CSEF
CENTRO STUDI IN ECONOMIA E FINANZA
CENTRE FOR STUDIES IN ECONOMICS AND FINANCE

WORKING PAPER NO. 81

Optimal Stabilization Policy When the Private Sector Has Information Processing Constraints

Klaus Adam

April 2002

DIPARTIMENTO DI SCIENZE ECONOMICHE - UNIVERSITÀ DEGLI STUDI DI SALERNO
Via Ponte Don Melillo - 84084 FISCIANO (SA)
Tel. 089-96 3167/3168 - Fax 089-96 3169 – e-mail: csef@unisa.it

CSEF WORKING PAPER NO. 81

Optimal Stabilization Policy When the Private Sector Has Information Processing Constraints[*]

Klaus Adam[**]

Abstract

This paper considers a linear-quadratic control problem and determines how optimal policy is affected when the private sector has finite (Shannon) capacity to process information. Such capacity constraints prevent private agents from perfectly observing the state variables and the policy choices. The first result is that the control problem when including these constraints remains to be of a linear-quadratic form, which makes the problem technically tractable. The main difference to a standard problem are the costs associated with the use of the policy instrument, which are now endogenous. Depending on parameters these costs might be either higher or lower and lead to less or more aggressive optimal policies, respectively. If shocks show persistence and are heteroskedastic then the costs of using the policy instrument are non-constant and generate either sluggish or overshooting optimal policy reactions.

JEL classification: E30, E60

Keywords: optimal policy, Shannon capacity, communication theory, sluggish and overshooting policy, measurement errors.

[*] Thanks go to Pietro Corretto for introducing me to the topic and to Chris Sims for very helpful comments on an earlier version of the paper. Any errors are mine.

[**] Department of Economics, University of Frankfurt and CSEF, Università di Salerno

1 Introduction

A long-standing debate in economics is concerned with the question how stabilization policy should be conducted in practice.

While optimal policies in dynamic economic models typically prescribe to condition the policy choice on exogenous shocks and endogenous state variables, prominent economists have argued for constant or unconditional policies (e.g. Milton Friedman (1970)).

Most informal arguments that have been put forward against active stabilization policy involve informational problems such as uncertainty about the precise effects and time lags associated with particular policy actions.[1] The implications of such uncertainty for optimal policy has received considerable attention in the recent literature on robust control. Yet, optimal policies have often been found to be even more reactive than in the absence of such uncertainty.[2]

This paper is also concerned with the implications of information problems for optimal policy but looks at the problem from a different angle. In particular, the paper abstracts from uncertainty about the precise effects of policy. Instead, it considers a situation where economic efficiency requires coordination between the policy maker's and private sector decisions. Such coordination is required whenever optimal private sector decisions are functions of the policy choice and the state of the economy, as will be assumed in the paper. Informational frictions about the policy and the state then determine agents' ability to coordinate with the policy maker's choices and thereby affect welfare.

It is easy to think of examples that require the coordination of decisions. This is so because general equilibrium models by nature postulate the consistency between the policy maker's and the private sector's plans.[3] If monetary policy, for example, affects production costs then optimal price setting decisions of monopolistically competitive firms depend on it; similarly, if exchange rate policy affects the relative price of foreign goods then optimal production plans of domestic firms should depend on exchange rate policy; finally, when fiscal policy affects the composition of aggregate demand then the optimal allocation of labor and capital across sectors will be affected by it.

The novelty of the paper consists of endogenizing the degree of information frictions by explicitly modeling a link between the policy rule pursued by the policy maker and the informational frictions that exist on the side of the private sector. The modelling of this link is based on Shannon's (1948) theory of information transmission which postulates that agents receive information through

[1] Further arguments against stabilization policy are based on credibility problems or policy-irrelevance propositions in models with rational expectations, no nominal rigidities, and no informational asymmetries.

[2] See, for example, Onatski and Stock (2000) or Rudebusch (2001).

[3] Note that such coordination may be necessary even if systematic policy has no real effects.

so-called communication channels. These channels may either be interpreted in terms of physical channels, e.g. telegraph or telephone lines, or alternatively in terms of a model of the information flow in the mind of economic agents.

Communication channels transmit information only with finite capacity and thereby cause residual noise in agents' information sets, which I suggest to interpret in terms of a measurement errors of the variables of interest.[4] The paper shows that optimal use of the information channel by private agents implies that their measurement error has a non-classical feature: measurement errors are proportionate to the variance of the variables that agents wish to observe, i.e. measurement errors of the state (policy) turn out to be a fixed proportion of the variance of the state (policy) itself.

As a result of the link between policy and information frictions, policy not only affects the economic state but also the degree of coordination between the policy maker and the private sector. Policies that ignore this relationship or that postulate that measurement errors are exogenous will then be sub-optimal.[5]

Technically speaking the paper considers a linear-quadratic policy problem where a benevolent policy maker chooses a control variable to influence the state of the economy. In addition, private agents choose decision variables whose optimal values depend on the policy and on the state of the economy. Agents (optimally) use communication channels to obtain information about these two variables.

The paper shows that the control problem remains to be of simple linear quadratic form even when taking account of the costs associated with the information frictions. In particular, it is shown that there exists a stationary equilibrium in which the policy maker solves an alternative control problem where a different cost-weight is associated with the use of the instrument. Given the resulting policy, the private sector's optimal use of the communication channels generates coordinations costs which can be interpreted in terms of the changed cost-weight.

This implies that information frictions of the kind introduced in this paper have a simple reduced form representation, which should make the setup attractive for economic applications.

The paper finds that informational frictions can push policy either way: depending on parameters it may become either more or less aggressive when compared to a benchmark policy that assumes that information frictions are independent from policy.[6] In particular, I find that optimal policy is more

[4] Alternatively, the noise could be interpreted as resulting from information processing errors.

[5] This statement holds with proability one under generic parameterizations of the economy.

[6] Alternatively, the benchmark policy can be interpreted as the policy that is optimal in the absence of information frictions.

(less) aggressive, i.e. it reacts stronger (less strong) to past state deviations and to current shocks, whenever observation errors about the economic state (policy) constitute the dominant costs.

The intuition for these results are straightforward. Suppose observation errors about the policy are less costly than errors about the state. To minimize the costs of observation errors, the policy maker should choose a more variable policy to stabilize the state and to reduce agents' observation errors about the state. Clearly, this comes at the expense of increased observation errors about the policy choice. However, by assumption the latter costs are less important.

A similar logic leading to more attenuated policies applies when observation errors about the policy constitute the dominant costs.

The paper also considers a non-stationary situation where the shocks that hit the economy are heteroskedastic. This is of interest because information frictions are then equally non-stationary because agents use 'quite times' to improve information about past 'busy times'. The paper shows that this causes the cost-weights associated with use of the policy instrument to change over time, which generates auto-correlated policy *changes.* In particular, sluggish or overshooting optimal policy adjustment are shown to be optimal in such cases.

The use of Shannon's (1948) model of information transmission in economic contexts is not new to this paper. Sims (2001) has recently applied the theory in a model without a policy maker to study consumption behavior in a permanent income model. Earlier uses of the theory in microeconomic settings are due Marschak (1964) and Marschak and Miyasawa (1968).

The structure of the paper is as follows. Section 2 presents the policy problem. As a benchmark, Section 3 calculates the policy that is optimal when information frictions are independent from the policy rule. Section 4 then introduces Shannon's (1948) communication channels and derives basic results. Optimal policy when agents use these channels to inform themselves about the state and the policy is determined in Section 5. Finally, Section 6 presents an extension to heteroskedastic shocks. A conclusion gives a brief outlook of work ahead. The appendix collects most of the technical details.

2 The Policy Problem

This section constructs a linear quadratic policy problem, which describes the following economic situation: There is a benevolent policy maker who chooses a policy instrument $i_t \in R$ to influence the state $x_t \in R$ of the economy. Atomistic private agent with quadratic preferences over x_t and i_t take decisions $d_{1,t}$ and $d_{2,t}$ where $d_{1,t}$ and $d_{2,t}$ optimally depend on the state x_t and the policy i_t, respectively. Measurement errors prevent agents from perfectly observing x_t and i_t and might cause suboptimal private sector decisions $d_{1,t}$ and $d_{2,t}$.

The model is presented generically in terms of state and control variables. A monetary policy making example that fits into the presented framework is given in Adam (2001).

The state x_t of the economy evolves according to a linear law of motion given by

$$x_t = a_1 x_{t-1} + a_2 i_t + a_3 \overline{d}_{1,t} + a_4 \overline{d}_{2,t} + \varepsilon_t \tag{1}$$

The parameter a_1 determines how past states affect the current state; a_2 captures the influence of the policy on the current state; a_3 and a_4 determine how the private sectors' average decisions $\overline{d}_{1,t}$ and $\overline{d}_{2,t}$ affect the current state. I assume $a_2 \neq -a_3$ and $a_4 \neq 1$. Unless otherwise indicated the shock term ε_t is a Gaussian white noise process with

$$\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$$

The private sector consists of a continuum of identical agents. Each agent maximizes a quadratic utility function of the form

$$\max_{\{d_{1,t}, d_{2,t}\}} E_0 \left[\sum_{t=0}^{\infty} \beta^t (-x_t^2 - \gamma_0 i_t^2 - \gamma_1 (d_{1,t} - i_t)^2 - \gamma_2 (d_{2,t} - x_t)^2) \right] \tag{2}$$

Private agents' utility depends on the state, the policy, and on the deviations of their decisions from these variables. Under suitable assumptions, see appendix 8.1 for details, the objective function (2) can be seen as a quadratic approximation to a utility function of the form

$$U(x_t, i_t, d_{1,t}, d_{2,t})$$

When agents take the aggregate law of motion (1) and the policy i_t as given, the solution to (2) is trivially given by

$$d_{1,t} = E_t[x_t] = x_t + \delta_{x_t} \tag{3}$$
$$d_{2,t} = E_t[i_t] = i_t + \delta_{i_t} \tag{4}$$

where δ_{i_t} and δ_{x_t} denote measurement or observation errors, which are mean zero random variables with variance $\sigma^2_{\delta(i_t)}$ and $\sigma^2_{\delta(x_t)}$, respectively. The utility of the representative private agent is then given by

$$E_0 \left[\sum_{t=0}^{\infty} \beta^t \left(-x_t^2 - \gamma_0 i_t^2 - \gamma_1 \sigma^2_{\delta(i_t)} - \gamma_2 \sigma^2_{\delta(x_t)} \right) \right] \tag{5}$$

The last two terms in (5) indicate that agents dislike the fact that they cannot observe x_t and i_t perfectly. Intuitively, noise in agents' information about the state and the policy stance causes them to take suboptimal decisions, which in expectation generate a loss that is proportionate to the variance of the observation error.

Assuming that across agents observation errors about i_t and x_t are independent, the law of large numbers implies $\overline{d}_{1,t} = i_t$ and $\overline{d}_{2,t} = x_t$, which allows to express the aggregate law of motion (1) as

$$x_t = b_1 x_{t-1} + b_2 i_t + \eta_t \tag{6}$$

where

$$b_1 = \frac{a_1}{1-a_4}, \qquad b_2 = \frac{a_2 + a_3}{1-a_4}, \qquad \eta_t = \frac{1}{1-a_4}\varepsilon_t$$

The policy maker's problem now consists of choosing a policy $\{i_t\}$ that maximizes the utility (5) of the representative private agent subject to (6).

3 Optimal Policy with Classical Measurement Error

Suppose agents observe x_t and i_t disturbed by so-called classical measurement errors δ_{i_t} and δ_{x_t} in equations (4) and (3). With measurement error being classical the random variables δ_{x_t} and δ_{i_t} are independent from past, current, or future values of x_t and i_t. This in turn implies that the variances $\sigma^2_{\delta(x_t)}$ and $\sigma^2_{\delta(i_t)}$ in the policy objective (5) are independent of policy and, thus, can be dropped from the maximization problem. As a result, the policy maker's problem (5) simplifies to

$$\max_{\{i_t\}} E_0 \left[\sum_{t=0}^{\infty} \beta^t (-x_t^2 - \gamma_0 i_t^2) \right] \tag{7}$$

subject to (6). As is well known, optimal policy i_t is then of the form

$$i_t = c_0 x_{t-1} + c_1 \eta_t \tag{8}$$

In a stationary equilibrium maximizing (7) is identical to

$$\max_{c_0, c_1} \left(-\sigma_x^2 - \gamma_0 \sigma_i^2\right) \tag{9}$$

where σ_x^2 and σ_i^2 are the stationary variances of x and i, respectively. This delivers the following optimal reaction coefficients:

$$c_0 = \frac{1}{2\gamma_0 b_1 b_2}\left(-\gamma_0 b_1^2 + \gamma_0 + b_2^2 - \sqrt{\tau}\right) \tag{10}$$

$$c_1 = \frac{c_0\left(\gamma_0 b_1^2 - \gamma_0 - b_2^2\right) - 2b_1 b_2}{b_1\sqrt{\tau}} \tag{11}$$

with

$$\tau = (\gamma_0 b_1^2 + \gamma_0 + b_2^2)^2 - 4\gamma_0^2 b_1^2 \tag{12}$$

4 A Micro-Model of Information Transmission

This section presents a micro-model of information transmission based on Shannon's (1948) communication theory. As will be shown, the model generates a non-classical measurement error. In particular, the variance of the observation errors turns out to be a fixed proportion of the variance of the variable that agents want to observe, i.e.

$$\sigma^2_{\delta(y)} = \lambda \sigma^2_y \tag{13a}$$

where σ^2_y denotes the variance of the variable agents want to observe, $\sigma^2_{\delta(y)}$ the variance of the observation error, and $\lambda \in [0,1]$ is a constant which depends on the technology available to agents.

4.1 Quantifying Uncertainty

Consider the following economic situation. An agent wants to choose a decision D to maximize the quadratic objective

$$-E\left[D - Y\right]^2$$

where Y is a real-valued random variable whose stochastic properties are known to the decision maker. Suppose that initially the decision maker does not know anything about the particular realization y of Y.

I now define a measure, called entropy, that quantifies the uncertainty involved by not possessing information about the realizations y of Y.[7] The entropy $H(Y)$ of a random variable Y is defined as[8]

$$H(Y) = -\int_R \ln(p(y))p(y)dy$$

This measure of uncertainty has two intuitively appealing properties.[9] First, entropy is equal to zero if and only if there is one realization that occurs with probability one, i.e. in the absence of uncertainty. Otherwise, entropy is strictly positive. Second, for a given bounded set of realizations entropy is maximal if all the realizations occur with equal probability, which is the situation intuitively corresponding to the situation of highest uncertainty.

4.2 Information Transmission via Communication Channels

I shall now assume that information about the realization y of Y can be transmitted through a communication channel to the decision maker. The channel

[7] Defining uncertainty in terms of an entropy measure is just convenient and by no means crucial for the discussion that follows.

[8] For y with $p(y) = 0$ let $\ln(p(y))p(y) = 0$ in the following definition.

[9] See Shannon (1948) for 'hard' properties that make this measure unique.

can be thought of a communication line connecting the information source with the economic agent. The channel is fed with input signals which are delivered at the other end of the line as output signals which are observed by the agent. This setup could be interpreted literally in terms of a physical communication channel, e.g. a telegraph line. Alternatively, and probably more relevant, the channel may be interpreted as a model of the information flow from the outside world into the mind of agents.

For illustrative purposes consider the following binary communication channel:[10] suppose the channel is fed with zeros or ones as input signals and delivers zeros and ones as outputs signals at the other end of the line. An important feature of the channel is that it is less than perfect in the sense that it occasionally delivers a zero as output when a one has been entered as input or vice versa. Let $s \in \{0, 1\}$ and $r \in \{0, 1\}$ denote the signal sent as input and received as output, respectively. The structure of the noise in the channel can be described by a non-degenerate conditional pdf $n(s|r)$ that describes the likelihood of s having been sent when r has been received.

Next, let the random variable S with pdf $q(s)$ describe the likelihood with which input s is sent at a particular point in time. The stochastic structure of the channel inputs together with the noise structure $n(\cdot|\cdot)$ define a random variable R with pdf $q(r)$ describing the channel outputs. Suppose we can observe the channel output R and know the stochastic properties of the channel's noise. Then we can compute a measure of residual uncertainty about the channel input S, which is called the conditional entropy of S after observing R. Formally,

$$H(S|R) = -\sum_{r=0}^{1}\left[\sum_{s=0}^{1}\ln(n(s|r))n(s|r)\right]q(r) \tag{14}$$

The conditional entropy averages the entropies for a given observation r, as given by terms in the square brackets, weighting them with the likelihood of observing r.

In the absence of noise, i.e. if $n(0|0) = n(1|1) = 1$, the conditional entropy is equal to zero indicating that there is no residual uncertainty about the channel input after observing the channel output. Otherwise the conditional entropy is positive indicating that residual uncertainty about the channel input still exists due to the channel noise. Also, the conditional entropy can never be larger than the unconditional one, which is intuitive because in the worst case the output signals are completely uninformative about the input signals.

The reduction in entropy (or uncertainty) about the channel input from observing the channel output is given by

$$H(S) - H(S|R) \tag{15}$$

[10] The restriction to two signals is not crucial in any way. Alternatively, one could use as many signals as there are letters in the alphabet, or even continous signal spaces, if seen as limits of discretized signal spaces.

which is a positive number.

Clearly, the entropy reduction in (15) depends on the pdf $q(s)$ of channel inputs: The probability of observing a certain output signal $q(r)$ in equation (14), for example, depends on the probability $q(s)$ with which the respective inputs are sent. Therefore, expression (15) not only captures the properties of the channel itself but also the properties of the signals that are sent.

One can now define a measure, called the channel capacity, which is independent from the distribution of input signals. The channel capacity $\widetilde{C}$ is the maximum possible entropy reduction per sent input signal that can be achieved via the channel where maximization occurs with respect to the pdf of the channel inputs, i.e.

$$\widetilde{C} = \sup_{q(s)} \left(H(S) - H(S|R) \right)$$

Intuitively, one might think of the maximization operation as choosing that probability distribution of input signals that optimally takes account of the channel's noise to maximize the reduction in entropy per sent channel input signal. If the channel has a transmission rate of T signals per period, then the channel transmission capacity C is given by

$$C = \widetilde{C} \cdot T$$

which is the maximum achievable entropy reduction per period.

Mathematical communication theory now establishes a link between the channel transmission capacity C and the ability to transmit information about the realizations of Y through the communication channel. The following results are due to sections 13 and 24 in Shannon (1948):

There exists a coding system that maps realizations of Y into channel input strings of zeros and ones of length t such that the entropy reduction $H(Y) - H(Y|R)$ approaches C as $t \to \infty$, where R denotes the observed channel output string of length t. Moreover, there exists no coding system that achieves a higher entropy reduction than C for any t.

I now assume that the channel's transmission rate T is large such that the above limiting results provide reasonably good approximations to the problem.[11] The preceding results then imply that the channel's transmission capacity C is a sufficient statistic of the channel, independent from the alphabet of available input and output signals and the structure of the channel's noise: The entropy about Y can be at most reduced by at most C units via the channel.

[11] By letting $T \to \infty$ I avoid issues that are related to the information delay caused by coding systems that seek maximum entropy reduction. However, this feature of information transmission is interesting in its own right and might be explored in another paper.

Now reconsider our original problem where the decision maker wishes to

$$\max_D -E\,[D-Y]^2$$

The results above imply that if the decision maker observes information about the realizations of Y through a channel with transmission capacity C, then D is constrained by

$$H(Y) - H(X|D) \leq C \tag{16}$$

which says that the additional information about Y contained in D cannot exceed C.

Suppose that

$$Y \sim N(0, \sigma_y^2)$$

Then the entropy of Y is given by

$$H(Y) = \frac{1}{2}\left(\ln 2\pi e + \ln \sigma_y^2\right)$$

Appendix 8.2 shows that efficient use of the channel implies that observation errors about Y are independent normal random variables, i.e. agents observe

$$y + \delta$$

where δ is normal with variance. Given this information agents optimally choose

$$D = y + \delta$$

which implies that

$$H(Y|D) = H(\delta) = \frac{1}{2}\left(\ln 2\pi e + \ln \sigma_\delta^2\right)$$

Constraint (16) then implies that

$$\sigma_\delta^2 = e^{-2C}\sigma_y^2 \tag{17}$$

Equation (17) confirms the initial claim that communication channels generate non-standard observation errors: the variance of the observation error depends on the variance of the variable that agents want to observe where $\lambda = e^{-2C}$ in equation (13a).

As the channel capacity becomes infinite, observation errors disappear, indicating that the channel transmits all available information; as the channel capacity approaches zero, the variance of the observation errors become as large as the variance of Y itself, which indicates that the channel transmits no information about the realization of Y.

Equation (17) also reveals that there are decreasing returns to capacity: the additional reduction in the variance of the observation error that can be achieved by a marginal increase in capacity is a decreasing function of the existing channel capacity. This is an important feature, which will be exploited below.

5 Optimal Policy with Communication Channels

This section considers the control problem when agents have to use an information channel with fixed capacity C to obtain information about the current state x_t and the policy i_t.

5.1 Optimal Use of the Channel by Private Agents

A first problem that has to be considered is how agents should split the available capacity C to collect information about x_t and i_t.

Suppose that x_t and i_t are stationary processes. Furthermore, suppose that optimal policy when agents use information channels remains to be of the form (8). Obviously, this fact will have to be verified later on.

Using the law of motion (6) and the policy rule (8) both x_t and i_t can be expressed as functions of current and past values of the shocks η_t:

$$x_t = \sum_{i=0}^{\infty} (b_1 + b_2 c_0)^i (b_2 c_1 + 1)\eta_{t-i} \tag{18}$$

$$i_t = c_1 \eta_t + c_0 \sum_{i=0}^{\infty} (b_1 + b_2 c_0)^i (b_2 c_1 + 1)\eta_{t-i-1} \tag{19}$$

This suggests that collecting information about x_t and i_t in each period is equivalent to collecting information about the shock terms η_t in each period and using this information to reconstruct the values of x_t and i_t.

Suppose for a moment that agents allocate each period all available capacity to receiving information about the current shock η_t. I will now discuss under which conditions this is optimal.

First, consider the task of collecting information about x_t. The stationarity of x_t implies that $|b_1 + b_2 c_0| < 1$, which says that the influence of past shocks η_{t-i} on x_t decreases with i. This coupled with the fact that there are decreasing returns to capacity shows that agents have no incentive to allocate capacity to observing past shocks if they have observed these with capacity C in previous periods. Clearly, such a deviation would only increase the variance of the observation error of x_t.

Second, consider the task of collecting information about i_t. Clearly, if

$$|c_1| \geq |c_0 (b_2 c_1 + 1)| \tag{20}$$

then the influence of the current shocks is larger than the influence of past shocks and the allocation is equally optimal for i_t. However, if inequality (20) does not hold, then agents could be tempted to reallocate capacity from the current shock to previous shocks, even if they have observed these with capacity C in previous periods.

We summarize the previous results in the following lemma:

Lemma 1 *Suppose x_t is a stationary process, optimal policy is of the form (8), and inequality (20) holds. Then it is optimal for agents to use all available capacity to observe the current shock η_t. The variances of the observation errors are then given by*

$$\sigma^2_{\delta(x)} = e^{-2C}\sigma^2_x \tag{21}$$

$$\sigma^2_{\delta(i)} = e^{-2C}\sigma^2_i \tag{22}$$

Equations (21) and (22) directly follow from equations (17), (18) and (19).

5.2 Optimal Policy

I now consider the implications of agents' capacity allocations for optimal policy. I first establish that the structure of the optimal policy reaction function remains unchanged when information processing constraints are introduced:

Lemma 2 *If x_t is a stationary process and agents allocate all available capacity to observe the current shock η_t, then optimal policy is of the form (8).*

Proof. When agents observe only current shocks lemma 1 establishes that observation errors are given by equations (21) and (22). Substituting these into the objective function (5) and exploiting the stationarity of x_t, the policy maker's maximization problem can be written as

$$\begin{aligned} &\max E\left[\sum_{t=0}^{\infty}\beta^t\left(-x_t^2-\gamma_0 i_t^2-\gamma_1 e^{-2C}\sigma_i^2-\gamma_2 e^{-2C}\sigma_x^2\right)\right] \\ &= \max E\left[\sum_{t=0}^{\infty}\beta^t\left(-(1+\gamma_2 e^{-2C})x_t^2-(\gamma_0+\gamma_1 e^{-2C})i_t^2\right)\right] \end{aligned} \tag{23}$$

which is again a quadratic objective function. Thus, optimal policy continues to be of the form (8), as claimed. ■

Equation (23) implies that in a stationary equilibrium the policy maker's maximization problem can be written as

$$\max_{c_0,c_1}\left(-\sigma_x^2-\gamma\sigma_i^2\right) \tag{24}$$

where

$$\gamma = \frac{\gamma_0+\gamma_1 e^{-2C}}{1+\gamma_2 e^{-2C}} \tag{25}$$

and where the states evolves according to (6).

The solution to (24) is given by equations (10) and (11) with γ_0 substituted by γ. To establish that this is indeed an equilibrium it remains to show that the optimal reaction coefficients c_0 and c_1 satisfy inequality (20), which is a sufficient condition for the optimality of the assumed capacity allocation by the private sector. Appendix 8.4 shows that this is the case. We can summarize:

Proposition 3 *There exists a stationary equilibrium where private agents use all available capacity to observe current shocks, where optimal policy is of the form (8), and where the optimal reaction coefficients are given by equations (10) and (11) with γ_0 substituted by γ, as defined in (25).*

The only difference to a control problem without measurement error, as given in (9), is that the weight γ_0 attached to the instrument is replaced by the weight γ.

Whether $\gamma \gtrless \gamma_0$ depends on whether $\frac{\gamma_1}{\gamma_2} \gtrless \gamma_0$. The fraction $\frac{\gamma_1}{\gamma_2}$ can be interpreted as the costs of observation errors of i_t relative to the costs of observation errors of x_t, while γ_0 denotes the direct utility costs of deviations of i_t relative to direct costs of deviations of x_t. If

$$\frac{\gamma_1}{\gamma_2} > \gamma_0 \tag{26}$$

then $\gamma > \gamma_0$ for any $C < \infty$, which indicates that the use of the policy instrument is more costly than it was in the absence of information processing constraints. As information processing constraints disappear ($C \to \infty$), the weight γ monotonically decreases towards γ_0. Correspondingly, if $\frac{\gamma_1}{\gamma_2} < \gamma_0$, then $\gamma < \gamma_0$ and the use of the policy instrument is less costly than in the absence of information processing constraints.

The preceding results suggest that to characterize the effect of information processing constraints on optimal policy it is sufficient to consider the derivatives of the reaction coefficients with respect to the parameter γ. The next proposition establishes how the optimal reaction coefficients change in response to γ. The proof can be found in Appendix 8.3.

Proposition 4 *The absolute value of the optimal reaction coefficients c_0 and c_1 decrease with γ, i.e.*

$$sign(\frac{\partial c_0}{\partial \gamma}) = -sign(c_0)$$
$$sign(\frac{\partial c_1}{\partial \gamma}) = -sign(c_1)$$

If inequality (26) holds, information processing constraints prompt policy to react less aggressively than in the absence of these constraints. This is intuitive because (26) implies that observation errors about the policy instrument are relatively more important. To reduce these costs the policy maker reduces the variance of the policy instrument and thereby increases the precision with which agents can observe the instrument choice. Of course, this comes at the expense of tolerating higher variability in the state variable.

If inequality (26) does not hold, then information processing constraints cause policy to become more aggressive, i.e. to react stronger to past state deviations and current shocks because observation errors about the state are relatively more important. To reduce the observation errors about the state variable the policy maker reduces the variance of the state variable by increasing the variance of the policy instrument.

6 Optimal Policy with Heteroskedastic Shocks

This section considers a situation where heteroskedastic shocks hit the economy.

From the viewpoint of information theory this can be interpreted as there being 'busy' times with lots of new information arriving and 'quite' times with little or no news. Such a situation is of interest because in quite times agents have an incentive to allocate some of their capacity to past shocks to improve information the previous busy times where they could not follow developments very precisely.

As we shall see this provides an incentive for the policy maker to react with a delay to the past news because as information about the past improves the information costs of reacting to past news decrease.

I consider a simple situation, which allows for straight-forward analytical solutions. First, suppose that past states have no influence on the present state such that the aggregate law of motion is given by

$$x_t = b_2 i_t + \eta_t \tag{27}$$

Regarding the shock process I simply assume that with probability $1 - \lambda$ a new shock arrives and that with probability λ the previous shock continues to be effective, i.e.

$$\eta_t = \begin{cases} \eta_{t-1} \text{ with probability } \lambda \\ N(0, \frac{\sigma_\varepsilon^2}{(1-a_4)^2}) \text{ with probability } 1 - \lambda \end{cases}$$

This implies that in periods in which the previous shock continues to be effective, agents optimally use all available capacity to obtain more accurate information about the old shock.

As a consequence the simple setup is that there are no intertemporal linkages between periods except for the information collection process of private agents.

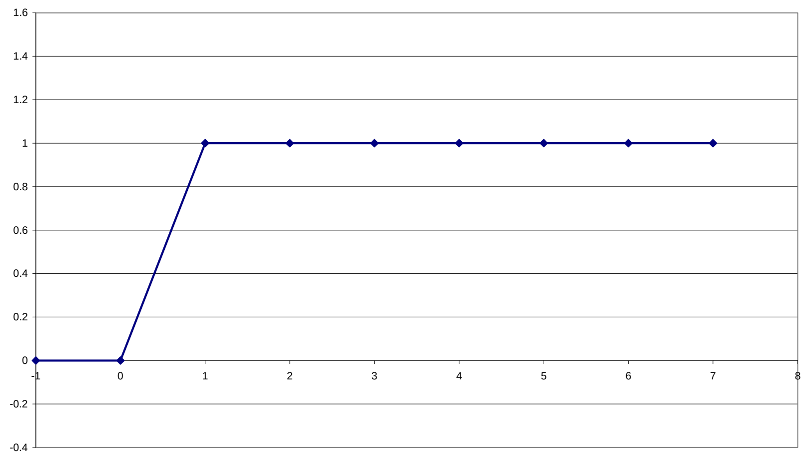

Figure 1: Optimal policy without information constraints

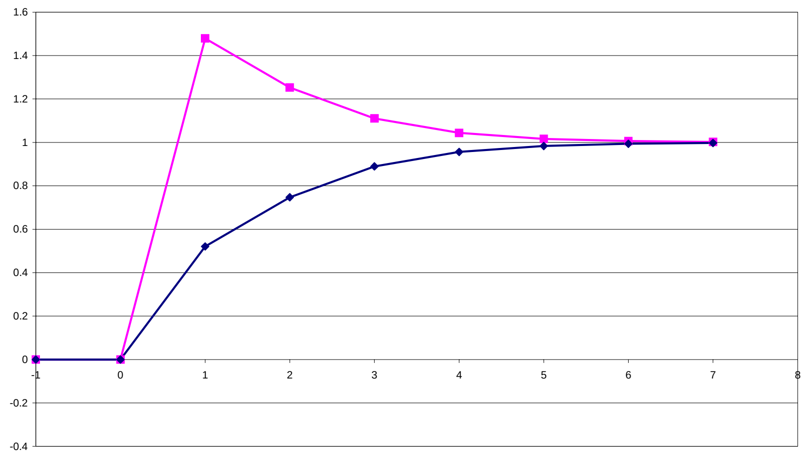

Figure 2: Sluggish and overshooting optimal policies with information constraints

Therefore, we can think of the policy maker's maximization problem in terms of separate maximization problems: one problem for periods in which a shock arrived, another one for periods in which a shock arrived one period ago, and so on. Clearly, in the absence of an impact of past states on the current state the policy maker has no incentive to react to past states, which implies that $c_0 = 0$ in (8).

Consider a period in which the current shock is effective since $n \geq 1$ periods. Agents already gathered information about this shock in the previous $n-1$ periods and can use the information capacity that is available in the current period to further reduce the observation error. It is easy to show that collecting for n periods information with capacity C about the same random variable is identical to collecting for one period with capacity nC.

As a result, optimal policy in a period when the current shock is effective since n periods can be calculated by maximizing

$$-x_t^2 - \gamma(n) i_t^2$$

where

$$\gamma(n) = \frac{\gamma_0 + \gamma_1 e^{-2nC}}{1 + \gamma_2 e^{-2nC}}$$

Consider the case $\frac{\gamma_1}{\gamma_2} > \gamma_0$ where the initial policy reaction to a shock is weaker than in the absence of information processing constraints. Clearly, as n increases $\gamma(n)$ will approach γ_0. This implies that the policy maker's reaction to the shock becomes stronger over time since improved information about the value of the shock reduces the costs associated with observation errors. Figure 2 illustrates a sluggish policy adjustment path for a shock that hits the economy in period 1 and that remains to be effective for the 7 periods. This should be confronted with figure 1 which shows the optimal policy adjustment in the absence of information processing constraints.

For the case $\frac{\gamma_1}{\gamma_2} < \gamma_0$ the initial policy reaction is stronger than in the absence of information processing constraints because information costs about the new state are relatively more important. Since $\gamma(n)$ again approaches γ_0 over time, this implies that there is an initial overshooting reaction in optimal policy, see figure 2 for an illustration.

7 Conclusions

This paper presents a first assessment of the impact of information processing constraints on optimal policy. Clearly, many open questions remain: What happens if agents face a trade-off between speed and accuracy of information transmission? What happens if the policy maker is subjected to similar information processing constraints as the private sector? What happens if the private sector has to take decisions that depend on both the state and the policy variable? Future research will have to explore these questions.

8 Appendix

8.1 Appendix 1

Agents in the private sector have the following utility function

$$\max_{\{d_t\}} \sum_{t=0}^{\infty} \beta^t U(d_{1,t}, d_{2,t}, i_t, x_t)$$

The one period utility function U is assumed to be twice continuously differentiable and convex in $(d_{1,t}, d_{2,t})$ for all (i, x). When the following second partial derivatives are equal to zero[12]

$$U_{d_1 x} = U_{d_2 i} = U_{d_1 d_2} = U_{ix} = 0 \tag{28}$$

optimal decisions can be expressed as

$$\begin{aligned} d_{1,t} &= f(i_t) \\ d_{2,t} &= g(x_t) \end{aligned}$$

Moreover, normalize x and i such that the one period utility

$$U\left(f(i_t), g(x_t), i_t, x_t\right)$$

is maximized for $i_t = x_t = 0$. Letting δ_i and δ_x denote mean zero observation errors a second-order Taylor approximation to the utility function around $i_t = x_t = \delta_i = \delta_x = 0$ is given by

$$\begin{aligned} & E\left[U(f(i_t + \delta_i), g(x_t + \delta_x), i_t, x_t) - U(f(0), g(0), 0, 0)\right] \\ & \approx \nabla_{(\delta_i, \delta_x, i, x)} U \cdot E\left[(\delta_i, \delta_x, i, x)'\right] + E\left[(\delta_i, \delta_x, i, x)\, \nabla^2_{(\delta_i, \delta_x, i, x)} U\, (\delta_i, \delta_x, i, x)'\right] \\ & = U_{xx} x_t^2 + U_{ii} i_t^2 + U_{\delta_i \delta_i} E\left[\delta_i^2\right] + U_{\delta_x \delta_x} E\left[\delta_x^2\right] \end{aligned}$$

where all first order terms are zero because the expected observation errors are zero and because of the normalization of x and i. Second order terms disappear because observation errors are independent from each other and from x and i and because of the assumptions in (28). Normalization of the utility function delivers (2) when substituting $d_{1,t}$ and $d_{2,t}$ by (4) and (3).

8.2 Appendix 2

Consider the problem

$$\max -E\left[D - Y\right]^2$$

[12] An example where this is the case is given by

$$U(d_{1,t}, d_{2,t}, i_t, x_t) = u(d_{1,t}, i_t) + v(d_{2,t}, x_t)$$

subject to

$$H(Y) - H(Y|D) \leq C$$

Letting σ_D^2, σ_Y^2 denote the variances of D and Y, respectively, and σ_{DY} their covariance, one can express this problem as

$$\min_{\sigma_D^2, \sigma_{DY}^2} -\sigma_D^2 + 2\sigma_{DY} - \sigma_Y^2$$

$$s.t.$$

$$H(Y) - \int_D H(Y|D = \Delta)\, p(\Delta) d\Delta \leq C \tag{29}$$

Now fix some choices for σ_D^2 and σ_{DY}^2. Then one can show that the functional form of the pdf $p(Y|D = \Delta)$ that maximizes the entropy $H(Y|D = \Delta)$ is a normal distribution, see Shannon (1948) section 20. Thus, by choosing $Y|D = \Delta$ normal for each Δ, the left-hand side of (29) is minimized and the constraint relaxed as much as possible.

The entropy of the normal variable $Y|D = \Delta$ with variance $\sigma_{Y|D=\Delta}^2$ is given by

$$H(Y|D = \Delta) = \frac{1}{2}\left(\ln 2\pi e + \ln \sigma_{Y|D=\Delta}^2\right)$$

and maximizing

$$H(Y|D) = \frac{1}{2}\int_D \left(\ln 2\pi e + \ln \sigma_{Y|D=\Delta}^2\right) p(\Delta) d(\Delta)$$

subject to

$$\sigma_Y^2 = \int_D \sigma_{Y|D=\Delta}^2 p(\Delta) d\Delta + \int_D (\mu_{Y|D=\Delta} - \mu_Y) p(\Delta) d\Delta$$

delivers that for any choice of the conditional means $\mu_{Y|D}$ one should choose $\sigma_{Y|D=\Delta}^2$ independent from Δ. If Y is normal then this implies that we can choose D to be jointly normal with Y. This implies that D has a representation as

$$D = X + \delta$$

where δ is a normal random variable, which is independent of X, as claimed.

8.3 Appendix 3

I first calculate the signs of the coefficients c_0 and c_1. Recall the optimal reaction coefficient c_0 in (10) and the definition of τ given in (12). Since

$$\left(-\gamma b_1^2 + \gamma + b_2^2 - \sqrt{\tau}\right) < 0$$

it follows that the sign of c_0 is given by

$$sign(c_0) = sign(-b_1 b_2) \tag{30}$$

Next, recall the optimal reaction coefficient c_1 in (11). If $\gamma b_1^2 - \gamma - b_2^2 \geq 0$ then (11) and (30) imply

$$sign(c_1) = sign(-b_2) \tag{31}$$

Now, suppose $\gamma b_1^2 - \gamma - b_2^2 < 0$. Substituting c_0 by (10) in the numerator of (11) delivers the following expression for the numerator

$$\frac{1}{2\gamma b_1 b_2}\left[\left(-\gamma b_1^2 + \gamma + b_2^2 - \sqrt{\tau}\right)\left(\gamma b_1^2 - \gamma - b_2^2\right) - 4\gamma b_1^2 b_2^2\right]$$

I will now show that the terms in the square brackets above are smaller than zero, which establishes that also for this case the sign of c_1 is given by (31):

$$\begin{aligned}
\left(-\gamma b_1^2 + \gamma + b_2^2 - \sqrt{\tau}\right)\left(\gamma b_1^2 - \gamma - b_2^2\right) - 4\gamma b_1^2 b_2^2 < 0 \Leftrightarrow \\
-\left(\gamma b_1^2 + \gamma + b_2^2\right)^2 + 4\gamma^2 b_1^2 - \sqrt{\tau}\left(\gamma b_1^2 - \gamma - b_2^2\right) < 0 \Leftrightarrow \\
\sqrt{\tau}\left(-\sqrt{\tau} - \left(\gamma b_1^2 - \gamma - b_2^2\right)\right) < 0 \Leftrightarrow \\
-\sqrt{\tau} - \left(\gamma b_1^2 - \gamma - b_2^2\right) < 0 \Leftrightarrow
\end{aligned}$$

Since both $-\left(\gamma b_1^2 - \gamma - b_2^2\right) > 0$, this is equivalent to

$$\begin{aligned}
-\left(\gamma b_1^2 - \gamma - b_2^2\right) < \tau \Leftrightarrow \\
0 < 4\gamma b_1^2 b_2^2
\end{aligned}$$

For the borderline cases $\gamma = 0$ it is $c_1 = -\frac{1}{b_2}$, similarly for $b_1 = 0$ we have $c_1 = -\frac{b_2}{b_2^2+\gamma}$ such that also in these case (31) holds.

Next, I calculate the sign of the derivatives of c_0 and c_1 with respect to γ. Taking the derivative of c_0 delivers

$$\frac{\partial c_0}{\partial \gamma} = \frac{b_2}{b_1}\frac{\gamma + \gamma b_1^2 + b_2^2 - \sqrt{\tau}}{2\sqrt{\tau}\gamma^2}$$

Since

$$\gamma + \gamma b_1^2 + b_2^2 - \sqrt{\tau} > 0$$

it follows that

$$sign(\frac{\partial c_0}{\partial \gamma}) = sign(-b_1 b_2) = sign(-c_0)$$

Next, take the derivative of c_1 with respect to γ:

$$\frac{\partial c_1}{\partial \gamma} = -b_2 \left(\frac{\left(\sqrt{\tau} - \left(\gamma b_1^2 + \gamma + b_2^2\right)\right)\tau + \gamma^2 b_1^2 b_2^2}{\gamma^2 b_1^2 \tau^{\frac{3}{2}}} \right) \tag{32}$$

Now consider the first bracket in the numerator above. Using a first order Taylor approximation to the root and exploiting its concavity one gets

$$\begin{aligned} \left(\sqrt{\tau} - \left(\gamma b_1^2 + \gamma + b_2^2\right)\right) &= \sqrt{\left(\gamma b_1^2 + \gamma + b_2^2\right)^2 - 4\gamma^2 b_1^2} - \left(\gamma b_1^2 + \gamma + b_2^2\right) \\ &< \frac{-2\gamma^2 b_1^2}{\left(\gamma b_1^2 + \gamma + b_2^2\right)} \end{aligned}$$

Using this expression, a sufficient condition for the numerator in (32) to be negative is given by

$$\frac{-2\gamma^2 b_1^2}{\left(\gamma b_1^2 + \gamma + b_2^2\right)}\tau + \gamma^2 b_1^2 b_2^2 < 0$$

Multiplying by $\left(\gamma b_1^2 + \gamma + b_2^2\right)$, substituting the expression for τ, and collecting terms delivers

$$-2\gamma^4 g_1^2 (g_1^4 - 2g_1^2 + 1) - 3\gamma^3 g_1^2 g_2^2 - \gamma^2 g_1^2 g_2^4 - 3\gamma^3 g_1^4 g_2^2 < 0$$

Since $(g_1^4 - 2g_1^2 + 1) \geq 0$, the previous inequality is satisfied, which shows that the numerator in (32) is negative. Equation (32) then implies

$$sign\left(\frac{\partial c_1}{\partial \gamma}\right) = sign(b_2) = sign(-c_1)$$

8.4 Appendix 4

We want to show that

$$c_1^2 - \left(c_0(b_2 c_1 + 1)\right)^2 \geq 0 \tag{33}$$

Using the expressions for reaction coefficients in (10) and (11), one obtains

$$\begin{aligned} & c_1^2 - \left(c_0(b_2 c_1 + 1)\right)^2 \\ &= \frac{1}{2}\frac{-g_2^4\tau + \left(\gamma^2 g_1^2 - \gamma^2\right)\tau - 2\gamma g_2^2 \tau + \sqrt{\tau}\left(g_2^2\tau + \gamma\tau - \gamma g_1^2 g_2^2(\gamma + \gamma g_1^2 + g_2^2)\right)}{\gamma^4 g_1^6 g_2^2} \end{aligned} \tag{34}$$

where τ is given by (12). Equation (34) is larger than zero if and only if

$$\left(g_2^2 + \gamma\right)\tau + \left(-g_2^4 + \gamma^2 g_1^2 - \gamma^2 - 2\gamma g_2^2\right)\sqrt{\tau} + -\gamma g_1^2 g_2^2 (\gamma + \gamma g_1^2 + g_2^2) \geq 0$$

which is obtained by dividing the numerator in (34) by $\sqrt{\tau} > 0$. Using (12) to substitute τ, this is equivalent to

$$0 \leq g_2^6 + 3\gamma g_2^4 + \gamma g_2^4 g_1^2 + 3\gamma^2 g_2^2 - \gamma^2 g_2^2 g_1^2 + \gamma^3 g_1^4 \tag{35}$$

$$- 2\gamma^3 g_1^2 + \gamma^3 + \left(-g_2^4 + \gamma^2 g_1^2 - \gamma^2 - 2\gamma g_2^2\right)\sqrt{\tau} \tag{36}$$

Now consider two cases:

- Case A: $\left(-g_2^4 + \gamma^2 g_1^2 - \gamma^2 - 2\gamma g_2^2\right) \geq 0$
- Case B: $\left(-g_2^4 + \gamma^2 g_1^2 - \gamma^2 - 2\gamma g_2^2\right) < 0$

First, consider case A. Using $\sqrt{\tau} > g_2^2$, a sufficient condition for (35) to hold is given by

$$\gamma g_2^4(1+g_1^2) + 2\gamma^2 g_2^2 + \gamma^3(g_1^4 - 2g_1^2 + 1) \geq 0$$

Since$(g_1^4 - 2g_1^2 + 1) \geq 0$, equation (35) and, therefore, equation (33) holds in this case.

Next, consider case B. Rewrite (35) as

$$-g_2^2(-g_2^4 + \gamma^2 g_1^2 - \gamma^2 - 2\gamma g_2^2) + 2\gamma^2 g_2^2 + \gamma g_2^4 + \gamma g_1^2 g_2^2 + \gamma^3(g_1^4 - 2g_1^2 + 1) \tag{37}$$

$$\geq -\left(-g_2^4 + \gamma^2 g_1^2 - \gamma^2 - 2\gamma g_2^2\right)\sqrt{\tau} \tag{38}$$

Clearly, both sides of equation (37) are positive. Now divide equation (37) by $-\left(-g_2^4 + \gamma^2 g_1^2 - \gamma^2 - 2\gamma g_2^2\right) > 0$ and square the result:

$$\left(\gamma g_1^2 + \gamma + g_2^2 - 2\gamma^2 g_1^2 \frac{\gamma g_1^2 - \gamma - g_2^2}{-g_2^4 + \gamma^2 g_1^2 - \gamma^2 - 2\gamma g_2^2}\right)^2 \geq \tau$$

Using the definition of τ in (12) this can be shown to be equivalent to

$$-4\gamma^4 g_1^6 \frac{g_2^4}{\left(-g_2^4 + \gamma^2 g_1^2 - \gamma^2 - 2\gamma g_2^2\right)^2} \leq 0$$

which is satisfied. Thus, also in this case equations (35) and (33) hold.

References

Adam, Klaus, "Optimal Monetary Policy with Information Processing Constraints," 2001. mimeo.

Friedman, Milton, "A Theoretical Framework for Monetary Analysis," *Journal of Political Economy*, 1970, *80*, 193–238.

Marschak, Jacob, "Problems in Information Economics," *in Charles P. Bonini, Robert K. Jeadicke, and Harvey M. Wagner (eds), Seminar on Basic Research in Management Controls: "Management Controls: New Directions in Basic Research", New York, McGraw-Hill*, 1964, pp. 38–74.

___ **and Koichi Miyasawa**, "Economic Comparability of Information Systems," *International Economic Review*, 1968, *9*, 137–174.

Onatski, Alexei and James Stock, "Robust Monetary Policy under Model Uncertainty in a Small Model of the U.S. Economy," *NBER Working Paper 7490*, 2000.

Rudebusch, Glenn D., "Is the Fed Too Timid? Monetary Policy in an Uncertain World," *Review of Economics and Statistics (forthcoming)*, 2001.

Shannon, C. E., "A Mathematical Theory of Communication," *The Bell System Technical Journal*, 1948, *27*, 623–656.

Sims, Christopher, "Implications of Rational Inattention," 2001. Princeton University Mimeo.