5-1-2006

Optimal Electoral Timing: Exercise Wisely and You May Live Longer

Jussi Keppo

Lones Smith

Dmitry Davydov

OPTIMAL ELECTORAL TIMING:
EXERCISE WISELY AND YOU MAY LIVE LONGER

By

Jussi Keppo, Lones Smith and Dmitry Davydov

May 2006

COWLES FOUNDATION DISCUSSION PAPER NO. 1565

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

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

May 3, 2006

Abstract

In many democratic countries, the timing of elections is flexible. We explore this potentially valuable option using insights from option pricing in finance.

The paper offers three main contributions on this problem. First, we derive a rationally-based mean-reverting political support process for the parties, assuming that politically heterogeneous voters continuously learn over time about evolving party fortunes. We solve for the long-run density for this process and derive the polling process from it by adding polling noise.

Second, we explore optimal timing using the political support process. The incumbent sees its poll support, and must call an election within five years of the last election to maximize its expected total time in office. This resembles the optimal exercise rule for an American financial option. This option is recursive, and the waiting and stopping values subtly interact. We prove the existence of the optimal exercise rule in this setting, and show that the expected longevity is a convex-then-concave function of the political support. Our model is tractable enough that we can analytically derive how the exercise rule responds to parametric shifts.

We 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 quite well explains when the elections were called, and beats simple linear regressions. 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.

---

[*]This supersedes a manuscript Optimal Electoral Timing in a Parliamentary Democracy (2000) that was an unpublished PhD thesis chapter of Dmitry Davydov, joint with Lones Smith. That 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. 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), the University of London, the University of Michigan, Yale University, Georgetown University, and the University of Iowa. 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.

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 tradeoff, we 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 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.

We assume an immense number of politically heterogeneous voters wish to vote for the "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; 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 current posterior chance that the right is the best party. 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 transforms our Bayesian story into a 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, closer to finance. The government's timing decision is analogous to the optimal exercise time of an American option. 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 nonlinear stopping barrier. Unlike the usual optimal stopping exercises in economics, our stopping value is recursively defined in terms of future waiting values. We assume that when calling an election, the process further evolves during a campaign period delay, rendering the outcome uncertain. Also, the government only has access to noisy polling data, and does not know its actual support. Not surprisingly, 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 prove existence of the solution of this recursive option (Proposition 2), and characterize it by variational equations (Proposition 3). We show that 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, eg., tend to be later with more volatile political support. We analytically motivate many other comparative statics.

Our third contribution is an empirical test of our timing model, and our finding that timing matters, i.e., 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 a long time. Since democracy is so young, we choose the provinces of Canada and states of the USA. 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., because it is the parliamentary democracy with the longest time series of voting intention polls. Moreover, it has just two serious parties. We use public polling data from 1943–2005 and the seventeen elections 1945–2005, and assume a left-right rivalry between Labour and Tory.[1] 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 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 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 call elections to maximize the expected time in power, using information summarized by the public polling process.

Our paper then 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 can quantify the long-run average magnitude of this advantage. We find that election options have great value, more than doubling 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.

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 more than three for Labour (from 36 to 10 years), and by slightly more than two for Tory (from 17 to 7 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. We show that the option is worth twenty-four years for Labour and only eight for Tory.

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 model's 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.

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 complications like minority governments and coalitions, in favour of a focused attack on the electoral timing problem.

There is a large and less related literature on timing and political business cycles.[3] For instance, Palmer (2000) finds that macroeconomic performance and political context both affect election timing. Better economic indicators lead to early elections. While not inconsistent with our approach, we argue that better economic indicators intuitively improve a government's standing in the polls, and *thereby* lead to earlier elections. Our governments take the polling process as a given, and optimize against it.

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.

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

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

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 (2005)). 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 nonstationary decision rule over time, and is the source of interest in this paper. We solve for the nonlinear exercise boundary for the electoral timing problem.

Structure of the Paper. In Section 2, we motivate our paper, showing that this electoral flexibility has been useful in practice. 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 electoral option and §8 concludes. 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.[5] 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

[4]Sanders (2003) analyzes polling error, and Alvarez and Keppo (2001) study the effect of delays.

[5]Gubernatorial term limits (e.g. www.termlimits.org/Current_Info/State_TL/gubernatorial.html) apply in several states. About 10% of all governorships after 1930 ended due to the term limits. The estimated probability that the ruling party changes after the term limit is active is 0.58 and 0.44 when the term limit is not active. With 5% significance level we cannot reject the hypothesis that these estimates are the same. Therefore, we ignore the term limits here.

(so that a shift must exist), and for which power has alternated between just two parties.[6] 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, 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 $38/5 = 7.6$ 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.[7,8]

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

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 left L and right R. Two *big parties*, denoted also by L and R, win all elections.[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.

[6]The first criterion just eliminates the Democratic bastion of Georgia. The latter prunes Connecticut, Maine, Minnesota, and Oregon, and the provinces of B.C., Saskatchewan, Manitoba, Ontario, and Quebec.

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

[8]The 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.

[9]Our approach is flexible extends to any number of parties, and in fact, Davydov and Smith (2000) considered three. To avoid the complexity of matrices, we simply allow two here.

B. The Information Process. The political state $\theta(t)$ is unobservable. We develop a tractable stylized "informational representative agent" voter model of the unobserved θ.

We assume a continuum $[0,1]$ of voters, and focus only on big party B-*voters*. While the number of B-voters could be stochastic, this would not matter since only proportions matter. 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 political state by perusing the newspaper, watching television, and listening to radio. We assume that news revelation is a continuously unfolding process of constant intensity. Such a non-lumpy stationary news process is tractably and optimally summarized by assuming a *Gaussian public information process* ξ in continuous time: namely, the stochastic differential equation $d\xi(t) = \beta_{\theta(t)}dt + \gamma dZ(t)$, where $Z(t)$ is a Wiener process and $\beta_R > \beta_L$.[10] More concretely, in state θ, in any Δt time interval, $\Delta\xi(t)$ is normally distributed, with mean $\beta_\theta \Delta t$, variance $\gamma^2 \Delta t$, and signals conditionally independent of other times. It is obvious that information $\xi(\cdot)$ is sufficient for beliefs. But the reverse is true too: The political slant process $p(\cdot)$ is sufficient for information $\xi(\cdot)$.

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

Proof:[11] The state is $\theta(t) = L$ and switches to $\theta(t + dt) = R$ with chance $(1 - p(t))\lambda_L dt$, while $\theta(t) = R = \theta(t + dt)$ with chance $p(t)(1 - \lambda_R dt)$. The drift is then the expected new posterior belief that $\theta(t + dt) = R$ less the old one, or:

$$E[dp(t)] = E[p(t + dt)] - p(t) = (1 - p(t))\lambda_L dt + p(t)(1 - \lambda_R dt) - p(t) = a(b - p(t))dt$$

The variance 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. Bolton and Harris (1999) offer an intuitive derivation by Bayes rule, yielding the first order terms:

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

[10] Moscarini and Smith (2001) prove this nice property of the observation process.

[11] Keller and Rady (1999) derive this law of motion. We include it here for a self-contained treatment.

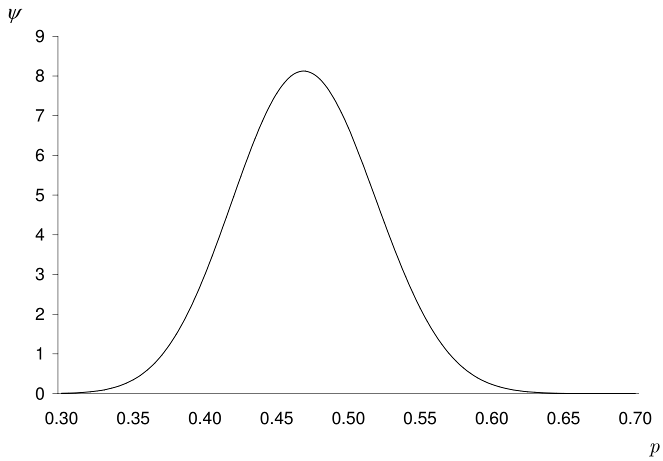

Figure 1: **The Long Run Density of the Political Slant p in the U.K.** The estimated parameters are $a = 1.59, b = 0.47, \sigma = 0.35$. The chance that L wins is $P(p \leq 0.5) = 75\%$.

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.*

Proof: The formula for the mean $m(p,t)$ obtains just as if the political slant process was noiseless, since it is an expectation, and the noise drops out: Thus, solve the differential equation $\dot{p}(t) = a(b - p(t))$. The comparative static in σ is proved in Appendix A.1. $\square$

Parameters a and b describe the unobserved 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" $(\beta_R - \beta_L)/\gamma$ — the more volatile are public beliefs. The parameter a captures the *speed of convergence* towards the *mean* b. For example, starting at $p = 0$, the expected slant after 3 years lies within 1% of the mean b, by Lemma 2 for the estimated U.K. parameter $a = 1.59$ (see §5.C). This fast mean reversion speed will later mean that parties do not have to proceed in an extremely far-sighted manner, since winning big is not much better than winning small.

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) *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:*

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

Figure 1 depicts the long run density for the U.K. parameters estimated in §5.C. Since this density is single-peaked, this in itself is a finding of the model, because one can

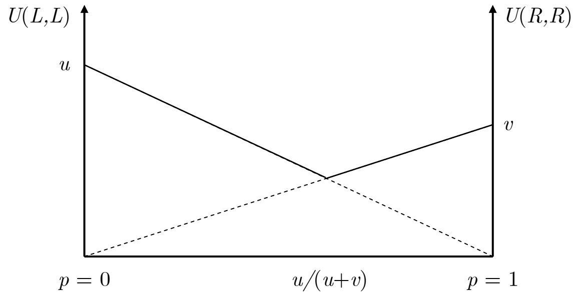

Figure 2: **A Voter's Preferences.** The figure schematically depicts a typical voter's utility maximization: he votes for L if $p < u/(u+v)$, the cross-over level, and otherwise he votes for R.

show that not all densities of the form (2) 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 $b < 0.5$ 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. 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, utility $u > 0$ if L is rightly elected, and utility $v > 0$ 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). A far-sighted 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 $p(t) > u/(u+v)$ and 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 4 (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 party R is the total fraction of the parameters (u, v) for which $v > [1 - p(t)]u/p(t)$. This equals the double integral

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

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.*

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 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.* 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.

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 by $F^i(p, t)$ its expected time in power at time t, and its expected time in power *once an election is called* $\Omega^i(p)$. (We drop the superscript whenever possible.) These are the dynamic programming *waiting* and *stopping values*, 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 $p > 1/2$. In other words, party R is in power, and so enjoys a duration $F^R(p, 0) > 0$ iff the p process starts out above $1/2$. The payoffs at the moment τ the election is called are $\Omega(p(\tau))$, for the recursively defined function:

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

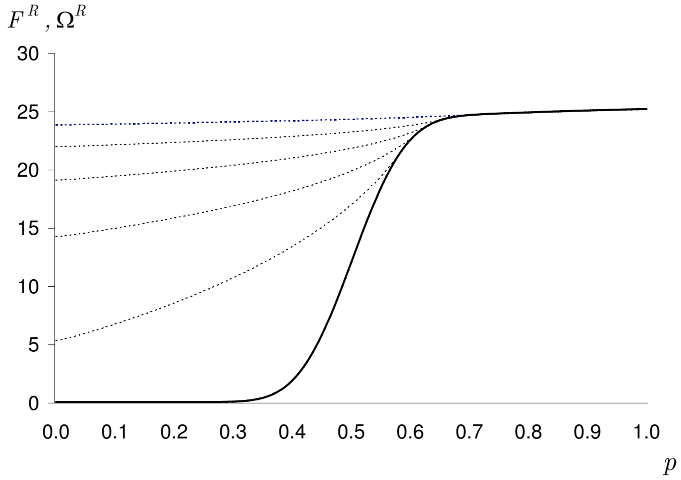

Figure 3: **Numerical Waiting and Stopping Values F^R and Ω^R for the U.K.** $\Omega^R(p)$ is the solid curve. From top to bottom, dotted lines are the numerically-computed waiting value functions $F^R(p, 0+)$, $F^R(p, 1)$, $F^R(p, 2)$, $F^R(p, 3)$, and $F^R(p, 4)$. The expected time in power once elected $F^R(p, 0+)$ 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 for each time t that $F^R(p, t) > \Omega^R(p)$, the stopping value. *In a wonderful numerical illustration of smooth pasting, observe how each $F^R(p, t)$ is tangent to $\Omega^R(p)$ at each time t.*

Since the stopping value is defined recursively, the option exercise is greatly enriched.

The value of a standard put or call option is continuous in the underlying price. Thus, the option is not worth much when the option is only slightly "in the money" at the expiration date. By contrast, with an election, a single vote can separate the glory of victory from the sting of defeat: Landing slightly "in the money" is discontinuously better than slightly "out of the money". In other words, Ω includes a *binary option* in (2) paying at maturity the "asset or nothing" (the value of the underlying asset if it expires) — here, paying F or zero. As Ω involves no 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 behavior generally entails an earlier election, since we care about the expected value $\Omega(p)$ once the election is called.[12] Since this is the sum of the time until the election and the continuation value, we have:

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

By recursive equations (2) and (3), $F(p, t)$ is then an American option on the binary

[12]Maximizing the expected time in power corresponds to a zero interest rate, since a day now is worth a day later (Smith, 1997). We do not consider discounting, since it adds little, and we have no justifiable interest rate. Also, discounting would decrease the expected value and postpone elections because the utility from the future falls. This pushes barriers away from realized political support levels (see Figure6).

European option Ω, i.e., it is *an option on an option.* In the appendix, we argue by recursive means that a solution exists to these stopping and continuation values.

Proposition 2 (Existence) *There exists a continuous, smooth, and monotone solution Ω^i, F^i to equations (2)–(3) where $F^R(p,t), \Omega^R(p)$ rise in p, and $F^L(p,t), \Omega^L(p)$ fall in p.*

Let us next parse the waiting value into $F(p,t) = \Gamma(p,t) + \Omega(p)$, where $\Gamma(p,t) \geq 0$ is the *time value* of the electoral option. A standard American option is more valuable with a longer exercise time horizon, as it is optimized over a larger domain. This holds here:

Lemma 5 (More Time Helps) *The waiting value $F(p,t)$ and the time value $\Gamma(p,t)$ both fall in the elapse time t, and therefore rise in the horizon T.*

To say anything more definite about the option values, we must exploit the stochastic process of beliefs. Define the *expected drift* of the waiting value $\mathcal{A}F$ as follows: The expected change $\mathcal{A}F(p,t)dt$ of the waiting value $F(p,t)$ in $[t, t+dt]$ is given by

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

Intuitively, F falls in t by $F_t dt$; the political slant drift moves F by $F_p dp = F_p a(b-p)dt$, and its volatility changes F by $\frac{1}{2}F_{pp}(dp)^2 = \frac{1}{2}F_{pp}\sigma^2 p^2(1-p)^2 dt$. This final Ito term reflects that volatility matters when F is nonlinear, improving option values when $F_{pp} > 0$.

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

Proposition 3 (Optimality) *The best election time for party $i = L, R$ 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 e.g. Øksendal, 1998) for the value (3), assuming that Ito's Lemma applies. They jointly imply that when waiting is called for because it is better than stopping, or $F(p,t) > \Omega(p)$, the unit flow utility balances the expected time lost in office $\mathcal{A}F(p,t) = -1$. Here, the waiting value $F(p,t)$ falls (see Figure 3) a 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. From then on, the value function would fall daily by more than one day, $\mathcal{A}F(p,t) < -1$. Figure 4

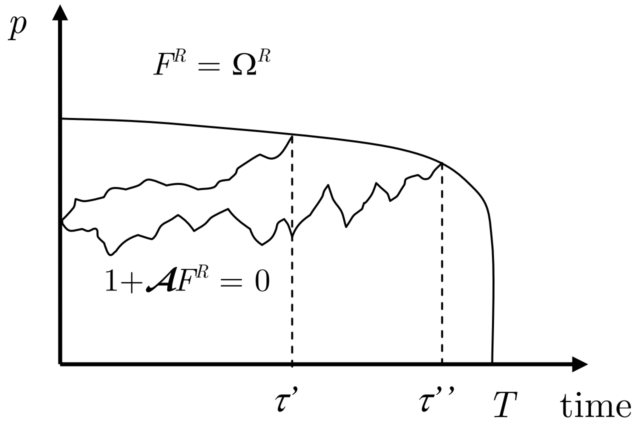

Figure 4: **Electoral Timing for Party R.** The figure schematically depicts the electoral option exercise problem facing party R. It calls an election when the political slant $p(t)$ hits the election barrier $\bar{p}^R(t)$. Two paths for $p(t)$ are shown, with resulting election times τ' and τ''.

schematically illustrates the situation for the ruling party R. By complementary slackness, the government either waits or calls an election, i.e., one of inequalities (a) or (b) is tight.

Assume $t < T$. Let the *optimal stopping barrier* $\bar{p}^R(t)$ be the time-dependent least solution p to $F^R(p,t) = \Omega^R(p)$, and $\bar{p}^L(t)$ the maximum solution p to $F^L(p,t) = \Omega^L(p)$.

Corollary 1 (Optimal Election Barriers) *Assume that the electoral term is not over.*
(a) *Party R delays an election iff $p(t) < \bar{p}^R(t)$ and party L delays iff $p(t) > \bar{p}^L(t)$.*
(b) *The stopping barrier $\bar{p}^R(t)$ falls in time, while the barrier $\bar{p}^L(t)$ rises.*

Proof: Along the barrier, $F^i = \Omega^i$, and thus the time value vanishes: $\Gamma^i \equiv F^i - \Omega^i = 0$. Total differentiation yields $d\Gamma = \Gamma_p d\bar{p}^i(t) + \Gamma^i_t dt = 0$ along the barrier. Also, since the time value is positive away from the barriers and vanishes along each barrier, approaching it horizontally tells us that $\Gamma^i_t < 0$. Next, approaching them vertically yields $\Gamma^R_p(\bar{p}^R(t), t) < 0$ and $\Gamma^L_p(\bar{p}^L(t), t) > 0$. Altogether, we have $d\bar{p}^R(t) < 0 < d\bar{p}^L(t)$. □

4.2 The Expected Time in Power

A. Analytical Results. Stock option values are convex in the underlying price, simply because more risk pushes weight into the exercise tail. This convenient property holds for the election option:

Lemma 6 (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 follows from insights in Bayesian learning (see Easley and Kiefer, 1988). First, because it can be ignored, information must have nonnegative value — raising the expected time in power. But information also causes a mean-preserving spread in beliefs p, since posterior beliefs are riskier than priors (by the law of iterated expectations, for instance). In other words, such spreads must be valuable in expected terms. How can this be? Obviously, by Jensen's inequality, the waiting value is a convex function of 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 $V^R(p)$ that R wins the election called when the political slant equals p.* Clearly, this is the expectation of the step function indicator that $p(t)$ lands above $1/2$ on election day. One might expect that $V^R(p)$ smooths out this step. The Appendix proves just this:

Lemma 7 (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 $V^L(p)$.*

Intuitively, once the expected election day political slant exceeds $1/2$, there is diminishing returns to raising p any further. Recalling the expression from Lemma 2, the Appendix proves that concavity therefore begins when $m(p, \delta) = e^{-a\delta}p + (1 - e^{-a\delta})b \geq 1/2$.

As seen in Figure 4, the government calls an election when the political slant hits the stopping barrier $\bar{p}(\tau)$. Upon winning, it acquires a new waiting value function $F(p, 0+)$, calculated from (3) by taking the limit $t \downarrow 0$. This makes sense because the next election is so far in the future that the current political support should not matter much. Ipso facto, $F(p, 0+) \approx F(0+)$. Indeed, since fast mean reversion pushes all beliefs p within 1% after three years (§3-B), whether one wins big or small has an insignificant impact on the expected time in power; the margin of victory has a second order impact on the expected time in power. One can thus well approximate the stopping value by $\Omega(p) \approx \delta + V(p)F(0+)$. So *governments can act as if they are merely trying to win back a single fixed term of length $F(0+)$.* They can, in effect, behave partially myopically, not looking beyond the next election. Intuitively then, Ω inherits the convex-concave shape of the win chance V, and this is numerically true (as seen in Figure 3).

Figure 3 also reveals that the election barrier hits Ω on its concave portion.

Lemma 8 (Local Concavity) *The stopping value $\Omega(p)$ is locally concave at $\bar{p}(t)$.*

This is analytically established in the Appendix. The key proof ingredient is that just as in Dixit (1993), *smooth pasting* $F_p(p, t) = \Omega_p(p)$ obtains along the stopping barrier — as seen in the numerical simulation in Figure 3. Further, the maximization (3) is equivalent to the solution of the PDE in Proposition 3 with value matching and smooth pasting. We show that any tangency $F_p(p,t) = \Omega_p(p)$ with $\Omega(p)$ locally convex implies a later higher tangency with a greater value function.

Because the waiting value is convex and the stopping value concave in p, the difference of slopes $F^i_p(\bar{p}^i(t),t) - \Omega^i_p(\bar{p}^i(t))$ rises in σ for $t < T$. To restore the smooth pasting optimality condition, the barrier must be hit later. In response, both barriers shift out.

Corollary 2 (Barriers and Risk) *The barrier $\bar{p}^R(t)$ rises and $\bar{p}^L(t)$ falls in volatility σ.*

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 6 and 8. It is thus ambivalent about risk: Neither greater σ nor a longer election period δ (which also raises election risk) has a clear effect on the expected time in power.

B. Analytically-Motivated Numerical Results. We now provide intuitions for the comparative statics that we cannot prove but have strong numerical support: specifically, for the campaign period δ, the mean reversion level b, and the speed a.

• THE CAMPAIGN PERIOD δ. When δ rises by $d\delta$, this lengthens the expected time in power directly by $d\delta$, which lifts both Ω^i and F^i. Also, the support drifts longer, affecting both terms by $F^i_p(p,t)a(b-p)d\delta = \Omega^i_p(p,t)a(b-p)d\delta$, equally so by smooth pasting. The difference is the variance effect. As with σ in Corollary 2, greater δ elevates election period uncertainty, thereby raising F^i relative to Ω^i. So $\bar{p}^R(\tau)$ rises and $\bar{p}^L(\tau)$ falls in δ.

• THE MEAN REVERSION SPEED a. Greater speed a is like smaller σ 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 $\mathcal{A}F^i$ 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 $\mathcal{A}F^R$ 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 nonetheless know whom they should 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.[13] Governments can then learn from polls, because these record voting preferences, summarizing otherwise unavailable information. By Lemma 4, one can view the political slant as the support process for party R. Also, from Proposition 1, equation (1) is the law of motion for the support for party R. We next deal with poll 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 η obeys a t-distribution, with variance $\sigma_\eta^2(\pi) \equiv \pi(1-\pi)/N$, and so is asymptotically normal. Hence, η behaves approximately like $\sigma_\eta \epsilon$, where ϵ 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 9 (Poll Dynamics) *The polling process is approximately:*

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

where $\varepsilon_j \sim N(0,1)$ and $\hat{\sigma}(\pi, N\Delta) > \sigma$ is falling in $N\Delta$, with $\lim_{N\Delta\uparrow\infty} \hat{\sigma}(\pi, N\Delta) = \sigma$.

Lemma 9 (proof in Appendix) exploits the fortuitous fact that the variance $\sigma_\eta^2(\pi)$ shares the nonlinear form $p(1-p)$ of the volatility of $p(t)$. With this addition of noise, we

[13] 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, the 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.

have derived the polling process law of motion. In other words, the polling process (5) consists of discrete time snapshots of a noisier political slant process — the *completely facetious* continuous-time process $\pi(t)$.[14]

Even though the best way to predict election outcomes is to exploit the polling process, 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:[15] A government riding high in the polls indeed believes that its trend is most likely down,[16] with mean reversion.

Lemma 9 implies that the PDE (4), for the poll π with polling sample size N and time difference Δ, is given by $\mathcal{A}F(\pi,t) = F_t + F_\pi a(b-\pi) + \frac{1}{2}F_{\pi\pi}\hat{\sigma}^2(\pi, N\Delta)\pi^2(1-\pi)^2$. Along with Proposition 3, this gives the optimal election barriers in (π,t)-space when we use any constant approximation for $\hat{\sigma}^2(\pi, N\Delta)$, as we estimate below in subsection C.

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.[17] 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 (mori.com) 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.

We first calculate the realized values of π, i.e., the Tory polling support among the B-voters, excluding small parties. Figure 5 depicts the poll levels π from June 1943 – May 2005. The polls on average have favored Labour — the average poll being 0.46.

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 (π for R and $1-\pi$ for L), comparing theoretical and realized political support levels at the times of election calls. 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.

[14]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 $\Delta = 1$.

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

[16]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.

[17]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.

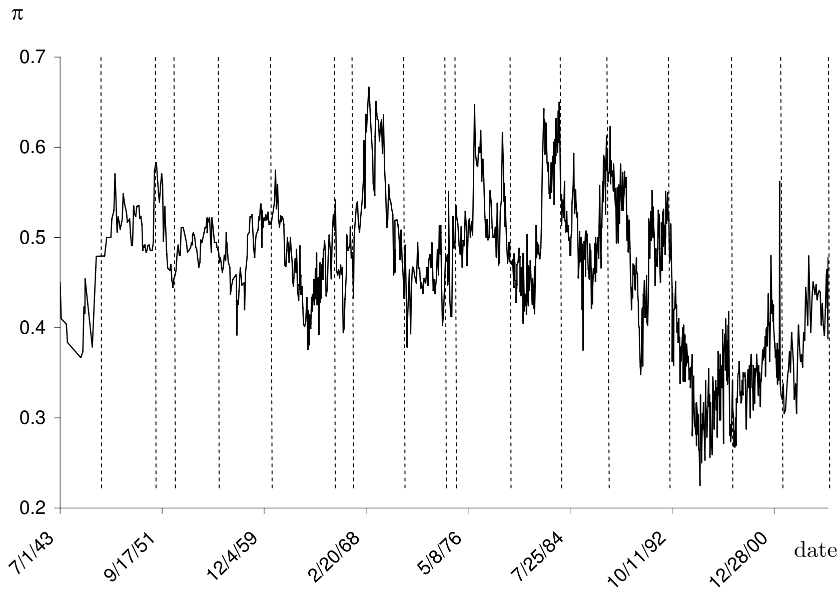

Figure 5: **Proportion of Tory B-voters, 6/1943–5/2005.** The polling process π has averaged 0.46 (i.e. left-leaning), and ranged from 0.23 to 0.67. Elections are the dashed lines.

Equation (5) is an autoregressive model.[18] The parameters a, b, as well as the standard deviations σ, σ_η, 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 $Z_j = -\sqrt{\Delta_j}/[1-\pi(t_j)]$. We estimate parameters a and b in (5) from the regression:[19]

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

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

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

where $\{\epsilon_j\}, \{e_j\}$ are standard normal iidrv's. For simplicity, we assume σ_η is constant.[20]

Table 1 summarizes the estimated process with four parameter sets: overall, outside and inside the δ-period, and inside the δ-period without three no-choice governments that

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

[19]The delta-method (see e.g. Casella and Berger, 2002) gives the standard deviation of b, σ, and σ_η.

[20]According to (8) and our data set, this is justified since about 97% of π 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: **The Estimated Polling Parameters.** 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 "no short" refers to the absence of the no-choice governments.)

δ-period	$R^2: 3.61\%, 3.30\%$			
	a	b	σ	σ_η
estimate	6.43	0.52	0.80	0.11
st. dev.	3.32	0.04	0.66	0.03

δ-period (no short)	$R^2: 3.56\%, 3.06\%$			
	a	b	σ	σ_η
estimate	6.39	0.52	0.82	0.10
st. dev.	3.62	0.05	0.73	0.03

Pre-δ	$R^2: 2.61\%, 26.43\%$			
	a	b	σ	σ_η
estimate	1.49	0.46	0.27	0.12
st. dev.	0.29	0.02	0.10	0.00

Overall	$R^2: 2.12\%, 11.12\%$			
	a	b	σ	σ_η
estimate	1.59	0.47	0.35	0.12
standard deviation	0.33	0.02	0.16	0.01

were called early due to the lack of clear majority (see §6.1).[21]

The parameters are not significantly different outside and inside the δ-period.[22] As the 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 σ and σ_η are 2.23 and 23.40, and thus significant.[23] The average polling time difference outside the δ-period is 0.059 years (about 22 days) and its standard deviation is 0.043 (about 16 days). The numbers inside the δ-period are 0.022 years (about 8 days) and 0.024 (about 9 days).[24] So the average poll volatilities are different inside and outside the δ-period. But this difference owes to the greater poll frequency — i.e. smaller elapse time Δ between polls — inside than outside the δ-period. So we assume a, b, σ, and σ_η constant.

We assume that the ruling party understands that the volatility during the δ-period is not higher than outside it.[25] 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.

[21]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)).

[22]The t-statistic for the test that a coincides outside and inside the δ-period (without no-choice governments) is 1.48 (1.35). The analogous t-statistics for b, σ, and σ_η are 1.34 (1.11), 0.79 (0.75), and -0.33 (-0.67). The joint hypothesis that parameters don't change cannot be rejected at the 1% significance level.

[23]For an internal consistency check, our estimate $\sigma_\eta = 0.12$ in Table 1 is near a direct computation of the standard deviation using our t-distribution formula in (8): $\sigma_\eta^2 = 2/[N\pi(t_j)(1-\pi(t_j))] \approx 1/125$, 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).

[24]We are able to reject the null hypothesis that the average polling time differences are equal inside and outside the δ-period with a 1% level of significance.

[25]The volatility estimates inside and outside the δ-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 §6.2; further, the average ruling periods for L and R in §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.[26] Table 2 lists the outcomes of 17 British elections from 1945–2005. While the delay time ranges from 21–45 days, we set δ 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.[27]

[26] 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 §4.2.

[27] The 1979 election was also forced by losing a nonconfidence vote (by just one 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 (π, t) columns lists the poll level π 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. Our theory cannot possibly explain such 'no-choice' elections.

Election date	Announced	Winner	Tory	Labour	(π, t)
7/5/45	5/23/45	Labour	39.8 (33.3)	47.8 (61.4)	
2/23/50	1/11/50	Labour	43.5 (47.8)	46.1 (50.4)	(0.52,4.52)
10/25/51*	9/19/51	Tory	48.0 (51.3)	48.8 (47.2)	(0.57,1.57)
5/26/55	4/15/55	Tory	49.7 (54.8)	46.4 (44.0)	(0.51,3.47)
10/8/59	9/8/59	Tory	49.4 (57.9)	43.8 (41.0)	(0.54,4.29)
10/15/64	9/15/64	Labour	43.4 (48.3)	44.1 (50.3)	(0.49,4.94)
3/31/66*	2/28/66	Labour	41.9 (40.2)	47.9 (57.6)	(0.46,1.37)
6/18/70	5/18/70	Tory	46.4 (52.4)	43.0 (45.6)	(0.46,4.13)
2/28/74	2/7/74	Labour	37.9 (46.8)	37.1 (47.4)	(0.48,3.64)
10/10/74*	9/18/74	Labour	35.8 (43.6)	39.2 (50.2)	(0.44,0.55)
5/3/79*	3/29/79	Tory	43.9 (53.4)	36.9 (42.4)	(0.54,4.47)
6/9/83	5/9/83	Tory	42.4 (61.1)	27.6 (32.2)	(0.61,4.02)
6/11/87	5/11/87	Tory	42.2 (57.9)	30.8 (35.2)	(0.58,3.92)
4/9/92	3/11/92	Tory	41.9 (51.6)	34.4 (41.6)	(0.49,4.75)
5/1/97	3/17/97	Labour	30.7 (25.0)	43.2 (63.4)	(0.37,4.94)
6/7/01	5/8/01	Labour	31.7 (25.2)	40.7 (62.5)	(0.36,4.02)
5/5/05	4/5/05	Labour	32.3 (30.5)	35.2 (55.0)	(0.46,3.83)

6.2 Election Timing

Next we analyze election timing for the overall parameters in Table 1. As Proposition 3 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 π space, we analyze the election times by comparing polls π 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, π for Tories and $1-\pi$ for Labour. As seen in Figure 6 and proved in Corollary 1, 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

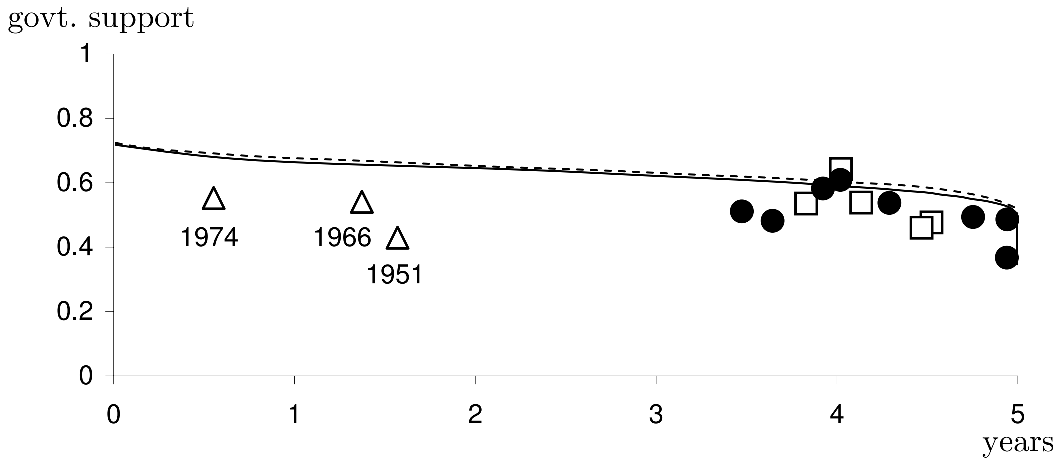

Figure 6: **Election Barriers for Tories (solid line) and Labour (dashed line).** According to our model a ruling party should call election when the support hits its barrier. Circles are support levels when the Tories called elections. Triangles are poll levels when Labour government did not have clear majority. Squares are poll levels corresponding to other Labour election times. On average, the barriers for L and R are 0.64 and 0.63, respectively.

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.[28]

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 cannot possibly take into account. Likewise, Blair should have called the election of 2001 eight months earlier.

The above discussion might suggest comparing 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. Comparing the indicators by using all sixteen elections up to their announcement times, we find that the indicators coincide

[28]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.

92.4% of the time. 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 simple measure, 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 equals 1), and analyze how well our model explains the those support levels.

The regression models 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 at the election announcement time. This analysis is motivated by cross-sectional asset pricing tests (see e.g. Cochrane, 1996). No-choice governments aside, we find that $Y = -0.49 + 1.76X$. This intercept is insignificant, and the slope significant. Since $R^2 = 43\%$, our model clearly explains a large portion in the variation of election times through the regression. Put differently, the correlation of the theoretical and realized support levels at the election times is 0.65. At the very least, *we have correctly identified the polls π and elapse time t as important decision factors*[29] (and below we consider a completely naive linear approach using π, t).

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 of a unit slope, since the t-statistic is 3.58.[30]

The slope test is obviously a joint test on the model and its parameters. To ensure a unified paper built on our fleshed-out theory, we have consistently employed an extremely 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.[31]

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

[29]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.

[30]We would likewise reject an extremely good model with slope, say, 1.01 and $R^2 = 100\%$, for then the slope would have zero standard deviation. In other words, we reject the slope test partly due to our high R^2.

[31]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 §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.

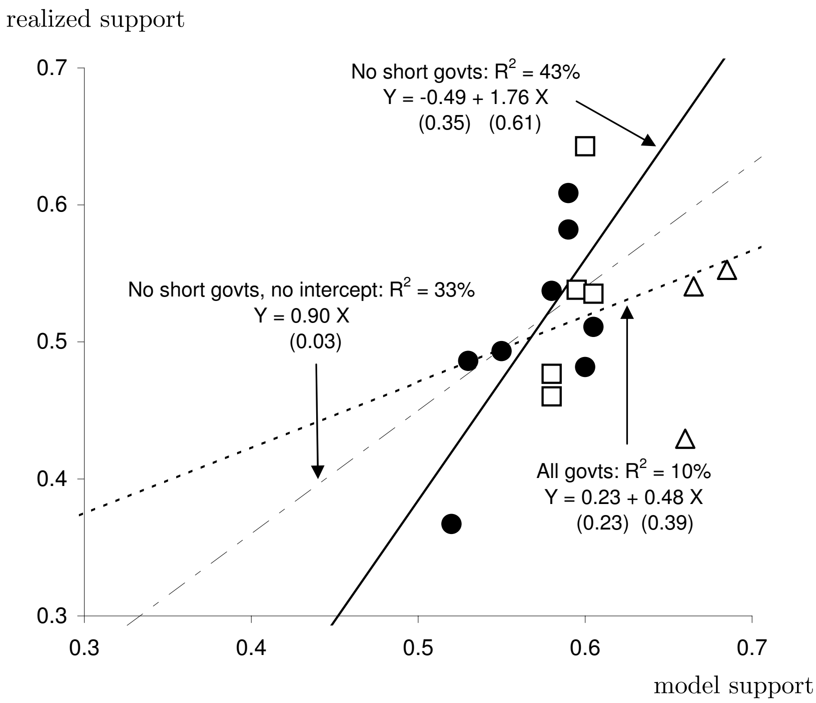

Figure 7: **Actual and Theoretical Poll Support at Election Calls.** The triangles are the elections that Labour called early due to slim majority or minority governments (10/25/51, 3/31/66, and 10/10/74). The squares are all other Labour election announcement support levels. The circles are the election announcement Tory support levels. In the regressions Y = actual support level, X = 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.

party. Do these effects matter? The coefficients were insignificant: All t-statistics were less than 0.9, $R^2 = 9\%$ with the intercept, and $R^2 = 13\%$ with no intercept. So 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 finally, *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.

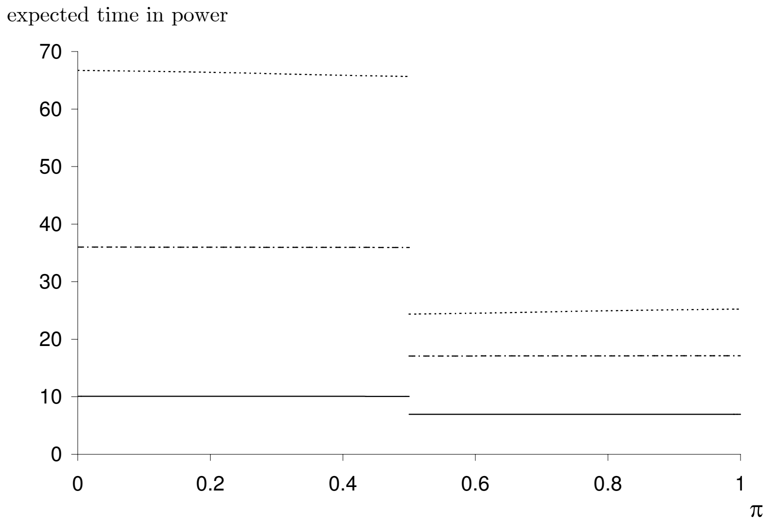

Figure 8: **Expected Time in Power Given Initial Support Levels.** The top graph now depicts the expected time in power for Labour (at left, for $\pi < 0.5$) and Tory (at right, for $\pi > 0.5$), as the initial poll varies. The middle graphs assume that the election cannot be called within three years, as is protocol in the U.K. for majority governments. The bottom graphs finally assume that elections are called every four years. The value of this timing option is seen to be substantial.

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 π (i.e. the last election outcome). Integrating these expected times over the long-run polling density in Lemma 3 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. In other words, we use the process parameters in Table 1 with $\sigma_\eta = 0$, as there is no polling error with p.

The predicted times are quite high, by historical standards[32] — 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.[33] We thus reformulated our timing exercise, asking

[32]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.

[33]For instance, Blais *et al* (2004) argue that voters punished Jean Chretien for calling a snap election in

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.

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

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.[34]

Secondly, we have discovered that the polls π 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.[35] This period 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.*

One may also be interested in a simple welfare analysis: Does endogenous timing on average help or hurt the voters? We can intuitively conclude the former, since the political

November 2000 after just three years and four months.

[34]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 $R^2 = 39\%$.

[35]Specifically, $a = 2.2, b = 0.49, \sigma = 0.19, \sigma_\eta = 0.11$, constant volatility $\hat{\sigma} = 0.69$. That is, the π process mean reverts faster about a higher mean, with less volatility.

slant process is Markovian and independent of the ruling party and its actions. Consider two cases: an early or a late election. The ruling party optimally calls the election early only if its support is high. This choice is welfare-neutral for the voters, since the best party is already in power. On the other hand, the ruling party might not call an election early if its support is low. Here, the wrong party stays in power, and the voters are hurt.

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 asking whether governments might simply be solving for 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 party in power holds sequential finite time horizon American options and because there is a delay in the exercise decisions. Each difficulty 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. We derived the comparative statics by indirect means, without recourse to closed forms. The optimization was done by numerical means, as happens for finite time horizon American put options.

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 tractability owes to some simplifying assumptions.[36] 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. Obviously, no maximizing theory can be expected to explain something that is not a choice.

A minority government is less desirable for the winner, since it must compromise its preferred governing choices to satisfy its coalition partners.[37] 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.[38] We are very upfront about this weakness, but it would clearly entail writing a different paper. A more involved model might assume higher flow payoffs to governing with a strong than a weak majority or minority government, and lead to early elections called with minority of weak governments.

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,[39] and is a justified simplification with little impact.

Third, our objective function of the decision makers is quite straightforward, positing that governments only maximize their expected time in power. We have crucially rendered this a decision-theoretic exercise, assuming that the government is unable to affect the course of the political slant. Namely, it takes the polling process as given. Like all other simplifications, we do not in any way claim their irrelevance; however, they are best studied elsewhere (eg. 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 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

[36] 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)). 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 behavior by the government, and not voters.

[37] 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).

[38] 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.

[39] 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.

prolong his time in power, delaying the election past its best date. Discounting — a year now is worth more than one later — may seem reasonable, but we have no prior reason to choose any particular discount factor. Moreover, any impatience delays elections, because the utility from the future terms falls. In light of Figure 6, this worsens the model fit.

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 π. 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

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

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

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

where we suppress the σ^2 argument whenever clear. Thus, we have the partial derivatives:

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

We find $v_t(0) > 0$ and $v_{\sigma^2}(0) > 0$ for all $\sigma^2 > 0$ and $t \geq 0$. Now $v_{\sigma^2}(t) = v_{\sigma^2}(0) + \int_0^t v_{\sigma^2 t}(y)dy$. If $v_{\sigma^2}(t) = 0$ for some $t > 0$, then $v_{\sigma^2 t}(t) > 0$. 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 σ.

A.2 Proof of Lemma 3: 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 $\frac{1}{2}[\sigma(p)\psi(p)]'' - [\mu(p)\psi(p)]' = 0$. 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:

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

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 similarly analyzed using $\bar{p}(t) = 1 - p(t)$ and noting $d\bar{p}(t) = a\left((1-b) - \bar{p}(t)\right)dt - \sigma\bar{p}(t)(1-\bar{p}(t))dW(t)$ 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 $\lim_{y\downarrow 0}\int_y^p s(z)dz = \infty$ 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 Monotone Values

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

Put $F_0^i = 0$ 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 $\Upsilon : F^i \mapsto \Omega^i$ and $\Phi : \Omega^i \mapsto F^i$, the iterations obey $0 \le \Omega_0^i \le \Omega_1^i \le \Omega_2^i \le \cdots$ and $0 \le F_0^i \le F_1^i \le F_2^i \le \cdots$. Their limits exist, and obey (2)–(3).

Furthermore, each of the maps Υ, Φ preserve monotonicity in p. To see this of Φ, use the stopping time τ' 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, Υ 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 ψ in p. As F^R is boundedly finite, Ω^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 7: The Shape of the Win Chance

If $p \ge \frac{1}{2}e^{a\delta} + (1 - e^{a\delta})b$, then the expected election outcome at the end of the delay period is $m(p, \delta) > \frac{1}{2}$. Now, by Lemma 1, increasing σ 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 σ for all $p \ge \frac{1}{2}e^{a\delta} + (1 - e^{a\delta})b$, and so must be (locally) concave. Likewise, changing variables $p \leftarrow 1 - p$, we find that $V^L(p)$ is concave for all $p \le \frac{1}{2}e^{a\delta} + (1 - e^{a\delta})b$. But $V^R(p) \equiv 1 - V^L(p)$. Altogether, we deduce the convexity of $V^R(p)$ for all $p \le \frac{1}{2}e^{a\delta} + (1 - e^{a\delta})b$ and $V^L(p)$ for all $p \ge \frac{1}{2}e^{a\delta} + (1 - e^{a\delta})b$. □

A.5 Proof of Lemma 8: Local Concavity

Claim 2 (Convexity at the Barrier) $F_{pp}^i(\bar{p}^i, t) \ge \Omega_{pp}^i(\bar{p}^i)$ *for all* $t < T$.

Proof: We consider only R. From Lemma 6 and Proposition 3, $F^R(p, t) - \Omega^R(p) \ge 0$ and $F_p^R(\bar{p}^R, t) - \Omega_p^R(\bar{p}^R) = 0$. Clearly, $F^R(\bar{p}^R, t) - \Omega^R(\bar{p}^R)$ is convex. □

Let p_l be the optimal barrier at t, so that $\bar{p}^R(t) = p_l$. By Proposition 3, it obeys the value matching and smooth pasting optimality conditions $\Omega^R(p_l) = F^R(p_l, t)$ and

$\Omega_p^R(p_l) = F_p^R(p_l,t) \geq 0$, where $\bar{p}^R(t) = p_l$. For a contradiction, assume that $\Omega^R(p)$ is locally convex at $p = p_l$. Because $\Omega_p^R(p) = \Omega_p^R(0) + \int_0^p \Omega_{pp}^R(y)dy$, if there exists $p_h > p_l$ with $\Omega_p^R(p_h) = \Omega_p^R(p_l)$, then Ω^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 Ω^R, and define $\hat{F}^R$ by value matching and smooth pasting at p_h, and the PDE in Proposition 3 — which is equivalent to the maximization (3). Since $\hat{F}_p^R(p,t) < \hat{F}_p^R(p_h,t) = \Omega_p^R(p_h) \leq \Omega_p^R(p)$ on $[p_l, p_h]$, we have

$$\hat{F}^R(p_l,t) = \hat{F}^R(p_h,t) - \int_{p_l}^{p_h} \hat{F}_p^R(p,t) > \Omega^R(p_h) - \int_{p_l}^{p_h} \Omega_p^R(p) = \Omega(p_l)$$

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

A.6 Proof of Lemma 9: 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 $\eta_{j+1} - \eta_j = \int_{t_j}^{t_{j+1}} \varphi(t)dW^\eta(t)$ and where W^η is independent of W. Hence, for all $j \in \{1, 2, \ldots, n\}$:

$$\pi_{j+1} - \pi_j = \int_{t_j}^{t_{j+1}} a(b - p(t))dt + \int_{t_j}^{t_{j+1}} \sigma p(t)(1-p(t))dW(t) + \int_{t_j}^{t_{j+1}} \varphi(t)dW^\eta(t) \quad (7)$$

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

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

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

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

$$\begin{aligned} & a\,(b - p_j)\,\Delta_j + \sigma p_j(1-p_j)\sqrt{\Delta_j}e_j + \tfrac{1+o(1)}{\sqrt{N}}\sqrt{\pi_j(1-\pi_j)k_j}\xi_j \\ = \; & a\,(b - \pi_j + \eta_j)\,\Delta_j + \sigma(\pi_j - \eta_j)(1 - \pi_j + \eta_j)\sqrt{\Delta_j}e_j + \tfrac{1+o(1)}{\sqrt{N}}\sqrt{\pi_j(1-\pi_j)k_j}\xi_j, \end{aligned}$$

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

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

where the two last terms of the variance owe to the independence of e_j and ξ_j. If N is high and Δ_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 ε_j is a standard normal variable and

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

Thus, $\lim_{N\Delta\downarrow 0}\hat{\sigma}(\pi, N\Delta) = \infty$ and $\lim_{N\Delta\uparrow\infty}\hat{\sigma}(\pi, N\Delta) = \sigma$. $\square$

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 π values and the discrete time period Δt. These define the grid in (π, t) space $(0,1)\times[0,5]$. Calculate the transition matrix M of π over the discrete time from (5) and Table 1. Set $F_i^0(0) = [50\cdots 50]^T$, where $F_i^j(t)$ is the j'th value function (column vector) for different π 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: $F_i^{j+1}(5) = M_\delta F_i^j(0)$, where $M_\delta = M^n$ is the transition matrix of π for δ-period, $M_{i,j}^n = \sum_{k=1}^{\infty} M_{i,k}^r M_{k,j}^s$ 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 $F_i^{j+1}(5)$ is also the value function if the election is called before the end of the period, i.e., it models Ω^i.

Step 2 (The Value Function). For each $n \in \{1,\ldots,5/\Delta t\}$, calculate first the waiting value $\hat{F}_i^{j+1}(5-n\Delta t) = MF_i^{j+1}(5-(n-1)\Delta t)$, and then check the 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 $z \in \{1,\ldots,1+1/\Delta\pi\}$, where $[F]_z$ is the z'th element of F.

Step 3 (Convergence Test). If

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

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

By §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:

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

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

Alvarez, Luis H.R. and Jussi Keppo (2002). "The impact on delivery lags on irreversible investment under uncertainty", *European Journal of Operational Research*, **136**: 173–180.

Alvarez, Luis H.R. and Larry A. Shepp (1998). "Optimal harvesting of stochastically fluctuating populations", *Journal of Mathematical Biology*, **37**: 155–177.

Austen-Smith, David and Jeffrey S. Banks (1988). "Elections, coalitions, and legislative outcomes", *American Political Science Review*, **82**: 405–422.

Balke, Nathan S. (1990). The Rational Timing of Parliamentary Elections, *Public Choice*, **65**: 201–216.

Blais, Andre, Elisabeth Gidengil, Neil Nevitte and Richard Nadeau (2004). "Do (Some) Canadian Voters Punish a Prime Minister for Calling a Snap Election?" *Political Studies*, **52**: 307–323.

Bolton, Patrick and Chris Harris (1999). "Strategic Experimentation", *Econometrica*, **67**: 349–74.

Campbell, John and Samuel Thompson (2005). "Predicting the Equity Premium out of Sample: Can Anything Beat the Historical Average?", Harvard mimeo.

Casella, George and Roger L. Berger (2002). Statistical Inference. 2nd edition, Duxbury, Pacific Grove, California.

Chowdhury, Abdur R. (1993). “Political surfing over economic waves: parliamentary election timing in India”, *American Journal of Political Science*, **37**: 1100–18.

Cochrane, John R. (1996). “A Cross-Sectional Test of an Investment-Based Asset Pricing Model”, *Journal of Political Economy*, **104**: 572–621.

Davydov, Dmitry and Lones Smith, “Optimal Electoral Timing in a Parliamentary Democracy”, PhD thesis, Dmitry Davydov.

Diermeier, Daniel and Antonio Merlo (2000). “Government Turnover in Parliamentary Democracies”, *Journal of Economic Theory*, **94**: 46–79.

Dixit, Avinash (1993). The Art of Smooth Pasting. Fundamentals of Pure and Applied Economics 55. Harwood Academic Publishers. Chur, Switzerland.

Dixit, Avinash and Robert S. Pindyck (1994). Investment under Uncertainty. Princeton University Press. Princeton, NJ.

Duan, Jin-Chuan and Jean-Guy Simonato (2001). “American option pricing under GARCH by a Markov chain approximation”, *Journal of Economic Dynamics and Control*, **25**: 1689–718.

Dybvig, Philip H. and Mark Loewenstein (2003). “Employee Reload Options: Pricing, Hedging, and optimal Exercise”, *Review of Financial Studies*, **16**: 145–171.

Easley, David and Nicholas M. Kiefer (1988). “Controlling a stochastic process with unknown parameters”, *Econometrica* **56**: 1045–64.

Ellis, Christopher and Mark Thoma (1991). “Partisan Effects in Economies with Variable Electoral Terms”, *Journal of Money, Credit, and Banking*, **23**: 728–41.

Feddersen, Timothy J., and Wolfgang Pesendorfer (1996). “The Swing Voter’s Curse”, *American Economic Review*, **86**: 408–24.

Hurst, Erik and Frank Stafford (2004). “Home is Where the Equity Is: Liquidity Constraints, Refinancing and Consumption”, *Journal of Money, Credit and Banking*, **36**: 985–1014.

Karlin, Samuel and Howard M. Taylor (1981). *A Second Course in Stochastic Processes.* Academic Press, New York, NY.

Kayser, Mark Andreas (2005a). “Who Surfs, Who Manipulates? The Determinants of Opportunistic Election Timing and Electorally Motivated Economic Intervention”, *American Political Science Review*, **99**: 17–28.

Kayser, Mark Andreas (2005b). “Trade and the Timing of Elections”, *British Journal of Political Science*, forthcoming.

Keller, Godfrey and Sven Rady (1999). “Optimal Experimentation in a Changing Environment”, *The Review of Economic Studies*, **66** (3) p. 475–507.

Kou, Steven G. and Michael E. Sobol (2004). “Forecasting the Vote: A Theoretical Comparison of Election Markets and Public Opinion Polls”, *Political Analysis*, **12**: 277–295.

Lesmono, Julius Dharma , Elliot J. Tonkes, and Kevin Burrage (2003). “An Early Political Election Problem”, *ANZIAM Journal*, **45**: C16–C33.

Liptser, Robert C. and Albert N. Shiryaev (2001). Statistics of Random Processes. 2nd edition, Springer-Verlag, New York, NY.

Moscarini, Giuseppe and Lones Smith (2001). “The Optimal Level of Experimentation”, *Econometrica*, **69** (6), 1629–44.

Palmer, Harvey (2000). “Government competence, economic performance and endogenous election dates”, *Electoral Studies*, **19** (2-3): 413–426.

Peura, Samu and Jussi Keppo (2005). “Optimal bank capital with costly recapitalization”, *Journal of Business*, forthcoming.

Sanders, David (2003). “Pre-election polling in Britain, 1950-1997”, *Electoral Studies*, **22**: 1-20.

Seydel, Rüdiger (2003). Tools for Computational Finance. Springer-Verlag, New York, NY.

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

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

Smith, Lones (1997), “Time-consistent optimal stopping”, *Economics Letters*, 1997, **56** (3), 277-79.