THE REVIEW OF ECONOMIC STUDIES

Review of Economic Studies (2008) **75**, 597–628 0034-6527/08/00250597$02.00

Optimal Electoral Timing: Exercise Wisely and You May Live Longer[1]

JUSSI KEPPO, LONES SMITH
University of Michigan

and

DMITRY DAVYDOV
Merrill Lynch

First version received May 2006; *final version accepted October* 2007 (*Eds.*)

The timing of elections is flexible in many countries. We study this optimization by first creating a Bayesian learning model of a mean-reverting political support process. We then explore optimal electoral timing, modelling it as a renewable American option with interacting waiting and stopping values. *Inter alia,* we show that the expected longevity is a convex, then concave, function of the support. Finally, we calibrate our model to the post-1945 Labour–Tory U.K. rivalry. Our story quite well explains when the elections were called. We also show that election options approximately double the expected time in power in the current streak.

1. INTRODUCTION

Timing lies at the heart of many economic decisions, and the option to choose when to act often has immense value. This has been the subject of a large literature in economics, most especially in finance. Building on the insights from finance about option pricing, this paper revisits, instead, a classic political economy question: optimal electoral timing.

In a key thread of a parliamentary democracy's fabric, the incumbent often has some flexibility in choosing when it faces the electorate. We first develop a theoretical model of the decision problem facing the government in deciding when to call an election. We then proceed to illustrate it using the post-World War II experience of the U.K. A newly elected government there must call an election within five years, but generally acts in advance of this binding terminal constraint. While the tradition is to call the election around the four-year mark, the actual exercise time has ranged from six to 60 months. In theory, we find that optimally exercising this option has tremendous value, approximately doubling the expected time in power vs. running the term out. And in practice, it offers insights into the electoral success of the Conservatives (simply: the Tories).

For some context, imagine a government in power that sees its monthly standing in the polls and must choose to call an election before its mandate expires. Suppose that an encouraging confluence of events sees its standing surging by 8%. Should it call a snap election now? Obviously, this depends on a host of considerations, ranging from the practical (perhaps it must first pass a budget) to the sociological (maybe the electorate will punish it for "opportunism"). We focus on just one consideration, as we assume that the government simply wishes to maximize its expected

1. This supersedes a manuscript "Optimal Electoral Timing in a Parliamentary Democracy" (2000) that was an unpublished Ph.D. thesis chapter of Dmitry Davydov jointly with Lones Smith. The paper was a three-party calibration exercise and was not testable. This paper differs in many other respects and so must be considered wholly new.

total time in power in the current streak. We find that this has significant explanatory power for the election times.

Naturally, the government should call election (i) the closer to the end of the term, and (ii) the higher its political support. To characterize this trade-off, we draw the analogy of the electoral timing choice to the optimal exercise time of an American option—that is, the right to buy or sell a stock in a fixed window of time at a moment of one's choosing. Yet the theory underlying our story is harder in several dimensions. First, an election is not at all like an asset sale: an investor choosing to exercise a financial option early need not ever think beyond its maximum term. On the other hand, a government that "sells its mandate" early in an election thereafter wins it back if it succeeds; this "renewal option" is forward-looking over an infinite horizon. Second, asset prices are perfectly observed, while a government only sees a noisy signal of its standing from the polls. Third, the stochastic process of asset prices is well developed and tractable (geometric Brownian motion), but there exists no similar model of the popular standing of a government.

We begin by addressing this last omission first. Our model is tractable and captures three key features of the political process in a left–right rivalry: voter heterogeneity, the fickle fortunes of political parties, and the continuous onslaught of media information.

The theory of voting is itself an area of much research. For simplicity, we assume a continuum of politically heterogeneous voters wish to vote for the current "best governing party". This best party is assumed unobserved by all. To wit, right and left wing supporters alike wish to vote Tory if Labour is a mess, but right is far more readily convinced to vote Tory than is left. Namely, ordinal preferences coincide—that is, all prefer the best party—but cardinal preferences diverge. This blend subsumes political ideologues (extreme cardinal preferences) and a varying intensity of political allegiances across voters.

Next, towards a political ebb and flow, we assume that the best party periodically and randomly changes according to a continuous-time Markov process. Voters continuously learn over time about this unobserved Markovian state from the news media. This is achieved in our model with a simple Bayesian device: voters constantly observe the outcome of a Brownian motion with uncertain drift. This drift represents the best party—high when the best party is right, say, and low when the best party is left. This yields in Lemma 1, a simple continuous time stochastic process for the *political slant*, the current posterior chance that the best party is right. As the best party periodically switches, *this stochastic process is mean reverting.* Its long-run distribution is so well behaved that we are able to precisely compute it (Lemma 3). Once we assume an exponential distribution over the strength of political beliefs, the political slant equals the fraction of voters that support it (Lemma 4). This brings us to a Bayesian law of motion for political support. At the end of this exercise, *a party's support reflects political leanings, and yet evolves in a Bayes-rational fashion to reflect new information.* We have not found another rationally derived support process. Ours is so tractable that it should prove useful in future work.

This brings us to our second contribution on the timing problem itself. The government continuously entertains a waiting value depending on the political slant and time left and stops when it coincides with a slant-dependent stopping value. It calls an election when its political standing first hits a non-linear stopping barrier. Since the winning government reacquires the election-calling option, the stopping value is recursively defined in terms of future waiting values—a novel feature. Also, the government only has access to noisy polling data and does not know its true support. Because the optimal exercise time for the finite horizon American put option is not analytically known, our harder optimization problem can only be numerically solved. Still, we prove existence of the solution of this recursive option (Proposition 2), and then characterize it by variational equations (Proposition 3). The expected longevity is a convex then concave function of the support. We also analytically study how the optimal strategy responds to parametric shifts. Elections, for example, tend to be later with more volatile political support.

Our third contribution is an empirical test of our timing model and a finding that timing matters, that is, the option is valuable. We motivate the relevance in two ways. For a bigger quick motivational picture, we seek a large cross-section of similar two-party democracies that have been around for a long time. Since democracy is so young, we choose the provinces of Canada and states of the U.S. We find that provincial governments (with flexible electoral timing) have lasted significantly (50%) more than the state governors. To say anything stronger, our model must be calibrated to a specific case.

We next calibrate our polling process to the U.K., for it is the parliamentary democracy with the longest time series of voting intention polls, and its two big parties, Labour and Tory, have won all the elections after World War II. We use public polling data from 1943 to 2005 and the 17 elections 1945–2005. We estimate the polling process parameters from the polling data: they are statistically significant and do not statistically depend on whether an election campaign is in progress. We show that volatility amounts to 48% of the average poll vs. 14% for the S&P 500. We also document the extremely fast mean reversion that drives the polls: regardless of the initial poll, its expected value three years later lies within 1% of its mean level. This corresponds to an underlying 2·5-year "political cycle" for the unobserved political state to return its initial position.

We use the estimated polling process parameters to solve for the optimal election times. We compare the predicted and realized election times. With just one explanatory variable apart from the elapsed time, our theory explains 44% of the variation in the timing decisions of governments *not troubled by weak or minority governments*. Also, if we additionally assume sufficiently impatient prime ministers, who earn no flow utility from weak or minority governments, then our model explains 39% of all election timing variation. Both of these fits are consistent with our idea that a major determinant of when governments call elections is their desire to maximize their expected time in power (or their expected discounted time, in the latter case), using public polling information.

Our paper also offers a useful normative message. The freedom to optimally time the next election clearly confers upon an incumbent government an advantage unavailable in fixed election cycle regimes. For instance, one can postpone the election until the economy is looking up. Our model quantifies the long-run average magnitude of this advantage, about doubling the expected time in power in the U.K. If the U.K. implemented a fixed electoral cycle with four-year terms, then the expected duration in power would fall by a factor of 1·8 for Labour (from 15·9 to 9·0 years) and by slightly less for Tory (from 12·4 to 7·5 years). Flexible terms on average benefit the more popular party more than the less popular party. Constitutional designers should be aware of the magnitude of this differential effect in choosing amongst fixed and flexible electoral terms.

1.1. *Literature review*

Balke (1990) showed that majority governments trade off current time in power against uncertain future time in their election timing decisions. Following on this observation, Lesmono, Tonkes and Burrage (2003) is the closest paper to ours. They also analogize election timing to American option theory.[2] In contrast to their paper, our underlying political support process is different, which should come as no surprise as we derive it from a Bayesian learning foundation. Their models implied political support process mean reverts about 1/2 (*i.e.* the long-run mean is fixed to 1/2), it does not consider polling error, and their model is not well defined if the support process has a high volatility. Further, we prove the existence of the solution, characterize the value function and the optimal policy by using variational equations, and give comparative statics. We also test empirically how well the model explains the realized elections times.

2. This paper was unavailable when our precursor paper by Davydov and Smith (2000) was written.

There is a large literature on timing and political business cycles.[3] For instance, Palmer (2000) finds that macroeconomic and political variables affect election timing. Better economic indicators lead to early elections. In our paper, governments take the polling process as given and optimize their election timing. Diermeier and Merlo (2000) argue that majority governments are so common because minorities are fragile.

Our paper relates to work on sequential optimal stopping problems in finance and elsewhere. Sequential American options are studied in optimal harvesting problems (*e.g.* Alvarez and Shepp, 1998), executive options with the so-called "reload" feature (*e.g.* Dybvig and Loewenstein, 2003), mortgage refinancing (*e.g.* Hurst and Stafford, 2004), and firms' optimal recapitalization (*e.g.* Peura and Keppo, 2006). Putting aside two other difficulties of our option—measurement error and election delay[4]—we believe that ours is the first renewable American option studied with a finite exercise time horizon. This creates a non-stationary decision rule over time and is the source of interest in this paper. We solve for the non-linear exercise boundary for the electoral timing problem.

1.2. *Structure of the paper*

In Section 2, we show that electoral flexibility has been useful in practice. Section 3 describes the model, and Section 4 the theoretical election timing results. In Section 5, we estimate the model parameters with U.K. polling data and then test it in Section 6. In Section 7, we price the U.K. electoral option, and Section 8 concludes. Appendix A gathers some proofs, while Appendix B describes the numerical solution of the optimal stopping problem.

2. THE ELECTORAL TIMING OPTION IN HISTORY

The U.K. is an ideal candidate for exploring the electoral timing option—it has flexible electoral terms, a long polling series, and a long, two-party alternation. But since we claim that the timing option has value, it would be helpful to see this evidenced in a wider cross-section drawn from other countries with both fixed and flexible electoral terms. Alas, democracy is young, and the democratic countries of the world are diverse. Some are *de facto* one-party states (like Mexico or Japan), about which any electoral theory must be silent. Many are multi-party states where electoral streaks are rare.

We now find the value of the electoral option evidenced in a wide cross-section of the national and state/provincial governments of Canada and the U.S. Hereby, we compare two geographically and culturally close older democracies with two contending parties.[5] Since we do not control for a host of other factors, this section is *purely motivational*.

Canada has flexible election timing (between 0 and five years) and U.S. fixed terms (four years).[6] In Canada, the winner is the party supplying the prime minister or premiers, and for the

3. See also Ellis and Thoma (1991), Chowdhury (1993), and Kayser (2006). Kayser (2005) derives a model to predict the degree of opportunistic election timing and manipulation under alternate institutional structures. Smith (1996, 2004) considers election timing with strategic signalling by assuming that the choice of election date reveals information about the government.

4. Sanders (2003) analyses polling error, and Alvarez and Keppo (2002) study the effect of delays.

5. We eliminate the Democratic one-party state of Georgia and the states/provinces where three parties have won: Connecticut, Maine, Minnesota, and Oregon, and B.C., Saskatchewan, Manitoba, Ontario, and Quebec. A two-party alternation obtains in all other states, provinces, and national governments.

6. Gubernatorial term limits (www.termlimits.org/Current_Info/State_TL/gubernatorial.html) apply in several states. About 10% of all governorships after 1930 ended due to term limits. The estimated chance that the ruling party changes after the term limit is active is 0·58 and 0·44 when the term limit is not active. At a 5% significance, we cannot reject the hypothesis of identical estimates. So we ignore term limits here.

U.S., we restricted attention to the presidency and the governorships. Our theory also assumes an easy information flow to the electorate about the merits of the competing parties. We begin with the first regime shift after 1930 (so that a power shift exists). Canada became a fully autonomous country in 1931, which makes this a focal starting decade. Also, if we choose earlier years, the parties have different names.

For each state, province, or country, we ask how many consecutive years the same government is in power. Delaware, for example, had its first post-1930 change of power in 1967; the government parties then changed power in 1971, 1987, 1991, and 1999. This yields five "ruling periods" over 1967–2005, or an average duration of $38/5 = 7{\cdot}6$ years, or 1·9 terms. Altogether, we have 46 data points for the U.S. and six for Canada. We find that the average government duration is 8·19 years for the U.S. and 15·43 for Canada—in other words, 2·05 four-year terms for the U.S. and 3·09 five-year terms for Canada. Using a pooled t-test, we find that $t = 2{\cdot}58$; we can confidently reject the hypothesis of equal mean numbers of terms. Clearly, the electoral timing option has significant value.[7,8]

We now try to precisely analyse this option and then test it for the U.K.

3. THE DYNAMIC POLITICAL PROCESS

3.1. *The changing political state*

An underlying and uncertain state variable describes the best political party for the country. This state variable is unobservable, randomly switching between left L and right R. There are only two *parties*, denoted also by L and R.[9] Party L, R is best in the unobserved *political state* $\theta = L, R$, respectively.

The state is random and persistent. Specifically, it follows an exogenous Poisson stochastic process, intuitively governed by the evolution of the political and economic situation. The state switches from $\theta = L, R$ in a time interval of length Δt with chance $\lambda_\theta \Delta t > 0$. Without a changing political state, the voters would eventually discern the true state via the information process below and an optimal ruling party would emerge.

3.2. *The information process*

We assume that a continuum $[0, 1]$ of voters passively learn about the unobserved political state, denoted $\theta(t)$ at time t. To escape complexities, we develop a tractable "informational representative agent" voter model. To wit, voters share a common understanding—a *political slant*—$p(t) = P[\theta(t) = R]$ that the optimal party is R. The electorate can be viewed as "right leaning" exactly when $p(t) > 0{\cdot}5$. This informational filtering story yields a tractable process for our analysis.

Voters freely learn about the political state from the newspaper, television, or radio. Specifically, we posit a *Gaussian public information process* ξ in continuous time: in other words, it is

7. A private member's motion was introduced into Canada's House of Commons in 2004 to shift the country towards fixed four-year terms. Commenting on election timing, the bill's sponsor said anyone in power would "call the election in the most self-serving moment for ourselves—and you'd be a fool not to". The Canadian provinces of British Columbia and Ontario have recently informally changed to fixed four-year terms.

8. The Canadian province of Quebec had a separatist government from 1976 to 1999. It seemed agreed that a majority in a referendum would allow the provincial government to initiate political separation from the rest of Canada. Trying to best time this vote using polling data proved an important activity and resulted in pro-separation votes just shy of 40% and 50% in the referenda called in 1980 and 1990.

9. In the empirical analysis we focus only on voters of the two big parties in the U.K. While the number of the big party voters could be stochastic, this would not matter since we model only proportions there. Our model extends to any number of parties, and in fact, Davydov and Smith (2000) considered three. To avoid the complexity of a multidimensional stopping time problem, we simply allow two here.

captured by the stochastic differential equation $d\xi(t) = \mu_{\theta(t)}dt + \gamma\, dZ(t)$, for some Wiener process $Z(t)$ and slopes $\mu_R > \mu_L$. More concretely, in state θ, in any Δt time interval, $\Delta\xi(t)$ is normally distributed, with mean $\mu_\theta \Delta t$, variance $\gamma^2 \Delta t$, and signals conditionally independent over time. So when the process greatly drifts up, the slant $p(t)$ rises; when it greatly drifts down, the slant falls. But all movements in $\xi(t)$ are obscured by high frequency noise, and so updating occurs slowly. Moscarini and Smith (2001) argue that this has some nice properties. For instance, it is a continuously unfolding ("non-lumpy") news process—its informativeness almost surely vanishes in the length of the time interval—and it is a time stationary ("constant intensity") process.

Since beliefs are constructed from information, the information process $\xi(\cdot)$ is clearly sufficient for the political slant process $p(\cdot)$. But the reverse holds true too: theorem 9.1 of Liptser and Shiryaev (2001) and Keller and Rady (1999) derive the next law of motion.

Lemma 1 (Dynamics). *The political slant $p(t)$ given signal $\xi(t)$ obeys Bayes' rule:*

$$dp(t) = a(b - p(t))dt + \sigma p(t)(1 - p(t))dW(t), \tag{1}$$

where $a = \lambda_L + \lambda_R > 0$, $0 < b = \lambda_L/a < 1$, $\sigma = (\mu_R - \mu_L)/\gamma > 0$, for a Wiener process W.

The drift expression is intuitive. The mean slant b is the fraction of the time we switch into state R. The mean reversion speed a is the flow switching chance. The noise term reflects Bayes' rule—after a Gaussian signal ξ in $[t, t+\Delta)$ with "chances" q_L and q_R in states L and R, we have $p(t+\Delta) - p(t) = p(t)q_R/[p(t)q_R + (1-p(t))q_L] - p(t) \propto p(t)(1-p(t))$.

Parameters a and b describe the political dynamics, while σ summarizes the quality of the information process. The more revealing is the public information process $\xi(t)$—as measured by the "signal-to-noise ratio" $(\mu_R - \mu_L)/\gamma$—the more volatile is the slant process. The parameter a captures the *speed of convergence* to the *mean* b. Intuitively, the expected slant reverts to b also, and at the exponential rate a. In the appendix, we prove

Lemma 2 (Future Beliefs). *If the political slant starts at p, then the expectation of $p(t)$ is $m(p,t) = e^{-at}p + (1 - e^{-at})b$. The variance of $p(t)$ increases in the diffusion coefficient σ.*

For example, starting with full Labour support, that is at $p = 0$, the expected slant after three years lies within 1% of the mean b by Lemma 2, given the estimated U.K. parameter $a = 1{\cdot}59$ (see Section 5.3). With such fast mean reversion speed, parties need not be very farsighted, since winning big is not much better than winning small. This speaks to the brief U.K. "political cycle"—the expected time it takes for the state to switch from L to R and back to L, or vice versa, equals $(1/\lambda_L) + (1/\lambda_R) = 1/(ab)(1-b) \approx 4/a \approx 2{\cdot}5$ years. Thus, there is time for more than one reversal of fortune during a typical electoral term.

A particularly convenient property of this political slant process is that its long-run density is analytically quite tractable, as we now assert (and prove in the appendix).

Lemma 3 (The Long Run Density). *The political slant process $p(t)$ forever remains in $(0,1)$, and the stationary political slant density $\psi(p)$ is given by*

$$\psi(p) \propto \frac{\exp\left(-\frac{2a}{\sigma^2}\left(\frac{1-b}{1-p} + \frac{b}{p}\right)\right)\left(\frac{p}{1-p}\right)^{2a(2b-1)/\sigma^2}}{p^2(1-p)^2}.$$

Figure 1 depicts the long-run density for the U.K. parameters estimated in Section 5.3. Since this density is single peaked, this in itself is a finding of the model, because one can show that

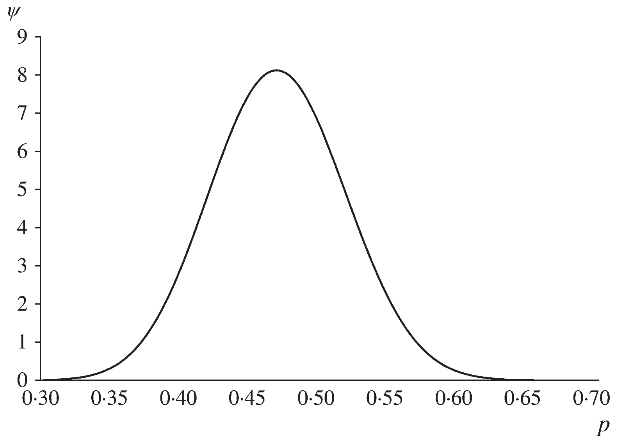

FIGURE 1

The long-run density of the political slant p in the U.K. The estimated parameters are $a = 1{\cdot}59, b = 0{\cdot}47, \sigma = 0{\cdot}39$. The chance that L wins is $P(p \leq 0{\cdot}5) = 70\%$. In the model we assume these parameters are known and derive the voting process below in Proposition 1. This result then allows us to estimate them in Section 5 from polling data

not all densities of Lemma 3 are hill shaped. Rather, the density $\psi(p)$ is U-shaped for high belief variances σ. For then, state switches quickly become known, and the political slant spends most of its time near 0 or 1. We have found that this is not true for the U.K. Also, since the estimated $b < 0{\cdot}5$ for the U.K. the process favours L—on average, L is ahead $P(p \leq 0{\cdot}5) = 70\%$ of the time. So the U.K. enjoys a left slant.

3.3. *Preference heterogeneity*

Voters agree on the best party in each state, but—uncertain of the political state—differ in their preference strength. Some are more willing to err on the side of left and some on the side of right. Turning to the *preference parameters*, type-(u, υ) voter has cardinal utility 0 if the wrong party is elected, utility $u > 0$ if L is rightly elected, and utility $\upsilon > 0$ if R is rightly elected. So he earns expected pay-off $[1 - p(t)]u$ from L, and $p(t)\upsilon$ from R (see Figure 2). A farsighted voter might rationally anticipate the mean reversion of the state and vote against his immediate preferences. We ignore such higher order rationalizations, assuming that voters choose R if $p(t) > u/(u+\upsilon)$ and L if $p(t) < u/(u+\upsilon)$. So a voter becomes more left leaning (or right leaning) as $u/\upsilon \to \infty$ (or 0), and in the limit, never votes R (or L). This framework subsumes doctrinaire voters as a special case.

Lemma 4 (Political Slants Become Electoral Support). *If preference parameters u and υ are independently and identically distributed across voters, and they have a common exponential density, then $p(t)$ is the fraction of voters for party R in any election at time t.*

Proof. The fraction of voters supporting party R is the total fraction of the parameters (u, υ) for which $\upsilon > [1 - p(t)]u/p(t)$. This equals the double integral

$$\int_0^\infty \lambda e^{-\lambda u} \int_{[1-p(t)]u/p(t)}^\infty \lambda e^{-\lambda \upsilon} d\upsilon\, du = \int_0^\infty \lambda e^{-\lambda u} e^{-\lambda[1-p(t)]u/p(t)} du = p(t) \int_0^\infty \lambda e^{-\lambda w} dw = p(t).$$

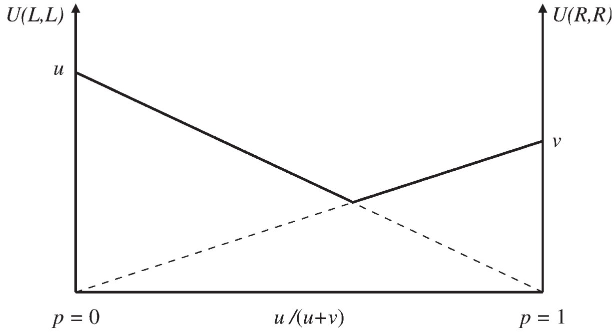

FIGURE 2

A voter's preferences. The figure schematically depicts a typical voter's utility maximization: he votes for L if $p < u/(u+v)$, the cross-over level, and otherwise he votes for R

The exponential distribution ideally captures the fact that extreme preferences are very rare. But its primary benefit is that it produces a tractable theory for which *the stochastic process of support for the right party R exactly coincides with the political slant p.* ‖

Proposition 1 (Dynamics). *The process* (1) *gives the electoral support dynamics for R.*

This result is key to the analytic and empirical tractability of our model. In other words, we now have a Bayesian learning based law of motion (1) for the support of the parties.

4. OPTIMAL ELECTORAL TIMING

4.1. *Stopping and waiting values*

We assume that the government seeks to maximize the expected total time in power *in the current streak*. One might think of this as the objective of the prime minister, since he usually is not around after falling from power. Alternatively, it is hard for a government to think beyond the current streak, since it is not able to affect the timing of an election for many years to come. But as it turns out, the difference between winning big and small is so negligible that concern for elections long after one is defeated has essentially no effect on electoral timing. The government opts whether to call an election or not, weighing the cost of losing the rest of the current term with an earlier election against the benefits of a higher re-election chance. After any election, the next must be called within T years. Once called, a fixed *delay time* $\delta > 0$ passes during the campaign. This delay is critical for us, since the ruling party does not know the outcome of the election when it is called.

The decision to call an election is an optimal stopping exercise. The *stopping time* τ is a function of the remaining time until the next election $T - t$ and the political slant $p(t)$. When the ruling party i follows an optimal strategy, we denote its expected time in power at time t by $F^i(p,t)$—the *waiting value*—and its expected time in power *once an election is called* by $\Omega^i(p)$—the *stopping value*. (We drop party superscript when possible.)

If party i wins when the political slant is p, then it enjoys an expected waiting value $F^i(p,0)$. Since R loses when $p < 1/2$, and L loses otherwise, we have $F^R(p,0) = 0$ for all $p < 1/2$ and $F^L(p,0) = 0$ for all $p > 1/2$. So party R enjoys the expected time in power $F^R(p,0) > 0$ iff the p process is above $1/2$ at the time of the election. The chance of a tie in our model is zero by the continuous probability density corresponding to (1).

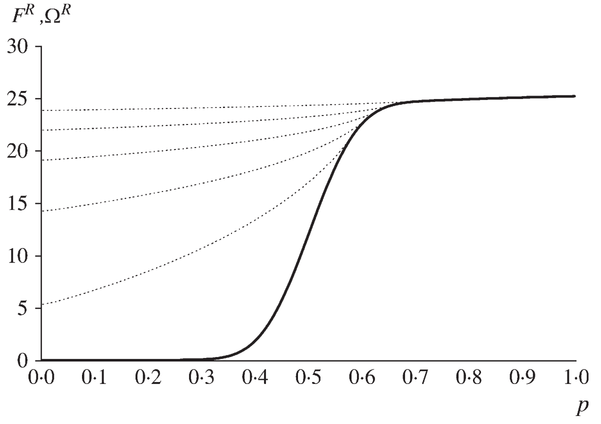

FIGURE 3

Waiting and stopping values F^R and Ω^R for the U.K. From top to bottom, the (dotted) numerically computed waiting value functions are $F^R(p, 0+)$ (just after winning), $F^R(p, 1)$, $F^R(p, 2)$, $F^R(p, 3)$, and $F^R(p, 4)$. Party R should not call an election at any time t with $F^R(p, t) > \Omega^R(p)$, the (solid) stopping value

The pay-offs at the stopping time τ are $\Omega(p(\tau))$, for the recursively defined function:

$$\Omega(p) \equiv \delta + E[F(p(\tau+\delta), 0) \mid p(\tau) = p]. \tag{2}$$

The value of a standard put or call option is continuous in the underlying price. Thus, the option is not worth much when it is only slightly "in the money" at the expiration date. By contrast, one vote can separate the glory of victory from the sting of defeat with elections: landing slightly "out of the money" is discontinuously worse than the alternative. To wit, Ω includes a *binary option* in (2) paying at maturity the "asset or nothing"—here, paying F or 0. As Ω lacks an optimal timing exercise, it is a "European option".

Easily, since a government has the option of running out its full term, this is a lower bound on its longevity: $F(p, t) \geq T - t$. Forward-looking behaviour generally entails an earlier election, since we care about the expected value $\Omega(p)$ once the election is called.[10] Since this is the sum of the time until the election and the continuation value, we have

$$F(p, t) = \sup_{t \leq \tau \leq T} E_{p(t)=p}[\tau - t + \Omega(p(\tau))]. \tag{3}$$

By recursive equations (2) and (3), $F(p, t)$ is an American option on the binary European option Ω. These must be solved numerically—since even the standard non-recursive American put option with a geometric Brownian motion is not analytically solved. This exercise is illustrated in Figure 3, and Appendix B gives the numerical algorithm.

Proposition 2 (Existence of Smooth Values). *There exist* Ω^i, F^i *solving* (2) *and* (3) *where* Ω^i *is smooth, and* F^i *is smooth when* $F^i(p, t) > \Omega^i(p)$.

This is proven by recursive means in Appendix A.

Define the *expected drift* $\mathcal{A}F(p, t)dt$ of the waiting value $F(p, t)$ on $[t, t + dt)$:

$$\mathcal{A}F(p, t) = F_t(p, t) + a(b - p)F_p(p, t) + \tfrac{1}{2}\sigma^2 p^2 (1 - p)^2 F_{pp}(p, t). \tag{4}$$

10. Maximizing the expected time in power corresponds to an interest rate $r = 0$. Later on, we consider $r > 0$, which is the only other time consistent objective function for optimal stopping (Smith, 1997).

By Proposition 2, $\mathcal{A}F(p,t)$ is well defined. Intuitively in (4), we see that F changes in t by $F_t dt$; the drifting political slant on average pushes F by $F_p E[dp] = a(b-p)F_p dt$, and its volatility changes F by $\frac{1}{2}F_{pp}(dp)^2 = \frac{1}{2}\sigma^2 p^2(1-p)^2 F_{pp}dt$. This final Ito term shows that volatility matters when F is non-linear, improving option values when $F_{pp} > 0$.

Next we analyse the optimal exercise strategy of this electoral option.

Proposition 3 (Optimality). *The best election time is the first time τ before T such that $F(p(\tau),\tau) = \Omega(p(\tau))$. Also, for all $(p,t) \in (0,1)\times(0,T)$, we have*

(a) *Calling an election is always an option:* $F(p,t) \geq \Omega(p)$

(b) *The value is expected to fall daily by at least one day:* $1 + \mathcal{A}F(p,t) \leq 0$,

where for each political slant p and time t, one of the inequalities (a) *or* (b) *is tight.*

These are standard variational inequalities (see, for example, Øksendal, 2003) for the value (3). For instance, that $F(p,t) > \Omega(p)$ implies $\mathcal{A}F(p,t) = -1$ says that "when waiting is better than stopping, the unit flow pay-off balances the expected time lost in office". Here, the waiting value $F(p,t)$ is expected to fall one day for every day in office until the election is called (while $t < \tau$). Once the waiting and continuation values coincide, $F = \Omega$, further delay hurts. Figure 4 illustrates the situation. By complementary slackness, the government either waits or calls an election, that is, one of inequalities (a) or (b) is tight.

Stock option values are convex in the underlying price, simply because greater risk pushes weight into the exercise tail (in the money). This convenient property holds for the electoral option. Observe that if the waiting value F is convex in p, then as an expectation, $E[F(P,t)] \geq F(E(P),t)$ by Jensen's inequality for all random variables P. The appendix shows how to reverse this logic, and deduces that because information both has value to the government, and adds variance to the belief, the function F is convex.

Lemma 5 (The Convex Waiting Value). *The waiting value $F(p,t)$ is a convex function of the political slant p for all times $t < \tau$. In particular, $F(p, 0+)$ is convex in p if $\tau > 0$.*

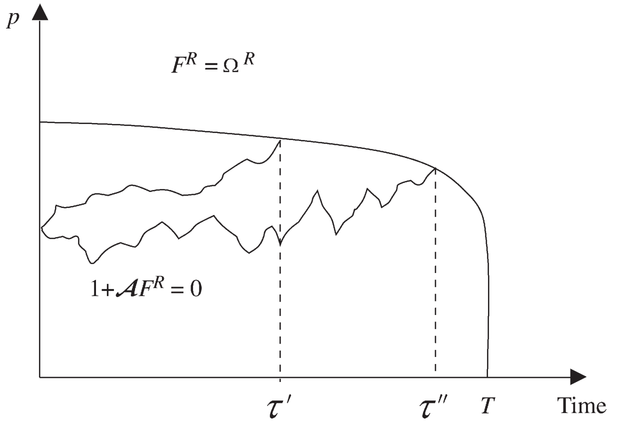

FIGURE 4

Electoral timing for party R. The figure schematically depicts the electoral option exercise problem facing party R (see Proposition 3). It calls an election when the political slant $p(t)$ hits the election barrier. Two paths for $p(t)$ are shown, with election times τ' and τ''

Define the *time value* of the electoral option $\Gamma(p,t) = F(p,t) - \Omega(p) \geq 0$.

Lemma 6 (Value Monotonicity). *Values are monotone in p, b, and t. Specifically*

(a) *The waiting and time values $F(p,t), \Gamma(p,t)$ fall in time t, and rise in the horizon T.*
(b) *The values $F^R(p,t)$ and $\Omega^R(p)$ rise in p, and $F^L(p,t)$ and $\Omega^L(p)$ fall in p.*
(c) *The values $F^R(p,t)$ and $\Omega^R(p)$ rise in b, and $F^L(p,t)$ and $\Omega^L(p)$ fall in b.*

To see part (a), recall that an American option is worth more with a longer time horizon in (3). The appendix proves (b). Part (c) follows from (b), as each path $p(t)$ rises in the parameter b.

4.2. *Election barriers and election timing*

For $t < T$, the *election barrier* $p = \beta^R(t)$ is the least solution to $F^R(p,t) = \Omega^R(p)$, and $p = \beta^L(t)$ the maximum solution to $F^L(p,t) = \Omega^L(p)$. The *continuation regions* are those (p,t) where respectively $p < \beta^R(t)$ and $p > \beta^L(t)$. Lemma 7 claims that at the barrier the waiting and stopping values coincide and are smoothly matched (see Figure 3): the appendix reformulates the optimal stopping problem using smooth pasting.

Lemma 7 (The Boundary Value Problem). *In the continuation region, the waiting value F solves the partial differential equation $1 + \mathcal{A}F(p,t) = 0$, for the boundary conditions*

(a) *value matching: $F(\beta(t),t) = \Omega(\beta(t))$;*
(b) *smooth pasting: $F_p(\beta(t),t) = \Omega_p(\beta(t))$.*

While Ω is globally smooth in p, the stopping value F is only smooth in the continuation region. Thus, the second derivatives need not be matched, but may be strictly ranked.

Corollary 1 (Second Derivatives). *$F^R_{pp}(p-,t) \geq \Omega^R_{pp}(p)$ and $F^L_{pp}(p+,t) \geq \Omega^L_{pp}(p)$ at $p = \beta(t)$.*

Proof. By Proposition 3 and Lemma 7, $F^R(p,t) - \Omega^R(p) \geq 0$ left of the barrier, and $F^R_p(\beta(t),t) - \Omega^R_p(\beta(t)) = 0$ at the barrier. So $F^R(\beta(t),t) - \Omega^R(\beta(t))$ is locally convex. Since both derivatives exist by Proposition 2, inequality follows by taking limits. For party L, the argument proceeds right of the barrier. ||

The appendix proves that optimal election barriers are monotone, in the following sense:

Proposition 4 (Monotonicity). *The election barriers are both continuous and monotonic. In particular, $\beta^R(t)$ falls in t, and $\beta^L(t)$ rises in t.*

The stopping value Ω is a non-trivial recursive function beyond the reach of most formal analysis. For insight into Ω, it helps to study the *victory chance* $V^R(p)$, namely, the chance that party R wins the election which is called when the political slant equals p. Intuitively, $V^R(p)$ smoothes out the step function that $p(\cdot)$ exceeds $1/2$ on election day:

Lemma 8 (Victory Chance). *The chance $V^R(p)$ is convex when $p < \frac{1}{2}e^{a\delta} + (1 - e^{a\delta})b$, and concave when $p > \frac{1}{2}e^{a\delta} + (1 - e^{a\delta})b$—and conversely for $V^L(p)$.*

Loosely, if victory is expected on election day, then there are diminishing returns to improving the political slant p. Recalling Lemma 2, the appendix proves that concavity begins at the slant where we expect to win the next election—that is, for party R, when $p \geq \frac{1}{2}e^{a\delta} + (1-e^{a\delta})b$, so that the expected poll after the δ time elapse is $m(p,\delta) \geq 1/2$.

The government calls an election when the political slant hits the stopping barrier $\beta(\cdot)$. Upon winning, it acquires a new waiting value function $\phi(p) \equiv F(p, 0+)$, calculated from (3) by taking the limit $t \downarrow 0$. As the next election is so far into the future, the margin of victory should have an insignificant impact on the expected time in power: for as we will verify in Section 5.3, the mean reversion in the U.K. is so fast that the slant p is expected to lie within 1% of the mean within three years. So one can closely approximate the stopping value by

$$\Omega(p) \approx \delta + V(p) \int_0^1 \phi(y)dy. \tag{5}$$

Governments can essentially act as if they are just trying to win back a single term of fixed length, not looking past the next election. We see that Ω intuitively inherits the convex–concave shape of the victory chance V, as in Figure 3.

From (2) and the victory chance definition, we get a sandwich for the stopping values

$$\begin{aligned} \phi^R(0)V^R(p) \leq \Omega^R(p) - \delta \leq \phi^R(1)V^R(p) \\ \phi^L(1)V^L(p) \leq \Omega^L(p) - \delta \leq \phi^L(0)V^L(p). \end{aligned} \tag{6}$$

By Lemma 8, these upper and lower bounds of the stopping value Ω share a convex–concave shape. In Figure 3, we see that $\phi(0)$ and $\phi(1)$ are close. Pushing a little harder on the fast mean reversion of the slant, the initial margin of victory only slightly affects the expected time in power. This suggests writing (5) as $\Omega(p) \approx \delta + V(p)\phi$, where $\phi(p) \approx \phi$ (a constant). The above sandwich inequality therefore offers some analytical support for the observed convex–concave shape of Ω.

The numerical simulation in Figure 3 also suggests that elections are called at political slant levels where Ω is concave. We can formally establish a slightly weaker result:

Lemma 9 (Waiting). *Party R never calls an election if $\Omega^R(p)$ is locally convex and $p \leq b$, while party L never calls an election if $\Omega^L(p)$ is locally convex and $p \geq b$.*

Proof. By Proposition 2, $\Omega^R_p(p) \geq 0 \geq \Omega^L_p(p)$. If $a(b-p)\Omega_p + \frac{1}{2}\sigma^2 p^2(1-p)^2\Omega_{pp} + 1 > 0$ then waiting is profitable. This holds under the given conditions. ‖

As an application, if in addition to (5), the second derivative approximation $\Omega''(p) \approx V''(p)\phi$ also holds, then by Lemma 8, $\Omega^R(p)$ should be convex for $p \leq 1/2$ when $b < 1/2$. Likewise, $\Omega^L(p)$ should be convex for all $p \geq 1/2$ when $b > 1/2$. Lemma 9 would then imply that party R never calls an election if $p < b < 1/2$, and party L never if $p > b > 1/2$. Loosely, the above conditions guarantee that the expected downward drift in its electoral standing is more than compensated by the extra day in power. This conclusion is consistent with Figure 3, suggesting that the second derivative approximation is valid.

The recursive structure of our model has denied us proofs for many intuitive and numerically true comparative statics. We now provide intuitions for these.

- THE VOLATILITY σ. By Lemma 2, the variance of the political slant at the barrier rises in σ. Since by Corollary 1, the waiting value is more convex than the stopping value at the barrier, $F_p^i(\beta^i(t),t) - \Omega_p^i(\beta^i(t))$ rises in σ for $t < T$. To restore smooth pasting, the election barriers intuitively should shift out. The effect on the expected ruling time is unclear. Indeed, as we saw in Figure 3, and reasoned after (6), the government acts like a decision maker with a convex–concave "utility function", and so is ambivalent about risk.

- THE MEAN REVERSION SPEED a. By Lemma 2 and equation (9), the variance of $p(t)$ falls in the speed a, because the slant reverts faster towards the mean level b. This either helps or hurts the government depending on whether $b > 1/2$ or $b < 1/2$. Since smaller a is tantamount to greater σ, the election barriers shift out, by the above volatility analysis.
- THE CAMPAIGN PERIOD δ. The election outcome depends less on the initial campaign period poll level, the longer is the campaign period—due to the mean reversion. This lowers the benefit from an early election, and thereby pushes out election barriers. Yet the government may still prefer higher δ, as the maximum time in power $T + \delta$ rises.
- THE MEAN LEVEL b. By Lemma 6 (c), the expected time in power rises in the mean level b for party R, and falls for L. Calling an early election is a calculated gamble that weighs the marginal benefit of waiting $\mathcal{A}F$ against the marginal cost of losing, namely the extra day in power. With a greater b, we see in (4) that the marginal benefit rises for party R and falls for party L. Altogether, the barriers should be pushed up for both parties.

5. POLLS: THEORY, DATA, AND ESTIMATION

We now estimate our support process in Section 3 using the noisy realized polls.

5.1. *The discrete time polling process*

Politicians enjoy a variety of ways to take the pulse of the electorate—many quite qualitative. We wish to assume that governments time their elections using monthly voting intention polls. These surveys ask individuals who are planning to vote whom they would pick in a hypothetical election the next day. Since the government consists of citizens privy to the information process $\xi(t)$, our model possibly accords no informative value to the polls (see Section 3.2). To escape a filtering exercise (see Section 8), we venture a story with a mild boundedly rational flavour. Imagine that voters cannot operate Bayes' rule, but nonetheless know whom they would vote for. Indeed, voting is a simple binary decision, and requires less introspection than deducing a probability via the non-linear Bayes' rule.[11] Governments can then learn from the polls, since these are noisy observations of the true but unobserved political support $p(t)$. By Lemma 4, one can view the political slant as the support process for party R—so that (1) is the law of motion for the support for party R, from Proposition 1.

In a given time-t poll with sample size N, let $\pi(t) = p(t) + \eta(t)$ be the fraction of voters that support R. The poll error η obeys a t-distribution with variance $\pi(1-\pi)/N$.[12] So η is approximately $\sqrt{\pi(1-\pi)/N}\varepsilon$, where ε is a mean 0 and variance 1 normal r.v.

To estimate the model, we now write the polling process at the discrete poll times $\{t_j\}$. We wish to imagine these polls as periodic observations of a phantom continuous time polling

11. Two other non-behavioural stories come to mind. Polls may be relevant if there are "noise voters"—who vote randomly, unswayed by the slant. The support would behave approximately like the political slant. Polls would then be useful as they record the actual voting intentions and follow the law of motion (7). More subtly, we may diverge from the informational representative agent and assume heterogeneously informed agents. In aggregate, the voting intentions would again obey approximately the same law of motion as the political slant, and (7) would apply. The complexity of neither approach is justified.

12. Polls enjoy a binomial distribution, which is asymptotically normal, by the Central Limit Theorem. The variance of this normal distribution is unknown, and so the t-distribution applies.

process. Since the polling error does not depend on the gap $\Delta_j \equiv t_{j+1} - t_j$ between polls, more frequent polls corresponds to a greater *polling volatility* ς.

Lemma 10 (Poll Dynamics). *The discrete time polling process is approximately*

$$\pi(t_{j+1}) - \pi(t_j) \approx a(b - \pi(t_j))\Delta_j + \varsigma(\pi(t_j), N\Delta_j)\pi(t_j)(1 - \pi(t_j))\sqrt{\Delta_j}\varepsilon_j, \tag{7}$$

where $\varepsilon_j \sim N(0,1)$ *and* $\varsigma(\pi, N\Delta) > \sigma$ *falls in* $N\Delta$, *with* $\lim_{N\Delta\uparrow\infty} \varsigma(\pi, N\Delta) = \sigma$. *If we have an election at time* t_{j+1} *or at time* t_j *then the volatility equals* $\varsigma(\pi(t_j), \sqrt{2}N\Delta_j)$.

In other words, the polling process (7) consists of discrete time snapshots of a noisier political slant process from Lemma 1.[13] Lemma 10, proved in the appendix, follows because the $p(1-p)$ volatility term in $p(t)$ dominates with large poll sample sizes.

Observe how today's polling result is not a best forecast of the election outcome, since the electoral process mean reverts:[14] a government riding high in the polls believes that its trend is most likely down,[15] with mean reversion. Further, the election day corresponds to poll volatility with a slightly longer elapse time $\sqrt{2}\Delta$.

5.2. *Polling history and data*

Our data set from the U.K. consists of two poll time series of voting intentions dating from June 1943 to May 2005.[16] The sample sizes are large, mostly between 1000 and 1500. The average time between consecutive polls is 21 days. Our data set begins with Gallup polls from June 1943 to May 2001. Gallup's voting intention polls were discontinued in 2001. We thus add a second data set, the MORI Political Monitor (ipsos-mori.com), spanning August 1979–May 2005; its sample size varies from 500 to 17,000. We average same-day polling results of Gallup and MORI.

We study only the voters of the big parties in the U.K.—Labour (L) or Tory (R). While the number of such voters is stochastic, this does not matter as we consider proportions. Thus, from the polls we calculate the realized values of π, which is now the Tory polling support among the big party voters. Figure 5 depicts the poll levels π from June 1943 to May 2005. The polls on average have favoured Labour, and the average poll is 0·46.

5.3. *Estimating the polling process*

Equation (7) is an autoregressive model.[17] We estimate the model parameters by OLS: transform the dependent variable of (7) into $Y_j = (\pi(t_{j+1}) - \pi(t_j))/[\pi(t_j)(1 - \pi(t_j))\sqrt{\Delta_j}]$ and its explanatory variables into $X_j = \sqrt{\Delta_j}/[\pi(t_j)(1 - \pi(t_j))]$ and $Z_j = -\sqrt{\Delta_j}/[1 - \pi(t_j)]$. We first estimate the parameters a and b in (7) from the regression[18]

$$Y_j = (ab)X_j + aZ_j + \varsigma\varepsilon_j,$$

13. In finance theory, prices may be modelled as if in continuous time, despite discrete time observations. This corresponds to a process with a certain fixed elapsed time, such as $\Delta = 1$.

14. Kou and Sobel (2004) find that election financial markets better predict election outcomes than polls.

15. This is consistent with Smith (2003), who finds that when calling an early election, one experiences a decline in one's popular support relative to pre-announcement levels.

16. Since polls ask "If there were a general election tomorrow, which party would you vote for?", we assume that each is simply a noisy observation of the actual election outcome that would have obtained that day. Respondents saying "don't know", "none", or who refused are removed from the base.

17. Sanders (2003) shows that such an autoregressive model gives accurate forecasts for the U.K. polls.

18. The delta-method (see, for example, Casella and Berger, 2002) gives the S.D. of b, σ, and σ_η.

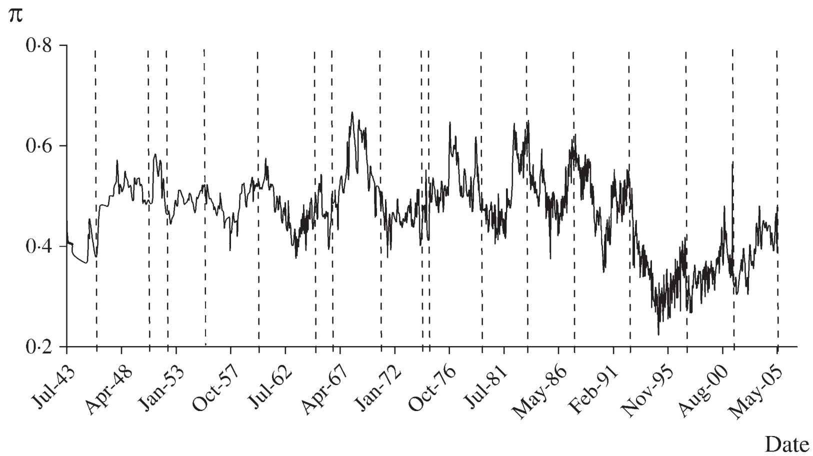

FIGURE 5

Proportion of Tory among the big party voters, 6/1943–5/2005. The polling process π has averaged 0·46 (*i.e.* left), and ranged from 0·23 to 0·67. Dashed lines are elections

TABLE 1

The estimated polling parameters

	R^2 : 3·61%, 1·92%					R^2 : 3·35%, 1·68%			
δ-period	a	b	σ	σ_η	δ-period (no weak)	a	b	σ	σ_η
Estimate	6·43	0·52	1·01	0·09	Estimate	6·09	0·52	1·08	0·09
S.D.	3·32	0·04	0·55	0·03	S.D.	3·90	0·05	0·60	0·04
Pre-δ	R^2 : 2·61%, 26·43%					R^2 : 2·12%, 9·88%			
	a	b	σ	σ_η	Overall	a	b	σ	σ_η
Estimate	1·49	0·46	0·27	0·12	Estimate	1·59	0·47	0·39	0·12
S.D.	0·29	0·02	0·10	0·00	S.D.	0·33	0·02	0·14	0·01

Notes: The "overall model" uses all the data; the pre-δ-period uses data before the election time is announced; the δ-period uses data after the election is announced. The first R^2 is for the Y regression and the second for the V regression. All parameters are significant (the label "weak" refers to the governments without a clear majority).

where $\varepsilon_j \sim N(0,1)$. The estimation square error $V_j = (Y_j - (ab)X_j - aZ_j)^2$ captures the polling process error. Using (10) in Appendix 10, write $\varsigma = \sqrt{\sigma^2 + \sigma_\eta^2 d_j}$, where $d_j = 1/(\sqrt{2}\Delta_j)$ if an election (by Lemma 10), and $d_j = 1/\Delta_j$ otherwise. We estimate

$$V_j = \sigma^2 + \sigma_\eta^2 d_j + \varsigma e_j,$$

where each $e_j \sim N(0,1)$, hereby implicitly assuming σ_η is constant.[19]

Table 1 gives four estimated parameter sets: overall, outside and inside the δ-period, and finally inside the δ-period without four governments that we call "weak" in Section 6.1.[20]

19. According to (10) and our data set, this is justified since about 97% of π values lie in $[0{\cdot}3, 0{\cdot}7]$, which implies that $\sqrt{\pi(1-\pi)} \in [0{\cdot}46, 0{\cdot}50]$. By (10) and assuming $N \approx 1000$, $\sigma_\eta = \sqrt{k(\pi)/(N\pi(1-\pi))} \approx \sqrt{1/500}/(\pi(1-\pi))$, which is between $[0{\cdot}126, 0{\cdot}137]$, that is, close to our estimate (0·12).

20. These R^2 levels may seem low, but are very good by comparison to the best empirical work in financial time series (see, for example, Table 3 in Campbell and Thompson, 2005).

While the parameter estimates for a, b, σ, and σ_η are different outside and inside the δ-period, these differences are not statistically significant.[21] On the other hand, as seen in Table 1 all parameter estimates are statistically significant. We can offer two internal consistency checks on these estimates. First, the mean poll level $b = 0{\cdot}47$ is near the average poll level 0·46 in Figure 5. Second, our estimate $\sigma_\eta = 0{\cdot}12$ in Table 1 is near a direct computation of the S.D. using our t-distribution formula in (10): $\sigma_\eta^2 \approx 2/[N\pi(t_j)(1-\pi(t_j))] \approx 1/125 \approx 0{\cdot}09^2$, for $N \approx 1000$ and $\pi(t_j) \approx 1/2$.[22]

Next, the average polling time gap outside the δ-period is 0·059 years (about 22 days) and its S.D. is 0·043 (about 16 days). Inside the δ-period, these numbers are 0·023 years (about eight days) and 0·026 (about nine days).[23] Thus, the average poll volatilities differ inside and outside the δ-period. But this owes to the smaller elapsed time Δ between polls prior to an election. In the next section, we ignore this, and assume a constant polling volatility. This is for the conceptual simplicity, since had we proceeded with the richer model, our main results would still be significant, as we discuss later (footnote 31).

To estimate the constant polling volatility ς, we use the polling time differences over the entire data set (average: 0·056 years $\approx$ 21 days, S.D.: 0·054 $\approx$ 20 days). This gives the constant volatility 0·89. For some perspective, this amounts to a fraction 48% of the average polls, using the quick approximation $\varsigma\pi(1-\pi)/\pi \approx 0{\cdot}48$. By comparison, the volatility of the S&P 500 stock market index has been about 14% over 1950–2005. Thus, volatility looms as a significantly greater factor in political than financial markets, even in a stable democracy like the U.K.

Altogether, we have established that our model in Sections 3–4 can be implemented with a fictitious continuous time polling process corresponding to our estimated discrete time process. The waiting value then satisfies the modified PDE $\mathcal{A}F(\pi,t) = -1$, where $\mathcal{A}F(\pi,t) = F_t + F_\pi a(b-\pi) + \frac{1}{2}F_{\pi\pi}\varsigma^2\pi^2(1-\pi)^2$. Further, the stopping value now solves

$$\Omega(\pi) = \delta + E[F(\pi(\tau+\delta),0) \mid \pi(\tau) = \pi].$$

With Proposition 3 and Lemma 7, this gives the optimal election barriers in (π,t)-space.

6. ACTUAL VS. OPTIMAL ELECTION TIMING

We now explain the variation of the U.K. governments' election decisions, comparing theoretical and realized political support levels at the times of election calls. Because the comparison is done for the parameter estimates from Section 5, the poll history can be understood as the sample data, and the support levels as the out of sample data.

6.1. *Election history and outcomes*

The Prime Minister chooses when to call an election by asking the Queen to dissolve parliament. She then issues a Royal Proclamation for writs to be sent out for a new parliament, starting the election timetable. According to the Parliament Act of 1911, the election must be called within $T = 5$ years. This has been extended twice—during the world wars, just after which our data set starts. The election timetable lasts 18 days, plus weekends and public holidays. It starts with

21. The t-statistic for the test that a coincides outside and inside the δ-period (without the weak governments) is 1·48 (1·18). The analogous t-statistics for b, σ, and σ_η are 1·34 (1·11), 1·32 (1·33), and $-1{\cdot}00$ ($-0{\cdot}75$). The joint hypothesis that parameters do not change cannot be rejected at the 1% significance level.

22. The S.D. of the polls is $\sqrt{\pi(1-\pi)/N} \approx \sigma_\eta\pi(1-\pi)/\sqrt{2} = (0{\cdot}12)(0{\cdot}25)/\sqrt{2} \approx 2\%$. This is consistent with Sanders (2003).

23. Here we consider the polling frequency change due to election outcomes (Lemma 10). We reject equal average polling time differences inside/outside the δ-period at the 1% significance level.

TABLE 2

U.K. Election results, 1945–2005 Tory and Labour columns are the vote percentage (seats percentage) for the two main parties

Election date	Announced	Winner	Tory	Labour	(π, t)
5.7.1945	23.5.1945	Labour	39·8 (33·3)	47·8 (61·4)	
23.2.1950	11.1.1950	Labour	43·5 (47·8)	46·1 (50·4)	(0·52,4·52)
25.10.1951*	19.9.1951	Tory	48·0 (51·3)	48·8 (47·2)	(0·57,1·57)
26.5.1955	15.4.1955	Tory	49·7 (54·8)	46·4 (44·0)	(0·51,3·47)
8.10.1959	8.9.1959	Tory	49·4 (57·9)	43·8 (41·0)	(0·54,4·29)
15.10.1964	15.9.1964	Labour	43·4 (48·3)	44·1 (50·3)	(0·49,4·94)
31.3.1966*	28.2.1966	Labour	41·9 (40·2)	47·9 (57·6)	(0·46,1·37)
18.6.1970	18.5.1970	Tory	46·4 (52·4)	43·0 (45·6)	(0·46,4·13)
28.2.1974	7.2.1974	Labour	37·9 (46·8)	37·1 (47·4)	(0·48,3·64)
10.10.1974*	18.9.1974	Labour	35·8 (43·6)	39·2 (50·2)	(0·44,0·55)
3.5.1979*	29.3.1979	Tory	43·9 (53·4)	36·9 (42·4)	(0·54,4·47)
9.6.1983	9.5.1983	Tory	42·4 (61·1)	27·6 (32·2)	(0·61,4·02)
11.6.1987	11.5.1987	Tory	42·2 (57·9)	30·8 (35·2)	(0·58,3·92)
9.4.1992	11.3.1992	Tory	41·9 (51·6)	34·4 (41·6)	(0·49,4·75)
1.5.1997	17.3.1997	Labour	30·7 (25·0)	43·2 (63·4)	(0·37,4·94)
7.6.2001	8.5.2001	Labour	31·7 (25·2)	40·7 (62·5)	(0·36,4·02)
5.5.2005	5.4.2005	Labour	32·3 (30·5)	35·2 (55·0)	(0·46,3·83)

Notes: The (π, t) columns lists the poll level π and the time t from the last election at the time election was announced.

*The elections were called for because of minority or weak governments.

the dissolution of Parliament and the issue of writs on day 0, and ends on day 17, election day (a Thursday, since 1935). While election season starts with the dissolution, one may extend this period by announcing an election before dissolution, as has been done just once.[24] Table 2 lists the outcomes of 17 British elections from 1945 to 2005. While the delay time ranges from 21 to 45 days, we fix δ at the average delay time 33 days (or 0·09 years) in our numerical solution.

We see that on average, governments have called elections after 3·65 years in our data set. We single out three unusually short governments: 23 February, 1950 to 25 October, 1951 (609 days), 15 October, 1964 to 31 March, 1966 (532 days), and 28 February, 1974 to 10 October, 1974 (224 days). Each had a slim victory, with seat proportions ranging from 0·485 to 0·515. Since all other governments lie outside this range, these are the only ones that we call *initially weak*.[25] The lifespan for all other governments has averaged 4·23 years. In 1951, Attlee's Labour government called an election only 20 months into his term. For he deemed his narrow majority of just five MPs insufficient to sustain his radical programme creating the welfare state that was started with the large majority that Labour enjoyed from 1945 to 1950. Attlee lost the election to Churchill, ushering in 13 years of Tory rule. A Labour election in 1966 after two years, given a slimmer majority of four, led to a win. Finally, beset by a minority government, Labour held and won an election after just seven months in 1974.

In addition to these three governments, a fourth one merits special consideration. The 1974–1979 government expired just four months before its legal term expired, after losing a non-confidence vote by just one vote. After a sequence of by-election defeats, it had drifted into a minority position in March 1977 and was thereafter sustained by the "Lib–Lab" pact. The Tories under Thatcher defeated Callaghan's Labour government in 1979, initiating 18 years of

24. In 1997, John Major announced the election on March 17 but did not dissolve parliament until April 8. As he was *behind in the polls* and just weeks away from the terminal date, this is one case where a longer campaign period is actually desired, notwithstanding the comparative static for δ in Section 4.2.

25. John Major's 1992–1997 Tory government won a small majority of just 21 seats—its seat proportion was 0·516. We consider this a regular government, but this choice is moot, as we show in footnote 33.

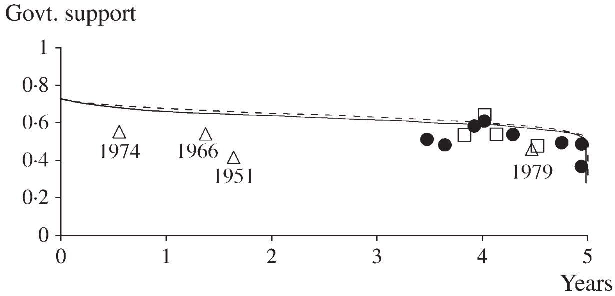

FIGURE 6

Election barriers. Depicted are the election barriers in support space (π for R and $1-\pi$ for L)—a solid line for Tories and a dashed line for Labour. The triangles are the elections that weak Labour governments called: namely, 25 October, 1951, 31 March, 1966, 10 October, 1974 (began with seat proportions below 0·515) and 3 May, 1979 (whose election was forced after it fell into a minority). The squares are all other Labour election announcement support levels, and circles are all Tory ones. The barriers for L and R averaged 0·64 and 0·63, respectively

Tory rule. Altogether, we simply call this and the three initially weak governments (coincidentally, all Labour) *weak*. Since weak governments clearly do not enjoy their time in power as much, we modify their objective function in Section 6.3, to capture why they might call earlier elections; for now, we keep them in the data set.[26]

6.2. *Election timing*

We now analyse election timing for the overall polling process estimated in Table 1. Given the complexity of the continuous time polling process, only a numerical solution of the finite time horizon stopping exercise is possible (see Appendix B). The election is called the first time the polling process hits an election barrier. We compare polls π when elections are called and the theoretical barrier polls computed from the estimated process.

To simultaneously analyse the optimal Labour and Tory election strategies, we draw the optimal barriers as a function of the polling support π for Tories and $1-\pi$ for Labour. As seen in Figure 6 and proved in Proposition 4, these barriers fall over time, gradually at first, and then steeply near the end of the term. Since the polling process favours Labour ($b = 0{\cdot}47$), the Tory barrier is everywhere lower, that is, the Tories optimally call elections at lower support levels. The average vertical deviation between the barrier and the realized support levels is 8·8% for all governments, 6·7% without the weak Labour governments, 11·0% for Labour (6·8% without the weak governments), and 6·7% for Tory. Including the weak Labour governments, the Tory election calls have evidently been closer to the optimal policy. This might afford some insight into why the Tories have led the polls about 33% of the time from 1945 to 2005, but have ruled about 58% of the time.[27]

Using the entire path of the polling process before the elections, we can explore how the actual election times diverged from our theoretical predictions. While most elections were early,

26. The close 1979 election is different, since it was forced near the end of its term. Consistent with our model, there was no early election because the political support did not hit the election barrier. But if we did not classify the 1974–1979 government as weak, then our results would be unchanged: the same overall parameters as in Table 2 would ensue, since they are not significantly different for weak and regular governments. Further, the election timing test in Section 6.3 would not change significantly (see footnote 33).

27. With only eight data points for each party, the distances from the barriers corresponding to L and R in Figure 6 are not significantly different. Further, as will be discussed in Section 7, the ruling time difference is not statistically significant. Thus, good luck might explain the ruling time difference just as well.

just two were more than a month late: according to our model, Thatcher should have called the 1983 election 11 months earlier, immediately after her triumph in the March–June 1982 Falklands War. But it might have been deemed opportunistic to take advantage of this patriotic upsurge—a fact that our model cannot possibly account for. Likewise, Blair should have called the election of 2001 about two months earlier.

For a different test, we compare the model and realized election indicators through time. Call the election indicator 0 before the next election is called, and 1 on the day the election is called. The model indicator is 0 if the political support falls below the ruling party's election barrier, and otherwise 1. We find that the indicators for all 16 elections coincide 92·4% of the time up to their announcements. One might think that this means that ours is close to the true model. But all models that call elections late will perform well by this test, simply because the realized indicator equals 0 most of the time. To avoid this problem, we next focus only on the election announcement times (when the realized indicator is 1) and analyse how well our model explains those support levels.

The regression lines in Figure 7 illustrate our support level analysis *at the election announcement times.* The dependent variable Y is the realized support level, and the explanatory variable X is the theoretical barrier value. This analysis is motivated by cross-sectional asset pricing tests (see, for example, Cochrane, 1996). Weak governments aside,[28] we find that $Y = -0{\cdot}42 + 1{\cdot}62X$. This intercept is insignificant, and the slope is significant. Since $R^2 = 44\%$, our model explains much of the variation of election times through the regression. Put differently, the correlation of the theoretical and realized support levels at the election times is 0·66. At the very least, *we have correctly identified the polls π and elapsed time t as important decision factors* (and see the naive linear regression below).

If the elections were called solely using our model with the estimated parameters, then the regression line would coincide with the diagonal $Y = X$. Since the intercept of the choice governments is insignificant, Figure 7 also includes the best zero-intercept regression, $Y = 0{\cdot}90X$. The t-statistic on this slope is now 31·53 and $R^2 = 35\%$. This regression agrees with the message of Figure 6, that the model barriers exceed the realized support levels. The average forecast is correct if we scale the barriers down by 0·9. But we must reject the null hypothesis of a unit slope, whose t-statistic is 3·38.[29]

The slope test is obviously a joint test on the model and its parameters. To ensure a unified paper built on our stylized theory, we have consistently employed an extremely conservative econometric exercise—for instance, assuming parametric constancy over the time period 1945–2005 (see Section 7) and introducing no other explanatory factors (see Sections 6.3 and 8). Any additional degrees of freedom would surely have improved the fit.[30,31]

Further testing the result, we ask whether our model can be significantly improved by adding new variables. We thus regressed the residuals of the regressions without the weak governments on the realized election time, election year, and incumbent party. The coefficients were

28. We consider weak governments in the next subsection. As can be seen, the 1979 election is close to the regular barrier and if we include that (most of the time it was a regular government), then $R^2 = 43\%$.

29. We reject the slope test partly due to our high R^2. We would likewise reject an extremely good model with slope, say, 1·01 and $R^2 = 100\%$, for then the slope would have zero S.D.

30. By the same token, tests on the Black and Scholes model with historical volatility fail in many option markets, and so in practice the model is used with the so-called implied volatility that is estimated from option prices. By Section 4.2 and the regression $Y = 0{\cdot}90X$, in our model the corresponding implied parameters involve lower ς or greater a. In Sections 6.3 and 8, we discuss other factors that could improve the model.

31. Let us briefly return to our assumption of a constant volatility. The volatility estimates inside and outside the δ-period are $\varsigma_i = 1{\cdot}48$ and $\varsigma_o = 0{\cdot}77$. These give $R^2 = 39\%$ in the regression analysis (without the weak governments) in Figure 7; also, the average ruling periods for L and R in Section 7 are 17 and 12 years (without discounting and zero flow utility for weak governments).

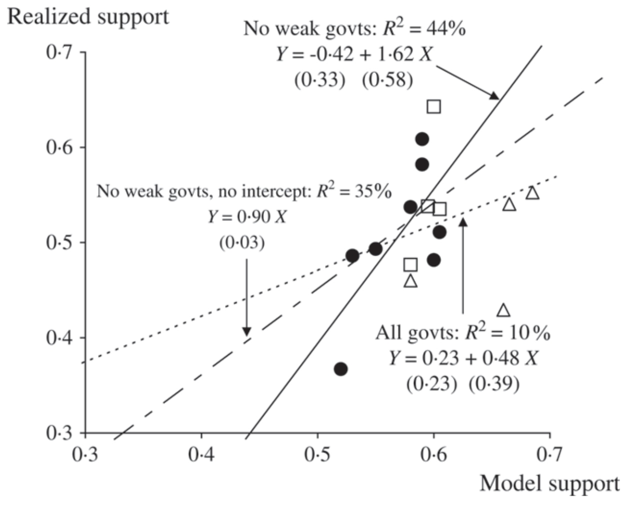

FIGURE 7

Actual and theoretical poll support at election calls. This depicts three regression lines corresponding to elections in Figure 6. Here, Y is the actual support level and X the model support level; S.D. are parenthesized. The solid line is without the weak governments, the dot–dash line is without the weak governments and without the intercept, and the dashed line is with all the governments

insignificant: all t-statistics were less than 0·9, while $R^2 = 9\%$ with the intercept, and $R^2 = 12\%$ with no intercept. Summing up, neither the party, the election year, nor the elapsed time offer any further significant predictive power.

Our contribution rests on our derivation of a rationally founded non-linear stopping barrier. But finally, *might a simpler naive model have done better?* How important is the optimization? To this end, we re-ran the regressions in Figure 7 assuming that election times can be linearly explained in (π, t)-space. This gives the regression $Y = 0{\cdot}84 - 0{\cdot}08t$ without the weak governments. While the t-statistics were 5·0 and $-1{\cdot}9$, the R^2 of this regression drops to 27% (and just 2% with all governments). We are reassured that our optimizing non-linear model is not only rationally justified, but also better explains the variation in the election times than an a-theoretical linear regression.

6.3. *Weak governments and discounting*

As seen in Figure 6, the barrier strategy does not well explain weak governments. We thus mildly modify our model. Intuitively, a minority government or a slim majority government may constantly fret that it might lose that next key vote. Still, even if it does not enjoy governing, it can aspire for a stronger government some day. We now assume that while in power, normal governments have unit flow utility and weak governments zero flow utility; further, parties discount future utility. While a weak government will not enjoy its time in power until it wins an election, its optimization is still well-defined, since the election timing choice is a pure option. In the interests of simplicity, we do not model forced elections (*i.e.* 1979), nor do we distinguish between weak and initially weak governments. The election barrier only depends on the initial status (weak or not).

Let $\Gamma^w(t, \pi)$ be the time value of a weak government. Since it earns no flow utility in the current term, $\Gamma(t, \pi) \geq \Gamma^w(t, \pi)$. As the election is called when Γ vanishes, we have

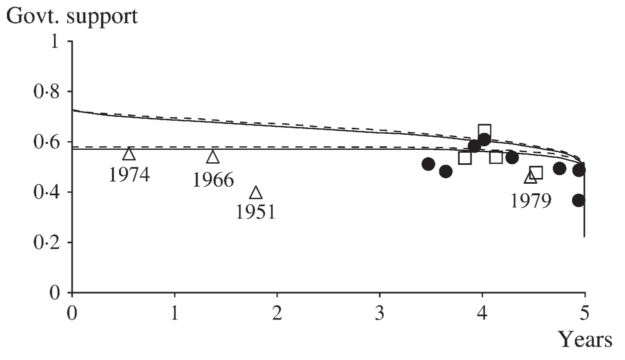

FIGURE 8

Election barriers with no flow utility for weak governments and 40% discounting. This revises Figure 6. The two upper barriers are for strong governments and the lower barriers for weak governments. Circles are support levels when the Tories called elections. Triangles are the support levels for weak Labour governments. Squares are the support levels corresponding to other Labour election times. On average, the weak governments' barriers for L and R are 0·573 and 0·568, and the regular governments' barriers for L and R are 0·652 and 0·645

Lemma 11 (Weak Timing). *Weak governments call earlier elections than regular ones.*

Next, consider the role of time discounting.[32] Intuitively, a weak government only earns pay-offs in the future, and thus its discounted value falls if it calls an election later. On the other hand, a regular government enjoys unit flow utility and discounting lessens the present value of future flow utilities. Thus, its benefit from an earlier election falls.

Lemma 12 (Discounting). *Discounting postpones regular governments' elections and advances weak governments' elections.*

In financial option markets, the Black and Scholes model is used with implied volatility, rather than historical volatility. Similarly here, we find that an annual discount rate of 40% best explains the realized elections. To make sense of such a high rate, recall that the Prime Minister's term in power may be shorter than his party's, and so he may less heavily weight future benefits of power. For instance, Tony Blair stepped down before his term is over, just as did Churchill (1955), Wilson (1976), and Thatcher (1990).

We do not re-derive all of our results for this revised model, but proceed directly to the empirical analysis. As can be seen in Figure 8, the weak governments have called elections near their revised optimal barriers. By Figures 6 and 8, these changes for weak governments have little effect on strong governments' stopping barriers. The average vertical deviation between the barriers and election support levels is 6·9% for Labour and 7·7% for Tory. Thus, the realized

32. In other words, one day of power at time t is now worth e^{-rt} days, where $r > 0$ is the discount rate. If we define $\mathbf{1}_w = 0$ for weak governments, and otherwise 1, then the value equations (2) and (3) become

$$\Omega(p) = \left(\int_0^\delta e^{-rs} ds\right)\{\mathbf{1}_w\} + \left\{e^{-r\delta} E\left[F(p(\tau+\delta), 0) \mid p(\tau) = p\right]\right\}$$

$$F(p,t) = \sup_{t \le \tau \le T} E_{p(t)=p}\left[\left(\int_t^\tau e^{-rs} ds\right)\mathbf{1}_w + e^{-r(\tau-t)}\Omega(p(\tau))\right].$$

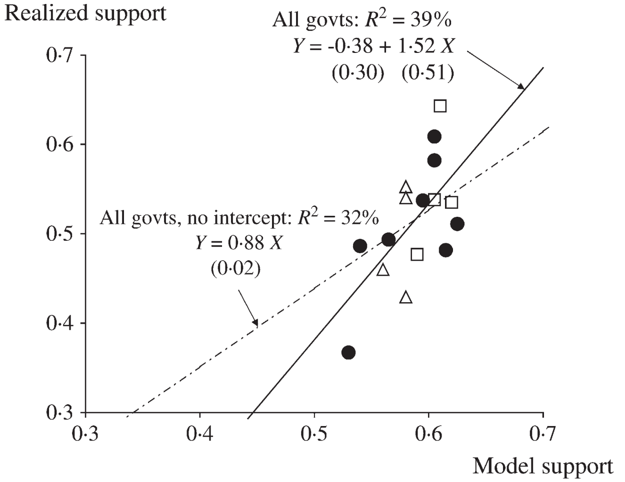

FIGURE 9

Actual and theoretical poll support at election calls with no flow utility for weak governments and 40% discounting. This depicts two regression lines for Figure 8, simultaneously incorporating weak and regular governments. Here, Y is the actual support level and X the model support level; the parameter S.D. are parenthesized. The solid line is with the intercept and the dot–dash line without the intercept

Labour strategy is closer to the policy that maximizes the expected discounted time in power, given the weak Labour governments.

Figure 9 shows the support level analysis at the election announcement times. For all the governments we now have $Y = -0{\cdot}38 + 1{\cdot}58X$, where Y is the realized political support level when the election is announced, and X is the corresponding model support. As in Section 6.2, the intercept is insignificant and the slope is significant. We now have $R^2 = 39\%$ for the revised model with all governments, far above the earlier level of 10%.[33] Once again, the polls π and elapsed time t are clearly important decision factors for understanding election timing, even when we admit a distinction between weak and regular governments. With a zero intercept, we have $Y = 0{\cdot}88X$, almost the same as the corresponding regression in Section 6.2. And we still reject a unit slope, given the high R^2. Since we use an implied discount rate, this section is not an out-of-sample test, but is rather the best fit parsimonious model. Without discounting we have $R^2 = 14\%$ for all the governments, which is only slightly better than in Figure 7.

7. THE OPTION VALUE OF ELECTION TIMING

The option to freely time an election obviously raises a party's expected time in power. We now measure the value of this option, by comparing the expected ruling times $F^L(\pi, 0)$ and $F^R(\pi, 0)$ with their analogues for fixed electoral timing. We take this expectation using the long-run density of the polls π from Lemma 3 with the estimated parameters in Section 5.3.[34] The predicted times are too high by historical standards—about 65·8 years for Labour and 24·4 for Tory. While parties have diverged from our basic optimal exercise rule (Tory less so, see Figure 6), the resulting ruling times have been much shorter. Labour governments have averaged 6·3 years in our

33. We did not call John Major's Tory 1992–1997 government weak since its initial seat proportion was 0·516. Otherwise, the R^2 in Figure 9 would *rise* to 42% from 39%. Further, if we considered Callaghan's 1974–1979 Labour government (whose election was forced) as regular, then the R^2 would fall to 35%.

34. The density is almost moot: the expected ruling times $F^L(\pi, 0)$, $F^R(\pi, 0)$ are nearly constant in π.

sample, and Tory 11·6 years.[35] We next offer three explanations for this shortfall: (i) weak governments calling early elections, (ii) precluding "opportunistic elections", and (iii) non-constant parameters.

First, we consider the model of Section 6.3 where weak governments do not enjoy their time in power, and the future is discounted by 40%. Table 3 reports the resulting expected times in power with the revised election barriers: 57·7 years for Labour and 22·4 years for Tory. Evidently, this alone cannot account for the divergent expected times in power.

Second, we assume that elections cannot be called within the first three years of a term since governments fear punishment for opportunism.[36] We thus reformulate our timing exercise in the numerical optimization, asking that elections be called in years 3–5; this eliminates repeatedly calling an election when riding high in the polls and would lessen the expected time in power to 36 years for Labour and 17·1 years for Tory (see Table 3).[37]

Third, we consider two parametric regime shifts. Polls π averaged 0·49 from 1943 to 1992, but just 0·37 in 1992–2005 (see Figure 5), a difference which is significant (t-statistic is 32·48). If we simply focus on the pre-1992 time span with the higher mean reversion level $b = 0{\cdot}49$,[38] then Labour's average win chance would fall from 70% to 65%. Accordingly, its expected time in power would fall to 51·9 years, while Tory's would rise to 34·3 years.

We consider an additional form of parametric non-constancy—distinct parameters inside and outside the δ-period.[39] More specifically, we employ the δ-period and pre-δ parameters in Table 1 and their polling frequencies. As can be seen from Table 3, the distinct parameters would dramatically cut the expected times in power to 15·9 years for Labour and 12·4 years for Tory. These expected times are much closer because the campaign periods have favoured Tory—the polls have averaged 0·52 (see Table 1).[40]

Since 1945, the Tories have led the polls about 33% of the time but have ruled about 58% of the time. We see that this might be mostly explained by a campaign-period edge that Tory enjoys. But we have that this cannot be the whole story. It might just owe to dumb luck (see footnote 35), since Labour has been so often beset by weak governments.

Motivated by Table 3, we focus on our best explanation above, and assume distinct pre-δ and δ-period polling process parameters in Table 1. If the U.K. implemented a fixed electoral cycle with four-year terms, then the expected duration in power given an optimal policy would fall by a factor of 1·7—from 15·9 to nine years for Labour, and 12·4 to 7·5 years for Tory. Labour's expected percentage time in power would drop from 56·2% to 54·5%. *An overarching observation here is that flexible electoral timing favours the dominant party more than does fixed election cycles.*

We conclude with a simple welfare analysis: Does endogenous timing on average help or hurt the voters? Barring considerations of weak governments, we can intuitively conclude it

35. Between 1945 and 2005, there have been only seven ruling periods, and so the S.D. are high: 1·2 for Labour and 7·3 for Tory. In other words, these differences are not significant.

36. We ignore the possibility of a government falling due to a vote of non-confidence. We use three years since it seems focal, and because the shortest regular government ended in 1955, just shy of the 3·5-year mark. For instance, Blais, Gidengil, Nevitte and Nadeau (2004) argue that voters punished Jean Chretien for calling a snap election in November 2000 after just three years and four months.

37. This does not greatly move our stopping barriers, and the resulting regression in Figure 7 for past elections without the weak governments is only a slightly worse fit, with $R^2 = 40\%$.

38. In fact, any break point in 1970–1995 produces a significant difference in the poll mean b. Additionally, $a = 2{\cdot}2, \sigma = 0{\cdot}20, \sigma_\eta = 0{\cdot}11$, and thus constant volatility is $\varsigma = 0{\cdot}69$. That is, the π process mean reverts faster about a higher mean, with less volatility.

39. By Section 5, these parameters are not statistically significantly different. This, of course, does not at all suggest that they must coincide, which cannot be proved.

40. Had we used these parameter values also in the support level analysis in Section 6.2, we would have $R^2 = 36\%$ for the regular governments.

TABLE 3

Expected time in power

Flexible timing 0–5 years	L rules 87·6% of time	L rules 65·8 years	R rules 24·4 years
Flexible timing 0–5 years *	L rules 74·4% of time	L rules 51·9 years	R rules 34·3 years
Flexible timing 0–5 years †	L rules 56·2% of time	L rules 15·9 years	R rules 12·4 years
flexible timing 0–5 years ‡	L rules 72·0% of time	L rules 57·7 years	R rules 22·4 years
Flexible timing 3–5 years	L rules 84·6% of time	L rules 36·0 years	R rules 17·1 years
Flexible timing 3–5 years*	L rules 72·5% of time	L rules 31·9 years	R rules 23·2 years
Flexible timing 3–5 years †	L rules 56·0% of time	L rules 15·0 years	R rules 11·8 years
Flexible timing 3–5 years ‡	L rules 68·4% of time	L rules 31·6 years	R rules 14·6 years
Elections every 5 years	L rules 79·2% of time	L rules 12·5 years	R rules 8·6 years
Elections every 4 years	L rules 79·2% of time	L rules 10·1 years	R rules 6·9 years
Elections every 4 years*	L rules 69·1% of time	L rules 8·9 years	R rules 7·6 years
Elections every 4 years †	L rules 54·5% of time	L rules 9·0 years	R rules 7·5 years
Elections every 4 years ‡	L rules 58·9% of time	L rules 9·6 years	R rules 6·7 years

Notes: Under different regime assumptions, we compute (a) the long-run fraction of time in power, and (b) the expected time in power immediately after winning an election.
*Rows use the pre-1992 polling process.
†Rows are with different parameters inside/outside the δ period.
‡ Rows are with weak governments and 40% discounting.

hurts. If the election is called at the four-year mark, when it would have been forced, then voters are unaffected. If the government is standing high in the polls earlier, then it calls an election. This choice is welfare neutral for the voters, since the best party is already in power. Finally, if a government has low support at the four-year mark, then it delays the election. But in this case, it is most likely that the wrong party is in power, and delay hurts voters. Thus, flexible electoral timing hurts voters.

8. CONCLUSION

8.1. *Summary*

Optimal timing of votes and elections is an important subject and periodically the topic of great media speculation in some countries. In this paper, we have modelled electoral timing as an optimization by the ruling party. We have developed and analysed a tractable electoral timing model capturing the informational richness of the political setting: namely, a forward-looking optimizing exercise using an informationally derived mean-reverting polling process. The election is called the first moment the polling process hits a non-linear stopping barrier. We think that this is a substantively novel timing exercise for economics (see Dixit and Pindyck, 1994). The ruling party holds a renewable finite time horizon American option and the exercise decision is delayed.

We then fit the polling process to the post-war Labour–Tory rivalry of the U.K. We found a high correlation between the realized political support levels and the model support levels at the election call dates. The weak governments aside, parties in power do indeed try to maximize their expected time in power, and election times are triggered by the polls and the time from the last election. We show that the value of the option to choose the election time can be very substantial, and favours the dominant party. Further, weak governments are explained by introducing zero flow utility and a high discount rate.

While this paper was written, Gordon Brown became Prime Minister and the U.K. press actively speculated on an early election. Our model predicts that if the Prime Minister follows the dynamically optimal strategy, then there is a high chance for an early election, most likely in 2008. By using the poll results on 11 August, 2007 (ipsos-mori.com), the parameter estimates from Section 5, and a Monte Carlo Simulation with 100,000 random outcomes, we find that the respective chances of early election in 2007, 2008, 2009, and 2010 are 0·34, 0·40, 0·17, and 0·09. By contrast, the respective market odds[41] on 20 August, 2007 were 7/2, 7/4, 6/4, and 8/1, so that the market's best guess is 2009.

8.2. *Some caveats*

As usual, our tractability owes to some simplifying assumptions.[42] Our objective function is straightforward, positing that governments maximize their expected (possibly discounted) time in power. We have rendered this a decision-theoretic exercise, assuming that the government cannot affect the polling process. These simplifications are not irrelevant, but have been best studied elsewhere (*e.g.* Austen-Smith and Banks, 1988, deal with the richer picture). Our single-minded theory explains much of the variation in election timing decisions with just the polls and time since the last election.

The U.K. employs the standard "first-past-the-post" electoral system. There are now 646 seats in the House of Commons, so that a party must capture 324 for an overall majority. But our theory assumes that when calling an election, the government *acts as if* it must win the popular vote. This almost holds in our data set. In October 1951, the Tories formed the government but lost the popular vote by 0·8%. In February 1974, the reverse occurred: Labour formed the government, but trailed the popular vote by 0·8%. The errors above are small and of opposite parity, and so this is not inconsistent with our assumption. Dealing with this more formally would require a treatment of the seat proportions. As seen in Table 2, the winning big party has consistently had the greatest seat proportion. But the deviation of the seat and vote proportions is positively correlated with the vote proportion, and so lies outside our model.[43]

We have only employed the public voting intention polls. In fact, the government surely has more accurate information, possibly from private polls, etc. This would raise the polling sample size and lower the polling volatility and so the election barriers. The government would then require a filtering exercise, estimating the political support at each moment in time. Also, testing our barriers with an unobserved state would be very hard. We have avoided this non-trivial and unprofitable exercise, but have verified that our barriers are quite robust to the variance specification. But the normative predictions of the model—the expected durations in power—are sensitive to the variance specification.

These limitations of our theory notwithstanding, we capture the central element of this crucial timing decision of a parliamentary democracy. Attesting to this, our empirical analysis explains a significant proportion of the variation in the election timing decisions.

APPENDIX A. OMITTED PROOFS

A.1. *Proof of Lemma* 2*: variance of the political slant process*

Let us rewrite (1): $dp(t)+ap(t)dt = abdt+\sigma p(t)(1-p(t))dW(t)$, which gives

$$e^{at}dp(t)+ae^{at}p(t)dt = abe^{at}dt+e^{at}\sigma p(t)(1-p(t))dW(t).$$

41. See http://odds.betrescue.com. These are gambling odds and do not directly represent the true chances that the event will occur since they include a profit margin (see http://en.wikipedia.org/wiki/Odds). If we assume a constant profit margin then the odds' implied probabilities are 0·17, 0·35, 0·40, and 0·08.

42. We have also ignored any strategic incentives to vote, but these are surely quite minuscule in a national election (see, for example, Feddersen and Pesendorfer, 1996). As noted, we also assume that voters simply myopically vote for the best current party, and do not anticipate the scandals or laurels to come. Ours is a theory of strategizing and forward-looking behaviour by the government, and not voters.

43. For the elections in our sample, an additional 1% in the polls raises the seat count just over 2%.

Note that the L.H.S. equals $d(e^{at}p(t))$ and therefore, by integration,

$$p(t) = e^{-at}p + (1-e^{-at})b + \sigma \int_0^t e^{-a(t-s)} p(s)(1-p(s))dW(s), \tag{8}$$

where $E[p(t)|p(0)=p] = e^{-at}p + (1-e^{-at})b$, which we denote by $m(p,t)$.

By (8), the variance of $p(t)$

$$\upsilon(t) = E[(p(t)-m(p,t))^2] = \sigma^2 \int_0^t e^{-2a(t-s)} E[p^2(s)(1-p(s))^2]ds \tag{9}$$

and using $p(t) = m(t) + \sqrt{\upsilon(t)}\varepsilon$, where ε is a standard normal variable, gives

$$\upsilon(t) = \sigma^2 \int_0^t e^{-2a(t-s)} \{m^2(s)(1-m(s))^2 + [1-6m(s)(1-m(s))]\upsilon(s) + 3(\upsilon(s))^2\} ds$$

suppressing the p argument. Thus, we have the partial derivatives

$$\upsilon_t(t) = \sigma^2 E[p^2(t)(1-p(t))^2] - 2a\upsilon(t),$$

$$\upsilon_{\sigma^2}(t) = \upsilon(t)/\sigma^2 + \sigma^2 \int_0^t e^{-2a(t-s)} [1 + 6(\upsilon(s) - m(s)(1-m(s)))]\upsilon_{\sigma^2}(s)ds,$$

$$\begin{aligned}\upsilon_{\sigma^2 t}(t) &= \sigma^{-2}[\upsilon_t(t) + 2a\upsilon(t)] + \sigma^2[1 + 6\{\upsilon(t) - m(t)(1-m(t))\} - 2a]\upsilon_{\sigma^2}(t)\\ &= E[p^2(t)(1-p(t))^2] + \sigma^2[1 + 6\{\upsilon(t) - m(t)(1-m(t))\} - 2a]\upsilon_{\sigma^2}(t).\end{aligned}$$

We find $\upsilon_t(0) > 0$ and $\upsilon_{\sigma^2}(0) > 0$ for all $\sigma^2 > 0$. Now $\upsilon_{\sigma^2}(t) = \upsilon_{\sigma^2}(0) + \int_0^t \upsilon_{\sigma^2 t}(y)dy$. If $\upsilon_{\sigma^2}(t) = 0$ for some $t > 0$, then $\upsilon_{\sigma^2 t}(t) > 0$. Thus, $\upsilon_{\sigma^2}(t+\varepsilon) > 0$, where $\varepsilon > 0$ is small, and we get that the variance of $p(t)$ rises in the diffusion coefficient σ. ‖

A.2. *Proof of Lemma* 3*: derivation of the stationary density*

We appeal to Karlin and Taylor (1981, pp. 220 and 241). If $dp(t) = \tilde{\mu}(p)dp + \sigma(p)dW$ has a stationary density $\psi(y) = \lim_{t\to\infty}(\partial/\partial y)P(p(t) \leq y \mid p(0) = x)$, then it obeys the stationary forward Fokker–Planck equation $\frac{1}{2}[\sigma(p)\psi(p)]'' - [\tilde{\mu}(p)\psi(p)]' = 0$. In particular, for (1), we have $\frac{1}{2}[(\sigma p(1-p))^2\psi(p)]'' - [a(b-p)\psi(p)]' = 0$. Its solution is given by $\psi(p) = m(p)[C_1 S(p) + C_2]$, where $m(p) = 1/(\sigma^2 p^2(1-p)^2 s(p))$ is the speed measure, and $S(p) = \int_{p_0}^p s(y)dy$ is the scale function, whose density equals

$$s(p) = e^{-\int_{p_0}^p \frac{2\tilde{\mu}(y)}{\sigma(y)^2}dy} = e^{-\int_{p_0}^p \frac{2a(b-y)}{\sigma^2 y^2(1-y)^2}dy} = e^{\frac{2a}{\sigma^2}\left(\frac{1-b}{1-p} + \frac{b}{p}\right)} \left(\frac{p}{1-p}\right)^{2a(1-2b)/\sigma^2} - C_0,$$

where $p_0 \in (0,1)$ is arbitrary, and C_0, C_1, and C_2 are constants.

Claim 1 (Entrance boundary). The extremes 0 and 1 are entrance boundaries, *i.e.*, they cannot be reached from $(0,1)$ but the process can begin from the boundaries.

Proof. We consider the left boundary; the right is analysed using $q(t) = 1 - p(t)$ and noting $dq(t) = a((1-b) - q(t))dt - \sigma q(t)(1-q(t))dW(t)$ and $p(t) = 1$ iff $q(t) = 0$.

The sufficient conditions that 0 be an entrance (see Karlin and Taylor, 1981, pp. 226–242) are $\lim_{y\downarrow 0}\int_y^p s(z)dz = \infty$ and $\lim_{y\downarrow 0}\int_y^p m(z)dz < \infty$, where $p \in (0,1)$. The first condition holds since $\int_0^p s(z)dz \geq \int_0^p \exp(c_0 + c_1/z)z^{c_2}dz = \infty$ for all $p \in (0,1)$, where c_0, c_1, and c_2 are positive constants. Likewise, we get the second condition. ‖

Note that $S(p)$ is monotonic. Claim 1 gives $S(0) = -\infty$ and $S(1) = \infty$. Therefore, for $\psi(p) > 0$ throughout $(0,1)$ we must have $C_1 = 0$. The constant C_2 is selected to ensure that $\int_0^1 \psi(p)dp = 1$ and, thus, the stationary density $\psi(p) = m(p)/\int_0^1 m(z)dz$. ‖

A.3. *Proof of Proposition* 2*: existence of smooth values*

If the government stands at 100% on the day election is announced, it loses the election with a positive chance, say at least $\ell > 0$. This yields an upper bound $\Omega^i < \delta + (T+\delta)/\ell$.

EXISTENCE. Substitute $F_0^i = 0$ into (2) to compute Ω_0^i. Insert Ω_0^i into (3) to compute F_1^i. Since (2) and (3) define monotone maps $F^i \mapsto \Omega^i$ and $\Omega^i \mapsto F^i$, the iterations obey $0 \le \Omega_0^i \le \Omega_1^i \le \Omega_2^i \le \ldots$ and $0 \le F_0^i \le F_1^i \le F_2^i \le \cdots$. Their limits exist, and obey (2) and (3).

Ω IS SMOOTH. Write $\Omega^R(p) = \int_{1/2}^1 \psi(p,q,\delta)F^R(q,0)dq$, for the smooth transition density ψ in p. As F^R is bounded, Ω^R is smooth in p. Similarly with Ω^L.

F IS SMOOTH. Denote the initial time by t_0, and use t as a free time variable. Any election timing in (p,t)-space defines an open "continuation set" $U \subseteq (0,1) \times (0,T)$, such that the election is called at $\tau_U = \inf\{t \ge t_0 | (p(t),t) \in \partial U\}$. There exists a smooth function $f(p,t)$ which solves the PDE $\mathcal{A} f(p,t) = -1$ in U with $f(p,\tau_U) = \Omega(p)$ on ∂U, where $\mathcal{A} f(p,t)$ is given by (4) on U.[44] Ito's Lemma therefore applies:

$$\Omega(p(\tau_U)) = f(p(t_0),t_0) + \int_{t_0}^{\tau_U} \mathcal{A} f(p(t),t)dt + \int_{t_0}^{\tau_U} f_p(p(t),t)\sigma p(t)(1-p(t))dW(t).$$

Thus, $E[\tau_U - t_0 + \Omega(p(\tau_U))] = f(p(t_0),t_0)$. Finally, we restrict to the optimal continuation set U, namely, those (p,t_0) where continuing at time t_0 with $p = p(t_0)$ is optimal. Since it is also true that $E[\tau_U - t_0 + \Omega(p(\tau_U))] = F(p,t_0)$, we conclude that $F(p,t_0) = f(p,t_0)$. In other words, F inherits the smoothness properties of f. ‖

A.4. *Proof of Lemma* 5*: convex waiting values*

We want the map $p \mapsto F(p,t)$ to be convex for fixed t. Let P be the random variable formed by updating the prior p using an additional current (time-t) binary signal I sending p to P_1 with chance λ and P_2 with chance $1-\lambda$. Then $p = E_I[P] = \lambda P_1 + (1-\lambda)P_2$ by the Law of Iterated Expectations. Since the information set $\{I,\xi\}$ allows a weakly better decision than the Gaussian public information process ξ alone:

$$E_I\left(\sup_{t\le\tau\le T} E_{p(t)=P}[\tau - t + \Omega(p(\tau))]\right) \ge \sup_{t\le\tau\le T} E_{p(t)=E_I(P)}[\tau - t + \Omega(p(\tau))].$$

Therefore, $\lambda F(P_1,t) + (1-\lambda)F(P_2,t) \ge F(\lambda P_1 + (1-\lambda)P_2,t)$. So F is convex in p. ‖

A.5. *Proof of Lemma* 6 (*b*)*: value monotonicity*

Fix party R. Let us start at slant $p'' > p'$, but call the election when you would have called it with a slant p'. Then the political slant path that starts at p'' dominates the corresponding path for p'. As this is true on every path, the stopping value for p'' exceeds the stopping value for p'. But if we re-optimize the stopping policy for p'', we do even better. So the optimal stopping and continuation values are ranked $\Omega^R(p'') > \Omega^R(p')$ and $F^R(p'',t) > F^R(p',t)$. ‖

A.6. *Proof of Lemma* 7 (*b*)*: Smooth pasting*

Let us fix party R. We show the differentiability of the solution to Proposition 3.

Step A1. RIGHT DERIVATIVE. By Proposition 3, $F(p,t) = \Omega(p)$ for all $p \ge \beta(t)$ and, hence, the right derivatives agree: $F_p(\beta(t)+,t) = \Omega_p(\beta(t))$.

Step A2. LEFT DERIVATIVE LOWER BOUND. Since $F(p,t) > \Omega(p)$ for $p < \beta(t)$, and the left derivative $F_p(\beta(t)-,t)$ exists by Proposition 2, and likewise $\Omega_p(\beta(t))$ exists, we have $F_p(\beta(t)-,t) \le \Omega_p(\beta(t))$. Thus, we must show $F_p(\beta(t)-,t) \ge \Omega_p(\beta(t))$.

44. For instance, by Theorem 4 in Section 6.2.3 and Theorem 3 in Section 6.3.1 of Evans (1998), under our conditions, the PDE $Lu(x) = f$ in U and $u = 0$ on ∂U has a solution, given the operator $Lu = -\sum a_{ij}(x)u_{ij} + \sum b_i(x)u_i + c(x)u$. In our case, we set $x = (p,t)$ and $u(x) = F(p,t) - \Omega(p)$ for suitable PDE coefficients.

Step A3. A DIFFERENCE CALCULATION. Let $p(t|\bar{p})$ be the slant at time t that equals $\bar{p}$ at time $\bar{t} < t$. Let $\zeta(t) = p(t|\bar{p}) - p(t \mid \bar{p} - \varepsilon)$ and

$$I(t) = \sigma \int_{\bar{t}}^{t} e^{-a(t-y)} \left[p(y \mid \bar{p})(1 - p(y \mid \bar{p})) - p(s \mid \bar{p} - \varepsilon)(1 - p(s \mid \bar{p} - \varepsilon)) \right] dW(s).$$

Then (1) yields $\zeta(t) \equiv e^{-a(t-\bar{t})}\varepsilon + I(t)$. Write $I(t) \sim \varepsilon d(t)e$, where e is a standard $N(0,1)$ variable. Since $I(0) = 0$ and $I(t)$ is continuous in t, we have $\lim_{t \downarrow \bar{t}} I(t) = 0$ for all paths, and thus $\lim_{t \downarrow \bar{t}} d(t) = 0$.

Step A4. LEFT DERIVATIVE UPPER BOUND. By comparing the political slants path by path, the election time $\tau_{\bar{p}} = \inf\{t \geq \bar{t} \mid p(t \mid \bar{p}) \geq \beta(t)\}$ is a decreasing function of $\bar{p}$, and $\lim_{\bar{p} \uparrow \beta(\bar{t})} \tau_{\bar{p}} = \bar{t}$. If we start at $\bar{p} = \beta(\bar{t})$, then[45]

$$\begin{aligned} F(\bar{p} - \varepsilon, \bar{t}) &= E_{\bar{t}}[\Omega(p(\tau_{\bar{p}-\varepsilon} \mid \bar{p} - \varepsilon))] \\ &= E_{\bar{t}}[\Omega(p(\tau_{\bar{p}-\varepsilon} \mid \bar{p}))] - E_{\bar{t}}[\Omega(p(\tau_{\bar{p}-\varepsilon} \mid \bar{p})) - \Omega(p(\tau_{\bar{p}-\varepsilon} \mid \bar{p} - \varepsilon))] \\ &\leq F(\bar{p}, \bar{t}) - E_{\bar{t}}[\Omega(p(\tau_{\bar{p}-\varepsilon} \mid \bar{p})) - \Omega(p(\tau_{\bar{p}-\varepsilon} \mid \bar{p} - \varepsilon))]. \end{aligned}$$

The inequality followed from the optimality of the election time—since $\tau_{\bar{p}-\varepsilon}$ need not be optimal starting at slant $\bar{p}$ at time $\bar{t}$, we must have $E_{\bar{t}}[\Omega(p(\tau_{\bar{p}-\varepsilon}|\bar{p}))] \leq F(\bar{p}, \bar{t})$. So

$$\begin{aligned} F_p(\bar{p}-, \bar{t}) &= \lim_{\varepsilon \downarrow 0} \tfrac{1}{\varepsilon}[F(\bar{p}, \bar{t}) - F(\bar{p} - \varepsilon, \bar{t})] \\ &= \lim_{\varepsilon \downarrow 0} E_{\bar{t}}[\Omega(p(\tau_{\bar{p}-\varepsilon} \mid \bar{p})) - \Omega(p(\tau_{\bar{p}-\varepsilon} \mid \bar{p} - \varepsilon))] \\ &\geq \lim_{\varepsilon \downarrow 0} E_{\bar{t}}[\Omega'(p(\tau_{\bar{p}-\varepsilon} \mid \bar{p} - \varepsilon))(e^{-a(\tau_{\bar{p}-\varepsilon} - \bar{t})} + d(\tau_{\bar{p}-\varepsilon})e) + O(\zeta(\tau_{\bar{p}-\varepsilon}))^2/\varepsilon], \end{aligned}$$

using the Taylor expansion $\Omega(p(t \mid \bar{p})) - \Omega(p(t \mid \bar{p} - \varepsilon)) = \Omega'(p(t \mid \bar{p} - \varepsilon))\zeta(t) + O(\zeta(t))^2$. Since $\lim_{\varepsilon \downarrow 0} \tau_{\bar{p}-\varepsilon} = 0$, and $O(\zeta(\tau_{\bar{p}-\varepsilon}))$ is of order ε^2, we now get $F_p(\bar{p}-, \bar{t}) \geq \Omega_p(\bar{p})$. ‖

A.7. *Proof of Proposition* 4*: optimal election barriers*

Fix party R. Our argument is inspired by that in Jacka (1991) for American put options.

Step A1. MONOTONICITY. We have $F(p,t) - \Omega(p) \geq 0$, with equality along the barrier. Since $F(p,t)$ is decreasing in t by Lemma 6, we have for any $\Delta > 0$ and $\varepsilon > 0$:

$$F(\beta(t) - \varepsilon, t - \Delta) \geq F(\beta(t) - \varepsilon, t) > \Omega(\beta(t) - \varepsilon).$$

Therefore, $\beta(t - \Delta) > \beta(t) - \varepsilon$ for all $\varepsilon > 0$. Hence, $\beta(\cdot)$ is non-increasing.

Step A2. RIGHT CONTINUITY. By definition of $p = \beta(t)$ as the least solution to $F(p,t) = \Omega(p)$, the electoral continuation region assumes the form $U = \{(p,t) \in (0,1) \times (0,T) | p(t) < \beta(t)\}$, which is open. Since $(\beta(t_i), t_i) \notin U$ for any convergent sequence $t_i \to \hat{t}$, we have $(\hat{\beta}, \hat{t}) \notin U$, where $\hat{\beta} = \limsup_{i \to \infty} \beta(t_i)$. This gives a lower semi-continuous barrier, as $\hat{\beta} \geq \beta(\hat{t})$. Since it is also non-increasing, it is right continuous.

Step A3. LEFT CONTINUITY. Fix $0 < t < T$. Since β is monotone, the left limit $\beta(t-)$ exists. Assume, for a contradiction, that $\beta(t-) > \beta(t)$. Now, pick $0 < \varepsilon < (\beta(t-) - \beta(t))/2$. Then $F(\beta(t_i) - \varepsilon, t_i) > \Omega(\beta(t_i) - \varepsilon)$ for all increasing sequences $t_i \uparrow t$, and so by continuity, in the limit we have $F(\beta(t-) - \varepsilon, t) \geq \Omega(\beta(t-) - \varepsilon)$. Since $(\beta(t-) - 2\varepsilon, t) \in U$ by the structure of U, we have $F(\beta(t-) - 2\varepsilon, t) > \Omega(\beta(t-) - 2\varepsilon)$. In other words, $\beta(t-) - 2\varepsilon < \beta(t)$, contrary to our choice of ε. Thus, $\beta(t-) = \beta(t)$. ‖

A.8. *Proof of Lemma* 8*: the shape of the victory chance*

We first consider party R. Let $p(0) = p > \frac{1}{2}e^{a\delta} + (1 - e^{a\delta})b$. Then the expected election day slant is $m(p,t) = e^{-at}p + (1 - e^{-at})b > \frac{1}{2}$ by Lemma 2. A win is thus expected.

45. The argument is motivated by the proof of Lemma 7.8 in Section 2.7 in Karatzas and Shreve (1998).

Step A1. THE VICTORY CHANCE FALLS IN σ. The logic is like a converse to Jensen's inequality: greater variance only hurts an expectation if the function is concave. For since $p(\delta)$ is a random variable with mean $m(p,\delta)$, losing is a "tail event" that $p(\delta) < 1/2$. So $V(p)$ falls in σ for all $p > \frac{1}{2}e^{a\delta} + (1-e^{a\delta})b$, since greater σ weights the Wiener realization more heavily, pushing the political slant paths away from the mean.

More generally, let $\sigma(s)$ be a step function, equal to σ_i on the interval $J_i \equiv [a_i, b_i]$, where $J_1, \ldots, J_n$ partition $[0,\delta]$. The variance of $p(\delta)$ increases in all σ_i, since $p(\delta)$ is the σ-weighted sum of independent 0-mean random variables. Indeed, by equation (8),

$$\begin{aligned} p(\delta) &= m(p,\delta) + \int_0^{\delta} \sigma(s) e^{-a(\delta-s)} p(s)(1-p(s)) dW(s) \\ &= m(p,\delta) + \sum_{i=1}^{n} \sigma_i \int_{J_i} e^{-a(\delta-s)} p(s)(1-p(s)) dW(s). \end{aligned}$$

Party R's losing chance is the weight in the lower tail of $p(\delta)$, and thus rises in each σ_i.

Step A2. THE POLITICAL SLANT MOMENTS IN σ. Raise σ by $\varepsilon > 0$ on the time interval $[0,\Delta)$ and denote the resulting political slant by p_ε. Then from (8), we get

$$p_\varepsilon(\Delta) - p(\Delta) = \int_0^{\Delta} e^{-a(\Delta-t)}[(\varepsilon+\sigma)p_\varepsilon(t)(1-p_\varepsilon(t)) - \sigma p(t)(1-p(t))] dW(t).$$

The variance is $E_p[(p_\varepsilon(\Delta)-p(\Delta))^2] = \varepsilon^2 p^2 (1-p)^2 \Delta + o(\Delta)$ while the higher moments $n > 2$ obey $E_p[(p_\varepsilon(\Delta) - p(\Delta))^n] = o(\Delta)$ for the error term $\lim_{\Delta\downarrow 0} o(\Delta)/\Delta = 0$.

Step A3. LOCAL CONCAVITY. Define victory chances $\upsilon_\varepsilon(p,\delta+\Delta) = E_p[V(p_\varepsilon(\Delta))]$ and $\upsilon(p,\delta+\Delta) = E_p[V(p(\Delta))]$ for a hypothetical $\delta+\Delta$ election delay. Then

$$\begin{aligned} 0 &> [\upsilon_\varepsilon(p,\delta+\Delta) - \upsilon(p,\delta+\Delta)]/\Delta \\ &= E_p[V(p_\varepsilon(\Delta)) - V(p(\Delta))]/\Delta \\ &= \tfrac{1}{2} E_p[V''(p(\Delta))(p_\varepsilon(\Delta)-p(\Delta))^2]/\Delta + o(1), \end{aligned}$$

by Steps 1 and 2. Since $V''(p)$ is continuous by Proposition 2, the limit as $\Delta \downarrow 0$ yields

$$0 \geq \left.\frac{\partial \upsilon_\varepsilon(p,t)}{\partial t}\right|_{t=\delta} - \left.\frac{\partial \upsilon(p,t)}{\partial t}\right|_{t=\delta} = \tfrac{1}{2} V''(p)\varepsilon^2 p^2 (1-p)^2,$$

which gives $V''(p) \leq 0$ for all $p > \frac{1}{2}e^{a\delta} + (1-e^{a\delta})b$, that is, the local concavity of $V(p)$.

Likewise, $V^L(p)$ is concave for $p < \frac{1}{2}e^{a\delta} + (1-e^{a\delta})b$. Since $V^R(p) \equiv 1 - V^L(p)$, $V^R(p)$ is convex for $p < \frac{1}{2}e^{a\delta} + (1-e^{a\delta})b$, and $V^L(p)$ for $p > \frac{1}{2}e^{a\delta} + (1-e^{a\delta})b$. ‖

A.9. *Proof of Lemma* 10*: the polling process*

Step A1. THE POLLING ERROR. Denote $\pi_j \equiv \pi(t_j)$. Given footnote 12, the polling error η_j is approximately $e_{\eta_j}\sqrt{\pi_j(1-\pi_j)/N}$, where $\{e_{\eta_j}\}$ are i.i.d. standard normal r.v.'s independent of the $\{W(t)\}$. The approximation error vanishes as $N \to \infty$, by the Central Limit Theorem. Since the differences $\eta_{j+1} - \eta_j$ are Gaussian, and independent of $\{W(t_{j+1}) - W(t_j)\}$, we have $\eta_{j+1} - \eta_j = \varphi_j \sqrt{\Delta_j}\xi_j$, for some $\varphi_j > 0$, where $\{\xi_j\}$ are independent standard normal r.v.'s. Solving for $\text{Var}[(\eta_{j+1}-\eta_j) \mid \pi_j] = \varphi_j^2 \Delta_j$, we see

$$\varphi_j^2 \Delta_j = E[\eta_j^2 + \eta_{j+1}^2 \mid \pi_j] \approx \{\pi_j(1-\pi_j) + E[\pi_{j+1}(1-\pi_{j+1}) \mid \pi_j]\}/N \equiv \pi_j(1-\pi_j)k_j/N,$$

where $k_j = 1 + E[\pi_{j+1}(1-\pi_{j+1}) \mid \pi_j]/\pi_j(1-\pi_j)$.

Step A2. THE POLLING PROCESS. From the stochastic process (1), we now get

$$\begin{aligned}\pi_{j+1}-\pi_j &= (p_{j+1}-p_j)+(\eta_{j+1}-\eta_j)\\ &\approx \int_{t_j}^{t_{j+1}} a(b-p(t))dt + \int_{t_j}^{t_{j+1}} \sigma p(t)(1-p(t))dW(t)+\varphi_j\sqrt{\Delta_j}\xi_j\\ &\approx a(b-p_j)\Delta_j+\sigma p_j(1-p_j)\sqrt{\Delta_j}e_j+\xi_j\sqrt{\pi_j(1-\pi_j)k_j/N}\\ &= a(b-\pi_j+\eta_j)\Delta_j+\sigma(\pi_j-\eta_j)(1-\pi_j+\eta_j)\sqrt{\Delta_j}e_j+\xi_j\sqrt{\pi_j(1-\pi_j)k_j/N},\end{aligned}$$

where $\{e_j\}$ are independent standard normal r.v.'s. This has drift $a(b-\pi_j)$ and variance

$$\text{Var}[\pi_{j+1}-\pi_j \mid \pi_j]=a^2\pi_j(1-\pi_j)\Delta_j^2/N+\sigma^2\pi_j^2(1-\pi_j)^2\Delta_j+\pi_j(1-\pi_j)k_j/N,$$

by the independence of e_j and ξ_j. For small Δ_j, the last two terms dominate, and so

$$\pi_{j+1}-\pi_j\approx a(b-\pi_j)\Delta_j+\varsigma(\pi_j,N\Delta_j)\pi_j(1-\pi_j)\sqrt{\Delta_j}\varepsilon_j,$$

where ε_j is a standard normal r.v. that suitably combines e_j and ξ_j, and where

$$\varsigma(\pi,N\Delta)\equiv\sqrt{\sigma^2+k(\pi)/[N\Delta\pi(1-\pi)]}\equiv\sqrt{\sigma^2+\sigma_\eta^2(\pi,N)/\Delta}>\sigma. \tag{10}$$

So $\lim_{N\Delta\downarrow 0}\varsigma(\pi,N\Delta)=\infty$ and $\lim_{N\Delta\uparrow\infty}\varsigma(\pi,N\Delta)=\sigma$. Note that $\sigma_\eta^2\approx 2/(N\pi(1-\pi))$ when t_{j+1} and t_j are close since then $k\approx 2$.

Step A3. ELECTION "POLL". For an election at t_{j+1}, put $\pi_{j+1}=p_{j+1}$. So $\pi_{j+1}-\pi_j$ equals

$$\begin{aligned}p_{j+1}-p_j-\eta_j &\sim \int_{t_j}^{t_{j+1}} a(b-p(t))dt+\int_{t_j}^{t_{j+1}}\sigma p(t)(1-p(t))dW(t)+\xi_j\varphi_j\sqrt{\Delta_j/2}\\ &\approx a(b-p_j)\Delta_j+\sigma p_j(1-p_j)\sqrt{\Delta_j}e_j+\xi_j\sqrt{\pi_j(1-\pi_j)k_j/(2N)},\end{aligned}$$

as $\eta_{j+1}-\eta_j=\varphi_j\sqrt{\Delta_j}\xi_j$ implies $\eta_j\sim\xi_j\varphi_j\sqrt{\Delta_j/2}$. Since $p_j=\pi_j+\eta_j$, the above has mean $a(b-\pi_j)\Delta_j$ and variance $\varsigma^2(\pi_j,N\sqrt{2}\Delta_j)\pi_j^2(1-\pi_j)^2\Delta_j$. That is, the volatility equals the poll volatility with elapse time $\sqrt{2}\Delta_j$. The case of an election at t_j is similar. ‖

APPENDIX B. THE NUMERICAL OPTIMIZATION METHOD

Our numerical method is as follows (see Duan and Simonato, 2001; and Seydel, 2002):

Step B1. GRID, TRANSITION MATRIX, AND INITIAL VALUES. Select the discrete interval $\Delta\pi$ between π values and the discrete time period Δt. These define the grid in (π,t)-space $(0,1)\times[0,5]$. Compute the discrete time transition matrix M of π from (7) and Table 1. Set $F_i^0(0)=[50\ldots50]^T$, where $F_i^j(t)$ is the j-th value function (column vector) for varying π levels and $i\in\{L,R\}$. Fix the convergence variable $\chi>0$. Set $j=0$.

Step B2. THE VALUE FUNCTION WHEN AN ELECTION IS CALLED. Calculate the value function at the end of the five year period: $F_i^{j+1}(5)=M_\delta F_i^j(0)$, where $M_\delta=M^n$ is the transition matrix of π for the δ-period, $M_{i,j}^n=\sum_{k=1}^{\infty}M_{i,k}^rM_{k,j}^s$ for any non-negative integers r and s with $r+s=n$, and n is the closest integer to $\delta/\Delta t$. Note that $F_i^{j+1}(5)$ is also the value function if the election is called before the end of the period, that is, Ω^i.

Step B3. THE VALUE FUNCTION. For each $n\in\{1,\cdots,5/\Delta t\}$, calculate the waiting value $\hat{F}_i^{j+1}(5-n\Delta t)=MF_i^{j+1}(5-(n-1)\Delta t)$, and then check for an early election:

$$[F_i^{j+1}(5-n\Delta t)]_z=\max\{[F_i^{j+1}(5)]_z,[\hat{F}_i^{j+1}(5-n\Delta t)]_z\},$$

for all $z\in\{1,\cdots,1+1/\Delta\pi\}$, where $[F]_z$ is the z-th element of F.

Step B4. CONVERGENCE TEST. If

$$\frac{\Delta t\,\Delta\pi}{1+\Delta\pi}\sum_{z\in\{1,\ldots,1+1/\Delta\pi\}}\sum_{n\in\{1,\ldots,1/\Delta t\}} |\,[F_i^{j+1}(5-n\,\Delta t)]_z-[F_i^{j}(5-n\,\Delta t)]_z\,|<\chi$$

then stop. Otherwise set $j=j+1$ and return to step 1.

By adapting Section A.3, this algorithm converges, and F_i^{j+1} approximates the value function on the grid, by Proposition 3. The grid's optimal election time is found by

$$\tau^j(z)=\min\{n\in\{0,\cdots,5/\Delta t\}:[F_i^{j+1}(n\,\Delta t)]_z\le[F_i^{j+1}(5)]_z\},$$

It thus gives the exercise barrier in the grid. The grid's value function approximates the true value function as the mesh size increases, more finely covering $(0,1)\times[0,5]$.

Acknowledgements. We thank David Levine, Tim Maull, Xu Meng, Shinichi Sakata, Sophie Shive, Dan Silverman, Tuomo Vuolteenaho, Xiqiao Xu, the Editor, and four referees for useful feedback, and Daniel Buckley and Catherine Tamarelli for research assistance. We benefited from seminar feedback at the INFORMS Annual Meeting (2005), the University of London, Michigan, Yale, Georgetown, Iowa, Washington University in St Louis, and Wisconsin. We are also grateful to the Gallup organization for freely providing us their data. Insights into British parliamentary history were graciously given us by Neil Rhodes of Leeds Metropolitan University. The authors acknowledge financial support for this research from the National Science Foundation.

REFERENCES

ALVAREZ, L. H. R. and KEPPO, J. (2002), "The Impact of Delivery Lags on Irreversible Investment Under Uncertainty", *European Journal of Operational Research*, **136**, 173–180.

ALVAREZ, L. H. R. and SHEPP, L. A. (1998), "Optimal Harvesting of Stochastically Fluctuating Populations", *Journal of Mathematical Biology*, **37**, 155–177.

AUSTEN-SMITH, D. and BANKS, J. S. (1988), "Elections, Coalitions, and Legislative Outcomes", *American Political Science Review*, **82**, 405–422.

BALKE, N. S. (1990), "The Rational Timing of Parliamentary Elections", *Public Choice*, **65**, 201–216.

BLAIS, A., GIDENGIL, E., NEVITTE, N. and NADEAU, R. (2004), "Do (Some) Canadian Voters Punish a Prime Minister for Calling a Snap Election?", *Political Studies*, **52**, 307–323.

CAMPBELL, J. and THOMPSON, S. (2005), "Predicting the Equity Premium out of Sample: Can Anything Beat the Historical Average?" (Mimeo, Harvard).

CASELLA, G. and BERGER, R. L. (2002), *Statistical Inference*, 2nd edn (Pacific Grove, CA: Duxbury).

CHOWDHURY, A. R. (1993), "Political Surfing Over Economic Waves: Parliamentary Election Timing in India", *American Journal of Political Science*, **37**, 1100–1118.

COCHRANE, J. R. (1996), "A Cross-Sectional Test of an Investment-Based Asset Pricing Model", *Journal of Political Economy*, **104**, 572–621.

DAVYDOV, D. and SMITH, L. (2000), "Optimal Electoral Timing in a Parliamentary Democracy" (Ph.D. Thesis, Dmitry Davydov).

DIERMEIER, D. and MERLO, A. (2000), "Government Turnover in Parliamentary Democracies", *Journal of Economic Theory*, **94**, 46–79.

DIXIT, A. and PINDYCK, R. S. (1994), *Investment Under Uncertainty* (Princeton, NJ: Princeton University Press).

DUAN, J.-C. and SIMONATO, J.-G. (2001), "American Option Pricing Under GARCH by a Markov Chain Approximation", *Journal of Economic Dynamics and Control*, **25**, 1689–1718.

DYBVIG, P. H. and LOEWENSTEIN, M. (2003), "Employee Reload Options: Pricing, Hedging, and Optimal Exercise", *Review of Financial Studies*, **16**, 145–171.

ELLIS, C. and THOMA, M. (1991), "Partisan Effects in Economies with Variable Electoral Terms", *Journal of Money, Credit, and Banking*, **23**, 728–741.

EVANS, L. C. (1998), *Partial Differential Equations* (Providence, RI: American Mathematical Society).

FEDDERSEN, T. J. and PESENDORFER, W. (1996), "The Swing Voter's Curse", *American Economic Review*, **86**, 408–424.

HURST, E. and STAFFORD, F. (2004), "Home is Where the Equity Is: Liquidity Constraints, Refinancing and Consumption", *Journal of Money, Credit and Banking*, **36**, 985–1014.

JACKA, S. (1991), "Optimal Stopping and the American Put", *Mathematical Finance*, **1**, 1–14.

KARATZAS, I. and SHREVE, S. (1998), *Methods of Mathematical Finance* (New York: Springer).

KARLIN, S. and TAYLOR, H. M. (1981), *A Second Course in Stochastic Processes* (New York: Academic Press).

KAYSER, M. A. (2005), "Who Surfs, Who Manipulates? The Determinants of Opportunistic Election Timing and Electorally Motivated Economic Intervention", *American Political Science Review*, **99**, 17–28.

KAYSER, M. A. (2006), “Trade and the Timing of Elections”, *British Journal of Political Science*, **36**, 437–457.

KELLER, G. and RADY, S. (1999), “Optimal Experimentation in a Changing Environment”, *The Review of Economic Studies*, **66** (3), 475–507.

KOU, S. G. and SOBEL, M. E. (2004), “Forecasting the Vote: A Theoretical Comparison of Election Markets and Public Opinion Polls”, *Political Analysis*, **12**, 277–295.

LESMONO, J. D., TONKES, E. J. and BURRAGE, K. (2003), “An Early Political Election Problem”, *ANZIAM Journal*, **45**, C16–C33.

LIPTSER, R. C. and SHIRYAEV, A. N. (2001), *Statistics of Random Processes*, 2nd edn (New York: Springer-Verlag).

MOSCARINI, G. and SMITH, L. (2001), “The Optimal Level of Experimentation”, *Econometrica*, **69** (6), 1629–1644.

ØKSENDAL, B. (2003), *Stochastic Differential Equations: An Introduction with Applications*, 6th edn (New York: Springer-Verlag).

PALMER, H. (2000), “Government Competence, Economic Performance and Endogenous Election Dates”, *Electoral Studies*, **19** (2–3), 413–426.

PEURA, S. and KEPPO, J. (2006), “Optimal Bank Capital with Costly Recapitalization”, *Journal of Business*, **79**, 2163–2201.

SANDERS, D. (2003), “Pre-Election Polling in Britain, 1950–1997”, *Electoral Studies*, **22**, 1–20.

SEYDEL, R. (2003), *Tools for Computational Finance* (New York: Springer-Verlag).

SMITH, A. (1996), “Endogenous Election Timing in Majoritarian Parliamentary Systems”, *Economics and Politics*, **8**, 85–110.

SMITH, A. (2003), “Election Timing in Majoritarian Parliaments”, *British Journal of Political Science*, **33**, 397–418.

SMITH, A. (2004), *Election Timing* (Cambridge, UK: Cambridge University Press).

SMITH, L. (1997), “Time-Consistent Optimal Stopping”, *Economics Letters*, **56** (3), 277–279.