Dynamic Stochastic Learning in Regulatory Optimization: Deadlines and Error with Continuous and Discrete Evidence

Daniel Carpenter and Justin Grimmer*

April 21, 2007

Abstract

We elaborate a theoretical model of regulatory learning in an environment where the regulator must learn from evidence that unfolds both continuously and discretely, sometimes characterized by rare but stark events. The combination of continuous with discrete data, we believe, represents many of the learning problems with which regulators and risk and safety analysts must engage. We show how the addition of a discrete component to a continuous stochastic process can heighten the rate of error that a regulator might make. We also examine how exogenous deadlines might affect optimal stopping decisions of the sort undertaken by product regulators. The model offers predictions both about the induced stopping behavior of the regulator and the distribution of error under different deadline institutions. We show how flexible deadlines can both accelerate the decision making process and yet induce additional decision error. Simulations illustrate the induced behavior and error predicted in the model.

Introduction

The policy tasks faced by government agencies and regulators often take the form of stopping problems or dynamic optimization problems. The regulator is confronted with an application of sorts - a grant application, a license application, a drug or medical device submitted for approval - and features of its decision may be irreversible. The irreversibility may hold for several reasons. A decision may be technologically reversible in that it induces a set of other events (construction of a subdivision on a property abutting wetlands, removal of a dam, a large capital investment) that are usefully modeled as irreversible. A decision also may be procedurally irreversible because its announcement sets in motion other processes of regulation (rulemaking, the involvement of other actors) or government (expected litigation, judicial review). Often for regulators, decisions are irreversible from the standpoint of reputation. Having made a risky decision, the regulator may not wish to revisit it (whether by reversal or even an investigation) because revisitation will publicize a potentially severe error.[1]

*Department of Government, Faculty of Arts and Sciences, Harvard University. Corresponding author e-mail: *dcarpenter@gov.harvard.edu*. Grimmer e-mail: *jgrimmer@fas.harvard.edu*. Carpenter acknowledges the National Science Foundation (SES-0076452) and a Robert Wood Johnson Foundation Investigator Award in Health Policy Research. Carpenter and Grimmer jointly thank Harvard University (the Center for American Political Studies, the Institutional Development Initiative and Institute for Quantitative Social Science) for support. All arguments, errors and omissions are our sole responsibility.

[1]The assumption of strict irreversibility can be relaxed by imposing revisitation costs or constraints upon a partially reversible decision. We do not investigate any of these possibilities in the confines of this essay.

Despite decades of study in regulatory decision making, students of regulation and government organizations generally lack any sound theoretical guidance about these phenomena.[2] First, and at the the most basic level, we lack models and theories that offer compelling portraits of what these decisions look like. What sort of probabilistic characterizations describe well the stopping decisions of regulators and other government agencies? Second, we lack general guidelines for describing how regulators learn both within and across cases. How would regulators learn about the probable behavior of regulated firms and social organizations, using the cases before them, each of which cases is subject to some uncertainty itself? Third, despite years of research into bounded rationality in organizational and government settings (Padgett 1980, Bendor 1995), we lack insight as to how bounded rationality and other cognitive factors shaping regulatory choice might influence stopping decisions undertaken by regulators. Fourth, how might certain institutions – constraints and frames commonly employed by politicians and other overseers of regulators – shape decision making? We focus on one such institution here, namely the deadline, because it is a common feature of administrative processes and because it is a particularly relevant constraint upon dynamic optimization. Because politicians or bureaucratic superiors wish to limit regulatory delay, they often impose deadlines or one form or another upon agencies. Deadlines may be more or less flexible.

In the context of these persistent questions, we offer a model of regulatory choice in which an uncertain agent must learn and render an irreversible decision about a "case" (for example, a product submitted for regulatory approval). The principal contribution of the model is to imagine repeated regulation of an evidence process that is more general and hence more complicated than is usually posited by analysts of political learning or dynamic stochastic optimization.[3] Whereas the customary analysis of dynamic regulatory behavior proceeds by analyzing a diffusion process (a Gaussian process, a Brownian motion, or a discrete-time random walk), the evidence process about which the regulatory agent must learn in the present model is characterized by *both* continuous and discrete movements. The additive combination of these movements, which we believe more accurately represents the sorts of evidence processes with which "real-world" regulators must deal, renders the modeling problem harder, and it turns out that some of the neater predictions of earlier models of dynamic regulatory behavior do not hold in a more complicated but realistic setting.

1 Parameters and Structure of the Model: Reputation and Irreversibility

The conceptual structure underlying the statistical analysis rests on a model of product review as a stochastic learning process characterized by uncertainty, boundedly rational foresight and costly reversibility.[4] The model is quite generally conceived but as an illustrative (and running) example, it may be applied to the approval regulation of pharmaceuticals. Let products (drugs) be indexed by i, market niches (diseases) by j, and review time by t, with t_{app} denoting an approval time and t_{stop} denoting an optional stopping time or feasible stopping time. The agency wishes to approve products such that their (therapeutic) value outweighs both (1) the harms or risks of the product and (2) the value of waiting for more information.

[2]The empirical literature examining time-to-decision in regulation is small but growing. Analyses include Olson (1997, 2000), Carpenter (2002), Ando (1999) and Kosnik (2005).

[3]The evidence process described here is, for instance, more general and complex than that analyzed by Carpenter (2004), Carpenter and Ting (2005), Volden, Ting and Carpenter (2006).

[4]See Carpenter (2004) for an elaboration of a different model where a problem simpler than that here is repeated and where the regulator keeps an eye on the "pipeline" of future therapies. We do not consider pipeline values in the model here.

1.1 Stochastic Fundamentals and Bayes Equations

The regulator observes the unfolding of evidence on a space Ω (with elements or experimental realizations ω), which is structured by a set of σ-algebras $\Im$, and a probability measure P. In addition, $\Im$ can be ordered and expressed as a filtration $(\Im_t)_{0\leq t\leq\infty}$, which is a family of σ-algebras that is increasing in its index, hence $\Im_s \subset \Im_t$ as long as $s \leq t$. The filtration sequentially collects and orders all realizations $\omega = \omega_t$ on a time dimension from 0 to t. The collection $(\Omega, \Im, \Im_t, P)$ constitutes a filtered probability space, on which we assume that a set of "usual hypotheses" posited by theorists of stochastic differential equations holds. These hypotheses and a relatively clear explanation of their importance appear in Protter (2005: Chapter I, esp. pp. 34-36).

Products are characterized by two parameters - efficacy μ_{ij} and danger λ_{ij} - both of them unknown to the regulator. Observed benefit in regulatory review evolves according to a combined diffusion-jump process $X(t)$. For this (or any other) adapted stochastic process, let $X_{t-}(\omega) = \lim_{s\to t, s<t} X_s(\omega)$, and let $\Delta X_t = X_t - X_{t-}$ for any variable X. $X(t)$ is an additively separable function of a Wiener process (a linear function of underlying efficacy (μ_{ij}) plus a random component $w(t)$) and a negative jump process $J(t)$ which imposes per-failure costs according to a known distribution $G(Z_k|\Delta J_t > 0)$. Formally, $X(t)$ is a special case of a Levy process – which we shall occasionally call the "Levy evidence process" – which obeys the following law of motion:

$$X(t) = \mu_{ij}t + \sigma_{ij}w(t) - \sum_{k=1}^{J_t(\lambda_{ij})} Z_k \tag{1}$$

where σ_{ij} and μ_{ij} are constants and $\sigma_{ij} > 0$, $w(t)$ is a standard normal variable with mean zero and variance t, and λ_{ij} is the rate at which the product imposes costly losses (of size Z). The regulator wishes to learn μ_{ij} and λ_{ij} but observes only $X(t)$ and thus faces what might be called a dual signal extraction problem: there are two parameters to learn about, and information about one of the parameters does not, in and of itself, provide any insight about the other. The problem is made somewhat easier by our assumption that the regulator knows the distribution of "bad" events contingent upon their happening. This distribution is $G(z_k)$ and it is assumed integrable such that the regulator can take expectations over it, with $E(Z_{kt}|\Delta J_t > 0) = IG(z_k)$ with I denoting the relevant integral.

For each product considered, efficacy is given by a drawn from a normal prior distribution $\mu_{ij} \sim N(m_j, s_j)$, and danger is given by a draw from a gamma prior distribution, such that $f(\lambda_{ij}) = [\Gamma(\beta)]^{-1}\alpha^{\beta}\lambda^{\beta-1}e^{-\alpha\lambda}$, where $\Gamma(\dot{})$ is the gamma function. The mean of the normal distribution is m_j, and the mean of the Gamma distribution is $\alpha_j\beta_j^{-1} = \Lambda_j$. The distributions are assumed independent, such that $\text{cov}(\mu_{ij}, \lambda_{ij}) = 0$.[5]

Without loss of generality, the regulator can restrict attention to several Bayes statistics that are sufficient for optimal inference. These are

$$\text{Diffusion Process Posterior Mean} \equiv E_{xt}(\mu_{ij}) = \hat{\mu}_{ijt} = \frac{m_j/s_j + x/\sigma_{ij}^2}{1/s_j + t/\sigma_{ij}^2}$$

[5]This independence assumption is not as restrictive as it might seem. The idea is that if quality and danger were correlated and that correlation were known, then the regulator could always use information from the revelation of quality to reduce her uncertainty about danger, and vice versa. Without loss of generality we can restrict the independence to a conditional form, such that once the regulator takes the prior of quality and danger into account, "surprises" in quality are uncorrelated in expectation with "surprises" in danger.

$$\text{Diffusion Process Posterior Variance} \equiv V(t) = \frac{1}{1/s_j + t/\sigma_{ij}^2}$$

$$\text{Jump Process Posterior Mean} = \frac{\Lambda + J_t}{1+t}$$

A crucial feature of this set-up is that the jumps take place on a set of collective (time)-measure zero. This allows us to divide the stochastic history into continuous and discrete parts. Let X_t^B be continuous part (the Brownian diffusion) and let X_t^J be the discrete part (the jumps).[6] For any optional stopping time, the agent can rearrange the stochastic history into two separable elements: (1) the portion of the process composed by jump movements, and (2) the portion of the process composed by diffusion movements. The idea is then that the agent can apply (possibly different) Bayes equations to each of these elements separately. This requires not only the independence of the processes, but also the assumption (implicit in the Poisson process specification) that the jumps take place on a set of (time)-measure zero. For the Bayes-Chernoff equation to apply to X_t, it is critical that scale-invariance be invoked. Yet while the diffusion process has the scale-invariance property, the aggregate Levy process does not. The measure exclusivity of the jump and diffusion components of the Levy process is sufficient for these purposes, as between jumps the Levy process $X(t)$ has scale-invariance and Gaussian increments, whereas the jumps (which do not obey scale-invariance) have collective measure zero. The rearrangement of the stochastic history is presented in Lemma 1.

Lemma 1: Bayesian Rearrangement and Sufficient Statistics for the Levy Evidence Process $X(t)$. *Let* $\mathbf{F}_t = (\Im_t)$ *represent the filtration for the evidence process* $X(t)$ *as given in (1). Without loss of generality, for any* $X(t)_{t\geq 0}$, *any* $\mathbf{F}_t$ *can be broken into two separable and independent components: (1) the stochastic history of the continuous diffusion,* $\mathbf{F}_t^B$, *and (2) the stochastic history of the jumps* $\mathbf{F}_t^J$. *Then a sufficient statistic for* $\mathbf{F}_t^B$ *is the dual* (t, X_t^B), *where* X_t^B *has full scale-invariance, and a sufficient statistic for* $\mathbf{F}_t^J$ *is the dual* $(t, \sum_{k=1}^{J_t} \mathbf{1}(\Delta Z_k > 0))$.

Proof: Proofs of all lemmata and propositions appear in the Appendix.

While a rather technical feature of the stochastic history, this rearrangement has the intuitive property that the agent can at any time separate the continuous from the discrete movements of the random process being observed, much as a physician could separate more continuous outcome measures such as monthly pain or hypertension measurements from discrete events such as a myocardial infarction or an event that flagged severe hepatotoxicity (Olson 1998; Carpenter 2002). In the review of a dam licensing project by an agency like the Federal Energy Regulatory Commission (FERC) in the United States (Kosnik 2006; Spence 1999), a regulator might separate more continuous measures such as megawatt generation from more discrete outcomes such as failures or environmental snafus and catastrophes. Alternatively, the review of data from nuclear power plants, as part of an inspection or licensing operation, could include more continuous measurements for energy generation and rarer, discrete measurements for safety issues (Gordon and Hafer 2005).

[6] This follows Mordecki (1999). The model differs substantially from Mordecki's in that our agent applies rules of optimization directly to the Bayes-Chernoff filters of the state variables as defined by moments of their respective posterior distributions. In Mordecki and other literature, the optimization programs are applied directly to the untransformed state variables.

1.2 Value Functions and the Regulatory Objective

Let $F(\hat{\gamma})$ be a convex function mapping the Bayesian "filters" of the state variable $X(t)$ into value experienced by the regulator. The problem facing the regulator can be described as the optimal stopping of the filtered evidence process $\hat{\gamma}_t = \hat{\mu}_t - \hat{\lambda}_t \int Z_k G(z_k) dz$, with the following objective (suppressing some subscripts):

$$\sup E_{xt} e^{-\delta(t_{app})} \left\{ A^{-1} + E_{\hat{\mu},\hat{\lambda},t} \int_t^{\infty} e^{-\delta(y-t)} \left[\mu^*(s,\omega) - \lambda^*(s,\omega) \int_{\Re^+} Z_k G(z_k) dz \right] dy \right\}$$

$$= E e^{-\delta(t_{app})} \left\{ A^{-1} + \delta^{-1} \left(\mu^*[t_{app},\omega] - \lambda^*(t_{app},\omega) \int_{\Re^+} Z_k G(z_k) dz \right) \right\} \tag{2}$$

where δ is a discount factor, A is an approval payoff which is static, positive and known with certainty throughout the review, t^{app} is a given approval time, μ^* and λ^* are the agency's efficacy and danger estimates retrieved at the optimal stopping time, ω denotes an elementary event in the space Ω, and y is a variable of integration. The regulator's optimal policy is to observe a first-passage time policy for the relevant state variables, where the threshold (or boundary) combines dynamic optimality with recognition of the irreversibility of any approval decision. Upon approval of any product or case, the values μ_{ij} and λ_{ij} are fully revealed, and utilities ("payoffs") are realized. Proposition 1 states the optimal rule.

Proposition 1: *The product is approved when and only when, and if and only if, the stochastic process $\hat{\gamma}_t$ passes for the first time through the following barrier*

$$\eta^*(\gamma,t) = \frac{1}{2} V(t)^2 F_{\hat{\mu}\hat{\mu}}(\mu,t) - \Lambda \int_{\Re^+} Z_k G(z_k) dz + \delta A^{-1} \tag{3}$$

where $F_{\hat{\mu}\hat{\mu}}$ is the second partial derivative of the value function F with respect to the filtered state variable $\hat{\mu}$, given a realization of $\hat{\mu}$ at time t.

An important property of the regulator's approval barrier is that second-order terms enter only for the continuous component of the Levy process (that is, the Brownian diffusion). This is a consequence of the theory of stochastic integration and local times of Levy processes, in particular the Meyer-Ito Formula (Theorem 7.66 and Theorem 7.70, Corollary 1, of Protter (2005)). The result is convenient for exposition of the present model, as expected effects of different orders are rendered additively separable. As with other analyses of optimal stopping, the second-order term $\frac{1}{2} V(t)^2 F_{\hat{\mu}\hat{\mu}}(\mu,t)$ may be interpreted as the marginal value, evaluated at t, of waiting for more information on the continuous data before making an irreversible approval decision (Carpenter 2004; see also Dixit and Pindyck 1994).

The intuition behind this result is that continuous stochastic processes (as represented here by Brownian motion) generate continual data and have high variation within small time intervals, whereas discrete stochastic processes generate signals only occasionally and have low variation.[7] This conforms to the bifurcated data that regulators often face in their learning

[7] More technically, purely continuous semimartingales such as Brownian motion are examples of an "unbounded variation" process whereas jump processes such as a Poisson process are "finite variation" processes. See Theorem I.27 and II.26 of Protter (2005).

problems, as revelations of "safety" and "danger" are often occasional, whereas revelations of data on quality, price, efficacy and other dimensions of a product or project are often much more continuous on both "space" (measurement) and time dimensions. The result encodes the notion that learning about continuous processes occurs more quickly than learning about the discrete ones.

2 Regulatory Approval Behavior and Deadlines

2.1 The Regulator's Approval Behavior

The key to a stochastic description of the regulator's approval behavior is the fact that discrete events are always negatively valued and cannot induce approval.[8] Good news thus comes incrementally, whereas bad news can come both incrementally and "all at once." Hence approval decisions must always be conditioned on the event that no "bad news" has recently arrived.

For any set of state variables observed at t, let $\Psi^*(t) = \Pr[t_{app} \leq t]$ be the probability of case approval by time t, with $\psi^*(t)$ its associated density. Then we can describe the "hazard" function of the regulator's approval policy as $h^{\Psi^*}(t) = \frac{\psi^*(t)}{1-\Psi^*(t)}$. It can be easily shown that $\Psi^*(t)$ is defective in the sense that $\lim_{t\to\infty} \Psi^*(t) < 1$, unless $(\mu - \lambda) > \delta A^{-1}$. Some products will never be approved, and some regulatory cases will never terminate by means of final resolution.

The first-passage distribution can be expressed as the joint probability that the diffusion component passes through the barrier, given that a certain number of jumps have occurred, and given that before none of these previous jumps was the barrier surpassed.

Lemma 2: *The approval distribution obeys*

$$\Psi^*(t) < \Pr\left[\hat{\mu}_t \geq \eta(t) + \sum_k Z_k | \sup_{s\in\{0,t_{Z_k}\}} \hat{\mu}_s < \eta(s) + \sum_k Z_k\right] \tag{4}$$

The utility of Lemma 2 is that it allows an upper bound on the approval probability to be expressed as the first (upward) passage probability of a continuous diffusion process, given downward movement in the barrier via jumps. Hence the first-passage probability of the Levy evidence process can be approximated by the first-passage probability of the diffusion, where jumps shift the origin of the diffusion downward.

One relatively intuitive version of a joint event which generates case approval is the event that (a) no jump has occurred by a stopping time t_{stop}, and that at t_{stop}, $\hat{\mu}_t$ surpasses $\eta(t)$ for the first time. Here we can imagine the case where the regulator uses the occurrence of a single event by time $t = t_{stop}$ as sufficient evidence that the case should *not* be approved by that time. The probability that, by time t_{stop}, no jump has occurred for a Poisson process with intensity $\lambda^\diamond$ is $1 - \Pr[\text{at least one event occurs by} t] = e^{-\lambda^\diamond t_{stop}}$.

[8]The first-passage time form of the regulator's posited optimum behavior allows for a fuller description of the limiting distribution of the regulator's behavior. The complexity of the evidence process means that except in several relevant and illuminating cases, this description does not allow for neat parametric forms. The reason is that the occasional discrete jumps of the process introduce another form of randomness that must be conditioned upon. While there are distributions describing the first-passage time of Brownian motion (Folks and Chhikkhara 1979), and while the interarrival times of events in the Poisson process are exponentially distributed, the evidence process considered here is a convex combination of a diffusion and jump process. We are aware of no closed form parametric distribution that describes the first-passage time of a convex combination of such stochastic processes through a barrier (linear or otherwise). Hence the elaboration here will rely on functional analysis.

2.2 Deadlines and Deadline-Induced Behavior

Many government agencies, including regulatory bodies, face time constraints of different sorts, some imposed from without and some induced by the structural characteristics of their tasks. In particular, politicians and courts concerned about limiting the delay associated with regulatory processes (whether governmental or private) may impose deadlines for decision making upon the agency. In the case of pharmaceutical regulation in the United States, Congress has imposed "review-time goals" upon the U.S. Food and Drug Administration's Center for Drug Evaluation and Research (CDER), such that it is now expected to act upon ninety percent of all "standard" new drug applications within 10 months or less. In other settings, deadlines are used to constrain the behavior of licensing agencies, product review agencies outside of the United States, and other forms of decision for regulatory agencies.

There are many possible rationales for deadlines, and all of them are exogenous to the model elaborated here. One benefit of deadlines might be that the regulator values time – for example, discounts the future – in a way markedly different from the way that citizens and their representatives do.

We consider flexible deadlines in the form of a bonus payment if the product is approved by the deadline. Imagine a deadline bonus D $(0 < D < \infty)$, which is awarded with certainty to the regulator if and only if she approves the product by an exogenous deadline t^D, s.t. $0 < t^D < \infty$. The finiteness of the deadline bonus means that, in principle, the agency could allow some cases to endure past the deadline, depending upon specific values or evidence encountered in the case. The exogenous imposition of this deadline program transforms the regulator's problem in a rather simple manner. The barrier specified in 3 now takes one of two forms, depending on whether the deadline has elapsed. Before the deadline has been reached, the regulator's adjusted dynamic value function is

$$\forall t \leq t^D, \eta^* (\gamma, t) = \frac{1}{2} V (t)^2 F_{\hat{\mu}\hat{\mu}} (\mu, t) - \Lambda \mathrm{I} (G (y)) + \delta(A^{-1} - D)$$

Whereas after the deadline has elapsed, the barrier resumes the form it takes in equation (3). The regulator's behavior around the deadline is described in Proposition 2.

Proposition 2: *The hazard function $h^{\Psi^*}(t)$ obeys the following two properties near the deadline*

(a) $\lim_{t \downarrow t^D} h^{\Psi^*}(t|t > t^D) = 0$

(b) $\lim_{t \uparrow t^D} h^{\Psi^*}(t|t < t^D) = \sup_{t \in \left[t^D - c, t^D -\right)} h^{\Psi^*}(s)$ *for some* $c \in \Re^+$.

The first statement in Proposition 2 suggests that, given a deadline bonus and a deadline, the probability of case or product approval immediately after the deadline has elapsed is zero. This does *not* imply that the deadline will always be met, but suggests only that, when the deadline is not met, the disappearance of the deadline bonus will delay approval for some time after the deadline has elapsed. The second statement in Proposition 2 suggests that, "right before" the deadline elapses, the approval hazard will be at a local maximum. This Proposition thus generates testable hypotheses if a deadline time – though not necessarily the deadline bonus itself – can be observed.

3 Regulatory Error

The regulator can err for many reasons, and a full elaboration of the possible determinants of error is purposefully excluded from this model.[9] Intuitively, the regulator might apply a biased estimator, might discount the near-term and the long-term at (unreasonably) different rates, might enter the problem with bad priors, might process information inefficiently, or might draw upon its history poorly.

The types of errors made in an inferential setting such as this one depend on the reference point or "null hypothesis" defined. If the null hypothesis is that the case should not be approved – that, for instance, the drug submitted to the FDA is not "safe and efficacious" until proven so – then a Type I error is the rejection of the null by "approval" of the case when the proper decision should have been rejection or withholding of approval. A Type II error would then correspond to acceptance of the null hypothesis when it should have been rejected, that is, the rejection of, or failure to approve, a "good" case. If the null hypothesis is that the case should be approved, then this typology is inverted, with a Type I error corresponding to faulty rejection and a Type II error corresponding to faulty approval. In part due to convenience and consistency with earlier literature, and in part because many if not most regulatory procedures that involve optimal stopping are characterized by the first set-up – assume the case should not be approved until proven so, by some criteria – I adopt the null hypothesis that the case should not be approved until an evidentiary basis has been satisfied, and define Type I and Type II errors accordingly (see also Carpenter and Ting 2005).

The event that the case should or should not have been approved admits of clear description within the terms of the model. The case should not be approved – according to the regulator's own objectives and goals – if the "true" value of the case lies below the value of rejection (or, equivalently, the value of infinite continuation). In terms of the parameters and variables of the present model, this corresponds to the event that $\mu - \lambda I(G(z_k)) < \delta A^{-1}$. Accordingly, the case should be approved if $\mu - \lambda I(G(z_k)) > \delta A^{-1}$.[10]

Given the probabilistic description of regulatory behavior that emerges from Proposition 1 and Proposition 2, we can then state the probabilities of Type I and Type II error as follows. Let Φ^I denote the probability of Type I error given the null hypothesis that the case should not be approved until proven effective, valid or otherwise "good." Define Φ^{II} as the corresponding probability of Type II error. Then

$$\Phi^I = \Pr\left[\sup_{t\in[0,\infty)} \hat{\gamma}_t \geq \eta(t) | \mu - \lambda I(G(z_k)) < \delta A^{-1}\right] \tag{5}$$

and

$$\Phi^{II} = \Pr\left[\sup_{t\in[0,\infty)} \hat{\gamma}_t < \eta(t) | \mu - \lambda I(G(z_k)) > \delta A^{-1}\right] \tag{6}$$

It merits remark here that under this definition, either Type of error can, in principle, happen randomly, that is, according to chance alone. In particular, it can be shown that the probability

[9] These different "causes" of error can be thought of as useful extensions to the model.

[10] We will not deal with the case of equality here, as it is a knife-edge occurrence that has measure zero. It is sufficient to state a tie-breaking rule such that the drug should be approved if $\mu - \lambda I(G(z_k)) = \delta A^{-1}$. As it turns out, the case of equality is a non-trivially difficult one to analyze in terms of dynamic stochastic movements.

of Type I error is always non-zero in finite time, and given the current set-up of the model, there is no way of avoiding Type I error if the regulator's preferences and constraints allow it any discretion whatsoever.

Error from False 'Initial Beliefs.' When regulators begin with overly optimistic or pessimistic beliefs about a case before them, then learning error can result because the regulator's initial beliefs ("priors") inappropriately drop or hike the hurdle of evidence that must be surpassed in order for the case to receive approval. One benefit of the present model is that it permits erroneous priors to be described in flexible and understandable terms. The regulator may enter the problem with a pessimistic "first guess" as to the efficacy or quality of the case, which we can describe with an inflated M_j, or the regulator may begin with an overly optimistic sense of when discrete adverse events are likely to occur, which corresponds to a depressed Λ_j. At first glance, the error in initial quality (M) would seem to tradeoff equally with the errors in initial danger (Λ). When Λ_j is sufficiently large, this is an accurate understanding.

Yet Λ_j describes an initially expected rate of events that are, by their assumed nature, rare. (If they were not rare and discretely occurring then the utility of using a jump or Poisson process to model would be in doubt.) To the extent that it encodes intuition, then, we should think of Λ_j as "small." When this representation is considered, then there exists a real possibility that the bias for M_j (the prior of μ_{ij}), being linear and unbounded, may be of an order different from that of the bias for Λ (the prior of λ_{ij}). Recall that μ_{ij} is normally distributed (with mean M_j), but that λ_{ij} is distributed as a Gamma variable. Both the Normal and the Gamma distributions are two parameter distributions, but in the Normal case the expectation (or mean, or first moment) of the distribution is a function of only one of these parameters. In part because the Gamma often represents small, strictly positive quantities, its expectation is a function of two parameters. Given $\alpha_j > 0$ and $\beta_j > 0$, the expectation of the Gamma distribution is $\frac{\alpha_j}{\beta_j}$.

A Bayesian regulator sets the origin point of the evidence process at the value of the prior for the parameters being learned. Let the origin point of $\hat{\gamma}(= \hat{\gamma}_0)$ as Γ. The regulator might erroneously guess this prior by believing in a $\Gamma^{est} = \Gamma^{true} + \xi$. If this is true, then the evidence process will commence at too "low" or too "high" a level. When in particular the bias on Γ^{est} is positive ($\xi > 0$), a "bad" product (for which $\gamma < \delta A^{-1}$) may be more likely approved because the evidence process starts out ever closer to the barrier and can trip over it randomly with higher probability. We can think of $\xi > 0$ as reflecting "optimism" about the case. In the following discussion, we show that increasing regulatory optimism yields a high probability of Type I errors.

The linkage between optimistic priors and the introduction of a jump process comes in the fact that human agents often inflate or deflate the actual probability of 'rare' events (**citation**). A family setting out to travel 400 miles to visit relatives might wish to drive because they believe that the chance of a place crash is higher than it really is, while they might also believe that the probability of a fatal automobile accident in the same interval is lower than it really is. The Gamma-distributed prior for λ_j encodes these possibilities. Because $E[\lambda_j] = \frac{\alpha_j}{\beta_j}$, first-order movements in β_j can induce higher-order movements in Λ_j. Intuitively, this means that human error in initial guesses about rare events is of greater magnitude than human error in initial guesses about continuous events. Proposition 3 demonstrates three ways in which seeminly innocuous movements in the prior can generate Type I error by inducing optimism.

Proposition 3: Let the set of comparison and counterfactual approvals have equivalent histories ($\Im_{it}(X)$ is the same for all i). Then the **(a)** *For equivalent unidirectional parametric movements in mean parameters of quality (m_j) and danger parameter (β_j), Φ^I is weakly increasing in β_j.*

(b) *For equivalent opposite-directional parametric movements in mean parameters of quality (m_j) and danger parameter (β_j), Φ^I is weakly increasing in β_j over a finite interval.*

(c) *For equivalent unidirectional parametric movements in the secondary parameters of the quality (s_j) and danger parameter (β_j), Φ^I is weakly decreasing in β_j for "quick" approvals but is weakly increasing in β_j for "longer" approvals.*

The intuition of Proposition 3 is that "movements" (equivalently, "errors") in the initial case parameters are of different orders. When both the quality and (inverse) danger are increased by equal amounts, the error in danger is hiked more quickly than the error in quality, and the origin point of the evidence process $\Gamma^{est} = \hat{\gamma}_{t=0}$ will increase disproportionately. Figure Z4 demonstrates a sample relationship. In statement (b), the increases in (inverse) danger are offset by reductions in the quality distribution prior, and this can dampen the overall origin point of the evidence process, as illustrated in Figure Z5. Finally, statement (c) suggests that when both the second parameters for both the normal distribution (quality) and the Gamma distribution (danger) are increased, the probability of Type I error is reduced during early review, but increases for longer reviews. This is a result of the Meyer-Ito Lemma and the second-order valuation of continuous data. Early in the review (when approval is least likely anyway), first-order increases in s_j inflate $\eta(t)$ by second-order movements, but as the review proceeds and approval becomes more likely, the reduction in squared posterior variance ($V(t)^2$) dilutes this effect and the influence of a higher prior Γ^{est} dominates.

4 Deadline-Induced Error

Deadlines may induce error because they stop the learning process and create a potentially "artificial" bonus which reduces the barrier so that it is more easily reached. Without deadlines and with "long continuation," the evidence process would with certainty reach the true estimates of μ_{ij} and λ_{ij}. However, the evidence process might probabilistically surpass the barrier in the process, hence even under an 'optimal' procedure some amount of Type I error is to be expected. The key is that deadlines may increase the probability of this error. With deadlines and the deadline bonus D_j, the approval barrier is more easily tripped even if $\gamma_{ij} < \delta A^{-1}$.

In considering institutions such as deadlines, it merits consideration that they may influence both the 'credit' and the 'debit' side of the ledger. The 'credit' side of the ledger can be thought of as the acceleration of an approval for a case or product that would eventually have been approved anyway. One useful feature of the present model is that it permits a simple expression of some of the "costs" and "benefits" of deadline institutions. The debit side of the ledger comes from the Type I errors induced by the deadline bonus. In the following Proposition, we demonstrate two features of the interaction between deadline and error.

Proposition 4: **(a)** *For any set of cases with equivalent histories ($\Im_{it} \equiv \Im_t \forall i, t$), the probability of Type I error Φ^I_{it} is weakly increasing in D.*

(b) *Given any D, and for any set of cases with equivalent histories, the introduction of jump process to diffusion process generates more error. Monotonically, Φ^I_{it} is weakly increasing in $\int_{\Re^+} Z_k G(z_k) dz$.*

Proposition 4 allows us to parameterize the error probability by the expectation over the jump distribution. At the simplest level, this permits a demonstration that the introduction of a jump process into the information space creates more problems for deadlines. The "baseline" case of Brownian motion alone is simply the case where $E[Z_k] = 0$, and any jump process for which $E[Z_k] > 0$ introduces discreteness into the evidence process.

Just because error is induced does not mean that the deadlines are net negative in terms of welfare, because the benefits from speeding up regulatory decisions may outweigh the costs of added error. The benefit can be quantified by thinking of the set of cases that would eventually be approved $\gamma_{ij} \geq \delta A^{-1}$ but which are approved before the deadline and would otherwise have been approved afterwards. The value differential for this event is

$$F^{diff} = \exp\left[-\delta\left(t^{D}_{app,i=1} - t^{D=0}_{app,i=2}\right)|\gamma \geq \delta A^{-1}, \Im_{1t} \equiv \Im_{2t} \forall t\right]$$

When this exceeds the value lost from higher Type I error, then the deadline can be said to be welfare improving within the constraints of the model. However, if there are other costs to deadines (in that the benefits of the case are perhaps dependent upon the amount of time taken to learn about its parameters), or if there are other benefits to deadlines (in that the deadlines perhaps induce grater efficiencies by the regulator that spill over to other activities), then the policy calculation of the present model is inadequate and will fail to capture these benefits and costs.

5 Conclusion and Possible Empirical Applications

The combination of continuously unfolding evidence and discrete events is common in many forms of regulation and risk analysis. We have shown that the introduction of a jump process to a continuous evidence process makes two important differences for regulatory review: (1) learning about diffusion is quicker (a result which is encoded in the Meyer-Ito lemma), but (2) errors in initial guesses about jump process are more likely. The model thus encodes properties of data, as well as intutions about the bounded rationality of human agrnts. These patterns can interact – as in Proposition 3(c) – yielding even more counter-intuitive predictions.

The model has several empirical implications that deserve testing. The first of these, following from Proposition 2, is that deadlines of the sort discussed here should be observed to introduce discontinuties into the hazard function of regulatory decisions. Second, following from Proposition 3, observable error should be an increasing function of the discreteness of the data. Rarer adverse events should be subject to greater regulatory error – both Type I and Type II – because initial guesses about these processes are more easily erroneous. Finally, following from Proposition 4, error rates should be conditioned upon deadline times and deadline bonuses. Cases approved immediately before a deadline are more likely to experience Type I error than cases approved after the deadline. The higher the deadline bonus, moreover, the higher the rate of Type I error. This effect may be offset by the quicker review of cases that would have been approved in any case.

Another set of implications concerns the methods for studying the duration of regulatory decisions. Whereas learning processes characterized by continuous diffusions alone induce parametric forms for the distributions governing regulatory decision time (the inverse Gaussian, Carpenter 2002, Carpenter 2004), our model predicts a hazard function which is both non-parametric and violates the proportional hazards assumptions often used in semi-parametric analysis of duration data (for example, the Cox model). We leave analysis of a more appropriate estimator to another paper. For now, suffice it to say that the introduction of greater realism into the evidence processes about which regulators learn yields much greater complexity of behavior than is appreciated or embedded in standard statistical models.

Two limitations of this modeling framework strike us as ideal targets for thoughtful extension. First, a primary determinant of time-to-decision in organizational settings is not simply the duration of "optimal stopping," but organizational features such as the queue of cases coming

to the agency. These flows may depend upon strategic considerations (Carpenter and Ting 2005), while in other cases factors such as organizational efficiancy and the number of organizational units reviewing cases may also be influential (Bendor 1985, Heimann 1993, Ting 2002). Such queueing processes have been well studied in stochastic analysis, but to our awareness, models which embed *both* stopping behavior and queues have not been attempted. Second, it is quite possible that much of regulatory review amounts not simply to optimal stopping but also to "optimal control," in that the eventual quality or danger of the case may in fact depend upon the amount of time that the regulator or risk analyst has spent learning about it. Whenever this is true, the signal extraction metaphor governing our model leaves much to be desired. We suspect that introduction of queues and case-based optimization will yield rich theoretical progress, and until this happens, the conclusions of the present analysis should be taken with circumspection.

APPENDIX

Proof of Lemma 1. First we define $\Delta X(t) \equiv X(t) - X(t-)$ where $X(t-) \equiv \lim_{s \to t, s \leq t} X(s)$ and let $\mathcal{T}$ denote the set of stopping times.[11] During the continuous portions of the process $\Delta X(t) = 0$, while during the instants of the jumps $|\Delta X(t)| > 0$. Define $\Delta X^J(t) \equiv \{t \in \mathcal{T} || \Delta X(t)| > 0\}$, or the set of all times where a jump occurs. Similarly define, $\Delta X^B(t) \equiv \{t \in \mathcal{T} | \Delta X(t) = 0\}$, or the set of time where the process is continuous. Note, that these two sets are disjoint,which implies that the information arrives at different times. Now, take the $\Im_t^B \equiv \{\Im_s | s \in \Delta X^B(t)\}$ and $\Im_t^J \equiv \{\Im_s | s \in \Delta X^J(t)\}$, and note that these two sets are disjoint ($\Im_t^B \cap \Im_t^j = \emptyset$). By assumption, the arrival of jumps are independent of the draws from the Brownian motion, which implies that the σ-algebras are also independent, therefore, $\Im_t^B \perp \Im_t^J$ (Billingsley 1980). Note that $\Im_t^B$ and $\Im_t^J$ contain all the relevant information about Equation 1 at moment t. Further, every Levy Process is a Markov process, therefore the only information relevant to the future of the process is the current state of the process (Protter 2005, Theorem I. 32, p 23); Applebaum 2004, pp. 71, 121). It follows directly that $(t, X(t))$ is a sufficient statistic to summarize $\Im_t^B$ and that $(t, \sum_{k=1}^{J_t} Z_k)$ is sufficient to summarize the information in $\Im_t^J$ (Chernoff 1968).

Proof of Proposition 1 The regulator's dynamic stochastic optimum is a time- and moment-dependent function, derived from the following Bellman-Hamilton-Jacobi equation with "omicron term" o representing terms of order less than t as $dt \to 0$.

$$\delta F(\hat{\gamma}, t) = E\left[dF\right] = \left\{E_{\hat{\mu}, \hat{\lambda}, t} F(\hat{\gamma}(t + dt)) - F(\hat{\gamma}(t))\right\} + o(dt) \tag{7}$$

To express the differential more concretely, we use the Meyer-Ito formula (Theorem 4.70 of Protter 2005: 218-220), which holds for any convex f. The scale invariance of the diffusion component, combined with the measure-exclusivity of the continuous and discrete components of the Levy efficacy process, allows us to apply the Meyer-Ito formula not only to $X(t)$ (as in 1), but directly to the transformed state variables expressed as posterior moments. This possibility is stated formally in Lemma 1. Then

$$f(\hat{\gamma}_t) - f(\hat{\gamma}_0) = \int_0^\infty f'(\hat{\gamma}_{s-}) d\hat{\gamma}_s + \frac{1}{2} \int_0^\infty f''(\hat{\gamma}_{s-}) d\left\langle \hat{\gamma}, \hat{\gamma} \right\rangle_f + \sum_{0 \leq s \leq t} \left(f(\hat{\gamma}_s) - f(\hat{\gamma}_{s-} - f'(\hat{\gamma}_{s-}) \Delta \hat{\gamma}_s\right) \tag{8}$$

[11] A stopping time in this instance is the set of all time, given that the σ-algebra under consideration is the completed natural σ-algebra of the stochastic process.

Define $\kappa = \kappa(\omega, d\hat{\gamma}, ds)$ as the jump measure for $X(t)$ (or $X^J(t)$), and define $\nu = \nu(d\hat{\gamma}, ds) = \lambda ds F(d\hat{\gamma})$ as its compensator. Using the measure-exclusivity of the diffusion and the jump process, we may express the first-order terms of the Meyer-Ito equation as

$$\int_0^\infty f'(\hat{\gamma}_{s-}) d\hat{\gamma}_s + \sum_{0 \leq s \leq t} \left(f(\hat{\gamma}_s) - f(\hat{\gamma}_{s-} - f'(\hat{\gamma}_{s-}) \Delta \hat{\gamma}_s \right)$$

$$= \int_0^\infty f'(\hat{\mu}_{s-}) d\hat{\mu}_s + \int_0^t \int_{\Re} \{ f(\hat{\gamma}_{s-} + \hat{\gamma}_0) - f(\hat{\gamma}_{s-}) \} \times (\kappa(\omega, d\hat{\gamma} \times ds) - \nu(d\hat{\gamma} \times ds))$$

$$+ \int_0^t \int_{\Re} \{ f(\hat{\gamma}_{s-} + \hat{\gamma}_0) - f(\hat{\gamma}_{s-}) \} \times \nu(d\hat{\gamma} \times ds))$$

Define the infinitesimal generator of X_t (equivalently, of $\hat{\gamma}_t$) by $\left(L^{\hat{\gamma}} f \right)(w, z)$. With these results, we can express the differential of the Bellman-Hamilton-Jacobi equation as

$$E[dF] = \int_0^t \left(L^{\hat{\gamma}} f \right)(\hat{\gamma}_{s-}) \langle ds \rangle + Q(w, z)_t + o(dt)$$

where Q is a local martingale, defined as follows:

$$Q(f)_t = V(t) \int_0^t f'(\hat{\gamma}_{s-}) d\hat{\mu}_s + \int_0^t \int_{\Re} (f(\hat{\gamma}_{s-} + \hat{\gamma}_0) - f(\hat{\gamma}_{s-})) \times (\kappa - \nu)$$

By an important supermartingale property, $E[Q(f)_t] \leq 0$, and by the value-matching condition, the relation is one of equality (pure martingale status) at t^*_{app}. Invoking the Ito-Meyer Lemma, independence, and dividing through by the differential and taking limits as the differential vanishes yields a second-order stochastic differential equation for the infinitesimal generator of X_t, $\left(L^{\hat{\gamma}} f \right)(w, z)$:

$$\left(L^{\hat{\gamma}} f \right)(w, z) = F_{\hat{\mu}}(w, t) + F_t(w, z, t) + \tfrac{1}{2} V(t)^2 F_{\hat{\mu}\hat{\mu}}(w, t) - \hat{\lambda} \int_0^t (f(x + y) - f(x)) dG(y)$$

As shown by Miroschnichenko (1975), the term $F_{\hat{\mu}}$ is a mean-zero martingale. To solve the second-order stochastic differential equation, it is sufficient to note that any upward barrier is passable only by continuous movements and not discrete ones (see also Alili and Kyprianou 2004; Kyprianou and Surya 2005). Hence a combination of smooth pasting and value matching conditions will suffice to characterize the dynamic stochastic optimum (as in Carpenter 2004), and using (2) for a characterization of F, yields

$$\delta \left[\mu \delta^{-1} - \lambda \delta^{-1} - A^{-1} \right] = \frac{1}{2} V(t)^2 F_{\hat{\mu}\hat{\mu}}(w, t) - \Lambda \mathrm{I}$$

Proof of Lemma 2: The approval distribution may be expressed as the cumulative distribution of a potentially infinite series of events in which the continuous component of the Levy evidence process passes through a conditionally-adjusted non-linear barrier, as follows

$$\Pr[\hat{\gamma}_t \geq \eta^*(t)] = \Pr[\{\hat{\mu}_t \geq \eta(t) - \Lambda I(G(z_k))\} \& \{\text{no jumps in } [0, t)\}]$$

$$+\Pr\left[\{\hat{\mu}_t \geq \eta(t) - \Lambda I(G(z_k))\}\&\{\text{1 jump in } [0,t)\}|\sup_{s\in\{0,t_{Z_1}\}}\hat{\mu}_s < \eta(s) + Z_1 - (1+s)\Lambda I(G(z_k))\right]$$

$$+\Pr\left[\{\hat{\mu}_t \geq \eta(t) - \Lambda I(G(z_k))\}\&\{\text{2 jumps in } [0,t)\}|\sup_{s\in\{0,t_{Z_2}\}}\hat{\mu}_s < \eta(s) + Z_1 + Z_2 - (1+s)\Lambda I(G(z_k))\right]$$

$$+...$$

It is easily shown that the probability terms converge to zero as the number of jumps k gets large. Given the independence of the jump process from the Brownian diffusion, the left hand side of each probability term can be rewritten as $\Pr\left[\hat{\mu}_t > \eta(t) + \sum_k Z_k - (1+t)\Lambda I(G(z_k))\right]$, and the first-passage probability can be reduced to

$$\Psi^*(t) = \Pr\left[\hat{\mu}_t > \eta(t) + \sum_k Z_k - (1+t)\Lambda I(G(z_k))|\sup_{s\in\{0,t_{Z_k}\}}\hat{\mu}_s < \eta(s) + \sum_k Z_k - (1+s)\Lambda I(G(z_k))\right] \tag{9}$$

Equation (4) of the Proposition is achieved by subtracting the compensator term from the argument and the conditioning event.

Proof of Proposition 2: We begin with the proof of statement (a). Assume that as X_t and $\hat{\gamma}_t$ reach $\tau = t^D - s$ that the product has not yet been approved ($\sup_{t\in[0,\tau)}\hat{\gamma}_t < \eta(t)$). Then for any τ and X_τ and $\sum_k Z_{k\tau}$, the probability of approval between τ and the deadline t^D can be written as

$$\Psi^*(\tau, t^D) = \Pr\left[\sup_{t\in[\tau,t^D)}\hat{\gamma}_t > (\eta(t^D) - \eta(\tau))|\hat{\gamma}_\tau = 0\right] \tag{10}$$

$$\geq \Pr\left[\sup_{t\in[\tau,t^D)}\hat{\mu}_t > (\eta(t^D) - \eta(\tau))|\hat{\mu}_\tau = 0, \sum_{k,t\in[\tau,t^D)} Z_k = 0\right] \tag{11}$$

The difference between the quantities in (10) and (11) is very small. As the difference between τ and t^D gets small, the relation approaches equality because the probability of more than one jump in the interval approaches zero more rapidly than the interval itself does. Since statements (a) and (b) of the Proposition invoke limits as t tends to t^D, we use (11) to characterize the hazard rate h^*. In this case, given no jump, the hazard of case approval is a function of the first-passage time for the continuous component of the Levy process through the barrier. As shown in Karatzas and Shreve (1991: 95-6, 196-7), this induces an inverse Gaussian form such that, for any origin point X_0 and any barrier $B > X_0 \geq 0$,

$$\Pr\left[t_{app} \in ds\right] = \frac{(B - X_0)}{\sqrt{2\pi s^3}} e^{-(B-X_0)^2/2s} ds \tag{12}$$

Now immediately after t^D has elapsed, the barrier $\eta(t)$ rises by the quantity D as the deadline bonus disappears. Now imagine the event where the case's evidence ($\hat{\gamma}_t$) came "closest" to approval (within ϵ of reaching the barrier, where ϵ is small (the epsilon of analysis), such that $\sup_{t\in[0,t^D)}\hat{\mu}_t < (\eta(t))$ but $\sup_{t\in[0,t^D)}\hat{\mu}_t + \epsilon > (\eta(t))$. Then for any small time-interval of size ι, equation (12) can be written as

$$\Pr\left[t_{app} \in ds | t = t^D + \iota\right] = \frac{D+\epsilon}{\sqrt{2\pi s^3}} e^{-(D+\epsilon)^2/2s} ds \tag{13}$$

Equation (13) is an expression of the approval density $\psi^*(t|t > t^D)$, hence the hazard $h^*(t)$ is strictly increasing in this quantity. By the structure of the inverse Gaussian form, at the "beginning" of post-deadline review $\psi^*(t|t = t^D + \epsilon) = 0$, hence $h^*(t|t = t^D + \epsilon) = 0$. But because this is the closest case (ϵ is small), and because of the monotonicity of (13), any case that was not "as close" (with ϵ larger) would take even longer (or have lower probability of immediate approval). Formally, $\frac{\partial \psi(\eta(t))}{\partial \epsilon} > 0$. Hence the result is proved.

For (b), note that $\psi^*(0) = 0$, hence $h^*(0) = 0$. So for $c = (0, t^D)$, the result is trivially true, which is sufficient to demonstrate that there exists a c for which the statement is true. For a more exact result, let τ be the time of the last downward jump ($\tau = s, s.t. Z_{ks} > 0 \& \sum_{s=\tau}^{t^D} Z_{ks} = 0$). For any jump of size $Z_{k\tau}$, the approval density can be written as

$$\Pr\left[t_{app} \in ds\right] = \frac{Z_{k\tau}}{\sqrt{2\pi s^3}} e^{-Z_{k\tau}^2/2s} ds \tag{14}$$

Fixing τ and $Z_{k\tau}$, it is always possible to define c so that Proposition 2(b) holds. QED.

Proof of Proposition 3. We begin by showing that the probability of Type I error (Φ^I) is increasing in the origin point Γ. Let the error in Γ^{est} be $\xi > 0$. Let $i = 1, 2$ be any two products and without loss of generality let the prior for case 1 be erroneously optimistic such that $\Gamma_{1,j}^{est,\xi>0} - \xi \equiv \Gamma_{2,j}^{est,\xi=0}$. (By transitivity the set of such products can be countably infinite.) The assumption of Proposition 3 is equivalent to the statement that the evidence process for 1 and 2 is identical, net of the error in priors, hence $\gamma_{1,j}^{\xi>0} - \xi \equiv \gamma_{2,j}^{\xi=0}$. This means that

$$\Im\left(X_{1,j}\left(t\right)|\gamma_{1,j}^{\xi>0}\right) \equiv \Im\left(X_{2,j}\left(t\right)|\gamma_{2,j}^{\xi=0}\right), \forall t$$

Although the products' histories are identical their regulation is not, as

$$\hat{\gamma}_t^{\xi>0}\left\{\Im\left(X_{1,j}\left(t\right)|\gamma_{1,j}^{\xi>0}\right)\right\} > \hat{\gamma}_t^{\xi=0}\left[\Im\left(X_{2,j}\left(t\right)|\gamma_{2,j}^{\xi=0}\right)\right], \forall t < \infty$$

By Lemma 1 and Lemma 2 we can separate the histories into continuous and discrete components, such that the movements in the continuous components of the two evidence processes are rendered comparable, as follows

$$\hat{\mu}_t^{\xi>0}\left\{\Im\left(X_{1,j}^c\left(t\right)|\mu_{1,j}^{\xi>0}\right)\right\} > \hat{\mu}_t^{\xi=0}\left[\Im\left(X_{2,j}^c\left(t\right)|\mu_{2,j}^{\xi=0}\right)\right], \forall t < \infty$$

where X^c is the continuous (Brownian motion) component of X. Begin with the case of 'equality,' letting $\mu_{ij} = \delta A^{-1}$. Then by scale invariance, the distribution governing the first passage of the process $\hat{\mu}_t$ through the barrier $\eta(t)$ is equivalent to the distribution governing the first passage compensated process $Y_t^c = X_t^c - \mu_{ij} t$ through the barrier $\theta\delta A^{-1}$, where $\theta > 0$ is a scalar parameter reflecting the option value of the untransformed problem. This scenario is that in which the case is "closest" to "bad" ($\mu_{ij} < \delta A^{-1}$). Then by the reflection principle (Karatzas and Shreve 1991, Proposition 2.8.1) the probability that, by any stopping time t_{stop}, the running maximum of Y_t^c has hit the absorbing barrier $\theta\delta A^{-1}$ is

$$P^0\left[\sup_{0<s<t_{stop}} Y_t^c \geq \theta\delta A^{-1}\right] = \frac{2}{\sqrt{2\pi}} \int_{[\theta\delta A^{-1} - Y_{t=0}^c]/\sqrt{t_{stop}}}^{\infty} e^{-s^2/2} ds$$

This probability is increasing in the origin of the compensated diffusion $Y_{t=0}^c = \Gamma_j - \Lambda_j \int_{\Re^+} Z_k G(z_k) dz$. For any "worse" case, the drift of X_t^c will be further downward ($\mu_{ij} < \delta A^{-1}$), but the approval probability will still be an increasing function of Γ_j^{est}. QED

We can then restate each statement [**(a)- (c)**] of the Proposition as a claim about the origin of the evidence process Γ_j^{est}. For statement **(a)**, note that $\Gamma^{est} = m_j + \frac{\alpha_j}{\beta_j}\int_{\Re^+} Z_k G(z_k)dz$. Whereas $\frac{\partial^2\Gamma^{est}}{\partial\mu^2}$ and $\frac{\partial^2\Gamma^{est}}{\partial\alpha^2}$ do not exist, $\frac{\partial^2\Gamma^{est}}{\partial\beta^2} = \frac{-2\alpha\int_{\Re^+} Z_k G(z_k)dz}{\beta^3}$. Define m_j', β_j' and ρ such that we can represent unidirectional movements in the quality and safety parameters of Γ as $\frac{\partial\Gamma^{est,\rho^+}}{\partial\rho} = \frac{\partial}{\partial\rho}\left[\left(m_j'+\rho\right) - \frac{\alpha_j\int_{\Re^+} Z_k G(z_k)dz}{\beta_j'+\rho}\right] = 1 + \frac{\alpha\int_{\Re^+} Z_k G(z_k)dz}{(\beta_j'+\rho)^2}$. Hence for any m_j' and β_j' there always exists a $\rho > -\beta_j'$ such that $\Gamma_j(m_j', \alpha, \beta_j'+\rho) > \Gamma_j(m_j', \alpha, \beta_j') + \rho$. This is demonstrated in Figure Z4.

For statement **(b)**, the identical second derivative results hold, but now we seek values of m_j', β_j' and ρ such that $\frac{\partial\Gamma^{est,\rho^-}}{\partial\rho} = \frac{\partial}{\partial\rho}\left[\left(m_j'-\rho\right) - \frac{\alpha_j\int_{\Re^+} Z_k G(z_k)dz}{\beta_j'+\rho}\right] = \frac{\alpha\int_{\Re^+} Z_k G(z_k)dz}{(\beta_j'+\rho)^2} - 1$ is positive. This is true for any $\beta_j' < \sqrt{\alpha\int_{\Re^+} Z_k G(z_k)dz} - \rho$.

For statement **(c)**, an increase in s_j' to $s_j + \rho$, accompanied by an increase in β_j' yields an increase in the posterior variance to $V(t) = \frac{1}{1/(s_j'+\rho)+t/\sigma_{ij}^2} \Rightarrow V(0) = s_j' + \rho \Rightarrow \eta(0) = \delta A^{-1} + F_{\hat{\mu}\hat{\mu}}(s_j'+\rho)^2$.

Now for any counting process $J(t)$ and jump distance $\sum_{k=1}^{J_t(\lambda_{ij})} Z_k$, the agent can by Lemma 1 separate the discrete movements in X from the regulatory history and compute the difference in continuous histories as

$$\hat{\mu}_t^{\xi>0}\left[\Im_t\right] - \hat{\mu}_t^{\xi=0}\left[\Im_t\right] = \frac{(m+\xi)/s + x_{1t}/\sigma^2}{1/s + t/\sigma^2} - \frac{m/s + x_{2t}/\sigma^2}{1/s + t/\sigma^2} = \xi\sigma_i^2\left(\sigma_i^2 + st\right)^{-1}$$

Then product 1 is approved with higher probability than product 2 whenever

$$\left[\delta A^{-1} + V_1\left(t_{stop}\right)^2\right] - \left[\delta A^{-1} + V_2\left(t_{stop}\right)^2\right] \leq \hat{\mu}_{t_{stop},1}^{\xi>0}\left[\Im_t\right] - \hat{\mu}_{t_{stop},2}^{\xi=0}\left[\Im_t\right]$$

Hence a sufficient condition for earlier approval for product 1 is

$$\xi \geq \frac{\sigma_i^2 + st_{stop}}{\sigma_i^2}\left[V_2\left(t_{stop}\right)^2 - V_1\left(t_{stop}\right)^2\right]$$

The difference in posterior variances declines by a second-order factor as t_{stop} increases, hence the condition fails "early" (t_{stop} smaller) and holds "late" (t_{stop} larger).

Proof of Proposition 4: For statement **(a)**, it is sufficient to restate the Gaussian form for error probability used in the proof of Proposition 3. With the addition of a deadline bonus, that probability is

$$P^0\left[\sup_{0<s<t_{stop}} Y_t^c \geq \left[\theta\delta(A-D)^{-1}\right]\right] = \frac{2}{\sqrt{2\pi}}\int_{\left[\left[\theta\delta(A-D)^{-1}\right]-Y_{t=0}^c\right]/\sqrt{t_{stop}}}^{\infty} e^{-s^2/2}ds$$

For **(b)**, note that $\frac{\partial\Gamma^{est}}{\partial\beta} = \frac{\alpha\int_{\Re^+} Z_k G(z_k)dz}{\beta^2}$. Where there is no jump process ($E[Z_k] = 0$), this derivative is zero and there is no error induces from inflated priors for Γ. By inspection the second-order error in priors from movements in β_j is increasing in $\int_{\Re^+} Z_k G(z_k)dz$, which functions as a weight on Λ_j.

References

Alili, L., and Andreas Kyprianou. 2004. "Some Remarks on First Passage of Levy Processes, the American Put and Smooth Pasting," *Annals of Applied Probability* **15**:2062-80.

Ando, Amy Whritenour. 1999. "Waiting to be Protected under the Endangered Species Act: The Political Economy of Regulatory Delay," *Journal of Law and Economics* **42**: 29-60.

Carpenter, Daniel. 2002. "Groups, the Media, Agency Waiting Costs, and FDA Drug Approval," *American Journal of Political Science* **46** (July 2002)(3): 490-505.

Carpenter, Daniel. 2004. "Protection without Capture: Product Approval by a Politically Responsive, Learning Regulator," *American Political Science Review* **98**(4):613-31.

Carpenter, Daniel, and Michael M. Ting. 2005. "Regulatory Errors under Two-Sided Uncertainty," manuscript, Harvard University, and RWJF Scholars in Health Policy Working Paper.

Chernoff, Herman. 1968. "Optimal Stochastic Control," *Sankya – The Indian Journal of Statistics* Series A **30** (3): 221-52.

Karatzas, Ioannis and Steven E. Shreve. 1991. *Brownian Motion and Stochastic Calculus* Sevcond Edition (New York: Springer).

Kosnik, Lea-Rachel D. 2005. "Sources of Bureaucratic Delay: A Case Study of FERC Dam Relicensing," *Journal of Law, Economics and Organization* **22**(1) 258-88.

Kyprianou, Andreas E. and Budhi A. Surya. 2005. "On the Novikov-Shiryaev Optimal Stopping Problems in Continuous Time," *Electronic Communications in Probability* **10**:146-54.

Miroschnichenko, T. P. 1975. "Optimal Stopping of an Integral of a Weiner Process," *Theory of Probability and Its Applications* **9**(4) 355-9.

Mordecki, Ernesto. 1999. "Optimal Stopping for A Diffusion with Jumps," *Finance and Stochastics* **3**:227-36.

Mordecki, Ernesto. 2002. "Optimal Stopping and perpetual options for Levy processes," *Finance and Stochastics* **6**:473-93.

Novikov, A. and A. N. Shiryaev. 2004. "On an effective solution of the optimal stopping problem for random walks," *Theory of Probability and Its Applications* **48**

Olson, Mary K. 1997. "Firm Characteristics and the Speed of FDA Approval," *Journal of Economics and Management Strategy* **6**:377-401.

Olson, Mary K. 2000. "Regulatory Reform and Bureaucratic Responsiveness to Firms: The Impact of User Fees in the FDA," *Journal of Economics and Management Strategy* **9**:363-95.

Protter, Philip E. 2005. *Stochastic Integration and Differential Equations*, Second Version, Version 2.1 (Berlin: Springer-Verlag).

Spence, D. 1999. "Managing Delegation Ex Ante: Using Law to Steer Administrative Agencies," *Journal of Legal Studies* **28**:413-59.