

# When Should Policymakers Make Announcements?\*

Ricardo Reis  
Columbia University

September 2011

## Abstract

If a policymaker learns today that there will be a regime change in the future that affects everyone, at what time between now and then should he/she announce it to the public? This paper presents a dynamic model where agents have a limited amount of attention to allocate between learning about the new regime and everything else. They trade off the benefit of being better informed and making better decisions when the new regime arrives against the cost of paying less attention to current events and making worse decisions now. By choosing when to make the announcement, the policymaker can affect this decision. The policymaker also takes into account that later announcements are more precise, and that agents may inefficiently put too much weight on public signals due to strategic complementarities. I solve for the optimal timing of announcements and the conditions under which it is preferable to keep mum in spite of public interest. As a by-product, I characterize the life-cycle of attention following an announcement, before and after the regime changes.

---

\*I am grateful to Alberto Alesina, Marios Angeletos, Steve Davis, Robert Shimer, Chris Sims, and Ivan Werning and several seminar participants for useful comments. Neil Mehrotra provided excellent research assistance. Contact: rreis@columbia.edu.

# 1 Introduction

Economists usually focus on designing optimal policies, but policymakers are just as worried with their implementation and, in particular, with communicating them effectively. In the field of public policy, it is understood that the success of a policy intervention depends as much on the details and rules of this policy, as it does on the ability to inform the community and to co-opt its involvement and support (Kaid, 2004). One area of economics where more attention is being paid to communication is monetary policy. The adoption of inflation targeting by many central banks has led to new questions for researchers, such as how often should the central bank announce changes in targets, whether to speak in one voice or many, or how transparent should the central bank be.

This paper answers one very specific question on the theory of policy communication: what is the optimal timing for policy announcements? In particular, consider the following scenario: a policymaker learns (or decides) today that at a future date an aggregate change will happen. The policymaker can communicate this to the public right away, at the date of the event, or at anytime in between. However, it can only do so imperfectly as communication is always noisy and subject to differences of interpretation. When should the policymaker make the announcement?

In the model, agents can choose to invest to learn more about the regime change, but at the cost of paying less attention to current conditions. Each agent faces a private trade-off between becoming better informed about the future and responding to the change better when it arrives, versus being better informed about conditions today and making better decisions today. Because future benefits are discounted, there is an optimal cut-off date at which agents start paying some attention to learn more beyond the announcement. In the model, agents endogenously choose what to pay attention to in two ways. First, in the traditional Bayesian sense, the informativeness of signals gets weighted by their relative precision. Second, agents endogenously choose these relative precisions subject to an information capacity constraint, following Sims (2003).

On top of this private trade-off, the policymaker takes two other considerations into account when choosing the optimal timing of the announcement. First, as the date of the change gets closer, the policymaker's signal becomes more precise. This could be either

because the policymaker learns more about the change, or it can communicate it better, or because it makes progress in reaching a final decision about the change. Delaying an announcement therefore delays agents' learning but gives them a more precise starting point.

Second, if agents' actions are strategic complements, then I show that their decision to acquire information also has strategic complementarities. Therefore, they may put too much weight on the policy announcements as they permit coordination of actions and expectations. In this case, there may be an interval of time when agents find it privately advantageous to pay attention to the announcement, but this is socially inefficient.

This paper is structured as follows. The remainder of this section describes applications of the theory, and explains the contributions to the literature. Section 2 presents the model and discusses its assumptions. Section 3 solves for the equilibrium life-cycle of attention and section 4 for the optimal date of announcements. Section 5 considers several extensions and section 6 concludes.

**Applications.** There are several applications to this theory of the optimal timing of announcements and the trade-offs that it focuses on. For instance, consider a central bank that decides today that in a few months it will start disinflating the economy: when should it make this decision public? Similarly, imagine that the central bank learns that conditions in financial markets will change and that an "exit strategy" is in order: when should it announce this strategy? In both cases, getting the attention of agents, as well as making sure that it is not wasted before it is necessary, and that agents do not over-react, as is emphasized in the model, are typical concerns for policymakers. A closely related problem is how often should central bank governors meet and make decisions, and the model in this paper provides some of the ingredients to study it.

In the field of social policy, elected politicians often wait for the right time to announce what is their priority for the coming months. Gaining political support and undermining political rivals are perhaps their most important goal (more on this in the literature review), but making sure the message is "effective," that it "draws a crowd," and that the politician becomes "the face behind the policy" are often mentioned in choosing the date of the announcement, pointing to attention as a relevant consideration (Cobb and Elder, 2001, McNair, 2007). Moreover, important changes in the provision of public services are

frequently done ahead of time to give citizens enough time to adjust to the changes. The model in this paper provides a particular model of sluggish adjustment of citizens, here driven by attention and learning, justifying the early announcement that policymakers feel is important.

A final application of the model in this paper is to the law. Typically, after a law is approved, there is a lag of time before it is applied. Part of the reason for it is the need to adjust directives and applications, and another part to allow those to whom the law applies to get ready for it. In both cases, if the model of attention in this paper is understood more broadly as capturing all of the necessary adjustments to apply a new law, it can also be applied to supply some insights on what is the optimal lag between the voting of a law and it coming into effect.

**Related literature and summary of contributions.** This paper makes contributions to four existing literatures in economics. First, there is a growing empirical literature on the communication of monetary policy. In a recent survey, Blinder et al. (2008) summarize it as follows: “Over the last two decades, communication has become an increasingly important aspect of monetary policy. These real-world developments have spawned a huge new scholarly literature on central bank communication—mostly empirical and almost all of it written in the last decade.” This paper contributes to the theory of central bank communication by focussing on the question of when should central banks announce their actions.

Following the important work of Morris and Shin (2002), a few papers have characterized scenarios under which policy may want to be less than fully transparent because of strategic complementarities inducing an over-weighting of public signals.<sup>1</sup> In most of these papers the choice of releasing information is essentially static as the policymaker must choose *whether* to release the information or not, once and for all. In this paper, instead, the question is *when* to release the information, and the focus is on the dynamic trade-offs involved. Moreover, in this paper, the precision of signals is an endogenous choice variable, whereas

---

<sup>1</sup>Hellwig (2005), Roca (2006), Morris et al. (2006), Svensson (2006), and Lorenzoni (2010) discuss this result when agents learn from signals on exogenous variables, while Amador and Weill (2009) show that this is also possible when agents learn from the distribution of prices. Angeletos and Pavan (2007) characterize more generally when can releases of public information be socially harmful.

in this literature private agents take signals as given. Relative to the general analysis in Angeletos and Pavan (2007), one important difference is that the focus of this paper is not on the equilibrium choice of private actions with imperfect information, but rather on the equilibrium choice of attention. Myatt and Wallace (2008, 2009) also focus on the choice of attention, but they take a very different approach by letting agents choose whether to pay attention to public or private signals of a common fundamental.

Eusepi and Preston (2010) provides a theory of optimal communication by central banks with private agents that are less than fully rational. There are three key differences between this paper and theirs. First, their model of agent's expectations is least-squares learning, whereas in this paper agents are Bayesian learners with rational inattention. Second, they focus on the stability and determinacy of equilibrium, whereas this paper focus on optimal policy according to social welfare. Third, they ask *what* variables and coefficients the central bank should communicate, whereas this paper again asks *when* to communicate. Morris and Shin (2007) provide another theory of optimal communication as a trade-off between more precise information and common understanding. This paper shares with theirs the important role played by strategic complementarities and coordination. However, their analysis is static, while the question in this paper is inherently dynamic. Moreover, the key choice studied in their paper is how to allocate policy signals across the population, whereas the key choice in this paper is on privately allocating attention.

This paper also provides a contribution to the literature on rational inattention.<sup>2</sup> The model builds on Sims (2003) and Mackowiak and Wiederholt (2009), although they typically focussed on interior solutions to the attention allocation problem, whereas the main focus of this paper is on the corner solution where agent choose to completely ignore public announcements. Van Nieuwerburgh and Veldkamp (2010) also focus on cases where rationally inattentive agents choose to completely ignore some information, but their application is on asset pricing. Moreover, in this paper the focus is on how agents choose to gradually become informed. Most of the literature on rational inattention is set in discrete time (exceptions are Moscarini (2004) and Kasa (2006)), whereas the model in this paper is in

---

<sup>2</sup>Recent surveys of this literature are Sims (forthcoming), Mankiw and Reis (forthcoming) and Veldkamp (2010).

continuous time. An interesting result that should be useful to future model of rational inattention in continuous time is that, in spite of their apparent complexity, they reduce to optimal control problems without uncertainty that are part of the standard tool kit of economists. The particular optimal control problem that arises is interesting in its own right and may be useful in other applications.<sup>3</sup>

Finally, while this paper approaches the question of when to release information from the perspective of agents' limited attention, there are surely other considerations at play. The delay between policy announcements and their implementation may be due to a modified version of the war of attrition mechanism in the seminal paper of Alesina and Drazen (1991). As in Dewatripont and Roland (1992) and Dewatripont and Roland (1995), policy reforms may proceed gradually in order to gain support from different groups. Moreover, the optimal transparency in communication may be related to credibility and reputation as in Moscarini (2007) or Athey et al. (2005). These papers do not speak directly to the optimal timing of announcements that is the focus of this paper, but it would be useful to develop them in this direction in future research. The same applies to the model of bounded rationality of Bolton and Faure-Grimaud (2010), where agents take time to deliberate their current and future decisions. While they do not address the question of timing policy announcements, their model would provide a limited-rationality alternative to the limited-attention story put forward here.

## 2 The model

Time is continuous, indexed by  $t \in \mathfrak{R}_0^+$ . At date 0, a single policymaker learns that at date  $T > 0$  there will an aggregate shock that she may want to transmit to a continuum of private agents, indexed by  $i \in [0, 1]$ .

---

<sup>3</sup>The optimal control problem in this paper can be described as follow. Imagine a person who has the view from her window blocked by two trees, that grow exogenously at different rates. The person has a fixed amount of effort every instant that she can devote to chop down the two trees. How should she split her effort between the two trees over time, and how does their height optimally evolve?

## 2.1 Agents and their payoffs

Each private agent  $i$  wishes to maximize:

$$W_i = \mathbb{E}_{i0} \left[ \int_0^\infty e^{-\rho t} [a_{it} - \omega u_{it} - (1 - \alpha)a_t - \alpha r_t]^2 dt \right]. \quad (1)$$

Agents discount the future at rate  $\rho > 0$ , and  $\mathbb{E}_{it}(\cdot)$  denotes their expectations conditional on their information at date  $t$ .<sup>4</sup> Each agent chooses actions  $a_{it} \in \mathfrak{R}$  every instant, as well as the information set on which their expectations are formed, in order to track the movements in three variables,  $a_t$ ,  $r_t$ , and  $u_{it}$ , suffering a quadratic loss when actions deviate from target. As is common in many macro models, this loss function can be seen as a second-order approximation to a more general objective function.<sup>5</sup>

The exogenous individual-specific stochastic processes  $u_{it}$  capture idiosyncratic shocks. They are independent from other shocks, pairwise independent across agents with mean zero,  $\int u_{it} di = 0$ , and follow a stationary Ornstein-Uhlenbeck process:

$$u_{it} = u_{i0} - \eta \int_0^t u_{is} ds + \sqrt{\phi} B_t, \quad (2)$$

for  $t > 0$  with an initial condition  $u_{i0}$  that has mean zero and is normally distributed independently of  $B_t$ . The scalars  $\eta > 0$  and  $\phi$  capture the persistence and volatility of the process, respectively, and  $B_t$  is a standard Brownian motion.<sup>6</sup> These shocks are meant to capture all of the individual circumstances that each agent must pay attention to.

The other exogenous process  $r_t$  denotes the policy regime. At date 0,  $r_0 = 0$  and all agents know it, but the policymaker learns that at date  $T > 0$ , the aggregate regime will

---

<sup>4</sup>To be more precise, all random processes are defined in a probability space  $(\Omega, \mathcal{F}, \mathbb{P})$ , where  $\Omega$  is the space of outcomes,  $\mathcal{F}$  is the  $\sigma$ -field of events (or the measurable subsets of  $\Omega$ ), and  $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$  is a probability measure. A stochastic process (or random variable)  $X_t : \Omega \rightarrow \mathfrak{R}$ , that has a subscript  $t$ , is understood to be adapted to the sub-filtration  $\mathcal{F}_t$  with an associated measure  $\mathbb{P}_t$  and its expectation is defined as  $\mathbb{E}_t(X_t) = \int X_t(\omega) d\mathbb{P}_t$ . Each agent observes a different set of events (or signals) leading to individual-specific  $\mathcal{F}_i$  and therefore likewise individual specific  $\mathcal{F}_{it}$ ,  $\mathbb{P}_{it}$ , and  $\mathbb{E}_{it}(X_t)$ .

<sup>5</sup>One example of this problem is of monopolistic price-setting firms to track changes in productivity ( $u_{it}$ ), the price level ( $a_t$ ) and the money supply ( $r_t$ ), as in new Keynesian models. Another example is that of firms choosing how much to invest in their individual-specific technology given idiosyncratic productivity ( $u_{it}$ ), aggregate productivity ( $r_t$ ), and other firms' investment choices ( $a_t$ ) if there are externalities in capital, as in endogenous growth models with spillovers.

<sup>6</sup>One can show that  $u_{it}$  is normally distributed with mean equal to  $u_{i0}e^{-\eta t}$  and variance equal to  $(1 - e^{-2\eta t})\phi^2/2\eta$ . So, as  $t$  goes to infinity,  $u_{it}$  tends to a normal variable with mean zero and variance  $\phi^2/2\eta$ .

change from  $r_t = 0$  for  $t < T$ , to  $r_t = r \neq 0$  for  $t \geq T$ . Bar an announcement, agents have only a prior that with probability  $1 - p$  the regime will stay the same, whereas with probability  $p$  there will be a new regime with prior distribution  $N(0, \sigma^2)$ . This change in the policy regime may correspond to something external to the policymaker that she has learned about (e.g., a shock to money demand), or to a decision that she has made (e.g., a change in money supply). The focus of this paper is not on the change per se, which I will take as given, but rather on communicating it subject to only being able to make noisy announcements.

The final factor affecting the choices of each agent is the aggregate behavior of all:

$$a_t = \int_0^1 a_{it} di \tag{3}$$

The parameter  $\alpha \in [0, 1]$  determines the strength of strategic complementarities in this economy. If  $\alpha = 0$ , agents only care about the policy regime via what other agents are doing, and this becomes an extreme beauty game. If  $\alpha = 1$ , then the actions of others are unimportant, and actions are strategically independent. In between, the smaller is  $\alpha$  the stronger are strategic complementarities. The coefficient in front of the aggregate actions and the policy regime add up to one so that the parameter  $\omega > 0$  governs the relative weight of idiosyncratic versus aggregate conditions in agents' payoffs.

## 2.2 Announcements

When the policymaker makes an announcement at date  $\tau \in [0, T]$ , agents learn that there has been a change in regime. However, the policymaker cannot communicate perfectly what the new regime is. One intuitive justification for this imperfect communication is that almost any government announcement gets reported differently by different media, and is interpreted subjectively in even slightly different ways by different people. Another, more formal justification is that transmitting the exact value of a real number precisely requires an infinite flow of information to send the infinite number of digits that characterize it.<sup>7</sup>

I model this by assuming that each agent receives an unbiased idiosyncratic normal

---

<sup>7</sup>In contrast, transmitting that the regime has changed requires only one bit of information, a Yes/No answer to a question.



signal  $z_{i\tau}^r$  of the new regime  $r$ . Therefore, right after the announcement, agent  $i$  perceives the new regime to be:

$$r \sim N((1 - e^{-\mu\tau})z_{i\tau}^r, \sigma^2 e^{-\mu\tau - \mu_0}). \quad (4)$$

The parameter  $\mu_0$  measure the information about the value of  $r$  that is contained in an announcement at date 0. More importantly, the parameter  $\mu > 0$  measures the rate at which the informativeness of signals increases as we get closer to the regime change.<sup>8</sup> I assume that the later the announcement is made, the more precise it can be, for several reasons. First, future circumstances may be hard to describe to people but become less remote as the time draws nearer. Second, the source of noise may partly reflect the policymaker's inability to express herself, and as it gets nearer to the event she can communicate it more accurately. Third, it may be that as we get closer to the date of a possible regime change, the public channel through which communication flows between the policymaker and the agents becomes better as it gets a larger share of media focus.

At this point, and for the remainder of this and the next section, I make two further assumptions.

**Assumption 1:** *The policymaker can only issue one signal at one date and  $\mu_0 = 0$ .*

**Assumption 2:**  *$p \rightarrow 0$ , so before an announcement, agents believed the regime would never change.*

The first assumption is not important; section 5.2 allows for multiple signals and for arbitrarily large initial informativeness of announcements. The second assumption is more important. It implies that, absent an announcement, agents expend no attention on learning about the new regime. Relaxing it adds a new set of considerations into the problem. I defer to section 5.3 to study them more carefully.

### 2.3 Information and constrains to attention

Private agents can choose to allocate attention to either learning about the regime or to learning about the individual circumstances. In particular, each agent's information on the

---

<sup>8</sup>For readers more used to seeing this expression in terms of relative variances of signal and noise, if the announcement leads to a signal  $z_{i\tau}^{pr} = r + \varepsilon_{i\tau}^{pr}$  where  $\varepsilon_{i\tau}^{pr} \sim N(0, \varkappa_\tau^2)$  then  $\mu = \ln [(\sigma^2 + \varkappa_\tau^2)/\varkappa_\tau^2] / \tau$ .

vector of fundamentals  $\mathbf{f}_t = (r_t, u_{it})$  is composed of error-ridden signals  $\mathbf{z}_{it}$ . The signals are unbiased but contaminated with additive Gaussian noise:

$$dz_{it}^r = r_t dt + \sqrt{\theta_{it}^r} dB_{it}', \quad (5)$$

$$dz_{it}^u = u_{it} dt + \sqrt{\theta_{it}^u} dB_{it}''. \quad (6)$$

where  $B_{it}'$  and  $B_{it}''$  are standard Wiener processes, independent across any pair of  $i$ , independent of each other, and independent of  $B_t$ .

Crucially, the noise in the two channels at every instant,  $\theta_{it}^r$  and  $\theta_{it}^u$ , are choice variables of each private agent.<sup>9</sup> The lower each of these is, then the more precise is the information revealed about their associated fundamental. Therefore, I refer to the choice of these parameters as the choice of how much attention to pay to either aggregate regime or individual circumstances. The particular structure of the signals assumed above fits into the “rational inattention” literature as section 5.1 will discuss at length,

There is also an attention constraint that prevents the agent from ever learning everything, and that involves a trade-off between learning about the aggregate regime or the individual circumstances. I measure attention using entropy, and its constraint as a limited amount of joint information between the signal and the fundamentals. More precisely, letting  $\mathbf{f}^t = (\mathbf{f}_s, 0 \leq s \leq t)$  and  $\mathbf{z}_i^t$  likewise, and denoting their joint distribution by  $\mathbb{P}_{\mathbf{fz}}$ , and their respective marginals as  $\mathbb{P}_{\mathbf{f}}$  and  $\mathbb{P}_{\mathbf{z}}$ , the constraint on the informativeness of the signals is the Shannon constraint on the reduction in entropy in going from marginal to conditional:

$$\int \ln \left( \frac{d\mathbb{P}_{\mathbf{fz}}}{d(\mathbb{P}_{\mathbf{f}} \times \mathbb{P}_{\mathbf{z}})} \right) d\mathbb{P}_{\mathbf{fz}} \leq kt. \quad (7)$$

The parameter  $k > 0$  is the constant rate of information transmission per instant of time.<sup>10</sup>

---

<sup>9</sup>Note that agents receive a signal on  $u_{it}dt$  rather than on  $du_{it}$ . It is a fundamental result in information theory that, if signals were of the latter form, appropriate coding could transmit the fundamentals perfectly.

<sup>10</sup>Using the change in entropy as a measure of information has become common across a few fields. In information theory, it is often called mutual information, and has become the most common measure of the information that a channel can transmit. In statistics, it is equal to the Kullback-Leibler distance between the joint distribution and the product of the marginals. In complexity theory, the expected average Kolmogorov complexity of  $n$  i.i.d. draws of a random variable from a distribution converges to the entropy of that distribution as  $n$  goes to infinity.

## 2.4 The equilibrium and the announcement problem

Because all private agents are infinitesimal and identical at date 0, and since their problems are linear in the fundamentals and signals, I focus on competitive, symmetric, linear equilibrium.

**Definition 1:** A linear symmetric competitive equilibrium for all  $t \in \mathfrak{R}_0^+$  under assumptions 1 and 2 is a set  $\{a_t, \theta_t^r, \theta_t^u\}$  such that, given an announcement date  $\tau \in [0, T]$ :

(i) each private agent chooses  $\{a_{it}, \theta_{it}^r, \theta_{it}^u\}$  to maximize (1) subject to the information constraints at every date (7), given the signals in (5)-(6) and the initial uncertainty on  $r$  in (4), taking  $a_t$  as given.

(ii) aggregate actions are defined by (3).

(iii) all agents behave identically:  $\{a_{it}, \theta_{it}^r, \theta_{it}^u\} = \{a_t, \theta_t^r, \theta_t^u\}$ .

(iv) The aggregate action is a linear function of the aggregate fundamentals.

The answer to the question posed in the title and introduction then is:

**Definition 2:** The announcement problem is to choose  $\tau \in [0, T]$  to maximize social welfare  $E_{P0} \left( \int_0^1 W_i di \right)$  subject to the information of the policymaker at date 0, and within the set of competitive equilibrium.

The next section solves for the equilibrium and the following one solves for the announcement problem.

## 3 The competitive equilibrium

The equilibrium actions in this linear-quadratic economy is relatively standard. The problem of allocating attention is new.

### 3.1 Equilibrium actions

Maximizing (1) with respect to  $a_{it}$  gives the standard certainty-equivalence result that each agent chooses its action equal to the expectation of the variables it wants to track:

$$a_{it}^* = (1 - \alpha)\mathbb{E}_{it}(a_t) + \alpha\mathbb{E}_{it}(r_t) + \omega\mathbb{E}_{it}(u_{it}). \quad (8)$$

This requires an expectation of the aggregate actions taken by all. Since  $r_t$  is the only aggregate variable, I conjecture that in the linear equilibrium:

$$a_t = \gamma_t r_t, \tag{9}$$

where  $\gamma_t$  is a deterministic process that agents take as given. Then, because for  $t < T$ , the aggregate regime was  $r_t = 0$ , it follows that  $a_t = 0$ .

After the regime changes at  $T$ , using the conjecture and integrating over all agents, the equilibrium aggregate action is

$$a_t = [(1 - \alpha)\gamma_t + \alpha] \int \hat{r}_{it} di. \tag{10}$$

where I use the notation  $\hat{r}_{it} = \mathbb{E}_{it}(r)$ . The Kalman and Bucy (1961) filter provides the optimal forecast of the aggregate regime after the announcement date  $\tau$ :

$$d\hat{r}_{it} = \frac{\nu_{it}^r}{\theta_{it}^r} (dz_{it}^r - \hat{r}_{it} dt). \tag{11}$$

With the initial condition in (4), it follows that

$$\int \hat{r}_{it} di = \delta_t r, \text{ where } \delta_t \equiv 1 - e^{-\mu\tau - \int_{\tau}^t (\nu_s^r / \theta_s^r) ds} \tag{12}$$

Combining equations (9), (10) and (12) verifies the guess and gives the solution for  $\gamma_t$  for  $t \geq T$ :

$$\gamma_t = \frac{\alpha \delta_t}{1 - \delta_t + \alpha \delta_t}. \tag{13}$$

There are two noteworthy features of this equation. First,  $\gamma < 1$  so there is always incomplete reaction to the regime change. Since the agent cannot devote infinite capacity to learning about the aggregate regime, equilibrium actions will always be in between the zero prior and the actual new value for the regime.

Second, note that because agents behave competitively, they take  $\gamma_t$  as given when they allocate their attention. However, the more attention (lower  $\theta_s^r$ ) they pay to the aggregate regime, the better their signal about it, and so the more weight they put on it in choosing

their actions. In turn, this raises the equilibrium value  $\gamma$  from equation (13), and because this raises the variability of aggregate actions, it increases the importance of knowing the new regime in the individual's payoff function as long as  $\alpha < 1$ , and it will affect the choice of  $x_1^*$ . This payoff externality therefore makes paying attention to the policy regime more costly socially than it is privately, and affects the allocation of attention.

### 3.2 The attention allocation problem

Turning to the problem of allocating attention, at date  $T$  agents are aware that there has been a change in the regime. Letting  $\nu_{it}^r = \mathbb{E}_{it}(r - \mathbb{E}_{it}(r))^2$  and  $\nu_{it}^u = \mathbb{E}_{it}(u_{it} - \mathbb{E}_{it}(u_{it}))^2$ , the posterior variances, and using the solution for optimal actions in (8) to re-express (1) gives the objective function:

$$V = - \int_T^\infty e^{-\rho(t-T)} [\omega^2 \nu_{it}^u + [(1 - \alpha)\gamma_t + \alpha]\nu_{it}^r] dt. \quad (14)$$

At  $T$ , there is a pair of initial conditions  $v_{iT}^r$  and  $v_{iT}^u$ . From then onwards, the evolution of the variances is described by a standard application of the Kalman and Bucy (1961) filter:

$$\frac{d\nu_{it}^r/dt}{\nu_{it}^r} = -\frac{v_{it}^r}{\theta_{it}^r} \text{ and} \quad (15)$$

$$\frac{d\nu_{it}^u/dt}{\nu_{it}^u} = \frac{\phi}{\nu_{it}^u} - 2\eta - \frac{v_{it}^u}{\theta_{it}^u}. \quad (16)$$

To gain some intuition about these expressions, note that the smaller is the attention devoted to the aggregate regime (higher  $\theta_{it}^r$ ), then the slower the variance  $\nu_{it}^r$  falls. The same applies to  $\nu_{it}^u$  and  $\theta_{it}^u$ , with the extra additions that the arrival of new idiosyncratic shocks raises variance every instant by  $\phi$ , while the variance falls faster for more predictable process (higher  $\eta$ ).

Letting  $x_{it}$  and  $y_{it}$  denote the signal-to-noise ratios with respect to the aggregate regime and the idiosyncratic fundamentals, respectively, it is useful to keep in mind that variances cannot be negative:

$$\frac{v_{it}^r}{\theta_{it}^r} \equiv x_{it} \geq 0 \text{ and } \frac{v_{it}^u}{\theta_{it}^u} \equiv y_{it} \geq 0 \quad (17)$$

When  $x_{it} = 0$  this corresponds to an infinite variance of the signal, that is no attention devoted to learning about the regime change at all. When  $x_{it} = 2k$ , then the signal on the aggregate regime is as precise as possible while there is no new information about the individual circumstances. In between, the larger  $x_{it}$  the more attention is devoted to the aggregate regime and the more precise are the signals on it.

Finally, taking into account the limited attention constraint (7) may appear at first a daunting task. Fortunately, Duncan (1970) showed that:<sup>11</sup>

$$\int \ln \left( \frac{d\mathbb{P}_{\mathbf{fz}}}{d(\mathbb{P}_{\mathbf{f}} \times \mathbb{P}_{\mathbf{z}})} \right) d\mathbb{P}_{\mathbf{fz}} = \frac{1}{2} \int_0^t (x_{is} + y_{is}) ds. \quad (18)$$

Therefore, the attention constraint every period ends up taking a very simple form. Every instant, the precision of the two signals at each of the two dates, is subject to an upper bound:

$$x_{it} + y_{it} \leq 2k. \quad (19)$$

To conclude, the attention problem at the date of the regime change is:

$$\begin{aligned} V^*(v_{iT}^r, v_{iT}^u, T) &= \max_{\{x_{it}, y_{it}\}} \{V\} \text{ in equation (14)} \\ &\text{subject to (15), (16), (17) and (18),} \\ &\text{and taking as given } \gamma_t \text{ in (13) and (12).} \end{aligned} \quad (20)$$

Turning to the attention problem at the date of the announcement,  $\tau$ , by similar steps the attention problem is:

$$\begin{aligned} U^*(v_{i\tau}^r, v_{i\tau}^u, \tau) &= \max_{\{x_{it}, y_{it}\}} \left\{ - \int_{\tau}^T e^{-\rho(t-\tau)} \omega^2 v_{it}^u dt + e^{-\rho(T-\tau)} V^*(v_{iT}^r, v_{iT}^u, T) \right\} \\ &\text{subject to (15), (16), (17) and (18).} \end{aligned} \quad (21)$$

Between the announcement and the actual change of regime, the private agents are

---

<sup>11</sup>Kasa's (2006) proposition 3.2 introduced this result to the economics literature, and provides a simpler proof following Lipster and Shiryaev (2001).

only suffering losses from not knowing enough about their individual circumstances. The aggregate regime is known, and there is nothing to learn about it. However, they realize that positive attention to the future aggregate regime will lower their variance about it when the regime actually changes,  $v_{iT}^r$ , which will lower future losses from  $T$  onwards.

Finally, between date 0 and date  $\tau$ , the private agents have only a probability  $p$  that there will be a change in regime. Assumption 2 that  $p \rightarrow 0$ , simplifies the problem considerably, since then it follows immediately that:

$$x_{it}^* = 0 \quad \text{for } t \in [0, \tau], \quad (22)$$

as the agent perceives no benefit from spending attention learning about the aggregate regime. This in turn implies that  $v_{i\tau}^r = \sigma^2 e^{-\mu\tau}$ .

Collecting all of these results, gives the following intermediate result on the way to characterizing the competitive equilibrium

**Proposition 1** *Given an announcement date  $\tau \in [0, T]$ , the equilibrium attention is described by the triplet  $\{x_t^*, y_t^*, \gamma_t\}$  and the value functions  $V^*$  and  $U^*$  that solve the two equations (13), (12), and the two deterministic optimal control problems (20), and (21) for  $[\tau, +\infty)$ . For  $[0, \tau)$  the equilibrium is  $\{0, 2k, 0\}$ .*

This proposition has a surprising result. In spite of all the uncertainty and the multiple signals and shocks hitting the economy, the competitive equilibrium boils down to solving a sequence of deterministic control problems, which are part of the standard toolkit of most economists. This result is more general than the study of the optimal timing of announcements to which this paper is devoted. As long as the objective functions are quadratic, a pervasive assumption in the literature on imperfect information, then modelling information endogenously through the theory of rational inattention leads to tractable problems.

### 3.3 Solving for attention after the regime change

The control problem determining  $V^*(v_{iT}^r, v_{iT}^u, T)$  in (20) is complicated by the presence of the  $\gamma_t$  term. People take  $\gamma_t$  as given, but as long as attention changes over time, then so will  $\gamma_t$  and this implies that the relative weights in the payoff function change. The

Bellman equation associated with this problem is a partial differential equation, which can generally only be solved numerically.

I therefore start by assuming that  $\alpha = 1$ , and return to the general case only at the end. In this case,  $\gamma_t$  drops out of the individual problem, since agents no longer care about aggregate actions in choosing their own behavior. The appendix proves the following result:

**Proposition 2** *Define the following steady-state levels of attention:*

$$v_*^r \equiv \frac{\rho\omega\phi}{2(\eta+k)[\rho+2(\eta+k)]} \quad (23)$$

$$v_*^u = \frac{\phi}{2(\eta+k)} \quad (24)$$

After the regime change at date  $T$ ,

a) If the uncertainty on the aggregate regime is low,  $\nu_T^r \leq v_*^r$ , then  $x_t = 0$  and  $\nu_t^r = \nu_T^r$  for all  $t \geq T$ . If  $\nu_t^u > v_*^u$ , it approaches  $v_*^u$  asymptotically, whereas if  $\nu_t^u = v_*^u$  it stays there forever.

b) If  $\nu_T^r > v_*^r$ , then  $x_t$  approaches 0 asymptotically from above, and as it does  $\nu_t^u \rightarrow v_*^u$  and  $\nu_t^r \rightarrow v_*^r$  asymptotically as well. As for the initial path of attention, let  $f(\cdot)$  be the increasing function that solves to the first-order ordinary differential equation:

$$(f + v_u f') \left( \frac{v_u}{f} - \frac{v_*^u}{v_*^r} \right) = \frac{2\phi}{\rho} \left( 1 - \frac{v_u}{v_*^u} \right) \quad (25)$$

with particular solution  $v_*^r = f(v_*^u)$ . Then:

(b-i) If  $v_{rT} = f(v_{uT})$ , initially  $x_T \in (0, 2k)$ , and  $x_t$  and  $v_t^r$  decline monotonically over time.

(b-ii) If  $v_{rT} < f(v_{uT})$ , initially  $x_T = 0$ , and at some finite date after  $T$ ,  $x_t$  jumps up to a positive value between 0 and  $2k$ , after which  $x_t$  and  $v_t^r$  decline monotonically over time.

(b-iii) If  $v_{rT} > f(v_{uT})$ , initially  $x_T = 2k$ , and at some finite date after  $T$ ,  $x_t$  jumps down to a value between 0 and  $2k$ , after which  $x_t$  and  $v_t^r$  decline monotonically over time.

The dynamics can perhaps be best understood by looking at the phase diagram in figure 1. The function  $f(\cdot)$  in proposition 2 is the stable arm of this saddle-path stable system that can be easily found numerically by solving the differential equation.

The first part of proposition 2 describes the path represented by A in the plot. If we are



vertically below the steady-state, then depending on whether we are at the steady-state for the idiosyncratic variance or to its right, we will either stay at the steady state, or approach it horizontally by moving leftward in the diagram. Intuitively, if agents' knowledge of the new aggregate regime is sufficiently precise, then they will never allocate any attention to learning more about it. Ideally, they would like to trade some of their knowledge of the aggregate regime for a more precise knowledge of the individual circumstances, which are being hit by new shocks every instant. Not being able to do this, the next best thing is to devote all attention to learning about their aggregate circumstances, so uncertainty about them falls to the minimum possible given the arrival of new shocks:  $v_*^u$ .

The second part of proposition 2 covers the more interesting case where agents start above the steady-state for the variance of the aggregate regime. The main result, in this case, is that we still asymptotically approach the case where all the attention ends up devoted to the individual circumstances and uncertainty about them gets driven down to the feasible minimum  $v_*^u$ . A second interesting result is that, in spite of it,  $\nu_t^r$  does not go to zero, so agents never learn for sure what the new policy regime is. After they have a good enough estimate, it is just not worthwhile improving on it instead of devoting all their attention to learning about the new idiosyncratic shocks that arrive every instant.<sup>12</sup>

The third result on proposition 2, covered in cases i to iii, concerns the initial allocation of attention. Note that while the dynamic system is saddle-path stable, both variables are states that cannot jump. Because the saddle-path only applies if attention is interior, unless in the seemingly unlikely case where the initial conditions at date  $T$  happen to fall exactly on the saddle path, initially agents will be devoting all or none of their attention to learning about the regime. Results ii and iii show that if we are above or below the saddle path, and so whether uncertainty about the regime is too high or too low, then all of the attention will be devoted to the aggregate regime or to idiosyncratic circumstances, respectively. This holds for a finite period of time, until we hit the saddle path, at which point, attention jumps to whatever level is consistent with the saddle path from then onwards slowly converges to

---

<sup>12</sup>The reader might wonder why did I then assume that information about the old regime was perfect at date 0. This assumption made the presentation easier, but has no influence on the results. As long as the uncertainty about the old regime had reached its steady-state, agents would never expand any attention on it ever again, so this would not affect any of the attention allocation decisions studied in this paper.

zero. These two possible paths are denoted by B and C in figure 1.

Finally note that proposition 2 does not cover the case when  $\nu_T^r < v_*^r$ . Knowing that  $v_*^r$  is the lowest possible uncertainty on individual conditions that an agent could achieve, in the proposition and from now onward, I assume that  $\nu_t^r \geq v_*^r$  at all dates  $t$ .

### 3.4 Attention allocation after the announcement

I turn next to solving for attention between the announcement date  $\tau$  and the regime change date  $T$ . Recall that, we had already found that before date  $\tau$ , attention to the aggregate regime is zero. Then, it is natural to assume that before date  $\tau$ , we were arbitrarily close to the steady state described in proposition 2, so that  $\nu_{i\tau}^u = v_*^u$ .

The appendix proves the following result:

**Proposition 3** *Given an announcement at date  $\tau$ , at date  $T$ , all attention is devoted to the aggregate regime:  $x_T = 2k$ . Before,*

a) *if  $\sigma^2 e^{-\mu\tau} > \bar{\sigma}_\tau$ , where  $\bar{\sigma}_\tau$  is defined in the appendix and is increasing in  $\tau$ , then  $x_t = 2k$  for  $t \in [\tau, T]$ .*

b) *If  $\sigma^2 e^{-\mu\tau} < \bar{\sigma}_\tau$ , define*

$$d = \frac{2}{\rho} \ln \left[ \frac{2(\eta + \kappa)}{2\eta + (\rho/2)} \right]. \quad (26)$$

*Then, if  $\tau < T - d$ , the time path for attention and the posterior variance is: initially, for  $t \in [\tau, t^*]$ ,  $x_t = 0$ , so  $v_t^r = \sigma e^{-\mu t}$  and  $v_t^u = v_*^u$  for a date  $t^* < T - d$ ; at  $t^*$ , attention jumps to  $x_{t^*} = \rho/2$ ; for  $t \in (t^*, t^* + d)$ , attention  $x_t$  increases,  $v_t^r$  decreases, and  $v_t^u$  increases, all monotonically according to:*

$$x_t = \frac{\rho}{2} + 2(\eta + \kappa) - \frac{\phi}{v_{ut^*} e^{\frac{\rho}{2}(t-t^*)}} \quad (27)$$

$$v_{ut} = v_{ut^*} e^{\frac{\rho}{2}(t-t^*)} \quad (28)$$

$$v_{rt} = v_{rt^*} e^{-\left(\frac{\rho}{2} + 2(\eta + \kappa)\right)(t-t^*) + \frac{2}{\rho}(\eta + \kappa)(1 - e^{-(\rho/2)(t-t^*)})} \quad (29)$$

*Finally, for  $t \in (t^* + d, T)$ , attention is solely devoted to the aggregate regime  $x_t = 2k$ , so  $v_t^r$  falls at a rapid rate and  $v_t^u$  increases.*

c) *If  $\sigma^2 e^{-\mu\tau} < \bar{\sigma}_\tau$  and instead  $\tau > T - d$ , depending on parameters, either (i)  $x_t = 0$  for*

$t \in [\tau, t^*)$  and  $x_t = 2k$  for  $t \in [t^*, T]$ , (ii)  $x_t = 2k$  for  $t \in [\tau, t^*)$ , followed by  $x_t \in (0, 2k)$  according to (27) reaching  $2k$  before date  $T$  and staying there until  $T$ .

This proposition is best understood by looking at figures 2 to 4. They plot the case when the policymaker knew about the regime change well in advance (that is  $d < T$ ) and announces it immediately at date 0. As long as  $\sigma^2$  is not too large, then agents will not want to devote all of their attention to the aggregate regime. Instead, they allocate attention by following to a few stages.<sup>13</sup>

Initially, the optimal  $x_t = 0$ , so agents devote no attention to the announcement. In a second stage,  $x_t$  jumps to  $\rho/2$  and then increases continuously and monotonically until  $2k$ . Finally, in a third stage,  $x_t = 2k$  until the regime changes at date  $T$ . After this, proposition 2 takes over: attention is initially at  $2k$ , but at some finite date it jumps down and then monotonically falls asymptotically approaching zero. Figures 3 and 4 show the corresponding paths for the variances of the aggregate regime and the idiosyncratic circumstances.

Figure 2 therefore shows the life-cycle of attention to aggregate regime changes. As the announcement comes very early, people ignore it. The benefits of learning about the future are being discounted at rate  $\rho$ , while the costs of not having enough information on current individual circumstances are felt immediately. As the regime change approaches, discounting is not so heavy and people start investing some capacity in learning about the change. The amount of attention increases with time as the change gets nearer and peaks at the maximum just before the change. After the change, and when the agents have a precise enough estimate of the new regime, they start allocating attention slowly back to individual circumstances.

The date  $t^*$ , defined in proposition 3, is the date at which the announcement stops being ignored. The following result describes the determinants of attention

**Corollary 4** *The date  $t^*$  at which attention to the aggregate regime  $x_t$  jumps from 0 to a positive value arrives earlier if there is more: total attention ( $k$ ), uncertainty about the regime ( $\sigma^2$ ) importance of the aggregate regime ( $\omega^{-1}$ ) and patience ( $\rho^{-1}$ )*

---

<sup>13</sup>The parameters used in the figures were:  $\rho = 0.04$ ,  $\omega = 1$ ,  $\alpha = 1$ ,  $\phi = 0.01$ ,  $\eta = 0.5$ ,  $\sigma^2 = 1$  and  $k = 0.1$ .

The first three properties are natural. The fourth is a little more interesting. The more agents discount the future, the less costly is their imperfect knowledge of the future regime vis-a-vis the cost of being imperfectly informed about individual circumstances right now. Therefore, the sooner they stop ignoring the policy announcement.

### 3.5 The case with strategic complementarities

It is difficult to provide any sharp results with a general  $\alpha$ , similar to the previous two propositions. The optimal control problems are not time-invariant. One result that is shown in the appendix is:

**Proposition 5** *Near the steady state after the regime change, when the system is in its saddle path, the stronger are strategic complementarities ( $\alpha$ ), the larger is attention to the aggregate regime ( $x$ ).*

Moreover, following the hint of propositions 2 and 3 for the  $\alpha = 1$  case, the appendix proposes an algorithm to solve the problem for  $\alpha < 1$ . Extensive numerical investigation has confirmed that, the lower is  $\alpha$ , the higher is attention to the aggregate regime at all points in the life-cycle. Figure 5 illustrates this by plotting the life-cycle of attention for  $\alpha = 0.25$  and comparing it with the case  $\alpha = 1$  already displayed in figure 2.<sup>14</sup>

Intuitively, with strategic complementarities, when agents devote more attention to the policy regime, aggregate actions depend more on the regime. Because agents wish to track what others are doing in the aggregate, this raises the incentive to pay attention to the aggregate regime.

## 4 Optimal announcements

Recall that the policymaker learned at date 0 that a change in regime would occur at date  $T$ . To give the policymaker the largest set of possibilities to choose from when picking the optimal announcement date, referring to proposition 3, I assume that  $T$  is sufficiently distant, in particular that  $T > d$ .

---

<sup>14</sup>Aside from  $\alpha$ , all other parameters are the same as in figures 2 to 4.

## 4.1 Optimal announcement date

The announcement problem is to choose  $\tau \in [0, T]$  to maximize:

$$W = -\omega^2 v_u^* \left( \frac{1 - e^{-\rho\tau}}{\rho} \right) + e^{-\rho\tau} U^*(\sigma^2 e^{-\mu\tau}, v_u^*, \tau) \quad (30)$$

subject to the competitive equilibrium described in propositions 1, 2 and 3.

The first-order condition for this problem is:

$$-\omega^2 v_u^* e^{-\rho\hat{\tau}} + e^{-\rho\hat{\tau}} \left[ \frac{\partial U^*}{\partial t} - \frac{\partial U^*}{\partial \sigma_\tau^2} (\mu \sigma^2 e^{-\mu\tau}) \right] = 0 \quad (31)$$

The first term is the loss in terms of having an extra instant of time when attention to the problem is zero and the variance of the individual circumstance is at its steady-state. The second-term is the gain that comes from two sources: the allocation of private attention to the regime change and the improvement in the precision of the signal. Both sources are straightforward to calculate, since an important result in optimal control theory is that  $\partial U^*/\partial t$  is equal to the negative of the maximized Hamiltonian, while  $\partial U^*/\partial \sigma_\tau^2$  is one of the co-state variables.

Using these, leads to the following result, proven in the appendix:

**Proposition 6** *With  $\alpha = 1$ , let  $t^*$  be the date at which agents would start paying attention to an announcement made at date 0, defined in proposition 3. Then the optimal announcement date  $\tau = t^*$  if  $\mu \rightarrow 0$ , while  $\tau > t^*$  if  $\mu > 0$  and increases with  $\mu$ . At the date of the announcement  $x_\tau = \mu + \rho/2$ .*

Intuitively, because  $\mu > 0$ , the policymaker will never want to release an announcement that is ignored; by waiting an extra instant, she could make a better announcement. If  $\mu \rightarrow 0$  though, as soon as agents privately would choose to pay some attention to the aggregate regime, then the policymaker wants to allow them to do so. If  $\mu$  is strictly larger than zero, the faster the policymaker learns about the regime change, then the later she should make an announcement. Interestingly, note that as  $\tau > t^*$ , there is a period of time when agents would be curious to learn about a regime change, but the policymaker chooses to keep mum.

Second, the more agents discount the future, the later they will wish to start paying attention to the regime change. Therefore, again the later the optimal announcement should be.

Finally, comes the influence of strategic complementarities. Following Angeletos and Pavan (2007), the *Pareto optimal* allocation of attention  $\{\hat{x}_t, \hat{y}_t\}$  still solves the two deterministic optimal control problems (20), and (21) for  $[\tau, +\infty)$ , as in proposition 1, but now taking into account the evolution of the precision on the public signal (4). and the evolution of  $\gamma_t$  in (13) and (12). These two constraints show the two main differences relative to the private agent's problem. First, if no attention effort or resources are expanded at date 0, still the prior variance will shrink as long as  $\mu > 0$ . Therefore, even if an individual agent might have wished to pay attention at date 0, it would be better for social welfare in some cases to delay the announcement and reap the free arrival of information. Second, as long as  $\alpha < 1$ , there is a payoff externality in this economy, that the social planner now internalizes.

**Proposition 7** *If  $\alpha < 1$ , then attention to the aggregate regime in the competitive equilibrium is higher than in the Pareto Optimum:  $x_t^* > \hat{x}_t$ .*

*(Note: what follows is a conjecture) Moreover, the optimal announcement date  $\tau^*$  is later than when  $\alpha = 1$ .*

As long as  $\alpha = 1$ , there is no payoff externality, so conditional on an announcement date, the competitive equilibrium amount of attention coincides with the Pareto optimum. When  $\alpha < 1$  though, Angeletos and Pavan (2007) showed that, in this class of models, agents' actions put too much weight on public signals. The result in the proposition is different, but related: there is too much attention devoted to learning about the aggregate regime. Intuitively, agents do not privately take into account the extra cost from exerting more attention that comes from the excess sensitivity of aggregate actions to the policy regime.

The presence of strategic complementarities provides an alternative reason to keep mum: as announcements lead agents to start trying to forecast what others are forecasting about it, they became inefficiently too curious and are best left in the dark.

## 5 Relaxing the key assumptions

This section extends the basic model in several directions and relaxes some of its assumptions. The first subsection focusses on the attention constraint, first by making the link to the literature on rational inattention, and then by allowing for a third use of attention beyond the regime and individual circumstances. The following subsection expands on the role of the policymaker. First, I allow for multiple signals, wasteful signals, and public signals. Second, allows the policymaker to take actions in response to the regime change. Section 5.3 focuses on the information revealed in announcements of a new regime, by considering the case where assumption 2 does not hold

### 5.1 The model of attention

The rational inattention model led to expressing the attention constraint as the sum of the signal-to-noise ratios on the two fundamentals being below an exogenous amount  $2k$ .

#### 5.1.1 Rational inattention foundations

Started by Sims (2003), the literature on rational inattention has made two contributions to imperfect information models.<sup>15</sup> First, taking as given the set of signals, it argued that their information content can be measured using the change in entropy due to the signal.<sup>16</sup> The reduction in uncertainty afforded by the signal is then the fall in entropy in going from the marginal prior for  $u_{it}$  to its conditional distribution. The parameter  $k$  in this literature is called the “capacity” of the agent to receive signals, and is interpreted as a limit on the agents’ attention or ability to absorb new information.

A second contribution of rational inattention has been to provide micro-foundations to the structure of the signals. Instead of assuming that signals are normal, one can let the private agents choose the distribution  $f(z_{it}^u|u_{it})$ . It is an elementary result in information

---

<sup>15</sup>See Sims (forthcoming) for a recent survey. An alternative model of inattention was proposed by Reis (2006) and Mankiw and Reis (2002). Both models are surveyed in Mankiw and Reis (forthcoming).

<sup>16</sup>Entropy has a few useful properties: it is non-negative, it equals zero only if the distribution is degenerate in that  $u_{it}$  can only take one value with probability 1, and it is maximized if  $f(u_{it})$  is the improper uniform distribution over the whole real line. Shannon (1948) showed that it is the only measure of uncertainty that satisfies three axioms: continuity in probabilities, increasing in the number of possible events, and additive if choices are sequential.

theory that the normal distribution maximizes entropy over all distributions with support in the real line and a given mean and variance.<sup>17</sup> But recall that the private agents' problem is to minimize the variance subject to a lower bound on entropy. By duality, this problem is the same as maximizing entropy subject to an upper bound on the variance. Therefore, it follows that  $f(u_{it}|z_{it}^u)$  is normally distributed. Since  $f(u_{it})$  was assumed to be normal, it then follows that  $f(z_{it}^u|u_{it})$  is normal, just as we assumed. Two further assumptions, that have become standard in this literature, lead precisely to the information structure assumed in section 2.3. First, following Sims (2003) that the noise in the signals is idiosyncratic, so that while agents set up similar channels to receive the information, the noise that contaminates these channels is individual-specific. Second, following Mackowiak and Wiederholt (2009), that signals on independent fundamentals, here  $u_{it}$  and  $r_t$ , are also independent signals, since paying attention to aggregate and idiosyncratic conditions are separate activities.

Therefore, the assumptions on information and attention in section 2.3 follow from the weaker assumptions that agents are rationally inattentive. One can also ask how do the assumption on the policy announcement in section 2.2 fit in. First, regarding the policy signal on  $r_1$ , rational inattention interprets policymakers as stating  $r$  precisely, but agents interpret this with normally distributed noise through their optimal limited channel. The  $\mu_t$  is then the capacity of this channel. Similarly, the information that the policymaker reveals that there has been a change in regime, has an information content of  $p \ln(p) + (1-p) \ln(1-p)$ . Note that an implicit assumption is that agents cannot re-allocate capacity between the announcement channels and their private information channel. Capacity in the latter is fixed at  $k$ , and the trade-off this paper focuses on is on allocating capacity within this private channel. It would be interesting to extend the analysis in this regard.

### 5.1.2 Further sources of attention

In the basic model, there are only two uses of attention: to learn about  $u_{it}$  or about  $r$ . People, of course, pay attention to many things, including some that may be hard to relate to the individual circumstances affecting actions captured by  $u_{it}$ . A simple way to extend

---

<sup>17</sup>See chapter 11 in Cover and Thomas (1991) for a proof, and Sims (forthcoming) for further discussion.



the model to allow for this is to consider the modified objective function to replace  $W$  in (1):

$$W_i = \mathbb{E}_{i0} \left[ \int_0^\infty e^{-\rho t} \left\{ [a_{it} - \omega u_{it} - (1 - \alpha)a_t - \alpha r_t]^2 + g(w_{it}) \right\} dt \right] \quad (32)$$

where  $g(\cdot)$  is an increasing concave function of the attention devoted to other activities,  $w_{it}$  satisfying Inada conditions.<sup>18</sup> The new attention constraint is:

$$x_{it} + y_{it} + w_{it} \leq 2k \quad (33)$$

To solve this new problem, we can proceed in two steps. First, take  $2k - w_{it}$  and solve for the allocation of attention  $x_{it}$  and  $y_{it}$  as in the previous section. All of the results established there follow exactly, by just replacing  $k$  by  $k - w_{it}/2$  in all the expressions. In the second step solve for  $w_{it}$  using the indirect utility function from the first step. Given the assumed properties of  $g(\cdot)$  there will be an interior solution for this problem.

The main conclusion of the propositions would not be changed. However, in the same way that attention to the aggregate regime fell with less  $k$ , and the optimal announcement date becomes later with higher  $k$ , the same effect will be present with  $w_{it}$ . If people wish to pay attention to other activities, this reduces the attention available for policy announcements. This makes it more likely that these announcements will be ignored, reinforcing the case for delaying them.

### 5.1.3 Learning from other sources

Another extension of the basic model is to note that sometimes people learn about aggregate changes from seemingly unrelated sources. For instance, a person may be spending attention on a leisure activity, like watching television or going out for dinner, and in doing so bump into news in the form of commercials or billboards that are informative about the aggregate regime. One way to model this is to say that the signal associated with the leisure activities above is correlated with the aggregate regime. In this case, the model barely changes. All it requires is that we split the information associated with leisure, by

---

<sup>18</sup>To be more precise,  $g : \mathfrak{R}_0^+ \rightarrow \mathfrak{R}$ , such that  $g'(\cdot) \geq 0$ ,  $g''(\cdot) \leq 0$  and  $\lim_{w \rightarrow 0} g'(\cdot) = +\infty$  and  $\lim_{w \rightarrow \infty} g'(\cdot) = 0$ .

projecting it onto the aggregate regime, thereby orthogonalizing it into two components. The information associated with the component correlated with the aggregate regime can then be seen as part of  $x_{it}$ , while the information strictly associated with leisure would be part of  $w_{it}$ . As long as the policymaker can still, at the margin, choose whether to spend more attention to exclusively learning about the aggregate regime, all of the conclusions remain the same.<sup>19</sup>

## 5.2 Policymaker acts beyond announcements

Section 2.2 assumed that, after making an announcement at date  $t$  and lowering uncertainty to  $\sigma^2 e^{-\mu t - \mu_0}$ , the policymaker was not allowed to reveal any more information later. One way to relax this assumption changes nothing about the model, leading only to a slight re-interpretation of the set-up. Imagine that the policymaker can issue a signal every period, but that if the agent does not devote any extra attention to the aggregate regime, then it cannot use its memory at date 1 to recover the date 0 signal. Then, as long as the repeated signal at date 1 following a date-0 announcement is still less precise than a signal following a date-1 announcement ( $\mu > 0$ ), propositions 2 and 3 do not change.

*To Write: Already done, but have to write in simpler way. With ,multiple signals nothing changes but need maths to see it.*

A third extension is to consider what happens if the signals sent by the policymaker are not completely free in terms of private attention. In particular, assume now that the attention constraint is instead

$$x_{it} + y_{it} \leq 2k - \chi \mu_t, \quad (34)$$

so that when the policymaker sends her signals is exhausts some capacity at a rate  $\chi < 1$ .

From the perspective of individual choices (and so from the description of the competitive

---

<sup>19</sup>To be more precise, imagine that any attention  $\tilde{w}_{it}$  devoted to leisure leads to both information about leisure as well as about the aggregate regime. We can model this as  $\tilde{w}_{it} = w_{it}/(1 - \xi) + x_{it}^L/\xi$ , where  $x_{it}^L$  is the information about  $r$  that comes from leisure and  $\xi$  is its exogenous share in the total information. Letting  $x_{it}^P$  be the information about  $r$  that comes from the channel exclusively dedicated to it, then the attention constraint becomes:  $x_{it}^P + y_{it} + \tilde{w}_{it} \leq 2k$ . Defining  $x_{it} = x_{it}^P + \xi x_{it}^L$ , leads to (33) and the same model. The only difference is that the constraint that  $x_{it}^P \geq 0$  leads to a constraint that  $x_{it} \geq \xi w_{it}$ . If this constraint doesn't bind, as at the margin, agents do not get too much information about the aggregate regime from leisure activities, then nothing in the analysis changes. If it does bind, equation (??) must be solved jointly with the equations for optimal  $x_{it}^*$  in section 2.

equilibrium) nothing changes, as long as  $k$  gets replaced by  $k - \chi\mu_t/2$  in all the expressions. When it comes to the optimal timing of policy announcements, all of the qualitative predictions in proposition 2 remain, but there is a new consideration at play. Because late announcements are more informative, they also use more capacity. This extra cost raises the desirability of earlier announcements.

Finally, a fourth extension, is to drop the assumption that policy signals give each agent an idiosyncratic signal centered around the true regime  $r$ . The reason for this assumption was the rational inattention micro-foundations described section 3.1. However, the work in global games following Morris and Shin (2002) has made a different assumption: that policy announcements are public signals, noisy but shared by all. In terms of the model, now following an announcement, while the policymaker tries to communicate  $r$ , every agent observes instead  $\bar{r}$ , a draw from a normal distribution centered at  $r$  with positive variance. Following the announcement, and before investing any private attention, the agents' posterior variance about its estimate of  $r$  can still be represented as  $\sigma^2 e^{-\mu_t}$ , where  $\mu_t$  is again a decreasing function of the variance of the policy signal.

*To do: Already done in 2-period white-noise model. Effect of coordination of actions a la Angeletos-Pavan on top of the coordination of attention in my propositions.*

### 5.2.1 Actions beyond announcements

So far, this paper has focused on when to announce a change in regime, taking the change as given. In this section, I consider a few alternatives, where the policymaker may influence the change, always from the perspective of communication policy.

First, imagine that  $r = 0$ . That is, the new regime turns out to be just like the old one. If agents receive an announcement of a change, because they are uncertain about  $r$  they will spend precious attention to gradually learn that nothing has changed. Making unnecessary announcement confuses people and is wasteful.

Second, and related, imagine that  $r$  is partly determined exogenously partly affected by policymaker's actions. Then, in response to an exogenous shock to the regime, the optimal response is to offset it with actions, reestablishing the regime back to  $r = 0$ . Policy announcement are costly in scarce attention, unlike actions to change the regime. If

policymakers can fix it, they should just do it, without having to talk about it.

Third, imagine that a rumor emerges at date 0 that there will be a change in regime at date 1. Rumors can be treated just like announcements, as signals with some information content. One possibility is to think of rumors as leading to increases in the probability of a regime change,  $p$ . In this case, the analysis in the previous section implies that it becomes more likely that policymakers make announcements. Alternatively, assume that the rumor is credible, so people learn from it that there has been a change, but very imprecise, as it reveals less information about the value of the new regime than a policy announcement. In terms of the model, even without an announcement, now  $\pi = 1$ , but the uncertainty on  $r$  falls by less than  $e^{-\mu t - \mu_0}$ . Agents may now start devoting attention to the regime change even without an announcement: the policymaker has lost her control over this. The optimal policy is then to make an immediate announcement, in order to give agents the benefit of the information about the new regime that this comes with. In other words, rumors should be squashed early.

### 5.3 Non-zero prior probability of regime changes and signalling

In the model, the prior probability of a regime change is  $p$ . Assumption 2, that  $p = 0$ , implied that until there was an announcement, agents' never believed that a change could happen. This simplified the analysis because I did not have to track what happened to agents' perception of a regime change. I now relax this assumption.

I maintain the assumption that announcements are fully credible and use  $\pi$  to denote the posterior probability that there is a regime change. A lower  $\pi$ , pushes down the desire of agents to expend valuable attention at date 0 to learn about a regime change. At date  $t$ , after observing an announcement, the agent's posterior probability that there is a regime change is 1. If the agent does not observe an announcement though, this could be either because the policymaker chose not to announce it, or because no change occurred.

*This is still work in progress: with 2-period white noise problem characterized the Perfect Bayesian equilibrium, and showed that as long as  $p < \bar{p}$ , still  $x_{it} = 0$  for  $t < \tau$ , so all the propositions were unchanged. In the full continuous time model here, this is proving to be quite challenging.*

## 6 Conclusion

When should a policymaker make an announcement about an impending change that will affect everyone? This paper provided an answer to this question based on the notion that people have a limited amount of attention that they choose to allocate (or not) to policy announcements. I focussed on two trade-offs involved in optimally timing announcements, one private and the other public. The private trade-off is between obtaining more information early on about the future regime change to make better choices when it arrives versus in exchange paying less attention to individual circumstances today and making worse choices now. The public trade-off is between satisfying agents' demand for information versus waiting to obtain a more precise signal of the impending change and fighting the over-reliance of agents on public signals.

The overriding lesson is that announcements that are made too early are ignored and socially costly. Announcements that are made too late do not give people enough time to get ready for the coming changes. The optimal timing of the announcement balances these two forces, and occurs after the time when people would start paying attention to it for two reasons. First, because the policymaker can make more precise announcements as we get closer to the date of the change. Second, because people put too much weight on announcements as they try to figure out what others know or think they know.

The aim was to make three contributions. First, to suggest a new question to the literature on policy announcements, focusing on the optimal timing of announcements. Second, to propose a limited-attention theory to this question leading to several applied policy principles. Third, to study the life-cycle of attention by solving a continuous time rational inattention model. In none of these three areas, this paper will surely be the last word.

## Appendix

*The appendix is very incomplete.*

### Proof of proposition 2: Solving the $V^*(.)$ optimal control problem.

With  $\alpha = 1$ , the control problem is, from date  $T$  onwards to:

$$V(v_T^r, v_T^u) = \max_{\{x_t\}_T^\infty} - \int_T^\infty e^{-\rho(t-T)} (v_t^r + \omega v_t^u) dt \quad \text{subject to} \quad (35)$$

$$\dot{v}_t^r = -x_t v_t^r \quad (36)$$

$$\dot{v}_t^u = \phi - 2(\eta + k)v_t^u + x_t v_t^u \quad (37)$$

$$x_t \geq 0 \quad (38)$$

$$x_t \leq 2k. \quad (39)$$

with initial conditions  $v_T^r$  and  $v_T^u$ . The variance of the idiosyncratic components is bounded:

$$\frac{\phi}{2\eta + 2k} \leq v_T^u \leq \frac{\phi}{2\eta}, \quad (40)$$

where the lower bound corresponds to all attention being devoted to the individual circumstances for an infinite amount of time before date  $T$ , and the upper bound to zero attention being devoted to it for an infinite amount of time before date  $T$ .

To derive the sufficient and necessary optimality conditions, start by setting up the current-value Hamiltonian:

$$H_t = -v_t^r - \omega v_t^u + \lambda_t(-x_t v_t^r) + \psi_t [\phi - 2(\eta + k)v_t^u + x_t v_t^u] \quad (41)$$

where  $\lambda_t$  and  $\psi_t$  are the co-state variables, which equal the marginal value of an increase in  $v_t^r$  and  $v_t^u$  respectively, and so are always non-positive. The associated Lagrangean, to take into account the boundary conditions on attention, is:

$$\mathcal{L} = H_t + \xi_{1t} x_t + \xi_{2t} (2k - x_t). \quad (42)$$

The first-order sufficient and necessary conditions describing the optimal paths, which must

hold for all  $t \geq T$  are:

- The first-order and complementary slackness conditions:

$$\lambda_t v_t^r + \xi_{2t} = \psi_t v_t^u + \xi_{1t}, \quad (43)$$

$$\xi_{1t} x_t = 0, \quad \xi_{1t} \geq 0, \quad x_t \geq 0, \quad (44)$$

$$\xi_{2t}(2k - x_t) = 0, \quad \xi_{2t} \geq 0, \quad x_t \leq 2k. \quad (45)$$

- The dynamic equations for states and co-states:

$$\dot{v}_t^r = -x_t v_t^r \quad (46)$$

$$\dot{v}_t^u = \phi - 2(\eta + k)v_t^u + x_t v_t^u \quad (47)$$

$$\dot{\lambda}_t = (\rho + x_t)\lambda_t + 1 \quad (48)$$

$$\dot{\psi}_t = (\rho + 2\eta + 2k - x_t)\psi_t + \omega \quad (49)$$

- And the initial conditions and terminal conditions, which come from transversality conditions:

$$v_T^r \text{ and } v_T^u \text{ are given} \quad (50)$$

$$\lim_{t \rightarrow \infty} e^{-\rho t} \lambda_t v_t^r = 0 \quad (51)$$

$$\lim_{t \rightarrow \infty} e^{-\rho t} \psi_t v_t^u = 0 \quad (52)$$

By the properties of optimal control problems, there is a unique solution to this dynamic system of equations. To characterize the solution, I prove a sequence of lemmas.<sup>20</sup>

**Lemma A.1:** *If and only if the initial uncertainty on the regime is sufficiently low,  $v_s^r \leq v_*^r$ , the optimal allocation of attention is  $x_t = 0$  for all  $t \geq s$ , where  $s \geq T$  is an arbitrary finite*

---

<sup>20</sup>One result that I will use repeatedly is that for a variable  $y_t$  such that:  $\dot{y}_t = -ay_t + b$  where  $a$  and  $b$  are real scalars, then:

$$y_t = y_0 e^{-at} + \left(\frac{b}{a}\right)(1 - e^{-at})$$

date.

**Proof:** All we have to do is verify that all of the optimality conditions are satisfied. Starting with the dynamic equations for states and co-states, they are all first-order ODEs with constant coefficients so that for any  $t \geq s$ :

$$v_t^r = v_s^r \quad (53)$$

$$\lambda_t = e^{\rho(t-s)} \lambda_s - \left( \frac{1}{\rho} \right) \left( 1 - e^{\rho(t-s)} \right) \quad (54)$$

$$v_t^u = e^{-2(\eta+k)(t-s)} v_s^u + \left( \frac{\phi}{2\eta + 2k} \right) \left( 1 - e^{-2(\eta+k)(t-s)} \right) \quad (55)$$

$$\psi_t = e^{(\rho+2\eta+2k)(t-s)} \psi_s - \left( \frac{\omega}{\rho + 2\eta + 2k} \right) \left( 1 - e^{(\rho+2\eta+2k)(t-s)} \right). \quad (56)$$

Plugging these into the transversality conditions, shows that:

$$\lambda_t = -\frac{1}{\rho} \quad (57)$$

$$\psi_t = -\frac{\omega}{\rho + 2\eta + 2k}, \quad (58)$$

for all dates  $t \geq s$ . Finally, the complementarity slackness conditions are  $\xi_{2t} = 0$  and  $\xi_{1t} \geq 0$ , which combined with the first-order conditions implies that:

$$\lambda_t v_t^r \geq \psi_t v_t^u. \quad (59)$$

Replacing the solutions in (53)-(56) transforms this expression into:

$$\frac{v_s^r}{\rho} \leq \frac{\omega v_t^u}{\rho + 2\eta + 2k}. \quad (60)$$

But finally note that  $\dot{v}_t^u < 0$  and  $v_t^u \rightarrow v_*^u$ , so  $v_t^u \geq v_*^u$ . Therefore for the expression to hold for all  $t \geq s$ , a necessary and sufficient conditions is that:

$$\frac{v_s^r}{\rho} \leq \frac{\omega v_*^u}{\rho + 2\eta + 2k}, \quad (61)$$

which proves the if condition in the lemma. To prove the only if, note that if the condition



does not hold, at some finite date  $t$ ,  $\xi_{1t} < 0$  which violates the optimality conditions. ■

From now onwards, all the remaining results will apply to when this condition does not hold.

**Lemma A.2:** *Let  $s > T$  be some arbitrary finite date. Then, having  $x_t = x^* > 0$  for all dates  $t \geq s$  is not a solution.*

**Proof:** Start with the case  $x^* = 2k$ . Then, by precisely the same steps as those in the proof of the lemma A.1, we can show that at all dates  $t \geq s$ :

$$\lambda_t = -\frac{1}{\rho + 2k} \quad (62)$$

$$\psi_t = -\frac{\omega}{\rho + 2\eta}, \quad (63)$$

Combining the complementary slackness and optimality conditions, for any date  $t \geq s$ :

$$\frac{v_t^r}{\rho + 2k} \geq \frac{v_t^u}{\rho + 2\eta}. \quad (64)$$

But as  $t$  grows larger, since  $x^* > 0$ ,  $v_t^r$  becomes arbitrarily close to zero, while  $v_t^u$  is bounded below by  $v_*^u$  thus violating this condition.

For the case  $x^* < 2k$ , by similar steps, we can show that for  $t \geq s$ :

$$\lambda_t = -\frac{1}{\rho + x^*} \quad (65)$$

$$\psi_t = -\frac{\omega}{\rho + 2\eta + 2k - x^*}, \quad (66)$$

The optimality condition then becomes:

$$\lambda_t v_t^r = \psi_t v_t^u, \quad (67)$$

which plugging in these values for the co-states implies that:

$$\frac{v_t^r}{\rho + x^*} = \frac{\omega v_t^u}{\rho + 2\eta + 2k - x^*}. \quad (68)$$

But, again,  $v_t^r$  becomes arbitrarily close to zero, while  $v_t^u$  is bounded below, so this equality cannot hold for all  $t$ . ■

**Lemma A.3:** *If  $x_t \rightarrow x^*$ , then  $x^* = 0$ .*

**Proof:** The proof is essentially the same as that in lemma A.2. It consists of noting that if  $x_t \rightarrow x^*$ , then at an arbitrarily large date,  $x_t$  is arbitrarily close to  $x^*$ . But if  $x^* > 0$ , this implies that  $v_t^r$  gets arbitrarily close to zero, while  $v_t^u$  is bounded below, which violates the optimality condition. ■

The combination of these three lemmas, plus another lemma that will follow characterizes the limiting behavior of attention:

**Lemma A.4:** *Attention  $x_t \rightarrow 0$  asymptotically.*

**Proof:** Lemmas A.1 to A.3 show that either (i)  $x_t \rightarrow 0$  asymptotically, or (ii)  $x_s = 0$  at some date  $s$  at which  $v_s^r \leq v_*^r$ , and from then onwards  $x_t = 0$  forever. Lemma A.8 will show that (ii) is not a solution. ■

The next lemma characterizes the dynamics of the posterior variances when the allocation of attention is interior.

**Lemma A.5:** *If  $x_t \in (0, 2k)$ , then the posterior variances follow the dynamic system:*

$$\frac{\dot{v}_t^u}{v_t^u} = \frac{\rho}{2} - \left( \frac{\rho + 2\eta + 2k}{2\omega} \right) \frac{v_t^r}{v_t^u}, \quad (69)$$

$$\frac{\dot{v}_t^r}{v_t^r} = -\frac{\rho}{2} + \frac{\phi}{v_t^u} - 2\eta - 2k + \left( \frac{\rho + 2\eta + 2k}{2\omega} \right) \frac{v_t^r}{v_t^u}. \quad (70)$$

**Proof:** The optimality condition is:

$$\lambda_t v_t^r = \psi_t v_t^u \quad (71)$$

Taking time derivatives this becomes:

$$\frac{\dot{\lambda}_t}{\lambda_t} + \frac{\dot{v}_t^r}{v_t^r} = \frac{\dot{\psi}_t}{\psi_t} + \frac{\dot{v}_t^u}{v_t^u}. \quad (72)$$

Replacing each time derivative using the dynamic equations in (46)-(49) and simplifying gives:

$$\frac{1}{\lambda_t} = \frac{\omega}{\psi_t} + \frac{\phi}{v_t^u}. \quad (73)$$

But then, equations (71) and (73) are two equations in  $\lambda_t$  and  $\psi_t$  that can be solved to get:

$$\psi_t = \frac{v_t^r - \omega v_t^u}{\phi} \quad (74)$$

$$\lambda_t = \frac{\psi_t v_t^u}{v_t^r} \quad (75)$$

Take time derivatives of equation (74):

$$\phi \dot{\psi}_t = \dot{v}_t^r - \omega \dot{v}_t^u. \quad (76)$$

Using the dynamic equations (46), (47) and (49) to replace for each time derivative, and then (74) to replace for  $\psi_t$  leads, after some algebra, to:

$$x_t = \frac{\rho}{2} - \frac{\phi}{v_t^u} + 2\eta + 2k - \left( \frac{\rho + 2\eta + 2k}{2\omega} \right) \frac{v_t^r}{v_t^u}. \quad (77)$$

Replacing this into equations (46), (47) gives the solution. ■

**Lemma A.6:** The dynamic system in lemma A.5 is saddle-path stable, so there is a unique function  $f : [v_*^r, \infty) \rightarrow [v_*^u, \infty)$ , written as  $v_r = f(v_u)$ , such that only if the initial conditions satisfy this function, will the system asymptotically approach the steady state  $(v_*^r, v_*^u)$  with  $x_t$  asymptotically converging to 0.  $f(\cdot)$  is an increasing function that can be easily found numerically as the solution to the following first-order ODE:

$$(f + v_u f') \left( \frac{\rho v_u}{2f} - \frac{\rho + 2\eta + 2k}{2\omega} \right) = \phi - 2(\eta + k)v_u, \quad (78)$$

with particular solution:  $v_r^* = f(v_u^*)$ .

**Proof:** The dynamics of the phase diagram in figure 1 shows that the system is saddle path stable. (Linearizing it around the steady states and checking eigenvalues also confirms it.)

To derive the ODE, note that:

$$f'(v_t^u) = \frac{dv_t^r}{dv_t^u} = \frac{\dot{v}_t^r}{\dot{v}_t^u} \quad (79)$$

Using the dynamic equations in lemma A.5 then gives

$$f'(v_{ut}) = \frac{\left[ \frac{\phi}{v_t^u} - 2\eta - 2k - \frac{\rho}{2} + \left( \frac{\rho+2\eta+2k}{2\omega} \right) \frac{v_t^r}{v_t^u} \right] v_{rt}}{\left[ \frac{\rho}{2} - \left( \frac{\rho+2\eta+2k}{2\omega} \right) \frac{v_t^r}{v_t^u} \right] v_{ut}}, \quad (80)$$

which after rearranging gives the condition in the lemma. ■

**Lemma A.7:** If  $x_s > 0$  at any date  $s$  then at that date  $v_s^r = f(v_s^u)$ , so the system is in its saddle path and  $x_t \in (0, 2k)$  at all dates  $t > s$ .

**Proof:** From lemmas A.4 and lemma A.6, we know that after some date  $s'$  we must be in the saddle path. Towards a contradiction, say that for  $t \in (s'', s')$  then  $x_t = 0$  even though for  $t < s''$ , it was that  $x_t \in (0, 2k)$ . Well from lemma A.5, equation (77) we know that:

$$\lim_{t \searrow s'} x_t = 0 \Rightarrow \frac{\rho}{2} - \frac{\phi}{v_{s'}^u} + 2\eta + 2k - \left( \frac{\rho + 2\eta + 2k}{2\omega} \right) \frac{v_{s'}^r}{v_{s'}^u} = 0, \quad (81)$$

where we used the fact that the posterior variances are state variables that cannot jump. But, likewise, by the same argument:

$$\frac{\rho}{2} - \frac{\phi}{v_{s''}^u} + 2\eta + 2k - \left( \frac{\rho + 2\eta + 2k}{2\omega} \right) \frac{v_{s''}^r}{v_{s''}^u} = 0. \quad (82)$$

However, for  $t \in (s'', s')$ , we know that  $\dot{v}_t^u < 0$  while  $\dot{v}_t^r = 0$ . Therefore the two equations cannot both hold leading to a contradiction. The case for  $x_t = 2k$ , during  $t \in (s'', s')$  is similar. ■

Finally, the last result uses lemmas A.5 and A.6 and was used in the proof of lemma A.4:

**Lemma A.8:** The path where  $x_t > 0$  for date  $t < s$  and  $x_t = 0$  for any date  $t > s$  is not a solution. Therefore only part (i) of lemma A.4 is optimal.

**Proof:** Using the phase diagram, if  $x_t > 0$  then at that date, we are to the right of the  $\dot{v}^r = 0$  line. If at date  $s$  we hit this line, then at date  $t < s$  we must have been above the stable arm. But in this case, the dynamic of the system imply that a date  $s$ ,  $v_s^r > v_*^r$ . This contradicts lemma A1. an alternative proof, using mathematics, instead of the diagram notes that just before date  $s$ :

$$\frac{\dot{v}_t^u}{v_t^u} = \frac{\rho}{2} - \left( \frac{\rho + 2\eta + 2k}{2\omega} \right) \frac{v_t^r}{v_t^u} \quad (83)$$

$$= \frac{\rho}{2} \left( 1 - \left( \frac{v_*^u}{v_*^r} \right) \frac{v_t^r}{v_t^u} \right). \quad (84)$$

But then, for we to stay at  $x_t = 0$  for all  $t > s$ , lemma 1 implies that  $v_t^r < v_*^r$ . In turn, since  $v_*^u$  is a lower bound for  $v_t^u$ , then  $v_t^u > v_*^u$ . Therefore,  $\dot{v}_{ut} > 0$ . But this is a contradiction since for  $x_t$  arbitrarily close to zero, we know that  $\dot{v}_t^u < 0$ . ■

Combining lemmas A1 to A.8, we can now prove proposition 2.

**Proof of Proposition 2:** Lemmas A.4 and A.8 showed that  $x_t$  approaches zero from above asymptotically. Working backwards in time, this implies that there exists a finite date  $s$ , such that for  $t > s$ , the dynamics of the posterior variances are described by lemma A.5. If that date  $s = T$ , then we must have started in the saddle path, and that is case i) in the proposition.

If  $s > T$ , then by lemma A.7 we know that for  $t \in [T, s)$ , then either  $x_t = 0$  or  $x_t = 2k$ . To show that the first case corresponds to  $v_T^r < f(v_T^u)$ , note that if  $x_t = 0$ , then:

$$v_s^r = v_T^r \quad (85)$$

$$v_s^u = e^{-2(\eta+k)(s-T)} v_T^u + \left( \frac{\phi}{2\eta + 2k} \right) \left( 1 - e^{-2(\eta+k)(s-T)} \right) \quad (86)$$

Then date  $s$  is the zero of the function:

$$g(s) \equiv v_s^r - f(v_s^u). \quad (87)$$

By assumption  $g(T) < 0$ . Moreover,  $\lim_{s \rightarrow \infty} g(s) = v_T^r - f(v_*^u) = v_T^r - v_*^r > 0$ . Finally, by equations (85)-(86) and lemma A.6 characterizing the stable arm:  $g'(s) = 0 -$

$f'(\cdot)(\partial v_s^u/\partial s) > 0$ . Therefore by Bolzano's theorem, there is a unique, finite  $s$ .

All that remains to show is that in the case where  $x_t = 2k$  for  $t \in [T, s)$ , then  $v_T^r > f(v_T^u)$ . The proof is identical to the case in the previous paragraph. Now,  $g(T) > 0$ , and  $\lim_{s \rightarrow \infty} g(s) = 0 - f(\phi/2\eta) < 0$ . Moreover,  $g'(s) = (\partial v_s^r/\partial s) - f'(\cdot)(\partial v_s^u/\partial s)$  and since  $x_t = 2k$ , we know that  $\dot{v}_t^r < 0$  and that  $\dot{v}_t^u > 0$ , so that  $g'(s) < 0$ . Therefore, again there exists a unique finite date  $s$  where we reach the stable arm. ■

#### A.5. Algorithm to find the dynamics after $\mathbf{T}$ , when $\alpha \neq 1$ .

The problem is still described by equations (36)-(39), but equation (35) is replaced by:

$$\max_{\{x_t\}_T^\infty} - \int_T^\infty e^{-\rho(t-T)} (z_t v_t^r + \omega v_t^u) dt, \quad (88)$$

and each agent takes as given the evolution of:

$$z_t = \frac{\alpha}{1 - \delta_t + \alpha \delta_t} \quad (89)$$

$$\dot{\delta}_t / \delta_t = -\bar{x}_t \quad (90)$$

with a starting value for  $\delta_T$  (if the announcement was made at date  $T$ , then  $\delta_T = 1$ ), and in equilibrium:  $\bar{x}_t = x_t$ . The optimality conditions are just as in (43)-(52), ( but where the equation for the co-state variable  $\lambda_t$  in (48) becomes:

$$\dot{\lambda}_t = (\rho + x_t)\lambda_t + z_t. \quad (91)$$

Redoing lemma A.5 now:

$$\frac{z_t}{\lambda_t} = \frac{\omega}{\psi_t} + \frac{\phi}{v_t^u} \text{ and so:} \quad (92)$$

$$\psi_t = \frac{z_t v_t^r - \omega v_t^u}{\phi} \text{ and } \lambda_t = \frac{\psi_t v_t^u}{v_t^r} \quad (93)$$

Taking time derivatives,:

$$\phi \dot{\psi}_t = \dot{z}_t v_t^r + z_t \dot{v}_t^r - \omega \dot{v}_t^u, \text{ leading to} \quad (94)$$

$$x_t = \frac{\rho}{2} - \frac{\phi}{v_t^u} + 2\eta + 2k - \left( \rho + 2\eta + 2k - \frac{\dot{z}_t}{z_t} \right) \frac{z_t v_t^r}{2\omega v_t^u}. \quad (95)$$

But recalling the dynamics for  $z_t$ , note that:

$$\frac{\dot{z}_t}{z_t} = \left( \frac{1-\alpha}{\alpha} \right) z_t \dot{\delta}_t = - \left( \frac{1-\alpha}{\alpha} \right) \frac{\alpha \delta_t x_t}{1 - \delta_t + \alpha \delta_t} \quad (96)$$

which after replacing leads to the very messy solution for  $x_t$  :

$$x_t = \frac{\rho}{2} - \frac{\phi}{v_t^u} + 2\eta + 2k \quad (97)$$

$$- \left( \rho + 2\eta + 2k + \frac{(1-\alpha) \delta_t x_t}{1 - \delta_t + \alpha \delta_t} \right) \frac{\alpha v_t^r}{(1 - \delta_t + \alpha \delta_t) 2\omega v_t^u} \quad (98)$$

Note that since  $\dot{\delta}_t/\delta_t = \dot{v}_t^r/v_t^r$ , this implies that  $\delta_t/v_t^r = c$  where  $c$  is a constant. I can therefore re-write the expression for  $x_t$  as:

$$x_t = \frac{\rho}{2} - \frac{\phi}{v_t^u} + 2\eta + 2k \quad (99)$$

$$- \left( \rho + 2\eta + 2k + \frac{(1-\alpha) c x_t}{\frac{1}{v_t^r} - c + \alpha c} \right) \frac{\alpha}{\left( \frac{1}{v_t^r} - c + \alpha c \right) 2\omega v_t^u}, \quad (100)$$

Solving this for  $x_t$  and plugging into the dynamic equations in (46), (47), and dividing leads to an ODE for  $v^r = f_c(v^u)$  just as in lemma A.5:

$$\text{write ODE MISSING} \quad (101)$$

Then, the algorithm consists of the following:

Step 1: Choose parameters  $(\rho, \omega, \phi, \eta, k, \alpha)$ , and input the initial conditions  $\delta_T, v_T^r$ , and  $v_T^u$ .

Step 2: Calculate  $c = \delta_T/v_T^r$ .

Step 3: Solve the ODE to find the stable arm:  $v_r = f_c(v_u)$

Step 4a: If  $v_T^r = f_c(v_T^u)$ , then you are in the saddle path, so just run the dynamic system of two equations.

Step 4b: If  $v_T^r < f_c(v_T^u)$ , set  $x_t = 0$ , and look for date  $s$  when the function  $g(s)$  has a zero:

$$g(s) = v_T^r - f_c(e^{-2(\eta+k)(s-T)}v_T^u + \left(\frac{\phi}{2\eta + 2k}\right) \left(1 - e^{-2(\eta+k)(s-T)}\right)). \quad (102)$$

Then from  $s$  onwards use the saddle path solution whereas before know that  $x_t = 0$ .

Step 4c: If  $v_t^r > f_c(v_T^u)$  too high, set  $x_t = 2k$ , and look for  $s$  when the function  $g(s)$  has a zero:

$$g(s) = v_T^r e^{-2k} - f_c(e^{-2\eta(s-T)}v_T^u + \left(\frac{\phi}{2\eta}\right) \left(1 - e^{-2\eta(s-T)}\right)). \quad (103)$$

Then from  $s$  onwards use the saddle path solution whereas before know that  $x_t = 2k$ .



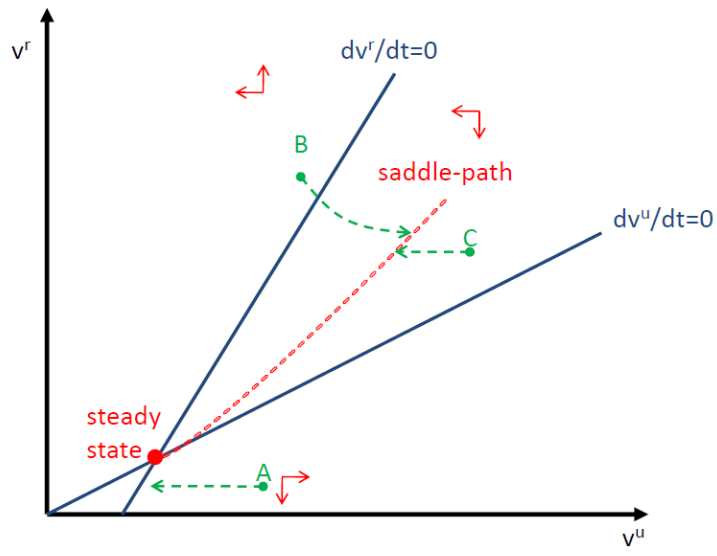
## References

- Alesina, Alberto and Allan Drazen (1991), “Why are stabilizations delayed?” *American Economic Review*, 81, 1170–88.
- Amador, Manuel and Pierre-Olivier Weill (2009), “Learning from prices: Public communication and welfare.” Stanford University manuscript.
- Angeletos, George-Marios and Alessandro Pavan (2007), “Efficient use of information and social value of information.” *Econometrica*, 75, 1103–1142.
- Athey, Susan, Andrew Atkeson, and Patrick J. Kehoe (2005), “The optimal degree of discretion in monetary policy.” *Econometrica*, 73, 1431–1475.
- Blinder, Alan S., Michael Ehrmann, Marcel Fratzscher, Jakob De Haan, and David-Jan Jansen (2008), “Central bank communication and monetary policy: A survey of theory and evidence.” *Journal of Economic Literature*, 46, 910–45.
- Bolton, Patrick and Antoine Faure-Grimaud (2010), “Thinking ahead: The decision problem.” *Review of Economic Studies*, xx, forthcoming.
- Cover, Thomas M. and Joy A. Thomas (1991), *Elements of Information Theory*. John Wiley and Sons, New York.
- Dewatripont, M and G Roland (1992), “Economic reform and dynamic political constraints.” *Review of Economic Studies*, 59, 703–30.
- Dewatripont, Mathias and Gerard Roland (1995), “The design of reform packages under uncertainty.” *American Economic Review*, 85, 1207–23.
- Duncan, Tyrone E. (1970), “On the calculation of mutual information.” *SIAM Journal of Applied Mathematics*, 19, 215–220.
- Eusepi, Stefano and Bruce Preston (2010), “Central bank communication and expectations stabilization.” *American Economic Journal: Macroeconomics*, forthcoming, forthcoming.
- Hellwig, Christian (2005), “Heterogeneous information and the welfare effects of public information disclosures.” UCLA manuscript.

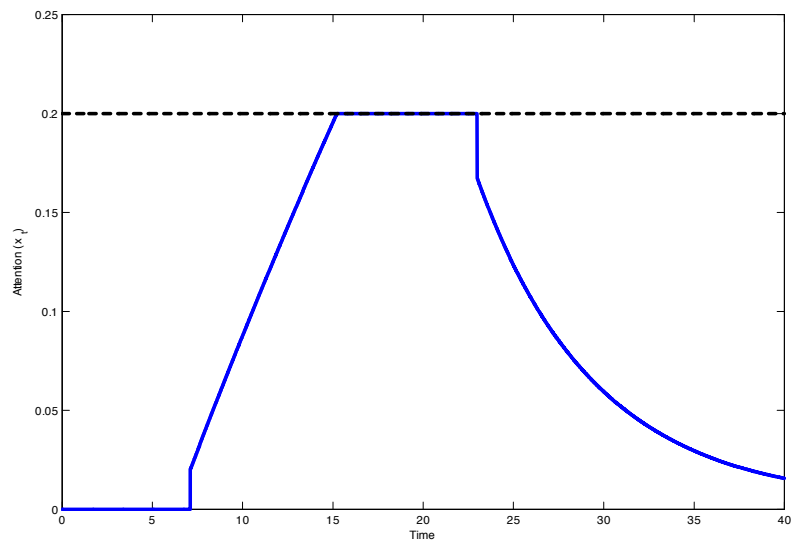
- Kalman, R. E. and R. S. Bucy (1961), “New results in linear filtering and prediction theory.” *Transactions of the ASME*, 83, 95–107.
- Lipster, Robert S. and Albert N. Shiryaev (2001), *Statistics of Random Processes*. Springer-Verlag, New York.
- Lorenzoni, Guido (2010), “Optimal monetary policy with uncertain fundamentals and dispersed information.” *Review of Economic Studies*, 77, 305–338.
- Mackowiak, Bartosz and Mirko Wiederholt (2009), “Optimal sticky prices under rational inattention.” *American Economic Review*, 99, 769–803.
- Mankiw, N. Gregory and Ricardo Reis (2002), “Sticky information versus sticky prices: A proposal to replace the new keynesian phillips curve.” *Quarterly Journal of Economics*, 117, 1295–1328.
- Mankiw, N. Gregory and Ricardo Reis (forthcoming), “Imperfect information and aggregate supply.” In *Handbook of Monetary Economics* (Benjamin Friedman and Michael Woodford, eds.), volume 3, Elsevier: North Holland.
- Morris, Stephen and Hyun Song Shin (2002), “Social value of public information.” *American Economic Review*, 92, 1521–1534.
- Morris, Stephen and Hyun Song Shin (2007), “Optimal communication.” *Journal of the European Economic Association*, 5, 594–602.
- Morris, Stephen, Hyun Song Shin, and Hui Tong (2006), “Social value of public information: Morris and shin (2002) is actually pro-transparency, not con: Reply.” *American Economic Review*, 96, 453–455.
- Moscarini, Giuseppe (2004), “Limited information capacity as a source of inertia.” *Journal of Economic Dynamics and Control*, 28, 2003–2035.
- Moscarini, Giuseppe (2007), “Competence implies credibility.” *American Economic Review*, 97, 37–63.
- Reis, Ricardo (2006), “Inattentive producers.” *Review of Economic Studies*, 73, 793–821.

- Shannon, Claude E. (1948), “A mathematical theory of communication.” *The Bell System Technical Journal*, 27, 379–423, 623–656.
- Sims, Christopher A. (2003), “Implications of rational inattention.” *Journal of Monetary Economics*, 50, 665–690.
- Sims, Christopher A. (forthcoming), “Rational inattention and monetary economics.” In *Handbook of Monetary Economics* (Benjamin Friedman and Michael Woodford, eds.), volume 3, Elsevier: North-Holland.
- Svensson, Lars E. O. (2006), “Social value of public information: Comment: Morris and shin (2002) is actually pro-transparency, not con.” *American Economic Review*, 96, 448–452.
- Van Nieuwerburgh, Stijn and Laura Veldkamp (2010), “Information acquisition and under-diversification.” *Review of Economic Studies*, forthcoming, forthcoming.

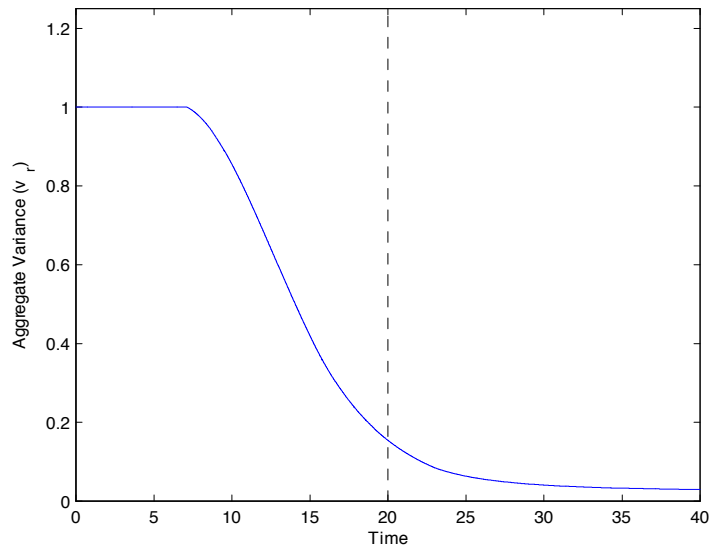
**Figure 1.** Phase diagram for dynamics with interior attention



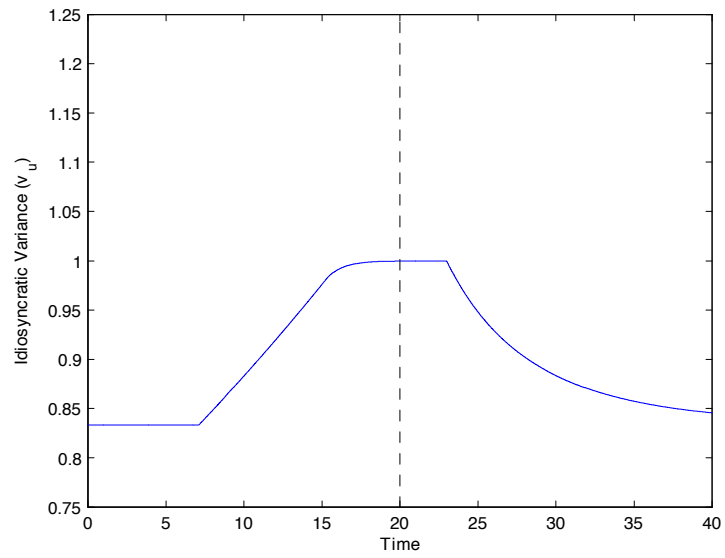
**Figure 1.** Phase diagram for attention allocation after regime change at date  $T$



**Figure 2.** Life-cycle of attention to the aggregate regime  $x_t$  after an announcement at date 0



**Figure 3.** Life-cycle of the posterior variance about the aggregate regime  $v_t^r$  after an announcement at date 0.



**Figure 4.** Life-cycle of the posterior variance about individual circumstances regime  $v_t^u$  after an announcement at date 0.