr(q_{C,1} \leq t_q \cap p_{C,1} \leq t_p, \text{Only 1 enters}|\mathbf{s}, N=2) \\
&\qquad + \Pr(q_{C,2} \leq t_q \cap p_{C,2} \leq t_p, \text{Only 2 enters}|\mathbf{s}, N=2) \\
&\qquad + \Pr(\text{Valence of Primary Winner} \leq t_q \cap \text{Policy} \leq t_p, \text{Both enter}|\mathbf{s}, N=2)).
\end{aligned} \tag{9}$$

The first term inside the summation can be expressed as follows.

$$\begin{aligned}
&\Pr(q_{C,1} \leq t_q \cap p_{C,1} \leq t_p, \text{Only 1 enters}|\mathbf{s}, N=2) \\
&= \mathbb{E}_{p_{C,1},p_{C,2}}[(F_{q_C|p_{C,1}}(t_q) - F_{q_C|p_{C,1}}(\bar{q}(\mathbf{s}, p_{C,1})))F_{q_C|p_{C,2}}(\bar{q}(\mathbf{s}, p_{C,2}))1\{p_{C,1} \leq t_p \cap \bar{q}(\mathbf{s}, p_{C,1}) \leq t_q\}].
\end{aligned} \tag{10}$$

The first term of Expression (10), $(F_{q_C|p_{C,1}}(t_q) - F_{q_C|p_{C,1}}(\bar{q}(\mathbf{s}, p_{C,1})))$, is the probability that candidate 1 enters and has valence less than $t_q$. The second term, $F_{q_C|p_{C,2}}(\bar{q}(\mathbf{s}, p_{C,2}))$, is the probability that candidate 2 does not enter. The last term, $1\{p_{C,1} \leq t_p \cap \bar{q}(\mathbf{s}, p_{C,1}) \leq t_q\}]$, corresponds to the event that the policy of candidate 1 is less than $t_p$ and that $t_q$ exceeds the threshold value for entry. Expression (10) is only a function of $\bar{q}(\cdot)$ and model primitives. We can show in a similar manner that other terms in Expression (9) are expressed only as a function of $\bar{q}(\cdot)$ and model primitives. Appendix 10.3 shows the results for general $N$.

Lastly, the probability of winning the primary, $p(\mathbf{s}, q_C, p_C)$, can be expressed as follows when $N = 2$:[15]

$$p(\mathbf{s}, q_{C,1}, p_{C,1}, N=2) = \mathbb{E}_{p_{C,2}}\left[\int_{\bar{q}(\mathbf{s}, p_{C,2})}^{\infty} \pi(q_{C,1}, p_{C,1}, q_{C,2}, p_{C,2}) dF_{q_{C,2}|p_{C,2}} + F_{q_C|p_{C,2}}(\bar{q}(\mathbf{s}, p_{C,2}))\right] \tag{11}$$

The first term inside the bracket is the event in which candidate 2 enters and candidate 1 wins. The second term corresponds to the event in which candidate 1 is the only entrant. Expression (11) is expressed only as a function of $\bar{q}(\cdot)$ and model primitives.

[15]The general expression is derived in Appendix 10.3.

**Open-seat Elections** Open-seat elections take place only when the current incumbent retires, which is assumed to be a terminal state for the incumbent. Hence, open-seat elections do not appear in any of the continuation games for incumbents and challengers of contested elections.

In an open-seat election, potential challengers from both parties make simultaneous entry decisions, and the candidate selection process for the out-party described above applies to both parties. Once candidates are selected as the party nominee, the candidates solve a problem that is similar to the one we defined earlier for the challengers that run against incumbents. We allow for the marginal effect of campaign spending on the vote share to be open-seat election specific, however. We denote the coefficient by $\beta_O$. Online Appendix 10.1 contains a full description of the model of open-seat elections.

**Deriving Model Properties Used in Identification** We now discuss two properties of the model that we exploit in identification. The first property is the injectivity of the policy function of uncontested incumbents:

**Proposition 1 (Injectivity)**: *Assume that the marginal cost of raising money, $\frac{\partial}{\partial x}\tilde{C}_I(x, q_I)$, is strictly decreasing with respect to $q_I$. Then, the policy functions of an uncontested incumbent, $\{d_I(\mathrm{s}), w'_I(\mathrm{s})\}$, are one-to-one from $q_I$ to $(d_I, w'_I)$, holding other state variables fixed.*

**Proof.** See Online Appendix 10.4. ■

Proposition 1 states that, if we have $\mathrm{s} = \{q_I, w_I, ten_I, p_I, p^*, X\}$ and $\mathrm{s}' = \{q'_I, w_I, ten_I, p_I, p^*, X\}$ such that $q_I \neq q'_I$, but all of the other elements of s and s′ are the same, the actions associated with s and s′ must be different, i.e., $\{d_I(\mathrm{s}), w'_I(\mathrm{s})\} \neq \{d_I(\mathrm{s}'), w'_I(\mathrm{s}')\}$. The injectivity of the policy function allows us to invert the policy functions of uncontested incumbents and express $q_I$ as a function of observed states and actions. This property is used to construct a control function for $q_I$ when we consider identification of the vote share equation. The property corresponds to the invertibility of the investment function in Olley and Pakes (1996).

The second model property is the presence of sufficient statistics for the valence distribution of the challengers, $G_{q_C}(\cdot|\mathrm{s}) \equiv G_{q_C,p_C}(\cdot, \infty|\mathrm{s})$, in elections with an incumbent. We say that $h = h(\mathrm{s})$ is a sufficient statistic for $f(\mathrm{s})$ if $h(\mathrm{s}') = h(\mathrm{s}'')$ implies $f(\mathrm{s}') = f(\mathrm{s}'')$. The following Proposition shows that the probability of challenger entry, $P_e(\mathrm{s})$, and the distribution of the policy positions among all entrants (including Primary losers), which we denote by $F_{p_C}(\tau|\mathbf{s}, \chi = 1)$, are sufficient statistics for $G_{q_C}(\cdot|\mathbf{s})$.

**Proposition 2 (Sufficient statistic)**: *Let* $m(\mathbf{s}) = \{P_e(\mathbf{s}), F_{p_C}(\cdot|\mathbf{s}, \chi = 1)\}$. *Then,* $m(\mathbf{s})$ *is a sufficient statistic for* $G_{q_C}(\cdot|\mathbf{s})$.

**Proof.** See Online Appendix 10.4 ■

Proposition 2 states that $F_{p_C}(\cdot|\mathbf{s}, \chi = 1) = F_{p_C}(\cdot|\mathbf{s}', \chi = 1)$ and $P_e(\mathbf{s}) = P_e(\mathbf{s}')$ imply $G_{q_C}(\cdot|\mathbf{s}) = G_{q_C}(\cdot|\mathbf{s}')$.

In our empirical application, we use the sufficient statistics property to control for $q_C$. Because we work with a discretized measure of candidate policy positions, our estimate of $F_{p_C}(\cdot|\mathbf{s}, \chi = 1)$ is a low dimensional vector corresponding to the probability mass on the support of the distribution. Our measure of candidate policy is based on that of Bonica (2023) which is available for incumbents and challengers, including those that lose Primary. The sufficient statistics property we derive here corresponds to the propensity score used in Olley and Pakes (1996) to control for firm exit.

# 3 Identification and Estimation

Our goal is to identify $q_I$ and $q_C$ for each candidate as well as the vote share equation. We first identify the vote share equation and the realization of $q_I$ for a subset of the incumbents who were uncontested in previous elections. In this step, we use the actions of the incumbents in past uncontested elections to construct a control function for $q_I$. In the second step, we identify $q_C$ by utilizing the first-order conditions associated with the candidates' problem. The valence terms of incumbents who were never uncontested, as well as the components of the candidates' payoffs (e.g., $C_I(\cdot)$ and $H_I(\cdot)$), are also recovered from the first-order conditions.

## 3.1 Identification of Incumbent's Valence and the Vote Share Equation

We first identify the vote share equation as well as the realization of $q_I$ for a subset of the incumbents who were uncontested in prior elections. Recall that the vote share equation is specified as follows:

$$\begin{aligned} vote_I =& \beta_I \ln d_I + \beta_C \ln d_C + \beta_P (p_I^2 - p_C^2) - 2\beta_P p^* (p_I - p_C) \\ &+ \beta_{ten} ten_I + \beta_X X + q_I - q_C + \varepsilon, \end{aligned} \tag{1'}$$

where we have expanded and rearranged the terms $\beta_P(p_I - p^*)^2$ and $\beta_P(p_C - p^*)^2$ in Expression (1).

The two main challenges in identifying the vote share equation and candidate valence are sample selection bias and endogeneity of spending. Sample selection problem arises, in part, from the challengers' endogenous entry decisions. Because the challengers know their own $p_C$, state s, etc., when making entry decisions, $q_C$ is potentially correlated with variables such as $p_C$ and s, including those that evolve exogenously. Similarly, incumbents' exit decisions create correlations between $q_I$ and s. Endogeneity of spending arises because the candidates choose $d_I$ and $d_C$ based, in part, on $q_I$ and $q_C$. For example, incumbents typically spend more against a challenger with higher valence. Note that because our main goal is to identify the candidate valence measures, we cannot rely on panel data methods that sweep out the valence terms as nuisance parameters.[16]

To overcome these challenges, we exploit a natural parallel between our setting and estimation of production functions. In particular, we adapt the control function approach developed by Olley and Pakes (1996) to our setting. We use Proposition 1 from the previous section to construct a control function that allows us to invert out $q_I$ as a function of observable terms and we use Proposition 2 to keep fixed the distribution of $q_C$ by conditioning on sufficient statistics. Note that we have two unobservable terms ($q_I$ and $q_C$) as opposed to just one in Olley and Pakes (1996).

**Control Function for $q_I$** We use the inverse of the policy function of uncontested incumbents as a control function to express $q_I$ as a function of observed variables. The policy functions associated with the problem of uncontested incumbents are how much to spend, $d_I(\mathbf{s})$, and how much to save, $w'_I(\mathbf{s})$. These policy functions can be viewed as

[16] Modeling voter decisions as discrete choice and applying the orthogonality arguments from Berry et. al. (1995) is also challenging. Voting and consumer choice are quite different in terms of the relationship between actions and outcomes. For example, in a consumer choice setting, buying good $A$ implies obtaining the utility associated with that good. However, in a voting context, a voter who votes for candidate $A$ may end up with candidate $B$ if other voters vote for $B$. These differences imply that, in a model in which voter's preferences are defined over outcomes, the standard intuition from Berry et. al. (1995) that we can recover the unobserved quality terms for each candidate using orthogonality conditions fails. This is shown in one of the author's earlier work (Kawai et. al. 2021).

mappings from $q_I$ to $(d_I, w'_I)$, holding the other state variables fixed. Because the mapping $q_I \mapsto (d_I, w'_I)$ is one-to-one (Proposition 1), we can uniquely solve for $q_I$ using these policy functions as $q_I = q_I(\bar{\mathbf{s}}_U)$, where $\bar{\mathbf{s}}_U$ denotes the vector of state variables and actions in the uncontested period.

Because the functional form of the policy functions, $d_I(\mathbf{s})$ and $w'_I(\mathbf{s})$ depends on the primitives of the model, so does $q_I(\cdot)$. Hence, $q_I(\cdot)$ is not a known object. Nevertheless, the fact that we can express $q_I$ as $q_I(\bar{\mathbf{s}}_U)$ allows us to substitute out $q_I$ in the vote share equation with a nonparametric function of observables, $\bar{\mathbf{s}}_U$. This allows us to identify $q_I(\cdot)$ by tracing out how the vote share varies with $\bar{\mathbf{s}}_U$, as all of the variables in $\bar{\mathbf{s}}_U$ are excluded from the original vote share equation.[17][18]

**Sample Selection Bias of $q_C$ and Sufficient Statistics** We next consider selection of challengers that choose to run against incumbents. We rewrite the vote share equation by decomposing $q_C$ into a part that depends on s and a part that is orthogonal to s.

$$\begin{aligned} vote_I =& \beta_I \ln d_I + \beta_C \ln d_C + \beta_P(p_I^2 - p_C^2) - 2\beta_P p^*(p_I - p_C) \\ &+ \beta_{ten} ten_I + \beta_X X + q_I(\bar{\mathbf{s}}_U) - \mathbb{E}[q_C|\mathbf{s}] - (q_C - \mathbb{E}[q_C|\mathbf{s}]) + \varepsilon. \end{aligned} \tag{1''}$$

In our setup, endogenous entry creates a sample selection problem in that the state variables such as $X$, $p_I$ and $ten_I$ affect the vote share through their effect on challenger valence $\mathbb{E}[q_C|\mathbf{s}]$ in addition to the direct effect. Following Olley and Pakes (1996), we use sufficient statistics to identify the direct effect of $X$, $p_I$ and $ten_I$ on the vote share while holding fixed the sample selection effect.[19]

As we show in Proposition 2, $m(\mathbf{s}) \equiv \{P_e(\mathbf{s}), F_{p_C}(p_C|\mathbf{s}, \chi = 1)\}$ is a sufficient statistic for the distribution of the challenger's valence in the general election, $G_{q_C}(\cdot|\mathbf{s})$. This

[17]Unlike in Olley and Pakes (1996) and Levinsohn and Petrin (2003), there are no collinearity issues when estimating the vote share equation (See Ackerberg Caves and Frazer 2006 and Gandhi, Navarro and Rivers 2020). This is because we express $q_I$ as a function of actions and state variables in some period $t$ which we then use to replace out $q_I$ in the vote share equation of some future period $t' > t$.

[18]In our empirical application, if an incumbent experiences multiple uncontested elections, we use the observations from the first uncontested election.

[19]We discuss selection problem with respect to $p_C$ in the next subsection. We need a different treatment of $p_C$ because $p_C \notin$ s. The sufficient statistics approach only works for variables in s.

implies that we can express $\mathbb{E}[q_C|\mathbf{s}]$ as a function of $m(\mathbf{s})$ as follows:

$$\begin{aligned}\mathbb{E}[q_C|\mathbf{s}] &= \mathbb{E}[q_C|m(\mathbf{s})] \\ &\equiv g(m(\mathbf{s})). \end{aligned} \tag{12}$$

Although $m(\mathbf{s})$ consists of endogenous objects, it is nonparametrically identified directly from the data. This is because the realization of candidate entry (i.e., whether or not $M \geq 1$) is observed and we use a data set that gives us direct measures of all entrant policy positions. Moreover, all elements of s are observed except for $q_I$, which we replace with functions of observables, $\bar{\mathrm{s}}_U$, as $q_I = q_I(\bar{\mathrm{s}}_U)$.

Replacing $\mathbb{E}[q_C|\mathrm{s}]$ as a function of $m(\mathbf{s})$ as in Expression (12) allows us to control for the indirect effect of s due to selection. By exploiting variation in s that leaves $m(\mathbf{s})$ fixed, $\mathbb{E}[q_C|\mathbf{s}]$ remains constant and we can identify the direct effect of s on the vote share.[20]

**Endogeneity of Spending with Respect to $q_C$** We next control for the endogeneity between $\{d_I, d_C\}$ and $(q_C - \mathbb{E}[q_C|\mathbf{s}])$. Because $(q_C - \mathbb{E}[q_C|\mathbf{s}])$ is the difference between the ex-post realization of the challenger's valence from its expectation, it is orthogonal to the set of predetermined variables, s. Hence we deal with the endogeneity by projecting the vote shares on s as follows:

$$\begin{aligned} vote_I =& \mathbb{E}[vote_I|\mathbf{s}] + \epsilon \\ =& \beta_I \mathbb{E}[\ln d_I|\mathbf{s}] + \beta_C \mathbb{E}[\ln d_C|\mathbf{s}] + \beta_P(p_I^2 - \mathbb{E}[p_C^2|\mathbf{s}]) - 2\beta_P p^*(p_I - \mathbb{E}[p_C|\mathbf{s}]) \\ & + \beta_{ten} ten_I + \beta_X X + q_I(\bar{\mathbf{s}}_U) - g(m(\mathbf{s})) + \epsilon, \end{aligned} \tag{1'''}$$

where $\epsilon \equiv (vote_I - \mathbb{E}[vote_I|\mathbf{s}])$. The term $\mathbb{E}[vote_I|\mathbf{s}]$ is the vote share equation evaluated *before* the challenger's valence $q_C$ realizes. Hence $\mathbb{E}[\epsilon|\mathbf{s}] = 0$ by construction. In particular, $\epsilon$ is uncorrelated with $\mathbb{E}[\ln d_I|\mathbf{s}]$ and $\mathbb{E}[\ln d_C|\mathbf{s}]$. Because $\mathbb{E}[\ln d_I|\mathbf{s}]$ and $\mathbb{E}[\ln d_C|\mathbf{s}]$ are identified directly from the data, the orthogonality condition guarantees identification of $\beta_I$ and $\beta_C$. Similarly, $\beta_P$ is identified because $(p_I^2 - \mathbb{E}[p_C^2|\mathbf{s}])$ is orthogonal to $\epsilon$.

The intuition behind the identification is as follows. Consider fixing $\bar{\mathrm{s}}_U$, the vector of state variables and actions in an uncontested period. This is equivalent to fixing $q_I$. Now consider $\mathrm{s}_1$ and $\mathrm{s}_2$ that keeps $P_e(\mathbf{s})$ and $F_{pc}(p_C|\mathbf{s}, \chi = 1)$ constant so that $P_e(\mathbf{s}_1) =$

[20]To be precise, $q_I(\cdot)$ and $g(\cdot, \cdot)$ are identified up to an additive constant. In our environment, shifting up or down the valence measures of all of the candidates by the same amount does not change the distribution of observable outcomes. We normalize the sample average of $q_I(\cdot)$ estimated from the control function to zero.

$P_e(\mathrm{s}_2)$ and $F_{p_C}(p_C|\mathrm{s}_1, \chi = 1) = F_{p_C}(p_C|\mathrm{s}_2, \chi = 1)$. The sufficient statistic property guarantees that the mean challenger valence will be the same in $\mathrm{s}_1$ and $\mathrm{s}_2$. Hence it is possible to use variation in expected candidate spending and variation in $X$ (across $\mathrm{s}_1$ and $\mathrm{s}_2$) to identify the coefficients of the vote share equation. Once all of the coefficients are identified, variation in $\bar{\mathrm{s}}_U$, $P_e(\mathrm{s})$ and $F_{p_C}(p_C|\mathrm{s}, \chi = 1)$ identifies $q_I(\cdot)$ and $g(\cdot)$.

Note that this approach requires that we observe the incumbents' actions in uncontested periods. Hence, in order to estimate the vote share equation, we only use a subset of elections in which the incumbent has experienced an uncontested election in the past.[21] We discuss identification of valence measures for incumbents who never experience uncontested elections in Section 3.2.

**Extensions** Our approach for estimating the vote share equation extends to settings with substantial outside spending such as recent House elections and to those with very few uncontested races, such as Senate elections.

Consider first an environment with outside spending as follows:

$$\begin{aligned} vote_I = \beta_I \ln d_I + \beta_C \ln d_C + \beta_{I,out} \ln d_{I,out} + \beta_{C,out} \ln d_{C,out} \\ + \beta_P (p_I - p^*)^2 - \beta_P (p_C - p^*)^2 + \beta_{ten} ten_I + \beta_X X + q_I - q_C + \varepsilon, \end{aligned}$$

where $d_{I,out}$ and $d_{C,out}$ denote outside spending supporting the incumbent and the challenger, respectively. Our approach directly extends to the identification of $\beta_{I,out}$ and $\beta_{C,out}$ because potential endogeneity between $\{d_{I,out}, d_{C,out}\}$ and $\{q_I, q_C\}$ can be controlled for by projecting all of the variables on the predetermined state variables s. Moreover, if there exist variables that impact outside groups' spending incentives that are orthogonal to $q_I$ and $q_C$ in a given district, including them as part of the state variable s provides extra source of variation to identify $\beta_{I,out}$ and $\beta_{C,out}$. An example of this type of variables is the number of other races that are predicted to be very close.[22] Note that sample selection bias can be dealt with in the same way as before even with outside spending.

Consider next an environment in which there are very few uncontested elections. In this case, we cannot invert the policy function of the incumbent in uncontested periods.

[21] We cannot use elections with an incumbent who experiences an uncontested election *in the future*, because it introduces selection on $\varepsilon$. Experiencing an uncontested election in the future means that the incumbent wins the current election, which implies a high $\varepsilon$ value.

[22] If outside groups have budget constraints, the presence of other close races will impact spending in a given election.

However, if the researcher has access to additional data on the predicted vote shares, it is possible to extend our approach to this case as well. In Online Appendix 10.8, we describe conditions under which our approach can be modified to estimate the vote share equation and to identify valence terms.

## 3.2 Identification and Estimation of Challengers' Valence and Components of Utility

We next consider identification and estimation of the challenger's valence, $q_C$, taking as given $q_I$ and the coefficients of the vote share equation. Because we do not have an adequate control function for $q_C$, we cannot directly identify $q_C$ from the vote share equation. To establish identification, we utilize the first-order conditions associated with the candidates' optimal spending and saving decisions. Because candidates observe $q_C$ when they make these decisions, the first-order conditions are informative about the value of $q_C$. We identify $q_C$ along with candidates' payoff terms, such as the cost of raising money, $C_I(\cdot;\theta)$, $C_C(\cdot;\theta)$, $\tilde{C}_I(\cdot;\theta)$ and the consumption value of spending, $H_I(\cdot;\theta)$, $H_C(\cdot;\theta)$, $\tilde{H}_I(\cdot;\theta)$ from the first-order conditions. We also identify the standard deviation of the shock in the vote share equation, $\sigma_\varepsilon$.[23]

We parameterize the cost of raising money and the consumption value of spending by $\theta$, where $\theta$ is a vector of unknown parameters. We also fix the discount factor to 0.9. Because Congressional elections take place every two years, $\delta = 0.9$ corresponds to an annual discount of roughly $0.95$. We also normalize the utility from winning, $B$, to one.[24]

**First-Order Conditions** The first-order conditions associated with the contested incumbent's spending and saving decisions are as follows:

$$\underbrace{\frac{\partial C_I}{\partial d_I}(w_I' + d_I - w_I, q_I;\theta)}_{\text{MC of fund-raising}} = \underbrace{\frac{\beta_I}{\sigma_\varepsilon d_I}\phi(K)\cdot\left(B + \delta\mathbb{E}_{\mathbf{s}'|\mathbf{s}}[V_I(\mathbf{s}')]\right) + \frac{\partial H_I}{\partial d_I}(d_I;\theta)}_{\text{MB of spending}} \tag{13}$$

[23]We do not estimate some of the model primitives regarding challenger entry, such as $\pi(\cdot)$, $F_N$, and $\kappa$. They are not necessary for recovering candidate valence measures, which is the focus of the paper.

[24]Identifying the discount factor in dynamic games is known to be difficult (Magnac and Thesmar 2002). We follow the literature in taking $\delta$ as given. Normalizing $B$ to one implies that costs and benefits are measured relative to the utility of winning the election.

$$\underbrace{\frac{\partial C_I}{\partial w_I'}(w_I' + d_I - w_I, q_I; \theta)}_{\text{MC of fund-raising}} = \underbrace{\delta \Phi\left(K\right) \frac{\partial}{\partial w_I'} \mathbb{E}_{\mathbf{s}'|\mathbf{s}}[V_I(\mathbf{s}')]}_{\text{MB of saving}}, \tag{14}$$

where

$$K = \frac{1}{\sigma_\varepsilon}(\beta_I \ln d_I + \beta_C \ln d_C + \beta_P(p_I - p^*)^2 - \beta_P(p_C - p^*)^2 + \beta_{ten} ten_I + \beta_X X + q_I - q_C). \tag{15}$$

$\Phi(\cdot)$ and $\phi(\cdot)$ are the c.d.f and the p.d.f of the standard normal distribution. Expression (13) equates the marginal cost of raising money to the marginal benefit of spending. The marginal benefit consists of an increase in the probability of winning the election, $(\frac{\beta_I}{\sigma_\varepsilon d_I})\phi(K)$, multiplied by the the value of winning, and the incremental consumption value of spending, $\frac{\partial}{\partial d_I} H_I$. Note that $K$ is proportional to the expected winning margin, and $\phi(\cdot)$ is highest when $K$ is close to zero. This implies that incumbents have the highest incentives to spend in close elections.

Expression (14) equates the marginal cost of raising money to the marginal benefit of saving, which is the incremental value of having more war chest next period. These expressions are obtained by using Expression (2) to substitute out $\Pr(vote_I > 0.5)$ from Expression (3), and taking derivatives.

Consider the first-order conditions of the incumbents for whom we can identify the value of $q_I$ using the control function, i.e., incumbents who are uncontested at least once. As we discuss below, we can simulate, as a function of $\theta$ and $\sigma_\varepsilon$, the continuation value $\mathbb{E}_{\mathbf{s}'|\mathbf{s}}[V_I(\mathbf{s}')]$ and compute its derivative $\frac{\partial}{\partial w_I'}\mathbb{E}_{\mathbf{s}'|\mathbf{s}}[V_I(\mathbf{s}')]$ for these incumbents. This allows us to solve for two values of $K$, one using equation (13) and the other using equation (14) for a given $\theta$ and $\sigma_\varepsilon$. Our identification of $\theta$ and $\sigma_\varepsilon$ relies on the restriction that the values of $K$ obtained from these two expressions for each incumbent coincide at the true parameter values.[25] Once $\theta$ and $\sigma_\varepsilon$ are identified, the value of $K$ for each election is identified, which implies identification of $q_C$ through Expression (15). Valence terms of the challengers who run against those incumbents (i.e., incumbents who experience uncontested elections) are hence identified.[26]

The payoff terms of uncontested incumbents, $\tilde{C}_I$ and $\tilde{H}_I$, are identified from the first-

[25] To the extent that $q_I$ can be recovered without any error, the first-order conditions must hold with equality for each observation whenever $(d_I, w_I') > 0$. Of course, in practice, $q_I$ is estimated nonparametrically, and the first-order conditions do not hold exactly at the estimated $q_I$ in finite sample.

[26] Once $q_C$ is identified, we can recover the implied residual $\varepsilon$ in the vote share equation (1). In the estimation, we also impose a restriction that the implied residuals have mean zero, and another restriction that the squared sum of the residuals equals $\sigma_\varepsilon^2$. See Online Appendix 10.6 for details.

order conditions of the uncontested incumbents whose $q_I$ are known. Similarly, the payoff terms of the challengers, $C_C$ and $H_C$, are identified by using the first-order conditions of the challengers whose $q_C$ are identified from the procedure outlined above. Parameters regarding open-seat elections are also identified from the first-order conditions of open-seat challengers.[27]

**Evaluating the Continuation Value by Simulation** We now discuss how to express the continuation value as a function of $\theta$ and $\sigma_\varepsilon$. Our approach is to adapt simulation methods developed by Hotz, Miller, Sanders and Smith (1994) and Bajari, Benkard and Levin (2007). These simulation methods involve estimating the transition of the state variables and the equilibrium policy functions nonparametrically in the first step, and using them to forward-simulate the value function for each parameter in the second step. Importantly, they do not require solving for an equilibrium at each candidate parameter value.

One challenge in applying these methods to our setting is that we do not observe $q_C$, one of the state variables in contested elections. Existing methods require that all of the state variables be observed when estimating the policy function. Nevertheless, it is still possible to forward-simulate the incumbent's continuation payoff in our setting because in our model, (i) the incumbent's utility does not depend directly on $q_C$ and only indirectly through actions and outcomes; and (ii) $q_C$ is independent across periods conditional on the observable state, s. These features of the model imply that we do not need to know the distribution of $q_C$ in future periods to forward-simulate the value function – we need only the distribution of actions (spending, saving and fund-raising) and outcomes (challenger's entry status and electoral outcome).

In our forward simulation procedure, we estimate the distribution of actions and outcomes conditional on just the state s, where $\mathrm{s} \equiv \{q_I(\bar{\mathrm{s}}_U), w_I, ten_I, p_I, p^*, X\}$. Note that this is not the same as estimating the policy functions in contested elections, which would be functions of both s and the challenger's valence and policy position, $\{q_C, p_C\}$. We can then draw a sequence of actions and outcomes to compute an associated sequence of flow payoffs which can then be averaged across simulation draws to evaluate $V_I(\mathrm{s})$. Once we simulate $V_I(\mathrm{s})$, we can obtain $\mathbb{E}_{\mathrm{s}'|\mathrm{s}}[V_I(\mathrm{s})]$ as well as $\frac{\partial}{\partial w_I'}\mathbb{E}_{\mathrm{s}'|\mathrm{s}}[V_I(\mathrm{s}')]$. Online Appendix 10.5 contains details on the simulation procedure.

[27]Because we do not estimate the vote share equation for open-seat elections directly, the coefficient of spending, $\beta_O$, is also identified from the first-order conditions.

**Recovering Valence for All Candidates** We now consider recovering the valence terms for incumbents who were never uncontested and those of challengers that run against them. Because the vote share equation and the payoff functions of the candidates are known, the four first-order conditions of the candidates in each contested election (two for each candidate) can be used to identify the valence terms of these candidates: the first-order conditions can be considered as a system of equations in $q_I$ and $q_C$. Similarly, for open-seat elections in which the valence terms of both candidates are yet to be identified, the first-order conditions can be considered as equations in $\mathbf{q_O}$, where $\mathbf{q_O}$ is a $2 \times 1$ vector that represents the valence of open-seat candidates. We recover the valence measures of all candidates by solving these first-order conditions.[28] Note that the first-order conditions we use in this stage were not used in any of the previous stages.

## 3.3 Estimation

Our estimation closely follows the step-by-step identification procedure described above. In step 1, we estimate the vote share equation using the control function approach. Specifically, we estimate the probability that a challenger enters an election, $P_e(\mathbf{s})$, and the distribution of entrant policy, $F_{p_C}(\cdot|\mathbf{s}, \chi = 1)$. We then estimate the vote share equation by sieve minimum distance estimator (Ai and Chen 2003). In step 2, we estimate candidates' payoffs and the challengers' valence terms by GMM in which we treat the candidates' first-order conditions as moments. To forward-simulate the continuation payoffs, we estimate the distribution of actions in contested elections by nonparametric maximum likelihood (Gallant and Nychka 1987). We also estimate the policy functions in uncontested elections, incumbents' retirement probability and choice of policy positions, as well as the evolution of the state variables. We then forward-simulate the continuation payoffs according to the procedure described in Online Appendix 10.5. In step 3, we estimate the parameters of the open-seat elections using GMM analogous to the case of contested elections. In step 4, we recover the valence terms of candidates whose valence measures are not recovered from the control function. We use GMM by stacking the candidates' first-order conditions as moments. The details of the estimation are described in Online Appendix 10.6.

[28] In practice, we minimize the sum of squared deviations. See Online Appendix 10.6 for details.

# 4 Data

We obtain campaign finance data from the Federal Election Commission (FEC 2011). The data contain information on the amount of fund-raising, spending and savings of all U.S. House candidates from 1984 to 2008. Data on electoral outcomes and candidate characteristics are from the database of the CQ Press. We obtain demographic characteristics of congressional districts from the Census (Census 2015) and the Bureau of Labor Statistics (BLS 2011). We also collect data on vote shares in Presidential elections to create a partisanship measure for each district. We use this measure to estimate $p^*$, the ideal policy position of the district. Vote shares in Presidential elections are from Adler (2003) and POLIDATA (POLIDATA 2015). Finally, we obtain data on incumbent and challenger policy positions from Bonica (2023).

From the set of regular-cycle House elections, we drop elections in Louisiana, elections in Texas in 1996 that are affected by Supreme Court rulings and elections involving major scandals. We also drop elections in which the spending and savings of one of the candidates are zero or very close to zero, and those in which one of the candidate's policy position is missing.[29] Lastly, we drop elections in which the incumbent saves more than \$1.2 million (in 1984 dollars).[30] Appendix 10.7 describes in more detail how the data are constructed.

Table 1 reports the summary statistics of the key variables. Dollar values are normalized to 1984 dollars and reported in units of \$1,000. Column (1) corresponds to sample statistics for the incumbents in contested elections and Column (2) corresponds to those for the challengers. The probability that an election is contested is about 79.6%. In contested elections, incumbents start out with an average war chest of about \$95,000 and raise about \$535,700. The incumbent spends about \$514,900 and saves about \$116,100, on average. The challengers, on the other hand, typically start out with zero war chest and raise about \$204,600, almost all of which is spent. Average incumbent vote share is 63%.

Column (3) of Table 1 corresponds to the sample statistics for the incumbents in uncontested elections. Uncontested incumbents start out with an average war chest of \$127,500, which is higher than the average war chest of contested incumbents. The average amount

[29]When savings and spending are zero, the first-order conditions of the candidates do not necessarily hold with equality. We drop elections in which one of the candidates spends and saves less than \$5,000, or raises less than \$5,000. These elections account for 16.0 percent of all observations. We further drop 3.8 percent of contested elections because the policy position of one of the candidates is missing.

[30]These elections account for 0.4 percent of all observations. Unusually large amount of savings are invariably for running for higher offices.

| | (1) | | (2) | | (3) | | (4) | |
|---|---|---|---|---|---|---|---|---|
| | Contested | | | | Uncontested | | Open-Seat | |
| | Incumbent | | Challenger | | | | | |
| Spending ($d$) | 514.9 | (366.3) | 202.1 | (313.3) | 275.7 | (202.7) | 509.3 | (432.0) |
| Amount Raised ($fr$) | 535.7 | (354.7) | 204.6 | (315.9) | 342.3 | (225.6) | 519.7 | (435.4) |
| War Chest ($w$) | 95.0 | (137.1) | 0.6 | (4.4) | 127.5 | (153.5) | 0.8 | (5.8) |
| Savings ($w'$) | 116.1 | (153.4) | 3.4 | (9.7) | 194.6 | (191.7) | 12.1 | (29.3) |
| Tenure ($ten$) | 4.9 | (3.9) | 0 | (0) | 5.7 | (4.2) | 0 | (0) |
| Vote Share | 0.633 | (0.084) | 0.367 | (0.084) | 1 | (0) | 0.5 | (0.419) |
| Sample Size | 3,065 | | 3,065 | | 787 | | 445 | |

Note: Spending, Amount Raised, War Chest and Savings are reported in units of $1,000. Dollar values are deflated to their values in 1984. Standard Errors are reported in parenthesis.

Table 1: Descriptive Statistics of Incumbents, Challengers and Open-Seat Candidates

of money raised in uncontested periods is about $342,300 and the average amount spent is about $275,700. Incumbents save more in uncontested races ($194,600) than in contested races ($116,100). Column (4) reports the sample statistics for open-seat elections.

Figure 2 plots the distribution of policy positions for incumbents (top panel) and for all entrants, including Primary losers (bottom panel). The bottom panel corresponds to $F_{p_C}(\cdot|\mathrm{s}, \chi = 1)$ which we use as sufficient statistics. The gray bars correspond to the Republicans and the white bars correspond to the Democrats. These policy positions, taken from Bonica (2023), are based on the source of campaign contributions received by each candidate. We discretize the measure into 10 symmetric bins around zero, based on the quantiles of the data.[31]

We find that the policy positions of the incumbents are more concentrated around the center while those of the entrants are more extreme. For Republicans, the average policy position of incumbents and entrants are 0.695 and 0.981, respectively. For Democrats, they are -0.445 and -0.823. The figure also shows that most Republicans are distributed to the right of zero and Democrats are distributed to the left. In the estimation, for the small number of Democrats whose policy is to the right of zero, we assume that their policy is at the negative bin closest to 0. Similarly, we assume that all Republicans are distributed to

[31]The points are $\{-1.242, -0.920, -0.738, -0.552, -0.179, 0.179, 0.552, 0.738, 0.920, 1.242\}$. We discuss the construction of our policy measures in more detail in Appendix 10.7.

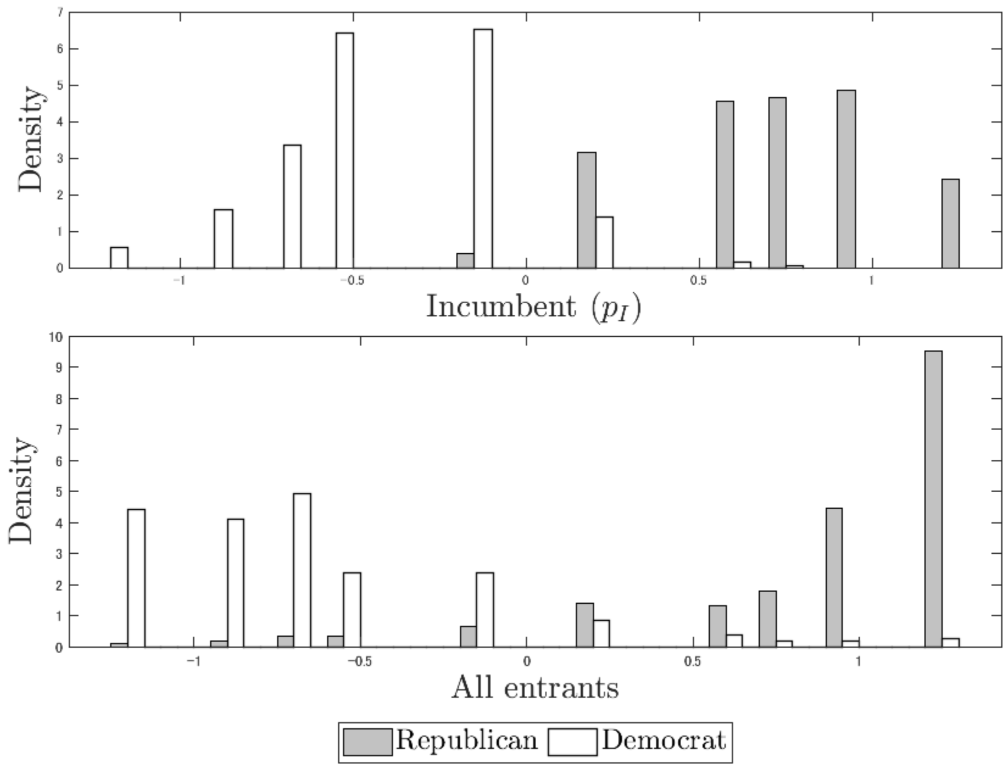



Figure 2: Distribution of Policy Positions

the right of zero. We do this to reduce computational complexity.

Table 2 reports the summary statistics of the characteristics of the Congressional Districts that we include as control variables ($X$) in the vote share equation. Column (1) corresponds to districts with an incumbent Democrat, Column (2) corresponds to those with an incumbent Republican, and Column (3) corresponds to open-seat elections. The Republican partisanship index is a measure of a district's Republican partisanship, constructed by following Levendusky et. al. (2008). In particular, we regress the log difference in the district-level vote shares of the Presidential election on demographic characteristics, year fixed effects and state fixed effects. Our measure of district partisanship is the fitted value of the regression for the concurrent or the most recent Presidential election. Positive (negative) values of the partisanship index correspond to an expected Presidential vote share above (below) 50% for the Republicans. Online Appendix 10.7 contains a detailed discussion of how this variable is constructed. Party of President is a dummy variable that equals 1 (-1) if the incumbent President is a Democrat (Republican).

| | (1) Democrat | | (2) Republican | | (3) Open-Seat | |
|---|---|---|---|---|---|---|
| % Unemployed | 0.061 | (0.022) | 0.054 | (0.019) | 0.057 | (0.020) |
| Population Density | 2917.6 | (7239.1) | 879.1 | (3849.6) | 1066.5 | (2410.8) |
| Republican Partisanship Index | -0.139 | (0.512) | 0.180 | (0.317) | 0.051 | (0.412) |
| Party of President | -0.406 | (0.914) | -0.316 | (0.949) | -0.303 | (0.954) |
| Sample Size | 2,047 | | 1,805 | | 445 | |

Note: These variables are used as controls in the vote share equation. Standard errors are reported in parentheses.

Table 2: Characteristics of Congressional Districts

# 5 Specification and Estimation Results

## 5.1 Specification

**Vote Share Equation** We specify the vote share equation as follows:

$$
\begin{aligned}
vote_I &= \beta_I \ln d_I + \beta_C \ln d_C + \beta_P (p_I - p^*)^2 - \beta_P (p_C - p^*)^2 + \beta_{ten} ten_I \\
&\quad + \underbrace{D_I \times (\beta_d + \beta_{dn} dn) + \beta_{ue}(ue \times D_I \times D_P) + \text{Election cycle FE}}_{\beta_X X} + q_I - q_C + \varepsilon,
\end{aligned}
$$

where $p^* = \beta_{\text{ID},0} + \beta_{\text{ID},1} pt$.

We specify the district's ideal point, $p^*$, as a linear function of the district's Republican partisanship index, $pt$, as $p^* = \beta_{\text{ID},0} + \beta_{\text{ID},1} pt$. The variable $pt$ is constructed from the district's Presidential vote shares as we discussed in Section 4.

The first term in the second line of the vote share equation is an interaction of the incumbent's party $D_I (\in \{-1, 1\})$ and a linear function of the population density of the district, $dn$. We include this term to capture the differential electoral strength of the parties in rural and urban districts.

In order to account for intertemporal variation in the popularity of the parties, we also include the term $ue \times D_I \times D_P$. This term captures the effect of retrospective voting. The variable $ue$ is the unemployment rate of the district, which proxies for the economic environment in the district.[32] Although retrospective voting can take many forms in principle, several studies find that voters express their satisfaction with the current administration by voting for or against the candidate from the President's party (Hibbing and Alford 1981, Stein 1990). For this reason, we interact $ue$ with a variable that indicates whether or not the incumbent and the President are from the same party. Specifically, we let $D_P$ be equal to $1$ $(-1)$ if the President is a Democrat (Republican). $D_I \times D_P$ is then equal to one if the candidate and the President are from the same party, and $-1$, otherwise.

Finally, we include several election cycle dummies. These dummies include one for a midterm election ($1\{\text{Midterm}\}$), one for a first-term President ($1\{\text{First Term}\}$), and an interaction term ($1\{\text{Midterm}\} \times 1\{\text{First Term}\}$). We also include a dummy for a midterm interacted with one for being in the same party as the President ($1\{\text{Midterm}\} \times D_I \times D_P$) to capture the well-known fact that Congressional candidates from the President's party fare poorly in midterm elections (See, e.g., Campbell 1985, Erikson 1988). We adopt a relatively parsimonious specification for the vote share equation in order to keep the number of state variables to a manageable level.

**State Variables and Their Transition** Formally, the vector of state variables s is

$$\mathrm{s} = \{q_I, w_I, ten_I, p_I, pt, dn \times D_I, ue \times D_I \times D_P, 1\{\text{First Term}\}, 1\{\text{Midterm}\}\}.$$

Because we assume that $p^* = \beta_{\text{ID},0} + \beta_{\text{ID},1} pt$, we keep track of $pt$ instead of $p^*$.

We assume that $ue$, $dn$ and $pt$ follow an $AR(1)$ process. We also assume the following process for $D_P$: (1) $D_P$ remains the same next period with probability 0.75 in a presidential election when the President is running for the second term; (2) $D_P$ remains the same with probability 0.5 when the incumbent President is at the end of his second term; and (3) $D_P$ remains the same next period with probability one if the election is a Midterm election.

**Incumbent Retirement and Policy Position** We specify the incumbent's retirement probability, $\lambda(\cdot)$, as a logit. The explanatory variables include $q_I, w_I', ten_I', p_I, pt'$ and $X'$. We specify the evolution of the incumbent policy position, $p_I(\cdot)$, as a multinomial logit over the discrete points in the distribution's support. See Appendix 10.5 for details.

[32] Unemployment rate is often used as a proxy for the economic environment, e.g., Ansolabehere et. al. (2014).

**Components of the Utility Function** We specify the cost of fund-raising and the benefit associated with spending for uncontested incumbents as follows:

$$\tilde{C}_I(fr_I; q_I) = c(q_I)(\ln fr_I)^2,$$
$$\tilde{H}_I(d_I) = \gamma_U \sqrt{\ln d_I},$$

where $fr_I$ denotes the amount raised and $c(\cdot)$ is a decreasing function of $q_I$. This specification implies that the cost of fund-raising is increasing and convex in $\ln fr_I$, and the benefit associated with spending is increasing and concave in $\ln d_I$. The assumption that $c(\cdot)$ is decreasing in $q_I$ guarantees that $q_I$ is invertible with respect to the actions of the incumbent in uncontested periods (see Proposition 1). In particular, in Online Appendix 10.4 we show that the functional form we specify for $\tilde{C}_I$ and $\tilde{H}_I$ allows us to express $q_I$ as a function of a scalar variable $z_U \equiv \frac{fr_I}{\sqrt{\ln d_I} d_I \ln fr_I}$ as $q_I = q_I(z_U)$, where $d_I$ and $fr_I$ are incumbent's spending and fund-raising amount in an uncontested period. Being able to express $q_I$ as a function of a scalar variable, $z_U$, rather than a function of a vector of all states and actions in the uncontested period, $\bar{\mathrm{s}}_U$, significantly reduces the data requirement for estimation. For estimation, we further specify $c(\cdot)$ as $c(q_I) = c_1 + c_2 \exp(-q_I)$, where $c_1 > 0$ and $c_2 > 0$ are parameters to be estimated. The functional form for $c(q_I)$ ensures that $c(q_I)$ is positive and decreasing for all $q_I$.

We specify the cost function of the contested incumbents, $C_I(\cdot; q_I)$, and the cost function of the challengers, $C_C(\cdot; q_C)$, in a similar way as $\tilde{C}_I$:

$$C_I(fr_I; q_I) = \eta_I \times c(q_I)(\ln fr_I)^\alpha$$
$$C_C(fr_C; q_C) = \eta_C \times c(q_C)(\ln fr_C)^\alpha,$$

where $\alpha$ is the curvature parameter.

We specify the benefit of spending for contested incumbents and challengers as follows:[33]

$$H_I(d) = H_C(d) = \gamma\sqrt{\ln d}.$$

[33]We assume $H_C(\cdot) = H_I(\cdot)$ because it is difficult to separately estimate $H_C(\cdot)$ from $C_C(\cdot)$ in our sample. The difficulty arises because (i) the probability of winning the election is very small for many challengers and (ii) many challengers save nothing, i.e., $w'_C \approx 0$. When the winning probability is very small, Expression (13) reduces to $C'_C(d_C) = H'_C(d_C)$. When $w'_C \approx 0$, Expression (14) becomes an inequality. Hence, the two first-order conditions reduce to a single restriction, $C'_C(d_C) = H'_C(d_C)$. This means that we need to normalize $C_C(\cdot)$ or $H_C(\cdot)$.

We assume that the costs of fund-raising and the benefit from spending for open-seat candidates are the same as those of challengers running against incumbents.

## 5.2 Parameter Estimates

| | $P_e(\mathbf{s})$ | Mean of $\|p_C\|$ | Std. Dev. of $\|p_C\|$ |
|---|---|---|---|
| Constant (Republican Challenger) | 2.452 (1.051) | 0.888 (0.137) | 0.598 (0.118) |
| Constant (Democrat Challenger) | 1.984 (1.051) | 0.630 (0.149) | 0.472 (0.123) |
| $\ln$ War Chest | -0.250 (0.197) | 0.002 (0.015) | -0.023 (0.013) |
| $(\ln$War Chest$)^2$ | 0.008 (0.010) | 0.0002 (0.001) | 0.002 (0.001) |
| Tenure | 0.027 (0.012) | 0.017 (0.004) | 0.006 (0.003) |
| $\|p_I\|$ | - | -0.068 (0.183) | -0.012 (0.151) |
| $p_I^2$ | 0.547 (0.184) | -0.091 (0.152) | 0.084 (0.121) |
| Republican Partisanship Index $\times D_I$ | 0.623 (0.174) | 0.103 (0.046) | -0.008 (0.036) |
| Population Density $\times D_I/10000$ | -0.149 (0.109) | -0.167 (0.056) | 0.001 (0.059) |
| Unemployment $\times D_I \times D_P$ | 3.571 (1.007) | 0.387 (0.260) | 0.341 (0.192) |
| Midterm $\times D_I \times D_P$ | 0.294 (0.087) | 0.006 (0.018) | -0.019 (0.018) |
| B-Spline of $z_U$ | ✓ | ✓ | ✓ |
| Election Cycle | ✓ | ✓ | ✓ |

Note: We take 7 knots corresponding to $(1/8,\cdots,7/8)$ quantiles of $z_U$ for the B-Spline of $z_U$. Election Cycle corresponds to a complete interaction of dummy variables $1\{$First Term$\}$ and $1\{$Midterm$\}$. $D_I$ is equal to 1 if the incumbent is a Democrat and $-1$ if the incumbent is a Republican. Standard errors are reported in parentheses.

Table 3: Estimates of $P_e(\mathbf{s})$ and $F_{p_C}(p_C|\mathbf{s}, \chi = 1)$

**Estimates of $m(\mathrm{s}) = \{P_e(\mathrm{s}), F_{p_C}(p_C|\mathrm{s}, \chi = 1)\}$** We first report our estimates of challenger entry probability, $P_e(\mathbf{s})$, and the distribution of policy positions among entrants, $F_{p_C}(p_C|\mathbf{s}, \chi = 1)$. Estimates of $m(\mathbf{s}) = \{P_e(\mathbf{s}), F_{p_C}(p_C|\mathbf{s}, \chi = 1)\}$ are used to control for the sample selection problem in Expression ($1'''$). Both $P_e(\mathbf{s})$ and $F_{p_C}(p_C|\mathbf{s}, \chi = 1)$ are functions of $\mathrm{s} = \{q_I, w_I, ten_I, p_I, p^*, X\}$. Because we specify $p^*$ to be a function of $pt$, and because $q_I$ is a function of the incumbent's actions in uncontested periods, as $q_I = q_I(z_U)$ (where $z_U \equiv \frac{fr_I}{\sqrt{\ln d_I} d_I \ln fr_I}$), we estimate $P_e(\mathbf{s})$ and $F_{p_C}(p_C|\mathbf{s}, \chi = 1)$ as functions of $(z_U, w_I, ten_I, p_I, pt, X)$. Note that all of these variables are observed.

The first column of Table 3 reports the estimation results for $P_e$. We use a Probit specification to estimate $P_e$. The first row corresponds to the estimate of the constant term for Republican challengers and the second row corresponds to that for Democratic challengers. We find that the entry probability is slightly higher for Republican challengers. In rows 3 and 4, we find that a higher incumbent war chest is associated with a lower entry probability, although the coefficients are not statistically significant. We find that the district partisanship index and the unemployment rate have statistically significant and sizeable effects on the entry probability. When the incumbent is a Democrat ($D_I = 1$), the entry probability is higher in a district with a higher Republican partisanship index, i.e., more conservative districts. When the incumbent is a Republican ($D_I = -1$), the entry probability decreases in a district with a higher partisanship index. We also find that higher unemployment rate increases entry when the incumbent is of the same party as the President.

For the estimation of $F_{p_C}(p_C|\mathbf{s}, \chi = 1)$, we make use of the way in which we discretize the policy position of the candidates, i.e., $p_C$ takes values on a discrete set that is symmetric around zero, and the policy position of a Democratic (Republican) candidate is always negative (positive). Instead of estimating the distribution of $p_C$ directly, we estimate the distribution of $|p_C|$, the distribution of policy extremity. In particular, we assume that $|p_C|$ follows a discretized normal distribution and specify the mean and the standard deviation as linear functions of the variables in the vote share equation, as well as $|p_I|$, $w_I$, and B-spline bases of $z_U$.

The second and third columns of Table 3 report the estimates of the mean and the standard deviation of the distribution of $|p_C|$. Focusing on the second column, we find that a Republican entrant on average holds a more extreme position (0.888) than a Democratic entrant (0.630). For the coefficients of variables that are not interacted with $D_I$, positive values indicate that an increase in the variable results in a more extreme entrant position. We find that larger incumbent tenure and war chest result in more extreme entrants, though the estimate for the latter is not statistically significant.

For variables that are interacted with $D_I$, a positive coefficient implies that an increase in the variable moves the entrants' positions toward the right. For instance, we find that an increase in the partisanship index (i.e., a more conservative district) moves both Democratic and Republican entrants' positions to the right.

The fact that variables such as the partisanship index affects $P_e(\mathbf{s})$ and $F_{p_C}(p_C|\mathbf{s}, \chi = 1)$ suggests that these variables indirectly affect the valence of the general election challenger,

$q_C$, through the challengers' endogenous entry decisions. Online Appendix 10.6 contains details on the estimation of $P_e(\mathbf{s})$ and $F_{p_C}(p_C|\mathbf{s}, \chi = 1)$ as well as those of other results in this section.

| | Control function | | OLS | |
|---|---|---|---|---|
| Incumbent Spending | 0.039 | (0.020) | -0.019 | (0.004) |
| Challenger Spending | -0.039 | (0.011) | -0.027 | (0.002) |
| Openseat Spending | 0.015 | (0.006) | - | - |
| Policy ($\beta_P$) | -0.031 | (0.021) | 0.010 | (0.004) |
| Policy ($\beta_{ID,0}$) | -0.670 | (4.556) | 0.673 | (0.378) |
| Policy ($\beta_{ID,1}$) | 0.660 | (8.372) | -1.878 | (0.764) |
| Tenure | -0.001 | (0.002) | -0.001 | (0.001) |
| $D_I$ | -0.055 | (0.027) | -0.020 | (0.010) |
| Population Density $\times D_I/10000$ | 0.020 | (0.020) | 0.031 | (0.006) |
| Unemployment $\times D_I \times D_P$ | -0.082 | (0.103) | 0.002 | (0.047) |
| Midterm $\times D_I \times D_P$ | -0.030 | (0.009) | -0.026 | (0.005) |
| $\sigma_\varepsilon$ | 0.069 | (0.003) | 0.059 | (0.012) |

Note: First column corresponds to the estimates obtained using the approach discussed in Section 3. Second column corresponds to OLS estimates. Standard errors are reported in parentheses. The standard errors of the first column are computed based on 500 bootstrap samples. While the standard errors of $\beta_{ID,0}$ and $\beta_{ID,1}$ are large, the standard errors of $\beta_P \times \beta_{ID,1}$ and $\beta_P \times \beta_{ID,1}$ are relatively small (0.016 and 0.015, respectively).

Table 4: Parameter Estimates of the Vote Share Equation

**Estimates of Vote Share Equation** In the first column of Table 4, we report our estimates of the vote share equation. The parameters are estimated by applying a sieve minimum distance estimator of Ai and Chen (2003) to Expression ($1'''$). Our point estimates of $\beta_I$ and $\beta_C$ are $0.039$ and -$0.039$, respectively. These estimates imply that a standard deviation increase in the spending of the incumbent increases the incumbent vote share by about 2.7 percentage points, while a standard deviation increase in the challengers' spending decreases the incumbent vote share by about 6.9 percentage points.[34] We find that the impact of spending on the vote share is smaller in open-seat elections. The estimate of

[34] Although the magnitude of the coefficients is the same, the marginal effect is different because we specify the vote share equation as a linear function of log spending (and, hence, the marginal return from spending diminishes), and challengers spend less than incumbents, on average.

$\beta_O$ is $0.017$, which implies that a standard deviation increase in the spending of an open-seat candidate increases the vote share of that candidate by about 1.5 percentage points.[35]

The effect of candidate policy positions on the vote share is captured by the coefficient, $\beta_P$, which is estimated to be $-0.031$. Given that candidate policy positions are between $-1.25$ and $1.25$, the coefficient estimates imply that the most extreme candidate running in a centrist district suffers about a 4.5 percentage point penalty in terms of vote share relative to a centrist candidate whose policy position matches with the district's ideal position. The district ideal point is estimated as a linear function of the partisanship index. Given that the partisanship index falls mostly between -1 and 1, our estimate of $\beta_{ID,1}$ (0.660) implies that the distance between the most extreme liberal district and the conservative district is about 1.32 in terms of Bonica (2023)'s ideology measure.[36]

Our estimates of $\beta_{ten}$ is small and statistically insignificant.[37] We also find that the President's party performs significantly worse in midterm elections, which has been documented in previous work (e.g., Abramowitz et. al. 1986). Our estimate of $\sigma_\varepsilon$, the standard deviation of the error term in the vote share equation, is 0.067.

In the second column of Table 4, we report the OLS estimates of the vote share equation for elections with incumbents for comparison. The OLS results correspond to a simple regression of Expression (1) in which the vote shares are regressed on observables without controlling for candidate valence. We find that the OLS estimate of incumbent spending is negative and statistically significant, reflecting the fact that incumbents choose higher $d_I$ against stronger challengers (i.e., $q_C$ and $d_I$ are positively correlated).

**Estimates of $C(\cdot)$ and $H(\cdot)$** We now report the estimates of $C(\cdot)$ and $H(\cdot)$, the candidates' cost of fund-raising and the personal benefit of spending.[38] Table 5 reports the parameter estimates and Figure 3 illustrates the shape of $C(\cdot)$ and $H(\cdot)$ at the estimated

[35]Note that we estimate $\beta_O$ from the first-order conditions associated with the problem of open-seat candidates.

[36]The standard errors of the estimates for $\beta_{ID,0}$ and $\beta_{ID,1}$ are large for the following reason. When we replace $p^*$ with $\beta_{ID,0} + \beta_{ID,1}pt$ in expression ($1'''$) and estimate the equation using Ai and Chen (2003), we obtain estimates of $\beta_P$, $\beta_P \times \beta_{ID,0}$ and $\beta_P \times \beta_{ID,1}$. While the standard errors of each of these coefficients are relatively small (0.016 and 0.015, respectively), the standard errors of $\beta_{ID,0}$ and $\beta_{ID,1}$ become large when we divide $\beta_P \times \beta_{ID,0}$ and $\beta_P \times \beta_{ID,1}$ by the estimate of $\beta_P$.

[37]This is consistent with existing work that finds that seniority leads to a small electoral *disadvantage*, e.g., Abramowitz (1991). This result is also robust to including a more flexible functional form for tenure.

[38]For the estimation of $C(\cdot)$ and $H(\cdot)$, we also need to estimate reduced-form policy functions (e.g., $\lambda(\cdot)$) and state transitions for forward simulation. These estimates are reported in Appendix 10.5.

parameters.[39] The vertical axes of the figure correspond to the cost of fund-raising and the benefit of spending measured relative to the utility from winning, which is normalized to 1. The horizontal axes correspond to the log amount raised and log amount spent. We find that, on average, the cost of fund-raising for contested incumbents is 0.238, or 24% of the utility from winning, and that for challengers is 0.162. Although we estimate $C_C(\cdot)$ to be higher than $C_I(\cdot)$, challengers incur lower costs on average because they raise significantly less money than incumbents. Fund-raising costs of uncontested incumbents are 0.155, on average. The average benefit from spending for contested incumbents, uncontested incumbents and challengers are 0.076, 0.074 and 0.070, respectively.

| $C(\cdot)$ | | | $H(\cdot)$ | | |
|---|---|---|---|---|---|
| $c_1$ | 0.001 | (0.000) | $\gamma$ | 0.021 | (0.005) |
| $c_2$ | 0.158 | (0.038) | | | |
| $\eta_I$ | 5.972 | (0.732) | | | |
| $\eta_C$ | 12.781 | (3.945) | | | |
| $\alpha$ | 2.670 | (0.158) | | | |

Note: Standard errors (reported in parentheses) are computed based on 500 bootstrap samples.

Table 5: Parameter Estimates of Fund-Raising Cost and Benefit from Spending

## 5.3 Estimates of Candidate Valence

Figure 4 plots our estimates of candidate valence measures. The top panel corresponds to the histogram of the estimated valence measures of the incumbents and the middle panel corresponds to those of challengers that run against incumbents.[40][41] We find that, on average, the valence measures of the incumbents are about 0.035 higher than those of the challengers, implying that the differences in candidate valence translate to a 3.5 percentage

[39] As we explain in Section 4, we keep candidates who raise more than $5,000 in contested elections for estimation. Because of this, we normalize fund-raising in contested periods by $5,000 when estimating $C_I(\cdot)$ and $C_C(\cdot)$. Letting the cost function start from $\log(5000)$ helps us better fit the dispersion of incumbent fund-raising.

[40] If a candidate competes in an election as a challenger and subsequently becomes an incumbent, the valence measure of the candidate is included in both panels. Similarly, candidates that compete in open-seat elections who later become incumbents appear twice.

[41] In Appendix 10.9 we report the corresponding distributions when we apply shrinkage to account for sampling error. We obtain the standard errors with 500 bootstrap sample draws. The bootstrap procedure we use is described in Appendix 10.10.

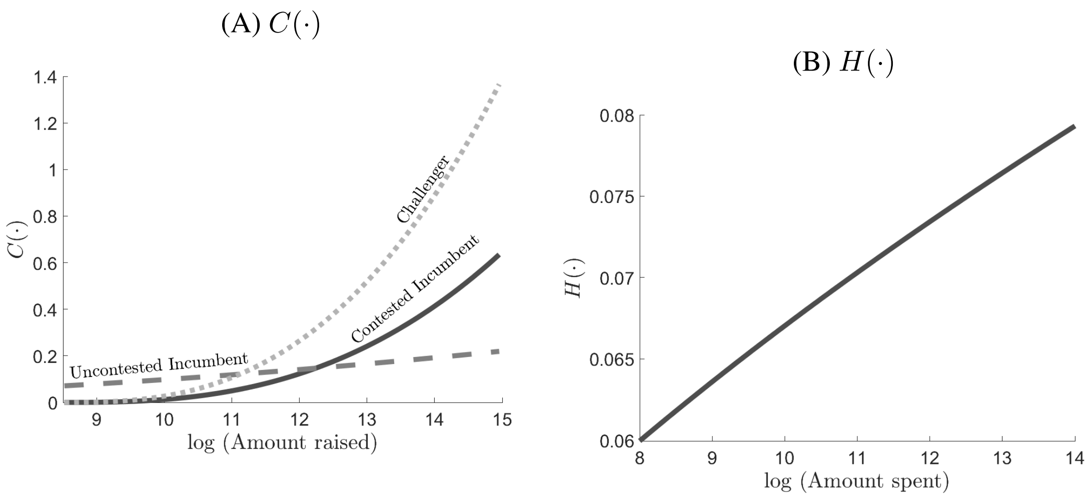



Figure 3: Cost of Fund-Raising and Benefit of Spending

Note: In Panel (A), the horizontal axis corresponds to the log amount raised and the vertical axis corresponds to the cost of raising money relative to the utility of winning. We use the mean of the estimated valence measures to draw the cost function. In Panel (B), horizontal axis corresponds to the log amount spent and the vertical axis corresponds to the benefit of spending money measured relative to the utility from winning.

point vote-share advantage for the incumbents. We also find a relatively small dispersion of valence measures among incumbents. The inter-quartile range of incumbent valence is about 3.8 percentage points. On the other hand, the valence measures of the challengers are more dispersed. The inter-quartile range is about 9.2 percentage points. Our finding that there is a longer tail of low-valence challengers is consistent with the fact that incumbents are selected partly by valence.

The bottom panel of Figure 4 plots the histogram of the estimated valence measures for open-seat candidates. We find that the upper tail of the distribution of the open-seat challengers resembles that of the incumbents. However, there is also a substantial mass of low valence open-seat challengers. The average valence measure of open-seat challengers is 1.5 percentage point lower than that of challengers that run against incumbents.[42] The inter-quartile range is about 7.8 percentage points.

**Valence of Winners and Losers, Democrats and Republicans** We now report the distribution of candidate valence by whether or not the candidate wins the election, and

[42] See Gowrisankaran et. al. (2008) for similar results regarding the U.S. Senate. In that paper, the authors estimate a parametric distribution of candidate valence measures for the U.S. Senate.

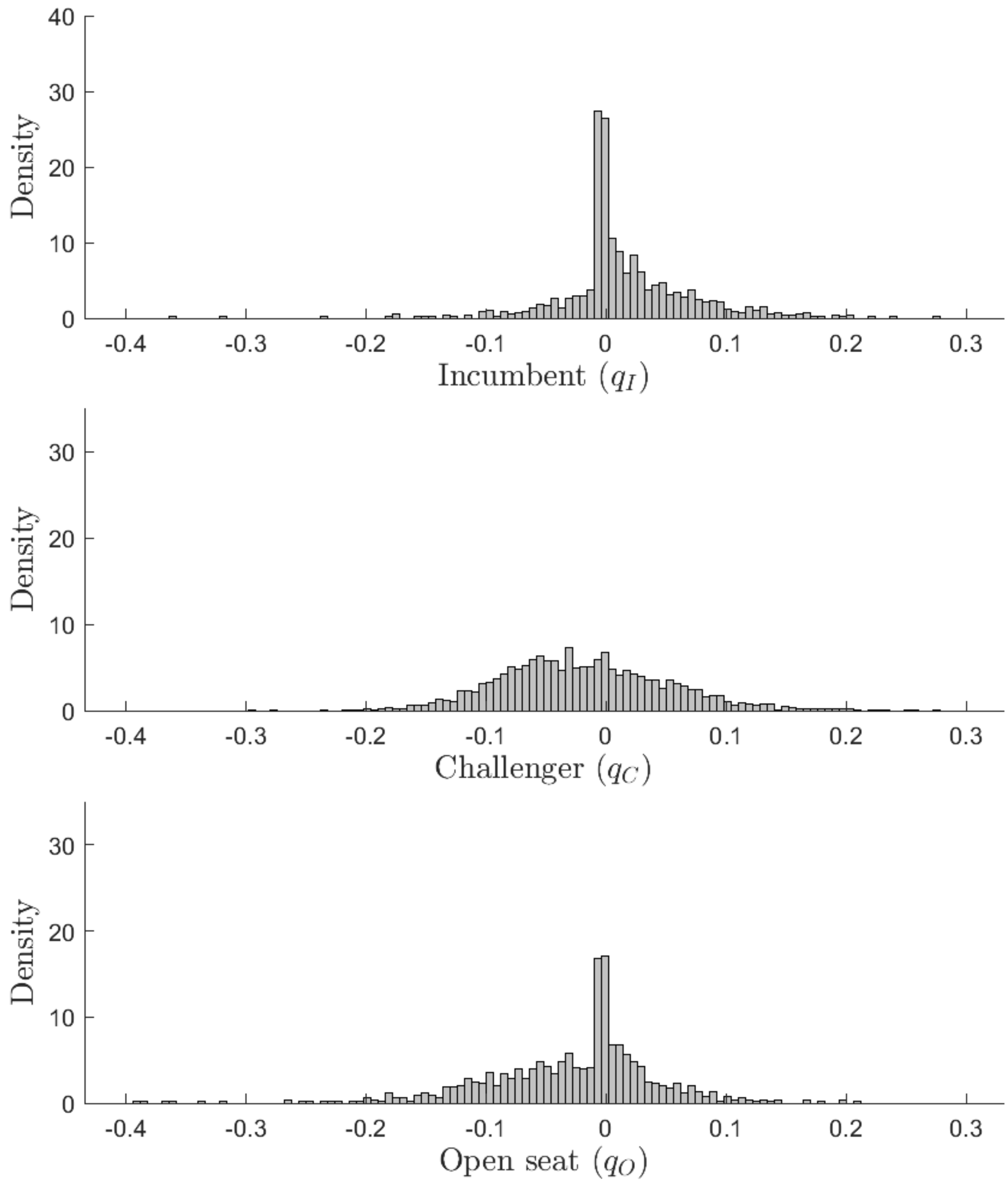



Figure 4: Distribution of Candidate Valence

Note: The top, middle and bottom panels correspond to the histogram of valence measures of incumbents, challengers running against incumbents, and open-seat candidates, respectively. The valence measures are scaled in unit of vote shares. The measures are normalized so that the average valence measure of incumbents estimated by the control function is zero.

by the party of the candidate. Panel (A) of Figure 5 illustrates the valence measure of incumbents, challengers, and open-seat candidates by whether or not the candidate wins the election. The gray bars correspond to the winners and the white bars correspond to the

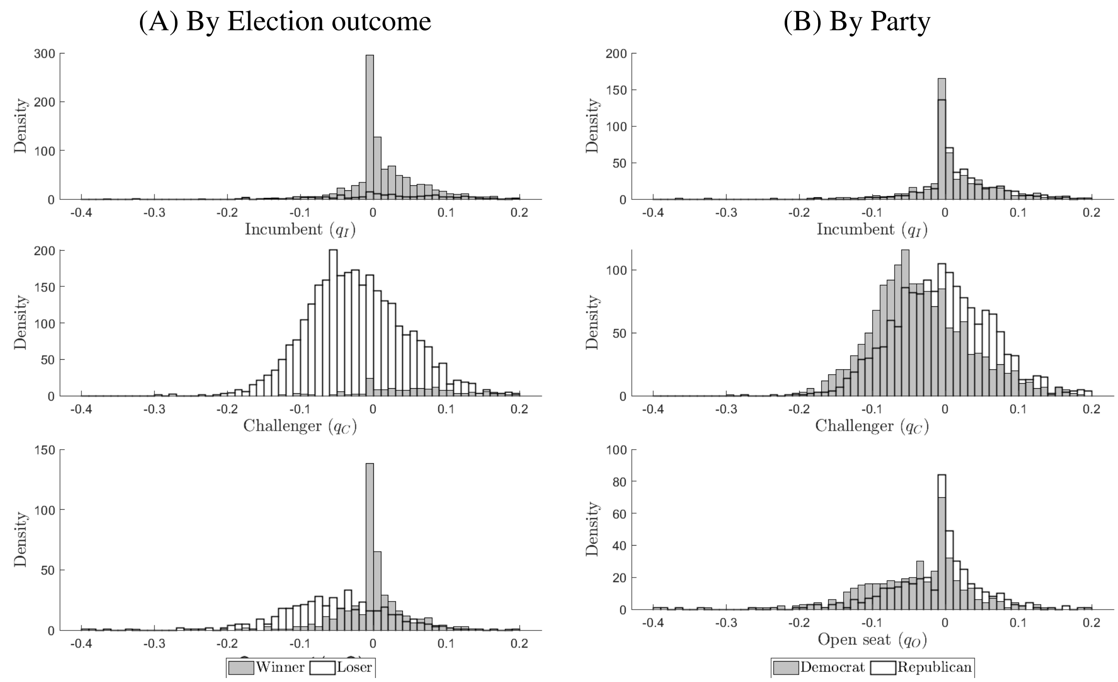



Figure 5: Distribution of Candidate Valence by Party and Election Outcome

Note: Panel (A) and (B) display the distribution of valence measures by the election outcome and the party of the candidate, respectively. The top, middle and bottom panels correspond to the histogram of valence measures of incumbents, challengers running against incumbents and open-seat candidates, respectively. The valence measures are in unit of vote share.

losers of the election. For incumbents, we find that the valence distribution of the winners and the losers are similar, although the mean valence is slightly higher for winners (0.017) than for losers (0.013). For challengers and open-seat candidates, we find that the valence measures of the winners are much higher than those of the losers. The average valence for challengers that win is 0.052, while it is -0.023 for challengers that lose. The average valence of open-seat candidates that win is -0.006 while the average for those that lose is -0.058.

Panel (B) of Figure 5 illustrates the valence measures broken down by party. The gray bars correspond to the Democrats and the white bars correspond to the Republicans. For incumbents, we do not find significant differences in the distribution of valence between the parties. For challengers, we find that Republican candidates have higher valence on average. The average valence of Republican and Democrat challengers that run against incumbents are -0.005 and -0.034, respectively, and the average valence of Republican and

| | | Serious 25 | Serious 50 | Serious 75 | Serious 90 |
|---|---|---|---|---|---|
| Spearman's $\rho$ | | | | | |
| | $q_C$ | 0.065 [0.008] | 0.128 [0.000] | 0.128 [0.000] | 0.152 [0.000] |
| Regression | | | | | |
| | Const. | 0.938 (0.007) | 0.804 (0.011) | 0.625 (0.013) | 0.513 (0.013) |
| | $q_C$ | 0.270 (0.088) | 0.705 (0.142) | 0.817 (0.169) | 0.984 (0.172) |
| Sample size | | 1703 | 1703 | 1703 | 1703 |

Note: We use challengers that run for election between 1992 and 2000. We report the Spearman's rank correlation coefficient and the regression coefficients. We report the p-values (in square brackets) for the Spearman's rank correlation coefficient and standard errors (in round brackets) for the regression.

Table 6: Correlation with Seriousness Measure of Maestas and Rugeley (2008)

Democrat open-seat challengers are -0.020 and -0.044.

**Comparison with Existing Valence Measures** In order to assess the validity of our measure of candidate valence, we compare our measure with a one constructed by Maestas and Rugeley (2008). In Maestas and Rugeley (2008), the authors construct four dummy variables (Serious 25, Serious 50, Serious 75, Serious 90) that capture the seriousness of the challengers that run for House seats between 1992 and 2000. The dummies are constructed based on observed characteristics of the candidates, such as previous political experience and extent of personal investment in the campaigns. The differences among the four dummies roughly reflect how much the candidate used his or her own personal funding in the campaign. Because these measures are specifically aimed at capturing factors that affect a candidate's probability of winning, they serve as good benchmarks of comparison.[43]

The first row of Table 6 reports the Spearman's rank correlation coefficient between our measure and each of the four measures of seriousness in Maestas and Rugeley (2008). We report the $p$-values in square brackets. We find that the rank correlation coefficients are positive and statistically significant. The second row of Table 6 reports the results from linear regressions in which we regress each of the four measures of seriousness on our measure of valence. We find that the coefficients on $q_C$ are positive and statistically significant. These results suggest that our measure captures an important aspect of candidate electability.

[43] We obtained the data from Maestas (2012).

# 6 Model Fit

In this section, we present the model fit. Figure 6 shows the histogram of the predicted (white bars) and the realized (gray bars) vote shares. The predicted vote share is computed using the estimated coefficients of the vote share equation (1), and plugging in the estimated valence measures of the candidates. The error term $\varepsilon$ is set to 0. The mean of the realized vote shares is 0.633 and that of the predicted vote shares is 0.632. In terms of winning probability, the model predicts 91.8% probability of incumbent winning whereas incumbents win 94.5% of times in the data.

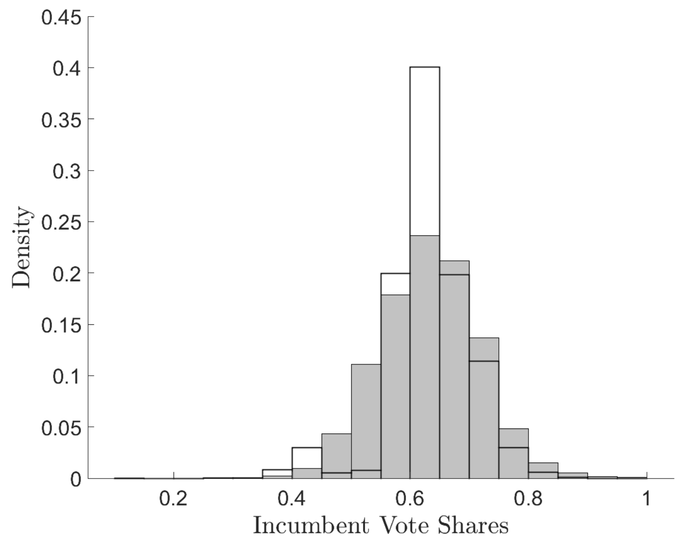
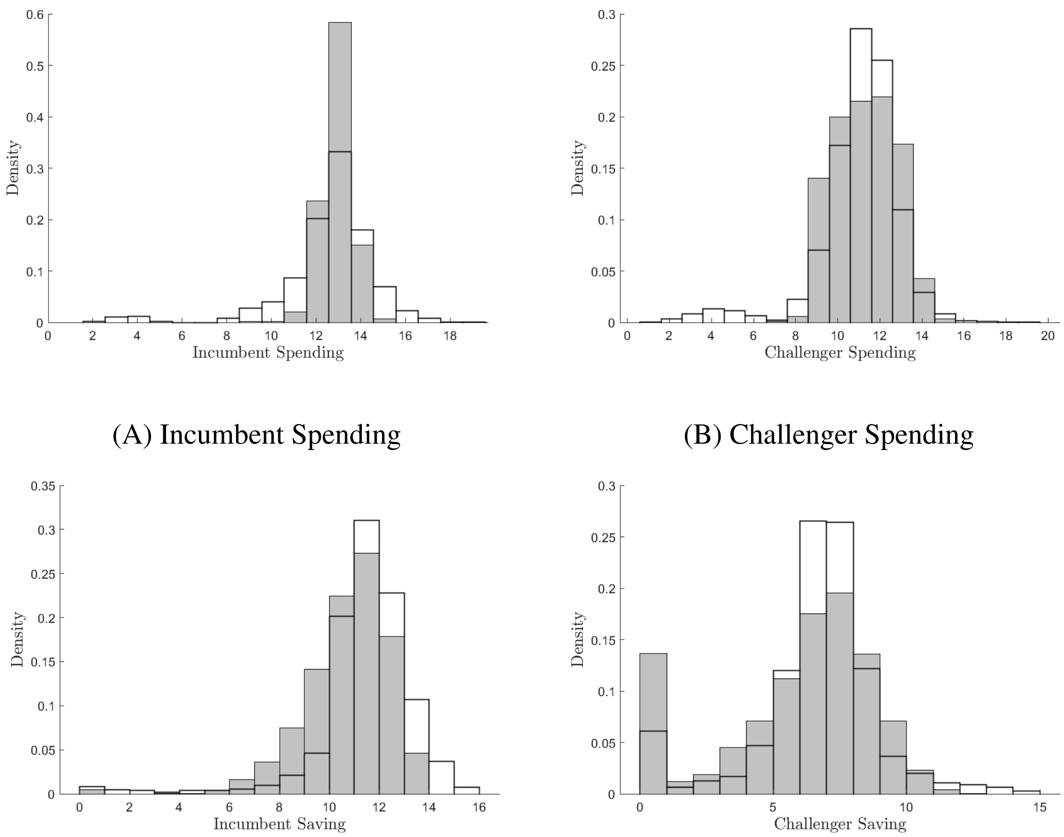



Figure 6: Model Fit - Incumbent Vote Shares

Note: Histogram of actual (gray bars) and predicted (white bars) incumbent vote shares.

Figure 7 reports the histograms of the predicted and the realized candidate actions in contested elections. The top two panels correspond to the spending of the candidates and the bottom two panels correspond to the savings. The predicted actions are computed by solving the first-order conditions of the candidates. The average predicted and actual log-spending of the incumbents are 12.60 and 12.94. For challengers, they are 11.02 and 11.30. The average predicted and actual log-saving of the incumbents are 11.42 and 10.73. For challengers, they are 6.64 and 5.89. We show the model fit for actions in unconstested elections in Appendix 10.11.




(A) Incumbent Spending

(B) Challenger Spending

(C) Incumbent Saving

(D) Challenger Saving

Figure 7: Model Fit - Actions in Contested Elections

Note: Histogram of observed actions (gray bars) and predicted actions (white bars). Predicted actions are obtained by solving the estimated first-order conditions election by election for optimal actions of incumbents and challengers.

# 7 Role of Valence and Competitiveness of House Elections

We now provide evidence that candidate valence plays a substantial role in determining electoral outcomes. To this end, we consider how the challengers' winning probabilities change if the out-party can field general election challengers whose valence is comparable to incumbents. Specifically, we consider a counterfactual in which for each challenger $C$, we replace $q_C$ with the corresponding percentile of the incumbent's valence distribution.[44]

In Figure 8, we plot the distribution of the challenger's estimated winning probabil-

[44]For example, if a challenger is located at the 10th percentile of the valence distribution among challengers, we replace her $q_C$ with the 10th percentile among incumbents.

ity for each contested election and the corresponding distribution under the counterfactual scenario. The solid curve corresponds to the distribution of the winning probability at the estimated values of challenger valence. The average winning probability is about 6.5%. There is a substantial mass of contested elections in which the challenger has less than 5% probability of winning.

The dotted line corresponds to the winning probability when we increase the challenger's valence, keeping fixed the spending of the incumbents and the challengers. We find a marked reduction in the proportion of contested elections in which the challenger's winning probability is very low, coupled with an increase in the proportion of elections with winning probabilities ranging from 10% to 25%. The average winning probability increases to about 12.1%, almost doubling the average winning probability of the baseline.

The dashed line in Figure 8 corresponds to the winning probability when we let candidates adjust their spending. Specifically, we compute the equilibrium spending levels by inserting the counterfactual values of $q_C$ into the policy functions of $d_I$ and $d_C$.[45] We find that the challenger's winning probability remains significantly above the baseline, but slightly less than the case with no adjustment. In other words, candidates' spending adjustments moderate the direct effect of increased challenger valence. The average winning probability of the challenger is 11.0%.[46]

Figure 9 plots the histogram of candidate spending at baseline (dark bars) and the corresponding histogram at the counterfactual values of $q_C$ when candidates are allowed to adjust their spending (white bars).[47] We find that the log spending of the incumbents increases by about 0.30 points, or about $144,600, and the log spending of the challengers increases by about 0.18 points, or about $16,400. This implies that the primary driver of the moderation effect we find in Figure 8 is the increase in incumbent spending.

# 8 Source of Incumbency Advantage in U.S. House Elections

The results in the previous section suggest that differences in candidate valence between the incumbents and the challengers play a substantial role in determining electoral outcomes.

[45] We discuss the estimation of policy functions in detail in Appendix 10.12.

[46] Incumbents' payoffs decline as a result of this shift in challenger valence and candidates' strategic responses to it. We find that on average, incumbents receive a value ($v_I$) of 3.18 at baseline, whereas they receive 2.97 in the counterfactual, both with and without spending adjustment.

[47] We use predicted values of spending for both the baseline and the counterfactual.

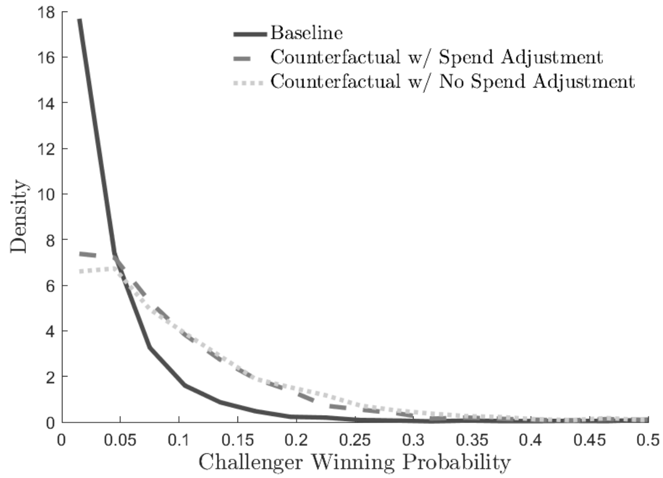



Figure 8: Challenger Winning Probability

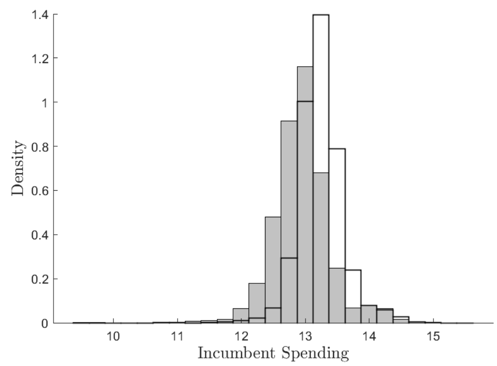
The results also suggest, however, that incumbents enjoy an electoral advantage that cannot be fully explained by differences in valence. In this section, we use the regression discontinuity (RD) framework introduced by Lee (2008) to decompose the incumbency advantage. Specifically, we decompose the incumbency advantage into differences in candidate valence, campaign spending, and policy positions, and quantify the contribution of each factor using the vote share equation we estimated. Since the work of Lee (2008) a large literature has developed that uses the RD design to identify and estimate the incumbency advantage.[48] Decomposing the incumbency advantage is important for understanding the effectiveness of various policies (e.g., subsidizing challengers' campaigns) to reduce the incumbency advantage and increase political competition.

Following Lee (2008), we define the incumbency advantage as the difference in the period $t+1$ vote share of the party who marginally won a seat in period $t$ and the party who marginally lost the seat in period $t$ as follows:

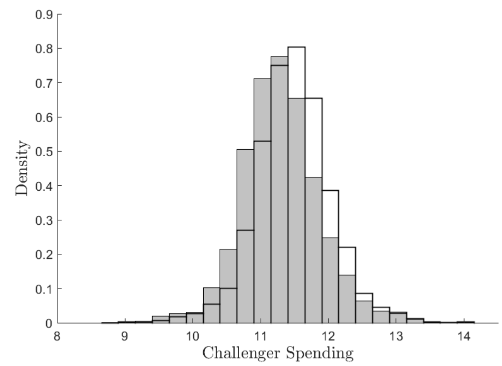
$$\lim_{\epsilon \to +0} \mathbb{E}[vote_{Dem,t+1} | vote_{Dem,t} = 0.5 + \epsilon] - \lim_{\epsilon \to +0} \mathbb{E}[vote_{Dem,t+1} | vote_{Dem,t} = 0.5 - \epsilon], \quad (16)$$

where $vote_{Dem,\tau}$ is the Democrat's vote share in period $\tau$. The RD estimate identifies the extra vote shares that a party gains from fielding an incumbent (who marginally won the previous election), relative to the case in which the party fields a challenger (against the rival party's marginal incumbent). Note that, because the Republican vote share is 1-

[48]See, e.g., Fouirnaies and Hall (2014) and Eggers et. al. (2015).




(A) Changes in $d_I$




(B) Changes in $d_C$

Figure 9: Campaign Spending at Baseline (Gray) and under the Counterfactual (White)

(Democrat vote share), we obtain an identical incumbency advantage measure if we use the Republican vote shares in equation (16).

Using the estimated measures of candidate valence and the parameters of the vote share equation, we decompose the incumbency advantage into a valence effect, a spending effect, a policy effect and a tenure effect. The valence effect is the difference in the valence between marginal winners and challengers fielded against incumbents. Similarly, the spending effect and the policy effect are the differences in the amount of spending and the adopted policy positions between marginal incumbents and average challengers. The tenure effect is the differences in tenure: a marginal incumbent has typically served several terms in office at the time of the election in period $t + 1$.

We first estimate the incumbency advantage in our sample by using the same regression discontinuity design as Lee (2008). Column (1) of Table 7 reports the results.[49] We find that the RD estimate of the incumbency advantage is 10.2 percentage points.[50] Figure 10 shows the binned scatter plot of the Democratic vote share in period $t + 1$ against the Democratic vote share in period $t$.

We now study how much of the incumbency advantage is explained by differences in candidate valence, spending, policy positions, and tenure. To do so, we estimate the same RD regression as Expression (16), but replace the outcome variable with valence, spending,

[49]We use the bias-correction estimator proposed by Calonico et. al. (2014) for all of our RD estimates.

[50]This result is reasonably close to the Lee's original result, which is around 8.0 percentage points. The difference of 2.2 percentage points is likely to reflect the fact that we only use elections from 1984, whereas Lee's data include elections from the 1950s. There is evidence that incumbency advantage is increasing over time (see, e.g., Gelman and King (1990)).

| | (1) Vote share | Valence | | Spending | | Policy | | (8) Tenure |
|---|---|---|---|---|---|---|---|---|
| | | (2) Dem | (3) Rep | (4) Dem | (5) Rep | (6) Dem | (7) Rep | |
| Estimate | 0.102 | 0.026 | 0.005 | 0.526 | -0.650 | 0.179 | 0.169 | 3.989 |
| | (0.012) | (0.011) | (0.009) | (0.138) | (0.133) | (0.055) | (0.049) | (0.599) |
| In vote share | 0.102 | 0.021 | | 0.043 | | 0.019 | | -0.003 |
| Bandwidth | 0.092 | 0.138 | 0.091 | 0.109 | 0.100 | 0.096 | 0.093 | 0.097 |
| Obs | 2320 | 2320 | 2320 | 2320 | 2320 | 2320 | 2320 | 2320 |

Note: The sample of elections used for this regression includes all election pairs between period $t$ and $t+1$ such that neither period is an uncontested election. Standard errors are reported in parentheses.

Table 7: RD Estimates of Incumbency Advantage

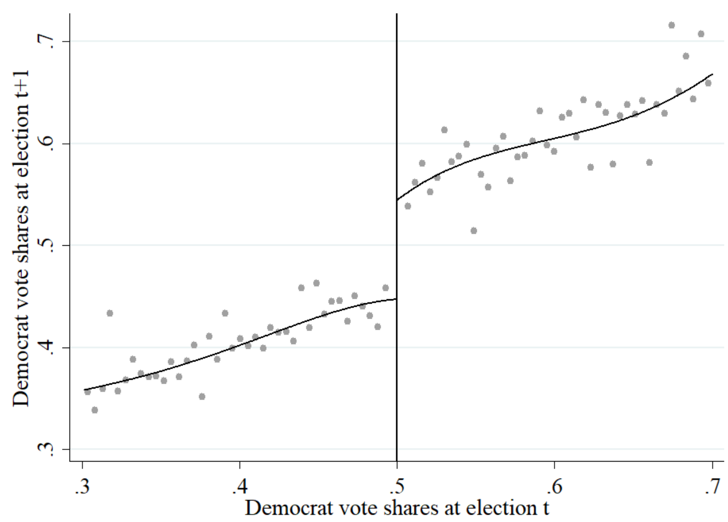



Figure 10: Binned Scatter Plot - Democratic Vote Share at t+1

Note: The figure plots the vote share of the Democratic candidates at $t+1$ on the vertical axis and the Democratic candidate's vote share in period $t$ on the horizontal axis. The curves in the figure correspond to a fourth-order polynomial approximation of the conditional expectation.

policy and tenure of the candidates. The running variable is always $vote_{Dem,t}$. Columns (2) through (9) of Table 7 report the results.

Columns (2) and (3) of Table 7 report the RD estimates for candidate valence. Column (2) corresponds to the case in which we take the outcome variable to be the valence of the Democratic candidate in period $t + 1$. The estimate (0.026) implies that a Democratic candidate who marginally wins in period $t$ (and hence is an incumbent in period $t + 1$) has higher valence measure than an average Democratic candidate that is fielded as a challenger by about 2.6 percentage points. Column (3) reports the corresponding RD estimate for the Republicans. We find that the valence of an average challenger is higher than that of marginal incumbents by 0.5 percentage points, a result consistent with the fact that Republican challengers have relatively high valence as we report in the right panel of Figure 5 in Section 5.3. However, the estimate is not statistically significant. The combined valence effect from the Democrats and the Republicans, reported in the second row, is about 0.024, implying that valence effect accounts for about 2.1 percentage points in terms of vote share.[51]

Panel (A) of Figure 11 is the binned scatter plot of the valence of the Democrats in period $t + 1$ and Panel (B) is that of the Republicans in period $t + 1$. In both panels, the horizontal axis is the Democratic party's two-party vote share in period $t$.

Columns (4) and (5) of Table 7 report the RD estimates for spending. For Column (4), we take the outcome variable in the RD regression to be the log spending of the Democratic candidate in period $t + 1$. The estimate implies that the log spending of a marginal Democratic incumbent is higher than the log spending of an average Democratic challenger by about 0.53 points. Similarly, the estimate in Column (5) implies that the log spending of an average Republican challenger is lower than a marginal Republican incumbent by about 0.65 points. The estimates correspond to about \$250,000 difference in spending for the Democrats and \$316,000 difference for the Republicans. Given our coefficient estimate on spending in the vote share equation, the combined spending effect accounts for about 4.3 percentage points in terms of vote share. Panel (A) of Figure 12 is the binned scatter plot of spending by the Democratic candidate in period $t + 1$ and Panel (B) is the corresponding plot for the spending of the Republicans.

Columns (6) and (7) report the coefficient estimates for the policy positions. We find

[51] The total effect in terms of vote share is estimated by a separate RD regression in which the dependent variable is set to $q_{Dem,t+1} - q_{Rep,t+1}$ for the case of $vote_{Dem,t} > 0.5$, and $q_{Rep,t+1} - q_{Dem,t+1}$, otherwise. In the absence of sampling error, the combined effect should be exactly equal to the effect for the Democratic candidates minus the effect for the Republican candidates.

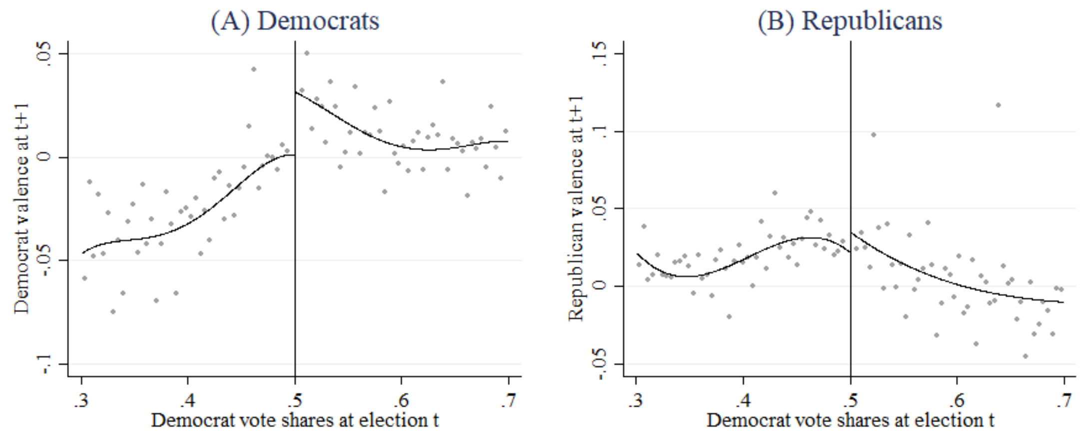



Figure 11: Binned Scatter Plot - Candidate Valence

Note: Panel (A) is the binned scatter plot of the valence of Democratic candidates at $t+1$ and Panel (B) is that of the Republican candidates at $t+1$. The horizontal axis for both panels is the Democratic candidate's vote share in period $t$. The curves in the figure correspond to a fourth-order polynomial approximation of the conditional expectation.

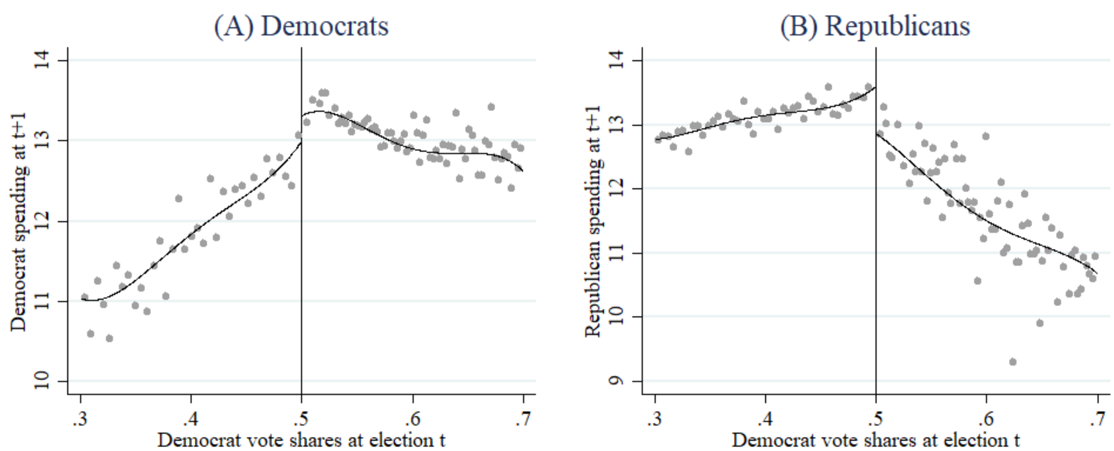



Figure 12: Binned Scatter Plot - Candidate Spending

Note: The left panel is the binned scatter plot of the log spending of Democratic candidates at $t+1$ and the right panel is that of the Republican candidates at $t+1$. The horizontal axis for both panels is the Democratic candidate's vote share in period $t$. The curves in the figure correspond to a fourth-order polynomial approximation of the conditional expectation.

that the marginal incumbent has a more centrist position than the average Democratic challenger by about 0.18 points. We also find that the average Republican challenger has a

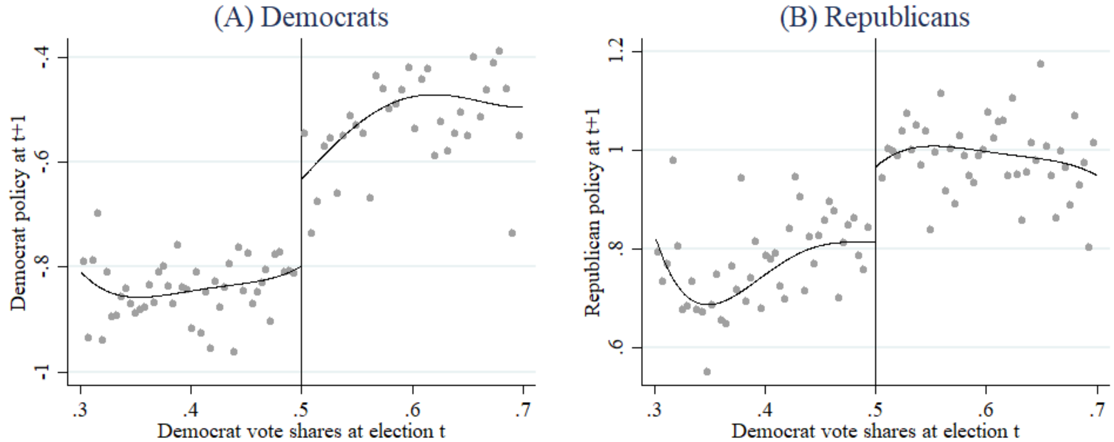



Figure 13: Binned Scatter Plot - Policy Positions

Note: The left panel is the binned scatter plot of the policy position of Democratic candidates at $t+1$ and the right panel is that of the Republican candidates at $t+1$. The horizontal axis for both panels is the Democratic candidate's vote share in period $t$. The curves in the figure correspond to a fourth-order polynomial approximation of the conditional expectation.

more conservative position than the marginal Republican incumbent by about 0.17 points. These findings imply that incumbents take more centrist positions than challengers for both parties and the effect on the vote share is about 1.9 percentage points. Figure 13 reports the corresponding binned scatter plots.

Lastly, we consider the component of the incumbency advantage that is attributable to the differences in the tenure of the candidates. The RD estimate of the tenure of the Democrats is 2.31 terms and the estimate for the Republicans is -1.69 terms. These estimates imply that a marginal Democratic winner has served about 2.31 terms in office at period $t+1$ and the marginal Republican winner has served about 1.69 terms in office. Column (8) of Table 7 reports the combined effect on tenure, which is 3.99 terms. Given that our coefficient estimate on tenure in the vote share equation is small (-0.001), the differences in tenure between the incumbent and the challenger translates to an incumbency advantage of about -0.3 percentage points in terms of vote share.

To summarize, we find that differences in candidate valence accounts for about 2.1 percentage points in terms of vote share. Differences in candidate spending accounts for about 4.3 percentage points and differences in the policy positions explain about 1.9 percentage points. Our results suggest that differences in candidate valence account for a substantial component of the incumbency advantage. This in turn suggests that policy interventions designed to eliminate incumbency advantage through the spending channel, such as subsidizing challengers' campaigns, may only be partially effective.

# 9 Conclusion

Although candidate valence plays a prominent role in many theoretical models of political competition, measuring candidate valence has been challenging. We think that the methods developed in this paper can serve as a starting point for testing and estimating models of political competition with vertical differentiation among candidates. We acknowledge that our valence measures are based on relatively strong assumptions on the model structure. For instance, we assume that the valence term is a scalar that enters additively into the vote share equation, an assumption akin to Hicks neutrality in the context of production function estimation. We think of relaxing the scalar unobservable assumption and parametric functional form restrictions on the vote share equation as potentially useful avenues of future research.

# References


[1] Abramowitz, A. I. (1991), "Incumbency, Campaign Spending, and the Decline of Competition in U.S. House Elections," *The Journal of Politics* 53(1), 34–56.

[2] Abramowitz, A. I., Cover, A. D. & Norpoth, H. (1986), "The President's Party in Midterm Elections: Going from Bad to Worse,"*American Journal of Political Science* 30(3), 562—576.

[3] Ackerberg, N., Caves, K. & Frazer, G. (2006), "Identification Properties of Recent Production Function Estimators," *Econometrica* 83(6), 2411–2451.

[4] Adams, J. & Merrill, S. III (2008), "Candidate and Party Strategies in Two-Stage Elections Beginning with a Primary," *American Journal of Political Science* 52(2), 344–359.

[5] Adler, S. E. (2003), "Congressional District Data File, 1984-2000." University of Colorado, Boulder, CO.

[6] Ai, C. & Chen, X. (2003), "Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions," *Econometrica* 71(6), 1795–1843.

[7] Ansolabehere, S., Snyder, J. M. Jr. & Stewart, C. III (2000), "Old Voters, New Voters, and the Personal Vote: Using Redistricting to Measure the Incumbency Advantage," *American Journal of Political Science* 44(1), 17–34.

[8] Ansolabehere, S., Meredith, M & Snowberg, E. (2014), "Mecro-Economics Voting: Local Information and Micro-Perceptions of the Macro-Economy," *Economics and Politics* 26(3), 380–410.

[9] Aragones, E. & Palfrey, T. R. (2004), "The Effect of Candidate Quality on Electoral Equilibrium: An Experimental Study," *American Political Science Review* 98(1), 77–90.

[10] Bajari, P. C., Benkard, L. & Levin, J. (2007), "Estimating Dynamic Models of Imperfect Competition," *Econometrica* 75(5), 1331–1370.

[11] Barwick, P. J. & Pathak, P. A. (2015), "The Costs of Free Entry: An Empirical Study of Real Estate Agents in Greater Boston," *RAND Journal of Economics* 46(1), 103–145.

[12] Berry, S., Levinsohn, J. & Pakes, A. (1995), "Automobile Prices in Market Equilibrium," *Econometrica* 63(3), 841–890.

[13] Bonica, A. (2023), Database on Ideology, Money in Politics, and Elections: Public version 3.1 [Computer file]. Stanford, CA: Stanford University Libraries. https://data.stanford.edu/dime.

[14] Buisseret, P. & Van Weelden, R. (2022), "Polarization, Valence, and Policy Competition," *American Economic Review: Insights* 4(3), 341–352.

[15] Bureau of Labor Statistics "Local Area Unemployment Statistics." (downloaded Jul 06, 2011). https://www.bls.gov/lau/.

[16] Calonico, S., Cattaneo, M. & Titiunik, R. (2014), "Robust Nnonparametric Confidence Intervals for Regression-Discontinuity Designs,"*Econometrica* 82(6), 2295–2326.

[17] Campbell, J. (1985), "Explaining Presidential Losses in Midterm Congressional Elections," *The Journal of Politics* 47(4), 1140–1157.

[18] Carter, J. & Patty, J. W. (2015), "Valence and Campaigns," *American Journal of Political Science* 59(4), 825–840.

[19] da Silveira, B. & de Mello, J. (2011), "Campaign Advertising and Election Outcomes: Quasi-natural Experiment Evidence from Gubernatorial Elections in Brazil," *Review of Economic Studies* 78(2), 590–612.

[20] Demirer, M. (2022), "Production Function Estimation with Factor-Augmenting Technology: An Application to Markups," Working Paper.

[21] Diermeier, D., Keane, M. & Merlo, A. (2005), "Political Economy Model of Congressional Careers," *American Economic Review* 95(1), 347–373.

[22] Doraszelski, U. & Jaumandreu, J. (2018), "Measuring the Bias of Technological Change," *Journal of Political Economy* 126(3), 1027–1084.

[23] Eggers, A. C., Fowler, A., Hainmueller, J., Hall, A. B. & Snyder Jr., J. M. (2015), "On the Validity of the Regression Discontinuity Design for Estimating Electoral Effects: New Evidence from Over 40,000 Close Races," *American Journal of Political Science* 59(1), 259–274.

[24] Erikson, R. S. (1988), "The Puzzle of Midterm Loss," *The Journal of Politics* 50(4), 1011–1029.

[25] Erikson, R. S. (1971), "The Advantage of Incumbency in Congressional Elections," *Polity* 3(3), 395–405.

[26] Erikson, R. S. & Palfrey, T. R. (2000), "Equilibria in Campaign Spending Games: Theory and Data," *American Political Science Review* 94(3), 595–609.

[27] Federal Election Commission "Candidates for House of Representatives" (downloaded Sep 27, 2011). https://www.fec.gov/data/candidates/house/.

[28] Fouirnaies, A. & Hall, A. B. (2014), "The Financial Incumbency Advantage: Causes and Consequence," *The Journal of Politics* 76(3), 711–724.

[29] Frey, A., López-Moctezuma, G., & Montero, S. (2023), "Sleeping with the Enemy: Effective Representation under Dynamic Electoral Competition," *American Journal of Political Science* 67(4), 915–931.

[30] Gallant, R. A. & Nychka, D. W. (1987), "Semi-Nonparametric Maximum Likelihood Estimation," *Econometrica* 55(2), 363–390.

[31] Gelman, A. & King, G. (1990), "Estimating Incumbency Advantage without Bias," *American Journal of Political Science* 34(4), 1142–1164.

[32] Gerber, A. (1998), "Estimating the Effect of Campaign Spending on Senate Election Outcomes Using Instrumental Variables," *American Political Science Review* 92(2), 401–411.

[33] Gandhi, A., Navarro, S. & Rivers, D. A. (2020), "On the Identification of Gross Output Production Functions," *Journal of Political Economy* 128(8), 2973-3016.

[34] Gordon, B. & Hartmann, W. (2013), "Advertising Effects in Presidential Elections," *Marketing Science* 32(1), 19–35.

[35] Gowrisankaran, G,, Mitchell, M. & Moro, A. (2008) "Electoral Design and Voter Welfare from the U.S. Senate: Evidence from a Dynamic Selection Model," *Review of Economic Dynamics* 11(1), 1–17.

[36] Green, D. & Krasno, J. (1988), "Salvation for the Spendthrift Incumbent: Reestimating the Effects of Campaign Spending in House Elections," *American Journal of Political Science* 32(4), 884–907.

[37] Hibing, J. R. & Alford, J. R. (1981), "The Electoral Impact of Economic Conditions: Who is Held Responsible?," *American Journal of Political Science* 25(3), 423–439.

[38] Hotz, J. V., Miller, R., Sanders, S. & Smith, J. (1994), "A Simulation Estimator for Dynamic Models of Discrete Choice," *Review of Economic Studies* 61(2), 265–289.

[39] Iaryczower, M., Montero, G, & Kim, G. (2023), "Representation Failure," Working paper.

[40] Jacobson, G. C. (1978), "The Effects of Campaign Spending in Congressional Elections," *American Political Science Review* 72(2), 469–491.

[41] Kawai, K., & Sunada, T. (2024): ”Supplement to ’Estimating Candidate Valence’,” Econometrica Supplemental Material.

[42] Kawai, K., Toyama, Y., & Watanabe, Y. (2021), “Voter Turnout and Preference Aggregation,” *American Economic Journal: Microeconomics* 13(4), 548–586.

[43] Kendall, C., Nannicini, T., & Trebbi, F. (2015), “How Do Voters Respond to Information? Evidence from a Randomized Campaign,” *American Economic Review* 105(1), 322–353.

[44] Lee, D. S. (2008), “Randomized Experiments from Non-Random Selection in U.S. House Elections,” *Journal of Econometrics* 142(2), 675–697.

[45] Levendusky, M. S., Pope, J. C. & Jackman, S. D. (2008), “Measuring District-Level Partisanship with Implications for the Analysis of U.S. Elections,” *The Journal of Politics* 70(3), 736–753.

[46] Levinsohn, J. & Petrin, A. (2003), “Estimating Production Function Using Inputs to Control for Unobservables,” *Review of Economic Studies* 70(2), 317–341.

[47] Levitt, S. D. (1994), “Using Repear Challengers to Estimate the Effect of Campaign Spending on Election Outcomes in the U.S. House,” *Journal of Political Economy* 102(4), 777–798.

[48] Levitt, S. D. & Wolfram, C. D. (1997), “Decomposing the Sources of Incumbency Advantage in the U.S. House,” *Legislative Studies Quarterly* 22(1), 45–60.

[49] Maestas, C. D. & Rugeley, C. R. (2008), “Assessing the “Experience Bonus” Through Examining Strategic Entry, Candidate Quality, and Campaign Receipts in U.S. House Elections,” *American Journal of Political Science* 52(3), 520–535.

[50] Maestas, C. D. ”Research.” (downloaded Aug 14, 2012). https://cheriemaestas.com/Research.html.

[51] Magnac, T. & Thesmar, D. (2002), “Identifying Dynamic Discrete Decision Processes,” *Econometrica* 70(2), 801–816.

[52] Maskin, E. & Tirole, J. (1988), “A theory of dynamic oligopoly, I: Overview and quantity competition with large fixed costs,” *Econometrica* 56(3), 549–569.

[53] Montero, S. (2023). “Going It Alone? A Structural Analysis of Coalition Formation in Elections ,” *Journal of Politics*, forthcoming.

[54] Olley, S. & Pakes, A. (1996), “The Dynamics of Productivity in the Telecommunications Equipment Industry,” *Econometrica* 64(6), 1263–1297.

[55] POLIDATA ”POLITICAL AND DEMOGRAPHIC DATASETS FOR SALE.” (downloaded Sep 22, 2015). https://polidata.org/data/default.htm.

[56] Poole, K. & Rosenthal, H. (1985), “A Spatial Model For Legislative Roll Call Analysis,”*American Journal of Political Science* 29(2), 357-–384.

[57] Rekkas, M. (2007), “The Impact of Campaign Spending on Votes in Multiparty Elections,” *The Review of Economics and Statistics* 89(3), 573–585.

[58] Serra, G. (2011), “Why Primaries? The Party’s Tradeoff between Policy and Valence,” *Journal of Theoretical Politics* 23(1), 21–51.

[59] Sieg, H. & Yoon, C. (2017), “Estimating Dynamic Games of Electoral Competition to Evaluate Term Limits in US Gubernatorial Elections,” *American Economic Review* 107(7), 1824–1857.

[60] Snyder, J. M. & Ting, M. M. (2011), “Electoral Selection with Parties and Primaries,” *American Journal of Political Science* 55(4), 782–796.

[61] Stein, R. M. (1990), “Voting for Governor and U.S. Senator: The Electoral Consequences of Federalism,” *The Journal of Politics* 52(1), 29–53.

[62] Stone, J. W., Futon, S. A., Maestas, C. D. & Maisel, L. S. (2010), “Incumbency Reconsidered: Prospects, Strategic Retirement, and Incumbent Quality in U.S. House Elections,” *The Journal of Politics* 72(1), 178–190.

[63] Stone, J. W. & Simas, E. N. (2010), “Candidate Valence and Ideological Positions in U.S. House Elections,” *American Journal of Political Science* 54(2), 371–388.

[64] Stratmann, T. (2005), “Some talk: Money in politics. A (partial) review of the literature,” *Public Choice* 124, 135–156.

[65] U.S. Census Bureau ”Census.” (downloaded Sep 21, 2015). https://data.census.gov/.