The Dissertation Committee for Urmee Khan
certifies that this is the approved version of the following dissertation:

Essays in Economic Design: Information, Markets and Dynamics

Committee:

Maxwell B. Stinchcombe, Supervisor

Takashi Hayashi, Supervisor

Marcin Peski

Robert P. Lieli

Gordan Zitkovic

Essays in Economic Design: Information, Markets and Dynamics

by

Urmee Khan, B.A.; M.A.; M. S.; M.S.; M.A.

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

May 2011

To my parents, Ayesha and Mortuza.

Acknowledgments

My deepest gratitude is to Maxwell B Stinchcombe for his invaluable guidance, infinite patience and unwavering support. Every page bears testimony to my immense intellectual debt to him. Thanks are also in order for Takashi Hayashi and Marcin Peski for their generosity with time and advice. Robert Lieli's contribution as a co-author to one of the essays is deeply appreciated. Of course none of it would be possible without the emotional support from my family and friends.

Essays in Economic Design: Information, Markets and Dynamics

Publication No. ____________________

Urmee Khan, Ph.D.
The University of Texas at Austin, 2011

Supervisors: Maxwell B. Stinchcombe
Takashi Hayashi

This dissertation consists of three essays that apply both economic theory and econometric methods to understand design and dynamics of institutions. In particular, it studies how institutions aggregate information and deal with uncertainty and attempts to derive implications for optimal institution design. Here is a brief summary of the essays.

In many economic, political and social situations where the environment changes in a random fashion necessitating costly action we face a choice of both the timing of the action as well as choosing the optimal action. In particular, if the stochastic environment possesses the property that the next environmental change becomes either more or less likely as more time passes since the last change (in other words the hazard rate of environmental change is not constant over time), then the timing of the action takes on special importance. In the first essay, joint with Maxwell B Stinchcombe, we model and solve a dynamic decision problem in a semi-Markov envi-

ronment. We find that if the arrival times for state changes do not follow a memoryless process, time since the last observed change of state, in addition to the current state, becomes a crucial variable in the decision. We characterize the optimal policy and the optimal timing of executing that policy in the differentiable case by a set of first order conditions of a relatively simple form. They show that both in the case of increasing and decreasing hazard rates, the optimal response may be to wait before executing a policy change. The intuitive explanation of the result has to do with the fact that waiting reveals information about the likelihood of the next change occurring, hence waiting is valuable when actions are costly. This result helps shed new light on the structure of optimal decisions in many interesting problems of institution design, including the fact that constitutions often have built-in delay mechanisms to slow the pace of legislative change. Our model results could be used to characterize optimal timing rules for constitutional amendments. The paper also contributes to generalize the methodology of semi-Markov decision theory by formulating a dynamic programming set-up that looks to solve the timing-of-action problem whereas the existing literature looks to optimize over a much more limited set of policies where the action can only be taken at the instant when the state changes.

In the second essay, we extend our research to situations, where the current choice of action influences the future path of the stochastic process, and apply it to the legal framework surrounding environmental issues, particularly to the ‘Precautionary Principle’ as applied to climate change legislation. We represent scientific uncertainty about environmental degradation using the concept of ’ambiguity’ and show that ambiguity aversion generates a ’precautionary effect’. As a result, justification is provided for the Precautionary Principle that is different from the ones provided by expected utility theory. This essay serves both as an application of the general theoretical results derived in the first essay and also stands alone as an analysis of a substantive question about environmental law.

Prediction markets have attracted public attention in recent years for making accurate predictions about election outcomes, product sales, film box office and myriad other variables of interest and many believe that they will soon become a very important decision support system in a wide variety of areas including governance, law and industry. For successful design of these markets, a thorough understanding of the theoretical and empirical foundations of such markets is necessary. But the information aggregation process in these markets is not fully understood yet. There remains a number of open questions.

The third essay, joint with Robert Lieli, attempts to analyze the direction and timing of information flow between prices, polls, and media coverage of events traded on prediction markets. Specifically, we examine the race between Barack Obama and Hillary Clinton in the 2008 Democratic primaries for presidential nomination. Substantively, we ask the following question: (i) Do prediction market prices have information that is not reflected in contemporaneous polls and media stories? (ii) Conversely, do prices react to information that appears to be news for pollsters or is prominently featured by the media? Quantitatively, we construct time series variables that reflect the "pollster's surprise" in each primary election, measured as the difference between actual vote share and vote share predicted by the latest poll before the primary, as well as indices that describe the extent of media coverage received by the candidates. We carry out Granger Causality tests between the day-to-day percentage change in the price of the "Obama wins nomination" security and these information variables. Some key results from our exercise can be summarized as follows. There seems to be mutual (two-way) Granger causality between prediction market prices and the surprise element in the primaries. There is also evidence of one-way Granger causality in the short run from price changes towards media news indices. These results suggest that prediction market prices anticipate at least some of the discrepancy between the actual outcome and the latest round of polls before the election. Nevertheless, prices also seem to be *driven* partly by election results, suggesting that there is an element of the pollster's surprise that is genuine news for the market as well.

Table of Contents

List of Tables

Chapter 1

The Virtues of Hesitation

[1]

1.1 Introduction

In many social, economic and political situations, there is a stochastic environment that changes at random points in time and changes in action are costly. Actions are costly, and we know the current state may give way to another new state at some random time in the future, potentially making today's optimal action again obsolete. The question is whether to take an action in response to a change in the environment or to delay any change (or changes).

Variants of this problem have been extensively analyzed in economics (for example Boyarchenko and Levendorski [2007], Stokey [2009] and the references therein). However, a crucial aspect of most existing analyses is that

[1]This essay is drawn from joint work with Maxwell B Stinchcombe. My contribution to this project include formulating the original research question, conducting the literature review, developing the theoretical model and construction of various proofs. Prof Stinchcombe collaborated on developing the model and some of the proofs.

the passage of time by itself does not reveal any information. By contrast, we study problems in which the passage of time without a change contains information about the arrival time of the next change. In such problems, there may be value to delaying decisions beyond the usual option value of waiting. We begin with examples where the time which a change has survived may be of crucial importance to its future longevity. We begin with some examples.

1.1.1 Political Change

Political process in a democratic system are driven by 'political issues' and the configuration of opinions and attitudes of the polity on these issues. Such configurations are hardly, if ever, static. There are slow and gradual changes that take place side by side with rapid and explosive changes. Some changes are long-lasting, some short-lived. As Carmines and Stimson [1990] say:

> … we shall see that issues, like species, can evolve to fit new niches as old ones disappear. But, unless they evolve to new forms, all issues are temporary. Most vanish at their birth. Some have the same duration as the wars, recessions, and scandals that created them. Some become highly associated with other similar issues or with the part system and thereby lose their independent impact. And some last so long as to reconstruct the political system that produced them … .

Vietnam War and the Watergate scandal seem to have very little traces left today either in public attitude or legislative response to the issues of war and executive power respectively. But they were the biggest issues of their day. On the other hand, the Civil Rights Movement and its aftermath marked a fundamental realignment in US politics. In general, some ideas and opinions "wear out their welcome" after a time, perhaps through changes in the conditions that gave rise to them, perhaps by the accumulation of counterarguments to their veracity. Hence, the likelihood that such an idea would become irrelevant increases with time. By contrast, some types of issues or opinions tend to get more entrenched the longer they live. Political actors in various capacities try to cope and make decisions in the face of such 'issue evolution' [Carmines and Stimson, 1990]. Legislatures choose whether or not to change a law, Supreme Courts decides whether or not to re-interpret or overturn past precedents, political parties decide whether or not to realign politically and redefine the agenda. Oftentimes, the most crucial ingredient in such decisions is the aspect of timing.

Each of these decisions entail some fixed cost either to the society at large or the actor herself.[2] One would have to trade off the immediate gains with substantial future losses if the initial change that triggered the costly action turns to be rather short-lived.

[2]In case of legislative changes and court decisions, the citizens have to re-adjust and re-optimize with respect to the new rules. In case of a political party, realignment may mean losing a traditional support base.

1.1.2 Constitutional Amendments

Constitutions establish the fundamental legal structures of a society. They are meta-institutions through which institutions are introduced, reformed and interpreted [Ostrom, 1990]. A constitution and the legal order it creates must have the support of, or at least tacit approval of, the governed to have legitimacy. Maintaining the legitimacy and relevance of a constitution require a certain degree of adaptability or flexibility to change because technology, environment and public opinion are forever changing. On the other hand, the basic value of a constitution lies in its stability because it coordinates the actions and expectations of people and reduces the uncertainty in the environment [Hardin, 2003]. Hence the basic tradeoff between 'commitment' and 'flexibility' lies at the heart of the constitution design problem, as encapsulated in the famous exchange between Thomas Jefferson and James Madison (Smith [1995]; Madison [1961]).[3]. It is also costly to change the constitution because it acts as a coordination device for peoples' behavior, and changes are likely to impose large adjustment costs on significant parts of the population [Hardin, 2003] and disrupt ancillary institutions that grow around the constitution.

From these perspectives, it is reasonable to presume that an optimal rule for constitutional change should be more sensitive to long-lasting changes than to transitory changes. It is clear that waiting longer will help answer

[3]for interesting empirical evidence to the effect that flexibility actually helps the sustainability of constitutions, see Elkins et al. [2009]

whether a change will have a longer or shorter total life, but what matters for decisions is the longer or shorter *future* life of the change. One tradeoff is between costly unneeded or ultimately unwanted changes (e.g. Prohibition) and undermining the legitimacy of the constitutional regime by ignoring new realities. It is from this perspective that we study the general question of why some changes in laws should be more difficult to implement, and what this should depend on. Under study is a class of explanations that we regard as complementary to the many previously offered ones, a class of explanations based on the observation that the persistence of changes in sentiment have predictive power for the future length of time the changes will last. For us the question becomes "How much longer should one wait before acting?"

The US constitution has had four different amendments that have extended voting rights to different parts of the population: Amendment XV (1870), which was passed at the end of the Civil War, extended suffrage to men independent of race or previous condition of servitude; XIX (1920) extended suffrage to women; XXIV (1964) made poll taxes illegal; and XXVI (1971) extended suffrage to those eighteen years of age or older. These formalized long-lived widely-shared changes in sentiment, but Amendment XVIII, Prohibition in 1919, was an expensive and short-lived failure, being repealed fourteen years later by Amendment XXI (1933).[4]

[4] For a detailed history of all amendments, see Amar [2006].

If one dates the beginning of the women's suffrage movement to the 1848 Seneca Falls Convention,[5] it took 72 years, until 1920, for the 19'th Amendment to pass. At various points in the political process, there was evidence that the recognition of women's rights to vote would be long-lasting: the passage of suffrage at the state level in western states by the early 20'th century;[6] the nation's westward expansion and the Civil War led to an expanded need for women both in industrial settings and as teachers; the slow increase in the numbers of college educated and professional women; unionization movements among female professions in the late 1800's and early 1900's. Even after one could perhaps clearly see that general sentiment had shifted in favor of the Nineteenth Amendment, there was (much) further delay in implementing what turns out to have been a long-lasting change in sentiment, perhaps consistent with unwillingness to believe that so drastic a change could be long-lasting.

By contrast, Amendment XVIII (Prohibition, 1919) proved to be very costly to society, and was short-lived, repealed fourteen years later.[7] The Temperance Movement had as long a history as the women's suffrage movement, and was even used by some women's suffrage organizers as an occasion to

[5]Flexner and Fitzpatrick [1996] emphasize the experience of female abolitionists and fighters for women's education in the early 19'th century as the roots of the suffrage movement.

[6]By 1915, Arizona, California, Colorado, Idaho, Illinois, Kansas, Montana, Nevada, Oregan, Utah, Washington, and Wyoming had granted full women's suffrage, and several other states or municipalities had granted suffrage in primary elections.

[7]Prohibition was repealed by the only Amendment to be passed by state ratifying conventions rather than by state votes.

teach women the necessity of having a voice in politics in order to achieve changes (Flexner and Fitzpatrick [1996]). From our point of view, this is a change of action that led to a change in the distribution of the time until general sentiment was reversed. This is an example of a more complicated scenario where one is not only wondering about how long the current state would last, but also has to consider the fact that her choice of action might actually affect the timing and nature of the next change.

1.1.3 Marketing Strategy

Research in consumer behavior has shown that when and how consumers switch brands depend on the last purchased brand and time since the last purchase. The inter-purchase time may exhibit increasing or decreasing hazard rates depending on the consumer characteristics named "inertia" or "variety seeking," and these change over time since the last purchase [Chintagunta, 1998]. It has been suggested that optimal timing of targeting consumers for marketing should depend on such considerations instead of the traditional demographic variables (Chintagunta [1998], see also Gonul and Ter Hofstede [2006] for an empirical approach to optimal timing for catalog mailing). The class of optimization models under study here are directly applicable to such situations.

1.1.4 Outline

The next section contains two simple examples that give a sense of what is involved in the more general analyses that follow. The essential aspects of the model include: a starting state and action, s_0 and a_0; random times Y_k at which the state changes from S_{k-1} to S_k according to an imbedded Markov process; and the option to engage in costly actions changes during the interarrival times $W_k = Y_{k+1} - Y_k$.

The within interval problems, from Y_k to Y_{k+1} will be central to the analysis, and the first example, on optimal search duration, highlights the role of the hazard rates for the W_k. The second example demonstrates how the value functions for the problem interact with the within interval problems.

The general model, existence of optima, and their recursive characterization through the value function follow. The following section develops the corresponding first order conditions (Euler equations) for a broad range of problems. The last section concludes.

1.2 Two Examples

A pair of non-stationary problems demonstrate the essential features of both the optimization problems under study and of their solutions. The first problem is about the determinants of optimal search duration and highlights the role of changing hazard rates in both first and second order conditions for an optimum. The second problem is about optimal adaptation to changing circumstances and highlights the role of stochastic intervals.

1.2.1 Optimal Search Duration

At a flow cost of $c > 0$, one can keep searching for a source of higher profits (a low cost source of a crucial input, a process breakthrough, a new product). If found, expected net profits of $\overline{\pi}$ result. If one abandons the search, the alternative yields expected net profits of $\underline{\pi}$, $\overline{\pi} > \underline{\pi} > 0$. Let W denote the waiting time till the source is found. We will assume throughout that waiting times have densities on $[0, \infty)$, hence having no atoms, except perhaps at ∞. If W has an atom at ∞, it is called an **incomplete** distribution, which corresponds, in the present search problem, to the object of search not existing or not being findable.

Since one optimally searches in the more likely locations or ideas first, we expect the arrival rate of W to be decreasing over time. The non-constancy of the hazard rate makes the problem non-stationary. The non-stationary choice problem is at what time, t_1, does one stop searching and accept the lower $\underline{\pi}$? The results are special cases of Theorem 4 (below), but we give both an intuitive and a more formal development of the first order and the second order conditions for an optimal $0 < t_1^* < \infty$ for this problem.

• First order conditions: the expected benefits of waiting an extra instant dt at t_1 are $(\overline{\pi} - \underline{\pi})h_W(t_1)$, the expected costs are $(c + r\underline{\pi})$ because $r\underline{\pi}$ is the perpetual annuity flow value of $\underline{\pi}$. At an interior optimum, $0 < t_1^* < \infty$, the necessary first order conditions are $(\overline{\pi} - \underline{\pi})h_W(t_1^*) = (c + r\underline{\pi})$.

• Second order conditions: in order for the solution just given to be a local maximum rather than a local minimum, the benefits of waiting must be positive before t_1^* and negative after t_1^*. For this to be true, the hazard rate must be decreasing, $h_W'(t_1^*) < 0$.

In order to give a more formal analysis, note the following:

1. if $1_{[0,t_1)}(W) = 1$, i.e. if $W < t_1$, one incurs the search cost $\int_0^W (-c)e^{-rt}\,dt$ and receives the discounted profits of $\overline{\pi}e^{-rW}$; and
2. if $1_{[t_1,\infty)}(W) = 1$, one incurs the search cost $\int_0^{t_1} (-c)e^{-rt}\,dt$ and receives the discounted profits of $\underline{\pi}e^{-rt_1}$.

Thus the problem is

$$\max_{t_1 \in [0,\infty]} \mathbb{E}\left[1_{[0,t_1)}(W)\left(\int_0^W (-c)e^{-rt}\,dt + \overline{\pi}e^{-rW}\right) + 1_{[t_1,\infty)}(W)\left(\int_0^{t_1} (-c)e^{-rt}\,dt + \underline{\pi}e^{-rt_1}\right)\right]. \quad (1.1)$$

Evaluating the terms in rounded brackets and rewriting yields

$$\psi(t_1) = \int_0^{t_1} \left(-c\frac{1}{r}(1 - e^{-rw}) + \overline{\pi}e^{-rw}\right) f_W(w) + \left(-c\frac{1}{r}(1 - e^{-rt_1}) + \underline{\pi}e^{-rt_1}\right) G_W(t_1). \quad (1.2)$$

Taking derivatives with respect to t_1, using $G_W' = -f_W$, and rearranging yields

$$\psi'(t_1) = \left(e^{-rt_1} G_W(t_1)\right)\left[(\overline{\pi} - \underline{\pi})\, h_W(t_1) - (c + r\underline{\pi})\right]. \quad (1.3)$$

As $e^{-rt_1}G_W(t_1) > 0$, $\psi'(t_1^\circ) = 0$ only if $(\overline{\pi} - \underline{\pi})h_W(t_1^\circ) - (c + r\underline{\pi}) = 0$, yielding

$$\psi''(t_1^\circ) = \left(e^{-rt_1^\circ}G_W(t_1^\circ)\right)'[0] + (\overline{\pi} - \underline{\pi})h'_W(t_1^\circ), \tag{1.4}$$

which can only be strictly negative if $h'_W(t_1^\circ) < 0$. Interior strict optima, t_1^*, are indicated by (1.3) being satisfied and $h'_W(t_1^*) < 0$, which makes the comparative statics of t_1^* immediate: decreasing in c, r, and $\underline{\pi}$, increasing in $\overline{\pi}$, and increasing in uniform upward shifts of the hazard rate.

If W has a negative exponential distribution with parameter λ, the hazard rate is constant at λ and $\psi'(t_1) \lesseqgtr 0$ as $\lambda \lesseqgtr \frac{c+r\underline{\pi}}{\overline{\pi}-\underline{\pi}}$, so that $t_1^* = 0, \infty$ or $[0, \infty]$ depending on the sign of ψ', which does not vary with t_1. Of particular interest are the cases of monotonically increasing and decreasing hazard rates.

A Weibull distribution with parameters λ and γ is of the form $W = X^\gamma$ where X has a negative exponential(λ) distribution. The associated hazard rate is $h_W(t) = \frac{\lambda}{\gamma}t^{\frac{1-\gamma}{\gamma}}$.

1. If $\gamma > 1$ in the Weibull case, then $h_W(0+) = \infty$ and the hazard rate is strictly decreasing to 0, which means that there is always a unique optimal strictly positive delay before ending search.

2. If $\gamma < 1$ in the Weibull case, then the hazard rate starts at 0 and increases without bound.[8] Depending on r, λ and γ, the optimal strat-

[8]An interpretation of the increasing hazard rate is that learning-by-doing in process of search makes search more and more effective over time.

egy at $t = 0$ may be to end search immediately, $t_1^* = 0$, or to wait until success, $t_1^* = \infty$. Even if $t_1^* = 0$ at $t = 0$, because $\gamma < 1$, there will always be a time T with the property that if one has already waited until T, then the conditionally optimal choice is $t_1^*(T) = \infty$.

1.2.2 Optimal Adaptation to Circumstances

We now study a simple model of the optimal timing of adaptations to a stochastic dynamic state. The first set of random variables used to describe the problem are a Markov process, $\{X_k : k = 0, 1, \ldots\}$ taking values in a two state set, $S = \{s', s''\}$, with the transition matrix $P = \left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]$. The second set of random variables are arrival times, $Y_0, Y_1, \ldots$, represent the switching times for the states. These are non-negative random variables that satisfy $Y_0 \equiv 0$ with $W_k := Y_k - Y_{k-1}$ being i.i.d. non-negative random variables with densities on $[0, \infty)$, hence having no atoms, except perhaps at ∞ when they are incomplete.

Our first use of stochastic intervals is to define the continuous time process that the decision maker is reactively adapting to. A **stochastic interval** is a subset of $\Omega \times [0, \infty)$ of the form $[\![Y_k, Y_{k+1}[\![= \{(\omega, t) : Y_k(\omega) \leq t < Y_{k+1}(\omega)\}$. The Markov chain and the waiting times combine to form the continuous time stochastic process $(\omega, t) \mapsto X(\omega, t)$ defined by

$$X(\omega, t) = \sum_k X_k 1_{[\![Y_k, Y_{k+1}[\![}. \tag{1.5}$$

Thus, if $X_0 = s''$, then the state remains s'' until Y_1, at which point it switches to s', where it stays until Y_2, when it switches back, and so on.

There are two possible actions, $A = \{a', a''\}$. The flow payoffs to being in state s and taking action a are given by

	s'	s''
a'	1	0
a''	0	2

. This means that one always wants the action to match the state, matching action to state when $X_t = s'$ earns a flow of 1, matching action to state when $X_t = s''$ earns a flow of 2, and mis-matching earns a flow of 0. What may make instantaneous adjustments suboptimal is the cost of switching actions, $C > 0$. If another change in the state is expected soon, it may not be worth incuring C to enjoy the extra flow. The question is which changes in state to react to? and after what amount of delay?

The **value function**, $V_*(a,s)$, gives the maximal expected discounted utility to starting at $t = 0$ in state $s \in S$ with the present action being a. From Theorem 2 (below), $V_*(a,s)$ is well-defined, and further, it is achievable by sequentially solving an optimization problem within each stochastic interval $[\![Y_k, Y_{k+1}[\![$.

For the present model, the decision problem within an interval $[\![Y_k, Y_{k+1}[\![$ is to pick a time $t_1 \in [0,\infty]$ at which to change actions and incur the cost C. There will be two cases: $Y_{k+1} < Y_k + t_1$, i.e. $1_{[Y_k, Y_k+t_1)}(Y_{k+1}) = 1$, corresponding to the new change in state arriving before the planned change in action; and $Y_{k+1} > Y_k + t_1$, i.e. $1_{[Y_k+t_1,\infty)}(Y_{k+1}) = 1$, corresponding to the new change in state arriving after the planned change in action. It is clear that if $(a,s) = (a',s')$ or $(a,s) = (a'',s'')$, then $t_1^* = \infty$ is optimal because any change both incurs C unnecessarily and loses flow payoff.

Subtracting Y_k, setting $k = 0$ and $W = Y_1 - Y_0$, and letting $\kappa = \mathbb{E}e^{-rW}$, the value function satisfies

$$\begin{aligned}
V_*(a',s') &= \tfrac{1}{r}(1-\kappa) + \kappa V_*(a',s''),\\
V_*(a',s'') &= \max_{t_1\in[0,\infty]} \mathbb{E}\Bigg[1_{[0,t_1)}(W)\left(\int_0^W 0e^{-rt}\,dt + e^{-rW}V_*(a',s')\right) + \\
&\qquad 1_{[t_1,\infty)}(W)\left(\int_0^{t_1} 0e^{-rt}\,dt + \int_{t_1}^W 2e^{-rt}\,dt - e^{-rY_k+t_1}C + e^{-rW}V_*(a'',s')\right)\Bigg],\\
V_*(a'',s') &= \max_{t_1\in[0,\infty]} \mathbb{E}\Bigg[1_{[0,t_1)}(W)\left(\int_0^W 0e^{-rt}\,dt + e^{-rW}V_*(a'',s'')\right) + \\
&\qquad 1_{[t_1,\infty)}(W)\left(\int_0^{t_1} 0e^{-rt}\,dt + \int_{t_1}^W 1e^{-rt}\,dt - e^{-rY_k+t_1}C + e^{-rW}V_*(a',s'')\right)\Bigg],\\
V_*(a'',s'') &= \tfrac{2}{r}(1-\kappa) + \kappa V_*(a'',s'),
\end{aligned} \tag{1.6}$$

which reduces to a system of two equations in two unknowns.

The value function equations involve two optimization problems, the one at (a',s'') and the one at (a'',s'). Monotone comparative statics show that the optimal $t_1^*(a',s'')$ at (a',s'') is smaller than the solution $t_1^*(a'',s')$ (because the flow payoffs of the switch are 2 rather than 1). Let us suppose that the solution at (a',s'') is strictly positive and less than ∞ and examine the determinants of the corresponding $t_1^*(a',s'')$. The tradeoff is between the gain in flow utility and the loss if C is incurred and the state changes back to s' in a short time, and the first order conditions should tell us that the marginal gain of switching at t_1^* is equal to the expected marginal opportunity cost.

From Theorem 4 (below), the first order conditions for $0 < t_1^*(a', s'') < \infty$ are

$$[u(a'', s'') - u(a', s'')] - rC = h_W(t_1^*)\mathbb{E}\left[C + (V_*(a', s') - V_*(a'', s'))\right]. \qquad (1.7)$$

This condition must capture indifference between switching and not switching at t_1^*. The LHS times dt is the next instant's net flow benefit from switching: the term $[u(a'', s'') - u(a', s'')]$ gives the change in flow benefit; and rC is the perpetual annuity flow value of C. To analyze the RHS times dt: $h_W(t_1^*)dt$ gives the probability that the state switches from s'' back to s' in the next instant; if this happens, then the decision maker has saved C plus the value difference $V_*(a', s') - V_*(a'', s')$. In this problem, it is necessary that the LHS be positive in order to ever justify switching to a'' at (a', s'').

1.3 The Model

We start with a brief description of the basic relations between incomplete waiting times and their hazard rates. We then turn to the class of stochastic processes describing the utility relevant parts of the changing environment in which the decision maker is immersed. We then describe the class of decision problems under consideration, reactive semi-Markovian decision problems. Following this we give an alternative interpretation of the decision maker knowing the hazard rate. The basic existence and characterization results for an optimal policy are in the subsequent section.

1.3.1 Hazard Rates of Incomplete Waiting Times

A random variable, $W \geq 0$, is **incomplete** if it has a mass point at ∞. For a possibly incomplete W with density on $[0, \infty)$, the following summarizes the relation between the **density**, $f_W(t)$, the **cumulative distribution function (cdf)**, $F_W(t)$, the **reverse cdf**, $G_W(t)$, the **hazard rate**, $h_W(t)$, the **cumulative hazard**, $H_W(t)$, and the **mass at infinity**, q_W, for $t \geq 0$:

$$F_W(t) = \int_0^t f_W(x)\,dx;\; G_W(t) = 1 - F_W(t);\; h_W(t) = \frac{f_W(t)}{G_W(t)};$$
$$H_W(t) = \int_0^t h_W(x)dx;\; G_W(t) = e^{-H_W(t)};\text{ and } q_W = e^{-H_W(\infty)}. \quad (1.8)$$

If $H_W(t) = \int_0^t h(x)\,dx \uparrow \infty$ as $t \uparrow \infty$, then $q_W = 0$ so that $W < \infty$ with probability 1, so that $F_W(t) \uparrow 1$ and $G_W(t) \downarrow 0$ as $t \uparrow \infty$.

From $G_W(t) = e^{-H_W(t)}$ one sees that any non-negative h can serve as the hazard rate for some waiting time, W, and W is incomplete iff h is integrable. The following are well-known examples.

1. An incomplete negative exponential has cdf $(1 - q_W)(1 - e^{-\lambda t})$ and everywhere decreasing hazard rate $\lambda\left[(q_W/(1-q_W))e^{\lambda t} + 1\right]^{-1}$. If $q_W = 0$, then the hazard rate is constant and the waiting time is memoryless.

2. A Weibull distribution is of the form $W = X^\gamma$, $\gamma > 0$, where X is a negative exponential(λ). The cdf of an incomplete Weibull is $(1 - q_W)(1 - e^{-\lambda t^{1/\gamma}})$, and the hazard rate is $h_W(t) = \frac{\lambda}{\gamma} t^{\frac{1-\gamma}{\gamma}} \left[(q_W/(1-q_W))e^{\lambda t^{1/\gamma}} + 1\right]^{-1}$. If $\gamma > 1$, the $h_W(0+) = \infty$ and the hazard rate strictly decreases to 0.

If $\gamma < 1$, then $h_W(0) = 0$ and the hazard rate is first increasing then decreasing if $q_W > 0$, otherwise it is strictly increasing.

1.3.2 Semi-Markov Processes

All random variables are defined on a probability space, $(\Omega, \mathcal{F}, P)$. A **semi-Markov process** is a piecewise constant, right-continuous stochastic process, $X : \Omega \times [0, \infty) \to S$ where S is a state space.

The crucial part of specifying X is the Markov process $(S_k)_{k=0}^{\infty}$, and the associated distribution of waiting times until the next transition, $(W_k)_{k=0}^{\infty}$. We assume that there is a continuous stochastic kernel $s \mapsto p(s) \in \Delta(S)$, where $\Delta(S)$ is the set of Borel probabilities on S.[9] For the reactive problems that we consider below, the Markov process $(S_k)_{k=0}^{\infty}$ takes values in S with transition probabilities $P(S_{k+1} \in E \big| S_k = s) = P(S_{k+1} \in E \big| S_k = s, S_{j<k}) = p(s)(E)$.

For each $s \in S$, there is a distribution Q_s on $\mathbb{R}_{++} \cup \{\infty\}$ determining the distribution of waiting times for changes, $(W_k)_{k=1}^{\infty}$.[10] The building blocks for X are a countable collection, $\Xi = (S_k, W_k)_{k=0}^{\infty}$ where: the $W_0 \equiv 0$; $(W_k)_{k \geq 1}$ is a sequence of strictly positive random times with distributions depending only on S_k, i.e. $P(W_k \in B \big| S_k) = Q_{S_k}(B) = P(W_k \in B \big| \Xi)$.

A crucial construct for both defining semi-Markov processes and analyzing the associated optimization problems is the **stochastic interval**.

[9]At the most general, the state space and the action space are assumed to be Polish, i.e. Borel subsets of complete separable metric spaces. $\Delta(S)$ has the weak* topology from S.

[10]In future research, we intend have both p and Q determined both by s and by a.

Definition 1. Define $Y_0 = W_0$ and $Y_{k+1} = Y_k + W_{k+1}$. The **stochastic interval** between Y_k and Y_{k+1} is a subset of $\Omega \times [0, \infty)$ defined as $[\![Y_k, Y_{k+1}[\![= \{(\omega, t) : Y_k(\omega) \le t < Y_{k+1}(\omega)\}$.

We assume $P(W_k > 0) = 1$ for all $k \in \mathbb{N}$ and that $\sup_k \mathbb{E} e^{-rW_k} < 1$. This implies that $P(Y_k \to \infty) = 1$ so that $\Omega \times [0, \infty) = \cup_{k=0}^{\infty} [\![Y_k, Y_{k+1}[\![$.

Definition 2. The **semi-Markov process (smp)** based on Ξ is the function $X(\omega, t) = \sum_k S_k(\omega) 1_{[\![Y_k, Y_{k+1}[\![}(\omega, t)$.

X_t denotes the random variable $\omega \mapsto X(\omega, t)$. By construction, $X_{Y_k} = S_k$.

Example 1. If $\Xi = (S_k, W_k)_{k=0}^{\infty}$ the $S_k \equiv k$ and the W_k are i.i.d. negative exponentials with parameter λ, then the associated smp X is a standard Poisson process with arrival rate λ. More generally, if $(S_k)_{k=0}^{\infty}$ is any Markov chain taking values in S, then $(S_k)_{k=0}^{\infty}$ is **embedded** in X and the transition times follow a Poisson process. For more general distributions of the W_k, the standard queueing models give rise to special cases of smp's.

Example 2. In the optimal search time problem in §1.2.1, take $S = \{s_0, s_1\}$ where s_0 is the initial state and s_1 is the state "source of higher profits has been found." The transition kernel for the states is given by $P = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}$, i.e. one transitions from s_0 to s_1 but never the reverse. Let $A = \{a_0, \underline{a}, \overline{a}\}$ where a_0 is the action "continue searching" and $\underline{a}$ is the action "use the alternative (older) technology." and $\overline{a}$ is the action "use the newly found technology." Q_{s_0} is the distribution of W, and $Q_{s_1}(\infty) = 1$, which correpsonds to s_1 being an absorbing state. We arrange for $\overline{a}$ not to be chosen in s_0 by making the flow utility sufficiently negative.

When actions a_k do not affect the distributions of the S_{k+1} nor the distribution of the W_k, the problem of picking the optimal a_k is inherently **reactive**.

1.3.3 Uncertainty About the Hazard Rate

We derive our formal results in a mathematical setting where the hazard rate functions are known to the decision maker. For expected utility maximizers, this is no loss of generality as any arbitrary prior distribution over hazard rates can be mimicked by a single hazard rate function.

Let $\mathcal{P}$ denote the set of distributions on $[0,\infty]$. The set of extreme points is the set of point masses $\{\delta_x : x \in [0,\infty]\}$. For any $p \in \mathcal{P}$ which is not an extreme point, there are uncountably many probability distributions, μ, on $\mathcal{P}$ such that $p = \int_{\mathcal{P}} q \, d\mu(q)$. This implies that any hazard rate can arise from updating a distribution over distributions. When we use subsets of $\mathcal{P}$, there are approximate versions of this. We give one such result, there are uncountably many similar ones.

Let $g_W(\cdot|\lambda,\gamma)$ denote the density function of a Weibull(λ,γ).

Theorem 1. *For every strictly positive C^1 density, f, on $(0,\infty)$, for every $\epsilon > 0$, and for every compact $K \subset (0,\infty)$, there exists a density g of the form $g_\mu(x) = \int g_W(x|\lambda,\gamma)\, d\mu(\lambda,\gamma)$, μ a probability distribution on $\mathbb{R}^2_{++}$, such that $\max_{x\in K} |f(x) - g_\mu(x)| < \epsilon$ and $\max_{x\in K} |h_f(x) - h_{g_\mu}(x)| < \epsilon$ where h_f and h_{g_μ} are the hazard rates associated with the densities f and g_μ.*

Proof. If the class of Weibulls included the point masses, we could simply draw them according to the distribution f. The first part of the proof shows that the class of Weibulls, which is C^∞, contains distributions arbitrarily close to the point masses.

Letting $r = 1/\gamma$, $g_W(x|\lambda, 1/r) = \frac{1}{\lambda^r} r x^{r-1} e^{-(x/\lambda)^r}$. For $r > 0$, $g_W(\cdot|\lambda, 1/r)$ is single-peaked, and achieves its maximum at $x^* = \lambda\left(\frac{r-1}{r}\right)^{1/r}$, at which point it takes the value $g_W(x^*|\lambda, 1/r) = r\frac{1}{\lambda}\left(\frac{r-1}{r}\right)^{(r-1)/r} e^{(r-1)/r}$. Further, as $r \uparrow \infty$, $x^* \uparrow \lambda$, $g_W(x^*|\lambda, 1/r) \uparrow \infty$, and for any positive $c \neq 1$, $g_W(cx^*|\lambda, 1/r) \downarrow 0$.

For $r > 1$, let μ_r be the distribution concentrated on r and having λ distributed according to the target density f. As $r \uparrow \infty$, the C^∞ density of g_μ converges uniformly to the density of f. Because f is bounded away from 0 on K, this directly implies the hazard rates are also uniformly close for r sufficiently large. □

1.4 Reactive Semi-Markovian Decision Problems

We study the problem of the optimal timing and choice of actions for reactive smp's when flow utility, $u(a, s)$, depends on the action a in an action space A, and the state, s in a state space S. A and S are metric spaces, A is compact, $u(\cdot, \cdot)$ is jointly continuous and bounded.[11]

[11] From an abstract point of view, the continuity assumption is without loss — if each $u(\cdot, s)$ is continuous and each $u(a, \cdot)$ is measurable, then u is jointly measurable and one can always find a metric on S making u jointly continuous.

Total reward is the expected discounted integral of the flows, and changes of action are costly. An action plan, $\boldsymbol{a}(\omega, t)$, will depend on t and on ω, but will only depend on ω through the realization of the process X at or before t.

1.4.1 Filtrations and Action Plans

Define a filtration $\{\mathcal{F}_t : t \in [0, \infty)\}$ by defining $\mathcal{G}_s \subset \mathcal{F}$ to be the smallest σ-field making $\omega \mapsto X(\omega, s)$ measurable and $\mathcal{F}_t = \sigma\{\mathcal{G}_s : s \le t\}$. Let $\mathcal{B}_t$ denote the Borel σ-field on the interval $[0, t]$.

Definition 3. A **policy** is a mapping $a : \Omega \times [0, \infty) \to A$ such that

a. for all t, $a : \Omega \times [0, t] \to A$ is $\mathcal{F}_t \otimes \mathcal{B}_t$-measurable, and

b. for all ω, $t \mapsto a(\omega, t)$ is cadlag, piecewise constant, and has at most finitely many discontinuities on any bounded interval.

Associated with any policy $a(\cdot, \cdot)$ are the **change times**, $\boldsymbol{T}(\omega)$, defined by $\boldsymbol{T} = \{t_{k+1}(\omega) = \min\{t \ge t_k : a(t, \omega) \ne a(t_k, \omega)\}$ where $t_0 \equiv 0$.

1.4.2 Stochastic Interval Problems

The decision problem is to pick the policy that maximizes

$$E\left(\int_0^\infty u(\boldsymbol{a}, X) e^{-rs}\, ds - C \sum_{t \in \boldsymbol{T}} e^{-rt}\right)$$

given that one starts in state s_0 with action a_0 at $t_0 := 0$. We will show that the bounded continuous value function is the unique fixed point of a contraction operator. As a step in that direction and to begin to characterize the associated value function, we treat the decision problem within each stochastic interval, $[\![Y_k, Y_{k+1}[\![$.

We suppose that the interval $[\![Y_k, Y_{k+1}[\![$ starts in a state s_k, i.e. $X_{Y_k} = s_k$, with the action being a_k. Let $\boldsymbol{\tau} = (t_n)_{n\in\mathbb{N}} \in [0,\infty]^{\mathbb{N}}$ be a non-decreasing sequence of times, and let $\boldsymbol{\alpha} \in A^{\mathbb{N}}$ be a sequence of actions. The interpretation is that $\boldsymbol{\tau}$ is the set of points in the interval at which the decision maker switches and $\boldsymbol{\alpha}$ is the associated set of choices. Specifically;

1. if $Y_k + t_1 < Y_{k+1}$, then the decision maker incurs the cost $Ce^{-r(Y_k+t_1)}$ of switching from a_0 to a_1, and during the interval $[Y_k, Y_k + t_1)$ the decision maker enjoys the flow utility $u(a_0, s_0)$;

2. if $Y_{k+1} < Y_k + t_1$, then during the interval $[Y_k, Y_{k+1})$, the decision maker enjoys the flow utility $u(a_0, s_0)$ and at Y_{k+1}, they receive the random utility $V_*(a_0, S_{k+1})$; more generally,

3. if $Y_k + t_{n+1} < Y_{k+1}$, then the decision maker incurs the cost $Ce^{-r(Y_k+t_{n+1})}$ of switching from a_n to a_{n+1}, and during the interval $[Y_k + t_n, Y_k + t_{n+1})$, the decision maker enjoys the flow utility $u(a_n, s_0)$; and

4. if $Y_k + t_n < Y_{k+1} < Y_k + t_{n+1}$, then during the interval $[Y_k + t_n, Y_{k+1})$, the decision maker enjoys the flow utility $u(a_n, s_0)$, and at Y_{k+1}, they receive the random utility $V_*(a_n, S_{k+1})$.

To discuss the problem of finding the optimal times to move and the optimal actions to choose in the stochastic interval $[\![Y_k, Y_{k+1}[\![$ with (something like) minimal notation, we condition on $Y_k = t$, subtract t, renormalize the k to 0, and start the discounting from 0. To this end, define

$$M_0(\boldsymbol{\alpha},\boldsymbol{\tau};V_\circ) = 1_{[0,t_1)}(Y_1)\left[\int_0^{Y_1} u(a_0,s_0)e^{-rs}\,ds + e^{-rY_1}V_\circ(a_0,S_1)\right], \quad (1.9)$$

$$M_n(\boldsymbol{\alpha},\boldsymbol{\tau};V_\circ) = 1_{[t_n,t_{n+1})}(Y_1)\left[\sum_{k=0}^{n-1}\int_{t_k}^{t_{k+1}} u(a_k,s_0)e^{-rs}\,ds+ \right. \quad (1.10)$$

$$\left. \int_{t_n}^{Y_1} u(a_n,s_0)e^{-rs}\,ds + e^{-rY_1}V_\circ(a_n,S_{k+1})\right], \quad (1.11)$$

$$C_n(\boldsymbol{\alpha},\boldsymbol{\tau};V_\circ) = -\sum_{i=1}^{n} Ce^{-rt_i}\cdot 1_{[t_n,t_{n+1})}(Y_1). \quad (1.12)$$

Definition 4. For a bounded continuous $(a,s)\mapsto V_\circ(a,s)$ starting action a_0 and initial state s_0, the associated **within-interval problem (winp)** is

$$\max_{\boldsymbol{\alpha},\boldsymbol{\tau}} E\left(M_0(\boldsymbol{\alpha},\boldsymbol{\tau};V_\circ) + \sum_{n=1}^{\infty}[M_n(\boldsymbol{\alpha},\boldsymbol{\tau};V_\circ) + C_n(\boldsymbol{\alpha},\boldsymbol{\tau};V_\circ)]\middle| s_0, a_0\right). \quad (1.13)$$

1.4.3 The Blackwell-Pontryagin Equation

A series of Lemmas leads to the basic existence and characterization results.

Lemma 1. *Solving each winp sequentially yields an action plan.*

Proof. Within each $[\![Y_k, Y_{k+1}[\![$, the optimal $(\boldsymbol{\alpha}^*,\boldsymbol{\tau}^*)$ only depends on S_k. □

Lemma 2. *Each winp has a solution.*

Proof. Because Q_s has no atoms except perhaps at ∞, the sum in (1.13) is continuous on the compact subset of $A^{\mathbb{N}} \times [0,\infty]^{\mathbb{N}}$ for which $t_n \leq t_{n+1}$ for all $n \in \mathbb{N}$. □

Lemma 3. *If* $(\boldsymbol{\tau}^*, \boldsymbol{\alpha}^*)$ *solves a winp, then for all* $n \in \mathbb{N}$, $[t_n^* < \infty] \Rightarrow [t_n^* < t_{n+1}^*]$, *and* $t_n^* \uparrow \infty$.

Proof. For the first part, note that moving twice at any $t < \infty$ incurs the fixed cost Ce^{-rt} twice with no advantage in flow payoffs. For the second part, if $\sup_n t_n^* < \infty$, then the cost incured is infinite, and never moving is strictly better than this policy. □

Theorem 2. *If* $\beta = \inf\{E\, e^{-r\tau} : \tau \sim \mathcal{L}(s),\ s \in S\} < 1$, *then the mapping from* $V_\circ$ *to* $T(V_\circ)$ *defined by*

$$T(V_\circ)(a_0, s_0) = \max_{\boldsymbol{\alpha},\boldsymbol{\tau}} E\left(M_0(\boldsymbol{\alpha},\boldsymbol{\tau}; V_\circ) + \sum_{n=1}^{\infty} [M_n(\boldsymbol{\alpha},\boldsymbol{\tau}; V_\circ) + C_n(\boldsymbol{\alpha},\boldsymbol{\tau}; V_\circ)] \middle| s_0, a_0 \right) \tag{1.14}$$

has contraction factor β, *and maps bounded continuous functions to bounded continuous functions. Further, the value to following the solutions to each winp is the unique fixed point of* T.

Proof. From Blackwell's Lemma for contraction mappings (e.g. Corbae et. al., Lemma 6.2.33, p. 282), it is sufficient to show that T is monotonic and that for any constant c, $T(V_\circ + c) \leq T(V_\circ) + \beta c$. Monotonicity is immediate. For the contraction, note that for all policies $(\boldsymbol{\alpha},\boldsymbol{\tau})$ with $t_m \uparrow \infty$, the difference between

$$E\left(M_0(\boldsymbol{\alpha},\boldsymbol{\tau}; V_\circ + c) + \sum_{n=1}^{\infty} [M_n(\boldsymbol{\alpha},\boldsymbol{\tau}; V_\circ + c) + C_n(\boldsymbol{\alpha},\boldsymbol{\tau}; V_\circ + c)] \right) \tag{1.15}$$

and

$$E\left(M_0(\boldsymbol{\alpha},\boldsymbol{\tau};V_\circ)+\sum_{n=1}^{\infty}[M_n(\boldsymbol{\alpha},\boldsymbol{\tau};V_\circ)+C_n(\boldsymbol{\alpha},\boldsymbol{\tau};V_\circ)]\right) \tag{1.16}$$

is

$$E\sum_{m=0}^{\infty}1_{[t_m,t_{m+1})}(Y_1)e^{-rY_1}c, \tag{1.17}$$

and because $t^*_m \uparrow \infty$, this difference is identically $e^{-rY_1}c$. Since the difference is independent of the choice of optimal policy, the solution (set) to the problem of maximizing the equation in (1.15) is exactly the same as the solution (set) for maximizing (1.16). Finally, since $e^{-rY_1} \le \beta$, $T(V_\circ + c) \le T(V_\circ) + \beta c$.

Finally, following the solution to the first winp associated with V_* yields value V_*. Inductively, following the solution to the first n winps also has value V_*. Since $Y_k \uparrow \infty$, the value to following the solution to each winp is V_*. □

1.5 Euler Equations

We give the Euler equations in reactive smp problems for the optimal policy within the class of policies where one moves at most once within a stochastic interval[12]. Recall that for a reactive problem, we consider semi Markov processes whose stochastic kernels do not depend on the choice of action.

[12]It pertains to situations where the cost is high enough so that the number of feasible moves within a stochastic interval is at most one, or the structure of the decision problem allows only a single action each time. A broad class of institutional decision making problems would fit this description.

Also, note that we now have two choice variables, namely the choice of action a and the optimal time to execute that action t, hence we now have two Euler equations, slightly different from the usual dynamic programming problem.

1.5.1 Memoryless Processes

Theorem 3. *If $(W_k)_{k=1}^{\infty}$ are i.i.d Exp(λ), then*

- *the optimal waiting time, if unique, is either 0 or ∞*
- *if the waiting time is non-unique, then it must be the whole interval $[0,\infty]$*
- *the optimal action only depends on the current state and is independent of the timing of action.*

Proof. Let us re-initialize time within each interval by setting $Y_k = 0$. With the condition that at most one move is feasible given C, we have the value function

$$\begin{aligned}
&V(a_k, s_k) \qquad (1.18)\\
&= \max_{t\in[0,\infty]} \max_{a\in A} \mathbb{E}\Big\{\Big[\int_0^{y_{k+1}} u(a_k,s_k)e^{-ru}du + V(a_k,s_{k+1})e^{-ry_{k+1}}\Big]\mathbf{1}_{\{y_{k+1}<t\}}\\
&+\Big[\int_0^{t} u(a_k,s_k)e^{-ru}du + \int_t^{y_{k+1}} u(a,s_k)e^{-ru}du + V(a,s_{k+1})e^{-ry_{k+1}} - C\Big]\mathbf{1}_{\{y_{k+1}>t\}}\Big|s_k\Big\}
\end{aligned}$$

First we establish that the optimal action does not depend on the timing of action, which is readily seen from the F.O.C. of the maximand in (1.18) with respect to a:

$$\frac{\partial}{\partial a}u(a,s_k) = -\frac{\lambda}{r}\frac{\partial}{\partial a}V(a,s_{k=1})$$

If $t = 0$ is chosen, then the Value Function evaluated at $t = 0$ is

$$V_0 = \max_{a \in A}[\frac{1}{\lambda + r}u(a,s_k) + \frac{\lambda}{\lambda + r}\mathbb{E}V(a,s_{k+1}) - C]$$

If $t = \infty$ is chosen, then the Value Function evaluated at $t = \infty$ is

$$V_\infty = \frac{1}{\lambda + r}u(a_k,s_k) + \frac{\lambda}{\lambda + r}\mathbb{E}V(a_k,s_{k+1})]$$

Finally, the derivative of the maximand in (1.18) with respect to t is

$$\int_0^t (V_\infty - V_0)e^{-(\lambda+r)u}du$$

where

$$\begin{aligned} V_\infty - V_0 &= \frac{1}{\lambda + r}[u(a_k,s_k) - u(a^*,s_k)] + \frac{\lambda}{\lambda + r}\mathbb{E}[V(a_k,s_{k+1}) - V(a^*,s_{k+1})] - C \\ &= \frac{1}{\lambda + r}\{(u(a_k,s_k) - u(a^*,s_k) - rc) + \lambda\mathbb{E}[V(a_k,s_{k+1}) - V(a^*,s_{k+1}) - C]\}. \end{aligned}$$

The sign of the derivative depends only on the sign of $V_\infty - V_0$, which is constant on $(0,\infty)$. If $V_\infty - V_0 > 0$, the value is increasing in time and $t^* = \infty$. If $V_\infty - V_0 < 0$, then the value is decreasing in time and $t^* = 0$. If $V_\infty - V_0 = 0$, then any time in $[0,\infty]$ is optimal. □

Hence, with memoryless distributions of arrival times, the optimal action depends only on the current state, and the optimal timing is either to change immediately after observing the state change, or wait until the next change.

If $V_\infty - V_0 = 0$, then

$$u(a_k, s_k) - u(a^*, s_k) + rC = \lambda \mathbb{E}[V(a^*, s_{k+1}) - V(a_k, s_{k+1})] - \lambda C \tag{1.19}$$

which is a special case of more general Euler equations derived in Theorem 4 that we turn to next.

1.5.2 Waiting Time Distributions with Memory

Theorem 4. *If $(W_k)_{k=1}^{\infty}$ have distributions with memory, i.e. if $P(W > t \mid W \geq t)$ depends non-trivially on t, then the optimal time of action depends on the hazard function. In particular, for complete waiting times, if the hazard rate is monotonic increasing, or decreasing, and at most one action is feasible then,*

- *any optimal interior time for action (t^*), and the optimal action a^* are characterized by the following system of Euler equations:*

$$u(a^*, s_k) - u(a_k, s_k) - rC = -h_W(t^*)\mathbb{E}[V(a^*, s_{k+1}) - V(a_k, s_{k+1}) - C]$$

$$\begin{aligned} &\int_{t^*}^{\infty} e^{-ry} \frac{d}{da} u(a, s_k)|_{a^*} (1 - F_W(s, y)) dy \\ &+ \int_{t^*}^{\infty} e^{-ry} f_W(s, y) \frac{d}{da} \mathbb{E} V(a, S_k)|_{a^*} \\ &= 0 \end{aligned} \tag{1.20}$$

- *the partial second order condition for interior optimal time t^* for both increasing and decreasing hazard rates are as follows,*

$$\frac{u(a^*, s_k) - u(a_k, s_k) - rC}{\mathbb{E}[V(a^*, s_{k+1}) - V(a_k, s_{k+1}) - C]} < \frac{f'_W(t^*)}{f_W(t^*)},$$

and in addition for decreasing hazard rates

$$\left|\frac{u(a^*, s_k) - u(a_k, s_k) - rC}{\mathbb{E}[V(a^*, s_{k+1}) - V(a_k, s_{k+1}) - C]}\right| > \left|\frac{f'_W(t^*)}{f_W(t^*)}\right|$$

Proof. For the derivation of the Euler equations, see Appendix. The first Euler equation makes it clear that optimal time of action depends on the hazard function.

The rest follows from the requirements of the second order conditions. Note first that (4) implies $u(a^*, s_k) - u(a_k, s_k) - rC$ and $\mathbb{E}[V(a^*, s_{k+1}) - V(a_k, s_{k+1}) - C]$ must have opposing signs at the optima, since for any t^*, $h_W(t^*) \geq 0$. Now, for an interior optimum for t^*, the second order condition (1.18) w.r.t. t at the optima is[13]:

$$f_W(t^*)[u(a^*, s_k) - u(a_k, s_k) - rC] - f'_W(t^*)\mathbb{E}[V(a^*, s_{k+1}) - V(a_k, s_{k+1}) - C] < 0 \quad (1.21)$$

which implies

$$\frac{u(a^*, s_k) - u(a_k, s_k) - rC}{\mathbb{E}[V(a^*, s_{k+1}) - V(a_k, s_{k+1}) - C]} < \frac{f'_W(t^*)}{f_W(t^*)} \quad (1.22)$$

It is easy to show that $h'_W(t) > 0 \Rightarrow f'_W(t) > 0$. Hence, given that the ratio on the L.H.S. must be negative if the (4) is satisfied, the S.O.C. is always

[13]easy calculation of the second derivative and then using the F.O.C

satisfied for an increasing hazard rate whenever the F.O.C is satisfied. For a decreasing hazard rate, the S.O.C. will be satisfied for small enough change in hazard rate, i.e. when

$$\left|\frac{u(a^*, s_k) - u(a_k, s_k) - rC}{\mathbb{E}[V(a^*, s_{k+1}) - V(a_k, s_{k+1}) - C]}\right| > \left|\frac{f_W'(t^*)}{f_W(t^*)}\right|. \tag{1.23}$$

□

1.5.3 Interpretation of the Euler Equations

The Euler equations have simple interpretations in terms of marginal benefits and marginal costs of change. First note that an equivalent representation of (4) is

$$\begin{aligned} &(1 - F_W(t^*))[u(a^*, s_k) - u(a_k, s_k) - rC] \\ &= f_W(t^*)\mathbb{E}[V(a_k, s_{k+1}) - V(a^*, s_{k+1}) + C] \end{aligned} \tag{1.24}$$

The L.H.S represents the net expected instantaneous flow cost/benefit of moving to action a^*, given that the state change has not happened yet. We would lose/gain $u(a^*, s_k) - u(a_k, s_k)$ in terms of flow utility and rC in terms of flow cost of action. The R.H.S shows expected instantaneous benefit of waiting if the state changes in the next instant. It is the expected net benefit of starting the next stochastic interval with the action a_k instead of a^*, net of cost saved, multiplied by the conditional probability that the next change of state would occur in the next instant, given that it has not occurred till now.

The second Euler equation has straightforward interpretation as an expression of the principle that the expected marginal increase in value with a change in action must equal zero.

The important thing to note is that for (4) to be satisfied at an interior point, $u(a^*, s_k) - u(a_k, s_k) - rC$ and $\mathbb{E}[V(a^*, s_{k+1}) - V(a_k, s_{k+1}) - C]$ must have opposite signs, given optimal action a^*. This is intuitive, because if the optimal action is such that both these terms are positive, then it means we gain utility by moving to a^* regardless of whether the state remains the same or changes, so one should move to a^* immediately. On the other hand, if both these terms are negative, then even the optimal change is not worth the cost, given the current state and the expected future state. Hence one must take no action, i.e. wait till the next change.

The next thing to note are the conditions there might exist an interior optimal time under both increasing and decreasing hazard rates, but under opposing circumstances. Observe that for a fixed t^*, we have a fixed a^*, and $u(a^*, s_k) - u(a_k, s_k) - rC$ is constant over time.

The following corollary characterizes the conditions under which interior optima exist for decreasing hazard rates.

Corollary 4.1. *For any interior optimal time t^* to exist in case of decreasing hazard rates, the following must be true:*

$$u(a^*, s_k) - u(a_k, s_k) - rC > 0$$

$$\mathbb{E}[V(a^*, s_{k+1}) - V(a_k, s_{k+1}) - C] < 0$$

and

$$u(a^*, s_k) - u(a_k, s_k) - rC < -h_W(t)\mathbb{E}[V(a^*, s_{k+1}) - V(a_k, s_{k+1}) - C] \text{ for } t < t^*$$

For any interior optimal time t^* to exist in case of increasing hazard rates, the following must be true:

$$u(a^*, s_k) - u(a_k, s_k) - rC < 0 \tag{1.25}$$

$$\mathbb{E}[V(a^*, s_{k+1}) - V(a_k, s_{k+1}) - C] > 0 \tag{1.26}$$

$$u(a^*, s_k) - u(a_k, s_k) - rC > -h_W(t)\mathbb{E}[V(a^*, s_{k+1}) - V(a_k, s_{k+1}) - C] \text{ for } t < t^* \tag{1.27}$$

Proof. Consider the various possible cases.

Case 1 Let,

$$u(a^*, s_k) - u(a_k, s_k) - rC > 0 \tag{1.28}$$

and

$$\mathbb{E}[V(a^*, s_{k+1}) - V(a_k, s_{k+1}) - C] < 0 \tag{1.29}$$

This is the scenario where a new optimal action provides net benefits in the current state, but it would be better to start the next stochastic interval with the status quo action. Then

$$-h_W(t^*)\mathbb{E}[V(a^*, s_{k+1}) - V(a_k, s_{k+1}) - C] > 0 \tag{1.30}$$

Clearly, there could not be an interior solution t^* with $h_W(t)$ decreasing in t if

$$u(a^*, s_k) - u(a_k, s_k) - rC > -h_W(t)\mathbb{E}[V(a^*, s_{k+1}) - V(a_k, s_{k+1}) - C] \text{ for } t < t^* \quad (1.31)$$

because then one would just change immediately. So, we would only have interior optima with decreasing hazard rate in *Case 1* if $-h_W(t)\mathbb{E}[V(a^*, s_{k+1}) - V(a_k, s_{k+1}) - C]$ cuts $u(a^*, s_k) - u(a_k, s_k) - rC$ from above, as a function of t.

On the other hand, we cannot have an interior optimal t^* in *Case 1* with $h_W(t)$ increasing in t if

$$u(a^*, s_k) - u(a_k, s_k) - rC < -h_W(t)\mathbb{E}[V(a^*, s_{k+1}) - V(a_k, s_{k+1}) - C] \text{ for} t < t^* \quad (1.32)$$

because then the benefit of waiting is higher than the cost at $t = 0$ and keeps increasing over time, and one would keep waiting forever. We also could not have an interior optimal time in this case with $h_W(t)$ increasing in t even if

$$u(a^*, s_k) - u(a_k, s_k) - rC > -h_W(t)\mathbb{E}[V(a^*, s_{k+1}) - V(a_k, s_{k+1}) - C] \text{ for} t < t^* \quad (1.33)$$

because although the benefit of waiting is less than the cost at $t = 0$, the more we wait, the marginal benefit of waiting one instant longer keeps going up. Hence optimal t^* is either 0 or ∞.

So there is no interior optimal time with increasing hazard rate in *Case 1*

Case 2 Let,

$$u(a^*, s_k) - u(a_k, s_k) - rC < 0 \quad (1.34)$$

and

$$\mathbb{E}[V(a^*, s_{k+1}) - V(a_k, s_{k+1}) - C] > 0 \tag{1.35}$$

This is the scenario when a new optimal action is worse than the status quo in the current state, but it would be better to start the next stochastic interval with the new optimal action. Then

$$-h_W(t^*)\mathbb{E}[V(a^*, s_{k+1}) - V(a_k, s_{k+1}) - C] < 0 \tag{1.36}$$

We cannot have an interior solution t^* with $h_W(t)$ decreasing in t if

$$u(a^*, s_k) - u(a_k, s_k) - rC > -h_W(t)\mathbb{E}[V(a^*, s_{k+1}) - V(a_k, s_{k+1}) - C] \text{ for } t < t^* \tag{1.37}$$

because the R.H.S, which now denotes the cost of waiting, keeps falling over time. And of course we cannot have an interior solution with

$$u(a^*, s_k) - u(a_k, s_k) - rC < -h_W(t)\mathbb{E}[V(a^*, s_{k+1}) - V(a_k, s_{k+1}) - C] \text{ for } t < t^* \tag{1.38}$$

because although the benefit of waiting is less than the cost at $t = 0$, the more we wait, the marginal cost of waiting one instant longer keeps going down. Hence optimal t^* is either 0 or ∞.

With increasing hazard rate, we have no interior optimal time t^* if

$$u(a^*, s_k) - u(a_k, s_k) - rC < -h_W(t)\mathbb{E}[V(a^*, s_{k+1}) - V(a_k, s_{k+1}) - C] \text{ for } t < t^* \tag{1.39}$$

one would change action immediately, since the cost of waiting is forever increasing. But if,

$$u(a^*, s_k) - u(a_k, s_k) - rC > -h_W(t)\mathbb{E}[V(a^*, s_{k+1}) - V(a_k, s_{k+1}) - C] \text{ for } t < t^* \tag{1.40}$$

then, although waiting gives net benefit at the beginning, the cost of waiting keeps going up, and at t^* it becomes optimal to take the new optimal action.

□

1.6 Extensions

We have presented results for reactive problems where the stochastic kernel of the smp does not depend on the choice of action. There are real life situations which are better modeled as controlled smp's where the choice of action effects the stochastic kernel. That would be the next step in our analysis. The next essay provides an example where controlled stochastic processes would be relevant in the context of Precautionary Principle and environmental degradation.

1.7 Conclusion

We began with the question as to what the optimal time is for changing a status quo policy in response to an environmental change when policy change is costly and one anticipates another change in the environment in an unknown, random time in the future. There is an immediate tradeoff between optimizing with respect to the current state and optimizing with respect to the expected future state, given that actions are costly and different actions become optimal in different states. But the main interest in such a problem stems from the fact that passage of time since the last observed change might contain information about how soon the next change is likely to occur. This would be the case when the distribution of inter-arrival times for environmental changes have hazard rates that are not constant over time. In case of increasing or decreasing hazard rates, the likelihood that the next change would happen in the next instant, given that it has not happened until now, goes up or down respectively. Hence we have an additional tradeoff in terms of timing of the action once we have seen an environmental change, namely, we lose utility every instant that the current action is not optimized to the current state, but every instant of passing time gives us more information about how far in the future the next change is likely to occur. Intuitively, it would suggest that there might be a place for 'informative waiting', i.e. delaying one's action in order to have more information about the time of the next environmental change. Our results show that for non-constant hazard rates, delaying your actions could be optimal under certain circumstances.

Chapter 2

Ambiguity and Hesitation

2.1 Introduction

This chapter looks at decision problems with ambiguity and extends the results of previous chapter to this setting. The aim is to provide a framework for posing questions about optimal institution design in the face of uncertain knowledge. We look at the particular example of the Precautionary Principle in environmental law.

The Precautionary Principle is the main anticipatory principle in international law for dealing with environmental degradation ([O'Riordan and Cameron, 1994]). It has formed the basis of many international treaties such as the Declaration of the Conference of Rio on Environment and Development, the Maastricht Treaty and the Kyoto Protocol. Article 15 of the Rio Declaration encapsulates the Precautionary Principle as follows: where there are threats of serious and irreversible damage, lack of full scientific certainty shall not be used as a reason for postponing cost-effective measures to prevent environmental degradation. Despite its widespread use, the Precautionary Principle has also been criticized for not having proper analytical basis [Sunstein, 2002, 2005]. There has been some recent attempts to provide economic or decision-theoretic foundations to justify the Precautionary Principle([Gollier et al., 2000] and references therein). This project aims to add to that literature.

Any analytical foundation for the precautionary principle must contain ingredients that address the following issues: scientific uncertainty, possibility of revelation of information over time (thereby the possibility of option value of information) and hysteresis. The existing literature deals with these issues primarily in an expected utility framework ([Arrow and Fisher, 1974; Kreps and Porteus, 1978; Gollier et al., 2000]) and the results are driven by decision maker's attitude towards risk. Our contribution is to introduce the concept and tools of ambiguity as a more reasonable model of scientific uncertainty and to delineate the relationship between the society's ambiguity aversion and timing of environmental action.

We follow Epstein [1999] in defining an ambiguous outcome as an incomplete description of the probability distribution over consequences. An 'incomplete description' is identified with a set of probability distributions that satisfy the description. A choice problem is uncertain if the decision maker is choosing between distributions, and is ambiguous if he is choosing between sets of distributions.

An example of decision making under ambiguity is environmental policy making in the absence of a scientific consensus about the possibility of man-made environmental disasters. Given that most scientific evidence about the causes and effects of global warming is presented as ranges of probabilities, well-defined probabilistic beliefs, which is a feature of most existing models, do not seem to be a suitable representation of the situation. Moreover, economic models of climate change have traditionally treated the process as one of gradual change to new, stable state. Recent research in climate science has found evidence of both very rapid changes over a short period of time (around a decade) and also periods of significant fluctuations or 'environmental flickering' over periods as short as a year (Hall and Behl [2006], Stern and Treasury [2007]). These phases of rapid change and/or flickering seem to be triggered once a threshold point is reached in the ecological system. In addition, how fast the threshold is approached seem to depend on cumulative effects over time or 'hysteresis'.

While 'gradual change' models have usually prescribed 'adaptation to climate change' as opposed to 'intervention to avert it' [Nordhaus and Boyer, 2003], the decision problem takes a new shape when we incorporate the uncertainty over the expected arrival time of a possible catastrophic change and over the issue of whether or not we are moving towards such a critical threshold. In this context, our approach is to introduce a set of probability distributions with differing hazard rates to represent the state of scientific knowledge.

2.2 The Precautionary Principle

We formalize the Precautionary Principle by setting up a problem of choosing an optimal time to stop doing an activity that yields a high payoff per unit of time, but puts us on a path with an increasing hazard rate of catastrophe. Once we stop the activity, we get a lower payoff per unit of time, but we move off the path of catastrophe.

Consider a situation where one can continue an activity and earn C_m per unit of time or stop it at any point of time t° and earn $c_n < C_m$ from then on. As long as the activity is going on, a catastrophe might happen at a random time W and the payoff would be $-K$ from then on.

2.2.1 Consequence Space

Let the consequence space M consists of history of payoffs, $h(t)$.

$M \subset \{h : (0, \infty) \to \{C_m, c_m, -K\}\}$

Possible histories given our formulation are: $h(t) = C_m 1_{(0,t')}(t) - K 1_{[t',\infty)}(t)$, or $h(t) = C_m 1_{(0,t')}(t) + c_n 1_{[t',\infty)}(t)$

2.2.2 Waiting Times

Let $p \in A \subset \Delta((0,\infty])$ such that p has a density on $(0,\infty)$ and $p(\{\infty\}) > 0$ is possible. W_p is the corresponding r.v.

For each $p \in A$ and each choice of $t^\circ \in [0, \infty]$, there is a distribution $q_{p,t^\circ} \in \Delta(M)$ given by the distribution of

$C_m 1_{\{0,\min(W,t^\circ)\}}(t) - K 1_{\{W<t^\circ\}} 1_{[W,\infty)}(t) + c_n 1_{\{W \geq t^\circ\}} 1_{[t^\circ,\infty)}(t)$. For each A a non-empty closed convex subset of $\Delta(M)$ and each choice of t°, the set of distributions on M is $A_{t^\circ} := \{q_{p,t^\circ} : p \in A\}$.

Note that $A_{t^\circ} \in \mathcal{K}(\Delta(M))$, i.e. is a closed convex subset of the distributions on M.

2.2.3 Optimization Problem

For each $h \in M$, let $u(h) := \int_0^\infty e^{-rt} h(t)\, dt$.

For a single well-defined probability distribution, i.e when A is a singleton set, we can write the optimization problem as:

$$\max_{t^\circ \in [0,\infty]} \mathbb{E}\left[1_{[0,t^\circ)}(W) \left(\int_0^W C_m e^{-rt}\, dt - e^{-rW} K \right) + 1_{[t^\circ,\infty)}(W) \left(\int_0^{t^\circ} C_m e^{-rt}\, dt - \int_{t^\circ}^\infty c_n e^{-rt}\, dt \right) \right]$$

The first order condition for the above problem is

$$h(t^\circ)[c_n - K] = r(C_m - c_n) \tag{2.1}$$

Note that for an interior optimum to exist, we need increasing hazard rate here.

Now to formalize the idea of a lack of scientific consensus about potential of man-made environmental disaster, we introduce ambiguity in the form of a set of probability distributions Athat are ordered by hazard rates.

We consider the Gilboa-Schmeidler representation of ambiguity aversion ([Schmeidler and Gilboa, 1989]): $U : \mathcal{K}(\Delta(M)) \to \mathbb{R}$ is

$$U(A) = \alpha \min_{p \in A} \int u(h)\, dp(h) + (1-\alpha) \max_{q \in A} \int u(h)\, dq(h)$$

Here α is the measure of ambiguity aversion.

Hence the decision problem is to choose the optimal stopping time t° to solve the problem

$$\max_{t^\circ} U(A_{t^\circ}) \tag{2.2}$$

We want to explore the relationship between ambiguity aversion and hesitation.

Consider an ambiguous set of probability distributions over arrival times in which the distributions are ordered by hazard rate ordering.

Proposition 1. *For any $A \in \mathcal{K}(D)$, the optimal t° is a decreasing function of α.*

Proof. Given that the set of probability distributions is ordered by hazard rate ordering, let $F^H(t)$ and $F^L(t)$ be the highest and lowest hazard rate functions in the set. We make the observation that $\min_{p \in A} \int u(h)\, dp(h)$ is solved by $F^L(t)$ and $\max_{q \in A} \int u(h)\, dq(h)$ is solved by $F^H(t)$. So we can write the first order condition at an interior optimum as:

$$\alpha u_L' + (1-\alpha)u_H' = 0 \tag{2.3}$$

Now, to get the cross-partial effect, differentiating the FOC and rearranging,

$$\frac{\partial}{\partial\alpha\partial t^\circ}U^2 = u_L' - u_H' + \frac{\partial t^\circ(\alpha)}{\partial\alpha}\left[\alpha u''(L)(t^\circ(\alpha)) - (1-\alpha)u''(H)(t^\circ(\alpha))\right] = 0$$

Now, the term inside the square brackets multiplying $\frac{\partial t^\circ(\alpha)}{\partial\alpha}$ is negative at an interior optimum, since it is the second order partial derivative of the maximand. Now consider the first term. It is the difference between the first order derivatives of the two components of the maximand. When the sum is at he maximum, this difference has to be negative. □

2.2.4 Conclusion

We introduce a representation of scientific uncertainty in terms of an 'ambiguous' set of probability distributions and show that ambiguity aversion creates a 'precautionary effect'. This provides decision-theoretic foundations for the Precautionary Principle that are of a different nature than those derived from expected utility theory.

Chapter 3

Information Aggregation in Prediction Markets: An Empirical Investigation

1

3.1 Introduction

In the last few decades, asset markets designed for and exclusively dedicated to gathering information about probable outcomes of future events have come to existence. These markets, popularly known as prediction or information markets, have attracted attention for making accurate predictions about election outcomes [Berg et al., 2003, 2008a,b], product sales [Chen and Plott, 2002], film box office and myriad other variables of interest. With some operational variations, the fundamental design of such markets is as follows. The future event of interest is formalized as a random variable whose outcome would depend on an unobserved underlying state about which information is assumed to be dispersed among potential par-

[1]This essay is derived from joint work with Robert Lieli. My specific contribution to this project includes formulation of the original research question, data collection and compilation, construction of the time series variables representing media coverage indices and part of the data analysis in terms of developing and executing the methodology.

ticipants in the market. A set of complete contingent assets are introduced, whose eventual payoffs are tied to the future realization of the outcome of interest. The participants are expected to trade in the contingent assets based on their private information about the underlying state. Subject to assumptions about arbitrage possibilities and risk attitudes of the traders, the resultant market prices could be considered as a probabilistic prediction of the outcome of interest, at least in theory.

The performance of such markets in predicting the respective events of interest have created an optimism about its potential as a forecasting technology (Arrow et al. [2008], Sunstein [2007], Berg and Rietz [2003]). There is a nascent industry around this technology, and a growing stream of literature trying to understand various aspects of the information aggregation process in prediction markets [Tziralis and Tatsiopoulos, 2007]. Such comprehensive understanding would be vital for future improvement in this technology. Nevertheless, there remains a number of open questions, both theoretical [Wolfers and Zitzewitz, 2004] and empirical. The set of open questions include, but are not limited to, the following: i) What kind of information is built into the price? Do prices contain information beyond that collected by traditional mechanisms such as polls and the news media? ii) How fast does new information get built into the price? iii) What properties (statistical moments) of the price process best exhibit the impact of new information? iv) How does the market perform in terms of forecasting relative to other mechanisms (like polls)?

In comparing polls and prediction markets, the existing empirical literature has focused mainly on the last question. In a series of papers, Berg et al. have shown, using Iowa Electronic Market (IEM) data, that prediction market prices have significantly lower forecast errors both in the short and the long run compared to contemporaneous polls[Berg et al., 2003, 2008a,b]. In response, Erikson and Wlezien [2008] point out that since polls and prediction markets ask inherently different questions, they are not directly comparable. The standard question asked in a pre-election poll is this: ”if the election were held today, who would you vote for?” Therefore, polls only capture *voters' preferences at the moment* and should not be interpreted as forecasts in themselves; rather, sophisticated forecasts should build on the information provided by the polls. Presumably, markets do just that, as traders take into account their private and public sources of information before placing their bid, providing, in effect, an answer to the question: "who do you think will win"? Nevertheless, other sophisticated forecasts may exist. Using data from the same time period as Berg et al. [2008b], Erikson and Wlezien [2008] construct projections of vote shares and win probabilities based on daily polls and show that those projections can outperform prediction markets.

In contrast, our goal is to investigate the information flow between prediction market prices and two other conventional aggregators of information, namely media and opinion polls. We are not directly concerned with forecasting performance; rather, we seek to establish stylized facts that go towards answering questions i), ii) and iii). Specifically, we examine the sequence of primary elections in the 2008 race for the Democratic nomination for president, and the evolution of prices in the market for the "Obama wins nomination" security. We construct time series variables that reflect the "pollster's surprise" in each primary election, measured as the difference between Obama's actual vote share and vote share predicted by the latest poll(s) before the primary, as well as indices that describe the extent of daily media coverage devoted to him. We conduct Granger causality tests between the daily percentage change in the price of the "Obama wins nomination" security and these information variables. These tests provide answers to the following operationalized versions of question i): Do prediction market prices have information that is not reflected in contemporaneous polls and media stories? Conversely, do prices react to information that appears to be news for pollsters or is prominently featured by the media? Further, the time horizons over which Granger causality relationships can be established provide answers to question ii). Finally, regarding question iii), the current version of the paper contains results about linear projections only. While under some auxiliary assumptions these projections can be interpreted as models of the conditional mean, we have not yet conducted tests aimed at detecting Granger causality in higher moments.

Our main empirical finding regarding the relationship between price changes and the pollster's surprise is strong two-way Granger causality. Thus, on the one hand, part of the pollster's surprise is predictable by previous price movements, so market prices appear to contain information not contained in the polls. On the other hand, the pollster's surprise also contains information about future price movements even when the history of prices is taken into account. Therefore, there is an element of the pollster's surprise that is genuine news for the market as well.

Regarding the relationship between prediction markets and media covarage, there is some evidence of one-way Granger causality in the short run from price changes towards some of the media news indices. Media coverage, particularly in the election season, contains both useful information and hype. Therefore, even if prediction markets are actually able to pick out the truly informative content from the media and discard the rest, the quantitative level of media coverage need not show a particularly strong empirical relationship with prices. Our baseline tests still deliver evidence that prices capture information earlier than its revelation in the news media, although the apparent strength of is sensitive to modeling decisions such as the inclusion autoregressive conditional heteroskedasticity terms in the price process.

3.2 Data

For prediction market prices, we use data from the IEM winner-take-all market for predicting the winner of the Obama vs. Clinton nomination race in 2008 for the period January 1, 2008 to June 15, 2008.[2] The raw data is a daily time series of prices (recorded every day at midnight, the single price being the last transaction price before midnight).

Data on state-based polls on primary outcomes was compiled from the website Real Clear Politics[3]. The polls are taken on the eve of the Democratic primaries in the respective states, conducted on likely Democratic voters, essentially asking the question who they would support in the upcoming primary (since polls were conducted by different organizations, the exact version of the question might have varied).

For each primary election, we consider the difference between Obama's actual vote share and his predicted vote share as measured by various polls in that state/territory. This raw difference is then multiplied by the fraction of total delegates at stake in the election (hence, a one percentage point difference in actual versus predicted vote share in a 'big' state counts as a larger surprise than the same observed difference in a 'small' state). To compute total surprise for a given day, we add the weighted surprise measures for all elections held that day. If there are no elections on the day in question, we set the surprise measure to zero.

[2] In the IEM winner-take-all market, one share of a candidate pays off 1 dollar if the candidate wins and nothing if the candidate loses. A portfolio of one unit of each candidate pays exactly 1 dollar. A trader who buys one unit of a candidate at, say 30 cents, wins either 1 dollar (a 70 cent profit) or nothing (a 30 cent loss) if the contract is held until market closing following the election. If the trader buys at 30 cents and sells at, say 70, the profit is 40 cents. For further details, consult the IEM website, http://www.biz.uiowa.edu/iem/..
.

[3] http://www.realclearpolitics.com

We measure the vote share predicted by polls in two different ways. First, we consider surprise relative to the average poll data published by the Real Clear Politics website. This average is taken over polls conducted by various polling agencies in the state in question over some period of time before the election. Second, we consider only the latest poll data available before the election (typically taken just one or two days before). If there is more than one such poll closing on the same day, then we take a weighted average with weights proportional to the sample size used in the polls. All raw poll data are taken from Real Clear Politics. There are a number of smaller states and territories where elections are held but for which no poll data is available. These are ignored altogether in constructing the surprise measures (i.e., their contribution to total surprise is set to zero).

For constructing the media variables, we use the data library of the Pew Research Centre Project for Excellence in Journalism (PEJ)[4]. They maintain a continually updated database of news stories in various kinds of media outlets which are monitored at a daily basis in regular intervals during the day. Stories are classified according to topic, lead newsmaker and placement prominence. Word counts for print and web stories and duration-in-seconds for broadcast/cable stories are also counted. For details about the sampling process, see Appendix.

For our study, we selected, from the PEJ daily sample for the relevant date

[4] `http://www.journalism.org`

range (January 1, 2008 to June 15, 2008), the stories that were categorized under the topic 'campaign', had Obama as the lead newsmaker, were news items as opposed to opinion pieces, and had relatively high prominence in their placements. For stories matching all the criteria above, we constructed time series variables representing total number of stories and total word counts per day for print media (Printcount and Wordcount respectively), total number of web stories (Webcount), total number of TV and radio stories (TV and Radio Count) and total duration-in-seconds for TV and Radio stories (Duration).

3.3 Methodology and baseline results

Let P_t denote the day t closing price of the Arrow-Debreau security that pays \$1 if Obama wins the Democratic nomination. As P_t exhibits a marked upward trend over the sample period, we work with the daily return series, denoted $p_t \equiv \log(P_t) - \log(P_{t-1})$, which does not show any apparent sign of being non-stationary. Further, let x_t denote any of the information variables (a measure of the pollster's surprise or a media index) described in Section 3.2.

A fairly narrow (but practical) definition of Granger causality states that x Granger causes p if the linear projection of p_t on $p_{t-1}, p_{t-2}, \ldots$ differs from the linear projection of p_t on the larger information set $p_{t-1}, p_{t-2}, \ldots, x_{t-1}, x_{t-2}, \ldots$ (Granger [1980]). Assuming that the linear projections involved in the definition can be represented by a finite number of lags, one can readily test for causality going from x to p by estimating the model

$$p_t = \alpha_0 + \alpha_1 p_{t-1} + \ldots + \alpha_K p_{t-K} + \beta_1 x_{t-1} + \ldots + \beta_K x_{t-K} + \epsilon_t, \tag{3.1}$$

and conducting a test of the hypothesis $H_0 : \beta_1 = \ldots = \beta_K = 0$.[5] In practice it is of course necessary to fix the lag length K. Given that observations are made at a daily frequency, and day of the week effects might be present, $K = 7$ does not, a priori, seem excessive. We in fact set the upper bound for K at 14 lags and report the results for $K = 1, 2, 3, 7, 10$. While small values of K (e.g., $K = 1, 2, 3$) are not sufficient to ensure that the residuals $\hat{\epsilon}_t$ are approximately white noise, these tests still provide information about which lags of x have the most predictive power, i.e. over what time horizons is Granger causality present.

If one is willing to assume that, for some finite value of K, ϵ_t in equation (3.1) is a martingale difference sequence w.r.t. to its own history and the history of x, then a stronger interpretation of the corresponding test is available. In particular, the systematic part of equation (3.1) will represent the conditional mean of p_t given $p_{t-1}, p_{t-2}, \ldots, x_{t-1}, x_{t-2}, \ldots$, and the test becomes a test of Granger causality in the mean.

To test for causality in the reverse direction, i.e. going from p to x, one can consider interchanging the role of x and p in equation (3.1). However, the

[5] We opt for a quasi-likelihood ratio test (under the normality assumption), which is asymptotically equivalent to a Wald test.

information variables constructed in Section 3.2 are nonnegative, and are given by a sequence of spikes with an interval of zero values in between. A finite order autoregressive model is unlikely to provide a good approximation to the projection of x_t on $x_{t-1}, x_{t-2}, \ldots, p_{t-1}, p_{t-2}, \ldots$. (Interpreting such a model as the conditional mean of x is especially problematic; see Engle [2002]) We circumvent this problem by implementing the Granger causality test as proposed by Geweke et al. [1983]. The method builds on the work of Sims (1972) and consists of regressing p_t on its own history as well as past, current, and *future* values of x:

$$p_t = \alpha_0 + \alpha_1 p_{t-1} + \ldots + \alpha_K p_{t-K} + \beta_1 x_{t-1} + \ldots + \beta_K x_{t-K} + \gamma_0 x_t + \gamma_1 x_{t+1} + \ldots \gamma_K x_{t+K} + \epsilon_t \tag{3.2}$$

Lack of Granger causality from p to x corresponds to the condition $H_0 : \gamma_1 = \ldots = \gamma_K = 0$. Using this equivalent formulation of Granger causality allows us to avoid building a model where x is a dependent variable rather than a regressor. Again, we consider models up to 14 lags/leads and report the results for $K = 1, 2, 3, 7, 10$.

The effective sample used for estimating all models described above ranges from January 15, 2008 to June 1, 2008 and consists of 139 observations. In addition, fourteen pre-sample and fourteen post-sample observations are used to create lags and leads of p_t and x_t. Results for these benchmark Granger causality tests are displayed in Table 3.1 (from p to x) and in Table 3.2 (from x to p).

Table 3.1: Does $\{p_t\}$ Granger cause $\{x_t\}$? Tests of H_0 : NO vs. H_1 : YES.

	Lags (K)				
Variable	1	2	3	7	10
Election surprise (RCP avg.)	0.006	0.002	0.002	0.000	0.000
Election surprise (latest poll)	0.037	0.034	0.037	0.000	0.000
Wordcount	0.019	0.047	0.080	0.047	0.338
Printcount	0.026	0.049	0.112	0.605	0.730
TV and radio count	0.697	0.700	0.689	0.667	0.679
Webcount	0.068	0.132	0.097	0.092	0.236
Duration	0.301	0.288	0.375	0.677	0.691

Note: The reported figures are p-values.

Looking at Table 3.1, there is strong Granger causality going from p_t to both measures of the pollster's surprise. That is, past price changes seem to be useful for predicting future election surprises (relative to polls) even when the history of surprises is taken into account. The result is very robust to the number of lags/leads included in model (3.2). Granger causality in this direction suggests that part of what appears to be a surprise for the pollster is already known to the market. On the other hand, in Table 3.2 we also find strong evidence of Granger causality from the pollster's surprise to future price changes. This suggests that some part of the pollster's surprise is genuine news for the market as well; moreover, it seems to take some time for this new information to be incorporated in the market price.

Regarding the media indices, Table 3.1 displays evidence of short to medium run causality from price changes to some of these variables. In particular, price changes today are informative about wordcount and printcount in the following day or two; in case of the former possibly even longer. As K in-

Table 3.2: Does $\{x_t\}$ Granger cause $\{p_t\}$? Tests of H_0 : NO vs. H_1 : YES.

	Lags (K)				
Variable	1	2	3	7	10
Election surprise (RCP avg.)	0.000	0.000	0.000	0.000	0.000
Election surprise (latest poll)	0.000	0.000	0.000	0.001	0.006
Wordcount	0.401	0.407	0.064	0.654	0.500
Printcount	0.641	0.626	0.111	0.595	0.731
TV and radio count	0.784	0.177	0.292	0.495	0.317
Webcount	0.724	0.610	0.761	0.676	0.812
Duration	0.952	0.417	0.550	0.434	0.415

Note: The reported figures are p-values.

creases, these effects are washed out by the addition of insignificant terms. The rest of the media variables (TV and radio count, webcount and duration) do not seem to respond to price changes conditional on their own history—there is maybe some weak evidence to the contrary for webcount, but none for the other two variables. As shown by Table 3.2, Granger causality in the other direction (from media indices to price changes) is completely missing. Though the evidence is not overwhelming, these findings suggests that the media coverage of the primary elections consists mostly of noise and stories that are already incorporated in the market price.

To sum up, there seems to be strong two-way Granger causality between prediction market prices and the surprise element in the primaries. There is also some evidence of one-way Granger causality in the short run from price changes towards media news indices. These results suggest that prediction market prices anticipate at least some of the discrepancy between the actual outcome and the latest round of polls before the election. Nevertheless, prices also seem to be *driven* partly by election results, suggesting that there is an element of the pollster's surprise that is genuine news for the market as well.

3.4 Sensitivity analysis and extensions

In order to check the robustness of the baseline results presented in Section 3.3 we consider some modifications to the model specifications (3.1) and (3.2), and sketch possible extensions of the basic linear Granger causality test.

The time series of returns on a financial asset often exhibits conditional heteroskedasticty. Casual visual evidence (the plot of p_t) is consistent with this phenomenon, and more formal tests also show some ARCH effects in the residuals of the models (3.1) and (3.2). We therefore repeat the tests described in Section 3.3 while explicitly modeling conditional heteroskedasticity in the errors as a GARCH(1,1) process. The results are displayed in Tables 3.3 and 3.4.

At first glance, explicit modeling of conditional heteroskedasticity changes the numerical results (p-values) quite substantially. Nevertheless, most of the qualitative findings of Section 3.3 remain true. In particular, there is still fairly strong evidence of mutual Granger causality between price changes and measures of the pollster's surprise, though some of the p-values are now higher (see especially the results in Table 3.3 concerning

Table 3.3: Does $\{p_t\}$ Granger cause $\{x_t\}$? Tests of H_0 : NO vs. H_1 : YES. GARCH effects in price model.

	Lags (K)				
Variable	1	2	3	7	10
Election surprise (RCP avg.)	0.004	0.001	0.000	0.000	0.000
Election surprise (latest poll)	0.323	0.447	0.001	0.000	0.000
Wordcount	0.014	0.031	0.033	0.353	0.002
Printcount	0.000	0.000	0.000	0.000	0.000
TV and radio count	0.008	0.003	0.008	0.001	0.000
Webcount	0.251	0.235	0.265	0.140	0.000
Duration	0.003	0.002	0.001	0.001	0.001

surprise relative to the latest polls). Furthermore, as shown by comparing Tables 3.1 and 3.3, accounting for conditional heteroskedasticity in price movements seems to amplify the extent to which price changes are informative about future values of the media indices. Printcount and duration, in particular, are now very strongly Granger caused by price movements.

There is however one rather baffling new effect that appears in Table 3.4. While media variables remain uninformative about future price movements in the short run, it appears that this is not so in the long run (i.e., a week and beyond). This finding is of course very hard to justify theoretically—it not only contradicts the efficient market hypothesis, but also raises the question why there is no short run effect given that there is a long run effect. Further investigation is needed to rule out the possibility that this result is due to a model specification issue.

As discussed in Section 3.3, the concept of Granger causality used in these tests is rather narrow. More specifically, the results can be interpreted, at

Table 3.4: Does $\{x_t\}$ Granger cause $\{p_t\}$? Tests of H_0 : NO vs. H_1 : YES. GARCH effects in price model.

	Lags (K)				
Variable	1	2	3	7	10
Election surprise (RCP avg.)	0.007	0.029	0.034	0.002	0.000
Election surprise (latest poll)	0.013	0.054	0.038	0.018	0.076
Wordcount	0.250	0.269	0.304	0.000	0.000
Printcount	0.168	0.093	0.099	0.002	0.000
TV and radio count	0.796	0.396	0.342	0.007	0.035
Webcount	0.870	0.984	1.000	0.014	0.019
Duration	0.691	0.505	0.472	0.000	0.000

best, as testing for causality in the conditional mean. Finance theory suggests however that the arrival of new information might be associated with higher price volatility. It is therefore of interest to try to test separately for causality in mean and causality in variance between the price change process and the information variables. One possibility is to use the methodology developed by Cheung and Ng (1994); this would however necessitate building an explicit univariate model for the information variables. Extending our research in this direction is work in progress.

3.5 Conclusion

Using data on the race between Barack Obama and Hillary Clinton in the 2008 Democratic primaries for presidential nomination, we investigate whether prediction market prices have information that is not reflected in contemporaneous polls and media stories and conversely, whether prices react to information that appears to be news for pollsters or is prominently featured by the media. We test for Granger causality between day-to-day percent change in prediction market prices and a constructed measure of the surprise element in primary results, i.e information that is not reflected in the polls. We also conduct Granger causality tests between price changes and indices constructed to capture the extent of media coverage received by a candidate.

The main qualitative finding of our exercise, based on the direction of Granger causality found in the data, is that prediction market prices seem to capture some, but not all, of the surprise element in the primary results. Also, there is some evidence that, at least in the short run, prediction market prices capture information that is not reflected in the media. As part of ongoing work, we try to separately test for causality in the conditional variance but this exercise is subject to a number of additional technical difficulties and our preliminary results are quite sensitive to the exact model specification used.

Appendices

Appendix A

Appendix to Chapter 1

A.1 Derivation of the Euler Equations in *Theorem 4*

With the condition that at most one move is feasible given C, we have the value function

$$\begin{aligned}
& V(a_k, s_k) \\
= \; & \max_{t\in[0,\infty]} \max_{a\in A} \mathbb{E}\{[\int_0^{y_{k+1}} u(a_k, s_k)e^{-ru}du + V(a_k, s_{k+1})e^{-ry_{k+1}}]\mathbf{1}_{\{y_{k+1}<t\}} \\
+ \; & [\int_0^t u(a_k, s_k)e^{-ru}du + \int_t^{y_{k+1}} u(a, s_k)e^{-ru}du + V(a, s_{k+1})e^{-ry_{k+1}} - C]\mathbf{1}_{\{y_{k+1}>t\}}\Big|s_k\}
\end{aligned}$$

Now we calculate each of the discounting terms;

$$\mathbb{E}\int_0^y e^{-ru}du\mathbf{1}_{\{y_{k+1}<t\}} = \frac{1}{r}(1-e^{-rt})F_W(t) - \int_0^t F_W(y)e^{-ry}dy$$

$$\mathbb{E}e^{-ry}\mathbf{1}_{\{y_{k+1}<t\}} = \int_0^t e^{-ry}dF_W y = e^{-rt}F_W(t) + r\int_0^t e^{-ry}F_W(y)dy$$

$$\mathbb{E}\int_0^t e^{-ru}du\mathbf{1}_{\{y_{k+1}>t\}} = \frac{1}{r}(1-e^{-rt})(1-F_W(t)$$

$$\mathbb{E}\int_t^y e^{-ru}du\mathbf{1}_{\{y_{k+1}>t\}} = \frac{1}{r}[e^{-rt}\int_t^\infty dF_W(y) - \int_t^\infty e^{-ry}dF_W(y)]$$

$$\mathbb{E}\int_t^\infty e^{-ry}dF_W(y) = -e^{-rt}F_W(t) + r\int_t^\infty e^{-ry}F_W(y)dy$$

Gathering terms, we have,

$$\begin{aligned}
& \frac{1}{r}(1-e^{-rt})F_W(t) - \int_0^t F_W(y)e^{-ry}dy + \frac{1}{r}(1-e^{-rt})(1-F_W(t) \\
= \; & \frac{1}{r}(1-e^{-rt}) - \int_0^t F_W(y)e^{-ry}dy \\
= \; & \int_0^t e^{-ry}dy - \int_0^t F_W(y)e^{-ry}dy \\
= \; & \int_0^t e^{-ry}(1-F_W(t))dy
\end{aligned}$$

Also,

$$\begin{aligned}
& \frac{1}{r}e^{-rt}(1-F_W(t)) - \frac{1}{r}\int_t^\infty e^{-ry}dF_W(y) \\
= \; & \frac{1}{r}e^{-rt}(1-F_W(t)) + \frac{1}{r}e^{-rt}F_W(t) - \int_t^\infty e^{-ry}F_W(y)dy \\
= \; & \frac{1}{r}e^{-rt} - \int_t^\infty e^{-ry}F_W(y)dy
\end{aligned}$$

Moreover,

$$\begin{aligned}
\mathbb{E}[-Ce^{-rt}\mathbf{1}_{\{y_{k+1}>t\}}] &= -Ce^{-rt}\int_t^\infty dF_W(y) \\
&= -Ce^{-rt}(1-F_W(t))
\end{aligned}$$

Replacing the discount terms in the value function and gathering terms, we have

$$
\begin{aligned}
V(a_k, s_k) &= [\int_0^t e^{-ry}(1 - F_W(t))dy]u(a_k, s_k) \\
&+ [e^{-rt}F_W(t) + r\int_0^t e^{-ry}F_W(y)dy]\mathbb{E}V(a_k, s_{k+1}) \\
&+ [\frac{1}{r}e^{-rt} - \int_t^\infty e^{-ry}F_W(y)dy]u(a, s_k) \\
&+ [-e^{-rt}F_W(t) + r\int_t^\infty e^{-ry}F_W(y)dy]\mathbb{E}V(a, s_{k+1}) \\
&- Ce^{-rt}(1 - F_W(t))
\end{aligned}
$$

Taking the derivative of the above expression with respect to t and a and some rearrangement give us the Euler equations.

Appendix B

Appendix to Chapter 3

B.1 Sampling and Coding Methodology of News Database

The following is a brief description of the sampling and coding methodology followed by PEJ in constructing their News Coverage Index data library. For further details, see the PEJ website `http://www.journalism.org`.

The main categories of news sources for the data are as follows: Network TV News, Newspapers, Online News Sites, Cable News and Radio News.

The major broadcast channels ABC, CBS, and NBC make up the broadcast segment. Stories are monitored through different time-slots during the day for 2 out of 3 channels on a rotation basis as follows:

- Commercial Evening News: Entire 30 minutes of 2 out of 3 programs each day (60 minutes)
- Commercial Morning News: 1st 30 minutes of 2 out of 3 programs each day (60 minutes)
- PBS NewsHour: Rotate to code the 1st 30 minutes one day, the 2nd 30 minutes the next day and then skip

This results in either 2 or 2.5 hours of programming each day. Similar method is used on a rotation basis for cable channels CNN, MSNBC and Fox News. During daytime, on a rotation basis, two out of three 30-minute daytime slots each day (60 minutes a day) are coded. During prime time, the following are included

- Two 30-minute segments for Fox News (60 minutes)
- Two 30-minute segments for CNN (60 minutes)
- Two 30-minute segments for MSNBC (60 minutes)

Newspapers are categorized into 3 tiers according to subscription levels, national prominence and regional location. Representative newspapers are chosen from each tier. Here is the list for newspapers in the sample:

- 1st Tier: The New York Times, LA Times, USA Today, Wall Street Journal
- 2nd Tier: Washington Post, Tampa Tribune, Seattle Times, Columbus Dispatch
- 3rd Tier: The Day, Rome News Tribune, Ventura News

For each of the papers selected, only articles that begin on page A1 (including jumps) are picked. This results in a newspaper sample of approximately 20 stories a day.

The websites included in the PEJ sample for our selected date range are as follows: Yahoo News, MSNBC.com, CNN.com, NYTimes.com, Google News, AOL News, Foxnews.com, USAToday.com, Washingtonpost.com, ABC-News.com, HuffingtonPost.com, and Wall Street Journal Online.

For the online news sites, the database captures each site once a day and code the top 5 stories that appear on the site at the time of capture. The time of the day that the sample captures the Web sites is rotated between 9-10 am Eastern time and 4-5 pm Eastern time. The 4-5 PM time-slot was added after April 28, 2008.

The sample of radio stories are collected as follows:

- News: 30 minutes of NPR each day, rotating between Morning Edition and All Things Considered, as broadcast on a selected member station.
- Talk: The first 30 minutes of either one or two talk programs each day. Every weekday, a total of 3 conservatives and 2 liberals were coded during the period of our sample.
- Headlines: Two headline segments each day (one from ABC Radio and one from CBS Radio), about 10 minutes total.

This results in a sample of roughly 1 or 2 hours of programming a day.

Bibliography

A. R. Amar. *America's constitution: a biography*. Random House Trade, 2006. ISBN 0812972724.

K. J. Arrow, R. Forsythe, M. Gorham, R. Hahn, R. Hanson, J. O. Ledyard, S. Levmore, R. Litan, P. Milgrom, and F. D. Nelson. Economics: The promise of prediction markets. *Science*, 320(5878):877, 2008. ISSN 0036-8075.

K.J. Arrow and A.C. Fisher. Environmental preservation, uncertainty, and irreversibility. *The Quarterly Journal of Economics*, 88(2):312–319, 1974. ISSN 0033-5533.

J. Berg, F. Nelson, and T. Rietz. Accuracy and forecast standard error of prediction markets. *Tippie College of Business Administration, University of Iowa*, 2003.

J. Berg, R. Forsythe, F. Nelson, and T. Rietz. Results from a dozen years of election futures markets research. *Handbook of Experimental Economics Results*, 1:742–751, 2008a. ISSN 1574-0722.

J. E. Berg and T. A. Rietz. Prediction markets as decision support systems. *Information Systems Frontiers*, 5(1):79–93, 2003. ISSN 1387-3326.

J. E. Berg, F. D. Nelson, and T. A. Rietz. Prediction market accuracy in the long run. *International Journal of Forecasting*, 24(2):285–300, 2008b.

S. I. Boyarchenko and S. Levendorski. *Irreversible decisions under uncertainty: optimal stopping made easy*. Springer Verlag, 2007.

E. G. Carmines and J. A. Stimson. *Issue evolution: Race and the transformation of American politics*. Princeton Univ Pr, 1990. ISBN 069102331X.

K. Y. Chen and C. R. Plott. Information aggregation mechanisms: concept, design and field implementation. *California Institute of Technology Social Science Working Paper*, 1131, 2002.

P. K. Chintagunta. Inertia and variety seeking in a model of brand-purchase timing. *Marketing Science*, 17(3):253–270, 1998. ISSN 0732-2399.

Z. Elkins, T. Ginsburg, J. Melton, and MyiLibrary. *The endurance of national constitutions*. Cambridge University Press, 2009. ISBN 0511640064.

R. Engle. New frontiers for arch models. *Journal of Applied Econometrics*, 17 (5):425–446, 2002. ISSN 1099-1255.

L.G. Epstein. A definition of uncertainty aversion. *Review of Economic Studies*, 66(3):579–608, 1999. ISSN 1467-937X.

R. S. Erikson and C. Wlezien. Are political markets really superior to polls as election predictors? *Public Opinion Quarterly*, 2008. ISSN 0033-362X. doi: 10.1093/poq/nfn010.

E. Flexner and E.F. Fitzpatrick. *Century of struggle: the woman's rights movement in the United States*. Belknap Press, 1996. ISBN 0674106539.

J. Geweke, R. Meese, and W. Dent. Comparing alternative tests of causality in temporal systems analytic results and experimental evidence. *Journal of Econometrics*, 21(2):161–194, 1983.

C. Gollier, B. Jullien, and N. Treich. Scientific progress and irreversibility: an economic interpretation of the Precautionary Principle. *Journal of Public Economics*, 75(2):229–253, 2000. ISSN 0047-2727.

F. F. Gonul and F. Ter Hofstede. How to compute optimal catalog mailing decisions. *Marketing Science*, 25(1):65–74, 2006.

C. W. J. Granger. Testing for causality:: A personal viewpoint. *Journal of Economic Dynamics and Control*, 2:329–352, 1980.

Darwin C. Hall and Richard J. Behl. Integrating economic analysis and the science of climate instability. *Ecological Economics*, 57(3):442 – 465, 2006. ISSN 0921-8009. doi: 10.1016/j.ecolecon.2005.05.001.

R. Hardin. *Liberalism, constitutionalism, and democracy*. Oxford University Press, USA, 2003. ISBN 0199261687.

D.M. Kreps and E.L. Porteus. Temporal resolution of uncertainty and dynamic choice theory. *Econometrica*, 46(1):185–200, 1978.

J. Madison. Federalist no. 49. *The federalist*, pages 338–347, 1961.

W. D. Nordhaus and J. Boyer. *Warming the world: economic models of global warming*. the MIT Press, 2003. ISBN 0262640546.

T. O'Riordan and J. Cameron. *Interpreting the precautionary principle*. Earthscan/James & James, 1994. ISBN 1853832006.

E. Ostrom. *Governing the commons: The evolution of institutions for collective action*. Cambridge Univ Pr, 1990. ISBN 0521405998.

D. Schmeidler and I. Gilboa. Maxmin expected utility with a non-unique prior. *Journal of Mathematical Economics*, 18(2):141–154, 1989.

J. M. Smith. The republic of letters: The correspondence between thomas jefferson and james madison 1776-1826, 3 vols. *New York: WW Norton*, 3: 1364, 1995.

N. H. Stern and G. B. Treasury. *The economics of climate change: the Stern review*. Cambridge Univ Pr, 2007. ISBN 0521700809.

N. L. Stokey. *The economics of inaction: Stochastic control models with fixed costs*. Princeton Univ Pr, 2009. ISBN 0691135053.

C. R. Sunstein. The paralyzing principle. *Regulation*, 25(4):32–37, 2002.

C. R. Sunstein. Deliberating groups vs. prediction markets (or hayek's challenge to habermas). *Episteme: A Journal of Social Epistemology*, 3(3):192–213, 2007. ISSN 1750-0117.

C.R. Sunstein. *Laws of fear: Beyond the precautionary principle*, volume 6. Cambridge Univ Pr, 2005. ISBN 0521848237.

G. Tziralis and I. Tatsiopoulos. Prediction markets: An extended literature review. *The Journal of Prediction Markets*, 1(1):75–91, 2007.

J. Wolfers and E. Zitzewitz. Prediction markets. *Journal of Economic Perspectives*, 18(2):107–126, 2004.