Optimal Electoral Timing:
Exercise Wisely and You May Live Longer*

Jussi Keppo†  Lones Smith‡  Dmitry Davydov§
IOE Department  Economics Department  Equities Division
University of Michigan  University of Michigan  Goldman Sachs

December 14, 2005

Abstract

Using insights from option pricing in finance, we explore the electoral timing problem in a flexible-term democracy. We assume that politically heterogeneous voters continuously learn over time about the evolving party merits. This yields a mean-reverting stochastic process for the support of the parties. The incumbent sees its poll support, and must call an election within five years of the last election. We assume that it wishes to maximize its expected total time in office, so that this resembles the optimal exercise rule for an American financial option. We characterize how the renewable election option behaves, as waiting and stopping values interact.

We then calibrate our model to the Labour-Tory rivalry in the U.K., with polling data from 1943–2005 and the 16 elections after 1945. Excluding three elections essentially forced by weak governments, our maximizing story explains when the elections were called quite well. We also measure the value of election options, finding that over the long run they should more than double the expected time in power of a fixed term electoral cycle.

*We thank Tim Maull, Xu Meng, Shinichi Sakata, Sophie Shive, Dan Silverman, Tuomo Vuolteenaho, and Xiqiao Xu for considered feedback, and Daniel Buckley and Catherine Tamarelli for research assistance. We benefited from seminar feedback at the INFORMS Annual Meeting (2005) and Michigan. 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. Lones acknowledges financial support for this research from the National Science Foundation.

†email: keppo@umich.edu; web: umich.edu/~keppo
‡email: lones@umich.edu; web: umich.edu/~lones
§email: dmitry.davydov@gs.com
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 instead revisits a classic political economy question: optimal electoral timing.

In a key thread of a parliamentary democracy’s fabric, the incumbent often has 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-WWII experience of the United Kingdom. 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 sixty months. In theory, we find that optimally exercising this option has tremendous value there, more than doubling the expected time in power versus 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 rigorously on just one consideration, as we assume that the government simply wishes to maximize its expected total time in power, and find that this has significant explanatory power on when elections have been called.

We wish to draw the analogy of the electoral timing choice to the optimal exercise time of an American option — i.e. the right to buy or sell a stock in a fixed window of time. 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 an 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 the 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 both well-developed and tractable, 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.

We assume an immense number of politically heterogeneous voters who wish to vote for the “best governing party”. Uncertainty plays a major role in our model, as 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; however, right is far more readily convinced to vote Tory than is left. In other words, ordinal preferences coincide — i.e. all prefer the best party — but cardinal preferences diverge. This blend not only subsumes political ideologues for extreme cardinal preferences, but also captures the fact that the intensity of political allegiances differs across voters.

Next, towards a political ebb and flow, we assume that the best party periodically and randomly changes according to a 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 right party is best, say, and low when the left party is. This yields in Lemma 1 a simple continuous time stochastic process for the political slant, the chance that the right is the best party. As the best party periodically switches, this stochastic process is mean-reverting. In fact, its long-run distribution is so well-behaved that we are able to precisely compute it (Lemma 2). And once we assume an exponential distribution over the strength of political beliefs, the political slant equals the chance that it is the best party (Lemma 3). This transforms our Bayesian story into a simple law of motion for political support.

As noted, the government’s timing decision is analogous to the optimal exercise time of an American option, and our solution methodology so adapts methods from finance. Ours is an optimal stopping exercise. At each moment, the government entertains a waiting value depending on the political slant and time left, and stops when it coincides with a slant-dependent stopping value (Proposition 2). So a government calls an election when its political standing first hits a nonlinear stopping barrier. Unlike usual optimal stopping exercises in economics, our stopping value is recursively defined in terms of future waiting values. But as the optimal exercise time for the finite horizon American put option is not analytically known, our harder optimization problem can only be found numerically. Still, we characterize the optimal strategy and analytically derive its comparative statics. For instance, when the political support is more volatile, elections tend to be later. Also, the expected time in power has a convex-then-concave shape, as a function of the support.
We calibrate our model to the U.K., the parliamentary democracy with the longest time series of voting intention polls. We use the public polling data from 1943–2005 and the seventeen elections 1945–2005, and assume a left-right rivalry between Labour and Tory.\(^1\) Given the question posed, we assume that each poll is simply a noisy observation of the actual election outcome that would have obtained that day. We estimate the learning process parameters from the polling data: They are statistically significant, and do not statistically depend on whether an election campaign is in progress.

We use the estimated polling process parameters to solve for the optimal election times. Our contribution here is the formulation and characterization of this as a nonlinear stopping barrier. We hereby assume that the best information about political support is given by the public polling process; this ignores private polls and non-polling information. Even though the best way to predict election outcomes is to exploit the polling process to date, this does not mean that today’s polling result is sufficient for future elections — nor is it even a best forecast, since the electoral process mean reverts:\(^2\) A government riding high in the polls indeed believes that its trend is most likely down.

We then compare the predicted and realized election times. With just one explanatory variable apart from the elapse time, our theory explains 43% of the variation in the timing decisions of governments not troubled by weak or minority governments. Further, the theoretical and actual election decisions have correlation coefficient 0.65. This fit augurs in favor of our main thesis that governments time elections to maximize the expected time in power, using information summarized by the polling process.

Our paper also offers a compelling normative message. The freedom to optimally time the next election clearly confers upon an incumbent government an advantage unavailable in fixed election cycle regimes. The British government can postpone the election until the economy is looking up, while this is impossible in the U.S. Our model can quantify the long-run average magnitude of this advantage. We find that election options have great value, more than doubling the expected time in power in the U.K. Specifically, British governments have averaged 3.65 years since 1945. But 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 more than three for Labour (from 36 to 10 years), and by slightly more than two for Tory (from 17 to 7 years). Variable terms on average benefit the more popular party far more than the less popular party. Constitutional designers should be aware of the

---

\(^1\)We finally test our model by using the 16 elections after World War II since, as will be seen later, the time difference between elections is one factor in the election timing and, thus, we lose one observation.

\(^2\)Kou and Sobel (2004) find that election financial markets better predict election outcomes than polls.
magnitude of this incumbent’s advantage and its differential effect on the parties — when wisely exercised — in choosing amongst fixed and flexible electoral electoral terms. We show that the option is worth twenty-four years for Labour and only eight for Tory.

**Literature Review.** Diermeier and Merlo (2000) is an equilibrium model of electoral timing. They assume that many types of governments (like minority) can form, early elections can happen, and that some governments are less stable than others. We ignore important complications like minority governments and coalitions, in the interests of a focused formulation of the electoral timing problem. Moreover, we assume that the party with the most votes wins. This is true in all but two extremely close British elections.

Palmer (2000) finds that macroeconomic performance and political context both affect election timing. Better economic indicators lead to early elections. While this finding is not inconsistent with our approach, we argue that better economic indicators intuitively improve a government’s standing in the polls, and *thereby* lead to an earlier election.

There is also less related work on timing with a political element. Ellis and Thoma (1991) and Chowdhury (1993) explored the link between timing and political business cycles. Smith (1996) considers election timing with strategic signalling by assuming that the choice of election date reveals information about the government. This is empirically analyzed in Smith (2003), who finds that when calling an early election, one experiences a decline in one’s popular support relative to pre-announcement levels. This holds here, as a government calling an election riding high in the polls is fully aware of the mean reversion it faces. *Yet despite the expected fall in support*, we show quite clearly that the electoral timing option is quite valuable.

Unlike all work that we have seen, our novelty is to formulate and solve the election timing decision problem, exploiting its formal similarity to the exercise timing of options. In so doing, we show how the polling process and the time from the last election induce a nonlinear stopping boundary for calling the next election. We hope that this usage of modern finance in political economy will prove generally profitable.

**Structure of the Paper.** In Section 2, we motivate our paper, showing that this electoral timing option has been useful in practise. Section 3 describes the model, and §4 the theoretical election timing results. In §5, we estimate the model parameters with British polling data. We then test the model in §6 with post-war support levels at the election times. In §7, we price the option value. Appendix A gathers analytical derivations, while Appendix B describes the numerical solution of the optimal stopping problem.
2 THE ELECTORAL TIMING OPTION IN HISTORY

This section is purely motivational. As already outlined, the United Kingdom has flexible electoral terms, a long polling series, and a long two party alternation. This makes it an ideal candidate for exploring the electoral timing option. However, 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 harder to maintain.

To address this heterogeneity, we explore the national and state or provincial governments of Canada and the U.S.A. In Canada, the winner is the party supplying the prime minister or premiers, and for the U.S.A., 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 thus begin with the first regime shift after 1930 (so that a shift must exist), and for which power has alternated between just two parties.³

For each state, province, or country, we ask how many consecutive years the same government is in power. Delaware, eg., 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 $\frac{38}{5} = 7.6$ years, or 1.9 terms. Altogether, we have 46 data points for the USA, and six for Canada. We find that the average government duration is 8.19 years for the U.S.A. and 15.43 for Canada — in other words, 2.05 four year terms for the U.S.A. and 3.09 five year terms for Canada. Using a pooled $t$-test, we find that $t = 2.58$; we can confidently reject the hypothesis of equal mean numbers of terms. Clearly, the electoral timing option has significant value.⁴,⁵

We now try to precisely model and analyze this option, and carefully calibrate and test it in the specific case of the U.K.

---

³Canada 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. The first criterion just eliminates the Democratic bastion of Georgia. The latter prunes Connecticut, Maine, Minnesota, and Oregon, as well as the Canadian provinces of B.C., Saskatchewan, Manitoba, Ontario, and Quebec.

⁴A 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 B.C. and Ontario have recently informally changed to fixed four year terms.

⁵The Canadian province of Quebec had a separatist government from 1976–1999. 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.
3 THE DYNAMIC POLITICAL PROCESS

A. 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 \( L \) and \( R \). We assume that two big parties, denoted also by \( L \) and \( R \),
win all elections. Party \( L, R \) is best in the unobserved political state \( \theta = L, R \). We assume
a continuum \([0, 1] \) of voters, and focus only on such big party \( B \)-voters. While the number
of \( B \)-voters could be stochastic, this would not matter as we only analyze proportions.

The state is random. It follows an exogenous Poisson stochastic process, intuitively
governed by the evolution of the political and economic situation. That is, given each
state \( \theta = L, R \), the state switches in a time interval of length \( \Delta t \) 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.

B. The Information Process. The political state \( \theta(t) \) is unobservable. We develop a
tractable stylized “informational representative agent” voter model of the unobserved \( \theta \).
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
\( p(t) > 0.5 \). From this, voters can compute their expected utility from voting for \( L \) or \( R \).

Voters freely learn about the current political state from a public information process \( \xi(t) \)
in continuous time. In state \( \theta \), in any \( \Delta t \) time interval, \( \Delta \xi(t) \) is normally distributed, with
mean \( \beta_\theta \Delta t \), variance \( \gamma^2 \Delta t \), and conditionally independent of signals at other times,\(^6\) with
\( \beta_R > \beta_L \). The political slant stochastic process is a sufficient statistic for \( \xi \).

Lemma 1 (Dynamics) The political slant \( p(t) \) given signal \( \xi(t) \) obeys Bayes rule:

\[
dp(t) = a(b - p(t))dt + \sigma p(t)(1 - p(t))dW(t),
\]

where \( a = \lambda_L + \lambda_R > 0 \), \( b = \lambda_L/a \in (0, 1) \), \( \sigma = (\beta_R - \beta_L)/\gamma > 0 \), and where \( W \) is a
Wiener process. The variance of the process \( p(t) \) increases in the diffusion coefficient \( \sigma \).

Proof: The drift term is \( E_{p(t)}[p(t + dt)] - p(t) = (1 - \lambda_R dt)p(t) + \lambda_L dt(1 - p(t)) - p(t) \),
which can be written as \( a(b - p(t))dt \). By Bayes rule, the belief difference is:

\[
p(t + \Delta) - p(t) = \frac{p(t)P(\Delta \xi(t)|R)}{p(t)P(\Delta \xi(t)|R) + (1 - p(t))P(\Delta \xi(t)|L)} - p(t) \propto p(t)(1 - p(t))
\]

\(^6\)In other words, \( \xi(t) \) satisfies the stochastic differential equation \( d\xi(t) = \beta_\theta dt + \gamma dZ(t) \), where \( Z(t) \)
is a standard Wiener process.
Figure 1: The Long Run Density of the Political Slant $p$ in the U.K. The estimated parameters are $a = 1.59$, $b = 0.47$, $\sigma = 0.35$. The chance that $L$ wins is $P(p \leq 0.5) = 75\%$.

The diffusion term $\sigma p(t)(1 - p(t))$ is formally derived in Theorem 9.1 of Liptser and Shiryayev (1977), who show that $\sigma = (\beta_R - \beta_L)/\gamma$, the signal-to-noise ratio. The comparative static in $\sigma$ is proved in Appendix A.1. It exploits the fact that if the political slant starts at $p$, then its expectation after an elapse time $t$ is $m(p, t) = e^{-at}p + (1 - e^{-at})b$, just as if the political slant process was noiseless: $\dot{p}(t) = a(b - p(t))$. □

Parameters $a$ and $b$ describe the unobserved political dynamics, while $\sigma$ 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 $(\beta_R - \beta_L)/\gamma —$ the more volatile are public beliefs. The parameter $a$ captures the speed of convergence towards the mean $b$. For instance, even if an $R$-government starts out with $p = 0$, by 3 years it is within 1% of the mean $b$, using the estimated U.K. parameter $a = 1.59$ (see §5.C) and the expectation $m(p, t) = e^{-at}p + (1 - e^{-at})b$ of $p(t)$. This speaks to the brief British “political cycle” — namely, the expected time it takes for the state to switch from $L$ to $R$ and back to $L$, or vice versa, equals $(1/\lambda_L) + (1/\lambda_R) = (\lambda_L + \lambda_R)/\lambda_L\lambda_R = 1/(ab)(1 - b) = 2.52$ years.

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 2 (The Long Run Density)*** Assume $\sigma^2 > 0$. Then the political slant process $p(t)$ forever remains in $(0, 1)$, where its stationary probability density $\psi(p)$ is given by:

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

Figure 1 depicts the long run density for the U.K. parameters estimated in §5.C. Since
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 $R$. This density is single-peaked, this in itself is a finding of the model. The density $\psi(p)$ is U-shaped for high belief variance $\sigma$. For then, state switches quickly become known, and the political slant spends most of its time near 0 or 1. This is not true for the U.K. Further, since the estimated $b < 0.5$ for the U.K, the process favors $L$ — on average, $L$ is ahead $P(p \leq 0.5) = 75\%$ of the time. So the U.K. generally enjoys a left-slant.

C. 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 right. A type-$(u, v)$ voter has (cardinal) utility 0 if the wrong party is elected, $u > 0$ if $L$ is rightly elected, and $v > 0$ if $R$ is rightly elected. He earns expected payoff $[1 - p(t)]u$ from voting $L$, and $p(t)v$ from $R$ (see Figure 2). He votes for $R$ if $p(t) > u/(u + v)$ and for $L$ if $p(t) < u/(u + v)$. So a voter becomes more left-leaning (or right leaning) as $u/v \to \infty$ (or 0), and in the limit, never votes $R$ (or $L$). This framework subsumes doctrinaire voters as a special case.

Lemma 3 (Political Slants Become Electoral Support) If preference parameters $u$ and $v$ 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 $t$.

Proof: The fraction of $B$-voters supporting $R$, for whom $v > [1 - p(t)]u/p(t)$, equals

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

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 $p$. 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 The Stopping and Waiting Values for the Timing Model

We assume that the government seeks to maximize the expected total time in power in the current streak. It 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.

The decision to call an election is an optimal stopping exercise. The stopping time $\tau$ is a function of the remaining time until the next election $T - t$ and the realized political slant. When the ruling party $i$ follows an optimal strategy, its expected time in power at time $t$ is $F^i(p, t)$, while its expected time in power once an election is called is $\Omega^i(p)$. These are the waiting and stopping values of the dynamic programming exercise, and each admits an expression in terms of the other.

If party $i$ wins when the political slant is $p$, then it enjoys an expected waiting value $F^i(p, 0)$. As the process $p(t)$ is continuous, we assume that $F^R(p, 0) = 0$ for all $p < 1/2$, and $F^L(p, 0) = 0$ for all $p > 1/2$. So,

$$\Omega^i(p) \equiv \delta + E(\text{remaining time in power } F^i(p \text{ election day}, 0) \mid \text{initial } p)$$ (2)

Unlike other stopping exercises in economics and finance, the stopping value itself is defined recursively. This greatly enriches the option exercise below in Figure 3. Even if the government stands at 100% on election day, it loses the election with a boundedly positive chance, say at least $\ell > 0$. This yields an upper bound $\Omega^i < \delta + (T + \delta)/\ell$.

The value of a standard put or call option is continuous in the price. Thus, the option is not worth much when the price only slightly exceeds the strike price at the expiration date. This is obviously not the case with an election, where a single vote can separate winning and losing. Given (2), $\Omega^i$ includes a binary option paying at maturity the “asset or nothing” (the value of the underlying asset if it expires in the money) — here, paying $F^i$ or zero. As $\Omega^i$ involves no optimal timing exercise, it is a “European option”.

Easily, a government has the option of running out its full term, so that $F^i(p, t) \geq T$.

---

7One 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. Thus, the objective function can be seen as a maximization of the long-run average time in power.
Figure 3: Waiting and Stopping Values $F^R$ and $\Omega^R$ for the U.K. $\Omega^R(p)$ is the solid line. From top to bottom, the dotted lines are the numerically-computed waiting value functions $F^R(p, 0^+)$, $F^R(p, 1)$, $F^R(p, 2)$, $F^R(p, 3)$, and $F^R(p, 4)$. The expected time in power once elected $F^R(p, 0^+)$ is the limit as $t \downarrow 0$ of $F^R(p, t)$ by (3), and so is continuous in $p$. It is only slightly increasing in the current support. The optimizing party $R$ does not call an election at time $t$ iff $F^R(p, t) > \Omega^R(p)$, the stopping value. Observe how $F^R(p, t)$ is smoothly pasted onto $\Omega^R(p)$.

Forward-looking behavior may entail an earlier election, since we also wish to maximize the expected value $\Omega^i(p)$ once the election is called:

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

(3)

By (2) and (3), $F^i(p, t)$ is then an American option on the binary European option $\Omega^i$, i.e., it is an option on an option. In the appendix, we argue by recursive means that:

**Proposition 1 (Existence)** There exists a continuous, smooth, and monotone solution $\Omega^i, F^i$ to (2) and (3) where $F^R(p, t), \Omega^R(p)$ rises in $p$, and $F^L(p, t), \Omega^L(p)$ falls in $p$.

Write the waiting value as $F^i(p, t) = F^i(p, t) + \Omega^i(p)$, where $F^i(p, t) \geq 0$ is the time value of the electoral option. Recall that an American option is more valuable with a longer exercise time horizon, as it is optimized over a larger domain. By the same token,

**Lemma 4 (More Time Helps)** The waiting value $F^i(p, t)$ and stopping value $\Gamma^i(p, t)$ both fall in the elapse time $t$, and therefore rise in the horizon $T$.

We define the expected drift of the waiting value $\mathcal{A}F^i$ as follows: The expected change of the waiting value $F^i(p, t)$ in $[t, t + dt]$ is $\mathcal{A}F^i(p, t)dt$, which is given by

$$\mathcal{A}F^i(p, t) = F^i_t(p, t) + F^i_p(p, t)a(b - p) + \frac{1}{2} F^i_{pp}(p, t)\sigma^2 p^2(1 - p)^2$$

(4)
Intuitively, $F^i$ falls in $t$ by $F^i_t dt$; the political slant drift moves $F^i$ by $F^i_p dp = F^i_p a(b-p) dt$, and its volatility changes $F^i$ by $\frac{1}{2}F^i_{pp}(dp)^2 = \frac{1}{2}F^i_{pp}\sigma^2 p^2 (1-p)^2 dt$, if $F^i$ is nonlinear in $p$.

Next we analyze the optimal electoral exercise strategy of the American option.

**Proposition 2 (Optimality)** The best election time for party $i = L, R$ is the first time $\tau$ before $T$ such that $F^i(p(\tau), \tau) = \Omega^i(p(\tau))$. Also, for all $(p, t) \in (0, 1) \times (0, T)$, we have:

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

(b) The value is expected to fall daily by at least one day: $1 + AF^i(p, t) \leq 0$,

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

These are standard variational inequalities (see e.g. Øksendal, 1998) for the value (3), assuming that Ito’s Lemma applies. They jointly imply that waiting yields a unit flow utility, balancing the expected time lost in office: $AF^i(p, t) = -1$ while $t < \tau$. The waiting value $F^i(p, t)$ therefore falls as time $t$ advances (see Figure 3), until the day an election is called, when $F^i = \Omega^i$. Figure 4 schematically illustrates the situation for the ruling party $R$. If the election is called too early then this is suboptimal, for then $F^i(p, t) > \Omega^i(p)$. On the other hand, if it is called late, then the value function falls daily by more than one day: $AF^i(p, t) < -1$. By complementary slackness, the government either waits or calls an election, i.e., one of the inequalities (a) or (b) is tight, as claimed.

Assume $t < T$. Party $R$ delays an election with the political slant 0, since time can only help winning, and so $\Gamma^R(0, t) > 0$. Since $\Gamma^R(p, t)$ is continuous, it is positive for all low enough $p$. Let the optimal stopping barrier $p^R(t)$ be the time-dependent minimum solution $p$ to $F^R(p, t) = \Omega^R(p)$, and $p^L(t)$ the maximum solution $p$ to $F^L(p, t) = \Omega^L(p)$. As in Dixit (1993), we have smooth pasting along the stopping barrier: $F^i(p, t) = \Omega^i(p)$. Since the time value $\Gamma^R(p, t) > 0$ is locally falling in $p$ as we approach the barrier from below, and is falling in time $t$, the barrier must be falling in time.\(^8\) Altogether:

**Corollary 3 (Optimal Election Barriers)** Assume that the election term is not over.

(a) Party $R$ delays an election iff $p(t) < p^R(t)$ and party $L$ delays iff $p(t) > p^L(t)$.

(b) The stopping barrier $p^R(t)$ falls in time, while the barrier $p^L(t)$ rises.

\(^8\)To see this more formally, observe that $d\Gamma^i = \Gamma^i_t dp^i(t) + \Gamma^i_p dp^i(t) dt = 0$ along the barrier, since $\Gamma^i = 0$. We have shown that $\Gamma^i_t < 0$. Also, $\Gamma^R > 0$ for $p < p^R$, and $\Gamma^R$ vanishes along the barrier; therefore, $\Gamma^R(p^R(t), t) < 0$, and likewise, $\Gamma^L(p^L(t), t) > 0$. Thus, $dp^R(t) < 0 < dp^L(t)$.
4.2 The Expected Time in Power

Stock option values are convex in the price, as more risk pushes weight into the exercise tail. This property holds for our election option, despite our general stopping value:

**Lemma 5 (The Waiting Value)** The waiting value $F_R(p, t)$ is a convex function of the political slant $p$ for $p < \bar{p}_R(t)$, whereas $F_L(p, t)$ is convex in $p$ for $p > \bar{p}_L(t)$.

The logic is standard from Bayesian learning (see Easley and Kiefer, 1988). First, information has value, raising the expected time in power, as it can be ignored. But it also causes a mean-preserving spread in beliefs $p$. Ipso facto, such spreads must be valuable. So the waiting value is convex in the political slant $p$. For the analogous reason, we might add, one’s utility function is concave iff one is averse to zero-mean wealth gambles.

To understand better the shape of the stopping value, it helps to study the chance $V^i(p)$ that $i$ wins the election given political slant $p$. Recall from §3-B that the expected election day political slant is $m(p, \delta) = e^{-a\delta}p + (1 - e^{-a\delta})b$. The Appendix proves:

**Lemma 6 (The Win Chance)** The chance $V^R(p)$ is convex when $p \leq \frac{1}{2}e^{a\delta} + (1 - e^{a\delta})b$, and concave when $p \geq \frac{1}{2}e^{a\delta} + (1 - e^{a\delta})b$ — and conversely for $V^L(p)$.

As seen in Figure 4, party $i$ calls an election when the political slant hits the stopping barrier $\bar{p}_i(\tau)$. Upon winning, party $i$ acquires a new waiting value function $F^i(p, 0+)$, calculated from (3) by taking the limit $t \downarrow 0$. This value is nearly flat in $p$ (as in Figure 3) — for the next election is far in the future, and the current support should not matter much, so that $F^i(p, 0+) \approx F^i(0+)$. (Recall from §3-B that the fast mean reversion pushes all beliefs $p$ within 1% after three years.) One can approximate the stopping value by $\Omega^i(p) \approx \delta + V^i(p)F^i(0+)$. So governments can act as if they are merely trying to win
back a single fixed term of length \( F_i(0+) \). They can in effect behave partially myopically. Intuitively then, \( \Omega^i \) inherits the convex-concave shape of the win chance \( V^i \), and this is numerically true (as seen in Figure 3). This figure also reveals that the election barrier hits \( \Omega^i \) on its concave portion. We analytically argue this in the Appendix.

**Lemma 7 (Local Concavity)** The stopping value \( \Omega^i(p) \) is locally concave at \( \bar{p}^i(t) \).

Indeed, the maximization (3) is equivalent to the solution of the PDE in Proposition 2 as well as value matching and smooth pasting. Consistent with Lemma 7, elections are only called early if \( p \) is high, precisely where the convex-concave value \( \Omega^R(p) \) is concave.

As the waiting value is convex and stopping value concave in \( p \), \( F_i(p, \bar{p}^i(t), t) - \Omega^i(p, \bar{p}^i(t)) \) rises in \( \sigma \) for \( t < T \). To restore smooth pasting optimality condition, the barrier changes:

**Corollary 4 (Barriers and Risk)** The barrier \( \bar{p}^R(t) \) rises and \( \bar{p}^L(t) \) falls in volatility \( \sigma \).

The effect of volatility on the expected time in power is ambiguous. The government behaves like a decision maker with a utility function that is convex and then concave, by Lemmas 5 and 7. It is thus ambivalent about risk: Neither greater \( \sigma \) nor a longer election period \( \delta \) (which also raises election risk) has a clear effect on the expected time in power.

We provide intuitions for the comparative statics that we cannot prove but have strong numerical support: the campaign period \( \delta \), and the mean reversion level \( b \) and speed \( a \).

- **The Campaign Period** \( \delta \). When \( \delta \) rises by \( d\delta \), this lengthens the expected time in power directly by \( d\delta \), which lifts both \( \Omega^i \) and \( F_i \). Also, the support drifts longer, affecting both terms by \( F_i(p, t)a(b - p)d\delta = \Omega^i(p, t)a(b - p)d\delta \), equal by smooth pasting. The difference is the variance effect. As with \( \sigma \) in Corollary 4, greater \( \delta \) elevates election period uncertainty, thereby raising \( F_i \) relative to \( \Omega^i \). So \( \bar{p}^R(\tau) \) rises and \( \bar{p}^L(\tau) \) falls in \( \delta \).

- **The Mean Reversion Speed** \( a \). Greater \( a \) is tantamount to smaller \( \sigma \) since it lowers the variance of \( p(t) \), but returns it faster towards the mean reversion level \( b \). Depending on whether \( b \) lies above or below \( 1/2 \), this can help or hurt the expected value.

But for the election barriers we are more definite. Consider a moment past time \( t \) at the stopping barrier \( \bar{p}^R(\tau) \), when the waiting value is falling at rate faster than 1, and stopping is best. If one considers the standard case when the barrier bracket the mean-reversion level, \( \bar{p}^R(\tau) > b > \bar{p}^L(\tau) \), then the waiting value drift \( AF_i \) rises when the speed \( a \) falls, by (4). Thus, one no longer needs to stop at time \( t \). So \( \bar{p}^R(\tau) \) rises and \( \bar{p}^L(\tau) \) falls.

- **The Mean Support Level** \( b \). When the mean \( b \) rises, party \( R \)'s winning chance rises in any election, and thus so does its expected time in power. Clearly, \( F^R \) then rises. Since the drift in the waiting value \( AF^R \) rises by (4), the barrier \( \bar{p}^R(\tau) \) rises, by the same reasoning as with the mean reversion speed \( a \). The opposite results hold for party \( L \).
5  POLLS: THEORY, DATA, AND ESTIMATION

A. Theory. Politicians enjoy a variety of ways to take the pulse of the electorate — many quite qualitative. We assume that governments time their elections using monthly voting intention polls. These survey individuals planning to vote asking whom they would pick if an election were called that day. They are noisy observations of the true political support $p(t)$. But as the government consists of citizens privy to the information process $\xi(t)$, our model in principle affords no role for polls. We therefore venture a story with a mild boundedly rational flavor. Imagine that individuals cannot operate Bayes rule, but can nonetheless deduce whom they would vote for. This is a simple binary decision, and requires far less introspection than the production of a probability by a method that they may employ but not understand. Governments can then learn from polls, as these record voting preferences, and thereby summarize information. By Lemma 3, one can view the political slant as the support process for party $R$. Also, from Lemma 1, equation (1) is the law of motion for the support for party $R$. We next deal with polling noise.

In a given time-$t$ poll with sample size $N$, let $\pi(t) = p(t) + \eta(t)$ be the proportion of $B$-voters that support $R$. As is well-known, the poll error $\eta$ obeys a $t$-distribution, with variance $\sigma^2_\eta(\pi) \equiv \pi(1 - \pi)/N$, and so is asymptotically normal. Hence, $\eta$ behaves approximately like $\sigma_\eta \epsilon$, where $\epsilon$ is a mean 0 and variance 1 normal r.v.

Let us denote the polling times by $\{t_j\}$. Since the polling error does not depend on the gap $\Delta_j \equiv t_{j+1} - t_j$ between polls, its dynamic effect increases in the poll frequency.

**Lemma 8 (Poll Dynamics)** The poll process obeys a stochastic differential equation similar to the election process (1). Its discrete-time process is approximately:

$$\pi(t_{j+1}) - \pi(t_j) \approx a(b - \pi(t_j))\Delta_j + \hat{\sigma}(\pi(t_j), N\Delta_j)\pi(t_j)(1 - \pi(t_j))\sqrt{\Delta_j} \epsilon_j$$  (5)

where $\epsilon_j \sim N(0, 1)$ and $\hat{\sigma}(\pi, N\Delta) > \sigma$, $\lim_{N\Delta \downarrow 0} \hat{\sigma}(\pi, N\Delta) = \infty$, $\lim_{N\Delta \uparrow \infty} \hat{\sigma}(\pi, N\Delta) = \sigma$.

Let us denote the polling times by $\{t_j\}$. Since the polling error does not depend on the gap $\Delta_j \equiv t_{j+1} - t_j$ between polls, its dynamic effect increases in the poll frequency.

**Lemma 8** (proof in Appendix) exploits the fortuitous fact that the variance $\sigma^2_\eta(\pi)$ shares the nonlinear form $p(1 - p)$ of the volatility of $p(t)$. This implies that the polling

---

9 Two other non-behavioral stories present themselves. First, polls may be relevant because there exist some “noise voters” — those who vote in a random fashion, uninfluenced by the political slant. However, given this additional layer of noise, support behaves approximately like the political slant. Polls are then useful as they record the actual voting intentions, and follow the law of motion that we identify in (5). Second, and more subtly, we may diverge from the informational representative agent, and assume heterogeneously-informed agents. In aggregate, the voting intentions again obey approximately the same law of motion as the political slant, and (5) still applies. The complexity of neither approach is justified.
process has a similar law to the political slant process. To the discrete-time process \((5)\) corresponds a completely facetious continuous-time polling process \(\pi(t)\) that is noisier than \(p(t)\), having the diffusion coefficient \(\hat{\sigma} > \sigma\).\(^{10}\) The theory of \(\S3–4\) applies to \(\pi(t)\).

**B. Polling History and Data.** Our data set from the United Kingdom (the ‘British’) consists of two poll time series of voting intentions dating from June 1943 – May 2005.\(^{11}\) The sample sizes are large, mostly between 1000–1500. The first poll time series are Gallup polls from June 1943 to May 2001, after which they were discontinued. In 1997, it shifted from face-to-face interviews to telephone surveys. The second time series, the MORI Political Monitor,\(^{12}\) spans August 1979 – May 2005 and its sample size varies from 500 to 17,000. We average any same-day polling results of Gallup and MORI.

As noted, we only study \(B\)-voters, i.e., supporters of Labour \((L)\) or Tory \((R)\). We first calculate the realized values of \(\pi\), i.e., the Tory polling support among the \(B\)-voters, excluding small parties. Figure 5 illustrates the poll levels \(\pi\) from June 1943 – May 2005. The polls on average have favored Labour, since the average poll value is 0.46 during this time span (even though we later compute that \(b = 0.47\)).

**C. Estimating the Polling Process.** In our empirical analysis, we wish to explain the variation of governments’ election decisions. To answer this question, we work in the political support space \((\pi \text{ for } R \text{ and } 1-\pi \text{ for } L)\), comparing theoretical and realized political support levels at the times of election calls — i.e., we let our theory explain the variation in the realized support levels. Before this analysis, we estimate the model parameters from the historical polling data. The poll history can be understood as the sample data, and the support levels as the out of sample data, since the comparison between the model and actual support levels is done by using the parameter estimates.

Equation \((5)\) is an autoregressive model.\(^{13}\) The parameters \(a, b\), as well as the standard deviations \(\sigma, \sigma_{\eta}\) are estimated by ordinary least squares (OLS), as follows. Transform the dependent variable of \((5)\) 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 \(Z_j = -\sqrt{\Delta_j}/[1-\pi(t_j)]\).

\(^{10}\)In finance theory, prices may be modeled as if in continuous-time, despite discrete time observations. This corresponds to a process with a certain fixed elapse time, such as \(\Delta = 1\).

\(^{11}\)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.


\(^{13}\)This is consistent with Sanders (2003) who shows that a long term autoregressive model provides accurate forecasts for government support in the U.K.
Figure 5: Proportion of Tory B-voters, 6/1943–5/2005. The polling process $\pi$ has averaged 0.46 (i.e. left-leaning), and ranged from 0.23 to 0.67. Elections are the dashed lines.

We estimate parameters $a$ and $b$ in (5) from the regression:\footnote{The delta-method (see e.g. Casella and Berger, 1990) gives the standard deviation of $b$, $\sigma$, and $\sigma_\eta$.}

$$Y_j = (ab)X_j + aZ_j + \hat{\sigma} \epsilon_j$$

Next, define $V_j = (Y_j - (ab)X_j - aZ_j)^2$. The polling error affects the polling process, and since the polling frequency peaks during the last weeks before the election, so does the volatility. Since $\hat{\sigma} = \sqrt{\sigma^2 + \sigma_\eta^2/\Delta_j} = \sqrt{\sigma^2 + \sigma_\eta^2 d_j}$, where $d_j = 1/\Delta_j$, we then estimate

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

where $\{\epsilon_j\}$, $\{\epsilon_j\}$ are standard normal iidrv’s. For simplicity, we assume $\sigma_\eta$ is constant.\footnote{According to (8) and our data set, this is justified since about 97% of $\pi$ values lie in $[0.3, 0.7]$, which imply that $\sqrt{\pi(1-\pi)} \in [0.46, 0.50]$. By (8) and assuming $N \approx 1000$, $\sigma_\eta = \sqrt{k(\pi)/(N\pi(1-\pi))} \approx \sqrt{1/500/(\pi(1-\pi))}$ which is between $[0.126, 0.137]$, i.e., close to our estimate (0.12).}

Table 1 summarizes the estimated process with four parameter sets: outside the $\delta$-period, inside the $\delta$-period, inside without the no-choice governments, and overall. These $R^2$ levels may seem low, but are very good by comparison to the best empirical work in financial time series. See, eg., Table 3 in Campbell and Thompson (2005).

The parameters are not significantly different outside and inside the $\delta$-period.\footnote{The $t$-statistic for the test that $a$ coincides outside and inside the $\delta$-period (without no-choice govern-}

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure5.png}
\caption{Proportion of Tory B-voters, 6/1943–5/2005. The polling process $\pi$ has averaged 0.46 (i.e. left-leaning), and ranged from 0.23 to 0.67. Elections are the dashed lines.}
\end{figure}
Table 1: The Estimated Polling Parameters. The ‘overall model’ uses all the data; the pre-
\( \delta \)-period uses data before the election time is announced; the \( \delta \)-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 “no short” refers to the absence of the no-choice governments.)

<table>
<thead>
<tr>
<th>( \delta )-period</th>
<th>( R^2 ) : 3.61%, 3.30%</th>
<th>( \delta )-period (no short)</th>
<th>( R^2 ) : 3.56%, 3.06%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( a ) ( b ) ( \sigma ) ( \sigma_\eta )</td>
<td>estimate</td>
<td>( a ) ( b ) ( \sigma ) ( \sigma_\eta )</td>
</tr>
<tr>
<td>estimate</td>
<td>6.43 0.52 0.80 0.11</td>
<td>estimate</td>
<td>6.39 0.52 0.82 0.10</td>
</tr>
<tr>
<td>st. dev.</td>
<td>3.32 0.04 0.66 0.03</td>
<td>st. dev.</td>
<td>3.62 0.05 0.73 0.03</td>
</tr>
<tr>
<td>Pre-( \delta )</td>
<td>( R^2 ) : 2.61%, 26.43%</td>
<td>Pre-( \delta ) (no short)</td>
<td>Overall</td>
</tr>
<tr>
<td>estimate</td>
<td>1.49 0.46 0.27 0.12</td>
<td>estimate</td>
<td>1.59 0.47 0.35 0.12</td>
</tr>
<tr>
<td>st. dev.</td>
<td>0.29 0.02 0.10 0.00</td>
<td>standard deviation</td>
<td>0.33 0.02 0.16 0.01</td>
</tr>
</tbody>
</table>

\( t \)-statistics of \( a \) and \( b \) are 4.85 and 27.29, they significantly differ from 0. As expected, the
mean poll level \( b \) is near the average \( B \)-poll level in Figure 5. The \( t \)-statistics of \( \sigma \) and \( \sigma_\eta \)
are 2.23 and 23.40 and so are significant.\(^{17}\) The average polling time difference outside the \( \delta \)-period is 0.059 years (about 22 days) and its standard deviation is 0.043 (about 16
days). The analogous numbers inside the \( \delta \)-period are 0.022 years (about 8 days) and
0.024 (about 9 days).\(^{18}\) So the average poll volatilities are different inside and outside the
\( \delta \)-period. But this difference owes to the greater poll frequency — i.e. smaller elapse time
\( \Delta \) between polls — inside than outside the \( \delta \)-period. Altogether, the process parameters
\( a \), \( b \), \( \sigma \), and \( \sigma_\eta \) can be assumed constant.

We assume that the ruling party understands that the volatility during the \( \delta \)-period is
not higher than outside it.\(^{19}\) We then calculate the constant volatility from the polling time
differences over the entire data set (average: 0.056 years \( \approx \) 21 days, standard deviation:
0.054 \( \approx \) 20 days). This gives that the constant volatility is 0.89.

\(^{17}\)For an internal consistency check, note that our estimate \( \sigma_\eta \approx 0.12 \) in Table 1 is near a direct computa-
tion of the standard deviation using our \( t \)-distribution formula in (8): \( \sigma_\eta^2 = \frac{2}{N \pi(t_j)(1-\pi(t_j))} \approx 1/125, \)
assuming \( N \approx 1000 \) and \( \pi(t_j) \approx 1/2 \), so that \( \sigma_\eta \approx 0.09 \). Further, the standard deviation of the polls equals
\( \sqrt{\pi(1-\pi)/N} = \sigma_\eta \pi(1-\pi)/\sqrt{2} = (0.12)(0.25)/\sqrt{2} \approx 2\% \). This is consistent with Sanders (2003), who
admittedly does not exclude small parties.

\(^{18}\)We are able to reject the null hypothesis that the average polling time differences are equal inside and
outside the \( \delta \)-period with a 1% level of significance.

\(^{19}\)The volatility estimates inside and outside the \( \delta \)-period are \( \hat{\sigma}_i = 1.48 \) and \( \hat{\sigma}_o = 0.77 \). These give
\( R^2 = 34\% \) in the regression analysis (without the no-choice governments) in \S 6.2; further, the average
ruling periods for \( L \) and \( R \) in \S 7 are 34 years and 15 years.
6 ACTUAL VERSUS OPTIMAL ELECTION TIMING

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 eighteen days, plus weekends and public holidays. It starts with 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.\textsuperscript{20} Table 2 lists the outcomes of 17 British elections from 1945–2005. While the delay time ranges from 21–45 days, we set $\delta$ to the average delay time 33 days (or 0.09 years).

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 win 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 this 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.

We see that on average, governments have called elections after 3.65 years in our data set. There are three unusually short governments: 2/23/50 – 10/25/51 (609 days), 10/15/64 – 3/31/66 (532 days), and 2/28/74 – 10/10/74 (224 days). Excluding these, the average lifespan has been 4.23 years. In 1951, the Labour government of Clement Attlee called an election only twenty months into his term, forced by a razor thin majority of just five MPs. For this was deemed insufficient to sustain his radical program creating the welfare state that was started with the large majority that Labour enjoyed from 1945–50. 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.\textsuperscript{21}

\textsuperscript{20}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 $\delta$ in §4.2.

\textsuperscript{21}The 1979 election was also forced by losing a nonconfidence vote. Since it was just four months before the legal term expired, we do not isolate this election.
Table 2: United Kingdom Election Results, 1945–2005. Tory and Labour columns are the vote percentage (seats percentage) for the two main parties. The \((\pi, t)\) columns lists the poll level \(\pi\) and the time \(t\) from the last election at the time election was announced. The starred elections were called because of weak governments, and thus not completely free choice variables.

<table>
<thead>
<tr>
<th>Election date</th>
<th>Announced</th>
<th>Winner</th>
<th>Tory</th>
<th>Labour</th>
<th>((\pi, t))</th>
</tr>
</thead>
<tbody>
<tr>
<td>7/5/45</td>
<td>5/23/45</td>
<td>Labour</td>
<td>39.8 (33.3)</td>
<td>47.8 (61.4)</td>
<td></td>
</tr>
<tr>
<td>2/25/50</td>
<td>1/11/50</td>
<td>Labour</td>
<td>43.5 (47.8)</td>
<td>46.1 (50.4)</td>
<td>(0.52, 4.52)</td>
</tr>
<tr>
<td>10/9/51*</td>
<td>9/19/51</td>
<td>Tory</td>
<td>48.0 (51.3)</td>
<td>48.8 (47.2)</td>
<td>(0.57, 1.57)</td>
</tr>
<tr>
<td>5/26/55</td>
<td>4/15/55</td>
<td>Tory</td>
<td>49.7 (54.8)</td>
<td>46.4 (44.0)</td>
<td>(0.51, 3.47)</td>
</tr>
<tr>
<td>10/8/59</td>
<td>9/8/59</td>
<td>Tory</td>
<td>49.4 (57.9)</td>
<td>43.8 (41.0)</td>
<td>(0.54, 4.29)</td>
</tr>
<tr>
<td>10/15/64</td>
<td>9/15/64</td>
<td>Labour</td>
<td>43.4 (48.3)</td>
<td>44.1 (50.3)</td>
<td>(0.49, 4.94)</td>
</tr>
<tr>
<td>3/31/66*</td>
<td>2/28/66</td>
<td>Labour</td>
<td>41.9 (40.2)</td>
<td>47.9 (57.6)</td>
<td>(0.46, 1.37)</td>
</tr>
<tr>
<td>6/18/70</td>
<td>5/18/70</td>
<td>Tory</td>
<td>46.4 (52.4)</td>
<td>43.0 (45.6)</td>
<td>(0.46, 4.13)</td>
</tr>
<tr>
<td>2/28/74</td>
<td>2/7/74</td>
<td>Labour</td>
<td>37.9 (46.8)</td>
<td>37.1 (47.4)</td>
<td>(0.48, 3.64)</td>
</tr>
<tr>
<td>10/10/74*</td>
<td>9/18/74</td>
<td>Labour</td>
<td>35.8 (43.6)</td>
<td>39.2 (50.2)</td>
<td>(0.44, 0.55)</td>
</tr>
<tr>
<td>5/3/79*</td>
<td>3/29/79</td>
<td>Tory</td>
<td>43.9 (53.4)</td>
<td>36.9 (42.4)</td>
<td>(0.54, 4.47)</td>
</tr>
<tr>
<td>6/9/83</td>
<td>5/9/83</td>
<td>Tory</td>
<td>42.4 (61.1)</td>
<td>27.6 (32.2)</td>
<td>(0.61, 4.02)</td>
</tr>
<tr>
<td>6/11/87</td>
<td>5/11/87</td>
<td>Tory</td>
<td>42.2 (57.9)</td>
<td>30.8 (35.2)</td>
<td>(0.58, 3.92)</td>
</tr>
<tr>
<td>4/9/92</td>
<td>3/11/92</td>
<td>Tory</td>
<td>41.9 (51.6)</td>
<td>34.4 (41.6)</td>
<td>(0.49, 4.75)</td>
</tr>
<tr>
<td>5/1/97</td>
<td>3/17/97</td>
<td>Labour</td>
<td>30.7 (25.0)</td>
<td>43.5 (52.0)</td>
<td>(0.37, 4.94)</td>
</tr>
<tr>
<td>6/7/01</td>
<td>5/8/01</td>
<td>Labour</td>
<td>31.7 (25.2)</td>
<td>40.7 (62.5)</td>
<td>(0.36, 4.02)</td>
</tr>
<tr>
<td>5/5/05</td>
<td>4/5/05</td>
<td>Labour</td>
<td>32.3 (30.5)</td>
<td>35.2 (55.0)</td>
<td>(0.46, 3.83)</td>
</tr>
</tbody>
</table>

6.2 Election Timing

Next we analyze election timing for the overall parameters in Table 1. As Proposition 2 applies to a finite time horizon stopping exercise of a complex underlying process (1), only a numerical solution is possible. Appendix B describes our numerical method. The election time is the first hitting time of the polling process. Since the elections are triggered in \(\pi\) space, we analyze the election times by comparing polls \(\pi\) when elections are called and the theoretical barrier polls computed from the estimated process parameters.

To distinguish between the optimal Labour and Tory election strategies, we draw the optimal barriers as a function of the polling support of each party: namely, \(\pi\) for Tories and \(1 - \pi\) for Labour. As seen in Figure 6 and proved in Corollary 3, these barriers fall over time, first gradually and then steeply tending to 0.5 in the last few months of the term. But the Tory barrier is globally nearer 0.5 as the polling process favors Labour \((b = 0.47)\), and thus Tories optimally call elections at lower support levels. The average distance between the barrier and the realized support levels is 8.8% for all governments and 7.1% without
the no-choice Labour governments. These numbers are 11.0\% for Labour (7.9\% without the no-choice governments) and 6.7\% for Tory. The Tory election calls have evidently been closer to the optimal policy. This might afford insight into why the Tories have led the polls about 33\% of the time from 1945–2005, but have ruled about 58\% of the time.\textsuperscript{22}

Using the polling process path before the elections, we can check how the actual election times diverged from our theoretical predictions. While most elections were called early, just two were more than a month late: Thatcher should have called the 1983 election eleven 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 does not directly take into account. Likewise, Blair should have called the election of 2001 eight months earlier.

Next, we analyze how well our model explains the election calls. We do this using the regression models in Figure 7, where the dependent variable $Y$ is the realized support level and the explanatory variable $X$ is the theoretical barrier value at the election announcement time. The no-choice governments aside, we find that $Y = -0.49 + 1.76X$. This intercept is insignificant, while the slope is significant. Since $R^2 = 43\%$, our model

\textsuperscript{22}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 §7, the ruling time difference is not statistically significant. Thus, good luck might explain the ruling time difference just as well.
Figure 7: **Actual and Theoretical Poll Support at Election Calls.** The triangles are the elections that Labour called early due to slim majority or minority governments (10/25/51, 3/31/66, and 10/10/74). The squares are all other Labour election announcement support levels. The circles are the election announcement Tory support levels. In the regressions $Y = \text{actual support level}$, $X = \text{model support level}$, and the parameter standard deviations are parenthesized. The solid line is without the no-choice (“short”) governments, the dot-dash line is without the no-choice governments and without the intercept, and the dashed line is with all the governments.

clearly explains a large portion in the variation of election times through the regression. Also, the correlation of the theoretical and realized support levels at the election times is 0.65. At the very least, we have correctly identified $\pi, t$ as important decision factors.\(^{23}\)

If the elections were called solely using our model with the estimated parameters, then the regression line should coincide with the diagonal $Y = X$. As the intercept of the choice governments is insignificant, Figure 7 also includes the best zero-intercept regression, $Y = 0.90X$. The $t$-statistic on this slope is now 32.25 and $R^2 = 33\%$. 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 that the slope is one, since the $t$-statistic is 3.58.

The slope test is obviously a joint test on the model and its parameters. To ensure

\(^{23}\)But as seen in Figure 7, our model fares poorly if the no-choice governments are also included. As argued, ours is not a theory of when minority or bare majority governments call elections.
a unified paper built on our fleshed-out theory, we have engaged in a very conservative econometric exercise — for instance, assuming constancy over the time period 1945–2005 (see §7), and introducing no other explanatory factors (see the conclusion). Any additional degrees of freedom would surely have improved the fit.24

Further testing the result, we regressed the residuals of the regressions without the no-choice governments on the realized election time, the election year, and the incumbent party. The coefficients were insignificant: All $t$-statistics were less than 0.9, $R^2 = 9\%$ without the no-choice governments, and $R^2 = 13\%$ with no intercept. Thus, neither the party, the election year, nor the elapse time offer any further significant predictive power.

Our contribution rests on our derivation of a rationally-founded nonlinear stopping barrier, just as with American options in the stock market. But might a simpler naive model have done better? How important is the nonlinearity? To this end, we tossed aside the theory, and re-ran the regressions in Figure 7 assuming that election times can be linearly explained (via OLS) using only the elapse time from the last election. This gives the regression $Y = 0.85 - 0.08t$ without the no-choice governments, where $Y$ is the actual support level. Both parameters are significant, their $t$-statistics are 5.3 and $-2.1$. But the $R^2$ of this regression drops to $28\%$ (and just $2\%$ with all governments). So our optimizing nonlinear model is not only rationally justified, but also better explains the variation in the election times than does an a-theoretical linear regression. This comparison is reassuring.

7 THE OPTION VALUE OF ELECTION TIMING

The option to freely time an election increases the expected time in power — because the incumbent can always ignore the option and hold elections at their term’s end. We now measure the value of having optimal electoral timing. Figure 8 illustrates the expected times in power $F^L(\pi, 0)$ and $F^R(\pi, 0)$ as a function of the initial polling process $\pi$ (i.e. the last election outcome). Integrating these expected times over the long-run polling density in Lemma 2 reveals the average worth of these options. We weight these expectations using the derived (lower variance) political slant process $p(t)$ in Table 3 (see also Figure 1), since we are taking the perspective of the true driving process, i.e., we use the process parameters in Table 1 with $\sigma_\eta = 0$, since there is no polling error with $p$.

24By 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 §4.2 and the regression $Y = 0.90X$, in our model the corresponding implied parameters involve lower $\hat{\sigma}$ or greater $a$. In Section 8, we discuss other factors that could improve the model.
The predicted times are quite high, by historical standards\textsuperscript{25} — almost 68 years for Labour, and over 24 for Tory. While it is true that parties have diverged from our optimal exercise rule, one might imagine a lesser cost of this suboptimal behavior. For this reason, we have explored various alternative explanations. Firstly, barring a weak government, elections have never been called within the first three years of a term. Governments might well fear punishment for opportunism.\textsuperscript{26} We thus reformulated our timing exercise, asking that elections be called in years 3–5; this eliminates repeatedly calling an election when riding high in the polls, and lessens the expected time in power, as we see in Table 3.\textsuperscript{27}

Secondly, we have discovered that the polls $\pi$ averaged 0.49 from 1943–92, but just 0.37 in 1992–2005. The $t$-statistics for the difference of these average values equals 32.48, and is obviously significant. The year 1992 may seem arbitrary, but any break point between 1970–1995 produces an extremely significant difference in the poll mean $b$. Of course, we do not model regime shifts, since that would be another topic. But suppose we use the statistically different parameters $a, b, \hat{\sigma}$ for the span before 1992.\textsuperscript{28} This period

\textsuperscript{25}The average historical ruling period (standard deviation) for all governments is 8.6 years (5.1 years), for Labour 6.3 years (1.2 years), and for Tory 11.6 years (7.3 years). Between 7/5/1945-5/5/2005, there has been only 7 ruling periods. Clearly, the average ruling periods are not significantly different.

\textsuperscript{26}For instance, Blais et al (2004) argue that voters punished Jean Chretien for calling a snap election in November 2000 after just three years and four months.

\textsuperscript{27}This does not greatly move our stopping barriers, and the resulting regression for past elections without the no-choice governments is only a slightly worse fit, with $R^2 = 39\%$.

\textsuperscript{28}Specifically, $a = 2.2, b = 0.49, \sigma = 0.19, \sigma_{\eta} = 0.11$, constant volatility $\hat{\sigma} = 0.69$. That is, the $\pi$ process mean reverts faster about a higher mean, with less volatility.
Table 3: **Expected Time in Power.** Under different regime assumptions, we compute (a) the long-run fraction of time in power, and (b) the expected time in power conditional on have just been elected. The starred rows use the polling process that obtained until 1992.

<table>
<thead>
<tr>
<th>flexible timing 0–5 yrs</th>
<th>L rules 87.6% of time</th>
<th>L rules 65.8 yrs</th>
<th>R rules 24.4 yrs</th>
</tr>
</thead>
<tbody>
<tr>
<td>flexible timing 0–5 yrs *</td>
<td>L rules 74.4% of time</td>
<td>L rules 51.9 yrs</td>
<td>R rules 34.3 yrs</td>
</tr>
<tr>
<td>flexible timing 3–5 yrs</td>
<td>L rules 84.6% of time</td>
<td>L rules 36.0 yrs</td>
<td>R rules 17.1 yrs</td>
</tr>
<tr>
<td>flexible timing 3–5 yrs *</td>
<td>L rules 72.5% of time</td>
<td>L rules 31.9 yrs</td>
<td>R rules 23.2 yrs</td>
</tr>
<tr>
<td>elections every 5 years</td>
<td>L rules 79.2% of time</td>
<td>L rules 12.5 yrs</td>
<td>R rules 8.6 yrs</td>
</tr>
<tr>
<td>elections every 4 years</td>
<td>L rules 79.2% of time</td>
<td>L rules 10.1 yrs</td>
<td>R rules 6.9 yrs</td>
</tr>
<tr>
<td>elections every 4 years *</td>
<td>L rules 69.1% of time</td>
<td>L rules 8.9 yrs</td>
<td>R rules 7.6 yrs</td>
</tr>
</tbody>
</table>

less strongly favors Labour, and sees Labour’s average win chance fall from 75% to 69%. Indeed, Tories only win for political slants \( p > 0.5 > 0.47 = b \), the average poll. With a more favorable process, this tail event happens more often. Consequently, Labour’s expected time in power falls to 31.9 while Tory’s rises to 23.2.

Assume now that elections by protocol are called in the 3–5 year window, and use the overall polling process parameters in Table 1. In that event, 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 more than two: from 36 to 10.1 years for Labour, and 17.1 to 6.9 years for Tory. Labour’s expected percentage time in power would drop slightly from 84.6% to 79.2%. An overarching observation here is that flexible electoral timing favors the dominant party far more than does fixed election cycles.

### 8 CONCLUSION

**Summary.** Optimal timing of votes and elections is an important subject, and the periodic topic of great media speculation in some countries. We have sought to demystify this exercise for all concerned, by computing the time maximizing strategies with flexible electoral timing. What we have done takes inspiration from the financial theory of options.

We have designed and analyzed a tractable model capturing the informational richness of the political economy setting: namely, a forward-looking optimizing exercise using an informationally-derived mean-reverting polling process. The optimal election time in this framework is the first moment the polling process hits a nonlinear stopping barrier. We believe that this is a substantively novel optimal timing exercise for economics (see Dixit and Pindyck, 1994). Its execution is also quite unlike other optimal stopping exercises in economics and finance, because the stopping value is recursive. This in itself presented
special hurdles. We pushed the analysis as far as possible, deriving the timing comparative statics and the convex-concave shape of the value function. Again, unlike other comparative statics we have seen in economics, ours were derived by indirect means, without recourse to closed forms. The optimization was done by numerical means, as is the case with finite time horizon American put options in finance.

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 also show that the value of the option to choose the election time can be very substantial, and favors the dominant party.

**Some Caveats.** As usual, our success owes to some simplifying assumptions. The first and foremost difficulty we avoid concerns the size of a win. We sidestep complications of minority or slim majority government, but in our regression, we do identify three elections well-understood to have resulted from weak governments.

A minority government is less desirable for the winner, since it must compromise its preferred governing choices to satisfy its coalition partners. Such a government is more fragile, and may struggle maintaining its winning margin with each vote. Also, a small majority may erode with the passage of time. We are very upfront about this weakness, and a far more involved model might assume higher flow payoffs to governing with a strong than a weak majority or minority government. This would lead to early elections called with minority of weak governments, as we observe.

The second key simplification we make is to ignore the quirks of the ‘first-past-the-post’ voting system, that the vote winner might not win the election. In the U.K., this has not proved critical, and is a justified simplification with little impact.

We also ignore unrelated objectives of the decision makers, positing that governments only maximize their expected time in power. Our single-minded theory explains much of

---

29 We have also ignored any strategic incentives to vote, but these are surely quite miniscule in a national election (see eg. Feddersen and Pesendorfer (1996)). Ours is a theory of strategizing and forward-looking behavior by the government, and not voters.

30 The Tories won in 1951 with a minority of seats, but formed a solid working alliance with the National Liberals. The 1974 Labour government began as a razor slim majority that soon evaporated into a minority government; from March 1977 to August 1978, it was sustained in the ‘Lib-Lab Pact’ (with the Liberals).

31 Even John Major’s 21 seat majority in 1992 slowly shrunk throughout his term, as many government ministers lost their seats. By the 1997 election, it was almost a minority government.

32 We did not use Canadian data in our main empirical analysis because the strong geographic concentration of the major parties produces a systematic divergence between the vote and election count winners.
the variation in election timing decisions with just two factors, polls and elapse time.

The objective function might be modified to incorporate different restrictions — such as certain dates when the elections must be called. This would lower stopping barriers (thereby improving the model fit) as support may be lost before the next open date.

Personal and party interests may conflict. A retiring Prime Minister may wish to prolong his time in power, delaying the election past its best date. Impatience — a year now is worth more than one later — delays elections, and worsens the model fit (Figure 6).

Fourthly, we assume constant parameters over 1945–2005. The parameters naturally change over this time span (see §7), and relaxing it would improve the model fit.

Finally, we posit that all decision-making depends on the polls and elapse time. In fact, the government surely has more accurate information, possibly from private polls, etc. This raises the polling sample size, and lowers the polling volatility and so the election barriers. Furthermore, the government must then engage in a filtering exercise, producing a posterior belief process with smaller variance than $\pi$. We have avoided this highly nontrivial exercise, but have verified that our model implications about electoral timing are reasonably robust to the variance specification. 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.

A OMITTED PROOFS

A.1 Proof of Lemma 1: Variance of the Political Slant Process

We claim

$$ p(t) = m(p, t) + \sigma \int_0^t e^{-a(t-s)} p(s)(1 - p(s)) dW(s), \quad (6) $$

where $m(p, t) = e^{-at}p + (1 - e^{-at})b$ is the expected time-$t$ posterior $E[p(t)|p(0) = p]$. To derive (6) and the $m(p, t)$ formula, differentiate (6). This gives (1), after manipulating:

$$ dp(t) = a \left[ (b - p(0))e^{-at} - \sigma \int_0^t e^{-a(t-s)} p(s)(1 - p(s))dW(s) \right] dt + \sigma p(t)(1 - p(t)) dW(t) $$

The variance of $p(t)$ is thus

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

which equals:

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

26
where we suppress the $\sigma^2$ argument whenever clear. Thus, we have the partial derivatives:

\[
\begin{align*}
v_i(t) &= \sigma^2 E \left[ p^2(t)(1 - p(t))^2 \right] - 2av(t), \\
v_{\sigma^2}(t) &= v(t)/\sigma^2 + \sigma^2 \int_0^t e^{-2a(t-s)} \left[ 1 + 6 (v(s) - m(s)(1 - m(s))) \right] v_{\sigma^2}(s)ds, \\
v_{\sigma^21}(t) &= \sigma^{-2} \left[ v_i(t) + 2av(t) \right] + \sigma^2 \left[ 1 + 6 \{ v(t) - m(t)(1 - m(t)) \} \right] - 2a] v_{\sigma^2}(t) \\
&= E[p^2(t)(1 - p(t))^2] + \sigma^2 \left[ 1 + 6 \{ v(t) - m(t)(1 - m(t)) \} \right] - 2a] v_{\sigma^2}(t).
\end{align*}
\]

We find $v_i(0) > 0$ and $v_{\sigma^2}(0) > 0$ for all $\sigma^2 > 0$ and $t \geq 0$. Now $v_{\sigma^2}(t) = v_{\sigma^2}(0) + \int_0^t v_{\sigma^21}(s)ds$. If $v_{\sigma^2}(t) = 0$ for some $t > 0$, then $v_{\sigma^21}(t) > 0$. Thus, $v_{\sigma^2}(t + \varepsilon) > 0$, where $\varepsilon > 0$ is small, and we get that the variance of $p(t)$ rises in the diffusion coefficient $\sigma$.

### A.2 Proof of Lemma 2: Derivation of the Stationary Density

We appeal to Karlin and Taylor (1981, pages 220 and 241). If $dp(t) = \mu(p)dp + \sigma(p)dW$ has a stationary density $\psi(y) = \lim_{t \to \infty} (\partial/\partial y)P(p(t) \leq y|p(0) = x)$, then it obeys the stationary forward Fokker-Plank equation $\frac{1}{2} \left[ \sigma(p)\psi(p) \right]' - [\mu(p)\psi(p)]' = 0$. In particular, for (1), we have: $\frac{1}{2} \left[ \sigma(p)(1 - p)^2\psi(p) \right]' - [a(b - p)\psi(p)]' = 0$. Its solution is given by $\psi(p) = m(p)[C_1S(p) + C_2]$, where $m(p) = 1/\left(\sigma^2 p^2(1 - p)^2s(p)\right)$ is the speed measure, and $S(p) = \int_{p_0}^p s(y)dy$ is the scale function, whose density equals:

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

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 similarly analyzed using $\bar{p}(t) = 1 - p(t)$ and noting $d\bar{p}(t) = a ((1 - b) - \bar{p}(t)) dt - \sigma \bar{p}(t)(1 - \bar{p}(t))dW(t)$ and $p(t) = 1$ iff $\bar{p}(t) = 0$.

The sufficient conditions that 0 be an entrance (see Karlin and Taylor (1981, pages 226–242)) are $\lim_{y \to 0} \int_y^p s(z)dz = \infty$ and $\lim_{y \to 1} \int_y^p m(z)dz < \infty$, 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. \hfill $\square$

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 $\psi(p) = m(p)/\int_0^1 m(z)dz$.  

27
A.3 Proof of Proposition 1: Existence of Smooth Monotone Values

Put $F^i_0 = 0$ into (2) to compute $\Omega^i_0$. Insert $\Omega^i_0$ into (3) to compute $F^i_1$. Since (2) and (3) define monotone maps $\Upsilon : F^i \mapsto \Omega^i$ and $\Phi : \Omega^i \mapsto F^i$, the iterations obey $0 \leq \Omega^i_0 \leq \Omega^i_1 \leq \Omega^i_2 \leq \cdots$ and $0 \leq F_0^i \leq F_1^i \leq F_2^i \leq \cdots$. Their limits therefore exist, and obey (2)--(3).

Furthermore, each of the maps $\Upsilon, \Phi$ preserve monotonicity in $p$. To see this of $\Phi$, use the stopping time $\tau'$ optimal for $p'$ at $p'' > p'$. This yields a higher expected stopping value $\Omega^R(p(\tau))$ with $p''$ than with $p'$, since the stopping belief $p(\tau)$ is higher, path by path. Once we optimize for $p''$, we therefore find that $F^R(t, p'') > F^R(t, p')$. Much more simply, $\Upsilon$ preserves monotonicity, as it involves no optimal stopping exercise. Since $F^R(0, q) = 0$ for all $q < 1/2$, $F^R$ is clearly (weakly) monotone increasing and not decreasing.

Write $\Omega^R(p) = \int_{1/2}^1 \psi(p, q, \delta) F^R(0, q) dq$, for the smooth transition density $\psi$ in $p$. As $F^R$ is boundedly finite, $\Omega^R$ is continuous and in fact smooth in $p$. Similarly with $L$.

Finally, as a boundedly finite solution to (3), $F^i$ is continuous.

□

A.4 Proof of Lemma 6: The Shape of the Win Chance

If $p \geq \frac{1}{2} e^{a\delta} + (1 - e^{a\delta})b$, then the expected election outcome at the end of the delay period is $m(p, \delta) > \frac{1}{2}$. Now, by Lemma 1, increasing $\sigma$ lifts the variance of $p(\delta)$ and since volatility can only hurt $R$ (as $R$ will most likely win the election), it makes losing more likely. Thus, $V^R(p)$ is falling in $\sigma$ for all $p \geq \frac{1}{2} e^{a\delta} + (1 - e^{a\delta})b$, and so must be (locally) concave. Likewise, changing variables $p \rightarrow 1-p$, we find that $V^L(p)$ is concave for all $p \leq \frac{1}{2} e^{a\delta} + (1 - e^{a\delta})b$. But $V^R(p) \equiv 1 - V^L(p)$. Altogether, we deduce the convexity of $V^R(p)$ for all $p \leq \frac{1}{2} e^{a\delta} + (1 - e^{a\delta})b$ and $V^L(p)$ for all $p \geq \frac{1}{2} e^{a\delta} + (1 - e^{a\delta})b$.

□

A.5 Proof of Lemma 7: Local Concavity

Claim 2 (Convexity at the Barrier) $F^i_{pp}(\bar{p}^i, t) \geq \Omega^i_{pp}(\bar{p}^i)$ for all $t < T$.

Proof: We consider only $R$. From Lemma 5 and Proposition 2, $F^R(p, t) - \Omega^R(p) \geq 0$ and $F^R_p(\bar{p}^R, t) - \Omega^R_p(\bar{p}^R) = 0$. Clearly, $F^R_p(\bar{p}^R, t) - \Omega^R_p(\bar{p}^R)$ is convex.

Let $p_t$ be the optimal barrier at $t$, so that $\bar{p}^R(t) = p_t$. By Proposition 2, it obeys the value matching and smooth pasting optimality conditions $\Omega^R(p_t) = F^R(p_t, t)$ and $\Omega^R_p(p_t) = F^R_p(p_t, t) \geq 0$, where $\bar{p}^R(t) = p_t$. For a contradiction, assume that $\Omega^R(p)$ is locally convex at $p = p_t$. Because $\Omega^R_p(p) = \Omega^R_p(0) + \int_0^p \Omega^R_{pp}(y) dy$, if there exists $p_h > p_t$
with \( \Omega^R(p_h) = \Omega^R(p_l) \), then \( \Omega^R \) must be locally concave at the smallest such \( p_h > p_l \).

Assume that \( p_h \) exists and select the smallest \( p_h > p_l \).

We show that this gives a contradiction: Fix \( \Omega^R \), and define \( \hat{F}^R \) by value matching and smooth pasting at \( p_h \), and the PDE in Proposition 2 — which is equivalent to the maximization (3). Since \( \hat{F}^R(p, t) < F^R(p_h, t) = \Omega^R(p_h) \leq \Omega^R(p) \) on \([p_l, p_h]\), we have

\[
\hat{F}^R(p_t, t) = \hat{F}^R(p_h, t) - \int_{p_l}^{p_h} \hat{F}^R(p, t) > \Omega^R(p_h) - \int_{p_l}^{p_h} \Omega^R(p) = \Omega(p_l)
\]

If \( p_h \leq 1 \) does not exist then the maximization (3) produces a higher waiting value \( \hat{F}^R \) with \( \hat{F}^R(t) = 1 \). By Claim 1, this means the election is not called at \( t < T \). \( \square \)

### A.6 Proof of Lemma 8: The Polling Process

Denote \( \pi_j \equiv \pi(t_j) \), etc. Let \( \eta_j = e_{\eta_j} \sqrt{\pi_j(1 - \pi_j) / N} (1 + o(1)) \) be the polling error for \( B \)-voters, where \( o(1) \) vanishes as \( N \to \infty \), and \( \{e_{\eta_j}\} \) are iid standard normal r.v.s independent of the \( \{W(t)\} \). Since the polling error and the political slant process are unobserved, they are random variables even though the poll outcome is known. Write the discrete-time poll differences as \( \pi_{j+1} - \pi_j = (p_{j+1} - p_j) + (\eta_{j+1} - \eta_j) \), where \( \eta_{j+1} - \eta_j = \int_{t_j}^{t_{j+1}} \varphi(t) dW^n(t) \) and where \( W^n \) is independent of \( W \). Hence, for all \( j \in \{1, 2, \ldots, n\} \):

\[
\pi_{j+1} - \pi_j = \int_{t_j}^{t_{j+1}} a(b - p(t)) dt + \int_{t_j}^{t_{j+1}} \sigma p(t)(1 - p(t)) dW(t) + \int_{t_j}^{t_{j+1}} \varphi(t) dW^n(t) \tag{7}
\]

If \( \varphi(t) = \varphi_j \) on \([t_j, t_{j+1}]\), then \( \varphi_j^2 \Delta_j = \text{Var}[\varphi_j(t) | \pi_j] \) by the Ito isometry, and so:

\[
\varphi_j^2 \Delta_j = E[\eta_j^2 + \eta_{j+1}^2 | \pi_j] = \frac{1 + o(1)}{N} \{ \pi_j(1 - \pi_j) + E[\pi_{j+1}(1 - \pi_{j+1}) | \pi_j] \} = \frac{1 + o(1)}{N} k_j \pi_j(1 - \pi_j),
\]

where \( k_j = 1 + E[\pi_{j+1}(1 - \pi_{j+1}) | \pi_j] / \pi_j(1 - \pi_j) \). This selection of \( \varphi_j \) is justified since \( \varphi_j \int_{t_j}^{t_{j+1}} dW^n(t) = \varphi_j[W^n(t_{j+1}) - W^n(t_j)] \) shares the mean and variance of \( \eta_{j+1} - \eta_j \).

Now we write the discrete-time polling difference (7) as follows:

\[
a(b - p_j) \Delta_j + \sigma p_j(1 - p_j) \sqrt{\Delta_j} e_j + \frac{1 + o(1)}{\sqrt{N}} \sqrt{\pi_j(1 - \pi_j) k_j \xi_j} = a(b - \pi_j + \eta_j) \Delta_j + \sigma(\pi_j - \eta_j)(1 - \pi_j + \eta_j) \sqrt{\Delta_j} e_j + \frac{1 + o(1)}{\sqrt{N}} \sqrt{\pi_j(1 - \pi_j) k_j \xi_j},
\]

where \( e_j \) and \( \xi_j \) are iid standard normal variables. This has drift \( a(b - \pi_j) \) and variance

\[
\text{Var}[\pi_{j+1} - \pi_j | \pi_j] = a^2 \pi_j(1 - \pi_j)(1 + o(1)) \Delta_j^2 / N + \sigma^2 \pi_j^2(1 - \pi_j)^2 \Delta_j + k_j \pi_j(1 - \pi_j) N \frac{1 + o(1)}{\Delta_j},
\]
where the two last terms of the variance owe to the independence of \( e_j \) and \( \xi_j \). If \( N \) is high and \( \Delta_j \) low, then the last terms dominate. Thus,

\[
\pi_{j+1} - \pi_j \approx a(b - \pi_j)\Delta_j + \hat{\sigma}(\pi_j, N\Delta_j)\pi_j(1 - \pi_j)\sqrt{\Delta_j}\varepsilon_j,
\]

where \( \varepsilon_j \) is a standard normal variable and

\[
\hat{\sigma}(\pi, N\Delta) = \sqrt{\sigma^2 + k(\pi)/[N\Delta\pi(1 - \pi)]} \equiv \sqrt{\sigma^2 + \sigma^2(\pi)/\Delta} > \sigma. \tag{8}
\]

Thus, \( \lim_{N\Delta \to 0} \hat{\sigma}(\pi, N\Delta) = \infty \) and \( \lim_{N\Delta \to \infty} \hat{\sigma}(\pi, N\Delta) = \sigma. \)

\[\square\]

**B THE NUMERICAL OPTIMIZATION METHOD**

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

**Step 0 (grid, transition matrix, and initial values).** Select the discrete interval \( \Delta \pi \) between \( \pi \) values and the discrete time period \( \Delta t \). These define the grid in \((\pi, t)\) space \((0, 1) \times [0, 5]\). Calculate the transition matrix \( M \) of \( \pi \) over the discrete time from (5) and Table 1. Set \( F^0_i(0) = [50 \cdots 50]^T \), where \( F^j_i(t) \) is the \( j \)'th value function (column vector) for different \( \pi \) levels and \( i \in \{L, R\} \). Select the convergence variable \( \chi > 0 \). Set \( j = 0 \).

**Step 1 (The value function when an election is called).** Calculate the value function at the end of the five year period:

\[
F^{j+1}_i(5) = MF^j_i(0),
\]

where \( M^\delta = M^n \) is the transition matrix of \( \pi \) for \( \delta \)-period, \( M^n_{i,j} = \sum_{k=1}^{\infty} M^r_{i,k}M^s_{k,j} \) for any fixed pair of nonnegative integers \( r \) and \( s \) with \( r + s = n \), and \( n \) is the closest integer to \( \delta/\Delta t \). Note that \( F^{j+1}_i(5) \) is also the value function if the election is called before the end of the period, i.e., it models \( \Omega^i \).

**Step 2 (The Value Function).** For each \( n \in \{1, \ldots , 5/\Delta t\} \), calculate first the waiting value \( \hat{F}^{j+1}_i(5 - n\Delta t) = MF^{j+1}_i(5 - (n - 1)\Delta t) \), and then check the early election:

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

for all \( z \in \{1, \ldots , 1 + 1/\Delta \pi\} \), where \( [F]_z \) is the \( z \)'th element of \( F \).

**Step 3 (Convergence Test).** If

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

then stop. Otherwise set \( j = j + 1 \) and go to step 1.
By §A.3, this algorithm converges, and $F_{i+1}^{j+1}$ approximates the value function on the grid, by Proposition 2. The grid’s optimal election time is found by:

$$\tau^j(z) = \inf \{ n \in \{0, ..., 5/\Delta t\} : [F_{i+1}^{j+1}(n\Delta t)]_z \leq [F_{i+1}^{j+1}(5)]_z \},$$

and so it gives the exercise barrier in the grid. The grid’s value function approximates the true value function as the mesh size increases since then the grid approximates $(0, 1) \times [0, 5]$.

References


