February 7, 2012

Sequential or Simultaneous Elections? A Welfare Analysis*

Abstract

This paper contributes to a debate over a key question in the design of electoral systems. Should all voters go to the polling booth on the same day or should elections be staggered, with late voters observing the choices of early voters before making their decisions? Using a model of voting and social learning, we identify a key trade-off between simultaneous and sequential elections. In particular, sequential election systems place too much weight on the preferences and information of early states but also provide late voters with information that is valuable in terms of selecting high quality candidates. Under simultaneous elections, voters equally weigh the available information but place too little weight on the information in aggregate and thus place too much weight on their priors. Given this trade-off, either sequential or simultaneous elections might be welfare-preferred. We provide a quantitative evaluation of this trade-off based upon an application to the 2004 presidential primary. The results suggest that simultaneous systems outperform sequential systems although the difference in welfare is relatively small.

Patrick Hummel
Yahoo! Research

Brian Knight
Brown University

*Thanks to Nageeb Ali and to participants at the Princeton Conference on Political Economy for helpful comments on an earlier draft.

"...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

While many elections are held on the same day, other elections are staggered, with different voters casting their votes on different days. 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 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. During the 2008 season, there was significant front-loading, with 22 states moving their primary to the first Tuesday in February, and this date was dubbed by some commentators as a "national primary". According to the 2012 schedule, by contrast, only a handful of states vote in early February, with a large group voting on the first Tuesday in March.

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 towards a more efficient and fair system. Hybrid systems, which move towards 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.

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 the current system argue that early states have disproportionate influence, while supporters argue that it is 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. While 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 paper, 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. In the context of a simple version of the model with two candidates, we show that neither system is optimal and that there is indeed a trade-off between voters equally weighing the 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 the simultaneous election tends 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 the 2004 election, and the estimates suggest that simultaneous election systems outperform sequential election systems, at least in the context of this election.

The paper 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 based on the 2004 Democratic primary calendar and the associated pool of candidates competing in this election.

2 Literature Review

This paper is at the intersection of four literatures: social learning, theoretical analyses of sequential voting systems, empirical analyses of presidential primary systems, and optimal electoral institutions.

The literature on social learning began with Welch (1992), Bikhchandani, Hirshleifer, and Welch (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 would rationally ignore the behavior of early agents were it in their best interests to do so. In the voting context, by contrast, individual payoffs depend upon the actions of all agents and thus whether a simultaneous or sequential calendar is preferred from a welfare perspective is less clear.

Several papers have examined this issue of social learning in the electoral context, with a focus on binary elections. In a model with 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, Callandar (2007), who examines sequential elections in the context of a model in which voters prefer to vote for winners, 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.

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. 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 are a series of papers, including Adkins and Dowdle (2001), Steger, Dowdle, and Adkins (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.

In closely related work, Deltas, Herrera, and Polborn (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, rather than 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 Condercet winner. Given that the underlying advantages of sequential elections are different in their model, we view our work as complementary to this paper.

Finally, this paper is related to a broader literature on the normative analysis of electoral institutions. Hummel and Holden (2012) 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 paper. 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, Roland, and Tabellini (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 ($s = 1, 2, ..., S$) choosing between candidates ($c = 0, 1, ..., C$). 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 that is assumed to be drawn independently from a type-I extreme value distribution 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.

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, we can write the vote shares for candidate c, relative to the baseline candidate 0, in state s voting at time t as follows:

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

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:

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

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

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 under 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:

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

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. 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:

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

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

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

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,

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

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 amongst 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 while the simultaneous and public information systems* *place equal weight on the preferences and information of the early and late states. That is,* $\frac{\partial z^{\text{seq}}/\partial \theta_A}{\partial z^{\text{seq}}/\partial \theta_B} > 1$, $\frac{\partial z^{\text{seq}}/\partial \eta_A}{\partial z^{\text{seq}}/\partial \eta_B} > 1$ *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.,* $\omega_\mu^{sim} > \omega_\mu^{\text{seq}} > \omega_\mu^{public}$*. Moreover, the front-runner has a higher probability of winning the simultaneous election than the sequential election, i.e.,* $P^{sim} > P^{\text{seq}}$.

Proofs of all propositions are in the appendix. The intuition for the first result in Proposition 2 ($\omega_\mu^{sim} > \omega_\mu^{\text{seq}}$) 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 ($\omega_\mu^{\text{seq}} > \omega_\mu^{public}$), 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 ($P^{sim} > P^{\text{seq}}$) 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.[1]

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.

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, we then have that:

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

[1] $P^{sim} > P^{public}$ 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 $P^{\text{seq}} > P^{public}$.

where $P = \Phi\left(\frac{\mu_1}{\sigma_z}\right)$ 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.

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

$$\begin{aligned}\Delta &= E^{sim}(W) - E^{\text{seq}}(W) \\ &= \mu_1\left(P^{sim} - P^{\text{seq}}\right) + \rho^{sim}_{y,z}\sigma_y\phi\left(\frac{\mu_1}{\sigma^{sim}_z}\right) - \rho^{\text{seq}}_{y,z}\sigma_y\phi\left(\frac{\mu_1}{\sigma^{\text{seq}}_z}\right) \qquad (25)\end{aligned}$$

The first term measures the expected benefit from electing the front-runner (μ_1) multiplied by the difference in the probabilities the front-runner will be 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 be interpreted as the difference between the informational gain associated with implementing the simultaneous system instead of the sequential system. This term can either be positive or negative and depends on $\rho^{sim}_{y,z}$ and $\rho^{\text{seq}}_{y,z}$, 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 $\rho^{\text{seq}}_{y,z}$, 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 the sequential system, i.e.,* $\rho^{sim}_{y,z} > \rho^{\text{seq}}_{y,z}$.

The fact that the correlation between aggregate utility and the index of support for the front-runner is greater under the simultaneous system than 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.

5 Numerical Analysis

Returning to the more general case of 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 particular, we aim to evaluate the welfare expression in equation (14) under all three systems. 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). 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 ($\mu_{D1}, \mu_{E1}, \sigma_1, \sigma_\varepsilon$, and σ_η) from this analysis are summarized in Table 1. As shown, 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 a quality value, relative to Kerry, for Dean (q_D) and Edwards (q_E) from the normal distributions with means μ_{D1} and μ_{E1}, respectively, and common variance σ_1^2.

2. For each state s, randomly draw a signal noise value, relative to Kerry, for Dean (ε_{Ds}) and Edwards (ε_{Es}), from the normal distribution with mean 0 and variance σ_ε^2.

3. Calculate the state-level signal for Dean ($\theta_{Ds} = q_D + \varepsilon_{Ds}$) and for Edwards ($\theta_{Es} = q_E + \varepsilon_{Es}$).

4. For each state s, randomly draw a preference, relative to Kerry, for Dean (η_{Ds}) and Edwards (η_{Es}) from the normal distribution with 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 three systems as the average voter welfare across these 50,000 replications.

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, in which all voters have access to all signals. In particular, while 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, we see that the simultaneous system produces higher welfare levels than does the sequential system, suggesting that the benefits to the simultaneous system, the equal weighting of voter preferences and information, outweigh any benefits from the sequential system, which provides dark horse candidates of unexpectedly high quality with an opportunity to emerge from the field of candidates.

Given the finite number of replications, we next provide confidence intervals for the welfare difference of 0.0026. With 50,000 replications and a standard deviation for the welfare difference of 0.1113, as estimated across replications, we have a 95-percent confidence interval of (0.0016 , 0.0036). Thus, using conventional significance levels, the number of replications is sufficient to reject the hypothesis that there is no difference in welfare between the simultaneous and sequential systems.

In terms of the magnitude of any welfare gains associated with moving from our current system to a simultaneous system, there are several relevant benchmarks. First, these welfare gains can be compared to the welfare difference between the all-public and sequential systems. That is, we calculate $[E^{sim}(W) - E^{\text{seq}}(W)]/[E^{pub}(W) - E^{\text{seq}}(W)]$. 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 17 percent of maximal gains that can be achieved. 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 $[E^{sim}(W) - E^{\text{seq}}(W)]/[E^{sim}(W) - \mu_{D1}]$. This difference, as expressed in the denominator, can be interpreted as 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 percent of the maximal possible gains. This small gain reflects the fact that both systems, simultaneous and sequential, substantially outperform to 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 (σ_ϵ^2=1.197) relative to the variance in the initial prior (σ_1^2=3.577). Thus, voters learn a substantial amount from a single piece of information.[2]

Another natural benchmark for comparing the difference in welfare between the simultaneous and sequential systems is to note how this welfare difference compares to the welfare difference that would arise if one happens to randomly draw a high quality candidate from the distribution of candidate qualities rather than a low quality candidate. The variance in candidate qualities is σ_1^2, so the standard deviation in candidate qualities is σ_1 and differences in random draws of candidate quality are likely to affect average voter welfare by an amount on the order of σ_1. Since the parameter estimates from Knight and Schiff (2010) indicate that σ_1^2 is about 3.577 (and thus σ_1 is about 1.891), randomly drawing a high quality candidate rather than a low quality candidate from the distribution of candidate qualities is likely to affect voter welfare by an amount on the order of a full unit of utility. By contrast, the

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

difference in expected welfare between the simultaneous system and the sequential system is 0.0026. Thus the benefit from randomly drawing a high quality candidate to run for office instead of a low quality candidate is several hundred times greater than the expected benefit from switching to a simultaneous system from a sequential system. This again indicates that the expected welfare difference between the simultaneous and sequential systems is small.

To provide further context to these differences in voter welfare, we provide a quantitative evaluation, in the context of this simulation exercise, of the relative advantages of the simultaneous and sequential systems. We first compute the odds of each of the three candidates winning the election. As shown in Table 2, the simultaneous system does give too much advantage to the front-runner, with Dean, who led prior to Iowa, winning in 69 percent of cases. Under the full information system, by contrast, Dean wins in only 61 percent of cases, and the sequential system, in which Dean wins in 62 percent of 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 percent of 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 illustrate and quantify the disadvantages of the sequential system, we next provide quantitative evidence on the disproportionate influence of early states. While 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 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 1 and 2 plot the coefficients on the signals and preferences, respectively, from these regressions. As shown in Figure 1, the sequential system does substantially overweight 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 signal of the first state has over 45 times the influence as that 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 2 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 preferences of the first state have roughly 9 times the influence as those of the last state in the sequence. 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. However, the simultaneous and all-public systems again place equal weight on the preferences of each state.

To provide further interpretation of these results, we consider three alternative sequential systems. First, we consider the 2008 calendar, when nearly half of the states moved their primary to the first Tuesday of February, and the 2012 calendar. As shown in Table 2, these alternative calendars still fall short of the simultaneous election in terms of voter welfare and yield welfare levels that are slightly lower than those using the 2004 calendar. 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 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 towards a simultaneous system tend to increase voter welfare.

Finally, to illustrate the trade-off identified in Proposition 4, we 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, while 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 $\lambda = \{0, 0.5, 2, 3, 4, 5\}$. Thus λ can be considered a measure of the electoral advantage of the front-runner. As shown in Figure 3, the welfare gains 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 small as the advantage grows larger, reflecting the fact that the front-runner is increasingly likely to win under either system.

Figure 3 also provides additional interpretation for the small size of the documented welfare gain. As noted above, the welfare gain under the baseline is smaller than the welfare gain when the advantage of the front-runner is small ($\lambda = 0$ and $\lambda = 0.5$). Moreover, the welfare gain under the baseline is smaller in absolute value than the welfare loss when the advantage of the front-runner takes on moderately larger values (i.e. $\lambda = \{2, 3, 4\}$).

This suggests that, at the baseline, the cost and benefits associated with a movement from sequential to simultaneous nearly offset one another, and the welfare gain is thus relatively small.

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. While 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, with signals of the first state having 45 times the weight of those of the last state, and to early preferences, with those of the first state having 9 times the weight of those of the last state. This is not a general result, however, in the sense that the sequential system tends to dominate as the advantage of the front-runner grows larger.

6 Conclusion

While 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 assumed that all candidates stay in the race for all states under the sequential system, but in reality some candidates may drop out if they have a poor early performance. It would be interesting to analyze an alternative model which incorporates candidate exit under the sequential system. While our analysis abstracts from exit, allowing for exit may give further disproportionate influence to early states if voting returns force candidates to exit from the race.

Second, we have not allowed for endogenous candidate strategies, which may differ between the two systems. For instance, in a sequential election, candidates typically focus their campaign efforts on early states,[3] and as a result candidates may focus on issues that are important to voters in early states. By contrast, in a simultaneous election, candidates

[3] This is noted in Knight and Schiff (2010).

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 paper 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. While 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. Focusing on the 2004 calendar and associated pool of candidates, we then conduct an empirical welfare analysis. While the results suggest that the simultaneous system outperforms the sequential system, the difference in welfare is relatively small.

References

Randall Adkins and Andrew Dowdle. 2001. "How Important Are Iowa and New Hampshire to Winning Post-Reform Presidential Nominations?" *Political Research Quarterly* 54 (June):431-444.

John Aldrich. 1980. "A Dynamic Model of Presidential Nomination Campaigns." *American Political Science Review* 74 (September): 651-669.

Nageeb Ali and Navin Kartik. 2012. "Herding with Collective Preferences." *Economic Theory* (in press).

Abhijit Banerjee. 1992. "A Simple Model of Herd Behavior." *Quarterly Journal of Economics* 107 (August): 797-817.

Larry Bartels. 1985. "Expectations and Preferences in Presidential Nomination Campaigns." *American Political Science Review* 79 (September): 804-815.

Larry Bartels. 1987. "Candidate Choice and the Dynamics of the Presidential Nominating Process." *American Journal of Political Science* 31 (February): 1-30.

Larry Bartels. 1988. *Presidential Primaries and the Dynamics of Public Choice*. Princeton: Princeton University Press.

Battaglini, Marco. 2005. "Sequential Voting with Abstention." *Games and Economic Behavior* 51 (May): 445-63.

Marco Battaglini, Rebecca Morton, and Thomas Palfrey. 2007. "Efficiency, Equity, and Timing of Voting Mechanisms." *American Political Science Review* 101 (August): 409-424.

Sushil Bikhchandani, David Hirshleifer, and Ivo Welch. 1992. "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades." *Journal of Political Economy* 100 (October): 992-1026.

Sushil Bikhchandani, David Hirshleifer, and Ivo Welch. 1998. "Learning from the Behavior of Others: Conformity, Fads, and Informational Cascades." *Journal of Economic Perspectives* 12 (Summer): 151-170.

Christophe Chamley. 2004. *Rational Herds: Economic Models of Social Learning*. Cambridge: Cambridge University Press.

Stephen Coate. 2004. "Pareto-Improving Campaign Finance Policy". *American Economic Review* 94: 628–655.

Stephen Coate and Brian Knight. 2007. "Socially Optimal Districting: A Theoretical and Empirical Exploration". *Quarterly Journal of Economics* 122: 1409–1471.

Eddie Dekel and Michele Piccione. 2000. "Sequential Voting Procedures in Symmetric Binary Elections." *Journal of Political Economy* 108 (February): 34-55.

George Deltas, Helios Herrera, and Mattias Polborn. 2010. "A Theory of Learning and Coordination in the Presidential Primary System", working paper, University of Illinois.

Patrick Hummel. 2011. "Sequential Voting When Long Elections Are Costly." *Economics and Politics* 23 (March): 36-58.

Patrick Hummel. 2012. "Sequential Voting in Large Elections with Multiple Candidates. "*Journal of Public Economics* (in press).

Patrick Hummel and Richard Holden. 2012. "Optimal Primaries", working paper, Yahoo! Research.

Patrick Kenney and Tom W. Rice. 1994. "The Psychology of Political Momentum." *Political Research Quarterly* 47 (December): 923-938.

Tilman Klumpp and Mattias Polborn. 2006. "Primaries and the New Hampshire Effect." *Journal of Public Economics* 90 (August): 1073-1114.

Brian Knight and Nathan Schiff. 2010. "Momentum and Social Learning in Presidential Primaries." *Journal of Political Economy* 118*:* 1110-1150.

Allesandro Lizzeri and Nicola Persico. 2001. "The Provision of Public Goods under Alternative Electoral Incentives," *American Economic Review* 91: 225-339.

Eric Maskin and Jean Tirole. 2004. "The Politician and the Judge: Accountability in Government" *American Economic Review* 94: 1034-1054.

Rebecca Morton and Kenneth Williams. 1999. "Information Asymmetries and Simultaneous Versus Sequential Voting". *American Political Science Review* 93 (March): 51-67.

Rebecca Morton and Kenneth Williams. 2001. *Learning by Voting: Sequential Choices in Presidential Primaries and Other Elections.* Ann Arbor: University of Michigan Press.

Torsten Persson, Gerard Roland and Guido Tabellini. 2000. "Comparative Politics and Public Finance" *Journal of Political Economy* 108: 1121-1161.

Torsten Persson and Guido Tabellini. 2004. "Constitutions and Economic Policy" *Journal of Economic Perspectives* 18: 75-98.

Andrea Prat. 2002. "Campaign Advertising and Voter Welfare" *Review of Economic Studies* 69, 999-1018.

Wayne Steger. 2008. "Forecasting the Presidential Primary Vote: Viability, Ideology and Momentum." *International Journal of Forecasting* 24 (April-June): 193-208.

Wayne Steger, Andrew Dowdle, and Randall Adkins. 2004. "The New Hampshire Effect in Presidential Nominations." *Political Research Quarterly* 57 (September): 375-390.

Koleman Strumpf. 2002. "Strategic Competition in Sequential Election Contests." *Public Choice* 111: 377-397.

7 Appendix

Proof of Proposition 2: The proof consists of three parts. First, we show that $\omega_\mu^{sim} > \omega_\mu^{\text{seq}}$. Second, we show that $\omega_\mu^{\text{seq}} > \omega_\mu^{public}$. Third, we show that $P^{sim} > P^{\text{seq}}$.

Part 1: To show that $\omega_\mu^{sim} > \omega_\mu^{\text{seq}}$, 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 re-written 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 that 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:

$$(1 - \alpha_2) = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + \sigma_1^2(1 - \beta_1)} \tag{27}$$

And, the RHS 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 re-arranging, 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 $\omega_{\mu}^{\text{seq}} > \omega_{\mu}^{public}$, 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 result from equation (27) of 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 re-arranging, 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 and standard deviation σ_z, we have that $P = \Phi(\mu/\sigma_z)$. Thus, to show that $P^{sim} > P^{\text{seq}}$, we only need to show that $\sigma_z^{\text{seq}} > \sigma_z^{sim}$. 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 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 simultaneous. Finally, the third component is larger under sequential than 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 $z = \omega(\theta_A)\theta_A + \omega(\theta_B)\theta_B + \omega(\mu)\mu + \omega(\eta_A)\eta_A + \omega(\eta_B)\eta_B$, we have $z = (\omega(\theta_A) + \omega(\theta_B))q + \omega(\mu)\mu + \omega(\eta_A)\eta_A + \omega(\eta_B)\eta_B + \omega(\theta_A)\epsilon_A + \omega(\theta_B)\epsilon_B$. 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 $\rho_{y,z}^{\text{seq}} = \frac{\frac{1}{2}(\alpha_1+\alpha_2+(1-\alpha_2)\beta_1)\sigma_1^2+\frac{1}{2}(1+\frac{1}{2}(1-\alpha_2)\frac{\beta_1}{\alpha_1})\sigma_\eta^2}{\sigma_y\sqrt{\frac{1}{4}(\alpha_1+\alpha_2+(1-\alpha_2)\beta_1)^2\sigma_1^2+\frac{1}{4}(1+(1+(1-\alpha_2)\frac{\beta_1}{\alpha_1})^2)\sigma_\eta^2+\frac{1}{4}((\alpha_1+(1-\alpha_2)\beta_1)^2+\alpha_2^2)\sigma_\varepsilon^2}}$. Thus in order to prove that $\rho_{y,z}^{sim} > \rho_{y,z}^{\text{seq}}$, it suffices to prove that $(\alpha_1\sigma_1^2 + \frac{1}{2}\sigma_\eta^2)^2(\frac{1}{4}(\alpha_1 + \alpha_2 + (1 - \alpha_2)\beta_1)^2\sigma_1^2 + \frac{1}{4}(1 + (1 + (1 - \alpha_2)\frac{\beta_1}{\alpha_1})^2)\sigma_\eta^2 + \frac{1}{4}((\alpha_1 + (1 - \alpha_2)\beta_1)^2 + \alpha_2^2)\sigma_\varepsilon^2) > (\frac{1}{2}(\alpha_1 + \alpha_2 + (1 - \alpha_2)\beta_1)\sigma_1^2 + \frac{1}{2}(1 + \frac{1}{2}(1 - \alpha_2)\frac{\beta_1}{\alpha_1})\sigma_\eta^2)^2(\alpha_1^2\sigma_1^2 + \frac{1}{2}\sigma_\eta^2 + \frac{1}{2}\alpha_1^2\sigma_\varepsilon^2)$, which is equivalent to proving $(\alpha_1^2\sigma_1^4 + \alpha_1\sigma_1^2\sigma_\eta^2 + \frac{1}{4}\sigma_\eta^4)((\alpha_1 + \alpha_2 + (1 - \alpha_2)\beta_1)^2\sigma_1^2 + (1 + (1 + (1 - \alpha_2)\frac{\beta_1}{\alpha_1})^2)\sigma_\eta^2 + ((\alpha_1 + (1 - \alpha_2)\beta_1)^2 + \alpha_2^2)\sigma_\varepsilon^2) > ((\alpha_1 + \alpha_2 + (1 - \alpha_2)\beta_1)^2\sigma_1^4 + 2(\alpha_1 + \alpha_2 + (1 - \alpha_2)\beta_1)(1 + \frac{1}{2}(1 - \alpha_2)\frac{\beta_1}{\alpha_1})\sigma_1^2\sigma_\eta^2 + (1 + \frac{1}{2}(1 - \alpha_2)\frac{\beta_1}{\alpha_1})^2\sigma_\eta^4)(\alpha_1^2\sigma_1^2 + \frac{1}{2}\sigma_\eta^2 + \frac{1}{2}\alpha_1^2\sigma_\varepsilon^2)$.

Expanding this expression then indicates that it suffices to prove that $\alpha_1^2(\alpha_1 + \alpha_2 + (1 - \alpha_2)\beta_1)^2\sigma_1^6 + \frac{1}{4}(1 + (1 + (1 - \alpha_2)\frac{\beta_1}{\alpha_1})^2)\sigma_\eta^6 + (\alpha_1(\alpha_1 + \alpha_2 + (1 - \alpha_2)\beta_1)^2 + \alpha_1^2(1 + (1 + (1 - \alpha_2)\frac{\beta_1}{\alpha_1})^2))\sigma_1^4\sigma_\eta^2 + \alpha_1^2((\alpha_1 + (1 - \alpha_2)\beta_1)^2 + \alpha_2^2)\sigma_1^4\sigma_\varepsilon^2 + (\frac{1}{4}(\alpha_1 + \alpha_2 + (1 - \alpha_2)\beta_1)^2 + \alpha_1(1 + (1 + (1 - \alpha_2)\frac{\beta_1}{\alpha_1})^2))\sigma_1^2\sigma_\eta^4 + \frac{1}{4}((\alpha_1 + (1 - \alpha_2)\beta_1)^2 + \alpha_2^2)\sigma_\eta^4\sigma_\varepsilon^2 + \alpha_1((\alpha_1 + (1 - \alpha_2)\beta_1)^2 + \alpha_2^2)\sigma_1^2\sigma_\eta^2\sigma_\varepsilon^2 > \alpha_1^2(\alpha_1 + \alpha_2 + (1 - \alpha_2)\beta_1)^2\sigma_1^6 + \frac{1}{2}(1 + \frac{1}{2}(1 - \alpha_2)\frac{\beta_1}{\alpha_1})^2\sigma_\eta^6 + (\frac{1}{2}(\alpha_1 + \alpha_2 + (1 - \alpha_2)\beta_1)^2 + 2\alpha_1^2(\alpha_1 + \alpha_2 + (1 - \alpha_2)\beta_1)(1 + \frac{1}{2}(1 - \alpha_2)\frac{\beta_1}{\alpha_1}))\sigma_1^4\sigma_\eta^2 + \frac{1}{2}\alpha_1^2(\alpha_1 + \alpha_2 + (1 - \alpha_2)\beta_1)^2\sigma_1^4\sigma_\varepsilon^2 + (\alpha_1^2(1 + \frac{1}{2}(1 - \alpha_2)\frac{\beta_1}{\alpha_1})^2 + (\alpha_1 + \alpha_2 + (1 - \alpha_2)\beta_1)(1 + \frac{1}{2}(1 - \alpha_2)\frac{\beta_1}{\alpha_1}))\sigma_1^2\sigma_\eta^4 + \frac{1}{2}\alpha_1^2(1 + \frac{1}{2}(1 - \alpha_2)\frac{\beta_1}{\alpha_1})^2\sigma_\eta^4\sigma_\varepsilon^2 + \alpha_1^2(\alpha_1 + \alpha_2 + (1 - \alpha_2)\beta_1)(1 + \frac{1}{2}(1 - \alpha_2)\frac{\beta_1}{\alpha_1})\sigma_1^2\sigma_\eta^2\sigma_\varepsilon^2$.

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

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

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

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

Also note that for general values of μ_1, we have

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

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

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

Now $\frac{1}{2}(\alpha_1 + \alpha_2 + (1-\alpha_2)\beta_1)\sigma_1^2 + \frac{1}{2}(1 + \frac{1}{2}(1-\alpha_2)\frac{\beta_1}{\alpha_1})\sigma_\eta^2 > \frac{1}{4}(\alpha_1 + \alpha_2 + (1-\alpha_2)\beta_1)^2\sigma_1^2 + \frac{1}{4}(1 + (1 + (1-\alpha_2)\frac{\beta_1}{\alpha_1})^2)\sigma_\eta^2 + \frac{1}{4}((\alpha_1 + (1-\alpha_2)\beta_1)^2 + \alpha_2^2)\sigma_\varepsilon^2$ holds if and only if $2(\alpha_1 + \alpha_2 + (1-\alpha_2)\beta_1)\sigma_1^2 + 2(1 + \frac{1}{2}(1-\alpha_2)\frac{\beta_1}{\alpha_1})\sigma_\eta^2 > (\alpha_1 + \alpha_2 + (1-\alpha_2)\beta_1)^2\sigma_1^2 + (1 + (1 + (1-\alpha_2)\frac{\beta_1}{\alpha_1})^2)\sigma_\eta^2 + ((\alpha_1 + (1-\alpha_2)\beta_1)^2 + \alpha_2^2)\sigma_\varepsilon^2 \Leftrightarrow 2(\alpha_1 + \alpha_2 + (1-\alpha_2)\beta_1)\sigma_1^2 > (\alpha_1 + \alpha_2 + (1-\alpha_2)\beta_1)^2\sigma_1^2 + ((1-\alpha_2)\frac{\beta_1}{\alpha_1} + (1-\alpha_2)^2\frac{\beta_1^2}{\alpha_1^2})\sigma_\eta^2 + ((\alpha_1 + (1-\alpha_2)\beta_1)^2 + \alpha_2^2)\sigma_\varepsilon^2 \Leftrightarrow 2(\alpha_1 + \alpha_2 + (1-\alpha_2)\beta_1)\sigma_1^2 > (\alpha_1^2 + \alpha_2^2 + (1-\alpha_2)^2\beta_1^2 + 2\alpha_1\alpha_2 + 2\alpha_1(1-\alpha_2)\beta_1 + 2\alpha_2(1-\alpha_2)\beta_1)\sigma_1^2 + ((1-\alpha_2)\frac{\beta_1}{\alpha_1} + (1-\alpha_2)^2\frac{\beta_1^2}{\alpha_1^2})\sigma_\eta^2 + (\alpha_1^2 + 2\alpha_1(1-\alpha_2)\beta_1 + (1-\alpha_2)^2\beta_1^2 + \alpha_2^2)\sigma_\varepsilon^2 \Leftrightarrow (\alpha_1 + 2\alpha_2 + 2(1-\alpha_2)\beta_1)\sigma_1^2 > (\alpha_2^2 + (1-\alpha_2)^2\beta_1^2 + 2\alpha_1\alpha_2 + 2\alpha_1(1-\alpha_2)\beta_1 + 2\alpha_2(1-\alpha_2)\beta_1)\sigma_1^2 + ((1-\alpha_2)\frac{\beta_1}{\alpha_1} + (1-\alpha_2)^2\frac{\beta_1^2}{\alpha_1^2})\sigma_\eta^2 + (2\alpha_1(1-\alpha_2)\beta_1 + (1-\alpha_2)^2\beta_1^2 + \alpha_2^2)\sigma_\varepsilon^2 \Leftrightarrow (\alpha_1 + 2\alpha_2)\sigma_1^2 > (\alpha_2^2 + (1-\alpha_2)^2\beta_1^2 + 2\alpha_1\alpha_2 + 2\alpha_2(1-\alpha_2)\beta_1)\sigma_1^2 + ((1-\alpha_2)\frac{\beta_1}{\alpha_1} + (1-\alpha_2)^2\frac{\beta_1^2}{\alpha_1^2})\sigma_\eta^2 + ((1-\alpha_2)^2\beta_1^2 + \alpha_2^2)\sigma_\varepsilon^2 \Leftrightarrow (\alpha_1 + 2\alpha_2)\sigma_1^2 >$

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

But $\alpha_1-\alpha_2\beta_1+\frac{\alpha_2}{\alpha_1}+\beta_1 = \alpha_1+\frac{\alpha_2}{\alpha_1}+(1-\alpha_2)\beta_1 = \frac{\sigma_1^2}{\sigma_1^2+\sigma_\varepsilon^2}+\frac{\sigma_2^2(\sigma_1^2+\sigma_\varepsilon^2)}{\sigma_1^2(\sigma_2^2+\sigma_\varepsilon^2)}+\frac{\sigma_\varepsilon^2\beta_1}{\sigma_2^2+\sigma_\varepsilon^2} = \frac{\sigma_1^4(\sigma_2^2+\sigma_\varepsilon^2)+\sigma_2^2(\sigma_1^2+\sigma_\varepsilon^2)^2+\sigma_1^2\sigma_\varepsilon^2(\sigma_1^2+\sigma_\varepsilon^2)\beta_1}{\sigma_1^2(\sigma_1^2+\sigma_\varepsilon^2)(\sigma_2^2+\sigma_\varepsilon^2)} = \frac{\sigma_1^4\sigma_2^2+\sigma_1^4\sigma_\varepsilon^2+\sigma_1^4\sigma_2^2+2\sigma_1^2\sigma_2^2\sigma_\varepsilon^2+\sigma_2^2\sigma_\varepsilon^4+\sigma_1^4\sigma_\varepsilon^2\beta_1+\sigma_1^2\sigma_\varepsilon^4\beta_1}{\sigma_1^4\sigma_2^2+\sigma_1^4\sigma_\varepsilon^2+\sigma_1^2\sigma_2^2\sigma_\varepsilon^2+\sigma_1^2\sigma_\varepsilon^4} = 2-\frac{(1-\beta_1)\sigma_1^4\sigma_\varepsilon^2+(\sigma_1^2-\sigma_2^2)\sigma_\varepsilon^4+(1-\beta_1)\sigma_1^2\sigma_\varepsilon^4}{\sigma_1^4\sigma_2^2+\sigma_1^4\sigma_\varepsilon^2+\sigma_1^2\sigma_2^2\sigma_\varepsilon^2+\sigma_1^2\sigma_\varepsilon^4} < 2$. Thus the last inequality in the previous paragraphs holds, which in turn implies that $Cov(y, z^{\text{seq}}) > (\sigma_z^{\text{seq}})^2$. Thus $\Delta < 0$ for sufficiently large μ_1.

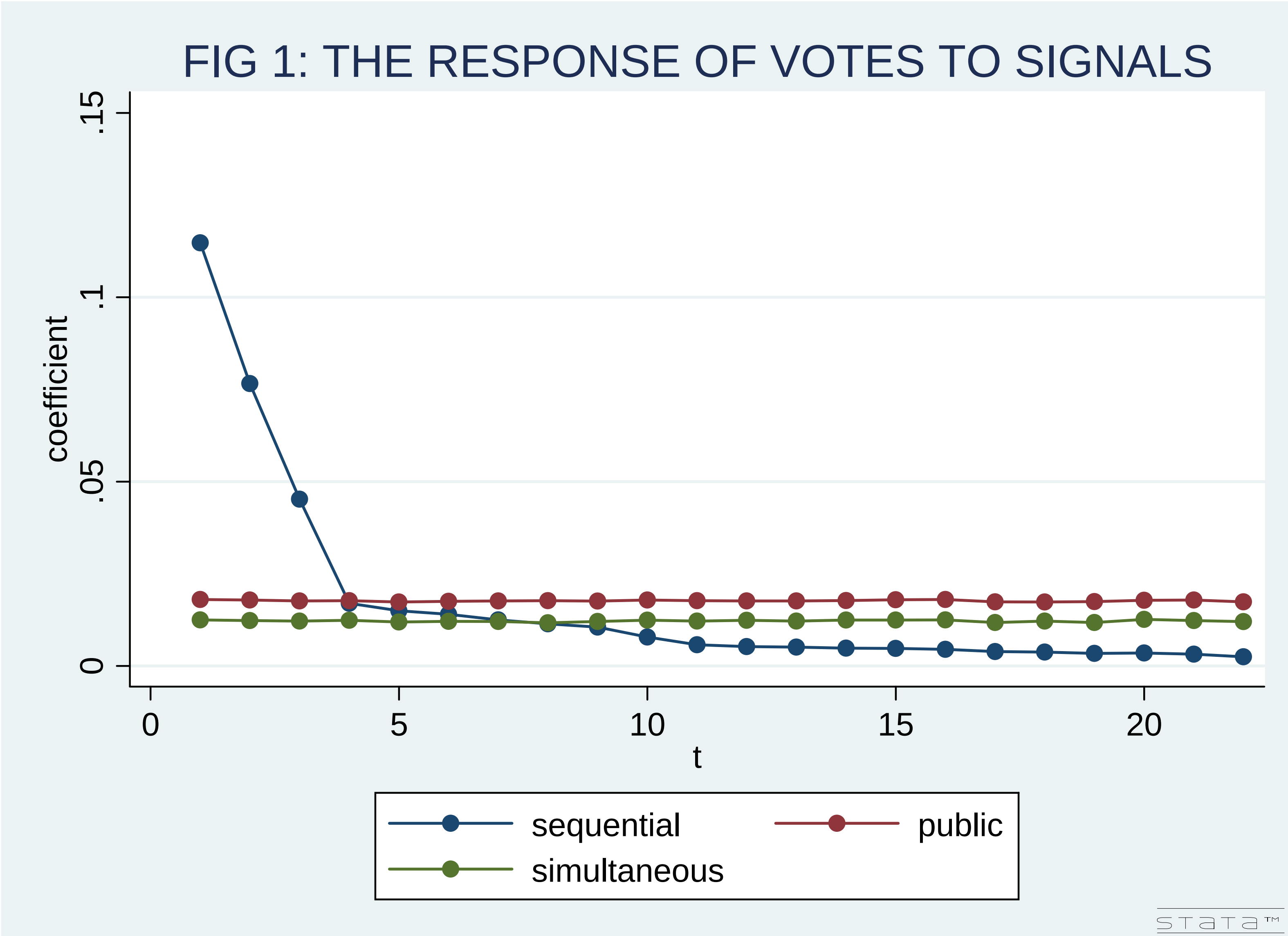
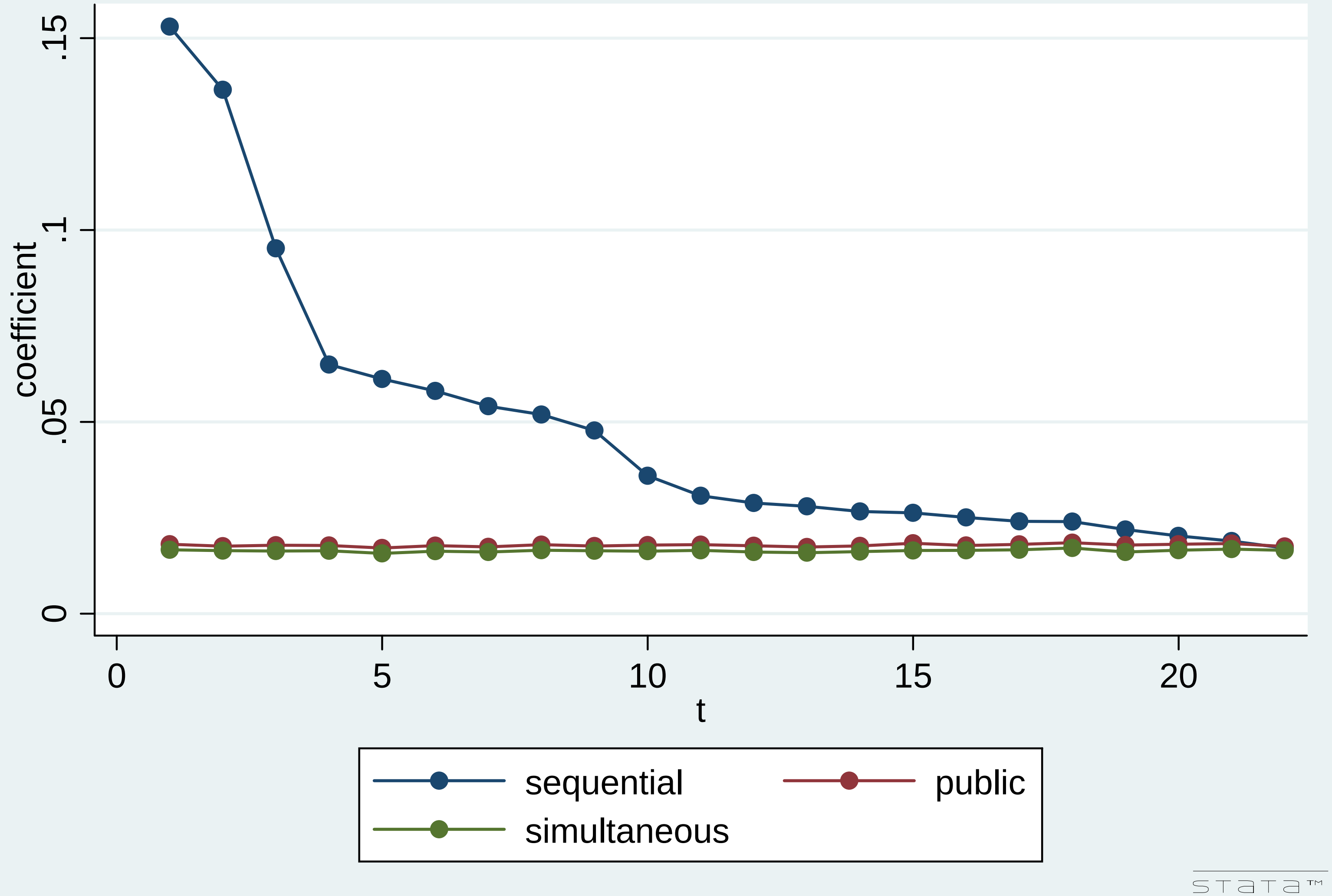


FIG 1: THE RESPONSE OF VOTES TO SIGNALS
coefficient
0
.05
.1
.15
0
5
10
15
20
t
sequential
public
simultaneous
stata™

FIG 2: THE RESPONSE OF VOTES TO PREFERENCES



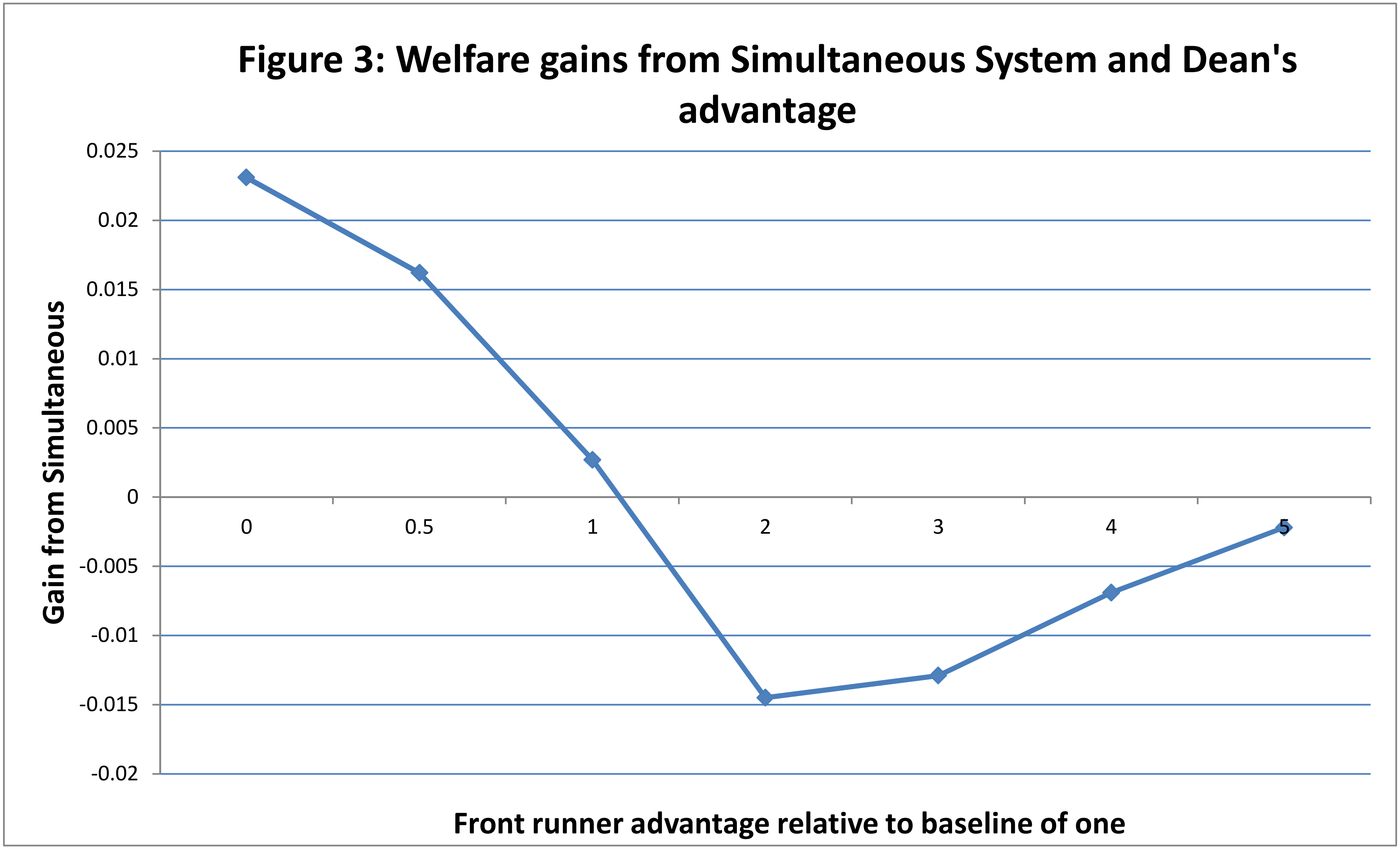


Figure 3: Welfare gains from Simultaneous System and Dean's advantage
0.025
0.02
0.015
0.01
0.005
0
-0.005
-0.01
-0.015
-0.02
Gain from Simultaneous
0
0.5
1
2
3
4
5
Front runner advantage relative to baseline of one

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]
σ_{η}^{2}	0.815**
	[0.551,1.194]
σ_{1}^{2}	3.577**
	[1.497,7.129]
σ_{ε}^{2}	1.197**
	[0.062,4.097]

[bootstrap 95% confidence interval], ** denotes significance at the 95-percent level

Table 2: Results from Numerical Welfare Analysis

system	average welfare level	Dean elected	Edwards elected	Kerry elected
all public	1.4150	60.88%	12.61%	26.50%
simultaneous	1.3978	68.91%	8.11%	22.98%
sequential (2004 calendar)	1.3952	62.24%	11.84%	25.92%

alternative systems	average welfare level
sequential (2008 calendar)	1.3951
sequential (2012 calendar)	1.3947
rotating regional primary	1.3967
pure sequential	1.3946