INTERNATIONAL ECONOMIC REVIEW
Vol. 56, No. 3, August 2015

SEQUENTIAL OR SIMULTANEOUS ELECTIONS? A WELFARE ANALYSIS*

BY PATRICK HUMMEL AND BRIAN KNIGHT[1]

Google Inc., U.S.A.; Brown University, U.S.A.

Should all voters vote on the same day or should elections be staggered? Using a model of voting and social learning, we illustrate that sequential elections place too much weight on early states but also provide late voters with valuable information. Simultaneous elections equally weigh states but place too much weight on voter priors, providing an inappropriate advantage to front-runners. Simultaneous elections are thus preferred if the front-runner advantage is small, but sequential elections are preferred if the advantage is large. Our quantitative welfare analysis of presidential primaries suggests that simultaneous systems slightly outperform sequential systems.

"... for years concerns have been raised regarding the calendar that some believe gives a disproportionate influence to these two early states."
— David Price, Commission on Presidential Nomination Timing and Scheduling, October 1, 2005

"We need to preserve the possibility for lesser known, lesser funded candidates to compete, and a national primary on February 5th will not do that."
— Terry Shumaker, Commission on Presidential Nomination Timing and Scheduling, December 5, 2005

1. INTRODUCTION

Although voting occurs simultaneously in many electoral and legislative settings, there are also settings in which voting occurs sequentially. Under a roll call ballot, for example, votes are recorded one-by-one with participants observing the votes of those preceding them before casting their own ballots. In the electoral context, presidential general elections were held on different days in different states prior to 1872. A related issue involves the release of voting returns on election day, especially in countries with multiple time zones.

This distinction between simultaneous and sequential systems is particularly salient in the design of presidential primary systems, which have traditionally followed a calendar in which Iowa and New Hampshire vote first, followed by a group of states on the first Tuesday in February and another group on the first Tuesday in March. This is followed by several months of further elections, with the process often continuing into early summer.

Given concerns associated with the current system, several alternatives have been proposed. At the extreme, advocates of a true national primary, in which every state would vote on the same date, point toward a more efficient and fair system. Hybrid systems, which move toward a simultaneous system but retain some features of the current sequential system, include the rotating regional primary system, under which Iowa and New Hampshire would vote first, followed by four weekly rounds of regional primaries, with the order of the regions rotating from election to election.

*Manuscript received July 2013; revised January 2014.

[1]We thank Nageeb Ali, Gerard Padro i Miquel, conference participants at the Princeton Conference on Political Economy, the Center for the Study of American Politics Summer Conference, and the NBER Summer Institute, and seminar participants at the University of Pennsylvania, the University of Miami, Miami University, the University of Michigan, and the New Economic School. Please address correspondence to: Brian Knight, 64 Waterman Street, Providence, RI 02912. E-mail: Brian_Knight@brown.edu.

Debates over the choice between traditional sequential calendars and these alternative, more compressed calendars typically focus on trading off the relative advantages of the two systems. In particular, opponents of sequential systems argue that early states have disproportionate influence, whereas supporters argue that it enhances competition since dark horse candidates can better emerge from the field of candidates. Under simultaneous elections, by contrast, states would have equal influence but dark horse candidates may not be provided with sufficient opportunity to compete. Although these factors have dominated the debate, there has been little formal analysis of this trade-off, and there have also been no attempts to weigh the relative importance of these advantages and disadvantages of the two systems.

In this article, we use the positive model of voting and social learning developed in Knight and Schiff (2010) in order to conduct a normative analysis of this trade-off. In the model, voters are uncertain over candidate quality but have some private information. Under sequential elections, voters in late states attempt to infer the information of voters in early states from voting returns. Using this model, we compare both simultaneous and sequential elections to a public information benchmark, under which all voters observe all relevant signals. We show theoretically that neither system is optimal and that there is indeed a trade-off between voters equally weighing preferences and information under simultaneous systems and late voters being better informed under sequential elections. We then develop welfare expressions based upon aggregate voter utility and show that simultaneous elections tend to dominate when the advantage of the front-runner is small. When this advantage is large, by contrast, sequential election systems tend to dominate as they provide greater opportunities for dark horse candidates of unexpectedly high quality to emerge from the field. Finally, we conduct an empirical welfare analysis based upon data from presidential primaries, and the estimates suggest that simultaneous election systems outperform sequential election systems.

The article proceeds as follows: We first discuss the related literature and then review the positive theoretical model of voting and social learning. Using this model, we provide a comparison of sequential and simultaneous systems and show that either system might be preferred from a welfare perspective. Finally, we conduct a numerical welfare analysis that examines a variety of electoral institutions, with a focus on comparing sequential and simultaneous systems.

2. LITERATURE REVIEW

This article is at the intersection of four literatures: social learning, theoretical analyses of sequential voting systems, empirical analyses of sequential voting systems in the context of presidential primaries, and optimal electoral institutions. We discuss each, in turn, below.

2.1. *Social Learning.* The literature on social learning began with Welch (1992), Bikhchandani et al. (1992), and Banerjee (1992). In these models, agents take actions in a predetermined sequence, individual payoffs depend only upon individual actions, and late movers have an opportunity to observe the actions of early movers. If actions are discrete and payoffs are sufficiently correlated, a herd may form in which agents ignore their private information and simply follow the actions of those earlier in the sequence. Note that despite the fact that information may be lost in this process, simultaneous choice never dominates a sequential order from a welfare perspective. This follows from the fact that individual payoffs depend only upon individual actions, and thus agents moving in a sequence that would rationally ignore the behavior of early agents were doing so in their best interests. In the voting context, by contrast, individual payoffs depend upon the actions of all agents. Thus, if it is optimal to do so, agents will rationally ignore the behavior of early agents.

2.2. *Theoretical Analyses of Sequential Voting.* Several papers have examined this issue of social learning in the electoral context with a focus on binary elections. In a model with two candidates and strategic voters, Dekel and Piccione (2000) show that every equilibrium of the simultaneous game is an equilibrium of the sequential game. This follows from the fact that voters condition on being pivotal and hence behave as if exactly half of the other voters favor one option over the other. Thus, the identity of the early voters is irrelevant, and voters do not condition on the behavior of those earlier in the sequence. The converse that every equilibrium of the sequential game is an equilibrium of the simultaneous game, however, is not necessarily true. In particular, Ali and Kartik (2012) construct equilibria in which late voters do condition on the behavior of early voters. Other theoretical analyses of sequential elections include Battaglini (2005), who focuses on voter turnout, Hummel (2012), who focuses on multicandidate elections, Morton and Williams (1999, 2001), who focus on learning about candidate ideology from early voters and conduct corresponding experimental tests, Hummel (2011), who addresses the desire to avoid a long and costly primary, Aldrich (1980) and Klumpp and Polborn (2006), who examine campaign finance in the context of sequential elections, and Strumpf (2002), who examines candidate incentives for exiting the election.

Two other theoretical papers have investigated the issue of competition when comparing simultaneous and sequential elections. In Selman (2010), there are two candidates, one of which is high quality and one of which is low quality, and voters receive private information about which candidate is of high quality. Loyal voters always vote for their preferred candidate, whereas uncommitted voters support the candidate of higher expected quality. Unlike our model, neither candidate is favored in terms of voter priors over quality. In the context of his model, Selman shows that the sequential system is preferred when loyal voters are imbalanced and the quality of information is low. Although competition also plays a role in our comparison between sequential and simultaneous elections, the mechanism is quite different. Unlike Selman (2010), candidates in our model are advantaged due to voter priors, and the advantage of the sequential system is that voters place less weight on these priors.

A second paper, Callandar (2007), compares sequential and simultaneous elections in the context of a model in which voters are strategic, in the sense of conditioning their vote on being pivotal, and also prefer to vote for winners. There are two candidates, one of which is high quality and one of which is low quality, and voters receive private information about which candidate is of high quality. In the context of this model, Callander shows that sequential elections outperform simultaneous elections when voter priors significantly favor one of the two candidates. Although this result is similar to our result regarding competition, the underlying mechanism is quite different. In particular, if voter priors significantly favor one of the two candidates, then under simultaneous voting, voters have an incentive to simply ignore their private information and vote for the candidate that is more likely to win based on their priors. However, when the preference for voting for the winner disappears, sequential and simultaneous are both efficient. Our comparison between the two systems, by contrast, is not based upon a preference for voting for the winner.

An additional contribution of our article relative to both Selman (2010) and Callander (2007) is our empirical analysis. In particular, our article represents the first attempt to quantify the trade-off between competition and equal weighting of information.

2.3. *Empirical Analyses of Sequential Voting.* Empirical analyses of presidential primary systems include Knight and Schiff (2010), who, using daily polling data from the 2004 presidential primary, document momentum effects and provide empirical support for a social learning interpretation. Note that the paper, Knight and Schiff (2010), is purely positive in nature and does not address the normative question of which system is welfare-preferred. Bartels (1987, 1988) examines polling data in 1984 and shows that candidate viability plays a key role in momentum effects. Bartels (1985) and Kenney and Rice (1994) also examine other possible empirical motivations for momentum effects using data from the 1980 and 1988 presidential primaries. Finally, there is a series of papers, including Adkins and Dowdle (2001), Steger et al. (2004), and Steger (2008), documenting that early states have a disproportionate influence in terms of selecting the winning candidate in presidential primaries. These papers are all relevant in the sense that they document important differences in electoral outcomes between simultaneous and sequential systems.

There is also a more qualitative literature that analyzes some of the consequences of using sequential or simultaneous elections. Mayer and Busch (2004) note that in the past few decades, states have been engaging in a process known as "front-loading" in which they attempt to hold their presidential primaries earlier in the election year to try to have more influence over the outcome, a process that would be unlikely to result under simultaneous elections. Redslawk et al. (2011) note that many voters support changing the current presidential primary system to a national or rotating regional system, and voters who vote in later states in the sequence tend to be more supportive of the reform than voters who vote earlier on. And Norrander (1992) notes that many party leaders who advocate switching to either a single national primary or a rotating regional primary do so because they believe that it will result in more electable candidates who are forced to address a broader audience.

Finally, in closely related work, Deltas et al. (2010) examine a model in which late voters learn about valence from the voting returns in early states. In addition to this vertical dimension, candidates are also distinguished by a horizontal dimension, and, when there are more than two candidates, their model thus introduces the potentially interesting issue of ticket-splitting. On the other hand, their model does not allow for candidates to differ in terms of the priors of voters over quality, and thus does not allow for front-runner and dark horse candidates. Thus, in their context, the advantage of sequential elections involves the ability of voters to better coordinate as the election unfolds, instead of allowing dark horse candidates of high quality to emerge from the field. After structurally estimating the model using aggregate, state-level voting returns data from the 2008 primary, they show that sequential elections tend to outperform simultaneous elections in terms of electing candidates of higher valence and being more likely to elect the Condorcet winner. Given that the underlying advantages of sequential elections are different in their model, we view our work as complementary to this article.

2.4. *Optimal Electoral Institutions*. Finally, this article is related to a broader literature on the normative analysis of electoral institutions. Hummel and Holden (2014) address the question of whether it is better to have small states vote before large states or well-informed states vote before less informed states in sequential elections, but do not analyze simultaneous elections, as we do in this article. Maskin and Tirole (2004) develop the optimal constitution in a model in which public officials can be held more or less accountable via reelection. Lizzeri and Persico (2001) compare the distribution of public goods under winner-take-all and proportional electoral systems. Coate and Knight (2007) develop the optimal districting plan for district-based legislative elections. Persson et al. (2000) and Persson and Tabellini (2004) compare presidential and parliamentary systems. And finally, Coate (2004) and Prat (2002) examine campaign finance from a voter welfare perspective.

3. BASIC MODEL

This section lays out our framework for comparing simultaneous and sequential elections. The notation follows Chamley (2004), and readers are referred to Knight and Schiff (2010) for additional details and discussion.

Consider a set of states ($s = 1, 2, \dots, S$) choosing between candidates ($c = 0, 1, \dots, C$). We allow for the possibility that multiple states may vote on the same day; in particular, let Ω_t be the set of states voting on date t and let $N_t \geq 1$ be the size of this set. This nests the case of sequential elections, where Ω_t is nonempty for multiple t, and simultaneous elections, where $N_t = 0$ if $t > 1$.

Within a state, there is a continuum of voters with unit mass. Voter i residing in state s is assumed to receive the following payoff from candidate c winning the election:

$$u_{cis} = q_c + \eta_{cs} + \nu_{cis}, \tag{1}$$

where q_c represents the quality of candidate c, η_{cs} represents a state-specific preference for candidate c, and ν_{cis} represents an individual preference for candidate c. This individual preference is assumed to be drawn from a mean-zero type-I extreme value distribution that is independent across both candidates and voters. We normalize utility from the baseline candidate to be zero for all voters ($u_{0is} = 0$).

We assume the following information structure: Voters know their own state-level preference (η_{cs}) but not those in other states. Voters do, however, know the distribution from which these state-level preferences are drawn. In particular, we assume that state-level preferences are drawn independently from a normal distribution [$\eta_{cs} \sim N(0, \sigma_\eta^2)$]. We further assume that voters are uncertain over candidate quality and are Bayesian. In particular, initial ($t = 1$) priors over candidate quality (q_c) are assumed to be normally distributed with a candidate-specific mean μ_{c1} and a variance σ_1^2 that is common across candidates. Under the assumptions to follow, the posterior distribution will be normal as well. Before going to the polls, all voters in state s receive a noisy signal (θ_{cs}) over the quality of candidate c:

$$\theta_{cs} = q_c + \varepsilon_{cs}, \tag{2}$$

where the noise in each state's signal is assumed to be drawn independently from a normal distribution [$\varepsilon_{cs} \sim N(0, \sigma_\varepsilon^2)$]. We assume that this signal is common across all voters within a state. Finally, we assume that the signal is unobserved by voters in other states.

Given the state-level signal (θ_{cs}), expected utility for voter i in state s from candidate c winning can be written as follows:

$$E(u_{cis}|\theta_{cs}, \eta_{cs}, \nu_{cis}) = E(q_c|\theta_{cs}) + \eta_{cs} + \nu_{cis}. \tag{3}$$

Finally, regarding voter behavior, we assume sincere voting. In particular, given the information available, voter i in state s at time t supports the candidate who provides the voter with the highest level of expected utility.[2]

Then, for voters in state s observing a signal over quality (θ_{cs}) and with a prior given by (μ_{ct}, σ_t^2), private updating over quality is given by

$$E(q_c|\theta_{cs}) = \alpha_t\theta_{cs} + (1 - \alpha_t)\mu_{ct}, \tag{4}$$

where the weight on the signal is given by

$$\alpha_t = \frac{\sigma_t^2}{\sigma_t^2 + \sigma_\varepsilon^2}. \tag{5}$$

Plugging Equation (4) into Equation (3), we have that

$$E(u_{cis}|\theta_{cs}, \eta_{cs}, \nu_{cis}) = \alpha_t\theta_{cs} + (1 - \alpha_t)\mu_{ct} + \eta_{cs} + \nu_{cis}. \tag{6}$$

Then, using the fact that ν_{cis} is drawn from a type-I extreme value distribution, Knight and Schiff (2010) show that we can write the vote shares for candidate c, relative to the baseline candidate 0, in state s voting at time t as follows:

$$\ln(v_{cst}/v_{0st}) = \eta_{cs} + \alpha_t\theta_{cs} + (1 - \alpha_t)\mu_{ct}. \tag{7}$$

[2] Ali and Kartik (2012) construct an equilibrium in the sequential game in which strategic voters condition on past votes by voting sincerely for the candidate that would afford them the highest expected utility given their current information, leading to momentum effects that are similar to those studied here. Although this is not a unique equilibrium, it does show that the key difference between sequential and simultaneous voting is present in a model with strategic voting.

Using the fact that $\theta_{cs} = q_c + \varepsilon_{cs}$, we can say that transformed vote shares provide a noisy signal of quality

$$\frac{\ln(v_{cst}/v_{0st}) - (1-\alpha_t)\mu_{ct}}{\alpha_t} = q_c + \frac{\eta_{cs}}{\alpha_t} + \varepsilon_{cs}, \tag{8}$$

where the noise in the voting signal includes the noise in the quality signal (ε_{cs}) but also the noise due to the unobserved state preferences (η_{cs}/α_t); the combined variance of the noise in the voting signal thus equals $(\sigma_\eta^2/\alpha_t^2) + \sigma_\varepsilon^2$. Given $N_t \geq 1$ such signals, the posterior distribution is also normal and can thus be characterized by its first two moments:

$$\mu_{ct+1} = \mu_{ct} + \frac{\beta_t/N_t}{\alpha_t} \sum_{s\in\Omega_t} [\ln(v_{cst}/v_{0st}) - \mu_{ct}], \tag{9}$$

$$\frac{1}{\sigma_{t+1}^2} = \frac{1}{\sigma_t^2} + \frac{N_t}{(\sigma_\eta^2/\alpha_t^2) + \sigma_\varepsilon^2}, \tag{10}$$

where the weight on the voting signals is given by

$$\beta_t = \frac{N_t\sigma_t^2}{N_t\sigma_t^2 + (\sigma_\eta^2/\alpha_t^2) + \sigma_\varepsilon^2}. \tag{11}$$

4. NORMATIVE ANALYSIS

Using this model, we first define voter welfare and then develop a public information benchmark under which all voters have access to all relevant signals. Focusing on a simple case of the model with two candidates and two states, we then compare electoral outcomes under this public information benchmark to those under sequential and simultaneous voting systems. Finally, we develop expressions for the welfare gain associated with moving from a sequential system to a simultaneous system, again focusing on the special case of two candidates and two states.

4.1. *Voter Welfare.* Our welfare measure is based upon average voter utility obtained under the winning candidate:

$$W = \frac{1}{S}\sum_{c=1}^{C} 1(c \text{ wins}) \sum_{s=1}^{S} \int_{i\in s} u_{cis} f(u_{cis})\,di, \tag{12}$$

where $1(c \text{ wins})$ indicates that candidate c received a plurality of votes and S is the total number of states. Since ν_{cis} is mean zero, we have that $\int_{i\in s} u_{cis} f(u_{cis})\,di = q_c + \eta_{cs}$. Substituting this in, we have that

$$W = \sum_{c=1}^{C} 1(c \text{ wins}) \left[q_c + \frac{1}{S}\sum_{s=1}^{S} \eta_{cs} \right]. \tag{13}$$

Then, for a given electoral system, we have that expected voter welfare is given by

$$E(W) = \sum_{c=1}^{C} \Pr(c \text{ wins}) E(q_c + \overline{\eta}_c | c \text{ wins}), \tag{14}$$

where $\overline{\eta}_c = \frac{1}{S}\sum_{s=1}^{S} \eta_{sc}$ measures the average state-level preference for candidate c.

4.2. *Public Information Benchmark.* As a welfare benchmark, we next consider electoral outcomes in the case in which voters have all of the relevant signals regarding candidate quality. That is, under this counterfactual system, voters in each state have access to the full set of signals and update over candidate c as follows:

$$E(q_c|\theta_{c1},\theta_{c2},\ldots,\theta_{cS}) = \frac{\sigma_1^2}{S\sigma_1^2+\sigma_\varepsilon^2}\sum_{s=1}^{S}\theta_{cs} + \frac{\sigma_\varepsilon^2}{S\sigma_1^2+\sigma_\varepsilon^2}\mu_{c1}. \tag{15}$$

The exact order of voting does not matter in this case since voters do not gather additional information from observing vote shares in other states, and we thus simply consider the case in which all states vote simultaneously after updating. In this case, vote shares in state s can be summarized as follows:

$$\ln(v_{cs}/v_{0s}) = \eta_{cs} + \frac{\sigma_1^2}{S\sigma_1^2+\sigma_\varepsilon^2}\sum_{s=1}^{S}\theta_{cs} + \frac{\sigma_\varepsilon^2}{S\sigma_1^2+\sigma_\varepsilon^2}\mu_{c1}. \tag{16}$$

4.3. *Electoral Outcomes.* To illustrate the key trade-offs involved and to demonstrate how the simultaneous and sequential systems compare to the public information benchmark, we next consider a special case, which we refer to as the two-by-two model, with two candidates (0 and 1) and two states (A and B). Without loss of generality, assume that state A votes earlier than state B under the sequential system. With only two candidates and normalizing candidate 0 to have quality of zero, we can drop all candidate subscripts (e.g., $\mu_{1t} = \mu_t$). Further, without loss of generality, assume that candidate 1 is not disadvantaged relative to candidate 0 ($\mu_1 \geq 0$). That is, candidate 1 can be considered the front-runner and candidate 0 the dark horse candidate.

With two candidates and two states, the first thing to note is that under any of the three systems, simultaneous, sequential, or all-public information, the front-runner is elected with the following probability:

$$P = \Pr\left[\frac{0.5\exp(E(q|I_A)+\eta_A)}{1+\exp(E(q|I_A)+\eta_A)} + \frac{0.5\exp(E(q|I_B)+\eta_B)}{1+\exp(E(q|I_B)+\eta_B)} > 0.5\right], \tag{17}$$

where I_s represents the information set of voters in state s.[3] Rearranging, we have that the front-runner wins if a front-runner support index (z), which is linear in the key expressions, is positive:

$$P = \Pr[z = 0.5E(q|I_A) + 0.5E(q|I_B) + 0.5\eta_A + 0.5\eta_B > 0]. \tag{18}$$

Then, under simultaneous voting, we have that $I_A = \{\theta_A\}$ and $I_B = \{\theta_B\}$, and using Equation (4) above, we have that

$$P^{sim} = \Pr[0.5\alpha_1(\theta_A+\theta_B) + (1-\alpha_1)\mu_1 + 0.5\eta_A + 0.5\eta_B > 0]. \tag{19}$$

Under the sequential system, we have that $I_A = \{\theta_A\}$ and $I_B = \{\theta_B, v_A\}$, and using the positive analysis above, one can show that

$$\begin{aligned} P^{seq} = \Pr[&(0.5\alpha_1 + 0.5(1-\alpha_2)\beta_1)\theta_A + 0.5\alpha_2\theta_B + (0.5(1-\alpha_1) \\ &+ 0.5(1-\alpha_2)(1-\beta_1))\mu_1 + (0.5 + 0.5(1-\alpha_2)(\beta_1/\alpha_1))\eta_A + 0.5\eta_B > 0]. \end{aligned} \tag{20}$$

[3] Note that ties occur with probability zero in this setting.

Finally, under the public information benchmark, we have that $I_A = I_B = \{\theta_A, \theta_B\}$, and, thus,

$$(21) \qquad P^{public} = \Pr\left[\left(\frac{\sigma_1^2}{2\sigma_1^2 + \sigma_\varepsilon^2}\right)(\theta_A + \theta_B) + \left(\frac{\sigma_\varepsilon^2}{2\sigma_1^2 + \sigma_\varepsilon^2}\right)\mu_1 + 0.5\eta_A + 0.5\eta_B > 0\right].$$

Thus, under all three systems, support for the front-runner can be summarized as a linear index of signals (θ_A, θ_B), the size of the advantage for the front-runner (μ_1), and the preferences of the two states (η_A, η_B). That is,

$$(22) \qquad P = \Pr[z = \omega(\theta_A)\theta_A + \omega(\theta_B)\theta_B + \omega(\mu)\mu_1 + \omega(\eta_A)\eta_A + \omega(\eta_B)\eta_B > 0].$$

Thus, in the two-by-two model, we can fully characterize these three systems according to the relative weights that they place upon signals, priors, and preferences.

SUMMARY OF THREE VOTING SYSTEMS

	Simultaneous	Sequential	Public information
$\omega(\theta_A)$	$0.5\alpha_1$	$0.5\alpha_1 + 0.5(1-\alpha_2)\beta_1$	$\frac{\sigma_1^2}{2\sigma_1^2+\sigma_\varepsilon^2}$
$\omega(\theta_B)$	$0.5\alpha_1$	$0.5\alpha_2$	$\frac{\sigma_1^2}{2\sigma_1^2+\sigma_\varepsilon^2}$
$\omega(\mu)$	$1-\alpha_1$	$0.5(1-\alpha_1) + 0.5(1-\alpha_2)(1-\beta_1)$	$\frac{\sigma_\varepsilon^2}{2\sigma_1^2+\sigma_\varepsilon^2}$
$\omega(\eta_A)$	0.5	$0.5 + 0.5(1-\alpha_2)(\beta_1/\alpha_1)$	0.5
$\omega(\eta_B)$	0.5	0.5	0.5

As shown in the above table, neither the simultaneous nor the sequential system implements the public information benchmark outcome in general. However, the simultaneous system does share the feature of the public information benchmark that the information and preferences of the different states are weighted equally. This feature is not present in the sequential system. These differences among the systems are summarized in the following proposition.

PROPOSITION 1. *The sequential system places disproportionate weight on the preferences and information of the early state, whereas the simultaneous and public information systems place equal weight on the preferences and information of the early and late states. That is,* $\frac{\partial z^{seq}/\partial\theta_A}{\partial z^{seq}/\partial\theta_B} > 1$, $\frac{\partial z^{seq}/\partial\eta_A}{\partial z^{seq}/\partial\eta_B} > 1$, *and* $\frac{\partial z^{sim}/\partial\theta_A}{\partial z^{sim}/\partial\theta_B} = \frac{\partial z^{sim}/\partial\eta_A}{\partial z^{sim}/\partial\eta_B} = \frac{\partial z^{public}/\partial\theta_A}{\partial z^{public}/\partial\theta_B} = \frac{\partial z^{public}/\partial\eta_A}{\partial z^{public}/\partial\eta_B} = 1$.

Thus, the sequential system has the disadvantage of providing disproportionate influence to the early state, both in terms of information and preferences. On the other hand, under the sequential system, voters make better informed choices, and this system thus has the advantage of placing more weight on information in aggregate and less weight on the prior. This leads to the front-runner being overly advantaged in the simultaneous election, relative to the sequential system. This advantage of the sequential system is summarized in the following proposition.

PROPOSITION 2. *The weight placed on the prior is higher under the simultaneous system than under the sequential system, which, in turn, places more weight on the prior than the all-public system, i.e.,* $\omega_\mu^{sim} > \omega_\mu^{seq} > \omega_\mu^{public}$.[4] *Moreover, the front-runner has a higher probability of winning the simultaneous election than the sequential election, i.e.,* $P^{sim} > P^{seq}$.

[4] Throughout we let ω_μ^{sim} denote the value of $\omega(\mu)$ corresponding to the simultaneous system, let ω_μ^{seq} denote the value of $\omega(\mu)$ corresponding to the sequential system, and let ω_μ^{public} denote the value of $\omega(\mu)$ corresponding to the public information benchmark.

Proofs of propositions are in the Appendix. The intuition for the first result in Proposition 2 ($\omega_\mu^{sim} > \omega_\mu^{seq}$) is as follows: Early voters place equal weight on their signals in the sequential and simultaneous systems. Late voters, by contrast, have an additional piece of information, returns from the early state, in the sequential election, when compared to the simultaneous election, and thus place less weight on their prior. Thus, in aggregate, the sequential system places more weight on the available information and less weight on the prior when compared to the simultaneous system.

Regarding the second result in Proposition 2 ($\omega_\mu^{seq} > \omega_\mu^{public}$), early voters have more information under the all-public system and thus place less weight on their prior than in the sequential election. Late voters also have more information under the all-public system since they observe the true signal of the early state. Under sequential voting, late voters only observe voting returns, which are a noisy signal of the state's information, and hence place more weight on their prior. Thus, both early and late voters place more weight on their prior under sequential voting.

The third result ($P^{sim} > P^{seq}$) follows from the three differences between the sequential and simultaneous systems. First, the sequential system places more weight on information in aggregate and less weight on the prior. Second, the sequential system places more weight on the information from the early state, relative to the late state. Finally, the sequential system places more weight on the preferences of the early state, relative to the late state. All three of these factors contribute to the sequential system having more variance, and hence being less predictable, than the simultaneous system. Thus, the front-runner has a smaller advantage under the sequential system than under the simultaneous system.[5]

To summarize, in the two-by-two model, the simultaneous system has the advantage of giving equal weight to state-level information and preferences, whereas the sequential system has the advantage of allowing dark horse candidates of unexpectedly high quality to emerge from the field of candidates. Complementing this analysis, the next section provides a comparison of welfare under the two systems.[6]

4.4. *Welfare Comparison.* We next compare welfare under the sequential system to welfare under the simultaneous system. In this two-by-two model, Equation (14) simplifies to

$$E(W) = E(y|z > 0)\Pr(z > 0), \tag{23}$$

where $y = q + 0.5\eta_A + 0.5\eta_B$ captures aggregate voter utility from the front-runner winning office instead of the dark horse candidate. Using the properties of the normal distribution, $E(q) = \mu_1$, and $z \sim N(\mu_1, \sigma_z^2)$, we then have that

$$E(W) = \mu_1 P + \rho_{y,z}\sigma_y \phi\left(\frac{\mu_1}{\sigma_z}\right), \tag{24}$$

where $P = \Phi(\frac{\mu_1}{\sigma_z})$ captures the probability of the front-runner winning the election, and $\rho_{y,z}$ represents the correlation between aggregate voter utility from the front-runner winning office and the index of support for the front-runner.

[5] $P^{sim} > P^{public}$ also holds because the simultaneous system places more weight on priors than the public information benchmark. However, it is unclear whether the front-runner has a higher or lower probability of winning under the sequential system than in the public information benchmark since the sequential system places more emphasis on preferences in addition to more heavily weighting priors. In the special case of no preference heterogeneity ($\sigma_\eta^2 = 0$), this second factor goes away, and $P^{seq} > P^{public}$.

[6] It is worth noting that if there are two candidates and voters are fully strategic, then the results in Dekel and Piccione (2000) guarantee that any strategies that are equilibrium strategies under simultaneous voting are also equilibrium strategies under sequential voting, and there would be no difference in welfare between the two systems. Feddersen and Pesendorfer's (1997) results would then guarantee that either voting system achieves the same welfare as the full information benchmark. These results do not apply in the realistic scenario in which at least some voters vote sincerely, and they also do not apply in multicandidate elections. We consider both of these possibilities in both the theoretical analysis and the empirical analysis in this article.

Using this welfare expression, we then have that the difference in expected welfare between the simultaneous and sequential systems is given by:

$$\Delta = E^{sim}(W) - E^{seq}(W)$$
$$= \mu_1 \left(P^{sim} - P^{seq}\right) + \rho^{sim}_{y,z}\sigma_y \phi\left(\frac{\mu_1}{\sigma^{sim}_z}\right) - \rho^{seq}_{y,z}\sigma_y \phi\left(\frac{\mu_1}{\sigma^{seq}_z}\right). \tag{25}$$

The first term measures the expected benefit from electing the front-runner (μ_1) multiplied by the difference in the probabilities of the front-runner being elected under the two systems. Since the front-runner is more likely to win under the simultaneous system, this first term is positive and can be interpreted as the reduction in risk associated with the dark horse candidate winning less often under the simultaneous system. The second term can either be positive or negative and depends on $\rho^{sim}_{y,z}$ and $\rho^{seq}_{y,z}$, the correlations between aggregate voter utility (y) and the index of support for the front-runner (z) under the two systems.

To understand how this welfare difference varies with the parameters of the model, it is necessary to understand how the correlations between aggregate voter utility (y) and the index of support for the front-runner (z), $\rho^{sim}_{y,z}$ and $\rho^{seq}_{y,z}$, compare under the two systems. This question is addressed in the following proposition.

PROPOSITION 3. *The correlation between aggregate utility and the index of support for the front-runner is greater under the simultaneous system than under the sequential system, i.e.,* $\rho^{sim}_{y,z} > \rho^{seq}_{y,z}$.

The fact that the correlation between aggregate utility and the index of support for the front-runner is greater under the simultaneous system than under the sequential system is due to how the two systems weigh the information and preferences of the different states. Since the sequential system gives disproportionate weight to the information and preferences of voters in the early state instead of weighing both states equally, vote shares under the sequential system are not as strongly correlated with aggregate utility as vote shares under the simultaneous system.

We now use Proposition 3 to prove the main result about when the simultaneous system is welfare-preferred to the sequential system.

PROPOSITION 4. *The simultaneous system is welfare-preferred when the front-runner's advantage is small, and the sequential system is welfare-preferred when the front-runner's advantage is large. In particular,* $\Delta > 0$ *for sufficiently small values of* μ_1*, and* $\Delta < 0$ *for sufficiently large values of* μ_1.

To understand the intuition behind this result, note that when the front-runner's advantage is small, the welfare comparison between the simultaneous and the sequential systems reduces to a comparison between which system has greater correlation between aggregate utility and vote shares. Since we have seen that this correlation is greater under the simultaneous system, the simultaneous system is welfare-preferred when the front-runner's advantage is small.

However, when the front-runner's advantage is large, this correlation difference becomes less relevant since the front-runner is very likely to win under either system. Instead, the most important factor becomes the fact that the sequential system gives the dark horse candidate a relatively greater chance of winning in circumstances when this candidate is actually the better candidate. For this reason, the sequential system is welfare-preferred when the front-runner's advantage is large.

The intuition for Proposition 4 emphasizes that the main benefit to simultaneous elections over sequential elections is that there is greater correlation between aggregate utility and vote shares under simultaneous elections due to the fact that simultaneous elections equally weigh the preferences and information of the voters in the different states. To further highlight this advantage, we now illustrate how, in a special case of the model with no preference heterogeneity across states, expected welfare under the sequential system can be improved by reweighting the votes of the voters in different states in such a way as to place more similar amounts of weight on the information of the voters in the different states.

PROPOSITION 5. *A weighted voting system that places marginally more weight on the votes of the later state results in higher expected welfare than a system which places equal weight on the votes of the two states under sequential elections when there is no preference heterogeneity* ($\sigma_\eta^2 = 0$).

The reason Proposition 5 holds is that the main disadvantage of sequential elections relative to simultaneous elections is that the sequential system fails to equally weigh the information and preferences of the different states. By using a weighted voting system in which slightly greater weight is placed on the votes of the voters in the later state, this disadvantage is mitigated, and the sequential system results in relatively higher expected welfare than before.

5. EXTENSIONS

Throughout our analysis so far, we have made use of the assumption that the state-level preferences η_{cs} are random draws from the same distribution for all states. Although this is a natural starting point for our analysis, the assumption that these preferences are drawn from the same distribution might not hold in real life. For instance, in practice, one might expect *ex ante* that a voter from Oklahoma is more likely to have right-wing preferences than a voter from Massachusetts, so η_{cs} is more likely to assume a right-wing value when the state is Oklahoma than when the state is Massachusetts. In this case, the assumption that the state-level preferences η_{cs} are random draws from the same distribution for all states might be inappropriate.

To address this possibility, we now consider an extension of the model in the previous section in which instead of assuming that $\eta_s \sim N(0, \sigma_\eta^2)$ for both states $s = A$ and $s = B$, we instead assume that $\eta_A \sim N(\eta, \sigma_\eta^2)$ and $\eta_B \sim N(-\eta, \sigma_\eta^2)$ for some value of η. Such a model has no net effect on total expected welfare from electing a given candidate because the distribution of $\overline{\eta}_c$, the average state-level preference for candidate c, is the same for any given η. However, this model does encompass the possibility that different states may be expected to have different private preferences for one candidate or the other because state A is more likely to have a private preference for candidate 1 than state B if $\eta > 0$.

The possibility that the state-level preferences for the candidates may be drawn from different distributions for different states in the manner described above turns out to have no substantive effect on any of the results of the article. In particular, we prove the following result.

PROPOSITION 6. *Suppose that* $\eta_A \sim N(\eta, \sigma_\eta^2)$ *and* $\eta_B \sim N(-\eta, \sigma_\eta^2)$. *Then, whether simultaneous elections are welfare-preferred to sequential elections is independent of the value of* η.

To understand the intuition behind this result, note that the distributions of $\eta_A - \eta$ and $\eta_B + \eta$ are the same for all η, and aggregate voter welfare for particular realizations of $\eta_A - \eta$ and $\eta_B + \eta$ is the same for any value of η. One can also show that the probabilities the various candidates will be elected under the two systems for particular realizations of $\eta_A - \eta$ and $\eta_B + \eta$ are the same for any value of η because for any given value of η, one can derive the appropriate expression for the probability a candidate is elected by simply replacing η_A with $\eta_A - \eta$ and η_B with $\eta_B + \eta$ in the expressions that were derived in the original model. By combining all of this, it follows that changes in the value of η have no effect on aggregate voter welfare under either system and thus have no effect on whether simultaneous elections are welfare-preferred to sequential elections.

With these results in mind, we now turn our attention to the question of endogenous candidate strategies. In the analysis so far, we have assumed that changing from a sequential system to a simultaneous system does not have any effect on voter preferences in any particular state.

This assumption might seem questionable because under simultaneous elections, the candidates must decide how to allocate campaign resources to various states, whereas under sequential systems, this is less relevant since candidates can always focus their effort on the next election at hand. These changed campaign strategies in the different types of elections could, in principle, affect the likely distribution of voter preferences in the states. We thus seek to address the question of how the possibility that candidates might endogenously vary their strategies could affect the results in this section.

To address this possibility, we extend the main model by considering what would happen if candidates could endogenously decide how much money to spend in a particular state prior to the election. We model the effect of campaign expenditures on voter preferences by assuming that by spending more money in a given state, a candidate can change the mean value of the distribution from which state-level preferences are drawn in that state in a manner that makes it more likely that voters will prefer that particular candidate. In particular, we assume that if the front-runner spends an amount c_j in state j, then this increases the mean value of the state-level preference in state j for the front-runner by an amount $v_j(c_j)$. Similarly, we assume that if the dark horse candidate spends an amount c_j in state j, then this decreases the mean value of the state-level preference for the front-runner in state j by an amount $v_j(c_j)$.

Formally, the strategies for the candidates are as follows: Each candidate has a budget C that represents the maximum total amount of money that the candidate may spend over the course of the campaign. The front-runner chooses an amount to spend in the first state c_1 and must then spend a total of $C - c_1$ in the second state. Similarly, the dark horse candidate chooses an amount to spend in the first state c_0 and must then spend a total of $C - c_0$ in the second state. If the front-runner spends c_1 in the first state, and the dark horse candidate spends c_0 in the second state, then the state-level preference for the front-runner in the first state, η_A, will be a random draw from the normal distribution $N(v_A(c_1) - v_A(c_0), \sigma_\eta^2)$ for some nondecreasing function $v_A(\cdot)$ satisfying $v_A(0) = 0$. Similarly, the state-level preference for the front-runner in the second state, η_B, will be a random draw from the normal distribution $N(v_B(C - c_1) - v_B(C - c_0), \sigma_\eta^2)$ for some nondecreasing function $v_B(\cdot)$ satisfying $v_B(0) = 0$. Each candidate's objective is to maximize the probability that he or she is elected.

In this setting, the possibility that candidates endogenously choose their campaign strategies in response to the type of election used to ultimately determine the identity of the elected candidate again has no effect on whether simultaneous elections are welfare-preferred to sequential elections. In particular, we prove the following result.

Proposition 7. *If candidates choose their strategies endogenously, then the distributions of voter preferences in each state will be the same after candidates follow these strategies as they were in the original model.*

The intuition for this result is that if candidates choose their strategies endogenously, then candidates are likely to have similar beliefs about which states are the important states and follow similar strategies in terms of how they allocate their campaign resources.[7] As a result, endogenous candidate strategies will have little net effect on the ultimate distribution of voter preferences. The result is proven formally by showing that candidates will choose the same resource allocation as one another in each of the states, and thus endogenous candidate strategies will have no net effect on the distribution of voter preferences.

An immediate consequence of Proposition 7 is that the possibility of endogenous candidate strategies also has no effect on whether simultaneous elections are welfare-preferred to sequential elections. Since endogenous candidate strategies have no effect on the distribution of voter preferences, they also have no effect on which system is better for voter welfare.

[7] This has been observed to be the case empirically in presidential primaries. For instance, Brams and Davis (1982) note that presidential candidates followed resource allocation strategies similar to one another in the 1976 presidential primaries.

6. MULTIPLE CANDIDATES AND STRATEGIC VOTING

Throughout our welfare analysis so far, we have restricted attention to cases in which there are two candidates and two states. Although it is nearly universal for research on sequential voting to restrict attention to elections with two candidates (e.g., Brams and Davis, 1982; Dekel and Piccione, 2000; Strumpf, 2002; Battaglini, 2005; Klumpp and Polborn, 2006; Battaglini et al., 2007; Callander, 2007; Selman, 2010; and Ali and Kartik, 2012 all restrict attention to models with exactly two candidates), in U.S. presidential primaries, it is quite common for there to be multiple candidates contending for their party's nomination. Furthermore, we have focused attention on elections in which there are exactly two states, but presidential primaries in the United States contain more than two states. In this section, we seek to address how the possibility that there are multiple candidates and more than two states that are voting would affect the main results of the article.

The existence of multiple candidates introduces a potential complication that was not present in the two candidate case. In our analysis of the case with two candidates, we made use of the assumption that voters vote sincerely by simply voting for whichever candidate they like best. Although it would be natural for voters to simply vote for the candidate they like best in two candidate elections, this might not necessarily hold in multicandidate elections because if a candidate is perceived as having little chance of winning, then voters might vote strategically by instead voting for the candidate they like best among the other candidates. We thus also take into account the possibility that voters might vote strategically in analyzing multicandidate elections in this section.

How strategic voting will influence voter strategies depends on the nature of the distribution of voter preferences. When voter preferences for the various candidates are drawn from distributions such that voters are roughly equally likely to like each of the candidates best, then work in Hummel (2012) illustrates that it will typically be an equilibrium for voters to simply vote sincerely by voting for whichever candidate they like best under simultaneous elections. Furthermore, it will also be an equilibrium for voters in the first few states to vote sincerely by voting for whichever candidate they like best under sequential voting. However, voters in later states will ultimately have an incentive to vote strategically by simply voting for whichever candidate they like best among the two candidates who have received the most votes in the first few states. Thus, when voter priors are similar for the different candidates, we model strategic voting by assuming that voters vote sincerely under simultaneous voting and in the first few states of sequential voting. However, we also assume that voters vote for the candidate they like best among the two candidates who received the most votes in the first few states after some number of states, m, have already voted.

When voters follow these strategies in multicandidate elections, it is still the case that simultaneous elections are welfare-preferred to sequential elections when voter priors about the likely qualities of the candidates are similar for the different candidates. This is illustrated in the following proposition.

PROPOSITION 8. *Suppose there are at least three candidates and voter priors about the likely quality of candidate c, μ_{c1}, are sufficiently close to 0 for all c. Then, simultaneous elections are welfare-preferred to sequential elections if there are a sufficiently large number of states.*

The intuition behind this result is that when voter priors about the likely qualities of the candidates are the same for the different candidates, then the highest quality candidate will win with probability arbitrarily close to one in the limit as the number of states becomes large under simultaneous voting. But under sequential voting, there is always some nonnegligible probability that the candidate that is actually the best candidate will have a poor showing in the first few states and strategic voters in later states will simply stop voting for this candidate by instead voting for whichever candidate they like best among the two candidates who had the strongest performance in the first few states. Thus, there is always some significant probability that the best candidate will fail to win under sequential elections, and simultaneous elections are welfare-preferred to sequential elections under the conditions given in Proposition 8.

Having illustrated how welfare compares under simultaneous and sequential voting in multicandidate elections when voter priors about the likely qualities of the candidates are similar for the different candidates, we now analyze this welfare comparison when voter priors about the likely qualities of the candidates differ significantly for the different candidates. When voter preferences for the various candidates are drawn from distributions such that voters are substantially more likely to like one of two candidates best than to like the other candidates best, then it will typically no longer be an equilibrium for voters to simply vote for whichever candidate they like best in either simultaneous voting or sequential voting. Instead, there will be an equilibrium in which voters vote for the candidates they like best among the two candidates for which voters have the strongest priors.

When voter priors about the likely qualities of the candidates are sufficiently different for the different candidates that voters vote strategically in the manner described above, the welfare comparison between simultaneous voting and sequential voting completely changes. In particular, we obtain the following result.

PROPOSITION 9. *Suppose there are at least three candidates and voter priors about the likely quality of candidates c and d, μ_{c1} and μ_{d1}, are different for all c and d. Also, suppose that voters vote strategically by voting for the candidate they like best among the two candidates for which voters initially have the strongest priors. Then, sequential elections are welfare-preferred to simultaneous elections if there are a sufficiently large number of states.*

The intuition behind this result is similar to the intuition behind why sequential elections are welfare-preferred to simultaneous elections when there are two candidates. When voters vote strategically by voting for the candidate they like best among the two candidates for which voters have the strongest priors, then the election effectively reduces to a two candidate election and the principles that governed which system is welfare-preferred under two candidate elections extend to multicandidate elections. Thus, sequential elections are indeed welfare-preferred to simultaneous elections when the front-runner has a large advantage.

The results in this section indicate that the substantive conclusions of Section 4, which focused on an environment with two candidates and two states, extend to the richer environment with multiple candidates, a large number of states, and strategic voting. When the front-runner initially has a sufficiently small advantage over the other candidates that strategic voters would have an incentive to vote sincerely under either simultaneous voting or the first few states of sequential voting, then simultaneous elections are welfare-preferred to sequential elections. But when the front-runners initially have a sufficiently large advantage over the other candidates that strategic voters would simply vote for the candidate they like best among the two strongest candidates, then sequential elections are welfare-preferred to simultaneous elections. Thus, whether simultaneous elections are welfare-preferred to sequential elections in multicandidate elections with strategic voting depends on the size of the front-runner's advantage in a manner similar to that in two candidate elections.

7. NUMERICAL ANALYSIS

Focusing on the case of sincere voting, many states, and more than two candidates, we next provide a quantitative evaluation of the welfare properties of the simultaneous and sequential elections, when compared to the all-public information benchmark. In addition to computing expected welfare under all three systems, this section aims to provide quantitative evidence on the trade-offs between the two systems.

7.1. *Methodology.* In order to conduct this evaluation, we use the key parameter estimates from the application to the 2004 Democratic presidential primary from Knight and Schiff (2010).

TABLE 1
PARAMETER ESTIMATES FROM KNIGHT AND SCHIFF (2010)

μ_{D1}	0.938** [0.773,1.14]
μ_{E1}	−0.701** [−0.913,−0.433]
$\sigma_\eta{}^2$	0.815** [0.551,1.194]
$\sigma_1{}^2$	3.577** [1.497,7.129]
$\sigma_\varepsilon{}^2$	1.197** [0.062,4.097]

NOTES: [bootstrap 95% confidence interval]; ** denotes significance at the 95% level.

This analysis focused on the three key candidates, Kerry, Dean (D), and Edwards (E), where Kerry was considered the baseline candidate. Estimates of the key parameters (μ_{D1}, μ_{E1}, σ_1, σ_ε, and σ_η) from this analysis are summarized in Table 1. Using daily polling data from the primary season, the parameters involving the mean of the prior (μ_{D1}, μ_{E1}) were identified by national support for candidates prior to the primary season, and the parameter σ_η is identified by the variation in such support across states. The key learning parameters (σ_1 and σ_ε) are identified by the degree to which voters in late states respond to actual voting outcomes in early states. As shown in Table 1, given his lead in the polls prior to the start of the primary season, Dean can be considered the front-runner in this analysis, followed by Kerry and then Edwards.

Using these parameter estimates, the numerical analysis proceeds in the following steps:

1. Randomly draw an absolute quality value for Kerry, Dean, and Edwards from the normal distribution so that the values, relative to Kerry, have means μ_{D1} and μ_{E1}, respectively, and common variance σ_1^2.[8]
2. For each state s, randomly draw an absolute signal noise value for Kerry, Dean, and Edwards from the normal distribution so that the values, relative to Kerry, have mean 0 and variance σ_ε^2.
3. Calculate the state-level signal, relative to Kerry, for Dean ($\theta_{Ds} = q_D + \varepsilon_{Ds}$) and for Edwards ($\theta_{Es} = q_E + \varepsilon_{Es}$).
4. For each state s, randomly draw an absolute preference for Kerry, Dean, and Edwards from the normal distribution so that the values, relative to Kerry, have mean 0 and variance σ_η^2.
5. Given these signals and preferences and using the models outlined above, compute the vote shares in each state s for Dean (v_{Dst}), Edwards (v_{Est}), and Kerry ($1 - v_{Dst} - v_{Est}$) under the sequential system, using the actual calendar from 2004, the simultaneous system, and finally, the all-public information system.
6. Compute the national vote shares as the average vote shares across states and identify the winner of the election as the candidate receiving a plurality of the vote.
7. Compute voter welfare in Equation (13).

Finally, steps 1–7 are repeated 50,000 times and we estimate expected welfare, as expressed in Equation (14), under each of the three systems as the average voter welfare across these 50,000 replications. Thus, by drawing new values of candidate quality, signals, and preferences in every replication, this is an ex ante measure of voter welfare. That is, while Kerry won the 2004 primary from an ex post perspective, Dean both wins most often and proves to be the

[8] More precisely, an absolute, standardized quality was first drawn for Dean ($\widetilde{q}_D$), Kerry ($\widetilde{q}_K$), and Edwards ($\widetilde{q}_E$) from the standard normal distribution. This was then translated into relative quality via $q_D = \mu_{D1} + (\sigma_1/\sqrt{2})(\widetilde{q}_D - \widetilde{q}_K)$ and $q_E = \mu_{E1} + (\sigma_1/\sqrt{2})(\widetilde{q}_E - \widetilde{q}_K)$. Then, q_D and q_E have a mean of μ_{D1} and μ_{E1}, respectively, and a variance of σ_1^2. A similar procedure was followed for the signal noise and preference draws in steps 2 and 4.

TABLE 2
BASELINE WELFARE CALCULATIONS

System	Average Welfare	Welfare Gains	% of Maximal Gains Starting from Seq	% of Maximal Gains from Moving to Sim	Pr(Dean)	Pr(Edwards)	Pr(Kerry)
All Public	1.4150 (0.0063)	0.0198 (0.0004))	100.00%	4.31%	60.88%	12.61%	26.50%
Simultaneous	1.3978 (0.0064)	0.0026 (0.0005)	13.13%	0.57%	68.91%	8.11%	22.98%
Sequential	1.3952 (0.0064)				62.24%	11.84%	25.92%
Details on Gains from Simultaneous System:							
Pr(gains=0)				90.45%			
Pr(gains>0)				4.72%			
Pr(gains<0)				4.84%			
E(gains|gains>0)				0.3207			
E(gains|gains<0)				−0.2577			
E(q_d|gains≠0)				0.3248			
E(q_e|gains≠0)				−0.435			

NOTES: Standard errors are in parentheses.

highest quality candidate most often in our simulations, given that quality is drawn from the distribution associated with voter priors.

7.2. *Baseline Results.* The results from this analysis are presented in Table 2. As shown, neither system produces the expected welfare levels associated with the all-public information benchmark, under which all voters have access to all signals. In particular, although the all-public system generates voter welfare of 1.4150, the simultaneous system generates welfare of 1.3978, and the sequential system generates welfare of 1.3952. Comparing the simultaneous and sequential systems, welfare under the simultaneous system increases by 0.0026 units. Although this difference is based upon a finite number of replications, the standard error associated with this difference is 0.0005, and we can thus say that this welfare gain is statistically significant at conventional levels.

In terms of the magnitude of any welfare gains associated with moving from our current system to a simultaneous system, there are two relevant benchmarks. First, these welfare gains can be compared to the welfare difference between the all-public and sequential systems. That is, we calculate $[E^{sim}(W) - E^{seq}(W)]/[E^{pub}(W) - E^{seq}(W)]$. This difference, as expressed in the denominator, can be interpreted as the maximal possible gains when starting from the sequential system. According to this measure, the difference in welfare between the simultaneous and sequential systems represents about 13% of maximal gains that can be achieved from reforming the status quo electoral system. Second, these welfare gains can be compared to the difference between the simultaneous system and a no-information system, under which Dean would always be elected and expected welfare equals 0.938. That is, we calculate $[E^{sim}(W) - E^{seq}(W)]/[E^{sim}(W) - \mu_{D1}]$. This difference can be interpreted as relative to the maximal possible gains associated with moving to the simultaneous system. According to this measure, the difference in welfare between the simultaneous and sequential systems is less than 1% of the maximal possible gains. This small gain reflects the fact that both systems, simultaneous and sequential, substantially outperform the no-information case. This, in turn, follows from the fact that the noise in the signal, as estimated by Knight and Schiff (2010), is small ($\sigma_\epsilon^2 = 1.197$) relative to the variance in the initial prior ($\sigma_1^2 = 3.577$). Thus, voters learn a substantial amount from a single piece of information.[9]

[9] In particular, the precision in the prior ($1/\sigma_q^2$) increases from 0.280 to 1.115 after observing one signal.

TABLE 3
WELFARE CALCULATIONS REMOVING FRONT-RUNNER ADVANTAGE

System	Average Welfare	Welfare Gains	% of Maximal Gains Starting from Seq	% of Maximal Gains from Moving to Sim
All Public	1.1300	0.0236	100.00%	2.09%
	(0.0057)	(0.0005)		
Simultaneous	1.1295	0.0231	97.88%	2.05%
	(0.0057)	(0.0005)		
Sequential	1.1064			
	(0.0058)			

NOTE: Standard errors are in parentheses.

The bottom panel of Table 2 provides three additional facts that help to explain why the measured welfare gain is relatively small. First, the two systems, simultaneous and sequential, select the same candidate in roughly 90% of cases, and the welfare gains from simultaneous are necessarily zero in these cases. Second, in the subset of cases where the systems select different candidates, the sequential system is actually more likely to select the better candidate (4.84% of cases versus 4.72% of cases). However, the expected gains when simultaneous selects better candidates exceed the expected losses when sequential selects better candidates (0.3207 versus 0.2577). Thus, when the two systems do select different candidates, expected welfare is slightly larger under simultaneous. Finally, as shown in the final two rows, differences between candidate quality are relatively small in the cases when the two systems select different candidates. That is, whereas Dean's average quality, relative to Kerry, across all simulations equals 0.938, it averages only 0.3248 in cases where the two systems select different candidates, and a similar pattern occurs with respect to average quality for Edwards. Said differently, the distinction between the two systems is most salient in cases when the stakes associated with candidate selection are smaller. All three of these factors contribute to relatively small welfare gains associated with moving from sequential to simultaneous. Having said that, as will be shown below, there are situations in which welfare differences between electoral systems are substantially larger in magnitude.

7.3. *Differences in Competition.* To illustrate the trade-offs between sequential and simultaneous, we next provide quantitative evidence on the differences between the two systems. We begin by highlighting, consistent with the theoretical results, that the sequential system is more competitive than the simultaneous system. This difference is highlighted in the final three columns of Table 2. As shown, the simultaneous system does give too much advantage to the front-runner, with Dean, who led prior to Iowa, winning in 69% of the cases. Under the full information system, by contrast, Dean wins in only 61% of the cases, and the sequential system, in which Dean wins in 62% of the cases, gives dark horse candidates a substantially better chance of winning. Conversely, the simultaneous system disadvantages the dark horse candidates, Kerry and Edwards, who win in just 23% and 8% of the cases, respectively. These candidates have significantly higher chances of winning in the all-public and sequential systems. These probabilities highlight the advantage of the sequential system.

To further highlight this advantage of the sequential system, we next consider sequential and simultaneous elections in a counterfactual environment with no front-runner advantage. In particular, we eliminate ex ante differences between the candidates by setting $\mu_{D1} = 0$ and $\mu_{E1} = 0$. As shown in Table 3, the simultaneous system achieves welfare levels that are nearly identical to those under full information and achieves a substantially larger gain equal to 2.05% of the possible gains associated with moving from no information to full information. Thus,

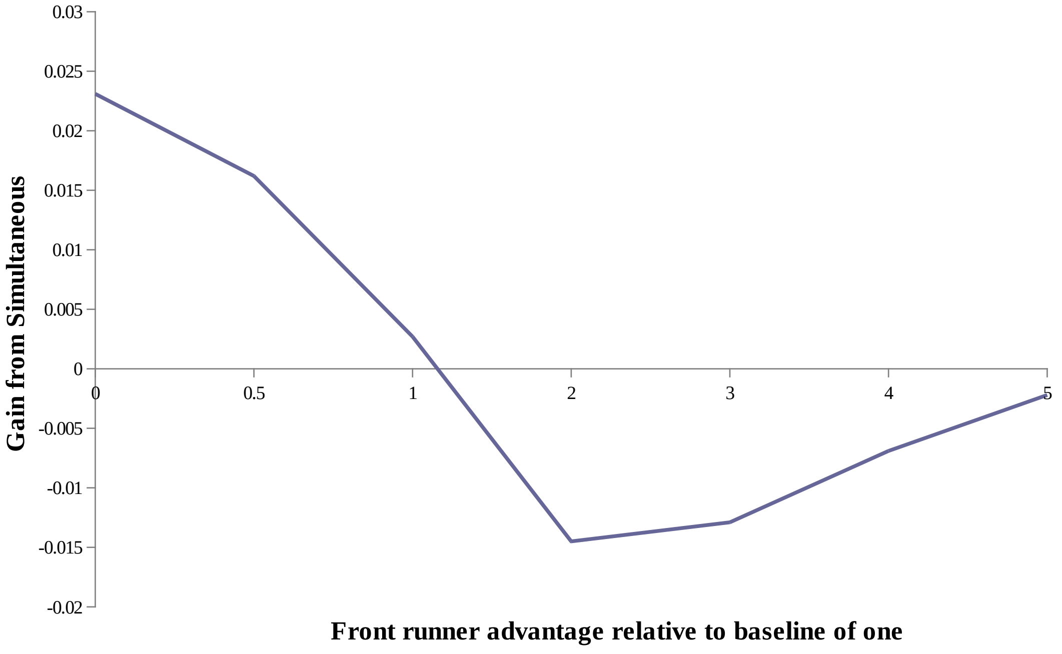

FIGURE 1

WELFARE GAINS FROM SIMULTANEOUS AND DEAN'S ADVANTAGE

by eliminating the advantages associated with the sequential system, the gains from moving to simultaneous are substantially larger.

To further explore the role of competition, we next calculate the welfare gains associated with moving from the sequential system to the simultaneous system under different alternative electoral advantages for the front-runner. In particular, whereas our baseline estimates are based upon $(\mu_{D1}, \mu_{E1}) = (0.938, -0.701)$, we next consider $(\mu_{D1}, \mu_{E1}) = (\lambda 0.938, -\lambda 0.701)$ for $\lambda = \{0.5, 2, 3, 4, 5\}$. Thus, λ can be considered a measure of the electoral advantage of the front-runner, where Table 2 presented baseline results for $\lambda = 1$, and Table 3 presented counterfactual results for $\lambda = 0$. As shown in Figure 1, the welfare gains from moving to simultaneous are positive and larger than the baseline ($\lambda = 1$) when the front-runner's advantage is small ($\lambda = 0$ and $\lambda = 0.5$), reflecting the fact that the advantage afforded to the front-runner under the simultaneous system is less salient in these cases. For front-runner advantages greater than the baseline ($\lambda > 1$), however, the sequential system outperforms the simultaneous system. This welfare difference, however, grows smaller as the advantage grows larger, reflecting the fact that the front-runner is increasingly likely to win under either system. Thus, the baseline case of $\lambda = 1$ appears to be the scenario under which the advantages and disadvantages of the two systems are roughly balanced.

7.4. *Differences in Weighting of Information and Preferences.* We next use the simulations to illustrate the key advantage of simultaneous, equal weighting of information between early and late states. In particular, we next provide quantitative evidence on the disproportionate influence of early states under sequential voting. Although analytic expressions for the relative vote shares are not a linear function of the players' private signals and preferences when there are more than two candidates and more than two states, we can approximate the extent to which changes in these signals and preferences affect the relative vote shares via a linear regression. In particular, using each of the 50,000 replications as an observation, we relate the cross-state

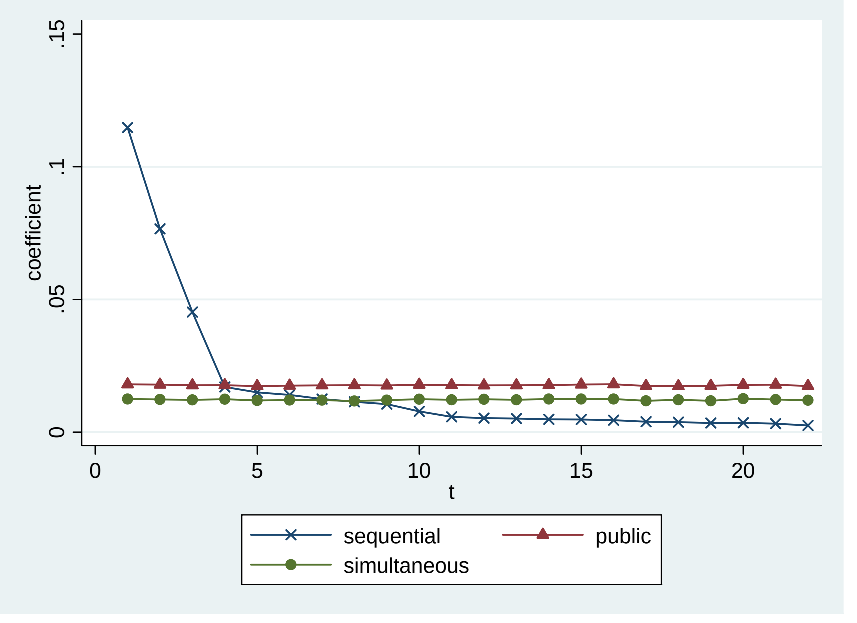

FIGURE 2

THE RESPONSE OF VOTES TO SIGNALS

average vote share of the front-runner, Dean, to the information and preferences of states at different points in the sequence by estimating the parameters of the following equation:

$$\ln\left(\frac{v_D}{1-v_D}\right) = \kappa + \sum_{t=1}^{t=22} \omega_t(\theta)\theta_t + \sum_{t=1}^{t=22} \omega_t(\eta)\eta_t, \tag{26}$$

where $\theta_t = (1/N_t)\sum_{s\in\Omega_t}\theta_{st}$ and $\eta_t = (1/N_t)\sum_{s\in\Omega_t}\eta_{st}$ represent the average signal and preference, respectively, among the set of states voting at time t. For comparison purposes, we run two additional regressions, both of which use the sequence from the sequential system but the vote shares for the all-public and simultaneous systems, respectively.

Figures 2 and 3 plot the coefficients on the signals and preferences, respectively, from these regressions. As shown in Figure 2, the sequential system does substantially overweigh the information of early states, with the first state having a coefficient of 0.1148 and the final state having a coefficient of 0.0025. Thus, the influence of the signal of the first state is over 45 times the influence of the signal of the last state. The simultaneous and all-public systems, by contrast, place equal weight on state-level information. Comparing the weights under the simultaneous and all-public systems, this figure also confirms the result that the simultaneous system places too little weight on information in aggregate and thus too much weight on the prior.

Figure 3 displays a similar pattern, with the preferences of the first state to vote having a weight of 0.1530 and the last state having a weight of 0.0171. Thus, the influence of the preferences of the first state is roughly nine times the influence of the preferences of the last state in the sequence.[10] The simultaneous and all-public systems again place equal weight on the preferences of each state. This overweighting of early preferences is sufficiently large such

[10] This indicates that the sequential system more severely overweights the information of early states relative to late states than it does the preferences of early states relative to late states. This makes sense intuitively since voters in later states have more of an incentive to ignore their private information when they have information about how early states voted than they do to ignore their private preferences.

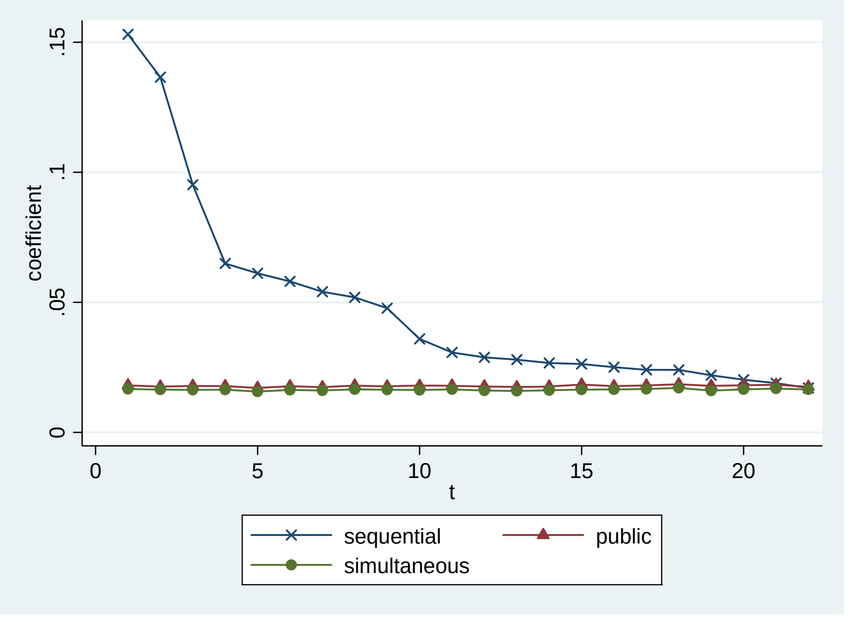

FIGURE 3

THE RESPONSE OF VOTES TO PREFERENCES

that the state-specific gains from simultaneous are negative for early states despite the fact that the gains are positive from a national perspective. This result is illustrated in Figure 4, which plots the average welfare gains for states at each point in the calendar.

To provide additional evidence on this disadvantage of sequential, we next consider a counterfactual weighted sequential system. In particular, we next consider a system in which the votes of early and late states are potentially weighed differently when determining the electoral outcome. As motivated by Proposition 5, by allowing for the possibility of overweighting of votes from late states, such a system may lessen the disadvantages associated with the documented overweighting of early information while retaining the advantages associated with heightened competition. In particular, using the results from the simulated sequential election, we first calculate the optimal set of weights on different states according to the voting order. To simplify the analysis, we consider the following parametric form for the weight placed upon the returns from a state voting at time t:

$$w_t = \frac{\exp(\kappa \times t)}{\sum_{\tau=1}^{\tau=22} N_t \exp(\kappa \times t)}, \tag{27}$$

where $\kappa = 0$ corresponds to equal weights, $\kappa < 0$ overweighs votes from early states, and $\kappa > 0$ overweighs votes from late states. By construction, these weights are nonnegative and sum to one. According to our estimates, the value of κ that maximizes voter welfare is 0.1367, and this implies that the latest state has almost 18 times the voting weight of the earliest state.[11] As shown in the first row of Table 4, this generates a welfare level of 1.3991, dominating both the baseline sequential system and the simultaneous system.

Finally, in Table 4, we also consider four additional sequential systems. First, we consider the 2008 calendar, when nearly half of the states moved their primary to the first Tuesday of

[11] In particular, note that the last state ($t = 22$), relative to the first state ($t = 1$), has a weight equal to $\exp(0.1367 \times 22)/\exp(0.1367 \times 1)$, or 17.65.

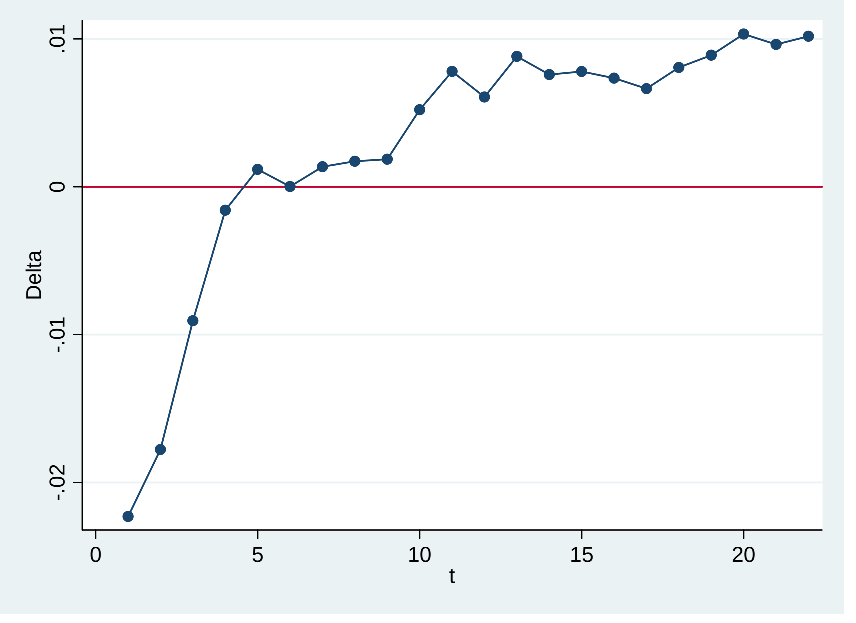

FIGURE 4

GAINS TO SIMULTANEOUS BY VOTING PERIOD

TABLE 4
ALTERNATIVE SEQUENTIAL SYSTEMS

System	Average Welfare	Welfare Gains	% of Maximal Gains Starting from Seq	% of Maximal Gains from Moving to Sim	Pr(Dean)	Pr(Edwards)	Pr(Kerry)
Weighted Sequential (2004 Calendar)	1.3991 (0.0064)	0.0039 (0.0003)	19.70%	0.85%	61.59%	12.23%	26.18%
Sequential (2008 Calendar)	1.3951 (0.0064)	0.0000 (0.0003)	0.00%	0.00%	62.48%	11.66%	25.86%
Sequential (2012 Calendar)	1.3947 (0.0064)	−0.0005 (0.0002)	−2.53%	−0.11%	62.22%	11.89%	25.89%
Rotating Regional	1.3967 (0.0064)	0.0015 (0.0003)	7.58%	0.33%	62.43%	11.76%	25.81%
Pure Sequential	1.3946 (0.0064)	−0.0006 (0.0002)	−3.03%	−0.13%	62.15%	11.91%	25.94%

NOTE: Standard errors are in parentheses.

February, and the 2012 calendar. As shown in Table 4, these alternative calendars still fall short of the simultaneous election in terms of voter welfare. The 2008 calendar produces welfare levels that are nearly identical to those using the 2004 calendar, and the 2012 calendar produces slightly lower welfare levels. We next consider a rotating regional primary system, under which Iowa and New Hampshire maintain their status as the first states to vote. Following these two states, there are then four rounds of voting with 12 states voting in round 1, 13 states in round 2, 12 states in round 3, and 12 states in round 4. As shown, this system also falls short of the simultaneous system in terms of voter welfare but dominates the 2004, 2008, and 2012 calendars. Finally, we consider a pure sequential system, under which every state votes on a different day. As shown, this system has the weakest performance of any system considered here, presumably

TABLE 5
2008 AND 2012 CANDIDATES

System	2008 Democratic Primary Clinton (0.515) Edwards (−0.700) Obama (0)		2008 Republican Primary Huckabee (0.234) McCain (−0.061) Romney (0)		2012 Republican Primary Paul (0) Romney (0.717) Santorum (−0.103)	
	Average Welfare	Welfare Gains	Average Welfare	Welfare Gains	Average Welfare	Welfare Gains
All Public	1.1781	0.0209	1.1946	0.0236	1.2322	0.0204
	(0.0059)	(0.0004)	(0.0058)	(0.0005)	(0.0060)	(0.0004)
Simultaneous	1.1690	0.0118	1.1928	0.0218	1.2173	0.0054
	(0.0059)	(0.0005)	(0.0058)	(0.0005)	(0.0061)	(0.0005)
Sequential	1.1572		1.1710		1.2118	
	(0.0059)		(0.0059)		(0.0061)	

NOTE: Standard errors are in parentheses

reflecting the fact that the disproportionate impact of early states is particularly extreme in this case. Taken together, the results from these alternative sequential calendars suggest that incremental steps toward a simultaneous system tend to increase voter welfare.

7.5. *Alternative Sets of Candidates.* Although our empirical application has focused on the 2004 election, we next extend the analysis by examining the set of candidates from the 2008 and 2012 primary elections. Although a full reestimation of the model parameters is beyond the scope of this article, we have developed a simple method for computing the differences in the mean of voter priors over candidate quality using aggregate results from polling during the month preceding the primary season; further details of the method are summarized in the Appendix. Although these simulations allow the mean of the prior to differ, the other key parameters (σ_1, σ_ε, and σ_η) are set at their baseline values in Table 1.

For comparability with the 2004 primary, we focus on the three candidates receiving the largest share of the delegates from an ex post perspective. During the 2008 Democratic primary, these candidates are Obama (the baseline), Clinton (expected quality, relative to Obama, of 0.515), and Edwards (expected quality, relative to Obama, of −0.700). During the 2008 Republican primary, these candidates are Romney (the baseline), Huckabee (expected quality, relative to Romney, of 0.234), and McCain (expected quality, relative to Romney, of −0.061). Finally, during the 2012 Republican primary, these candidates are Paul (the baseline), Romney (expected quality, relative to Paul, of 0.717), and Santorum (expected quality, relative to Paul, of −0.103). As shown in Table 5, the welfare gains from simultaneous are positive and statistically significant in all cases. Although the magnitude is small for the 2012 election, the gains are larger in both of the 2008 primary elections.

7.6. *Summary.* To summarize, the numerical analysis demonstrates that the counterfactual simultaneous system would have outperformed the sequential system in the context of the 2004 Democratic presidential primary. Although the simultaneous election overly advantages the front-runner, this is outweighed by the fact that the sequential system gives disproportionate weight to early states. In particular, the sequential system gives too much weight to both early information and early preferences. We show that a counterfactual weighted sequential system dominates both simultaneous and unweighted sequential. Finally, we show that the finding that simultaneous dominates sequential is robust to considering the set of candidates in the 2008 and 2012 primaries.

8. CONCLUSION

Although this analysis is meant to be a realistic description of presidential primaries, we have abstracted from several institutional details of these systems, and future work could thus extend the model in interesting directions. First, we have ignored noninformational reasons that momentum may develop for candidates. For example, a candidate who wins an early state may be able to garner more media coverage that helps the candidate in later states. It would be interesting to extend the model to allow for noninformational reasons for momentum. Although our analysis abstracts away from such noninformational reasons for momentum, incorporating these into the model might give further disproportionate influence to early states if voting returns force candidates to exit from the race.

Second, we have not allowed for the possibility of strategic policy choices, which may differ between the two systems. For instance, in a sequential election, candidates may focus on issues that are important to voters in early states. By contrast, in a simultaneous election, candidates may try to run on issues that are more likely to have a broad appeal to the average primary voter. Further research could reveal exactly how this affects the trade-off between simultaneous elections and sequential elections.

To summarize, this article provides a theoretical and empirical analysis of voter welfare under simultaneous and sequential voting systems. Using a model of voting and social learning, we first show that neither the simultaneous nor the sequential system achieves the all-public information welfare benchmark. Although the simultaneous system has the advantage of equally weighing the information and preferences of the different states, the sequential system has the advantage of allowing dark horse candidates of unexpectedly high quality to emerge from the field of candidates. These results imply that the simultaneous system is welfare-preferred if the front-runner is initially only thought of as a slightly better candidate, but the sequential system is welfare-preferred if the front-runner is initially thought of as a significantly stronger candidate. Focusing on the 2004 calendar and the associated pool of candidates, we then conduct an empirical welfare analysis. Although the results suggest that the simultaneous system outperforms the sequential system, the difference in welfare is relatively small.

APPENDIX

PROOF OF PROPOSITION 2. The proof consists of three parts. First, we show that $\omega_\mu^{sim} > \omega_\mu^{seq}$. Second, we show that $\omega_\mu^{seq} > \omega_\mu^{public}$. Third, we show that $P^{sim} > P^{seq}$.

Part 1: To show that $\omega_\mu^{sim} > \omega_\mu^{seq}$, we need the following condition to hold:

$$1 - \alpha_1 > (1 - \alpha_2)(1 - \beta_1).$$

We first use the fact that

$$\frac{1}{\sigma_2^2} = \frac{1}{\sigma_1^2} + \frac{1}{(\sigma_\eta^2/\alpha_1^2) + \sigma_\varepsilon^2}$$

can be rewritten as

$$\sigma_2^2 = \frac{\sigma_1^2[(\sigma_\eta^2/\alpha_1^2) + \sigma_\varepsilon^2]}{\sigma_1^2 + (\sigma_\eta^2/\alpha_1^2) + \sigma_\varepsilon^2}.$$

Next, we use the fact that $1 - \beta_1 = \frac{(\sigma_\eta^2/\alpha_1^2)+\sigma_\varepsilon^2}{\sigma_1^2+(\sigma_\eta^2/\alpha_1^2)+\sigma_\varepsilon^2}$ and substitute in above as follows:

$$\sigma_2^2 = \sigma_1^2(1 - \beta_1),$$

Given that $(1-\alpha_2) = \sigma_\varepsilon^2/(\sigma_\varepsilon^2+\sigma_2^2)$, we thus have that

$$(1-\alpha_2) = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+\sigma_1^2(1-\beta_1)}.$$

And the right-hand side of the original condition is thus given by

$$(1-\alpha_2)(1-\beta_1) = \frac{\sigma_\varepsilon^2(1-\beta_1)}{\sigma_\varepsilon^2+\sigma_1^2(1-\beta_1)}.$$

Plugging this into the original condition and using the definition of α_1, we require that

$$\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2+\sigma_1^2} > \frac{\sigma_\varepsilon^2(1-\beta_1)}{\sigma_\varepsilon^2+\sigma_1^2(1-\beta_1)}.$$

Cross-multiplying and rearranging, we require that

$$\begin{aligned} \sigma_\varepsilon^2+\sigma_1^2(1-\beta_1) &> (1-\beta_1)(\sigma_\varepsilon^2+\sigma_1^2) \\ \sigma_\varepsilon^2 &> (1-\beta_1)\sigma_\varepsilon^2, \end{aligned}$$

which establishes the result.

Part 2: To show that $\omega_\mu^{seq} > \omega_\mu^{public}$, we need the following condition to hold:

$$0.5(1-\alpha_1)+0.5(1-\alpha_2)(1-\beta_1) > \frac{0.5\sigma_\varepsilon^2}{2\sigma_1^2+\sigma_\varepsilon^2}+\frac{0.5\sigma_\varepsilon^2}{2\sigma_1^2+\sigma_\varepsilon^2}.$$

Since it is clear that $(1-\alpha_1) > \frac{\sigma_\varepsilon^2}{2\sigma_1^2+\sigma_\varepsilon^2}$, we only need that

$$(1-\alpha_2)(1-\beta_1) > \frac{\sigma_\varepsilon^2}{2\sigma_1^2+\sigma_\varepsilon^2}.$$

Using the definition of the left-hand side from Part 1, we need that

$$\frac{(1-\beta_1)}{\sigma_\varepsilon^2+\sigma_1^2(1-\beta_1)} > \frac{1}{2\sigma_1^2+\sigma_\varepsilon^2}.$$

Cross-multiplying and rearranging, we require that

$$\frac{(1-\beta_1)}{\beta_1} > \frac{\sigma_\varepsilon^2}{\sigma_1^2}.$$

Using the definition of β_1, we require that

$$\frac{(\sigma_\eta^2/\alpha_1^2)+\sigma_\varepsilon^2}{\sigma_1^2} > \frac{\sigma_\varepsilon^2}{\sigma_1^2},$$

which establishes the result.

Part 3: Note that $P = \Pr(z > 0)$. Since z is normal with mean μ_1 and standard deviation σ_z, we have that $P = \Phi(\mu_1/\sigma_z)$. Thus, to show that $P^{sim} > P^{seq}$, we only need to show that $\sigma_z^{seq} > \sigma_z^{sim}$. First, note that σ_z^2 can be written as follows:

$$\sigma_z^2 = [\omega(\theta_A) + \omega(\theta_B)]^2\sigma_q^2 + [\omega(\theta_A)^2 + \omega(\theta_B)^2]\sigma_\varepsilon^2 + [\omega(\eta_A)^2 + \omega(\eta_B)^2]\sigma_\eta^2.$$

To establish the result, we show that each of the three components of σ_z^2 are higher under the sequential system. Since we have previously shown that $\omega(\mu)$ is higher under simultaneous and given that $\omega(\theta_A) + \omega(\theta_B) = 1 - \omega(\mu)$, it follows that $[\omega(\theta_A) + \omega(\theta_B)]^2$ is higher under sequential than under simultaneous. The second component is also larger under sequential than under simultaneous since $\omega(\theta_s)^2$ is convex in $\omega(\theta_s)$ and since $\omega(\theta_A) + \omega(\theta_B)$ is higher under sequential than under simultaneous. Finally, the third component is larger under sequential than under simultaneous since $\omega(\eta_A)$ is higher under sequential than simultaneous and since $\omega(\eta_B) = 0.5$ under both systems. ■

PROOF OF PROPOSITION 3. Since $z = \omega(\theta_A)\theta_A + \omega(\theta_B)\theta_B + \omega(\mu)\mu_1 + \omega(\eta_A)\eta_A + \omega(\eta_B)\eta_B$, we have $z = (\omega(\theta_A) + \omega(\theta_B))q + \omega(\mu)\mu_1 + \omega(\eta_A)\eta_A + \omega(\eta_B)\eta_B + \omega(\theta_A)\epsilon_A + \omega(\theta_B)\epsilon_B$. Combining this with the fact that $y = q + \frac{1}{2}\eta_A + \frac{1}{2}\eta_B$ shows that $Cov(y, z) = (\omega(\theta_A) + \omega(\theta_B))\sigma_1^2 + \frac{1}{2}(\omega(\eta_A) + \omega(\eta_B))\sigma_\eta^2$.

Thus, $\rho_{y,z} = \frac{(\omega(\theta_A) + \omega(\theta_B))\sigma_1^2 + \frac{1}{2}(\omega(\eta_A) + \omega(\eta_B))\sigma_\eta^2}{\sigma_y\sqrt{(\omega(\theta_A) + \omega(\theta_B))^2\sigma_1^2 + [\omega(\theta_A)^2 + \omega(\theta_B)^2]\sigma_\varepsilon^2 + [\omega(\eta_A)^2 + \omega(\eta_B)^2]\sigma_\eta^2}}$, where σ_y denotes the standard deviation of the random variable $y = q + \frac{1}{2}\eta_A + \frac{1}{2}\eta_B$.

By substituting in the appropriate values for $\omega(\theta_A)$, $\omega(\theta_B)$, $\omega(\eta_A)$, and $\omega(\eta_B)$, we then see that

$$\rho_{y,z}^{sim} = \frac{\alpha_1\sigma_1^2 + \frac{1}{2}\sigma_\eta^2}{\sigma_y\sqrt{\alpha_1^2\sigma_1^2 + \frac{1}{2}\sigma_\eta^2 + \frac{1}{2}\alpha_1^2\sigma_\varepsilon^2}},$$

and

$$\rho_{y,z}^{seq} = \frac{\frac{1}{2}(\alpha_1 + \alpha_2 + (1-\alpha_2)\beta_1)\sigma_1^2 + \frac{1}{2}(1 + \frac{1}{2}(1-\alpha_2)\frac{\beta_1}{\alpha_1})\sigma_\eta^2}{\sigma_y\sqrt{\frac{1}{4}(\alpha_1 + \alpha_2 + (1-\alpha_2)\beta_1)^2\sigma_1^2 + \frac{1}{4}(1 + (1 + (1-\alpha_2)\frac{\beta_1}{\alpha_1})^2)\sigma_\eta^2 + \frac{1}{4}((\alpha_1 + (1-\alpha_2)\beta_1)^2 + \alpha_2^2)\sigma_\varepsilon^2}}.$$

Thus, in order to prove that

$$\rho_{y,z}^{sim} > \rho_{y,z}^{seq},$$

it suffices to prove that

$$\left(\alpha_1\sigma_1^2 + \frac{1}{2}\sigma_\eta^2\right)^2\left(\frac{1}{4}(\alpha_1 + \alpha_2 + (1-\alpha_2)\beta_1)^2\sigma_1^2 + \frac{1}{4}\left(1 + \left(1 + (1-\alpha_2)\frac{\beta_1}{\alpha_1}\right)^2\right)\sigma_\eta^2 + \frac{1}{4}((\alpha_1 + (1-\alpha_2)\beta_1)^2 + \alpha_2^2)\sigma_\varepsilon^2\right)$$
$$> \left(\frac{1}{2}(\alpha_1 + \alpha_2 + (1-\alpha_2)\beta_1)\sigma_1^2 + \frac{1}{2}\left(1 + \frac{1}{2}(1-\alpha_2)\frac{\beta_1}{\alpha_1}\right)\sigma_\eta^2\right)^2\left(\alpha_1^2\sigma_1^2 + \frac{1}{2}\sigma_\eta^2 + \frac{1}{2}\alpha_1^2\sigma_\varepsilon^2\right),$$

which is equivalent to proving

$$\left(\alpha_1^2\sigma_1^4 + \alpha_1\sigma_1^2\sigma_\eta^2 + \frac{1}{4}\sigma_\eta^4\right)\left((\alpha_1 + \alpha_2 + (1-\alpha_2)\beta_1)^2\sigma_1^2 + \left(1 + \left(1 + (1-\alpha_2)\frac{\beta_1}{\alpha_1}\right)^2\right)\sigma_\eta^2\right.$$
$$\left. + ((\alpha_1 + (1-\alpha_2)\beta_1)^2 + \alpha_2^2)\sigma_\varepsilon^2\right) > \left((\alpha_1 + \alpha_2 + (1-\alpha_2)\beta_1)^2\sigma_1^4 + 2(\alpha_1 + \alpha_2 + (1-\alpha_2)\beta_1)\right.$$

$$\times\left(1+\frac{1}{2}(1-\alpha_2)\frac{\beta_1}{\alpha_1}\right)\sigma_1^2\sigma_\eta^2+\left(1+\frac{1}{2}(1-\alpha_2)\frac{\beta_1}{\alpha_1}\right)^2\sigma_\eta^4\Bigg)\left(\alpha_1^2\sigma_1^2+\frac{1}{2}\sigma_\eta^2+\frac{1}{2}\alpha_1^2\sigma_\varepsilon^2\right).$$

Expanding this expression then indicates that it suffices to prove that

$$\begin{aligned}
&\alpha_1^2(\alpha_1+\alpha_2+(1-\alpha_2)\beta_1)^2\sigma_1^6+\frac{1}{4}\left(1+\left(1+(1-\alpha_2)\frac{\beta_1}{\alpha_1}\right)^2\right)\sigma_\eta^6\\
&+\left(\alpha_1(\alpha_1+\alpha_2+(1-\alpha_2)\beta_1)^2+\alpha_1^2\left(1+\left(1+(1-\alpha_2)\frac{\beta_1}{\alpha_1}\right)^2\right)\right)\sigma_1^4\sigma_\eta^2\\
&+\alpha_1^2((\alpha_1+(1-\alpha_2)\beta_1)^2+\alpha_2^2)\sigma_1^4\sigma_\varepsilon^2+\Bigg(\frac{1}{4}(\alpha_1+\alpha_2+(1-\alpha_2)\beta_1)^2\\
&+\alpha_1\left(1+\left(1+(1-\alpha_2)\frac{\beta_1}{\alpha_1}\right)^2\right)\Bigg)\sigma_1^2\sigma_\eta^4+\frac{1}{4}((\alpha_1+(1-\alpha_2)\beta_1)^2+\alpha_2^2)\sigma_\eta^4\sigma_\varepsilon^2\\
&+\alpha_1((\alpha_1+(1-\alpha_2)\beta_1)^2+\alpha_2^2)\sigma_1^2\sigma_\eta^2\sigma_\varepsilon^2\\
>\ &\alpha_1^2(\alpha_1+\alpha_2+(1-\alpha_2)\beta_1)^2\sigma_1^6+\frac{1}{2}\left(1+\frac{1}{2}(1-\alpha_2)\frac{\beta_1}{\alpha_1}\right)^2\sigma_\eta^6+\Bigg(\frac{1}{2}(\alpha_1+\alpha_2+(1-\alpha_2)\beta_1)^2\\
&+2\alpha_1^2(\alpha_1+\alpha_2+(1-\alpha_2)\beta_1)\left(1+\frac{1}{2}(1-\alpha_2)\frac{\beta_1}{\alpha_1}\right)\Bigg)\sigma_1^4\sigma_\eta^2+\frac{1}{2}\alpha_1^2(\alpha_1+\alpha_2+(1-\alpha_2)\beta_1)^2\sigma_1^4\sigma_\varepsilon^2\\
+\ &\left(\alpha_1^2\left(1+\frac{1}{2}(1-\alpha_2)\frac{\beta_1}{\alpha_1}\right)^2+(\alpha_1+\alpha_2+(1-\alpha_2)\beta_1)\left(1+\frac{1}{2}(1-\alpha_2)\frac{\beta_1}{\alpha_1}\right)\right)\sigma_1^2\sigma_\eta^4\\
&+\frac{1}{2}\alpha_1^2\left(1+\frac{1}{2}(1-\alpha_2)\frac{\beta_1}{\alpha_1}\right)^2\sigma_\eta^4\sigma_\varepsilon^2+\alpha_1^2(\alpha_1+\alpha_2+(1-\alpha_2)\beta_1)\left(1+\frac{1}{2}(1-\alpha_2)\frac{\beta_1}{\alpha_1}\right)\sigma_1^2\sigma_\eta^2\sigma_\varepsilon^2.
\end{aligned}$$

By collecting terms, we see that in order to prove this inequality, it suffices to prove the following:

(1) $\frac{1}{4}(1+(1+(1-\alpha_2)\frac{\beta_1}{\alpha_1})^2)\sigma_\eta^6>\frac{1}{2}(1+\frac{1}{2}(1-\alpha_2)\frac{\beta_1}{\alpha_1})^2\sigma_\eta^6$

(2) $\alpha_1^2((\alpha_1+(1-\alpha_2)\beta_1)^2+\alpha_2^2)\sigma_1^4\sigma_\varepsilon^2>\frac{1}{2}\alpha_1^2(\alpha_1+\alpha_2+(1-\alpha_2)\beta_1)^2\sigma_1^4\sigma_\varepsilon^2$

(3) $(\frac{1}{4}(\alpha_1+\alpha_2+(1-\alpha_2)\beta_1)^2+\alpha_1(1+(1+(1-\alpha_2)\frac{\beta_1}{\alpha_1})^2))\sigma_1^2\sigma_\eta^4+\frac{1}{4}((\alpha_1+(1-\alpha_2)\beta_1)^2+\alpha_2^2)\sigma_\eta^4\sigma_\varepsilon^2>(\alpha_1^2(1+\frac{1}{2}(1-\alpha_2)\frac{\beta_1}{\alpha_1})^2+(\alpha_1+\alpha_2+(1-\alpha_2)\beta_1)(1+\frac{1}{2}(1-\alpha_2)\frac{\beta_1}{\alpha_1}))\sigma_1^2\sigma_\eta^4+\frac{1}{2}\alpha_1^2(1+\frac{1}{2}(1-\alpha_2)\frac{\beta_1}{\alpha_1})^2\sigma_\eta^4\sigma_\varepsilon^2$

(4) $(\alpha_1(\alpha_1+\alpha_2+(1-\alpha_2)\beta_1)^2+\alpha_1^2(1+(1+(1-\alpha_2)\frac{\beta_1}{\alpha_1})^2))\sigma_1^4\sigma_\eta^2+\alpha_1((\alpha_1+(1-\alpha_2)\beta_1)^2+\alpha_2^2)\sigma_1^2\sigma_\eta^2\sigma_\varepsilon^2>(\frac{1}{2}(\alpha_1+\alpha_2+(1-\alpha_2)\beta_1)^2+2\alpha_1^2(\alpha_1+\alpha_2+(1-\alpha_2)\beta_1)(1+\frac{1}{2}(1-\alpha_2)\frac{\beta_1}{\alpha_1}))\sigma_1^4\sigma_\eta^2+\alpha_1^2(\alpha_1+\alpha_2+(1-\alpha_2)\beta_1)(1+\frac{1}{2}(1-\alpha_2)\frac{\beta_1}{\alpha_1})\sigma_1^2\sigma_\eta^2\sigma_\varepsilon^2$.

We prove each of these in turn:

To prove (1), note that $\frac{1}{4}(1+(1+(1-\alpha_2)\frac{\beta_1}{\alpha_1})^2)\sigma_\eta^6>\frac{1}{2}(1+\frac{1}{2}(1-\alpha_2)\frac{\beta_1}{\alpha_1})^2\sigma_\eta^6\Leftrightarrow(1+(1+(1-\alpha_2)\frac{\beta_1}{\alpha_1})^2)>2(1+\frac{1}{2}(1-\alpha_2)\frac{\beta_1}{\alpha_1})^2\Leftrightarrow 1+1+2(1-\alpha_2)\frac{\beta_1}{\alpha_1}+(1-\alpha_2)^2\frac{\beta_1^2}{\alpha_1^2}>2(1+(1-\alpha_2)\frac{\beta_1}{\alpha_1}+\frac{1}{4}(1-\alpha_2)^2\frac{\beta_1^2}{\alpha_1^2})\Leftrightarrow\frac{1}{2}(1-\alpha_2)^2\frac{\beta_1^2}{\alpha_1^2}>0$, which holds. Thus, (1) holds.

To prove (2), note that $\alpha_1^2((\alpha_1+(1-\alpha_2)\beta_1)^2+\alpha_2^2)\sigma_1^4\sigma_\varepsilon^2>\frac{1}{2}\alpha_1^2(\alpha_1+\alpha_2+(1-\alpha_2)\beta_1)^2\sigma_1^4\sigma_\varepsilon^2\Leftrightarrow(\alpha_1+(1-\alpha_2)\beta_1)^2+\alpha_2^2>\frac{1}{2}(\alpha_1+\alpha_2+(1-\alpha_2)\beta_1)^2\Leftrightarrow\alpha_1^2+2\alpha_1(1-\alpha_2)\beta_1+(1-\alpha_2)^2\beta_1^2+\alpha_2^2>\frac{1}{2}\alpha_1^2+\frac{1}{2}\alpha_2^2+\frac{1}{2}(1-\alpha_2)^2\beta_1^2+\alpha_1\alpha_2+\alpha_1(1-\alpha_2)\beta_1+\alpha_2(1-\alpha_2)\beta_1\Leftrightarrow\frac{1}{2}\alpha_1^2+\frac{1}{2}\alpha_2^2-\alpha_1\alpha_2+\alpha_1(1-\alpha_2)\beta_1-\alpha_2(1-\alpha_2)\beta_1+\frac{1}{2}(1-\alpha_2)^2\beta_1^2>0\Leftrightarrow\frac{1}{2}(\alpha_1-\alpha_2)^2+(\alpha_1-\alpha_2)(1-\alpha_2)\beta_1+\frac{1}{2}(1-\alpha_2)^2\beta_1^2>0$, which holds. Thus, (2) holds as well.

To prove (3), first note that

$$\left(\frac{1}{4}(\alpha_1+\alpha_2+(1-\alpha_2)\beta_1)^2+\alpha_1\left(1+\left(1+(1-\alpha_2)\frac{\beta_1}{\alpha_1}\right)^2\right)\right)\sigma_1^2\sigma_\eta^4$$
$$-\left(\alpha_1^2\left(1+\frac{1}{2}(1-\alpha_2)\frac{\beta_1}{\alpha_1}\right)^2+(\alpha_1+\alpha_2+(1-\alpha_2)\beta_1)\left(1+\frac{1}{2}(1-\alpha_2)\frac{\beta_1}{\alpha_1}\right)\right)\sigma_1^2\sigma_\eta^4$$

is equal to

$$\left(\frac{1}{4}\left(\alpha_1^2+\alpha_2^2+(1-\alpha_2)^2\beta_1^2+2\alpha_1\alpha_2+2\alpha_1(1-\alpha_2)\beta_1+2\alpha_2(1-\alpha_2)\beta_1\right)\right.$$
$$+\alpha_1\left(2+2(1-\alpha_2)\frac{\beta_1}{\alpha_1}+(1-\alpha_2)^2\frac{\beta_1^2}{\alpha_1^2}\right)-\alpha_1^2\left(1+(1-\alpha_2)\frac{\beta_1}{\alpha_1}+\frac{1}{4}(1-\alpha_2)^2\frac{\beta_1^2}{\alpha_1^2}\right)$$
$$\left.-\alpha_1-\alpha_2-(1-\alpha_2)\beta_1-\frac{1}{2}(1-\alpha_2)\beta_1-\frac{1}{2}\alpha_2(1-\alpha_2)\frac{\beta_1}{\alpha_1}-\frac{1}{2}(1-\alpha_2)^2\frac{\beta_1^2}{\alpha_1}\right)\sigma_1^2\sigma_\eta^4,$$

which is, in turn, equal to

$$\left(\frac{1}{4}\alpha_1^2+\frac{1}{4}\alpha_2^2+\frac{1}{4}(1-\alpha_2)^2\beta_1^2+\frac{1}{2}\alpha_1\alpha_2+\frac{1}{2}\alpha_1(1-\alpha_2)\beta_1+\frac{1}{2}\alpha_2(1-\alpha_2)\beta_1+2\alpha_1\right.$$
$$+2(1-\alpha_2)\beta_1+(1-\alpha_2)^2\frac{\beta_1^2}{\alpha_1}-\alpha_1^2-\alpha_1(1-\alpha_2)\beta_1-\frac{1}{4}(1-\alpha_2)^2\beta_1^2-\alpha_1-\alpha_2$$
$$\left.-(1-\alpha_2)\beta_1-\frac{1}{2}(1-\alpha_2)\beta_1-\frac{1}{2}\alpha_2(1-\alpha_2)\frac{\beta_1}{\alpha_1}-\frac{1}{2}(1-\alpha_2)^2\frac{\beta_1^2}{\alpha_1}\right)\sigma_1^2\sigma_\eta^4,$$

and this equals

$$\left(-\frac{3}{4}\alpha_1^2+\frac{1}{4}\alpha_2^2+\frac{1}{2}\alpha_1\alpha_2-\frac{1}{2}\alpha_1(1-\alpha_2)\beta_1+\alpha_1+\frac{1}{2}(1-\alpha_2)\beta_1+\frac{1}{2}\alpha_2(1-\alpha_2)\beta_1\right.$$
$$\left.-\frac{1}{2}\alpha_2(1-\alpha_2)\frac{\beta_1}{\alpha_1}+\frac{1}{2}(1-\alpha_2)^2\frac{\beta_1^2}{\alpha_1}-\alpha_2\right)\sigma_1^2\sigma_\eta^4,$$

which, in turn, equals

$$\left(\alpha_1-\alpha_2-\frac{3\alpha_1}{4}(\alpha_1-\alpha_2)-\frac{\alpha_2}{4}(\alpha_1-\alpha_2)+\frac{1}{2}(1-\alpha_1)(1-\frac{\alpha_2}{\alpha_1})(1-\alpha_2)\beta_1+\frac{1}{2}(1-\alpha_2)^2\frac{\beta_1^2}{\alpha_1}\right)\sigma_1^2\sigma_\eta^4,$$

and this is then equal to

$$\left(\left(1-\frac{3\alpha_1}{4}-\frac{\alpha_2}{4}\right)(\alpha_1-\alpha_2)+\frac{1}{2}(1-\alpha_1)(1-\frac{\alpha_2}{\alpha_1})(1-\alpha_2)\beta_1+\frac{1}{2}(1-\alpha_2)^2\frac{\beta_1^2}{\alpha_1}\right)\sigma_1^2\sigma_\eta^4$$
$$>(1-\alpha_1)(\alpha_1-\alpha_2)\sigma_1^2\sigma_\eta^4.$$

Also note that $\frac{1}{4}((\alpha_1+(1-\alpha_2)\beta_1)^2+\alpha_2^2)\sigma_\eta^4\sigma_\varepsilon^2-\frac{1}{2}\alpha_1^2(1+\frac{1}{2}(1-\alpha_2)\frac{\beta_1}{\alpha_1})^2\sigma_\eta^4\sigma_\varepsilon^2=(\frac{1}{4}(\alpha_1^2+2\alpha_1(1-\alpha_2)\beta_1+(1-\alpha_2)^2\beta_1^2+\alpha_2^2)-\frac{1}{2}\alpha_1^2-\frac{1}{2}\alpha_1(1-\alpha_2)\beta_1-\frac{1}{8}(1-\alpha_2)^2\beta_1^2)\sigma_\eta^4\sigma_\varepsilon^2=(\frac{1}{8}(1-\alpha_2)^2\beta_1^2-\frac{1}{4}\alpha_1^2+\frac{1}{4}\alpha_2^2)\sigma_\eta^4\sigma_\varepsilon^2>(\frac{1}{4}\alpha_2^2-\frac{1}{4}\alpha_1^2)\sigma_\eta^4\sigma_\varepsilon^2=-\frac{1}{4}(\alpha_1+\alpha_2)(\alpha_1-\alpha_2)\sigma_\eta^4\sigma_\varepsilon^2>-\frac{1}{2}\alpha_1(\alpha_1-\alpha_2)\sigma_\eta^4\sigma_\varepsilon^2$.

By combining the results in the previous two paragraphs, we see that the difference between the left-hand side and the right-hand side of the inequality in (3) is greater than $(1-\alpha_1)(\alpha_1-\alpha_2)\sigma_1^2\sigma_\eta^4 - \frac{1}{2}\alpha_1(\alpha_1-\alpha_2)\sigma_\eta^4\sigma_\varepsilon^2 = (\alpha_1-\alpha_2)\sigma_\eta^4((1-\alpha_1)\sigma_1^2 - \frac{1}{2}\alpha_1\sigma_\varepsilon^2) = (\alpha_1-\alpha_2)\sigma_\eta^4\frac{\sigma_1^2\sigma_\varepsilon^2}{2(\sigma_1^2+\sigma_\varepsilon^2)} > 0$. Thus, the inequality in (3) holds.

To prove (4), first note that

$$\left(\alpha_1(\alpha_1+\alpha_2+(1-\alpha_2)\beta_1)^2 + \alpha_1^2\left(1+\left(1+(1-\alpha_2)\frac{\beta_1}{\alpha_1}\right)^2\right)\right)\sigma_1^4\sigma_\eta^2$$
$$-\left(\frac{1}{2}(\alpha_1+\alpha_2+(1-\alpha_2)\beta_1)^2 + 2\alpha_1^2(\alpha_1+\alpha_2+(1-\alpha_2)\beta_1)\left(1+\frac{1}{2}(1-\alpha_2)\frac{\beta_1}{\alpha_1}\right)\right)\sigma_1^4\sigma_\eta^2$$

is equal to

$$\left(\alpha_1(\alpha_2-\alpha_1)(\alpha_1+\alpha_2+(1-\alpha_2)\beta_1) + \alpha_1^2\left(2+2(1-\alpha_2)\frac{\beta_1}{\alpha_1} + (1-\alpha_2)^2\frac{\beta_1^2}{\alpha_1^2}\right)\right.$$
$$\left.-\frac{1}{2}(\alpha_1^2+\alpha_2^2+(1-\alpha_2)^2\beta_1^2 + 2\alpha_1\alpha_2 + 2\alpha_1(1-\alpha_2)\beta_1 + 2\alpha_2(1-\alpha_2)\beta_1)\right)\sigma_1^4\sigma_\eta^2,$$

which, in turn, equals

$$\left(\alpha_1(\alpha_2-\alpha_1)(\alpha_1+\alpha_2+(1-\alpha_2)\beta_1) + \frac{3}{2}\alpha_1^2 - \alpha_1\alpha_2 - \frac{1}{2}\alpha_2^2 + \frac{1}{2}(1-\alpha_2)^2\beta_1^2 + (\alpha_1-\alpha_2)(1-\alpha_2)\beta_1\right)\sigma_1^4\sigma_\eta^2.$$

This is then equal to

$$\left((\alpha_1-\alpha_2)\left(-\alpha_1^2-\alpha_1\alpha_2-\alpha_1(1-\alpha_2)\beta_1\right) + \frac{3}{2}\alpha_1(\alpha_1-\alpha_2) + \frac{1}{2}\alpha_2(\alpha_1-\alpha_2)\right.$$
$$\left.+(1-\alpha_2)\beta_1(\alpha_1-\alpha_2) + \frac{1}{2}(1-\alpha_2)^2\beta_1^2\right)\sigma_1^4\sigma_\eta^2,$$

which also equals

$$\left((\alpha_1-\alpha_2)\left(\frac{3}{2}\alpha_1 + \frac{1}{2}\alpha_2 - \alpha_1^2 - \alpha_1\alpha_2 + (1-\alpha_1)(1-\alpha_2)\beta_1\right) + \frac{1}{2}(1-\alpha_2)^2\beta_1^2\right)\sigma_1^4\sigma_\eta^2,$$

which is, in turn, equal to

$$\left((\alpha_1-\alpha_2)\left(\alpha_1(1-\alpha_1) + \frac{1}{2}\alpha_1(1-\alpha_2) + \frac{1}{2}\alpha_2(1-\alpha_1) + (1-\alpha_1)(1-\alpha_2)\beta_1\right) + \frac{1}{2}(1-\alpha_2)^2\beta_1^2\right)\sigma_1^4\sigma_\eta^2.$$

From this, it follows that this expression is greater than

$$2\alpha_2(\alpha_1-\alpha_2)(1-\alpha_1)\sigma_1^4\sigma_\eta^2. \tag{A.1}$$

Also, note that

$$\alpha_1((\alpha_1+(1-\alpha_2)\beta_1)^2+\alpha_2^2)\sigma_1^2\sigma_\eta^2\sigma_\varepsilon^2 - \alpha_1^2(\alpha_1+\alpha_2+(1-\alpha_2)\beta_1)\left(1+\frac{1}{2}(1-\alpha_2)\frac{\beta_1}{\alpha_1}\right)\sigma_1^2\sigma_\eta^2\sigma_\varepsilon^2$$
$$=\alpha_1\left((\alpha_1+(1-\alpha_2)\beta_1)^2+\alpha_2^2-\alpha_1(\alpha_1+\alpha_2+(1-\alpha_2)\beta_1)\left(1+\frac{1}{2}(1-\alpha_2)\frac{\beta_1}{\alpha_1}\right)\right)\sigma_1^2\sigma_\eta^2\sigma_\varepsilon^2$$

$$= \alpha_1 \bigg(\alpha_1^2 + 2\alpha_1(1-\alpha_2)\beta_1 + (1-\alpha_2)^2\beta_1^2 + \alpha_2^2 - (\alpha_1+\alpha_2+(1-\alpha_2)\beta_1)\Big(\alpha_1 + \frac{1}{2}(1-\alpha_2)\beta_1\Big)\bigg)\sigma_1^2\sigma_\eta^2\sigma_\varepsilon^2,$$

which is, in turn, equal to

$$\alpha_1 \bigg(\alpha_1^2 + 2\alpha_1(1-\alpha_2)\beta_1 + (1-\alpha_2)^2\beta_1^2 + \alpha_2^2 - \alpha_1^2 - \alpha_1\alpha_2 - \alpha_1(1-\alpha_2)\beta_1 - \frac{1}{2}\alpha_1(1-\alpha_2)\beta_1 - \frac{1}{2}\alpha_2(1-\alpha_2)\beta_1 - \frac{1}{2}(1-\alpha_2)^2\beta_1^2\bigg)\sigma_1^2\sigma_\eta^2\sigma_\varepsilon^2,$$

which then equals

$$\alpha_1\bigg(\frac{1}{2}\alpha_1(1-\alpha_2)\beta_1 - \frac{1}{2}\alpha_2(1-\alpha_2)\beta_1 + \frac{1}{2}(1-\alpha_2)^2\beta_1^2 + \alpha_2^2 - \alpha_1\alpha_2\bigg)\sigma_1^2\sigma_\eta^2\sigma_\varepsilon^2$$
$$= \alpha_1\bigg(\frac{1}{2}(\alpha_1-\alpha_2)(1-\alpha_2)\beta_1 + \frac{1}{2}(1-\alpha_2)^2\beta_1^2 - \alpha_2(\alpha_1-\alpha_2)\bigg)\sigma_1^2\sigma_\eta^2\sigma_\varepsilon^2.$$

Thus, this expression is greater than

$$-\alpha_1\alpha_2(\alpha_1-\alpha_2)\sigma_1^2\sigma_\eta^2\sigma_\varepsilon^2. \tag{A.2}$$

By combining inequalities that were derived in expressions (A.1) and (A.2), we see that the difference between the left-hand side and the right-hand side of the inequality in (4) is greater than $2\alpha_2(\alpha_1-\alpha_2)(1-\alpha_1)\sigma_1^4\sigma_\eta^2 - \alpha_1\alpha_2(\alpha_1-\alpha_2)\sigma_1^2\sigma_\eta^2\sigma_\varepsilon^2$. But this difference is equal to $\alpha_2(\alpha_1-\alpha_2)\sigma_1^2\sigma_\eta^2(2(1-\alpha_1)\sigma_1^2 - \alpha_1\sigma_\varepsilon^2) = \alpha_2(\alpha_1-\alpha_2)\sigma_1^2\sigma_\eta^2\frac{\sigma_1^2\sigma_\varepsilon^2}{\sigma_1^2+\sigma_\varepsilon^2} > 0$. Thus, the inequality in (4) holds, and from this, it follows that $\rho_{y,z}^{sim} > \rho_{y,z}^{seq}$. ■

PROOF OF PROPOSITION 4. First, note that in the limit as $\mu_1 \to 0$, $\Delta \to [\rho_{y,z}^{sim} - \rho_{y,z}^{seq}]\sigma_y\phi(0) > 0$. Thus, $\Delta > 0$ for sufficiently small values of μ_1.

Also, note that for general values of μ_1, we have

$$\begin{aligned}
\Delta &= \mu_1\left[\Phi\left(\frac{\mu_1}{\sigma_z^{sim}}\right) - \Phi\left(\frac{\mu_1}{\sigma_z^{seq}}\right)\right] + \rho_{y,z}^{sim}\sigma_y\phi\left(\frac{\mu_1}{\sigma_z^{sim}}\right) - \rho_{y,z}^{seq}\sigma_y\phi\left(\frac{\mu_1}{\sigma_z^{seq}}\right)\\
&= \frac{1}{\sqrt{2\pi}}\left[\mu_1\int_{\mu_1/\sigma_z^{seq}}^{\mu_1/\sigma_z^{sim}} e^{-x^2/2}\,dx + \rho_{y,z}^{sim}\sigma_y e^{-\mu_1^2/2(\sigma_z^{sim})^2} - \rho_{y,z}^{seq}\sigma_y e^{-\mu_1^2/2(\sigma_z^{seq})^2}\right]\\
&= \frac{1}{\sqrt{2\pi}}\left[\sigma_z^{seq}\int_{\mu_1/\sigma_z^{seq}}^{\mu_1/\sigma_z^{sim}} \frac{\mu_1}{\sigma_z^{seq}} e^{-x^2/2}\,dx + \rho_{y,z}^{sim}\sigma_y e^{-\mu_1^2/2(\sigma_z^{sim})^2} - \rho_{y,z}^{seq}\sigma_y e^{-\mu_1^2/2(\sigma_z^{seq})^2}\right]\\
&\le \frac{1}{\sqrt{2\pi}}\left[\sigma_z^{seq}\int_{\mu_1/\sigma_z^{seq}}^{\mu_1/\sigma_z^{sim}} x e^{-x^2/2}\,dx + \rho_{y,z}^{sim}\sigma_y e^{-\mu_1^2/2(\sigma_z^{sim})^2} - \rho_{y,z}^{seq}\sigma_y e^{-\mu_1^2/2(\sigma_z^{seq})^2}\right]\\
&= \frac{1}{\sqrt{2\pi}}\left[\sigma_z^{seq}e^{-\mu_1^2/2(\sigma_z^{seq})^2} - \sigma_z^{seq}e^{-\mu_1^2/2(\sigma_z^{sim})^2} + \rho_{y,z}^{sim}\sigma_y e^{-\mu_1^2/2(\sigma_z^{sim})^2} - \rho_{y,z}^{seq}\sigma_y e^{-\mu_1^2/2(\sigma_z^{seq})^2}\right]\\
&= \frac{1}{\sqrt{2\pi}}\left[(\sigma_z^{seq} - \rho_{y,z}^{seq}\sigma_y)e^{-\mu_1^2/2(\sigma_z^{seq})^2} + (\rho_{y,z}^{sim}\sigma_y - \sigma_z^{seq})e^{-\mu_1^2/2(\sigma_z^{sim})^2}\right].
\end{aligned}$$

Now in the limit as $\mu_1 \to \infty$, $\frac{e^{-\mu_1^2/2(\sigma_z^{sim})^2}}{e^{-\mu_1^2/2(\sigma_z^{seq})^2}} \to 0$ since $(\sigma_z^{sim})^2 < (\sigma_z^{seq})^2$. Thus, if $\sigma_z^{seq} - \rho_{y,z}^{seq}\sigma_y < 0$, then it follows that $\Delta < 0$ for sufficiently large μ_1. But $\sigma_z^{seq} - \rho_{y,z}^{seq}\sigma_y < 0$ holds if and only if $\rho_{y,z}^{seq}\sigma_y > \sigma_z^{seq}$, which, in turn, holds if and only if $\frac{\mathrm{Cov}(y, z^{seq})}{\sigma_z^{seq}} > \sigma_z^{seq}$ or $\mathrm{Cov}(y, z^{seq}) > (\sigma_z^{seq})^2$.

Now $\mathrm{Cov}(y, z^{seq}) = \frac{1}{2}(\alpha_1 + \alpha_2 + (1-\alpha_2)\beta_1)\sigma_1^2 + \frac{1}{2}(1 + \frac{1}{2}(1-\alpha_2)\frac{\beta_1}{\alpha_1})\sigma_\eta^2$ and $(\sigma_z^{seq})^2 = \frac{1}{4}(\alpha_1 + \alpha_2 + (1-\alpha_2)\beta_1)^2\sigma_1^2 + \frac{1}{4}(1 + (1 + (1-\alpha_2)\frac{\beta_1}{\alpha_1})^2)\sigma_\eta^2 + \frac{1}{4}((\alpha_1 + (1-\alpha_2)\beta_1)^2 + \alpha_2^2)\sigma_\varepsilon^2$. Thus, $\mathrm{Cov}(y, z^{seq}) > (\sigma_z^{seq})^2$ holds if and only if

$$\begin{aligned}
&\frac{1}{2}\left(\alpha_1 + \alpha_2 + (1-\alpha_2)\beta_1\right)\sigma_1^2 + \frac{1}{2}\left(1 + \frac{1}{2}(1-\alpha_2)\frac{\beta_1}{\alpha_1}\right)\sigma_\eta^2 \\
&\quad > \frac{1}{4}\left(\alpha_1 + \alpha_2 + (1-\alpha_2)\beta_1\right)^2\sigma_1^2 + \frac{1}{4}\left(1 + \left(1 + (1-\alpha_2)\frac{\beta_1}{\alpha_1}\right)^2\right)\sigma_\eta^2 \\
&\quad + \frac{1}{4}((\alpha_1 + (1-\alpha_2)\beta_1)^2 + \alpha_2^2)\sigma_\varepsilon^2.
\end{aligned}$$

Now the above equation holds if and only if

$$\begin{aligned}
&2(\alpha_1 + \alpha_2 + (1-\alpha_2)\beta_1)\sigma_1^2 + 2\left(1 + \frac{1}{2}(1-\alpha_2)\frac{\beta_1}{\alpha_1}\right)\sigma_\eta^2 \\
&> (\alpha_1 + \alpha_2 + (1-\alpha_2)\beta_1)^2\sigma_1^2 + \left(1 + \left(1 + (1-\alpha_2)\frac{\beta_1}{\alpha_1}\right)^2\right)\sigma_\eta^2 + ((\alpha_1 + (1-\alpha_2)\beta_1)^2 + \alpha_2^2)\sigma_\varepsilon^2.
\end{aligned}$$

This, in turn, holds if and only if

$$\begin{aligned}
&2(\alpha_1 + \alpha_2 + (1-\alpha_2)\beta_1)\sigma_1^2 > (\alpha_1 + \alpha_2 + (1-\alpha_2)\beta_1)^2\sigma_1^2 + \left((1-\alpha_2)\frac{\beta_1}{\alpha_1} + (1-\alpha_2)^2\frac{\beta_1^2}{\alpha_1^2}\right)\sigma_\eta^2 \\
&\quad + ((\alpha_1 + (1-\alpha_2)\beta_1)^2 + \alpha_2^2)\sigma_\varepsilon^2,
\end{aligned}$$

which holds if and only if

$$\begin{aligned}
2(\alpha_1 + \alpha_2 + (1-\alpha_2)\beta_1)\sigma_1^2 > {}& (\alpha_1^2 + \alpha_2^2 + (1-\alpha_2)^2\beta_1^2 + 2\alpha_1\alpha_2 + 2\alpha_1(1-\alpha_2)\beta_1 \\
&+ 2\alpha_2(1-\alpha_2)\beta_1)\sigma_1^2 + \left((1-\alpha_2)\frac{\beta_1}{\alpha_1} + (1-\alpha_2)^2\frac{\beta_1^2}{\alpha_1^2}\right)\sigma_\eta^2 \\
&+ (\alpha_1^2 + 2\alpha_1(1-\alpha_2)\beta_1 + (1-\alpha_2)^2\beta_1^2 + \alpha_2^2)\sigma_\varepsilon^2,
\end{aligned}$$

and this then holds if and only if

$$\begin{aligned}
(\alpha_1 + 2\alpha_2 + 2(1-\alpha_2)\beta_1)\sigma_1^2 > {}& (\alpha_2^2 + (1-\alpha_2)^2\beta_1^2 + 2\alpha_1\alpha_2 + 2\alpha_1(1-\alpha_2)\beta_1 \\
&+ 2\alpha_2(1-\alpha_2)\beta_1)\sigma_1^2 + \left((1-\alpha_2)\frac{\beta_1}{\alpha_1} + (1-\alpha_2)^2\frac{\beta_1^2}{\alpha_1^2}\right)\sigma_\eta^2 \\
&+ (2\alpha_1(1-\alpha_2)\beta_1 + (1-\alpha_2)^2\beta_1^2 + \alpha_2^2)\sigma_\varepsilon^2.
\end{aligned}$$

This holds if and only if

$$\begin{aligned}
(\alpha_1 + 2\alpha_2)\sigma_1^2 > {}& (\alpha_2^2 + (1-\alpha_2)^2\beta_1^2 + 2\alpha_1\alpha_2 + 2\alpha_2(1-\alpha_2)\beta_1)\sigma_1^2 \\
&+ \left((1-\alpha_2)\frac{\beta_1}{\alpha_1} + (1-\alpha_2)^2\frac{\beta_1^2}{\alpha_1^2}\right)\sigma_\eta^2 + ((1-\alpha_2)^2\beta_1^2 + \alpha_2^2)\sigma_\varepsilon^2,
\end{aligned}$$

which holds if and only if

$$(\alpha_1 + 2\alpha_2)\sigma_1^2 > (\alpha_2^2 + 2\alpha_1\alpha_2 + 2\alpha_2(1-\alpha_2)\beta_1)\sigma_1^2 + (1-\alpha_2)\frac{\beta_1}{\alpha_1}\sigma_\eta^2 + \alpha_2^2\sigma_\varepsilon^2 + (1-\alpha_2)^2\beta_1\sigma_1^2,$$

or if and only if

$$(\alpha_1 + 2\alpha_2)\sigma_1^2 > (\alpha_2^2 + 2\alpha_1\alpha_2 - \alpha_2^2\beta_1 + \beta_1)\sigma_1^2 + (1-\alpha_2)\frac{\beta_1}{\alpha_1}\sigma_\eta^2 + \alpha_2^2\sigma_\varepsilon^2.$$

This, in turn, holds if and only if

$$\begin{aligned}(\alpha_1 + 2\alpha_2)\sigma_1^2 > {} & (\alpha_2^2 + 2\alpha_1\alpha_2 - \alpha_2^2\beta_1 + \beta_1 - \alpha_1(1-\alpha_2)\beta_1)\sigma_1^2 + (\alpha_2^2 - \alpha_1(1-\alpha_2)\beta_1)\sigma_\varepsilon^2 \\ & + \alpha_1(1-\alpha_2)\beta_1\left(\sigma_1^2 + \frac{\sigma_\eta^2}{\alpha_1^2} + \sigma_\varepsilon^2\right),\end{aligned}$$

which holds if and only if

$$\begin{aligned}(\alpha_1 + 2\alpha_2)\sigma_1^2 > {} & (\alpha_2^2 + 2\alpha_1\alpha_2 - \alpha_2^2\beta_1 + \beta_1 - \alpha_1(1-\alpha_2)\beta_1)\sigma_1^2 + (\alpha_2^2 - \alpha_1(1-\alpha_2)\beta_1)\sigma_\varepsilon^2 \\ & + \alpha_1(1-\alpha_2)\sigma_1^2.\end{aligned}$$

Finally, this holds if and only if

$$\begin{aligned}& 2\alpha_2\sigma_1^2 > (\alpha_2^2 + \alpha_1\alpha_2 - \alpha_2^2\beta_1 + \beta_1 - \alpha_1\beta_1 + \alpha_1\alpha_2\beta_1)\sigma_1^2 + (\alpha_2^2 - \alpha_1\beta_1 + \alpha_1\alpha_2\beta_1)\sigma_\varepsilon^2 \Leftrightarrow \\ & 2\alpha_2\sigma_1^2 > (\alpha_1\alpha_2 - \alpha_2^2\beta_1 + \beta_1)\sigma_1^2 + (\alpha_2^2 - \alpha_1\beta_1 + \alpha_1\alpha_2\beta_1)(\sigma_1^2 + \sigma_\varepsilon^2) \Leftrightarrow \\ & 2\alpha_2\sigma_1^2 > \left(\alpha_1\alpha_2 - \alpha_2^2\beta_1 + \beta_1 + \frac{\alpha_2^2}{\alpha_1} - \beta_1 + \alpha_2\beta_1\right)\sigma_1^2 \Leftrightarrow \\ & 2\alpha_2 > \alpha_1\alpha_2 - \alpha_2^2\beta_1 + \frac{\alpha_2^2}{\alpha_1} + \alpha_2\beta_1 \Leftrightarrow \\ & 2 > \alpha_1 - \alpha_2\beta_1 + \frac{\alpha_2}{\alpha_1} + \beta_1.\end{aligned}$$

But

$$\begin{aligned}\alpha_1 - \alpha_2\beta_1 + \frac{\alpha_2}{\alpha_1} + \beta_1 &= \alpha_1 + \frac{\alpha_2}{\alpha_1} + (1-\alpha_2)\beta_1 = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_\varepsilon^2} + \frac{\sigma_2^2(\sigma_1^2 + \sigma_\varepsilon^2)}{\sigma_1^2(\sigma_2^2 + \sigma_\varepsilon^2)} + \frac{\sigma_\varepsilon^2\beta_1}{\sigma_2^2 + \sigma_\varepsilon^2} \\ &= \frac{\sigma_1^4(\sigma_2^2 + \sigma_\varepsilon^2) + \sigma_2^2(\sigma_1^2 + \sigma_\varepsilon^2)^2 + \sigma_1^2\sigma_\varepsilon^2(\sigma_1^2 + \sigma_\varepsilon^2)\beta_1}{\sigma_1^2(\sigma_1^2 + \sigma_\varepsilon^2)(\sigma_2^2 + \sigma_\varepsilon^2)} \\ &= \frac{\sigma_1^4\sigma_2^2 + \sigma_1^4\sigma_\varepsilon^2 + \sigma_1^4\sigma_2^2 + 2\sigma_1^2\sigma_2^2\sigma_\varepsilon^2 + \sigma_2^2\sigma_\varepsilon^4 + \sigma_1^4\sigma_\varepsilon^2\beta_1 + \sigma_1^2\sigma_\varepsilon^4\beta_1}{\sigma_1^4\sigma_2^2 + \sigma_1^4\sigma_\varepsilon^2 + \sigma_1^2\sigma_2^2\sigma_\varepsilon^2 + \sigma_1^2\sigma_\varepsilon^4} \\ &= 2 - \frac{(1-\beta_1)\sigma_1^4\sigma_\varepsilon^2 + (\sigma_1^2 - \sigma_2^2)\sigma_\varepsilon^4 + (1-\beta_1)\sigma_1^2\sigma_\varepsilon^4}{\sigma_1^4\sigma_2^2 + \sigma_1^4\sigma_\varepsilon^2 + \sigma_1^2\sigma_2^2\sigma_\varepsilon^2 + \sigma_1^2\sigma_\varepsilon^4} < 2.\end{aligned}$$

Thus, the last inequality in the previous paragraph holds, which, in turn, implies that $\mathrm{Cov}(y, z^{seq}) > (\sigma_z^{seq})^2$. Thus, $\Delta < 0$ for sufficiently large μ_1. ■

PROOF OF PROPOSITION 5. Consider a weighted voting system that places weight $0.5 + \epsilon$ on the votes of the later state and weight $0.5 - \epsilon$ on the votes of the early state for some small value of $\epsilon > 0$. In this case, by analogy to Equation (17), it follows that candidate A is elected if and only if

$$\frac{(0.5-\epsilon)\exp(E(q|I_A))}{1+\exp(E(q|I_A))} + \frac{(0.5+\epsilon)\exp(E(q|I_B))}{1+\exp(E(q|I_B))} > 0.5,$$

which, in turn, holds if and only if

$$0.5\exp(E(q|I_A)+E(q|I_B)) + \epsilon[\exp(E(q|I_B)) - \exp(E(q|I_A))] > 0.5.$$

From this, it follows that if $\epsilon > 0$ is sufficiently small, then changing from a system in which the votes of the two states are weighted equally to one which places weight $0.5+\epsilon$ on the votes of the later state and weight $0.5-\epsilon$ on the votes of the early state can only change the outcome of the election if $E(q|I_A)+E(q|I_B)$ is arbitrarily close to 0 and the election would have been a virtual tie when the votes of the two states are weighted equally. In this case, placing slightly greater weight on the votes of the later state will result in the election of candidate A if $E(q|I_B) = -E(q|I_A) > 0$ and the election of candidate B if $E(q|I_B) = -E(q|I_A) < 0$.

Now when there is no preference heterogeneity, $E(q|I_B) = (1-\alpha_2)\alpha_1\theta_A + \alpha_2\theta_B + (1-\alpha_2)(1-\alpha_1)\mu_1$ and electing candidate A is better for aggregate expected voter welfare if and only if $\frac{\sigma_1^2}{2\sigma_1^2+\sigma_\epsilon^2}(\theta_A+\theta_B) + \frac{\sigma_\epsilon^2}{2\sigma_1^2+\sigma_\epsilon^2}\mu_1 > 0$. Thus, if $E(q|I_B) = -E(q|I_A) > 0$, then $(1-\alpha_2)\alpha_1\theta_A + \alpha_2\theta_B + (1-\alpha_2)(1-\alpha_1)\mu_1 > 0$. This, in turn, implies that $\alpha_2\theta_B > -(1-\alpha_2)(\alpha_1\theta_A + (1-\alpha_1)\mu_1)$, and since $\alpha_2 = \frac{\sigma_1^2}{2\sigma_1^2+\sigma_\epsilon^2}$ and $1-\alpha_2 = \frac{\sigma_1^2+\sigma_\epsilon^2}{2\sigma_1^2+\sigma_\epsilon^2}$, this further implies that $\frac{\sigma_1^2}{2\sigma_1^2+\sigma_\epsilon^2}\theta_B > -(\frac{\sigma_1^2}{2\sigma_1^2+\sigma_\epsilon^2}\theta_A + \frac{\sigma_\epsilon^2}{2\sigma_1^2+\sigma_\epsilon^2}\mu_1)$, meaning $\frac{\sigma_1^2}{2\sigma_1^2+\sigma_\epsilon^2}(\theta_A+\theta_B) + \frac{\sigma_\epsilon^2}{2\sigma_1^2+\sigma_\epsilon^2}\mu_1 > 0$ and candidate A is better for aggregate expected voter welfare.

Similarly, if $E(q|I_B) = -E(q|I_A) < 0$, then candidate B is better for aggregate voter welfare. Thus, regardless of whether $E(q|I_B) = -E(q|I_A) > 0$ or $E(q|I_B) = -E(q|I_A) < 0$, it follows that changing from a system in which the votes of the two states are weighted equally to one which places weight $0.5+\epsilon$ on the votes of the later state and weight $0.5-\epsilon$ on the votes of the early state would result in the election of the candidate that is best for expected voter welfare in the cases where this change can make a difference in the outcome of the election. The result then follows. ■

PROOF OF PROPOSITION 6. By rearranging Equation (18), we note that the front-runner is elected if and only if the following condition is satisfied:

$$z = 0.5E(q|I_A) + 0.5E(q|I_B) + 0.5(\eta_A - \eta) + 0.5(\eta_B + \eta) > 0.$$

Under simultaneous elections, whether this holds for particular realizations of $\eta_A - \eta$ and $\eta_B + \eta$ is independent of η. Under sequential elections, by the same logic used to derive Equation (9), we know that the posterior distribution of voters in the second state after they have observed the results of the first state but before they have considered their own signals is normal with mean

$$\mu_2 = \mu_1 + \frac{\beta_1}{\alpha_1}[\ln(v_1/v_0) - \mu_1 - \eta]$$

and variance given by the formula in Equation (10). By substituting in the equation for transformed vote shares as a noisy signal of quality given by (8), it then follows that

$$\mu_2 = \mu_1 + \beta_1\left[q - \mu_1 + \frac{\eta_A - \eta}{\alpha_1} + \epsilon_A\right].$$

But since $\theta_A = q + \epsilon_A$, it then follows that

$$\mu_2 = \mu_1 + \beta_1 \left[\theta_A - \mu_1 + \frac{\eta_A - \eta}{\alpha_1} \right] = (1 - \beta_1)\mu_1 + \beta_1 \theta_A + \frac{\beta_1}{\alpha_1}(\eta_A - \eta)$$

and

$$E[q|I_B] = \alpha_2 \theta_B + (1 - \alpha_2) \left[(1 - \beta_1)\mu_1 + \beta_1 \theta_A + \frac{\beta_1}{\alpha_1}(\eta_A - \eta) \right].$$

By using this and the fact that $E[q|I_A] = \alpha_1 \theta_A + (1 - \alpha_1)\mu_1$, it then follows that under the sequential system, candidate A is elected if and only if the following condition is satisfied:

$$z = [(0.5\alpha_1 + 0.5(1 - \alpha_2)\beta_1)\theta_A + 0.5\alpha_2\theta_B + (0.5(1 - \alpha_1) + 0.5(1 - \alpha_2)(1 - \beta_1))\mu_1 + (0.5 + 0.5(1 - \alpha_2)(\beta_1/\alpha_1))(\eta_A - \eta) + 0.5(\eta_B + \eta) > 0].$$

Whether this condition is satisfied for any particular realizations of $\eta_A - \eta$ and $\eta_B + \eta$ is independent of η. Furthermore, aggregate voter utility is equal to $y = q + 0.5(\eta_A - \eta) + 0.5(\eta_B + \eta)$, meaning that the value of aggregate utility for any particular realizations of $\eta_A - \eta$ and $\eta_B + \eta$ is independent of η. Finally, the distribution of $\eta_A - \eta$ and $\eta_B + \eta$ is also independent of η under this new model.

Thus, we have seen that whether a candidate is elected under either simultaneous elections or sequential elections for particular realizations of $\eta_A - \eta$ and $\eta_B + \eta$ is independent of η, aggregate utility for any particular realizations of $\eta_A - \eta$ and $\eta_B + \eta$ is independent of η, and the distributions of $\eta_A - \eta$ and $\eta_B + \eta$ are also independent of η. By combining these facts, it then follows that whether simultaneous elections are welfare-preferred to sequential elections is independent of η in this model. ■

PROOF OF PROPOSITION 7. By rearranging Equation (18), we can rewrite the probability that the front-runner is elected as

$$\Pr[z = 0.5E(q|I_A) + 0.5E(q|I_B) + 0.5(\eta_A - (v_A(c_1) - v_A(c_0))) + 0.5(\eta_B + (v_A(c_1) - v_A(c_0))) > 0].$$

Now under the simultaneous system, $0.5E(q|I_A)$, $0.5E(q|I_B)$, and the distribution of $\eta_A - (v_A(c_1) - v_A(c_0))$ are independent of $v_A(c_1) - v_A(c_0)$, but the distribution of $(\eta_B + (v_A(c_1) - v_A(c_0))$ is normal with mean $v_A(c_1) + v_B(C - c_1) - v_A(c_0) - v_B(C - c_0)$ and variance σ_η^2. Thus, in order to maximize their probabilities of being elected, the front-runner should choose c_1 in such a way to make $v_A(c_1) + v_B(C - c_1) - v_A(c_0) - v_B(C - c_0)$ as large as possible, and the dark-horse candidate should choose c_0 in such a way to make $v_A(c_1) + v_B(C - c_1) - v_A(c_0) - v_B(C - c_0)$ as small as possible. For both candidates, this then involves choosing an amount of money to spend in the first state c so that $v_A(c) + v_B(C - c)$ is as large as possible. But this means that both candidates will choose to spend the same amount of money in the first state and the overall distribution of voter preferences in each state will be the same as in the original model. Thus endogenous candidate strategies have no effect on voter preferences under simultaneous elections.

Under sequential elections, by the same logic used to derive Equation (9), we know that the posterior distribution of voters in the second state after they have observed the results of the first state but before they have considered their own signals is normal with mean

$$\mu_2 = \mu_1 + \frac{\beta_1}{\alpha_1}[\ln(v_1/v_0) - \mu_1 - (v_A(c_1) - v_A(c_0))],$$

and variance given by the formula in Equation (10). By substituting in the equation for transformed vote shares as a noisy signal of quality given by (8), it then follows that

$$\mu_2 = \mu_1 + \beta_1 \left[q - \mu_1 + \frac{\eta_A - (v_A(c_1) - v_A(c_0))}{\alpha_1} + \epsilon_A \right].$$

But since $\theta_A = q + \epsilon_A$, it then follows that

$$\mu_2 = \mu_1 + \beta_1 \left[\theta_A - \mu_1 + \frac{\eta_A - (v_A(c_1) - v_A(c_0))}{\alpha_1} \right] = (1-\beta_1)\mu_1 + \beta_1\theta_A + \frac{\beta_1}{\alpha_1}(\eta_A - (v_A(c_1) - v_A(c_0))),$$

and

$$E[q|I_B] = \alpha_2\theta_B + (1-\alpha_2)\left[(1-\beta_1)\mu_1 + \beta_1\theta_1 + \frac{\beta_1}{\alpha_1}(\eta_A - (v_A(c_1) - v_A(c_0))) \right].$$

By using this and the fact that $E[q|I_A] = \alpha_1\theta_A + (1-\alpha_1)\mu_1$, it then follows that under the sequential system, the probability that candidate A is elected is given by the following formula:

$$\begin{aligned} P^{seq} = \Pr[&(0.5\alpha_1 + 0.5(1-\alpha_2)\beta_1)\theta_A + 0.5\alpha_2\theta_B + (0.5(1-\alpha_1) + 0.5(1-\alpha_2)(1-\beta_1))\mu_1 \\ &+ (0.5 + 0.5(1-\alpha_2)(\beta_1/\alpha_1))(\eta_A - (v_A(c_1) - v_A(c_0))) + 0.5(\eta_B + (v_A(c_1) - v_A(c_0))) > 0]. \end{aligned}$$

Now under the sequential system, θ_A, θ_B, and the distribution of $\eta_A - (v_A(c_1) - v_A(c_0))$ are all independent of $v_A(c_1) - v_A(c_0)$, but the distribution of $(\eta_B + (v_A(c_1) - v_A(c_0))$ is normal with mean $v_A(c_1) + v_B(C - c_1) - v_A(c_0) - v_B(C - c_0)$ and variance σ_η^2. Thus, in order to maximize their probabilities of being elected, the front-runner should choose c_1 in such a way to make $v_A(c_1) + v_B(C - c_1) - v_A(c_0) - v_B(C - c_0)$ as large as possible, and the dark-horse candidate should choose c_0 in such a way to make $v_A(c_1) + v_B(C - c_1) - v_A(c_0) - v_B(C - c_0)$ as small as possible. For both candidates, this then involves choosing an amount of money to spend in the first state c so that $v_A(c) + v_B(C - c)$ is as large as possible. But this means that both candidates will choose to spend the same amount of money in the first state and the overall distribution of voter preferences in each state will be the same as in the original model. Thus, endogenous candidate strategies have no effect on voter preferences under sequential elections either. The result then follows. ■

PROOF OF PROPOSITION 8. In the limit as the number of states becomes arbitrarily large, the average state-level preference for candidate c, $\overline{\eta}_c$, becomes arbitrarily close to zero so expected voter welfare becomes arbitrarily close to the expected quality of the elected candidate. Thus, the question of which system is welfare-preferred reduces to the question of which system results in a higher expected quality of the elected candidate.

Now when $\mu_{c1} = 0$ for all c, then in the limit as the number of states becomes arbitrarily large, the probability that the candidate that is actually the highest quality candidate is elected goes to one. Thus, expected welfare under the simultaneous system approaches the expected quality of the highest quality candidate in the limit as the number of states becomes arbitrarily large. But under sequential elections, there is always some probability bounded away from zero that the candidate that is actually the best candidate will not be one of the two candidates that has received the most votes after m states have voted. Thus, under sequential elections, there is always some probability bounded away from zero that the candidate that is actually the best candidate will not win the election. From this, it follows that expected welfare remains bounded away from the expected quality of the highest quality candidate under sequential elections.

These results imply that simultaneous elections are strictly welfare-preferred to sequential elections when $\mu_{c1} = 0$ for all c if there is a sufficiently large number of states. But the expected welfare from either system varies continuously with the values of the priors μ_{c1}. From this, it follows that if voter priors about the likely quality of candidate c, μ_{c1}, are sufficiently close to 0 for all c, then simultaneous elections are welfare-preferred to sequential elections if there is a sufficiently large number of states. ■

PROOF OF PROPOSITION 9. In the limit as the number of states becomes arbitrarily large, the average state-level preference for candidate c, $\overline{\eta}_c$, becomes arbitrarily close to zero, so expected voter welfare becomes arbitrarily close to the expected quality of the elected candidate. Thus, the question of which system is welfare-preferred reduces to the question of which system results in a higher expected quality of the elected candidate.

Now under sequential elections, voters in later states eventually learn the true qualities of the two candidates that voters vote for in early states with arbitrary precision, so the probability that the majority of voters in states who vote after the first m states vote for the higher quality candidate of these two candidates goes to one for sufficiently large m. From this, it follows that if there is a sufficiently large number of states, then the probability that the majority of all voters vote for the higher quality candidate of these two candidates also goes to one in the limit. Thus, expected welfare under the sequential system approaches the expected quality of the highest quality candidate of these two candidates in the limit as the number of states becomes large.

But under simultaneous elections, if the candidate among the two candidates that voters vote for that initially has the lower prior is only a slightly higher quality candidate than the other such candidate, then the majority of voters vote for the lower quality candidate because all voters place positive weight on their priors and therefore act as if this candidate is a better candidate than the candidate actually is. Thus, the lower quality candidate will win with probability bounded away from zero in the limit as the number of states becomes arbitrarily large, and expected welfare remains bounded away from the expected quality of the highest quality candidate of these two candidates under simultaneous elections in the limit as the number of states becomes large. The result then follows. ■

Analysis of Candidates from 2008 and 2012 Elections: This section describes how the mean of candidate quality can be estimated using only aggregate polling data during the period preceding the primary season. Note first that according to the model, the vote share for candidate c in state s at time $t = 0$ is given by

$$v_{cs0} = \frac{\exp(\eta_{cs} + \mu_c)}{1 + \sum_d \exp(\eta_{ds} + \mu_d)}.$$

Note further that national support for candidate c time $t = 0$ is then given by

$$v_{c0} = (1/S) \sum_s v_{cs0},$$

where S is the total number of states. Then, expected national support is given by

$$E(v_{c0}) = (1/S) \sum_s E(v_{cs0}).$$

Given a value of μ_c for each candidate, one can then calculate $E(v_{cs0})$ by simulating draws of η_{cs} from the normal distribution. Following this procedure, we then choose the values of μ_c for each candidate, relative to a baseline candidate, which best match the observed national vote share to the predicted national vote shares.

To measure national support for the three candidates during the 2008 and 2012 primary seasons, we average all national polls conducted during the month leading up to the Iowa caucus as reported on the Web site www.pollingreport.com.

REFERENCES

Adkins, R., and A. Dowdle, "How Important Are Iowa and New Hampshire to Winning Post-Reform Presidential Nominations?" *Political Research Quarterly* 54 (2001), 431–44.

Aldrich, J., "A Dynamic Model of Presidential Nomination Campaigns," *American Political Science Review* 74 (1980), 651–69.

Ali, N., and N. Kartik, "Herding with Collective Preferences," *Economic Theory* 51 (2012), 621–6.

Banerjee, A., "A Simple Model of Herd Behavior," *Quarterly Journal of Economics* 107 (1992), 797–817.

Bartels, L., "Expectations and Preferences in Presidential Nomination Campaigns," *American Political Science Review* 79 (1985), 804–15.

———, "Candidate Choice and the Dynamics of the Presidential Nominating Process," *American Journal of Political Science* 31 (1987), 1–30.

———, *Presidential Primaries and the Dynamics of Public Choice* (Princeton: Princeton University Press, 1988).

Battaglini, M., "Sequential Voting with Abstention," *Games and Economic Behavior* 51 (2005), 445–63.

———, R. Morton, and T. Palfrey, "Efficiency, Equity, and Timing of Voting Mechanisms," *American Political Science Review* 101 (2007), 409–24.

Bikhchandani, S., D. Hirshleifer, and I. Welch, "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades," *Journal of Political Economy* 100 (1992), 992–1026.

Brams, S. J., and M. D. Davis, "Optimal Resource Allocation in Presidential Primaries," *Mathematical Social Sciences* 3 (1982), 373–88.

Callander, S., "Bandwagons and Momentum in Sequential Voting," *Review of Economic Studies* 74 (2007), 653–84.

Chamley, C., *Rational Herds: Economic Models of Social Learning* (Cambridge: Cambridge University Press, 2004).

Coate, S., and B. Knight, "Socially Optimal Districting: A Theoretical and Empirical Exploration," *Quarterly Journal of Economics* 122 (2007), 1409–71.

Dekel, E., and M. Piccione, "Sequential Voting Procedures in Symmetric Binary Elections," *Journal of Political Economy* 108 (2000), 34–55.

Deltas, G., H. Herrera, and M. Polborn, "A Theory of Learning and Coordination in the Presidential Primary System," Working Paper, University of Illinois, 2010.

Feddersen, T., and W. Pesendorfer, "Voting Behavior and Information Aggregation in Elections with Private and Common Values," *Econometrica* 65 (1997), 1029–58.

Hummel, P., "Sequential Voting When Long Elections Are Costly," *Economics and Politics* 23 (2011), 36–58.

———,"Sequential Voting in Large Elections with Multiple Candidates," *Journal of Public Economics* 96 (2012), 341–8.

———, and R. Holden, "Optimal Primaries," *Journal of Public Economics* 109 (2014), 64–75.

Kenney, P., and T. W. Rice, "The Psychology of Political Momentum," *Political Research Quarterly* 47 (1994), 923–38.

Klumpp, T., and M. Polborn, "Primaries and the New Hampshire Effect," *Journal of Public Economics* 90 (2006), 1073–114.

Knight, B., and N. Schiff, "Momentum and Social Learning in Presidential Primaries," *Journal of Political Economy* 118 (2010), 1110–50.

Lizzeri, A., and N. Persico, "The Provision of Public Goods under Alternative Electoral Incentives," *American Economic Review* 91 (2001), 225–39.

Maskin, E., and J. Tirole, "The Politician and the Judge: Accountability in Government," *American Economic Review* 94 (2004), 1034–54.

Mayer, W. G., and A. Busch, *The Front-Loading Problem in Presidential Nominations* (Washington, DC: Brookings Institute Press, 2004).

Morton, R., and K. Williams, "Information Asymmetries and Simultaneous Versus Sequential Voting," *American Political Science Review* 93 (1999), 51–67.

——— and ———, *Learning by Voting: Sequential Choices in Presidential Primaries and Other Elections* (Ann Arbor: University of Michigan Press, 2001).

Norrander, B., *Super Tuesday: Regional Politics and Presidential Primaries* (Lexington: University Press of Kentucky, 1992).

Persson, T., and G. Tabellini, "Constitutions and Economic Policy," *Journal of Economic Perspectives* 18 (2004), 75–98.

———, G. Roland, and G. Tabellini, "Comparative Politics and Public Finance," *Journal of Political Economy* 108 (2000), 1121–61.

Prat, A., "Campaign Advertising and Voter Welfare," *Review of Economic Studies* 69 (2002), 999–1018.

Redslawk, D. P., C. J. Tolbert, and T. Donavan, *Why Iowa? How Cacuses and Sequential Elections Improve the Presidential Nominating Process* (Chicago: University of Chicago Press, 2011).

Selman, D., "Optimal Sequencing of Presidential Primaries," Working Paper, Boğaziçi University, 2010.
Steger, W., "Forecasting the Presidential Primary Vote: Viability, Ideology and Momentum," *International Journal of Forecasting* 24 (2008), 193–208.
———, A. Dowdle, and R. Adkins, "The New Hampshire Effect in Presidential Nominations," *Political Research Quarterly* 57 (2004), 375–90.
Strumpf, K., "Strategic Competition in Sequential Election Contests," *Public Choice* 111 (2002), 377–97.
Welch, I., "Sequential Sales, Learning, and Cascades," *The Journal of Finance* 47 (1992), 695–732.