Sequential Primaries, Pandering, and Information Transfer*

N. D. Hill†

Johns Hopkins University

October 7, 2005

There are worse ways to choose a leader—hereditary monarchy, for example, or by consensus within an aged Politburo—but one wonders what history’s verdict will be as it records how...240 million people...allowed some 250,000 voters off in one odd corner of the country to winnow down...the list of serious competitors for the preeminent position in American government and American life.

The Almanac of American Politics 1988

Abstract

I analyze the effect of a sequential election system when the first voter has private information using a simple two candidate, two voter model in which the second voter is decisive. Both voters observe the candidates’ policy positions, but only the first voter observes which candidate is competent. I show that in equilibrium the candidates pander to the policy preferences of the first voter. Despite the pandering that it introduces, I show that a sequential election can be a Pareto improvement over a simultaneous election.

JEL classification: D71, D72.

Keywords: Sequential Election, Pandering

---

*I wish to thank Jimmy Chan, Joseph Harrington, Bruce Hamilton and the seminar audience at Johns Hopkins University.

†Department of Economics, Johns Hopkins University, 3400 N. Charles Street, Baltimore, Maryland 21218, United States of America. Email: nhill3@jhu.edu.
1 Introduction

Iowa and New Hampshire, which hold the first caucus and primary in the nation, respectively, exert tremendous influence on presidential nomination contests. Candidates campaign extensively in these states, and their results are watched closely by voters, potential donors and the media. In nine of the last ten nomination contests that did not include a sitting president, the eventual nominee won either Iowa or New Hampshire.\(^1\) Table 1, which shows the results of national tracking polls taken before and after the New Hampshire primary, illustrates the influence of New Hampshire in 2004; note the strong boost received by John Kerry, the winner of the primary and, eventually, the nomination.

<table>
<thead>
<tr>
<th>Candidate</th>
<th>Jan 20-22</th>
<th>Jan 29-31</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kerry</td>
<td>29</td>
<td>36</td>
</tr>
<tr>
<td>Edwards</td>
<td>17</td>
<td>11</td>
</tr>
<tr>
<td>Dean</td>
<td>14</td>
<td>15</td>
</tr>
<tr>
<td>Clark</td>
<td>11</td>
<td>9</td>
</tr>
<tr>
<td>Lieberman</td>
<td>9</td>
<td>7</td>
</tr>
</tbody>
</table>

*Source: Rasmussen Reports*

Table 1: 2004 National Tracking Polls

The influence of Iowa and New Hampshire on the nomination contest may introduce an incentive for candidates to tailor their policies towards the preferences of voters in these states. The case of ethanol subsidies is a prime example of this potential for pandering. Ethanol subsidies are widely viewed as poor policy for the country as a whole, but they are popular in Iowa, a major producer of corn (from which ethanol is produced). As a result, Presidential candidates—with few exceptions—fall over one another in their eagerness to praise ethanol subsidies. As John McCain put it in 2000 when announcing his refusal to support ethanol subsidies:\(^2\)

> I’m going to tell you the things you don’t want to hear. Ethanol is not worth it. It does not help the consumer. Those ethanol subsidies should be phased out, and everybody...if it wasn’t for the fact that Iowa was (sic) the first caucus state, would share my view.

McCain’s quote and the example of ethanol may make pandering seem inevitable, but the logic of pandering is complex. While pandering may increase a candidate’s chance of winning in Iowa or New Hampshire, it may also alienate voters in other states and decrease her chance of winning the nomination. In this paper I investigate whether rational candidates will pander to voters in early primaries using a simple model with two voters and two candidates. In my model there is a single representative voter in each state.\(^3\)

---

\(^1\) The exception was Bill Clinton in 1992; Tom Harkin won in Iowa and Paul Tsongas won in New Hampshire.

\(^2\) All of McCain’s opponents endorsed ethanol subsidies.
These voters share a taste for a valence characteristic—which I shall call competence—but have different policy preferences. My two voters constitute a sequential primary in which the voter in state one votes first, but the voter in state two is decisive. The two candidates cannot compete on valence grounds (indeed, neither candidate knows if she is competent) but they do compete on policy grounds by simultaneously selecting binding policy positions.

I make two critical assumptions in my model. First, I assume that the first voter observes which of the two candidates is competent and that second voter does not. This assumption is a simplification, but it reflects the fact that the small size and “first in the nation” status of Iowa and New Hampshire allows voters there enjoy extraordinary access to presidential candidates. These voters meet candidates in intimate and varied settings and engage in give-and-take, rather than listening to a well-honed stump speech. This unusual access to candidates allows voters in Iowa and New Hampshire to evaluate candidates’ competence in a manner not available to voters in other states. This point was forcefully enunciated by Bill Jones, then the Secretary of State of California, in the Los Angeles Times on May 6, 1996:

Earlier this year residents of a small New England state with a population roughly equivalent to that of California’s Alameda County were able to eat breakfast, share war stories and then intimately discuss the future of America one-on-one with a number of prominent presidential candidates. On Feb. 20, the nation watched as New Hampshire voters played a major role in determining who would be the nominees for president. Why New Hampshire? Why not Alameda County?

My second critical assumption is that the second voter calculates the candidates’ posterior probabilities of being competent based upon the result of the first primary and not the margin of victory. There are two complementary explanations for this assumption. First, drawing inferences based upon the result of an election is less demanding than drawing inferences based upon the margin of victory. In particular, since the result of an election is discrete while the margin of victory is continuous, if voters draw inferences based upon past experience, they must observe and process many more elections if they wish to draw conclusions using the margin of victory. Second, the nature of media coverage of primary elections, in which the identity of the winner is stressed and the margin of victory is glossed over, makes the cost of obtaining information about the margin of victory higher. Many Americans know that John Kerry won both Iowa and New Hampshire, but few remember—or ever knew—his margin of victory.

In addition to two critical assumptions, I impose a third, weak assumption that voters have random utility shocks and that these shocks have a mean zero, single-peaked distribution. Politically, this assumption means that the relationship between a candidate’s electoral advantage (i.e., the extent to which a voter prefers her to her rival) and her probability of victory displays diminishing returns: as her electoral advantage increases, her probability of victory increases but at an ever decreasing rate.

---

3 State two’s decisive role reflects the fact that the overwhelming majority of delegates come from states other than Iowa and New Hampshire.

4 This quote was taken from Morton and Williams (1999).
In my model, under these three assumptions, pandering has both benefits and costs. Since the first voter has private information about competence, victory in the first primary is a signal that a candidate is competent and improves her probability of victory in the second primary. Given this, pandering is beneficial because it increases the probability that a candidate will win the first primary and receive the associated positive signal.

Pandering is costly because it reduces a candidate’s chance of winning the second primary conditional on the outcome of the first. It does so in two ways: first, it moves a candidate away from the preferred policy position of the second voter, and second, it reduces the posterior probability that she is competent given the result of the first primary. If pandering decreases a candidate’s probability of winning the second primary conditional on having won the first primary and her probability of winning the second primary conditional on having lost the first primary, how can it improve her probability of winning the nomination? The answer is that the candidate’s chance of winning the second primary is higher when she wins the first primary than when she loses it, and pandering increases the chance that she will win the first primary. In short, pandering increases the chance of sending a positive signal while decreasing the value of this signal and making one’s policy position less attractive to the second voter.

But when—if ever—does the benefit of pandering outweigh its costs? I find, under mild additional assumptions, that in any pure strategy equilibrium, there will be pandering. This result is driven by my two critical assumptions. The first creates an incentive to pander by making the result of the first primary informative about the candidates’ competence. The second maintains this incentive by keeping the first primary from being too informative: if voters can interpret (and learn) margins of victory, they will simply scale up their expectations for a pandering candidate, removing the benefit to pandering.

The intuition behind the existence of pandering is best understood by initially imagining that the second voter does not care about competence. Under this condition, in equilibrium both candidates will adopt the second voter’s preferred policy position. From this starting point, increasing the weight placed upon competence by the second voter has three effects. First, it increases the importance of competence in the second primary which increases the influence of the first primary on the second. This increases the benefit of pandering because it increases the payoff to winning the first primary. Second, it increases the value of being seen to be competent by the second voter. This increases the cost of pandering because pandering reduces the posterior probability that a candidate is competent (i.e., it decreases her electoral advantage if she wins the first primary and increases her electoral disadvantage if she loses). Third, by increasing the influence of competence in the second primary, it makes the second primary less competitive conditional on the result of the first (i.e., the electoral advantage of the winner of the first primary—and hence her probability of winning the second primary—increases).

The first and second effects would cancel out except for the third effect, which blunts the second. In particular, since the third effect increases the electoral advantage of the winner of the first primary, it weakens the relationship between electoral advantage and

---

\[5\] In particular, I assume that both voters care about competence and that the second voter and the candidates share a common prior belief that each candidate has an equal chance of being competent. I relax the second assumption when I extend the model in section 6.
probability of victory—which reduces the impact of the second effect. The result is that the cost of pandering decreases relative to its benefit when the importance of competence to the second voter rises.

While I find that there is always pandering in equilibrium, I find that a pure strategy equilibrium does not always exist. Interestingly, the breakdown in equilibrium occurs when the second voter cares a great deal about competence while the first voter cares little. In such a situation, the incentive to pander is strong—since the second voter is hungry for information—but the quality of the signal is weak (because the first voter cares mostly about policy). This combination means that a candidate equilibrium in which candidates pander shamelessly is vulnerable to a deviation in which one candidate adopts a non-pandering or “principled” position.\textsuperscript{6} Such a candidate is likely to lose the first primary regardless of her competence, so the first primary is uninformative about competence. Further, because she doesn’t pander, her policy position is more attractive to the second voter, leaving her in an excellent position to win the second primary.

I find also that if the first voter cares sufficiently about competence relative to policy then a pure strategy equilibrium always exists. This condition ensures existence of equilibrium because it ensures that the first primary is informative (by making that contest primarily about competence) and because it reduces pandering to a bare minimum (when the first voter cares only about competence, there is little benefit to pandering).

Given that pandering is always present in equilibrium, I go on to investigate whether voters would be better off under a simultaneous primary system. I find that if the first voter cares sufficiently about competence relative to policy, both voters are better off under a sequential primary system than under a simultaneous primary system. Intuitively, if the first voter values highly competence, there is little incentive to pander and the quality of the signal that she sends is strong. The first primary therefore improves the ability of the second voter to identify the competent candidate while introducing little pandering. In this case, the sequential primary system transfers useful information while introducing only a small policy distortion.

Combining my results, I find that if the first voter cares sufficiently about competence, there exists a unique equilibrium characterized by little pandering and improved welfare. Thus, an influential first primary is desirable if its voters are sufficiently interested in competence, despite the pandering that it introduces.

The rest of this paper is organized as follows. Section 2 presents a review of the relevant literature. Section 3 lays out my model. Section 4 analyzes the effects of a first primary on candidates’ policy positions and presents simple comparative statics. Section 5 explores the welfare effects of a first primary. Section 6 extends the basic model model. Section 7 concludes.

2 Literature Review

Many papers in the economics and political science literatures examine the characteristics and effects of a sequential primary system. This existing literature can be broken down into three distinct strands that are relevant to the questions addressed in this paper.

\textsuperscript{6}For example, John McCain’s refusal to support ethanol subsidies.
The first stand concerns the importance of the results of early primaries. The momentum camp (see Aldrich (1980) for the classic treatment) argues that success breeds success: well-financed candidates win primaries and candidates that win primaries find it easier to raise money. The organizations camp (see Green and Hinckley (1996) for empirical evidence from the 1988 primaries) argues that success in fund raising is more a function of a strong organization (e.g., a stout donor base) than results in the primary season. This view weakens the momentum argument since it removes a natural mechanism for spinning victory into victory. There are other mechanisms to be sure—increased media coverage comes to mind—but no candidate can win the nomination without fund raising success. The sensible compromise view that a strong position before Iowa aids—but does not guarantee—success is succinctly presented by Mayer (1996).

Even if the organization view is correct though, it doesn’t diminish the importance of Iowa and New Hampshire, since there are typically several well funded candidates at the beginning of any primary contest that does not involve a sitting president. Iowa and New Hampshire often serve to differentiate between these advantaged candidates, while also occasionally giving a boost to a rank outsider such as Jimmy Carter.

A second strand of the literature concerns the relative merits of sequential and simultaneous elections. Since the primary system gained a central role in 1972, states have been moving their primaries towards the front of the election calendar in an attempt to increase their influence. Taken to its extreme, this trend could move the primary system away from a sequential election and towards a simultaneous election, with a possible loss in the aggregation of private information. This has led a number of authors to examine the effect of this trend on voters’ welfare.\(^7\)

Morton and Williams (1999) test the performance of sequential and simultaneous elections both theoretically and experimentally. In their model the Condorcet winner wins more often in a sequential election (which consists of two primaries) than in a simultaneous election. This is because voters in the second primary can infer information from the results of the first primary. There are two critical differences between their model and my model. First, there is no electoral competition in their model: candidates are assigned one of three fixed positions (liberal, moderate, or conservative). Second, voters in their model are uncertain about the policy positions of candidates, not about their competence.

In a related paper, Dekel and Piccione (2000) consider simultaneous versus sequential voting procedures in a model with \(n\) ex ante identical voters, two fixed alternatives, and incomplete information. They show that any informative symmetric equilibrium of the simultaneous game is an equilibrium of any sequential game. Their model differs from mine in that they consider fixed alternatives (i.e., there is no electoral competition) and the structure of the uncertainty is quite different (in their model, voters receive exogenous signals about the value of the alternatives).

Finally, a small literature considers the importance of personal and valence characteristics. Ansolabehere and Snyder (2000) show that differences in valence characteristics can lead to equilibria in multidimensional spatial competition. Groseclose (2001) and Aragones and Palfrey (2002) show that in a two candidate race, the candidate with superior valence characteristics will adopt a more moderate position than her opponent.

\(^7\)Closely related to these papers are the herding and information cascade literatures. See, for example, Banerjee (1992), Bikhchandani (1992), and Gul and Lundholm (1995).
and win more often. While both of these papers assume that voters have full information about candidates’ valence characteristics, Hess and Orphanides (1995)—which considers the relationship between war and economic performance—do not; indeed, the lynch pin of their model is uncertainty about these characteristics.

The existing literature, therefore, while skirting around this paper’s central questions, does not address them. No work so far has addressed the incentive for pandering that sequential primaries introduce in the presence of uncertainty and private information about candidates’ valence characteristics; and no work has investigated whether, in the presence of this incentive, voters are better off under a sequential or simultaneous primary system.

3 The Model

Voters 1 and 2 must select candidate A or candidate B by a sequential vote. Voter 1 votes first, with his selection observed by voter 2. Voter 2 votes as soon as voter 1’s choice is known, and the candidate for whom he votes wins the election.

Voter 1 gets utility \( U_{1A} \) from candidate A and utility \( U_{1B} \) from candidate B. He votes for candidate A if

\[
U_{1A} > U_{1B}
\]

and candidate B otherwise. The utility \( U_{1j}, j \in \{A, B\} \), contains a systematic component, \( V_{1j} \), and a stochastic component, \( \varepsilon_{1j} \). Specifically,

\[
U_{1j} = V_{1j} + \varepsilon_{1j}
\]

The term \( V_{1j} \) is equal to

\[
V_{1j} = \beta_1 - (\rho_1 - \rho_j)^2
\]

if candidate \( j \) is competent and

\[
V_{1j} = -(\rho_1 - \rho_j)^2
\]

otherwise. The parameter \( \rho_1 \in (0, \infty) \) is voter 1’s ideal policy position and \( \rho_j \in (-\infty, \infty) \) is candidate \( j \)’s platform. The parameter \( \beta_1 \in [0, \infty) \) measures the relative importance of policy and competence. When it is equal to zero, voter 1 cares only about policy; when it approaches infinity, voter 1 cares only about competence. Note that I assume that competence is considered desirable by voter 1 (and, later, voter 2); interpret it therefore as a characteristic or set of characteristics that all voters value (or, at worst, are indifferent too).

I have also assumed that the first voter observes the candidates enough to dispel any uncertainty about which candidate is competent. The assumption that voter 1 is fully informed while voter 2 has no information is a simplification, of course, but the thrust

---

8I shall assume throughout that candidates are female and voters male.

9The probability that \( U_{1A} = U_{1B} \) is zero because of subsequent random utility assumptions, so I shall ignore it.

10The specific functional form that I use for \( V_{ij} \) is not critical to the results. I adopt it because of its prevalence in the literature. It is critical however that \( V_{ij} \) be concave in \( \rho_j \).

11Competence is relative in this model so only one candidate can be competent.
of this assumption is reasonable: voters in Iowa and New Hampshire have extraordinary access to the candidates and can better evaluate their competence.

While voter 1 knows which candidate is competent, neither voter 2 nor the candidates knows (the candidates do not know which candidate is competent because competency is a judgment of the voters). I assume that voter 2 and the candidates share a prior belief that

\[ P_r\{j\} = P_r\{k\} = 0.5 \quad \text{(5)} \]

Turning to the stochastic part of \( U_1 \), let \( F(\cdot) \) be the cdf for \( \varepsilon_1 \), where \( \varepsilon_1 \equiv \varepsilon_{1k} - \varepsilon_{1j} \). I assume that \( \varepsilon_1 \) has a zero mean and a finite variance, that \( F(\cdot) \) is twice-differentiable and continuous, and that \( f(x) > f(y) \) if and only if \( |x| < |y| \). This last assumption implies that \( f(\cdot) \) has a symmetric distribution and that it has the entire real line as its support; it boils down to assuming that the probability of an electoral shock is decreasing in its absolute value. The political interpretation is that the relationship between electoral advantage and probability of victory displays diminishing returns.

The stochastic part of voter 1’s utility has three possible interpretations. First, one may view it as uncertainty about turnout. Second, one may view it as idiosyncratic tastes that are not common knowledge. For example, voter 1 may favor candidates with agrarian backgrounds; if voter 2 and the candidates are unaware of this, voter 1’s voting behavior will be stochastic from their perspective. Third, one may view it as uncertainty about the effectiveness of advertising: both candidates advertise, but only the first voter knows the relative effectiveness of their advertising campaigns.

The cdf and pdf of \( \varepsilon_1 \) allow me to calculate probabilities of victory in the first primary (from the perspective of voter 2 and the candidates). Given valuations \( V_{1j} \) and \( V_{1k} \), the probability that candidate \( j \) defeats candidate \( k \) in primary 1 is

\[ F(V_{1j} - V_{1k}) = \int_{-\infty}^{V_{1j} - V_{1k}} f(\varepsilon) \, d\varepsilon \quad \text{(6)} \]

The quantity \( V_{1j} - V_{1k} \) is candidate \( j \)'s electoral strength in the first primary.\(^{13}\) The probability that candidate \( j \) wins the first primary is

\[ P_r\{W_{1j}\} = P_r\{W_{1j} \mid j\} P_r\{j\} + P_r\{W_{1j} \mid k\} P_r\{k\} \quad \text{(7)} \]

While the second primary is decisive, the first primary matters to the candidates and voter 2 because it transmits information about the probability that each candidate is competent. The signal sent by voter 1 is noisy (since it has a stochastic component) but informative, so long as \( \beta_1 > 0 \). Once the result of the first primary is known, voter 2 can use Bayes’ rule to calculate the probability that each of the candidates is competent. Define \( \lambda_{jw} \) to be the probability that candidate \( j \) is competent given that she has won the first primary and \( \lambda_{jl} \) to be the probability that candidate \( j \) is competent given that she has lost the first primary. That is,

\[ \lambda_{jw} = \frac{P_r\{j \mid W_{1j}\}}{P_r\{W_{1j}\}} = \frac{P_r\{W_{1j} \mid j\} P_r\{j\}}{P_r\{W_{1j}\}} \quad \text{(8)} \]

\(^{12}\)I am assuming that the candidates are in symmetric positions. I relax this assumption in section 6.

\(^{13}\)If candidate \( j \)'s electoral strength is positive, I call it her electoral advantage; if it is negative, I call it her electoral disadvantage.
and
\[ \lambda_{jl} = Pr\{j \mid W_{1k}\} = \frac{Pr\{W_{1k} \mid j\}Pr\{j\}}{Pr\{W_{1k}\}} \]  \hspace{1cm} (9)

I can define analogously the variables \( \lambda_{kw} \) and \( \lambda_{kl} \) to describe the probabilities that candidate \( k \) is competent conditional on the result of the first primary. Voter 1 will receive the signal pair \((\lambda_{jw}, \lambda_{kl})\) if candidate \( j \) wins the first primary and the signal pair \((\lambda_{jl}, \lambda_{kw})\) otherwise. If candidate \( j \) wins, candidate \( k \) must lose (and vice versa), and one of them must be competent, so
\[ \lambda_{jw} + \lambda_{kl} = 1 \]  \hspace{1cm} (10)

and
\[ \lambda_{jl} + \lambda_{kw} = 1 \]  \hspace{1cm} (11)

Notice that I have implicitly assumed in the discussion above that voter 2 bases his deductions about competence on the identity of the winner in primary 1 and not on her margin of victory. This assumption reflects the difficulty of drawing inferences from the margin of victory as compared to the result of the election and, due to the nature of media coverage of primaries, the cost of learning the margin of victory as compared to the cost of learning the identity of the winner.

Once the result of the first primary is known, the second voter can calculate \( V_{2j} \) and \( V_{2k} \). If candidate \( j \) wins the first primary
\[ V_{2jw} = \lambda_{jw}^2 \rho_2 - (\rho_2 - \rho_j)^2 \]  \hspace{1cm} (12)
while if she loses,
\[ V_{2jl} = \lambda_{jl}^2 \rho_2 - (\rho_2 - \rho_j)^2 \]  \hspace{1cm} (13)

I assume without loss of generality that \( \rho_2 = 0 \).

Voter 2, like voter 1, votes for the candidate who gives him the highest expected utility. His utility from candidate \( j \) is
\[ U_{2j} = V_{2j} + \varepsilon_{2j} \]  \hspace{1cm} (14)

where \( \varepsilon_2 \) (with \( \varepsilon_2 = \varepsilon_{2k} - \varepsilon_{2j} \)) has the same distribution as \( \varepsilon_1 \). Of course, the value of \( U_{2j} \) is a function of the result of the first primary, which means that the probability that candidate \( j \) wins the second primary is a function of the result of the first primary. In particular, the probability that candidate \( j \) wins the second primary given that she has won the first is
\[ Pr\{W_{2j} \mid W_{1j}\} = F(V_{2jw} - V_{2kl}) = \int_{-\infty}^{V_{2jw} - V_{2kl}} f(\varepsilon) \, d\varepsilon \]  \hspace{1cm} (15)

The quantity \( V_{2jw} - V_{2kl} \) is candidate \( j \)'s electoral strength when she wins the first primary; her electoral strength when she loses the first primary is \( V_{2jl} - V_{2kw} \).

Note in passing that if both candidates adopt the same position, the probability \( Pr\{W_{2j} \mid W_{1j}\} \) quantifies the impact of winning the first primary. When its value is equal to one-half, winning the first primary has no effect on the nomination process. When its value is equal to one, the first primary is decisive.
The probability that candidate \( j \) wins the second primary given that her opponent has won the first primary is

\[
Pr\{W_{2j} \mid W_{1k}\} = F(V_{2j} - V_{2kw}) = \int_{-\infty}^{V_{2j} - V_{2kw}} f(\varepsilon) \, d\varepsilon
\]  

(16)

The probability that candidate \( j \) wins the second primary (her strategic object of interest) is

\[
Pr\{W_{2j}\} = Pr\{W_{2j} \mid W_{1j}\}Pr\{W_{1j}\} + Pr\{W_{2j} \mid L_{1j}\}Pr\{L_{1j}\}
\]  

(17)

Comparing equation 17 to equation 7, notice that the probability that candidate \( j \) is competent has been replaced by the probability that she wins the first primary. Once the first primary has been completed, competent and incompetent candidates are treated the same, conditional on having achieved the same result in primary 1.

4 Pure Strategy Equilibrium

Each candidate chooses her policy position to maximize her probability of winning the second primary.\(^{14}\) That is, holding her opponent’s position \( \rho_k \) fixed, candidate \( j \) solves

\[
\max_{\rho_j} Pr\{W_{2j}\} = Pr\{W_{1j}\}Pr\{W_{2j} \mid W_{1j}\} + Pr\{W_{1k}\}Pr\{W_{2j} \mid W_{1k}\}
\]  

(18)

Differentiating equation 18 with respect to \( \rho_j \), rearranging, and setting the result equal to zero yields

\[
\frac{\partial Pr\{W_{1j}\}}{\partial \rho_j} (Pr\{W_{2j} \mid W_{1j}\} - Pr\{W_{2j} \mid W_{1k}\}) = \\
- \frac{\partial Pr\{W_{2j} \mid W_{1j}\}}{\partial \rho_j} Pr\{W_{1j}\} - \frac{\partial Pr\{W_{2j} \mid W_{1k}\}}{\partial \rho_j} Pr\{W_{1k}\}
\]  

(19)

The left-hand side of equation 19 is the marginal benefit to candidate \( j \) of pandering to the first voter: she increases her probability of winning the first primary, an increase that is advantageous whenever \( Pr\{W_{2j} \mid W_{1j}\} \) is larger than \( Pr\{W_{2j} \mid W_{1k}\} \). The right-hand side is the marginal cost of pandering to the first voter: the probability of winning the second primary conditional on the result of the first falls (i.e., \( Pr\{W_{2j} \mid W_{1j}\} \) and \( Pr\{W_{2j} \mid L_{1j}\} \) decrease). More succinctly, the left-hand side means that candidate \( j \) is more likely to send the more favorable signal to voter 2, while the right-hand side means that both signals are less favorable.

In political terms pandering is beneficial to a candidate because it increases her chance of victory in the first primary, which increases the chance that she will be seen to be competent. Pandering is costly because it requires a candidate to move away from the preferred policy position of voters in later primaries and because it makes a victory in the first primary less impressive and a defeat more damaging. This second negative effect is critical: when a candidate panders she wins the first primary more often, but later voters ascribe less importance to a victory (and more significance to a

\(^{14}\)I shall consider only pure strategies in what follows.
defeat) because of her policy advantage. A candidate must be careful to avoid a Pyrrhic victory in which she wins the first primary but doesn’t impress the second voter.

To ease exposition, it is helpful to rewrite equation 18 as

$$\max_{\rho_j} Pr\{W_{2j}\} = Pr\{W_{1j}\} \delta + Pr\{W_{2j} \mid L_{1j}\}$$

(20)

where

$$\delta = Pr\{W_{2j} \mid W_{1j}\} - Pr\{W_{2j} \mid W_{1k}\}$$

(21)

I shall now state a brace of lemmas that, while not of enormous interest per se, allows me to establish this section’s key results.

**Lemma 1.** The probability that candidate j wins the second primary is the same whenever the two candidates adopt the same position. That is,

$$Pr\{W_{2j}\} = \theta, \forall \rho = \rho_j = \rho_k$$

(22)

**Proof.** See appendix A.

**Lemma 2.** There is a unique shared policy position $\rho^* = \rho_j = \rho_k$ at which $\frac{\partial W_{2j}}{\partial \rho_j} = \frac{\partial W_{2k}}{\partial \rho_k} = 0$. Further, $\rho^*$ is a local maximum.

**Proof.** See appendix A.

Taken together, lemmas 1 and 2 allow me to establish the following uniqueness result.

**Theorem 1.** If there is pure strategy equilibrium in the primary game, then it is symmetric with $\rho_j = \rho_k = \rho^*$.

**Proof.** Lemma 2 establishes that there is a unique symmetric policy $\rho^*$ at which $\frac{\partial W_{2j}}{\partial \rho_j} = \frac{\partial W_{2k}}{\partial \rho_k} = 0$. To see that this rules out all other potential equilibria, consider an equilibrium $\rho_j \neq \rho_k, \rho_k \neq \rho^*$. If this is to be an equilibrium, candidate j’s probability of winning the second primary must equal $\theta$ or adopting her opponent’s position would be a profitable deviation. This means that candidate j can adopt the position $\rho_k$ without changing her probability of winning the second primary. But lemma 2 says that $\frac{\partial W_{2j}}{\partial \rho_j} \neq 0$ if $\rho_j = \rho_k \neq \rho^*$, so candidate j can improve her probability of winning the second primary by adopting the position $\rho_k + \epsilon$, where $\epsilon$ is arbitrarily small. Thus, the strategies $\rho_j \neq \rho_k, \rho_k \neq \rho^*$ cannot constitute an equilibrium.

The same argument rules out any equilibrium in which $\rho_j \neq \rho_k, \rho_j \neq \rho^*$. Since lemma 2 rules out any equilibrium in which $\rho_j = \rho_k \neq \rho^*$, the only possible equilibrium is $\rho^* = \rho_j = \rho_k$.

Before turning to existence and examining when $\rho^*$ is an equilibrium strategy, it is worth performing some simple comparative statics.

**Theorem 2.**

1. $\rho^* \in [\rho_2, \rho_1]$, with $\rho^* = \rho_2 \iff \beta_1 = 0$ or $\beta_2 = 0$.

2. $\rho^* \to \rho_2$ as $\beta_1 \to \infty$.

3. $\rho^* \to \rho_1$ as $\beta_2 \to \infty$. 

11
Proof. See appendix A.

The first result in theorem 2 states that if both candidates care about competence, then $\rho^*$ is strictly between the preferred positions of the two voters. That is, if both voters care about competence and an equilibrium exists, then there is pandering in that equilibrium. To get the intuition behind this result, consider the case in which the value of $\beta_2$ is equal to zero. In this case, $\rho^*$ will be equal to $\rho_2$.

Now consider the three effects on $\rho^*$ of increasing the value of $\beta_2$. First, the distance between $Pr\{W_{2j} \mid W_{1j}\}$ and $Pr\{W_{2j} \mid W_{1k}\}$ increases because competence (and hence the first primary) has a larger impact on the second primary. This effect increases the value of pandering because it increases the payoff to winning the first primary. Second, increasing the value of $\beta_2$ increases the relative importance of the posterior beliefs $\lambda_{jw}$ and $\lambda_{jl}$ in the second voter’s utility function. This increases the cost of pandering because pandering reduces the values of $\lambda_{jw}$ and $\lambda_{jl}$, a reduction that now has a larger impact on candidate $j$’s electoral strength. Finally, increasing the value of $\beta_2$ increases the value of $Pr\{W_{2j} \mid W_{1j}\}$ and decreases the value of $Pr\{W_{2j} \mid W_{1k}\}$. This reduces the strength of the relationship between candidate $j$’s electoral strength and her probability of winning the second primary; and this decreases the cost of pandering, because the reduction in candidate $j$’s electoral strength has less impact on her probability of winning the second primary. The first and second effects would cancel out but for the third effect, which blunts the second. Hence, the benefit of pandering exceeds its cost and the value of $\rho^*$ moves towards $\rho_1$—there is pandering in equilibrium.

In political terms, candidate $j$’s electoral advantage could be interpreted as her expected margin of victory. In this formulation, increasing the value of $\beta_2$ has three effects. First, it increases the influence of the first primary, which increases the benefit of pandering. Second, by increasing the importance of competence to the second voter, it increases the cost of having one’s victory in the first primary reduced in importance because of pandering. This second effect is essentially a reduction in candidate $j$’s expected margin of victory in the second primary. As such, its seriousness is reduced by the third effect, which is a reduction in the impact of margin of victory on probability of victory due to the increase in the influence of competence. Since candidates care about victory and not about their margin of victory, this third effect ensures that there is pandering in equilibrium so long as voter 2 cares about competence.

The second result of theorem 2 establishes that there is very little incentive to panderm when the first voter cares overwhelmingly about competence. This is because pandering to a first voter who values competence above all will not influence his behavior but will make one’s opponent more attractive to voter 2.

This result obscures the interesting fact that the relationship between $\rho^*$ and $\beta_1$ is not monotone. When $\beta_1$ increases, there are two opposing effects: the effect of pandering on voter 1 falls because his relative interest in policy decreases, but the value of such pandering rises because the signal from the first primary becomes increasingly informative. Eventually the former effect dominates and $\rho^*$ approaches $\rho_2$, but there are regions where the latter effect is dominant and increasing $\beta_1$ moves $\rho^*$ towards $\rho_1$.

The complex relationship between $\rho^*$ and $\beta_1$ brings into sharp focus the dual role of $\beta_1$. From voter 1’s perspective, $\beta_1$ describes the extent to which he values competence. From voter 2’s perspective, $\beta_1$ describes how reliable a judge voter 1 is of the candidates’
competence. Perversely, the more voter 1 cares about competence vis-a-vis policy—which reduces the incentive to pander—the more weight voter 2 places on his judgment, which increases the incentive to pander!

Turning to the existence of equilibrium, the critical clause of theorem 1 reads “if a pure strategy equilibrium exists.” While \( \rho^* \) is the unique candidate equilibrium strategy, it is not always an equilibrium. An example will help illustrate why \( \rho^* \) need not be an equilibrium.

Figure 1 plots candidate \( j \)'s probability of winning the second primary, her probability of winning the second primary given that she loses the first primary and the probability that she is competent given that she loses the first primary against \( \rho^*_j \), holding candidate \( k \)'s position at \( \rho^* \). In this example, the value of \( \rho^* \) is approximately equal to one, which is the preferred policy position of the first voter.

![Figure 1: A Breakdown in Equilibrium](image)

As figure 1 shows, \( \rho^* \) is not an equilibrium: candidate \( j \) can improve her probability of winning the second primary by moving her position to \( \rho_2 \), voter 2’s preferred position. This deviation has two effects, both of which favor candidate \( j \). First, by moving her policy position to \( \rho_2 \), candidate \( j \) makes her policy position far more attractive to voter 2 than candidate \( k \)'s policy position. Second, by moving her policy position far from \( \rho_1 \) relative to candidate \( k \)'s position, candidate \( j \) makes the probability that candidate \( k \) wins the first primary high, regardless of which candidate is competent. As a result, a victory in the first primary says little about candidate \( k \)'s competence (at \( \rho_j = 0 \), \( \lambda_{jl} \approx 0.4 \)), and candidate \( j \) is likely to win the second primary (due to her more desirable policy position) even if she loses the first primary (\( Pr\{W_{2j} \mid W_{1k}\} \approx 0.75 \)). By adopting the second voter’s preferred policy position, candidate \( j \) makes the second primary about policy—since candidate \( k \) is likely to win the first primary regardless of which candidate is competent—while positioning herself to win such a contest.\(^{15}\)

\(^{15}\)Note in passing that candidate \( j \)'s best response to \( \rho^* \) is a policy position to the left of \( \rho_2 \). Adopting

13
What about the values of $\beta_1$ and $\beta_2$ has caused equilibrium to break down? The large value of $\beta_2$ creates a strong incentive to pander (because voter 2 values highly competence), pushing $\rho^*$ close to $\rho_1$. However, because $\beta_1$ is relatively small in this example, a deviation from $\rho^*$ to $\rho_2$ makes the signal from the first primary uninformative (since candidate $k$ wins regardless of which candidate is competent thanks to her policy advantage). Thus, while voter 2 values highly information about the candidates’ competence, candidate $j$ can make the signal from the first primary uninformative while positioning herself to win the second primary on policy grounds.

The example above shows that the candidate equilibrium will break down when there exists a deviation that turns both primaries into policy contests. Viewed in another light, when there is too much pandering equilibrium will break down because a candidate with “integrity” can win more than half the time.

Given that equilibrium need not exist, it is necessary to establish the conditions under which it is sure to exist.

**Theorem 3.** For all $\beta_2$, there exists a $\beta_1$ such that $\rho^*$ is the unique pure strategy equilibrium in the primary game. Further, $\rho^*$ is also the unique pure strategy equilibrium for all $\beta_1 > \beta_1$.

**Proof.** See appendix A. 

Theorem 3 establishes that, under certain conditions, the primary game has an equilibrium, that this equilibrium is unique, and that in it the candidates adopt a common policy $\rho^*$. The critical condition is that $\beta_1$ be sufficiently large, that the first voter care sufficiently about competence. When this condition is met, the signal from the first primary is sufficiently informative that no profitable deviation from $\rho^*$ exists.

5 Welfare

In this section I shall discuss the welfare implications of a first primary when a pure strategy equilibrium exists. I have shown that in such an equilibrium, politicians pander to the first voter by moving towards his preferred policy position whenever both voters care about competence. This pandering is disadvantageous to the second voter, but there is an offsetting effect that benefits the second voter: the competent candidate wins the election more frequently. A priori, it is not clear which of these effects is larger.

One way to evaluate the welfare effect of the first primary is to contrast the outcome under a sequential primary system with the outcome under a simultaneous primary system. Under a simultaneous primary system, the candidates will ignore the first primary and adopt the preferred policy position of the second voter. When the candidates adopt the same position and the second voter has no information about which is competent, the competent candidate will win half the time.

Given this, it is clear that the first voter is strictly better off under the sequential primary system: candidates move towards his preferred position and the competent candidate wins more often. I can therefore restrict my attention to the welfare of the positional deviation that a position so far from voter 1's preferred position raises the probability that candidate $k$ is competent given that she loses the first primary above one-half.
second voter. The expected payoff to the second voter under a simultaneous primary system is

\[
EP_{si} = \frac{1}{2}(\beta_2 - (\rho_2 - \rho_2)^2) - \frac{1}{2}(\rho_2 - \rho_2)^2 = \frac{1}{2}\beta_2 \tag{23}
\]

Evaluating welfare under a sequential primary system is complicated by the fact that equilibrium need not exist. I shall proceed therefore by examining welfare under the policy \(\rho^*\) (which always exists) and address the existence of equilibrium later. The expected payoff to the second voter under the sequential primary system is

\[
EP_{se} = Pr\{W_{2c}\}(\beta_2 - (\rho_2 - \rho^*)^2) - (1 - Pr\{W_{2c}\})(\rho_2 - \rho^*)^2 \tag{24}
\]

where \(Pr\{W_{2c}\}\) is the probability that the competent candidate wins the second primary. Now, it can be shown that

\[
(\rho_2 - \rho^*)^2 = \eta^2(\rho_1 - \rho_2)^2 \tag{25}
\]

where \(\eta \in [0, 1]\) is the degree of pandering, with a value of 0 representing no pandering and a value of 1 representing complete pandering.\(^\text{16}\) I can therefore rewrite equation 24 as

\[
EP_{se} = Pr\{W_{2c}\}\beta_2 - \eta^2(\rho_1 - \rho_2)^2 \tag{26}
\]

It is helpful at this stage to define \(\omega\) as the difference between voter 2’s expected welfare under the sequential primary system and his expected welfare under the simultaneous primary system. In particular,

\[
\omega = EP_{se} - EP_{si} = (Pr\{W_{2c}\} - \frac{1}{2})\beta_2 - (\rho_1 - \rho_2)^2\eta^2 \tag{27}
\]

In what follows I shall say that voter 2 enjoys an expected welfare gain when \(\omega > 0\), and suffers an expected welfare loss when \(\omega < 0\).

**Theorem 4.** If \(\beta_2 > 0\), as \(\beta_1 \to \infty\), \(\omega\) approaches a positive constant.

**Proof.** From theorem 2 I know that as \(\beta_1 \to \infty\), \(\rho^* \to \rho_2\). The term \((\rho_1 - \rho_2)^2\gamma^2\) therefore approaches zero. Now, if both candidates adopt \(\rho^*\)

\[
Pr\{W_{2c}\} = F(\beta_1)F([2\lambda_{jw} - 1]\beta_2) - F(-\beta_1)F([2\lambda_{jl} - 1]\beta_2) \tag{28}
\]

As \(\beta_1 \to \infty\), this quantity approaches \(F(\beta_2)\). Thus,

\[
\omega \to (F(\beta_2) - \frac{1}{2})\beta_2 \tag{29}
\]

\(^{16}\)A formal definition of \(\eta\) is found in appendix A.
Theorem 4 establishes that if $\beta_1$ is sufficiently large, then voter 2 will be better off under a sequential primary system. I know from theorem 3 that if $\beta_1$ is sufficiently large, then $\rho^*$ is an equilibrium. Thus, there exists a value of $\beta_1$ such that equilibrium exists and a sequential primary system is welfare improving. It is interesting that the critical condition is that voter 1 care sufficiently about competence. This condition has two effects, both of which make voter 2 better off: voter 1’s signal is highly informative (i.e., the competent candidate wins more often) and there is little incentive to pander.

It is important to note that the condition in theorem 4 is not redundant: there exist equilibria in which $\omega$ is negative. From a practical perspective, theorem 4 establishes that a desirable property for an early primary state is that its voters be interested in competence.

Another interesting aspect of theorem 4 is that increasing the value of $\beta_1$ can only increase player 2’s expected welfare gain by so much. Voter 2’s expected welfare gain is limited as the value of $\beta_1$ rises because while the signal from the first primary becomes increasingly informative, the value of the information that it contains is fixed. In the limit, the first primary is perfectly informative, but the value of that information is constrained by the value of $\beta_2$.

Finally, the proof of theorem 4 shows that the probability that the competent candidate wins the second primary is not a function of either voter’s preferred position. This establishes the following corollary.

**Corollary 1.** There is a negative relationship between $\omega$ and the absolute distance between $\rho_1$ and $\rho_2$.

Corollary 1 is a natural result: the closer the policy positions of the primary voters, the higher their welfare under a sequential primary system. Another desirable property for early primary state then is that its voters hold moderate political views.

### 6 Extensions

In this section I extend my basic model in several ways. First, I examine the model’s predictions when voter 2 and the candidates share a common prior belief that one candidate is more likely to be competent than the other. My second extension is to consider the case in which one of the candidates is favored for some exogenous reason by voters in the first state (i.e., she is the favorite daughter of the state). I find that in both extensions pandering need not exist, but that as the value of $\beta_1$ becomes arbitrarily large both models converge to the basic model.

#### 6.1 Prior Equality

I have assumed thus far that voter 2 and the two candidates share a prior belief that the candidates are equally likely to be competent (henceforth I shall call this assumption the prior equality assumption). This is appropriate when there is no clear favorite in the nomination race (as, for example, in the 2004 Democratic primary elections), but it is not suitable when one candidate is in a dominant position (as in the 2000 Democratic primary election). In this section I restrict voter 2 and the two candidates to having a common prior belief about the competence of candidate $j$, but I allow this common
prior to take any value strictly between zero and one. In particular, I assume that the common prior is

$$Pr\{j\} = \xi$$  \tag{30}

with $\xi \in (0,1)$. Since there are only two candidates, this means that

$$Pr\{k\} = 1 - \xi$$  \tag{31}

When the prior equality assumption is relaxed, there is still only one possible equilibrium, and this equilibrium exists so long as $\beta_1$ is sufficiently large and $\beta_2$ is not equal to zero.

**Theorem 5.** If there is a pure strategy equilibrium in the primary game when the prior equality assumption is relaxed, then it is symmetric with $\rho_j = \rho_k = \rho^*$. Further, if $\beta_1$ is sufficiently large, then an equilibrium exists.

*Proof.* See appendix A. \hfill $\square$

**Corollary 2.** $Pr\{W_2j\} > 0.5$ whenever $\xi \in (0.5,1)$ and $Pr\{W_2j\} < 0.5$ whenever $\xi \in (0,0.5)$.

While theorem 5 is identical to my previous results, relaxing the prior equality assumption changes the model’s predictions in an important way: there now exist equilibria in which either there is no pandering or there is reverse pandering (i.e., $\rho^*_\xi$ is negative). Figure 6.1 depicts the values of $\rho^*_\xi$ as a function of $\xi$. All of the values of $\rho^*_\xi$ that are depicted are equilibria.

![Equilibrium Policy Positions as a Function of $\xi$](image)

**Figure 2: Equilibrium Policy Positions as a Function of $\xi$**

There are two distinct regions in figure 6.1. In the first, as $\xi$ moves away from one-half, the degree of pandering in equilibrium initially falls because the impact of winning the first primary on voter 2’s posterior belief decreases. Eventually the disadvantaged
candidate’s probability of being competent given she wins the first primary becomes so low that she must reverse-pander to make a victory in the first primary a meaningful statement about her competence. In the second region, as \( \xi \) approaches zero or one, there is little reward to competing to win the first primary—since the advantaged candidate is believed to be competent regardless of the outcome—and both candidates’ policy positions will approach \( \rho_2 \).

While relaxing the prior equality assumption does not affect the uniqueness of equilibrium nor the requirement for existence, it can profoundly affect the equilibrium position. If one candidate has a strong advantage in public perception, there may be no pandering—or even reverse pandering—in equilibrium. This prediction matches well with one’s intuition. In races with several viable candidates the reward to winning an early primary is high (e.g., John Kerry in 2004), which increases the incentive to pander. In races with a clear favorite, early primary results can change the tenor of the race, but more often they are a minor impediment to the favorite (e.g., George H. Bush in 1992). This dampens the incentive to pander and may even motivate iconoclastic behavior.

While the prior equality assumption guarantees the existence of pandering, it is less important than my two critical assumptions because pandering can exist without it. Further, as theorem 6 shows, as the importance of competence to the first voter becomes large, the extended model converges to the basic model.

**Theorem 6.** As \( \beta_1 \) becomes arbitrarily large, \( \rho^\xi \) converges to \( \rho^* \).

*Proof.* See appendix A. \( \Box \)

### 6.2 A Favorite Daughter

It is common for politicians from Iowa, New Hampshire, or surrounding states to enter the nomination fight. These favorite daughter candidates hope to parlay their popularity with voters in an early primary into a nomination. In this section I examine the effect of being the favorite daughter of voters in the first primary. In particular, I assume that the valence utility that candidate \( j \) gives to voters in the first primary is \( \beta_1 + \tau, \tau > 0 \), if she is competent and \( \tau - \beta_1 \) if she is not. The valence utility that candidate \( k \) gives to voters in the first primary is \( \beta - \tau \) if she is competent and \( -((\beta_1 + \tau) \) if she is not.

**Theorem 7.** If there is pure strategy equilibrium in the primary game with a favorite daughter, then it is symmetric with \( \rho_j = \rho_k = \rho^*_e \). Further, if \( \beta_1 \) is sufficiently large, then an equilibrium exists.

*Proof.* See appendix A. \( \Box \)

As in the basic model, equilibrium is unique if it exists, and it exists when the value of \( \beta_1 \) is sufficiently large. The next question then is whether being a favorite daughter is an advantage or a disadvantage. When the candidates adopt the same policy position in equilibrium, being a favorite daughter has a benefit and a cost. The benefit is an increase in the probability of winning the first primary. The cost is that winning the first primary becomes so low that she must reverse-pander to make a victory in the first primary a meaningful statement about her competence. In the second region, as \( \xi \) approaches zero or one, there is little reward to competing to win the first primary—since the advantaged candidate is believed to be competent regardless of the outcome—and both candidates’ policy positions will approach \( \rho_2 \).

While relaxing the prior equality assumption does not affect the uniqueness of equilibrium nor the requirement for existence, it can profoundly affect the equilibrium position. If one candidate has a strong advantage in public perception, there may be no pandering—or even reverse pandering—in equilibrium. This prediction matches well with one’s intuition. In races with several viable candidates the reward to winning an early primary is high (e.g., John Kerry in 2004), which increases the incentive to pander. In races with a clear favorite, early primary results can change the tenor of the race, but more often they are a minor impediment to the favorite (e.g., George H. Bush in 1992). This dampens the incentive to pander and may even motivate iconoclastic behavior.

While the prior equality assumption guarantees the existence of pandering, it is less important than my two critical assumptions because pandering can exist without it. Further, as theorem 6 shows, as the importance of competence to the first voter becomes large, the extended model converges to the basic model.

**Theorem 6.** As \( \beta_1 \) becomes arbitrarily large, \( \rho^\xi \) converges to \( \rho^* \).

*Proof.* See appendix A. \( \Box \)

### 6.2 A Favorite Daughter

It is common for politicians from Iowa, New Hampshire, or surrounding states to enter the nomination fight. These favorite daughter candidates hope to parlay their popularity with voters in an early primary into a nomination. In this section I examine the effect of being the favorite daughter of voters in the first primary. In particular, I assume that the valence utility that candidate \( j \) gives to voters in the first primary is \( \beta_1 + \tau, \tau > 0 \), if she is competent and \( \tau - \beta_1 \) if she is not. The valence utility that candidate \( k \) gives to voters in the first primary is \( \beta - \tau \) if she is competent and \( -((\beta_1 + \tau) \) if she is not.

**Theorem 7.** If there is pure strategy equilibrium in the primary game with a favorite daughter, then it is symmetric with \( \rho_j = \rho_k = \rho^*_e \). Further, if \( \beta_1 \) is sufficiently large, then an equilibrium exists.

*Proof.* See appendix A. \( \Box \)

As in the basic model, equilibrium is unique if it exists, and it exists when the value of \( \beta_1 \) is sufficiently large. The next question then is whether being a favorite daughter is an advantage or a disadvantage. When the candidates adopt the same policy position in equilibrium, being a favorite daughter has a benefit and a cost. The benefit is an increase in the probability of winning the first primary. The cost is that winning the first
primary is a less informative signal of one’s competence. In this sense, in equilibrium, being a favorite daughter is akin to pandering except that one does not incur a policy cost.

**Theorem 8.** If one of the candidates is a favorite daughter, then her probability of winning the second primary is larger than her opponents in equilibrium whenever voters in both primaries care about competence (i.e., \( \beta_1 > 0 \) and \( \beta_2 > 0 \)).

*Proof.* See appendix A.

While being a favorite daughter is always an advantage, the relationship between \( \Pr \{ W_{2j} \} \) and \( \tau \) is not monotone, as figure 6.2 shows. Increasing \( \tau \) initially increases candidate \( j \)'s probability of winning the second primary because it increases her chance of victory in the first primary. Eventually, however, her advantage in the first primary becomes so overwhelming that the first primary becomes uninformative, and her probability of winning the second primary approaches one-half from above. In other words, the model predicts a large advantage for candidates with some ties to New Hampshire or Iowa and a small advantage for candidates who are strongly identified with either state.

This finding coincides well with the experience of favorite son candidates. Candidates who are strongly associated with Iowa or New Hampshire (e.g., Tom Harkin or Dick Gephardt) have found their victories heavily discounted, while candidates for whom Iowa and New Hampshire have a milder affinity (e.g., John Kerry, Michael Dukakis or Walter Mondale) have used victories to propel themselves to the nomination.

![Figure 3: \( \Pr \{ W_{2j} \} \) as a Function of \( \tau \)](image)

Finally, introducing a favorite daughter can change the model’s equilibrium position significantly, but there is no systematic relationship between \( \tau \) and \( \rho^* \). However, it is possible to derive two limit results.
Theorem 9.  

1. As \( \tau \) becomes arbitrarily large, \( \rho^*_1 \) converges to \( \rho_2 \).

2. As \( \beta_1 \) becomes arbitrarily large, \( \rho^*_2 \) converges to \( \rho^* \).

Proof. See appendix A. \( \square \)

When \( \tau \) is arbitrarily large, the first primary is not informative because it is won by the first daughter regardless of which candidate is competent. As a result, the second voter ignores the first primary and candidates compete only on policy. And, since the second voter is decisive, both candidates will adopt a position arbitrarily close to \( \rho_2 \). In essence, a strong favorite daughter makes the transfer of information impossible, and the sequential primary system collapses to a simultaneous primary system. When the value of \( \beta_1 \) is arbitrarily large, \( \tau \) becomes unimportant, the first primary is informative, and the favorite daughter model approaches the basic model.

7 Conclusion

I have analyzed the effect of a sequential primary system using a simple two voter, two candidate model. I have found that if the voter in the first primary has private information and values competence sufficiently highly relative to policy, then a unique pure strategy symmetric equilibrium exists. In this equilibrium both candidates pander to the policy preferences of the voter in the first primary. However, while the amount of pandering in equilibrium is always positive, it becomes negligible when the voter in first primary cares sufficiently about competence. The critical assumptions that drive this result are that voters in the first primary have private information about competence and that voters draw inferences using the identity of the winner of the first primary but not her margin of victory. I find also that although there is always pandering in equilibrium when competence is sufficiently important to the first voter, for intermediate cases, there may be no pandering—or even reverse-pandering—if the candidates are in some way unequal a priori.

I find also that a sequential primary system increases voter 1’s expected welfare relative to a simultaneous primary system. The welfare of voter 2 may increase or decrease, but if voter 1 cares sufficiently about competence, it will increase. Combined with my first result, this suggests that a reasonable criterion for selecting a state to hold a first primary is that its electorate value competence highly relative to policy. Such a state would introduce little pandering and increase the probability of electing the competent candidate, making all voters better off.

Finally, note that while my model was motivated by a sequential primary system, it applies to any sequential election. Further, it generalizes naturally to any situation in which one entity tries to learn another’s private information by observing its voting behavior. Natural examples include a case in which a search committee vets candidates for a job and a case in which a parliamentary committee votes on a piece of legislation before the full parliament votes on it.
A Proofs of Selected Theorems and Lemmas

A.1 Proof of Lemma 1

I will proceed by showing that none of the terms in equation 20 are functions of $\hat{\beta}$. The probability that candidate $j$ wins the first primary is

$$Pr\{W_{1j}\} = Pr\{W_{1j} \mid j\}Pr\{j\} + Pr\{W_{1j} \mid k\}Pr\{k\}$$  \hspace{1cm} (32)

The probability that candidate $j$ wins the first primary given that she is competent is

$$Pr\{W_{1j} \mid j\} = \int_{-\infty}^{V_{1j} - V_{1k}} f(x) \, dx$$  \hspace{1cm} (33)

Since $\hat{\beta} = \beta_k$, this reduces to

$$Pr\{W_{1j} \mid j\} = \int_{-\infty}^{-\beta_1} f(x) \, dx = F(-\beta_1)$$  \hspace{1cm} (34)

which is not a function of $\hat{\beta}$. The probability that candidate $j$ wins the first primary given that candidate $k$ is competent reduces to

$$Pr\{W_{1j} \mid j\} = \int_{-\infty}^{-\beta_1} f(x) \, dx = F(-\beta_1)$$  \hspace{1cm} (35)

which is also not a function of $\hat{\beta}$. Thus, $Pr\{W_{1j}\}$ is not a function of $\hat{\beta}$. To see that $\delta$ is not a function of $\hat{\beta}$, note that when $\rho_j = \rho_k$,

$$Pr\{W_{2j} \mid W_{1j}\} = \int_{-\infty}^{V_{2j} - V_{2k}} f(x) \, dx = F([2\lambda_{jw} - 1]2)$$  \hspace{1cm} (36)

and

$$Pr\{W_{2j} \mid W_{1k}\} = \int_{-\infty}^{V_{2j} - V_{2k}} f(x) \, dx = F([1 - 2\lambda_{jw}]2)$$  \hspace{1cm} (37)

These expressions are not functions of $\hat{\beta}$ so long as $\lambda_{jw}$ is not a function of $\hat{\beta}$. Now,

$$\lambda_{jw} = \frac{Pr\{W_{1j} \mid j\}Pr\{j\}}{Pr\{W_{1j} \mid j\}Pr\{j\} + Pr\{W_{1j} \mid k\}Pr\{k\}}$$  \hspace{1cm} (38)

As $\rho_j = \rho_k$, I can use equations 34 and 35 to reduce this to

$$\lambda_{jw} = \frac{Pr\{j\}F(\beta_1)}{Pr\{j\}F(\beta_1) + Pr\{j\}F(-\beta_1)}$$  \hspace{1cm} (39)

which establishes that lambda is not a function of $\hat{\beta}$. \hfill $\Box$

A.2 Proof of Lemma 2

To save space, I shall use the prime symbol instead of writing out derivatives explicitly; all derivatives are with respect to $\rho_j$ unless otherwise stated. Differentiating equation 36 with respect to $\rho_j$ gives

$$Pr'\{W_{2j} \mid W_{1j}\} = 2(\rho_2 - \rho_j + \beta_2\lambda_{jw}) f([2\lambda_j - 1]2)$$  \hspace{1cm} (40)

Differentiating equation 37 with respect to $\rho_j$ gives

$$Pr'\{W_{2j} \mid W_{1k}\} = 2(\rho_2 - \rho_j + \beta_2\lambda_{jl}) f([2\lambda_{jl} - 1]2)$$  \hspace{1cm} (41)
Exploiting the symmetry of $\varepsilon_1$ and $\varepsilon_2$’s distribution and the fact that $\rho_j = \rho_k$, this can be rewritten as

$$Pr\{W_{2j} \mid W_{1k}\} = 2(\rho_2 - \rho_j + \beta_1 \lambda_{jw}^\prime) f([2\lambda_{jw} - 1]\beta_2)$$

(42)

or

$$Pr\{W_{2j} \mid W_{1k}\} = 2f([2\lambda_{jw} - 1]\beta_2)(\rho_2 - \rho_j) + 2\beta_2 f([2\lambda_{jw} - 1]\beta_2) \lambda_{jw}^\prime$$

(43)

Using the fact that

$$\lambda_{jw}^\prime = -2(\rho_1 - \rho_j)f(\beta_1) F(-\beta_1, \beta_1)$$

(44)

this can be written as

$$Pr\{W_{2j} \mid W_{1k}\} = 2f([2\lambda_{jw} - 1]\beta_2)(\rho_2 - \rho_j) - 4\beta_2 f([2\lambda_{jw} - 1]\beta_2) f(\beta_1) F(-\beta_1, \beta_1)(\rho_1 - \rho_j)$$

(45)

Introducing $\alpha_2$ and $\alpha_3$, this can be written as

$$Pr\{W_{2j} \mid W_{1k}\} = 2\alpha_2(\rho_2 - \rho_j) - 2\alpha_3(\rho_1 - \rho_j)$$

(46)

It follows immediately from equations 40 and 42 that when $\rho_j = \rho_k$,

$$\delta' = 0$$

(47)

Thus, differentiating $Pr\{W_{2j}\}$ with respect to $\rho_j$ gives

$$Pr'\{W_{2j}\} = Pr'\{W_{1j}\} \delta + Pr'\{W_{2j} \mid L_{1j}\}$$

(48)

Differentiating equation 34 with respect to $\rho_j$ and then using the fact that $\rho_j = \rho_k$ gives

$$Pr\{W_{1j} \mid j\} = 2(\rho_1 - \rho_j) f(\beta_1)$$

(49)

Doing the same with equation 35 gives

$$Pr\{W_{1j} \mid k\} = 2(\rho_1 - \rho_j) f(-\beta_1)$$

(50)

Combining these two results and again using the symmetry of $\varepsilon_1$ and $\varepsilon_2$’s distribution gives

$$Pr\{W_{1j}\} = 2(\rho_1 - \rho_j) f(\beta_1)[Pr\{j\} + Pr\{k\}] = 2(\rho_1 - \rho_j) f(\beta_1)$$

(51)

If $\alpha_1$ is defined as $f(\beta_1) \delta$, then

$$Pr'\{W_{2j}\} = 2\alpha_1(\rho_1 - \rho_j) + 2\alpha_2(\rho_2 - \rho_j) - 2\alpha_3(\rho_1 - \rho_j)$$

(52)

Setting this equal to zero and solving for $\rho_j$ gives

$$\rho^* = \rho_2 + \frac{\alpha_1 - \alpha_3}{\alpha_1 + \alpha_2 - \alpha_3}(\rho_1 - \rho_2)$$

(53)

where

$$\alpha_1 = f(\beta_1) F([1 - 2\lambda_{jw}]\beta_2, [2\lambda_{jw} - 1]\beta_2)$$

$$\alpha_2 = f([2\lambda_{jw} - 1]\beta_2)$$

$$\alpha_3 = 2\beta_2 f(\beta_1) f([2\lambda_{jw} - 1]\beta_2) F(-\beta_1, \beta_1)$$

(54)

Since none of the alphas are functions of $\rho_j$, this value is unique.

To see that $\rho^*$ is a maximum, note first that the second-order condition when both candidates adopt $\rho^*$ is

$$Pr''\{W_{2j}\} = Pr''\{W_{1j}\} \delta + \frac{1}{2} (Pr''\{W_{2j} \mid W_{1j}\} + Pr''\{W_{2j} \mid W_{1k}\})$$

(55)
since $\delta' = 0$ and $Pr\{W_{1j}\} = \frac{1}{2}$. Now, if $\rho_j = \rho_k$,
\[
Pr''(W_{1j})\delta = 2(\rho_1 - \rho^*)^2(f'(\beta_1) + f'(-\beta_1)) - (f(\beta_1) + f(-\beta_1))\delta
\] (56)
The first term cancels and the second reduces to
\[
Pr''(W_{1j})\delta = -2f(\beta_1)F([1 - 2\lambda_{jw}]\beta_2, [2\lambda_{jw} - 1]\beta_2) = -2\alpha_1
\] (57)
Further, if $\rho_j = \rho_k$,
\[
Pr''(W_{2j} | W_{1j}) = 4(\rho_2 - \rho^* + \beta_2\lambda_{jw})^2f'(2\lambda_{jw} - 1)\beta_2 - 2f(2\lambda_{jw} - 1)\beta_2(\alpha_3''\beta_2 - 1)
\] (58)
while
\[
Pr''(W_{2j} | W_{1k}) = 4(\rho_2 - \rho^* + \beta_2\lambda_{jw})^2f'(2\lambda_{jw} - 1)\beta_2 - 2f(2\lambda_{jw} - 1)\beta_2(\alpha_3''\beta_2 - 1)
\] (59)
When $\rho_j = \rho_k$, the first terms in each expression cancel one another. Adding the two remaining terms after multiplying each by one-half and simplifying gives
\[
\frac{1}{2}(Pr''(W_{2j} | W_{1k}) + Pr''(W_{2j} | W_{1k})) = f(2\lambda_{jw} - 1)\beta_2(\lambda_{jw}''\beta_2 + \lambda_{jw}''\beta_2 - 2) = 2(\alpha_3 - \alpha_2)
\] (60)
The second order condition is thus
\[
Pr''(W_{2j}) = 2(\alpha_3 - \alpha_2 - \alpha_1) < 0
\] (61)
where the inequality follows from a result in the proof of theorem 2 which shows that $(\alpha_1 + \alpha_2 - \alpha_3) > 0$.

**A.3 Proof of Theorem 2**

1. From the proof of lemma 2, I know that
\[
\rho^* = \rho_2 + \frac{\alpha_1 - \alpha_3}{\alpha_1 + \alpha_2 - \alpha_3}(\rho_1 - \rho_2) = \rho_2 + \eta(\rho_1 - \rho_2)
\] (62)
where
\[
\alpha_1 = f(\beta_1)F([1 - 2\lambda_{jw}]\beta_2, [2\lambda_{jw} - 1]\beta_2)
\alpha_2 = f(2\lambda_{jw} - 1)\beta_2
\alpha_3 = 2\beta_2 f(\beta_1) f(2\lambda_{jw} - 1)\beta_2 F(-\beta_1, \beta_1)
\] (63)
Note first that $\alpha_2 > 0$. Now, if $\beta_2 = 0$ then $\alpha_1 = \alpha_3 = 0$ so $\rho^* = \rho_2$. On the other hand, if $\beta_1 = 0$, then $\alpha_3 = 0$ and $\alpha_1 = 0$ (since $\lambda_{jw} = 0.5$) so $\rho^* = \rho_2$. Thus, $\rho^* = \rho_2$ if $\beta_1 = 0$ or $\beta_2 = 0$.
I shall now show that $\alpha_1 - \alpha_3 > 0$ whenever $\beta_1 > 0$ and $\beta_2 > 0$. If this is true, then $\rho^* \in (\rho_2, \rho_1)$ whenever $\beta_1 > 0$ and $\beta_2 > 0$ since $\alpha_2 > 0$.
I begin by picking an arbitrary value of $\beta_1 > 0$ and letting $\beta_2 = 0$. For these values, $\alpha_1 - \alpha_3 = 0$. The first derivative of $\alpha_1$ with respect to $\beta_2$ is
\[
f(\beta_1)2(2\lambda_{jw} - 1) f([2\lambda_{jw} - 1]\beta_2) = 2f(\beta_1)(2F(\beta_1) - 1)f([2\lambda_{jw} - 1]\beta_2) = 2f(\beta_1) f([2\lambda_{jw} - 1]\beta_2) F(-\beta_1, \beta_1)
\] (64)
The proof is organized as follows.

2. From the proof of lemma 2, I know that

$$\rho^* = \rho_2 + \frac{\alpha_1 - \alpha_3}{\alpha_1 + \alpha_2 - \alpha_3} (\rho_1 - \rho_2)$$

where

$$\alpha_1 = f(\beta_1) F([1 - 2\lambda_{yw}] \beta_2, [2\lambda_{yw} - 1] \beta_2)$$

$$\alpha_2 = f([2\lambda_{yw} - 1] \beta_2)$$

$$\alpha_3 = 2\beta_2 f(\beta_1) f([2\lambda_{yw} - 1] \beta_2) F(-\beta_1, \beta_1)$$

Now, as $$\beta_1 \to \infty$$, both $$\alpha_1$$ and $$\alpha_3$$ approach zero (since $$f(\beta_1) \to 0$$), while $$\alpha_2$$ approaches $$f(\beta_2)$$. Thus, $$\frac{\alpha_1 - \alpha_3}{\alpha_1 + \alpha_2 - \alpha_3} \to 0$$ and $$\rho^* \to \rho_2$$.

3. As $$\beta_2$$ approaches infinity $$f([2\lambda_{yw} - 1] \beta_2) \to 0$$, so $$\alpha_2 \to 0$$. Since the first derivative of $$(\alpha_1 - \alpha_3)$$ with respect to $$\beta_2$$ is positive $$\frac{\alpha_1 - \alpha_3}{\alpha_1 + \alpha_2 - \alpha_3} \to 1$$ and $$\rho^* \to \rho_1$$.

A.4 Proof of Theorem 3

To save space, I shall not write out derivatives explicitly, relying instead on the prime and double-prime symbols. All derivatives are with respect to $$\rho_j$$. I assume throughout that $$\rho_k = \rho^*$$. The proof is organized as follows.

1. Fix $$\beta_2$$ and then increase $$\beta_1$$ until $$(\rho^* - \rho_2) \to 0$$ (that a sufficiently large value of $$\beta_1$$ exists is guaranteed by theorem 2).

2. It can then be shown that $$Pr\{W_{2j}\}$$ is strictly concave for the range $$[\rho_2, \rho^*]$$, which establishes that $$\rho^*$$ is a maximum for this range.

3. Consider a deviation by candidate $$j$$ from $$\rho^*$$ to $$\rho_j < \rho_2$$. It is possible to show that the first derivative of $$Pr\{W_{2j}\}$$ with respect to $$\rho_j$$ is positive for all such values, ruling them out as profitable deviations.

4. Similar logic shows that deviations by candidate $$j$$ in which $$\rho_j > \rho^*$$ are not profitable.

5. It follows that $$\rho^*$$ is a global maximum.

1. Proved above.

2. Begin by noting that the second order condition is

$$Pr''\{W_{2j}\} = Pr''\{W_{1j}\} \Delta + Pr\{W_{1j}\} Pr''\{W_{2j} \mid W_{1j}\}$$

$$+ (1 - Pr\{W_{1j}\}) Pr''\{W_{2j} \mid W_{1k}\}$$

$$+ Pr'\{W_{1j}\} (Pr'\{W_{2j} \mid W_{1j}\} - Pr'\{W_{2j} \mid W_{1k}\})$$

I turn my attention first to $$Pr\{W_{1j}\}$$ which is shown below.

$$Pr\{W_{1j}\} = Pr\{W_{1j} \mid j\} Pr\{j\} + Pr\{W_{1j} \mid k\} Pr\{k\}$$
Define

$$\alpha_{1jc} = \beta_1 - (\rho_1 - \rho_j)^2 + (\rho_1 - \rho_k)^2$$  \hspace{1cm} (71)$$

As $\beta_1 \to \infty$, $\alpha_{1jc} \to \infty$ for $\rho_j \in [\rho_2, \rho^*]$. Thus,

$$Pr\{W_{ij} \mid j\} = \int_{-\infty}^{\alpha_{1jc}} f(\varepsilon) \, d\varepsilon$$  \hspace{1cm} (72)$$

approaches 1 as $\alpha_{1jc} \to \infty$ and

$$Pr'\{W_{ij} \mid j\} = 2(\rho_1 - \rho_j)f(\alpha_{1jc})$$  \hspace{1cm} (73)$$

and

$$Pr''\{W_{ij} \mid j\} = (2(\rho_1 - \rho_j)^2f'(\alpha_{1jc}) - 2f(\alpha_{1jc})$$  \hspace{1cm} (74)$$

approach zero.

Now define

$$\alpha_{1ji} = -\beta_1 - (\rho_1 - \rho_j)^2 + (\rho_1 - \rho_k)^2$$  \hspace{1cm} (75)$$

This term approaches $-\infty$ as $\beta_1 \to \infty$ for $\rho_j \in [\rho_2, \rho^*]$. Thus,

$$Pr\{W_{ij} \mid k\} = \int_{-\infty}^{\alpha_{1ji}} f(\varepsilon) \, d\varepsilon$$  \hspace{1cm} (76)$$

approaches zero and

$$Pr'\{W_{ij} \mid k\} = 2(\rho_1 - \rho_j)f(\alpha_{1ji})$$  \hspace{1cm} (77)$$

and

$$Pr''\{W_{ij} \mid k\} = (2(\rho_1 - \rho_j)^2f'(\alpha_{1ji}) - 2f(\alpha_{1ji})$$  \hspace{1cm} (78)$$

approach zero. This establishes the results about $Pr\{W_{ij}\}$.

Now,

$$\lambda_{jw} = Pr\{j \mid W_{ij}\} = \frac{Pr\{j\}Pr\{W_{ij} \mid j\}}{Pr\{W_{ij}\}}$$  \hspace{1cm} (79)$$

Define

$$M = Pr\{j\}Pr\{k\}(Pr'\{W_{ij} \mid j\}Pr\{W_{ij} \mid k\} - Pr'\{W_{ij} \mid k\}Pr\{W_{ij} \mid j\})$$  \hspace{1cm} (80)$$

and

$$M' = Pr\{j\}Pr\{k\}(Pr''\{W_{ij} \mid j\}Pr\{W_{ij} \mid k\} - Pr''\{W_{ij} \mid k\}Pr\{W_{ij} \mid j\})$$  \hspace{1cm} (81)$$

Then

$$\lambda_{jw}' = \frac{M}{Pr\{W_{ij}\}^2}$$  \hspace{1cm} (82)$$

The term in brackets in $M$ goes to zero as $\beta_1 \to \infty$ for $\rho_j \in [\rho_2, \rho^*]$, so the numerator of $\lambda_{jw}' \to 0$ while the denominator approaches one-fourth. Thus, $\lambda_{jw}' \to 0$. Also,

$$\lambda_{jw}'' = \frac{M'Pr\{W_{ij}\}^2 - 2Pr\{W_{ij}\}Pr'\{W_{ij}\}M}{Pr\{W_{ij}\}^4}$$  \hspace{1cm} (83)$$

The term in brackets in $M'$ goes to zero as $\beta_1 \to \infty$ for $\rho_j \in [\rho_2, \rho^*]$, so the numerator of $\lambda_{jw}'' \to 0$ while the denominator approaches one-sixteenth. Thus, $\lambda_{jw}'' \to 0$. 

25
Now,
\[ \lambda_{jl} = \text{Pr}(j \mid W_{1k}) = \frac{\text{Pr}(j) \text{Pr}(W_{1k} \mid j)}{\text{Pr}(W_{1k})} \]  
(84)

Define
\[ N = \text{Pr}(j) \text{Pr}(k) (\text{Pr}'(W_{1k} \mid j) \text{Pr}(W_{1k} \mid k) - \text{Pr}'(W_{1k} \mid k) \text{Pr}(W_{1k} \mid j)) \]  
(85)

and
\[ N' = \text{Pr}(j) \text{Pr}(k) (\text{Pr}''(W_{1k} \mid j) \text{Pr}(W_{1k} \mid k) - \text{Pr}''(W_{1k} \mid k) \text{Pr}(W_{1k} \mid j)) \]  
(86)

Then
\[ \lambda_{jl}' = \frac{N}{\text{Pr}(W_{1k})^2} \]  
(87)

The term in brackets in \( N \) goes to zero as \( \beta_1 \to \infty \) for \( \rho_j \in [\rho_2, \rho^*] \), so the numerator of \( \lambda_{jl}' \to 0 \) while the denominator approaches one-fourth. Thus, \( \lambda_{jl}' \to 0 \). Also,
\[ \lambda_{jl}'' = \frac{N' \text{Pr}(W_{1k})^2 - 2 \text{Pr}(W_{1k}) \text{Pr}'(W_{1k}) N}{\text{Pr}(W_{1k})^4} \]  
(88)

The term in brackets in \( N' \) goes to zero as \( \beta_1 \to \infty \) for \( \rho_j \in [\rho_2, \rho^*] \), so the numerator of \( \lambda_{jl}'' \to 0 \) while the denominator approaches one-sixteenth. Thus, \( \lambda_{jl}'' \to 0 \).

Now define
\[ \alpha_{2ju} = (2\lambda_{ju} - 1)\beta_2 - (\rho_2 - \rho_j)^2 + (\rho_2 - \rho_k)^2 \]  
(89)

As \( \lambda_{ju} \to 1 \), \( \alpha_{2ju} \to \beta_2 \) for \( \rho_j \in [\rho_2, \rho^*] \). Now,
\[ \text{Pr}(W_{2j} \mid W_{1j}) = \int_{-\infty}^{\alpha_{2ju}} f(\varepsilon) \, d\varepsilon \]  
(90)

So
\[ \text{Pr}'(W_{2j} \mid W_{1j}) = 2(\rho_2 - \rho_j + \beta_2 \lambda_{ju}')f(\alpha_{2ju}) \]  
(91)

Since \( \rho_2 - \rho_j \to 0 \) and \( \lambda_{ju}' \to 0 \), the first bracketed term approaches zero, making the whole expression equal to zero. The second derivative is
\[ \text{Pr}''(W_{2j} \mid W_{1j}) = (2(\rho_2 - \rho_j + \beta_2 \lambda_{ju}')^2 f'(\alpha_{2ju}) \]
\[ + 2f(\alpha_{2ju})(\beta_2 \lambda_{ju}' - 1) \]  
(92)

The first term equals zero for the reason given above. The second term approaches \(-2f(\beta_2)\), since \( \lambda_{ju}'' \to 0 \). This establishes results about \( \text{Pr}(W_{2j} \mid W_{1j}) \).

Now define
\[ \alpha_{2ji} = (2\lambda_{ji} - 1)\beta_2 - (\rho_2 - \rho_j)^2 + (\rho_2 - \rho_k)^2 \]  
(93)

As \( \lambda_{ji} \to 0 \), \( \alpha_{2ji} \to -\beta_2 \) for \( \rho_j \in [\rho_2, \rho^*] \). Define
\[ \text{Pr}(W_{2j} \mid W_{1k}) = \int_{-\infty}^{\alpha_{2ji}} f(\varepsilon) \, d\varepsilon \]  
(94)

Then
\[ \text{Pr}'(W_{2j} \mid W_{1k}) = 2(\rho_2 - \rho_j + \beta_2 \lambda_{ji}')f(\alpha_{2ji}) \]  
(95)
A.5 Proof of Theorem 5

The proof of this theorem is substantively identical to the proofs of theorems 1 and 3 and is omitted in the interest of brevity. It is available upon request from the author. I do however characterize the value of $\rho_2'$. It is given by

$$
\rho_2' = \rho_2 + \frac{\alpha_1 - \alpha_3}{\alpha_1 + \alpha_2 - \alpha_3} (\rho_1 - \rho_2)
$$

(101)
where
\[
\alpha_1 = f(\beta_1)F([2\lambda_{jl} - 1]\beta_2, [2\lambda_{jw} - 1]\beta_2) \\
\alpha_2 = f([2\lambda_{jw} - 1]\beta_2)Pr\{W_{1j}\} + f([2\lambda_{jl} - 1]\beta_2)Pr\{W_{1k}\} \\
\alpha_3 = 2\beta_2(1 - \xi)f(\beta_1)F(-\beta_1, \beta_1) \left( \frac{f([2\lambda_{jw} - 1]\beta_2)}{Pr\{W_{1j}\}} + \frac{f([2\lambda_{jl} - 1]\beta_2)}{Pr\{W_{1k}\}} \right)
\]

and
\[
Pr\{W_{1j}\} = \xi F(\beta_1) + (1 - \xi)F(-\beta_1) \\
\lambda_{jw} = \frac{\xi F(\beta_1)}{Pr\{W_{1j}\}} \\
\lambda_{jl} = \frac{\xi F(-\beta_1)}{Pr\{W_{1k}\}}
\]

### A.6 Proof of Theorem 6

As \( \beta_1 \) becomes arbitrarily large, \( \lambda_{jl} \) approaches zero and \( \lambda_{jw} \) approaches one, so \( \alpha_1 \) approaches \( f(\beta_1)F(-\beta_2, \beta_2) \). Also, \( Pr\{W_{1j}\} \) approaches \( \xi \) and \( Pr\{W_{1k}\} \) approaches \( 1 - \xi \), so \( \alpha_2 \) approaches \( f(\beta_2) \) and \( \alpha_3 \) approaches \( 2\beta_2f(\beta_1)f(\beta_2)F(-\beta_1, \beta_1) \). Namely, \( \rho^*_x \to \rho^* \).

### A.7 Proof of Theorem 7

The proof of this theorem is substantively identical to the proofs of theorems 1 and 3 and is omitted in the interest of brevity. It is available upon request from the author. I do however characterize the value of \( \rho^*_x \). It is given by
\[
\rho^*_x = \rho_2 + \frac{\alpha_1 - \alpha_3}{\alpha_1 + \alpha_2 - \alpha_3}(\rho_1 - \rho_2)
\]

where
\[
\alpha_1 = (f(\beta_1 + a) + f(-\beta_1 + a))F([2\lambda_{jl} - 1]\beta_2, [2\lambda_{jw} - 1]\beta_2) \\
\alpha_2 = 2f([2\lambda_{jw} - 1]\beta_2)Pr\{W_{1j}\} + 2f([2\lambda_{jl} - 1]\beta_2)Pr\{W_{1k}\} \\
\alpha_3 = \frac{\beta_2}{2} \left( \frac{f([2\lambda_{jw} - 1]\beta_2)(f(\beta_1 + a)F(-\beta_1 + a) - f(-\beta_1 + a)F(\beta_1 + a))}{Pr\{W_{1j}\}} + \frac{f([2\lambda_{jl} - 1]\beta_2)(f(\beta_1 - a)F(-\beta_1 - a) - f(-\beta_1 - a)F(\beta_1 - a))}{Pr\{W_{1k}\}} \right)
\]

and
\[
Pr\{W_{1j}\} = \frac{1}{2}F(\beta_1 + a) + \frac{1}{2}F(-\beta_1 + a) \\
\lambda_{jw} = \frac{\frac{1}{2}F(\beta_1 + a)}{Pr\{W_{1j}\}} \\
\lambda_{jl} = \frac{\frac{1}{2}F(-\beta_1 + a)}{Pr\{W_{1k}\}}
\]

### A.8 Proof of Theorem 8

Pick any positive values for \( \beta_1 \) and \( \rho_1 \) and let \( \beta_2 \) equal zero. Both candidates will adopt the position \( \rho_2 \) and will win the second primary half of the time. The probability that candidate \( j \)
wins the second primary when both candidates adopt the same position is
\[
Pr\{W_{2j}\} = Pr\{W_{1j}\}Pr\{W_{2j} \mid W_{1j}\} + Pr\{W_{1k}\}Pr\{W_{2j} \mid W_{1k}\}
\]
\[= \frac{1}{2}(F(\beta_1 + a) + F(-\beta_1 + a))F([2\lambda_{jw} - 1]\beta_2)
\]
\[+ \frac{1}{2}(F(\beta_1 - a) + F(-\beta_1 - a))F([2\lambda_{jl} - 1]\beta_2)
\]
(107)

Note that this quantity does not depend on the shared policy position and that in equilibrium that candidates always adopt the same policy position. Thus, while changing

\[\frac{\partial W_{2j}}{\partial \beta_2} = Pr\{W_{1j}\}[2\lambda_{jw} - 1]f([2\lambda_{jw} - 1]\beta_2) + Pr\{W_{1k}\}[2\lambda_{jl} - 1]f([2\lambda_{jl} - 1]\beta_2)
\]
(108)

Now,
\[2\lambda_{w} - 1 = \frac{F(\beta_1 + a)}{Pr\{W_{1j}\}} - \frac{\frac{1}{2}F(\beta_1 + a) + \frac{1}{2}F(-\beta_1 + a)}{Pr\{W_{1j}\}} = \frac{\frac{1}{2}F(\beta_1 + a) - \frac{1}{2}F(-\beta_1 + a)}{Pr\{W_{1j}\}}
\]
(109)

while
\[2\lambda_{jl} - 1 = \frac{F(-\beta_1 - a)}{Pr\{W_{1k}\}} - \frac{\frac{1}{2}F(-\beta_1 - a) + \frac{1}{2}F(\beta_1 - a)}{Pr\{W_{1k}\}} = -\frac{\frac{1}{2}F(-\beta_1 - a) - \frac{1}{2}F(\beta_1 - a)}{Pr\{W_{1k}\}}
\]
(110)

Thus,
\[Pr\{W_{1j}\}(2\lambda_{jw} - 1) = -Pr\{W_{1k}\}(2\lambda_{jl} - 1)
\]
(111)

and
\[\frac{\partial W_{2j}}{\partial \beta_2} > 0
\]
(112)

provided that
\[f([2\lambda_{jw} - 1]\beta_2) > f([2\lambda_{jl} - 1]\beta_2)
\]
(113)

which is the case so long as \(\beta_2\) is not equal to zero.\(^{18}\) Thus, whenever \(\beta_1\) and \(\beta_2\) are greater than zero, candidate \(j\)—the favorite daughter—wins more than half the time.

A.9 Proof of Theorem 9

1. As \(\tau \to \infty\), \(Pr\{W_{1j}\} \to 1\) and \(\lambda_{jw} \to \frac{1}{2}\). Thus, \(Pr\{W_{2j}\} \to Pr\{W_{2j} \mid W_{1j}\}\). Since \(\lambda_{jw} \to \frac{1}{2}\) and \(\rho_j = \rho_k\), \(Pr\{W_{2j} \mid W_{1j}\} \to \frac{1}{2}\).

2. As \(\beta_1\) becomes arbitrarily large, \(\lambda_{jl}\) approaches zero and \(\lambda_{jw}\) approaches one, so \(\alpha_1\) approaches \(f(\beta_1)F(-\beta_2, \beta_2)\). Also, \(Pr\{W_{1j}\}\) approaches \(\frac{1}{2}\) and \(Pr\{W_{1k}\}\) approaches \(\frac{1}{2}\), so \(\alpha_2\) approaches \(f(\beta_2)\) and \(\alpha_3\) approaches 2\(\beta_2f(\beta_1)F(\beta_2)F(-\beta_1, \beta_1)\). Namely, \(\rho^*_e \to \rho^*\).

\(^{18}\)Making candidate \(j\) the favorite daughter moves \(\lambda_{jw}\) towards one-half and \(\lambda_{jl}\) away from it. This moves \([2\lambda_{jw} - 1]\beta_2\) towards zero and \([2\lambda_{jl} - 1]\beta_2\) away from it and given the assumption on the distribution of the error, establishes the claim.
References


