Dynamic Stackelberg Game with Risk-Averse Players: Optimal Risk-Sharing under Asymmetric Information

Dan Protopopescu[1]
Center for the Study of Organizations and Decisions in Economics
Universitat Autònoma de Barcelona
08193 Bellaterra, Barcelona, Spain
E-mail: *dan.protopopescu@uab.cat*

November 2009

Abstract

The objective of this paper is to clarify the interactive nature of the leader-follower relationship when both players are endogenously risk-averse. The analysis is placed in the context of a dynamic closed-loop Stackelberg game with private information. The case of a risk-neutral leader, very often discussed in the literature, is only a borderline possibility in the present study. Each player in the game is characterized by a risk-averse type which is unknown to his opponent. The goal of the leader is to implement an optimal incentive compatible risk-sharing contract. The proposed approach provides a qualitative analysis of adaptive risk behavior profiles for asymmetrically informed players in the context of dynamic strategic interactions modelled as incentive Stackelberg games.

Keywords: Dynamic stochastic Stackelberg game, optimal path, closed-loop control, endogenous risk-aversion, adaptive risk management, optimal risk-sharing.

JEL Classifications: C71, C73, D81, D82.

1. Introduction

Many economic problems are characterized by the presence of both strategic behavior and asymmetric information. Stackelberg (or leader-follower) games are, in this sense, useful tools for studying dynamic behavior in equilibrium settings in which some player is dominant.

The leader is supposed to know the objective function of the follower while this last one knows the control strategy of the first. The optimal strategy of the follower depends on the strategy selected by the leader. The objective function of the leader may depend not only on his own decisions but also on the follower's.

The leader is able to make his decisions by estimating the follower's rational reactions, assuming that he behaves in such a way that he optimizes his objective function given the leader's actions.

The distribution of information among the players plays a crucial role in determining their actions and leads to discontinuities in their behavior. In general, the players differ in their information systems and beliefs.

[1]Financial support from the Spanish Ministry of Education and Science through grant **PTA-2003-02-00877** and **FEDER** is acknowledged.

This asymmetry of information between the leader (the uninformed party but the game form designer) and the follower (the informed party detaining some private information) as well as the insufficient knowledge of the leader is a source of adverse selection.

Optimal risk-sharing between two parties has first been analyzed by Borch (1962) in the context of a static reinsurance contracting problem. Understanding the extent of risk-sharing is important for assessing the performance of the follower as well as the efficiency of the leader's incentive policy. It is also important to the proper modelling of the economy. When both contracting parties are risk-averse, it is optimal for them to share risk.

In the real world, there generally exist conflicts between follower's efficiency and efficiency in risk-sharing. It is useful to distinguish between risk-sharing and incentives. The private information can be acquired before the signature of the contract or rather after the contract has been signed. In other words, the contract between both sides must rather be settled before the state of nature is known. Risk-sharing is thus an objective which must be solved before the signature of the contract, while allocative efficiency becomes the only concern of the parties ex-post. This point of view, specific to the traditional approach, is not totally justified when considering dynamic interactions in highly fluctuating environments.

The leader-follower problem can be viewed as a problem in economizing on information. This situation can generate a different qualitative behavior for the players. The leader's objective is to design a contract mechanism consistent with incentive-compatibility and individual rationality-constraints (the contract must produce truthful revelation of the follower's type, and the follower must find it profitable to accept the contract). Its role is to influence the allocation of ex-ante risks optimally. There is a complex trade-off between hidden information, moral hazard and risk-sharing.

For constructing such a mechanism, the leader needs very precise knowledge about the follower's preferences and the technology he controls. The incentive mechanism is generally subjected to inescapable informational constraints. This is sensitive to the description of the environment and the size of informational asymmetries.

Correct incentives are necessary to share information and act appropiately. These will typically be provided at each period of time. It is important to note that the leader does not absolutely incur the information acquisition cost in any period of the contract. This may be quite costly. Information asymmetry and costly acquisition of information impose restrictions on the leader's behavior.

Suppose that both players use a closed-loop strategy, that is, they set their control vectors as functions of the history of the state vector from the start of the game to the moment of decision. Although past behavior may be a poor guide for estimating the effects of policy actions, it is possible to control these effects by going back in time.

Closed-loop dynamic games are appropriate modelling approaches to policy analyses even if the closed-loop solutions are often computationally intractable. The closed-loop control can be viewed as an alternative method to economic planning.

The paper is organized as follows. Section 2 describes the model. Section 3 formulates the discrete-time incentive contracting problem. Section 4 deals with asymmetric risk-averse behavior. Section 5 presents several qualitative results on adaptive risk management. Section 6 makes distinction between more and less risk-averse leaders /followers. Section 7 improves the concept of closed-loop dynamic Stackelberg equilibrium by taking into account endogenous risks exhibited by the players. An analytical characterization of the closed-loop solution path is provided in this sense. Section 8 concludes and makes suggestions for further research.

2. The Model

We consider a dynamic risk-sharing contract between a leader and a follower. This is modelled by a discrete-time finite-horizon environmental game with coupling constraints.

The leader is an environmental authority whose objective is the minimization of the environmental risks associated with important pollution levels generated by a potential polluter company. He has to account for the sequential nature of the polluter's decision strategy in designing the optimal contract.

Environmental decisions are sequential decisions with different degrees of future commitment. These represent an important class of stochastic control problems that arise in many other fields of economics.

The specificity of the economic activity is to superimpose on the initial equilibrium of the system another one richer in complexity, and thus more instable. The flexibility of the system is not infinite, and the disequilibrium caused by the deviating behavior of the polluter can be irreversible after some critical threshold is reached. The stability of the system is thus compromised, resulting in potentially large losses from coordination. Damages can be partially or totally irreversible.

In the literature on environmental policy making, two types of pollution are most often analyzed:

i) punctual pollution, when the polluter is identified and his monitoring does not generally cause problems, only maybe from the financial point of view (Caussade and Prat 1990);

ii) diffuse (or non-point) pollution, when the polluter cannot be identified, and hence monitored. In this case, it is costly for the authority to evaluate the responsibility of the polluter, and therefore to negotiate with him (Shortle et al. 1998). Some incentive measures based on the pollution sources (Wu et al. 1995) or subventions for developing pollution control equipments can be adopted.

One can distinguish two alternatives for the authority to control pullution:

i) to fix the polluter's output targets. Players have generally different preferences on the outcomes, this asymmetric behavior inducing conflict. However, this impediment can be resolved by cooperation (the outcome is negotiated);

ii) to impose certain thresholds for the targets, which must not be exceeded by the polluter. The state variable can thus float freely until it will reach the absorbing barrier fixed by the authority, at which time she intervenes to keep that fixed thereafter.

For further interesting aspects on the analysis of strategic conflicts, see Myerson (1997).

Under imperfect information, the ex-ante optimal contract typically exhibits ex-post inefficient outcomes. The authority must be able to renegotiate away the ex-post inefficiencies. In the over-pollution context, the penalty problem is generally transformed into an incentive constraints problem.

Due to uncertainties and irreversibilities that are inherent in environmental problems, environmental policy design can involve important problems of timing or stopping (Pindyck 2001).

Externalities do occur. In general, the polluter knows more about expected externalities than the authority does. There is an intimate link between externalities, uncertainty, private information and different objectives among the players.

The most efficient mechanisms of regulation externalities consist in endogenous financial incentives based on the level of pollution measured in the environment. These have the role to compensate the risk and to limit the effects of adverse selection. Two cases are of interest:

i) when the information available to the authority is sufficient to determine the externalities produced by the polluter's actions;

ii) when the polluter's actions reveal relevant information to the authority.

Incentives are transfer payments that may take the form of penalties, taxes, compensations, and so on. The incentive transfers generally depend on the information detained by the authority as well as on the polluter's actions. Transfers are costly. These can be made at each stage of the game or at the conclusion of the process. This way to define the incentives permits to the polluter to concentrate his efforts on the environmental problem, being also adapted to the control of the effects due to the inherent stochasticity of the process (Cabe and Herriges 1992; Xepapadeas 1991, 1992, 1994; Shortle and Dunn 1986; Meran and Schwalbe 1987; Segerson 1988).

The polluter has the possibility to choose freely his optimal strategy of pollution reduction, this permiting to realize his objective at a lesser cost. He must measure in a reliable manner the concentrations of the pollution and to determine the setting and the period of sample the most appropiate. The disadvantage of this kind of incentive is that when it is applied without discrimination between several polluters, this will be expensive and implies a responsability of group, being difficult to manage within the regions where the polluters constitute some pressure groups. In this case, the authority's objective is to induce a group-optimal behavior (Krawczyk 1985; Bystrom and Bromley 1998).

The incentive policy implemented by the authority consists in a dynamic taxation when the output observed is superior to a fixed limit threshold (we call this negative externality) or to recompense the polluter when the fixed limit threshold is not exceeded (we call this positive externality).

When it is not profitable to invest into individual controls, the authority can only observe the state of the environment. The authority does not need to control continually the polluter. It may be possible to choose a stochastic monitoring (Roth et al. 1989). This is the case when the authority and the polluter interact on a long term, when it is often advantageous to commit to a long term incentive scheme, specifying all potential payments and taxes for the polluter as time and events unfold. It relieves the authority.

In real economic applications, it appears that to achieve environmental objectives, a combination of taxes with incentive-compatible regulatory schemes is often necessary. The optimal policy implemented by the authority is thus determined by the balance of control versus incentive.

The type of game we study is composed of two distinct periods, say $[-T_1, 0]$ and $[1, T_2]$. The second playing period corresponds to a renegociation of the contract terms based on the polluter's performances during the first playing period.

It is possible that after the contract is finished, the players renegotiate the original contract looking for mutually beneficial gains. The goal structure of the players may change. Other potential advantages may thus be exploited. This possibility can modify the dynamic incentive compatibility constraints and may have positive implications for the equilibrium of the game. It is useful to point out that negative externalities may cause delay in negotiation (Jehiel and Moldovanu 1995).

In what follows, we investigate the case where the authority does not intervene in the selection of the polluter's targets, but imposes certain thresholds over which the polluter will be taxed. We only analyze the second contract period $[1, T_2]$ by taking into account the history of the previous game during the playing period $[-T_1, 0]$.

The players are indexed by an ordered pair $(i, j) \in \{(1, 2), (2, 1)\}$, where i is the player on whom attention is focused and j is the opponent. Player 1 is the authority /leader.

The type of model we analyze corresponds to a data generating process which is dynamic, linear and managed by a system of discrete simultaneous equations.

We make the following basic assumptions:

Assumption 1. The evolution of the environment is modelled by a discrete-time multivariate linear stochastic process:

$$y_t = A_t y_{t-1} + C_t^{(1)} x_t^{(1)} + C_t^{(2)} x_t^{(2)} + B_t z_t + D_t + u_t,\ t = 1, ..., T_2$$

where:
i) $\beta_t \overset{not.}{=} (A_t, C_t^{(1)}, C_t^{(2)}, B_t, D_t) \in \mathbf{R}^k$ is the time-varying parameter to be estimated;
ii) $x_t^{(1)} \in \mathbf{R}^{q_1}$ represents the authority's control variable at period t;
iii) $x_t^{(2)} \in \mathbf{R}^{q_2}$ represents the polluter's control instrument at time t;
iv) $z_t \in \mathbf{R}^r$ is an exogenous variable observed outside the system under consideration, and hence not subjected to the control;
v) $u_t \in \mathbf{R}^p$ is an exogenous "white noise" disturbance modelled by a normal random variable with zero mean-vector and finite variance-covariance matrix Ψ;
vi) T_2 represents the end of the planning horizon.

The magnitude of the output values are important determinants of the environmental policy. Let y_t^a ($t = 1, ..., T_2$) be the output thresholds not to be exceeded (regarded as reference points-limits) that are fixed by the authority. These are strategic constraints to be satisfied by the polluter during the entire contract period.

Remark 1. There is presumably a natural level of pollution that exists in the absence of any economic activity. It may also exist situations where the level of pollution in t is added to the level of pollution already existent in $t-1$. Pollution can thus accumulate in the environment sufficiently to become irreversible.

Assumption 2. The polluter's objective is to constrain the system to follow a fixed path $\eta \overset{not.}{=} \{y_1^g, y_2^g, ..., y_{T_2}^g\}$ reflecting his preferences on the outcomes.

Since a real time control process is necessarily discrete, this cannot converge with precision to any target value, but only to a neighborhood of it. When the control process is finished, the polluter will only obtain a stochastic neighbouring-optimal trajectory which is expected to be close to the desired path η.

Taking into account foreseeable movements in y as well as possible economic constraints, the polluter will impose the following local restrictions on the targets:

$$y_t^{n.\ l.} < y_t^g \leq y_t^a \text{ and } \parallel y_t^a - y_t^g \parallel < l_t < 1,\ t = 1, ..., T_2$$

where l_t is a strategic parameter chosen according to the incentive mechanism implemented by the authority (for further details, see Section 3).

Typically, the environment is stochastic and non stationary by nature. There is thus a natural tendency for the polluter to overcome the fixed targets due to unobserved shocks which escape to his control, and hence to be taxed by the authority.

Remark 2. Consider that two deviations of the system with respect to the polluter's fixed targets are comparable in magnitude if and only if their ratio is very close to 1.

Assumption 3. The timing of the control is as follows: At each period t, the players implement their actions $x_t^{(i)}$ ($i = 1, 2$), which are a stimulus for the system. A shock u_t is realized and they observe the output (or impulse response) y_t. The uncertainty is reduced only ex-post, that is, only after the informative output-signal has been received.

The polluter employes this output for a strategic learning (specific to a closed-loop monitoring) in order to drive the system as close as possible to the desired path η.

Assumption 4. The optimality of the instruments $x_t^{(i)}$ $(i = 1, 2)$ is considered with respect to a global criterion $W_{[1,T_2]}^{(i)}$ (supposed twice continuously differentiable, strictly increasing and convex) which measures the deviations $\triangle^a\ y_t \overset{not.}{=} y_t - y_t^a$ (for the authority) and, respectively, $\triangle^p\ y_t \overset{not.}{=} y_t - y_t^g$ (for the polluter).

Using the traditional approach (Van der Ploeg 1984), we consider as sufficient a quadratic additive recursive criterion:

$$W_{[1,T_2]}^{(1)}(y_1, ..., y_{T_2}) \overset{def.}{=} \sum_{t=1}^{T_2} W_t^{(1)}(y_t) \text{ (for the authority)}$$
$$W_{[1,T_2]}^{(2)}(y_1, ..., y_{T_2}) \overset{def.}{=} \sum_{t=1}^{T_2} W_t^{(2)}(y_t) \text{ (for the polluter)}$$

where $W_t^{(i)}$ $(i = 1, 2)$ are quadratic asymmetric loss functions defined by:

$$W_t^{(1)}(y_t) \overset{def.}{=} (y_t - y_t^a)' K_{t,1}(y_t - y_t^a) + 2(y_t - y_t^a)' d_{t,1}$$
$$W_t^{(2)}(y_t) \overset{def.}{=} (y_t - y_t^g)' K_{t,2}(y_t - y_t^g) + 2(y_t - y_t^g)' d_{t,2}$$

with a prime denoting transpose.

The decision for choosing certain parameters $K_{t,i}$ and $d_{t,i}$ reflects the players' priorities and also depends on the available amount of information concerning the future development of the system parameters. At each period t, the parameters $K_{t,i}$ and $d_{t,i}$ are updated and new optimal values are chosen in order to satisfy the players' requirements.

This is the classical context, often employed in the literature, where both players are considered risk-neutral in their preferences during the entire period of the contract. In the following, we extend this restrictive point of view to the more realistic case where both players exhibit endogenous risk-averse behavior.

3. Potential Taxes and Recompenses

Optimal contract schemes when the relationship between a regulator (e.g., the authority) and a private agent (e.g., the polluter) is subject to asymmetric information are now widely studied in the literature (Laffont and Tirole 1993; Salanie 1997; Laffont and Martimort 2002; Laffont 2003, amongst others).

Private information to the polluter may be used for strategic purposes in the relationship with the authority. Most of the time, it is assumed that the private-information parameter is observed /revealed to the polluter when the contract is signed. However, in a number of practical situations, it may be the case that the polluter does not observe the private-parameter before engaging in a contract-based relationship with the authority. It comes to consider that the private information does entail uncertainty.

Simulating the optimal policy of the polluter, the authority can conclude if the equilibrium obtained corresponds with his regulation policy. If not, the space of the polluter's strategies can be restricted through some means of pressure (Krawczyk 1985, 1995).

In the case where the measures are not in concordance with the expectations of the authority, it can signify that there exist dynamic externalities, that is, some situations in which the most prefered outcome is unilateral non-cooperation (the polluter is ordinary selfish maximizer). In this case, the authority must design structures (or incentive schemes) which harmonize this tendancy. These are purported to incite the polluter to choose the cooperative instead of a free rider solution.

Taxes are decisions of the authority which generally depend on the features of the environment. These are employed to punish the deviating polluter in order to achieve fixed environmental objectives. Using taxes, the authority imposes constraints on the polluter's behavior.

Because the inputs are not public information (and hence verifiable in court), the authority does not tax directly the polluter's actions. These are not observable, and thus non contractible by the authority. Only the outputs (poor performances) are taxed.

The output is considered to be of strategic importance, not being allowed to exceed a specific magnitude. The tax level for the period t depends on the observable quantity $\parallel y_t - y_t^a \parallel$. One can imagine two distinct scenarios:

i) if $y_t > y_t^a$ such that $\parallel y_t - y_t^a \parallel > 1$, then the polluter is taxed with P_t;

ii) if $y_t > y_t^a$ such that $\parallel y_t - y_t^a \parallel \leq c_t < 1$ (with $c_t \geq l_t$, a strategic parameter fixed by the authority), then the polluter is taxed with $c_t P_t$.

Note that it may exist discontinuities in the taxation system, in the sense that for distinct periods of the planning horizon one can have different levels of taxation.

The authority can influence the choice of the polluter's action by conditioning his utility on the outcome, and hence offering a recompense which depends on the outcome measures that have been observed.

The decision cost increases with the incentive levels. The recompense mechanism is given by the following decision rule:

i) if $y_t < y_t^a$ such that $\parallel y_t - y_t^a \parallel > 1$, then the polluter is granted with S_t;

ii) if $y_t \leq y_t^a$ such that $\parallel y_t - y_t^a \parallel < c_t' < 1$ (with $c_t' > c_t$, a strategic parameter fixed by the authority), then the polluter is granted with $c_t' S_t$.

It may also exist discontinuities in the recompense system, in the sense that for distinct periods of the planning horizon, one can have different levels of recompense.

It is important for the authority to fix higher values for taxes than for recompenses (i.e., $P_t > S_t$ for $t = 1, ..., T_2$). The strategic parameters c_t and c_t' are selected such that $c_t P_t < c_t' S_t$ for each period t. The polluter has thus the possibility to fully benefit from recompenses.

We mention here three important strategic objectives of the environmental regulation:

i) to protect the environmental quality (it implies a cost);

ii) efficiency (maximizing net benefits);

iii) cost-effectiveness (i.e., less costly method for achieving the goal, and a cost-minimizing contract).

These criteria are by no means the only possible criteria for judging environmental policies. They can be wiewed as strategic priorities (or asymmetric rationing criteria) in preserving the environment and optimizing the authority's cost function. Note that it will always exist uncertainty over the future potential costs (specific to environmental damages and their reduction) and benefits of the policy adopted.

When modelling the polluter's risk behavior, both potential taxes and recompenses are taken into account. Taxes increases the polluter's risk aversion, while recompenses have a compensation effect.

Denote by $G_{[t-\overline{k},\, t-1]} \overset{def.}{=} \sum_{j=t-(\overline{k}-1)}^{t} (p_{j-1}\mathbf{P}_{j-1} + s_{j-1}\mathbf{S}_{j-1})$, with $\mathbf{P}_{j-1} = P_{j-1}$ (or $c_{j-1}P_{j-1}$) and $\mathbf{S}_{j-1} = S_{j-1}$ (or $c'_{j-1}S_{j-1}$), the cumulative taxes and recompenses during the period $[t-\overline{k}, t-1]$, where $\overline{k}$ is an optimal backward lag parameter correlated with the polluter's risk-averse type (see Section 4), and p_{j-1}, s_{j-1} ($j = t-(\overline{k}-1), ..., t$) are strategic weights reflecting the importance the polluter places on taxes and recompenses taken individually, and respectively together.

By definition:

$$0 \leq p_{t-\overline{k}} \leq ... \leq p_{t-1} \leq 1;\ 0 \leq s_{t-\overline{k}} \leq ... \leq s_{t-1} \leq 1$$

$$p_{t-\overline{k}}\mathbf{P}_{t-\overline{k}} \geq s_{t-\overline{k}}\mathbf{S}_{t-\overline{k}}, ..., p_{t-1}\mathbf{P}_{t-1} \geq s_{t-1}\mathbf{S}_{t-1}$$

$$p_{t-\overline{k}}\mathbf{P}_{t-\overline{k}} + s_{t-\overline{k}}\mathbf{S}_{t-\overline{k}} \leq ... \leq p_{t-1}\mathbf{P}_{t-1} + s_{t-1}\mathbf{S}_{t-1}$$

The absence of cooperation can result in a low performance of the system. In this case, the authority can choose draconian punishment strategies (Abreu 1988). These are proposed to punish the polluter to the maximum extent possible.

The contract is reinforced by punishments. The authority learns to optimally punish deviating behavior and the potential deviator learns that he will be punished.

The presence of draconian strategies in the perturbations can restrict the equilibrium payoffs. Non-draconian strategies do not generally induce cooperation (Evans and Thomas 2001).

4. Risk Averse Behavior

Although there is an extensive literature on incentive control of hierarchical bilevel planning models in environmental science (Filar and Carraro 1995; Tidball and Zaccour 2005; Breton et al. 2008; Dinar et al. 2008, among many others), there is no theoretical approach for analyzing, comparing, evaluating and predicting the degree of risk-aversion of asymmetric players in the context of dynamic strategic interactions stated as a Stackelberg game.

Games with a hierarchical decision-making structure are known as Stackelberg games and the solution concept used for this type of games is the Stackelberg equilibrium. Note that non-cooperative games with an additional structure of hierarchical decision-making were first studied by Von H. Stackelberg (1934).

The aim of this section is to introduce the key concept of time-varying endogenous risk premium in dynamic stochastic environments. Constant absolute risk-aversion is not a tenable assumption for a majority of environmental models. CARA utility functions cannot capture the full behavior of the polluter towards risk.

Polluter's risk aversion is essential in assessing correctly the choice of inputs and their impact on the environment. Ommiting attitude towards risk may result in incorrect interpretation and prediction of environment responses. Three distinct definitions of the polluter's risk aversion index are proposed:

Definition 1. Using t to denote time, the polluter's risk aversion index evolves according to:

$$\varphi_{t,1}^{(2)} \overset{def.}{=} \frac{\| y_{t-1} - y_{t-1}^{g} \|^2\, \overline{L}_{t-1} + ... + \ \| y_{t-\overline{k}} - y_{t-\overline{k}}^{g} \|^2\, \overline{L}_{t-\overline{k}}}{\sqrt{(\| y_{t-1} - y_{t-1}^{g} \|^2 + ... + \ \| y_{t-\overline{k}} - y_{t-\overline{k}}^{g} \|^2)^2 + \overline{l}}},\ t = 1, ..., T_2$$

where:

i) $-1 < \overline{L}_{t-1} \leq \overline{L}_{t-2} \leq ... \leq \overline{L}_{t-\overline{k}} < 0$ are strategic weights attached to the system deviations with respect to the reference path $\{y^g_{t-1}, ..., y^g_{t-\overline{k}}\}$.

ii) $\overline{k} \geq 1$ is the optimal number of backward periods taken into account for estimating the risk-aversion index.

iii) $\overline{l} \geq 1$ is a fixed integer which characterizes the polluter's type (more or less risk-averse by nature).

In this context, only a limited history of the process (the most informative for the polluter) is taken into consideration. The strategic weights are supposed to take smaller values in the context of a principal-agent relationship compared to the case of a single-player game.

Denote by $\{y_0, ..., y_{1-\overline{k}}, ...\}$ and $\{y^g_0, ..., y^g_{1-\overline{k}}, ...\}$ the history of the process and, respectively, the output targets fixed by the polluter during the first playing period $[-T_1, 0]$.

The set of parameters $\{\overline{L}_0, ..., \overline{L}_{1-\overline{k}}, ...\}$ represents strategic weights attached to the system deviations with respect to the polluter's optimal path for the period $[-T_1, 0]$.

In the case where the polluter has the interest to use a progressive history of the process, the following definition of the risk-aversion index is considered:

Definition 2. Using t to denote time, the polluter's risk aversion index evolves according to the following relationship:

$$\varphi^{(2)}_{t,2} \overset{def.}{=} \frac{\| y_{t-1} - y^g_{t-1} \|^2 \, \overline{L}_{t-1} + ... + \ \| y_0 - y^g_0 \|^2 \, \overline{L}_0}{\sqrt{(\| y_{t-1} - y^g_{t-1} \|^2 + ... + \ \| y_0 - y^g_0 \|^2)^2 + \overline{l}}}, \ t = 1, ..., T_2$$

where

$$-1 < \overline{L}_{t-1} \leq \overline{L}_{t-2} \leq ... \leq \overline{L}_0 < 0$$

are strategic weights attached to the system deviations with respect to the reference level $\{y^g_{t-1}, ..., y^g_0\}$.

This may be the case where the polluter needs more information in improving the risk assessment process. For further details, see Protopopescu (2007).

Let us now consider the case where the polluter's targets are fixed by the authority. In this particular context, the polluter's risk aversion index is given by the following definition:

Definition 3. Using t to denote time, the polluter's risk aversion index evolves according to:

$$\varphi^{(2)}_{t,3} \overset{def.}{=} \frac{\| y_{t-1} - y^l_{t-1} \|^2 \, \widetilde{L}_{t-1} + ... + \ \| y_{t-\overline{k}} - y^l_{t-\overline{k}} \|^2 \, \widetilde{L}_{t-\overline{k}}}{\sqrt{(\| y_{t-1} - y^l_{t-1} \|^2 + ... + \ \| y_{t-\overline{k}} - y^l_{t-\overline{k}} \|^2)^2 + \overline{l}}}, \ t = 1, ..., T_2$$

where

$$-1 < \widetilde{L}_{t-1} \leq \widetilde{L}_{t-2} \leq ... \leq \widetilde{L}_{t-\overline{k}} < 0, \ \widetilde{L}_{t-j} \leq \overline{L}_{t-j} \ \forall \ j = 1, ..., \overline{k}$$

are strategic weights attached to the system deviations with respect to the authority's fixed thresholds $\{y^l_{t-1}, ..., y^l_{t-\overline{k}}\}$.

Due to random shocks which are beyond the polluter's control, at least one amongst the deviations $\| y_{t-j} - y^g_{t-j} \|$and $\| y_{t-j} - y^l_{t-j} \|$ ($j = 1, ..., \overline{k}$ or $j = 1, ..., t$) is inherently strictly positive. The risk-aversion index is thus strictly negative by construction. In other words, the polluter can be considered risk-averse by nature.

Depending on the definition employed, the polluter's risk behavior and the authority's transfers may vary differently with the system fluctuations. We exemplify this possibility in the case $\overline{k} = 1$. We have:

$$\varphi_{t,1}^{(2)} \stackrel{def.}{=} \frac{\| y_{t-1} - y_{t-1}^{g} \|^2 \, \overline{L}_{t-1}}{\sqrt{\| y_{t-1} - y_{t-1}^{g} \|^4 + \overline{l}}}, \ t = 1, ..., T_2$$

$$\varphi_{t,3}^{(2)} \stackrel{def.}{=} \frac{\| y_{t-1} - y_{t-1}^{l} \|^2 \, \widetilde{L}_{t-1}}{\sqrt{\| y_{t-1} - y_{t-1}^{l} \|^4 + \overline{l}}}, \ t = 1, ..., T_2$$

One can distinguish several scenarios:

i) If $y_{t-1} < y_{t-1}^{g} < y_{t-1}^{a}$, then $\| y_{t-1} - y_{t-1}^{g} \| < \| y_{t-1} - y_{t-1}^{a} \|$, and hence $\varphi_{t,1}^{(2)} > \varphi_{t,3}^{(2)}$ if $\widetilde{L}_{t-1} = \overline{L}_{t-1}$.

In this case, the polluter is recompensed at time $t-1$ and will be less risk-averse at time t when employing the **Definition 1**. However, if $\widetilde{L}_{t-1} < \overline{L}_{t-1}$, one cannot decide about which of two definitions is the best solution in terms of risk allocation strategy.

ii) If $y_{t-1}^{g} < y_{t-1}^{a} < y_{t-1}$, then $\| y_{t-1} - y_{t-1}^{g} \| > \| y_{t-1} - y_{t-1}^{a} \|$, and so $\varphi_{t,1}^{(2)} < \varphi_{t,3}^{(2)}$.

In other words, the polluter is taxed at time $t-1$ and will be more risk-averse at time t when employing the **Definition 1**.

iii) If $y_{t-1}^{g} < y_{t-1} < y_{t-1}^{a}$, then either $\varphi_{t,1}^{(2)} > \varphi_{t,3}^{(2)}$ or $\varphi_{t,1}^{(2)} < \varphi_{t,3}^{(2)}$, this depending on the difference in magnitude between the norm-deviations $\| y_{t-1} - y_{t-1}^{g} \|$ and $\| y_{t-1} - y_{t-1}^{a} \|$. At time $t-1$, the polluter is recompensed.

In the literature on principal-agent issues, two distinct scenarios are considered when analyzing the players' strategic relationship:

i) the case of a risk-neutral principal;

ii) the case of a principal with a fixed risk premium.

In both cases, the assumptions adopted are very restrictive, being convenient only for modelisation purposes. The fact that the principal is the dominant player in the game does not justify to assume that he is less risk-averse compared to the private agent. Taking into account the effects of adverse selection and moral hazard, inherent in any principal-agent relationship, there may exist situations where the principal is more risk-averse compared to the private agent. One can imagine the extreme case where the agent is risk-neutral and the principal is risk-averse.

In what follows, two distinct definitions are proposed for the principal's /authority's risk aversion index:

i) when the most informative actions of the agent /polluter are taken into account;

ii) when a progressive history of the agent's /polluter's response-actions is considered in the analysis.

Definition 4. Using t to denote time, the authority's risk aversion index evolves according to:

$$\varphi_{t,1}^{(1)} \stackrel{def.}{=} \frac{\| x_{t-1}^{(1)} - x_{t-1}^{(2)} \|^2 \, \overline{\overline{L}}_{t-1} + ... + \| x_{t-\overline{\overline{k}}}^{(1)} - x_{t-\overline{\overline{k}}}^{(2)} \|^2 \, \overline{\overline{L}}_{t-\overline{\overline{k}}}}{\sqrt{(\| x_{t-1}^{(1)} - x_{t-1}^{(2)} \|^2 + ... + \| x_{t-\overline{\overline{k}}}^{(1)} - x_{t-\overline{\overline{k}}}^{(2)} \|^2)^2 + \overline{\overline{l}}}}, \ t = 1, ..., T_2$$

where:

i) $-1 < \overline{\overline{L}}_{t-1} \leq \overline{\overline{L}}_{t-2} \leq ... \leq \overline{\overline{L}}_{t-\overline{\overline{k}}} < 0$ are strategic weights attached to the distance between the players' actions on the time horizon $[t-1, t-\overline{\overline{k}}]$.

ii) $\overline{\overline{k}} \geq 1$ is an optimal number of backward periods (the most informative for the authority).

iii) $\overline{\overline{l}} \geq 1$ is a fixed integer which characterizes the authority's type (more or less risk-averse by nature).

Denote by $\{x_0^{(i)}, ..., x_{1-\overline{\overline{k}}}^{(i)}, ...\}$, $i = 1, 2$, the history of the players' actions during the first playing period $[-T_1, 0]$.

The set of parameters $\{\overline{\overline{L}}_0, ..., \overline{\overline{L}}_{1-\overline{\overline{k}}}, ...\}$ represents strategic weights attached to the distance between the players' actions for the period $[-T_1, 0]$.

Definition 5. Using t to denote time, the authority's risk-aversion index evolves according to:

$$\varphi_{t,2}^{(1)} \stackrel{def.}{=} \frac{\| x_{t-1}^{(1)} - x_{t-1}^{(2)} \|^2 \, \overline{\overline{L}}_{t-1} + ... + \| x_0^{(1)} - x_0^{(2)} \|^2 \, \overline{\overline{L}}_0}{\sqrt{(\| x_{t-1}^{(1)} - x_{t-1}^{(2)} \|^2 + ... + \| x_0^{(1)} - x_0^{(2)} \|^2)^2 + \overline{\overline{l}}}}, \quad t = 1, ..., T_2$$

where $-1 < \overline{\overline{L}}_{t-1} \leq \overline{\overline{L}}_{t-2} \leq ... \leq \overline{\overline{L}}_0 < 0$ are strategic weights attached to the distance between the players' actions on the time horizon $[t-1, 0]$.

The smaller the scalar weight, the higher the importance given by the authority to the polluter's deviation from his local objective.

One can imagine several scenarios when comparing the degree of risk-aversion exhibited by the players:

i) when both authority and polluter have similar coefficients of absolute risk-aversion during the period of contract (i.e., $\mid \varphi_{t,j}^{(1)} - \varphi_{t,j'}^{(2)} \mid$ is small, with $j \in \{1, 2\}$, $j' \in \{1, 2, 3\}$ and $t \in \{1, ..., T_2\}$);

ii) when both authority and polluter have the same coefficient of absolute risk-aversion at given periods of time (i.e., $\varphi_{t,j}^{(1)} = \varphi_{t,j'}^{(2)}$ for $j \in \{1, 2\}$, $j' \in \{1, 2, 3\}$, and $t \in \{1, ..., T_2\}$ taking fixed values);

iii) when the players have very different degrees of risk-aversion during the period of contract (i.e., $\mid \varphi_{t,j}^{(1)} - \varphi_{t,j'}^{(2)} \mid$ is high, with $j \in \{1, 2\}$, $j' \in \{1, 2, 3\}$ and $t \in \{1, ..., T_2\}$).

We point out the local character of the players' risk aversion. This is defined for a neighborhood of their fixed targets. There thus exist neighborhood effects of the system dynamics on the players' risk behavior.

The reality shows that the relationship between players' reaction to perceived states of nature and their attitude to risk is complex. In general, risk-aversion makes this reaction stronger than risk-neutrality.

For the agent /polluter, this is a consequence of the importance he places on the system states, while for the principal /authority this depends on the importance he places on the agent's /polluter's actions.

When the players' optimal actions are different in magnitude, the principal /authority is inevitably risk-averse.

The combination of the asymmetric information with the risk-aversion makes the analysis difficult. The classical theory of principal-agent predicts that the risk is shared inefficiently between a risk-neutral principal and a risk-averse agent.

When both players are risk-averse by nature, the objective is to optimally share the risk during the period of contract. This requires an efficient incentive mechanism and a powerful control strategy to be implemented in order that the agent does not exceed the thresholds set by the principal.

5. Adaptive Risk Management

In this section, we deal with important qualitative aspects of the players' risk behavior. More exactly, it is analyzed the polluter's risk aversion according to the deviation of the system from his fixed targets and, respectively, the authority's risk aversion according to the polluter's response-actions. It is also studied the effect of the system fluctuations on the authority's incentive policy. In this sense, it is examined the way the variation in outputs influences the allocation of taxes and recompenses.

Proposition 1. Small system deviations during the period $[t-\overline{k}, t-1]$ will annul the effect of potential taxes and decrease to zero the polluter's risk aversion at time t.

Proof. In the case where $\| y_{t-j} - y^g_{t-j} \| \to 0 \; \forall \; j = 1, ..., \overline{k}, \; \forall \; t = 1, ..., T_2$, we have $\varphi^{(2)}_{t,1} \to 0$, and hence an almost risk-neutral behavior of the polluter at time t.

One can write the inequality: $\| y_{t-j} - y^l_{t-j} \| < \| y_{t-j} - y^g_{t-j} \| < c_{t-j}$. The polluter will be thus recompensed during the period $[t-\overline{k}, t-1]$.

We have:

$$G_{[t-\overline{k},\, t-1]} = \sum_{j=t-(\overline{k}-1)}^{t} s_{j-1}\mathbf{S}_{j-1}, \text{with } \mathbf{S}_{j-1} = c'_{j-1} S_{j-1} \; (j = t-\overline{k}+1, ..., t).$$

Proposition 2. High system deviations during the period $[t-\overline{k}, t-1]$ will amplify the effect of potential taxes and increase the polluter's risk aversion at time t.

Proof. For ease of exposition, we consider the case where the system deviations $\| y_{t-j} - y^g_{t-j} \|$ are high and comparable in magnitude ($\| y_{t-j'} - y^g_{t-j'} \| \;/\; \| y_{t-j''} - y^g_{t-j''} \| \approx 1$ for $j', j'' = 1, ..., \overline{k}$; $j' \neq j''$). This type of behavior will annul the effect of potential recompenses and amplify the effect of potential taxes. In this case, we have:

$$G_{[t-\overline{k},\, t-1]} = \sum_{j=\, t-(\overline{k}-1)}^{t} p_{j-1}\mathbf{P}_{j-1}, \text{with } \mathbf{P}_{j-1} = P_{j-1} \; (j = t-\overline{k}+1, ..., t)$$

One can write:

$$\varphi^{(2)}_{t,1} \to \frac{\overline{L}_{t-1} + ... + \overline{L}_{t-\overline{k}}}{\overline{k}}$$

In the case of a principal-agent relationship, the strategic weights $\overline{L}_{t-j}$ $(j = 1, ..., \overline{k})$ are assumed to be smaller compared to those fixed in the case of a single-player game. The value of the ratio $\frac{\overline{L}_{t-1}+...+\overline{L}_{t-\overline{k}}}{\overline{k}}$ will be thus smaller in magnitude.

The agent /polluter is hence characterized by a higher risk-aversion in t. In particular, it may be possible that he becomes excessively risk-averse for the period t.

Consequence 1. In the case where $\| y_{t-j} - y^g_{t-j} \| \to 1 \; \forall \; j = 1, ..., \overline{k}$, one obtains:

$$\varphi^{(2)}_{t,1} \to \frac{\overline{L}_{t-1} + ... + \overline{L}_{t-\overline{k}}}{\sqrt{\overline{k}^2 + \overline{l}}} > \frac{\overline{L}_{t-1} + ... + \overline{L}_{t-\overline{k}}}{\overline{k}}$$

In other words, high system deviations with respect to the polluter's fixed targets signify that $\parallel y_{t-j} - y^g_{t-j} \parallel \gg 1$ for $t = 1, ..., T_2$ and $j = 1, ..., \overline{k}$.

Consequence 2. A smaller number of periods $\overline{k}$ implies a higher risk-aversion level $\varphi^{(2)}_{t,1}$ at time t.

Proof. Suppose that $\overline{k}_1 < \overline{k}_2$. One can write the inequality:

$$\frac{\overline{L}_{t-1} + ... + \overline{L}_{t-\overline{k}_2}}{\overline{k}_2} > \frac{\overline{L}_{t-1} + ... + \overline{L}_{t-\overline{k}_1}}{\overline{k}_1}$$

In other words, the smaller the number of backward periods, the higher the effect of taxes on the polluter's risk behavior when all system deviations are high and comparable in magnitude.

Proposition 3. The polluter's degree of risk-aversion will be less and less raised when he controls better and better the system evolution.

Proof. One can write the sequence of inequalities:

$$\frac{\overline{L}_{t-1} + ... + \overline{L}_{t-\overline{k}}}{\overline{k}} < \frac{\overline{L}_{t-2} + ... + \overline{L}_{t-\overline{k}}}{\overline{k} - 1} < ... < \frac{\overline{L}_{t-(\overline{k}-1)} + \overline{L}_{t-\overline{k}}}{2} < \frac{\overline{L}_{t-\overline{k}}}{1}$$

In other words, the polluter will become more and more confident over time when he manages better and better the system evolution. The conclusion follows.

For this type of scenario, the potential recompenses have a dominant effect on the potential taxes.

Proposition 4. The polluter becomes more and more risk-averse over time when he controls more and more hardly the system evolution.

Proof. One can write the sequence of inequalities:

$$\frac{\overline{L}_{t-1}}{1} < \frac{\overline{L}_{t-1} + \overline{L}_{t-2}}{2} < ... < \frac{\overline{L}_{t-1} + ... + \overline{L}_{t-(\overline{k}-1)}}{\overline{k} - 1} < \frac{\overline{L}_{t-1} + ... + \overline{L}_{t-\overline{k}}}{\overline{k}}$$

Each above ratio corresponds (in this order) to the polluter's risk-aversion index at time t in the case where he controls more and more hardly the system evolution. The conclusion follows.

For this type of scenario, the potential taxes have a dominant effect on the potential recompenses.

Proposition 5. If the last system deviation is high and all preceding deviations are close to zero in magnitude, then the polluter's risk aversion will be higher compared to the case when all system deviations are high and comparable.

Proof. Suppose that $\parallel y_{t-1} - y^g_{t-1} \parallel$ is high and $\parallel y_{t-j} - y^g_{t-j} \parallel \to 0 \; \forall \; j = 2, ..., \overline{k}$. It follows that:

$$\varphi^{(2)}_{t,1} \to \overline{L}_{t-1} < \frac{\overline{L}_{t-1} + ... + \overline{L}_{t-\overline{k}}}{\overline{k}}$$

A single taxation period can therefore generate a higher risk-aversion compared to the case of multiple taxation periods. This is an astonishing result, far from intuitive.

Proposition 6. The moment of time when a large deviation of the system is obtained in the past can change the current risk-aversion degree of the polluter.

Proof. Consider the case where the system deviation from the fixed target at time $t-\overline{k}$ is large but all other deviations are close to zero in magnitude. This is what we call accidental taxation at time $t-\overline{k}$. For this type of scenario, one obtains:

$$\varphi_{t,1}^{(2)} \rightarrow \overline{L}_{t-\overline{k}} > \ldots > \overline{L}_{t-1}$$

In the case where $\overline{k} < \overline{k}'$, we have $\overline{L}_{t-\overline{k}} < \overline{L}_{t-\overline{k}'}$, and hence a higher risk-aversion of the polluter for the same period of the contract. The more the tax is distant in the past, the more its effect on the polluter's risk behavior is weaker.

Consider now the case where the system deviation from the fixed target at time $t-\overline{k}$ is small but all other deviations are large and comparable in magnitude. This is what we call accidental recompense at time $t-\overline{k}$. One can write the inequality:

$$\frac{\overline{L}_{t-1} + \ldots + \overline{L}_{t-(\overline{k}-1)}}{\overline{k}-1} < \overline{L}_{t-\overline{k}}$$

In other words, the polluter's risk aversion at time t will be higher for an accidental recompense compared to the case of an accidental taxation at the same period $t-\overline{k}$.

Proposition 7. The more the players' optimal actions are distant during the period $[t-\overline{\overline{k}}, t-1]$, the more the authority is risk-averse at time t. In the case where the players' optimal actions are comparable in magnitude, the authority becomes almost risk-neutral.

Proof. If $x_{t-j}^{(1)}$ and $x_{t-j}^{(2)}$ $(j = 1, \ldots, \overline{\overline{k}})$ are comparable in magnitude, then one obtains $\varphi_{t,1}^{(1)} \rightarrow 0$. In other words, the authority becomes almost risk-neutral at time t. In contrast, when the players' optimal actions are distant, the authority's risk aversion is large.

In the real world, the incentive mechanism implemented by the principal /authority does not always ensure an optimal risk-sharing during the entire contract period. It may exist situations where the principal /authority is more risk-averse compared to the private agent /polluter. In particular, it may be possible for the principal /authority to become excessively risk-averse over time.

6. Players with Distinct Risk-Aversion Profile

Experimental evidence shows that strategic players overweight extreme events. The ability to assess future risks associated with extreme events is increasingly important in many strategic interactions modelled as a discrete-time dynamic Stackelberg game.

Let $\varphi_{\min}^{(2)}$ be an optimal risk-aversion threshold fixed by the polluter before starting the contract and for the entire period $[1, T_2]$. It characterizes the polluter's risk-averse type.

The objective is not to exceed this limit threshold. Otherwise, the polluter becomes excessively risk-averse for the current control period.

Note that $\varphi_{t,i}^{(2)}$ $(i = 1, 2, 3)$ characterize the polluter's local risk-aversion (at time t), while $\varphi_{\min}^{(2)}$ characterizes his global risk-aversion (over the whole period $[1, T_2]$). For further details, see Protopopescu (2007).

Proposition 8. In any dynamic principal-agent relationship there exist two types of risk-averse agents.

Proof. One can write:

$$\mid \varphi_{t,1}^{(2)} \mid \; < \frac{\| y_{t-1} - y_{t-1}^{g} \|^2 + \ldots + \| y_{t-\overline{k}} - y_{t-\overline{k}}^{g} \|^2}{\sqrt{(\| y_{t-1} - y_{t-1}^{g} \|^2 + \ldots + \| y_{t-\overline{k}} - y_{t-\overline{k}}^{g} \|^2)^2 + \overline{l}}}$$

The agent's objective is to obtain $\parallel y_{t-j} - y^{g}_{t-j} \parallel < 1 \; \forall \; j = 1, ..., \overline{k}$. It implies:

$$-\frac{\overline{k}}{\sqrt{\overline{k}^2 + \bar{l}}} < \varphi^{(2)}_{t,1} < 0$$

One can imagine two possible scenarios:

$$\varphi^{(2)}_{\min,\, less} < -\frac{\overline{k}}{\sqrt{\overline{k}^2 + \bar{l}}} \text{ (less risk-averse agent)}$$

and respectively,

$$-\frac{\overline{k}}{\sqrt{\overline{k}^2 + \bar{l}}} < \varphi^{(2)}_{\min,\, more} \text{ (more risk-averse agent)}$$

This type of analysis allows us to distinguish between common types of polluters. A total separation of distinct types of polluters is also possible. We deal with a continuum of polluter's types.

Both risk-aversion thresholds $\varphi^{(2)}_{\min,\, less}$ and $\varphi^{(2)}_{\min,\, more}$ are not exceeded during the period of the game if and only if the agent /polluter succeeds in controlling the fluctuating system.

By analogy, nothing impedes to consider an optimal risk-aversion threshold for the principal /authority, say $\varphi^{(1)}_{\min}$. This characterizes his risk behavior profile and must not be exceeded during the entire planning horizon. The fact that the principal /authority imposes the rules of the game does not necessarily mean that he will be less risk-averse (before and after starting the game) compared to the private agent /polluter.

Proposition 9. In a dynamic principal-agent relationship there generally exist two types of risk-averse principals.

Proof. The goal of the principal's incentive policy is to induce an optimal behavior at equilibrium, and thus to eliminate sub-optimal outcomes. In other words, the players' actions must satisfy the following inequality:

$$\parallel x^{(1)}_{t-j} - x^{(2)}_{t-j} \parallel < 1 \; \forall \; j = 1, ..., \overline{\overline{k}}, \; \forall \; t = 1, ..., T_2$$

We thus obtain an inferior limit for the principal's risk-aversion index:

$$-\frac{T_2 - 1}{\sqrt{(T_2 - 1)^2 + \bar{\bar{l}}}} < \varphi^{(1)}_{t,1} < 0$$

Two distinct scenarios are possible:

$$\varphi^{(1)}_{\min,\, less.} < -\frac{T_2 - 1}{\sqrt{(T_2 - 1)^2 + \bar{\bar{l}}}} \text{ (less risk-averse principal)}$$

and respectively,

$$-\frac{T_2 - 1}{\sqrt{(T_2 - 1)^2 + \bar{\bar{l}}}} < \varphi^{(1)}_{\min,\, more} \text{ (more risk-averse principal)}$$

This type of analysis allows us to obtain a differentiation of common /distinct types of risk-averse authorities.

Note that an excessive risk-averse behavior of the polluter does not necessarily induce an imitative behavior of the authority for the same period of the contract.

We have the inequality:

$$-1 < -\frac{T_2 - 1}{\sqrt{(T_2-1)^2 + \bar{\bar{l}}}} < -\frac{\bar{k}}{\sqrt{\bar{k}^2 + \bar{l}}} < 0 \text{ for } \bar{l} \geq \bar{\bar{l}}$$

In this case, the authority can be considered less risk-averse before starting the contract compared to the polluter. It may also be possible to have the contrary situation, when $\bar{l} < \bar{\bar{l}}$.

It is very important to distinguish between “nature” and “type”. Both players are considered risk-averse by nature. By “type”, one means more or less risk-averse players. The system evolution refines the polluter’s type, while the polluter’s response-actions refines the authority’s type.

The length of the game can influence the attitude to risk of the players before starting the game. Because a closed-loop strategy is employed for a short planning horizon, this implies a non-negligible risk-aversion level for both players at time $t = 0$. The shorter the planning horizon, the higher the players’ degree of risk-aversion before starting the game. In other words, increased uncertainty leads to a shorter duration of the game. The contract length is thus decreasing in uncertainty.

This result is in accordance with the classic empirical analysis of contract duration in labor markets (Wroman 1989; Murphy 1992, amongst others).

7. Stackelberg Strategic Equilibrium Sensitive to Endogenous Risk-Aversion

This type of equilibrium is appropiate and quite attractive for situations where one of the players is in possession of a sufficient quantity of information for being leader (or dominant player), but many times possibly a poor enough information about his opponent.

Even if the parties are initially symmetrically informed, the follower may later endogenously acquire some private information. The presence of asymmetry is not only in the information structure but also in the way the follower affects the decision process.

This brings the impact of the follower’s behavior as an issue of major importance in the construction of the optimal policy. For this type of situations, the follower adopts a passive strategy equilibrium, waiting for the leader’s optimal policy to reveal.

The interaction of the leader’s optimal policy with the dynamic learning process of the follower is a reality which must be fully recognized. At each step of the game, the follower can choose between cooperative and non-cooperative behavior.

Stackelberg games play an extremely important role in such fields as economics, management, politics and behavioral sciences. This game-theoretical concept is widely used in various domains such as market sharing, investment allocation, environmental policy, monetary problems, and so on.

This kind of game may be at the leader’s best payoff or at the follower’s best payoff. The dominant player has the ability to announce his strategy in advance. By doing this, he will send a signal to the follower, and thus will lose a private strategic information, but the goal is to dictate the game. A question arises: How this strategic signal will influence the follower’s behavior?

Both players derive individual utilities from their optimal actions. The follower can take a decision which affects his utility level and also has a direct impact on the leader's planning objective. In general, the leader's preferences tend towards non-stationarity due to possible changes in the follower's dynamic behavior.

We consider the case where both players take into account only a limited number of backward periods when estimating their risk-aversion parameters.

For each $t = 1, ..., T_2$, the players' risk preferences are described, respectively, by the strictly decreasing concave utility functions:

$$U_t^{(1)}(W_{[1,t]}^{(1)}, \varphi_{t,1}^{(1)}) \overset{def.}{=} \frac{2}{\varphi_{t,1}^{(1)}}[\exp(-\frac{\varphi_{t,1}^{(1)}}{2}W_{[1,t]}^{(1)}) - 1] \text{ (for the leader)}$$
$$U_t^{(2)}(W_{[1,t]}^{(2)}, \varphi_{t,1}^{(2)}) \overset{def.}{=} \frac{2}{\varphi_{t,1}^{(2)}}[\exp(-\frac{\varphi_{t,1}^{(2)}}{2}W_{[1,t]}^{(2)}) - 1] \text{ (for the follower)}$$

The exponential utility functions are the most widely utilized risk-sensitive utility functions because they can model a spectrum of risk attitudes for the players (Corner and Corner 1995).

By construction, the local objective function $U_t^{(i)}$ is strictly convex with respect to the control variable handled by the player i.

In order to avoid drastic changes from one period to the next, some upper and lower bounds on the players' strategies are imposed (Sandblom and Banasik 1985).

Define by:

$$\mathbf{R}^{q_2} \supset \Lambda_t^{(2)} \overset{def.}{=} \{x_t^{(2)} \mid \underbrace{0 < \nu_t^{(2)} \le x_t^{(2)} \le \xi_t^{(2)}}_{\text{(amplitude bounds)}} \text{ and } \underbrace{\lambda_t^{(2)} \le x_t^{(2)} - x_{t-1}^{(2)} \le \mu_t^{(2)}}_{\text{(change bounds)}}\}$$

the follower's feasible strategies space at time t.

Denote also by:

$$\mathbf{R}^{q_1} \supset \Lambda_t^{(1)} \overset{def.}{=} \{x_t^{(1)} \mid \underbrace{0 < \nu_t^{(1)} \le x_t^{(1)} \le \xi_t^{(1)}}_{\text{(amplitude bounds)}} \text{ and } \underbrace{\lambda_t^{(1)} \le x_t^{(1)} - x_{t-1}^{(1)} \le \mu_t^{(1)}}_{\text{(change bounds)}}\}$$

the admissible strategies space of the leader for the period t. The follower's reaction set in t for a given incentive compatible strategy $x_t^{(1)}$ is denoted by $R_2(x_t^{(1)})$.

Let $E_{t-1}(\cdot) \overset{def.}{=} E(\cdot \mid I_{t-1})$ be the operator of conditional expectation, where I_{t-1} represents the information available up to time $t-1$.

Definition 5. The set of strategies $(s_t^{g(2)} \mid s_t^{g(1)}, s_t^{g(1)})$ verifying the inequality:

$$E_{t-1}U_t^{(1)}(W_{[1,t]}^{(1)}(s_t^{g(2)} \mid s_t^{g(1)}, s_t^{g(1)})) \ge E_{t-1}U_t^{(1)}(W_{[1,t]}^{(1)}(s_t^{(2)} \mid s_t^{(1)}, s_t^{(1)}))$$

$\forall s_t^{(1)} \in \Lambda_t^{(1)}$, $\forall s_t^{(2)} \in R_2(s_t^{(1)}) \cap \Lambda_t^{(2)}$ is called incentive closed-loop Stackelberg equilibrium sensitive to the players' local risk-aversion.

The equilibrium of the game is a random vector whose statistic is not known a priori because it depends on what strategy the follower will adopt.

In the case of a leader-follower interaction process, the Pareto-optimality principle (Pareto 1909) is not generally valid for the leader's strategies. The equilibrium of the game needs not to be Pareto optimal. When the optimality principle is also valid for the leader, he will choose a strategy situated on the curve of reaction of the follower.

As opposed to a Nash (1950, 1951) equilibrium, the Stackelberg equilibrium is defined uniquely in the case of a linear model. In contrast, for a nonlinear specification, there may exist non-unique solutions, and hence it will be difficult for the leader to make decisions. More exactly, when the leader employes some strategy, there exist diverse decisions for the follower. Some strategies are beneficial for the leader, while others are not. Therefore, the leader has to be careful when selecting his strategy.

Note that the dynamic Stackelberg solution is appropriate in nonzero-sum dynamic games when a hierarchy in decision-making exists. By hierarchy, one means that the players hold non-symmetrical roles in the decision-making process.

The common feature of the closed-loop solutions is that they are incentive strategies in which the leader induces the follower to be on the optimal solution path. If the follower deviates from the ex-ante specified path for some reason, then the leader can induce him to return to this after one period (Basar and Selbuz 1979; Basar and Olsder 1980).

The leader knows that the follower will maximize his utility function at time t with respect to his optimal strategy $s_t^{g(1)}$ to obtain the best response-function $s_t^{g(2)}$. This type of information will be included in his expected utility function for the next period, before maximizing with respect to his control variable. Because the leader is not sure about the strategy the follower will play, his behavior will be uncertain. However, he can compute (ex-ante) up to a random shock the effects of the follower's policy changes on his optimal rule. It is precisely in this sense that one can say that the leader's strategy depends upon his beliefs.

A substantial effort has been devoted in the literature to various incentive Stackelberg solution concepts. However, most of the incentive strategies developed include either the follower's control, which may not be realistic in practice, or time-varying delays in the state variable, which makes the stabilization process more difficult to achieve.

Denote by $s_{t,e}^{(2)}$ the agent's strategy expected by the leader at time t. By taking into account the projected rational response of the follower, the leader seeks that policy which leads to a most favorable outcome for him.

If the leader will announce a strategy $s_t^{g(1)}$, by solving the maximization programme:

$$
\begin{aligned}
s_t^{g(1)} &= \arg\max_{s_t^{(1)}\in\Lambda_t^{(1)}} E_{t-1} U_t^{(1)}(W_{[1,t]}^{(1)}(s_{t,e}^{(2)}, s_t^{(1)}) \mid s_{t,e}^{(2)}) \\
s.t.: &\begin{cases} y_t = A_t y_{t-1} + C_t^{(1)} s_t^{(1)} + C_t^{(2)} s_{t,e}^{(2)} + B_t z_t + D_t + u_t \\ y_0,\ y_t > 0,\quad t = 1, ..., T_2 \quad \text{(economic constraints)} \end{cases}
\end{aligned}
$$

then the follower will compute his optimal policy as if the environment would be modelled by:

$$
y_t = A_t y_{t-1} + C_t^{(1)} s_t^{g(1)} + C_t^{(2)} s_t^{(2)} + B_t z_t + D_t + u_t,\quad t = 1, ..., T_2
$$

and will find $s_t^{g(2)} \mid s_t^{g(1)}$ (reaction strategy or response-function) by solving the asymmetric information problem:

$$
\begin{aligned}
s_t^{g(2)} &= \arg\max_{s_t^{(2)}\in\Lambda_t^{(2)}} E_{t-1} U_t^{(2)}(W_{[1,t]}^{(2)}(s_t^{(2)}, s_t^{g(1)}) \mid s_t^{g(1)}) \\
s.t.: &\begin{cases} y_t = A_t y_{t-1} + C_t^{(1)} s_t^{g(1)} + C_t^{(2)} s_t^{(2)} + B_t z_t + D_t + u_t \\ y_0,\ y_t > 0,\quad t = 1, ..., T_2 \quad \text{(economic constraints)} \end{cases}
\end{aligned}
$$

Each player takes the reactions of the opponent at least partially into account. In this type of relationship, the follower faces a double uncertainty:

i) about the action that the leader takes; and ii) about the action taken by himself.

The unsignalled information of the follower is probabilistically dependent on the unsignalled information of the leader.

When modelling the leader-follower relationships as a dynamic game, the optimization problem is naturally interactive. If the decision rule of the leader changes, the behavior of the follower will also change. It may be possible to exist policy shifts caused by adverse selection and imperfect information.

The leader must take into account the effect of his policy rules on the follower's actions. His decisions can be interpreted as signals sent to reveal the reactions of the follower. It generally exists a discrete tâtonnement process from the part of the leader. The information revealed by the follower's actions may be relevant to the leader's future decisions. The follower can transmit involuntarily (by his actions) a part of his private information to the leader. Moreover, he creates (by his actions) informational externalities which induce new information structure and new behavior.

Outside behavior can effect the equilibrium agreements. The follower may be able to gain more in average expected utility (conditional on his information up to that date) by deviating from the game equilibrium strategy. In this case, the leader will modify his strategy in order to discourage the follower from taking advantage from this situation. The follower is free to reject the contract without sharing risk with the leader.

The leader's ability to exploit the anticipative effects of his policy depends critically on the one-period delay inherent to the discrete-time modelling of the state variable. The timing effect plays an important role in this case. The solution which is optimal for the leader at the beginning of the game is thus time-inconsistent. It ceases to be optimal for the leader in subsequent periods (Simaan and Cruz 1973; Kydland 1975, 1977; Kydland and Prescott 1977; Miller and Salmon 1985).

The leader has thus an incentive to restart the game. Such a succession of restarted leader-follower solutions would be unsustainable and hence unappealing as a solution concept. The time inconsistency may arise because the follower takes certain actions before the leader chooses the optimal path of instruments. More exactly, the follower takes his decisions based on expectations of the leader's policy while the leader takes his decisions based on given expectations of the follower. This leads to an informational advantage for the leader (Cohen and Michel 1988).

Anticipative control is the leader's ability to indirectly influence the evolution of the state vector backwards in time by accounting for the follower's current reactions to change in his current and expected future control settings. It is this effect which causes dynamic Stackelberg equilibria to be time-inconsistent (Basar and Olsder 1995).

Note that generally one cannot say how much information is actually transmitted from the informed to the uninformed party in equilibrium. There may be equilibria in which all the relevant information is transmitted as well as equilibria in which no relevant information is transmitted.

We are now in a position to derive the closed-loop Stackelberg equilibrium sensitive to endogenous risk-aversion of the players.

Proposition 10. Suppose that the matrices $\Psi^{-1}+\varphi_{t,1}^{(i)}H_{t,i}$, $K_{t,i}-\varphi_{t,1}^{(i)}H_{t,i}(\Psi^{-1}+\varphi_{t,1}^{(i)}H_{t,i})^{-1}H_{t,i}$, and $C_t^{(i)\prime}[K_{t,i}-\varphi_{t,1}^{(i)}H_{t,i}(\Psi^{-1}+\varphi_{t,1}^{(i)}H_{t,i})^{-1}H_{t,i}]C_t^{(i)}$ are inversible for each $t=1,...,T_2$ and $i=1,2$. Under the hypotheses stated in Section 2, the equilibrium of the Stackelberg game is described by the following linear equations:

$$\left\{\begin{array}{l} s_t^{g(1)}(I_{t-1}, z_t, \beta_t, K_{t,1}, d_{t,1}, y_t^a, s_{t,e}^{(2)}) = G_t^{(1)} y_{t-1} + g_t^{(1)},\ t = 1, ..., T_2 \\ s_t^{g(2)}(I_{t-1}, z_t, \beta_t, K_{t,2}, d_{t,2}, y_t^g, s_t^{g(1)}) = G_t^{(2)} y_{t-1} + g_t^{(2)},\ t = 1, ..., T_2 \end{array}\right.$$

where the matrices $(G_t^{(i)}, g_t^{(i)})$, $i = 1, 2$, are given, respectively, by:

$$G_t^{(1)} \overset{not.}{=} -(C_t^{(1)\prime} \overline{H}_t^{(1)} C_t^{(1)})^{-1} (C_t^{(1)\prime} \overline{H}_t^{(1)} A_t)$$

$$g_t^{(1)} \overset{not.}{=} -(C_t^{(1)\prime} \overline{H}_t^{(1)} C_t^{(1)})^{-1} C_t^{(1)\prime} [\overline{H}_t^{(1)} (B_t z_t + D_t + C_t^{(2)} s_{t,e}^{(2)}) - (I_p - \varphi_{t,1}^{(1)} K_{t,1} M_{t,1}^{-1} (\varphi_t^{(1)})) h_{t,1}]$$

and

$$G_t^{(2)} \overset{not.}{=} -(C_t^{(2)\prime} \overline{H}_t^{(2)} C_t^{(2)})^{-1} (C_t^{(2)\prime} \overline{H}_t^{(2)} A_t)$$

$$g_t^{(2)} \overset{not.}{=} -(C_t^{(2)\prime} \overline{H}_t^{(2)} C_t^{(2)})^{-1} C_t^{(2)\prime} [\overline{H}_t^{(2)} (B_t z_t + D_t + C_t^{(1)} s_t^{g(1)}) - (I_p - \varphi_{t,1}^{(2)} K_{t,2} M_{t,2}^{-1} (\varphi_t^{(2)})) h_{t,2}]$$

with

$$\overline{H}_t^{(1)} \overset{not.}{=} K_{t,1} - \varphi_{t,1}^{(1)} H_{t,1} M_{t,1}^{-1} (\varphi_{t,1}^{(1)}) H_{t,1}$$

$$M_{t,1}(\varphi_{t,1}^{(1)}) \overset{not.}{=} \Psi^{-1} + \varphi_{t,1}^{(1)} H_{t,1}$$

$$H_{t,1} \overset{def.}{=} K_{t,1}, \quad h_{t,1} \overset{not.}{=} K_{t,1} y_t^a - d_{t,1}$$

and

$$\overline{H}_t^{(2)} \overset{not.}{=} K_{t,2} - \varphi_{t,1}^{(2)} H_{t,2} M_{t,2}^{-1} (\varphi_{t,1}^{(2)}) H_{t,2}$$

$$M_{t,2}(\varphi_{t,1}^{(2)}) \overset{not.}{=} \Psi^{-1} + \varphi_{t,1}^{(2)} H_{t,2}$$

$$H_{t,2} \overset{def.}{=} K_{t,2}, \quad h_{t,2} \overset{not.}{=} K_{t,2} y_t^g - d_{t,2}$$

Proof. At each period t, the leader will solve the following stochastic optimization programme:

$$s_t^{g(1)} = \underset{s_t^{(1)} \in \Lambda_t^{(1)}}{\arg\max} E_{t-1} U_t^{(1)} (W_{[1,t]}^{(1)} (s_{t,e}^{(2)}, s_t^{(1)}) \mid s_{t,e}^{(2)})$$

$$s.t. : \left\{\begin{array}{l} y_t = A_t y_{t-1} + C_t^{(1)} s_t^{(1)} + C_t^{(2)} s_{t,e}^{(2)} + B_t z_t + D_t + u_t \\ y_0,\ y_t > 0, \quad t = 1, ..., T_2 \quad \text{(economic constraints)} \end{array}\right.$$

One can write:

$$s_t^{g(1)} = \underset{s_t^{(1)} \in \Lambda_t^{(1)}}{\arg\max} E_{t-1} U_t^{(1)} (W_{[1,t]}^{(1)}(y_t), \varphi_{t,1}^{(1)}) = \underset{s_t^{(1)} \in \Lambda_t^{(1)}}{\arg\max} E_{t-1} [\exp(-\frac{\varphi_{t,1}^{(1)}}{2} W_t^{(1)}(y_t))]$$

For the computation of $E_{t-1}[\exp(-\frac{\varphi_{t,1}^{(1)}}{2} W_t^{(1)}(y_t))] \overset{not.}{=} V_t^{(1)}$ (which is supposed to exist), we proceed as follows:

$$E_{t-1}[\exp(-\frac{\varphi_{t,1}^{(1)}}{2} W_t^{(1)}(y_t))] = E_{t-1}[\exp(-\frac{\varphi_{t,1}^{(1)}}{2} (\triangle^l\, y_t' K_{t,1} \triangle^l\, y_t + 2 \triangle^l\, y_t' d_{t,1}))]$$

$$= E_{t-1}[\exp(-\frac{\varphi_{t,1}^{(1)}}{2}(y_t' H_{t,1} y_t - 2y_t' h_{t,1} + f_{t,1}))]$$

where:

$$\triangle^l y_t \overset{not.}{=} y_t - y_t^a, \quad H_{t,1} \overset{def.}{=} K_{t,1}, \quad h_{t,1} \overset{not.}{=} K_{t,1} y_t^a - d_{t,1}, \quad f_{t,1} \overset{not.}{=} y_t^{a\prime} (h_{t,1} - d_{t,1})$$

Substituting $A_t y_{t-1} + C_t^{(1)} s_t^{(1)} + C_t^{(2)} s_{t,e}^{(2)} + B_t z_t + D_t + u_t$ for y_t, one obtains:

$$V_t^{(1)} = E_{T-1}[\exp(\omega_2(u_t))] \exp \omega_1(I_{t-1}, s_{t,e}^{(2)}, s_t^{(1)}, z_t, \beta_t, K_{t,1}, d_{t,1}, y_t^a)$$

$$= \exp \omega_1(I_{t-1}, s_{t,e}^{(2)}, s_t^{(1)}, z_t, \beta_t, K_{t,1}, d_{t,1}, y_t^a) \int\limits_{R^p} (2\pi)^{-\frac{p}{2}} \mid \det \Psi \mid^{-\frac{1}{2}} \exp(-\frac{1}{2}\widetilde{u}_t' \Psi^{-1} \widetilde{u}_t) \exp \omega_2(\widetilde{u}_t)\, d\widetilde{u}_t$$

with $\omega_2(\widetilde{u}_t)$ a quadratic function in $\widetilde{u}_t$.

One can write:

$$\overline{\mathbf{I}}_t^{(1)} \overset{not.}{=} \int\limits_{R^p} (2\pi)^{-\frac{p}{2}} \mid \det \Psi \mid^{-\frac{1}{2}} \exp(-\frac{1}{2}\widetilde{u}_t' \Psi^{-1} \widetilde{u}_t) \exp \omega_2(\widetilde{u}_t)\, d\widetilde{u}_t$$

$$= \int\limits_{R^p} (2\pi)^{-\frac{p}{2}} \mid \det \Psi \mid^{-\frac{1}{2}} \exp(-\frac{1}{2}\widetilde{u}_t'(\Psi^{-1} + \varphi_{t,1}^{(1)} H_{t,2})\widetilde{u}_t + \text{linear in } \widetilde{u}_t') d\widetilde{u}_t$$

$$= \mid \det(\Psi^{-1} + \varphi_{t,1}^{(1)} H_{t,2}) \mid^{-\frac{1}{2}} \mid \det \Psi \mid^{-\frac{1}{2}} \int\limits_{R^p} (2\pi)^{-\frac{p}{2}} \mid \det(\Psi^{-1} + \varphi_{t,1}^{(1)} H_{t,2}) \mid^{\frac{1}{2}} \exp \omega_3(\widetilde{u}_t)\, d\widetilde{u}_t$$

with $\omega_3(\widetilde{u}_t)$ a quadratic function in $\widetilde{u}_t$. Now, we find $\overline{u}_t \in R^p$ such that:

$$\omega_3(\widetilde{u}_t) = -\frac{1}{2}(\widetilde{u}_t - \overline{u}_t)'(\Psi^{-1} + \varphi_{t,1}^{(1)} H_{t,1})(\widetilde{u}_t - \overline{u}_t) + \text{independent of } \widetilde{u}_t$$

It follows that:

$$\text{independent of } \widetilde{u}_t = \frac{1}{2}\overline{u}_t'(\Psi^{-1} + \varphi_{t,1}^{(1)} H_{t,1})\overline{u}_t - \frac{\varphi_{t,1}^{(1)}}{2} f_{t,1} \overset{not.}{=} \omega_4(\overline{u}_t)$$

and

$$-\varphi_{t,1}^{(1)} \widetilde{u}_t'[K_{t,1}(A_t y_{t-1} + C_t^{(1)} s_t^{(1)} + C_t^{(2)} s_{t,e}^{(2)} + B_t z_t + D_t) - h_{t,1}] = \widetilde{u}_t'(\Psi^{-1} + \varphi_{t,1}^{(1)} H_{t,1})\overline{u}_t$$

that is,

$$\overline{u}_t = -\varphi_{t,1}^{(1)}(\Psi^{-1} + \varphi_{t,1}^{(1)} H_{t,1})^{-1}[K_{t,1}(A_t y_{t-1} + C_t^{(1)} s_t^{(1)} + C_t^{(2)} s_{t,e}^{(2)} + B_t z_t + D_t) - h_{t,1}]$$

Thus, the integral becomes:

$$\overline{\mathbf{I}}_t^{(1)} = \mid \det(\Psi^{-1} + \varphi_{t,1}^{(1)} H_{t,1}) \mid^{-\frac{1}{2}} \mid \det \Psi \mid^{-\frac{1}{2}} \exp(\frac{1}{2}\overline{u}_t'(\Psi^{-1} + \varphi_{t,1}^{(1)} H_{t,1})\overline{u}_t - \frac{\varphi_{t,1}^{(1)}}{2} f_{t,1}) \cdot$$

$$\int\limits_{R^p} (2\pi)^{-\frac{p}{2}} \mid \det(\Psi^{-1} + \varphi_{t,1}^{(1)} H_{t,1}) \mid^{\frac{1}{2}} \exp(-\frac{1}{2}(\widetilde{u}_t - \overline{u}_t)'(\Psi^{-1} + \varphi_{t,1}^{(1)} H_{t,1})(\widetilde{u}_t - \overline{u}_t)) d\widetilde{u}_t$$

The last integral is equal to 1 because the integrand is the probability density function of a p-dimensional normal random variable:

$$\widetilde{u}_t \sim \mathcal{N}(\overline{u}_t, (\Psi^{-1} + \varphi_{t,1}^{(1)} H_{t,1})^{-1})$$

with -1 power denoting inverse.

If we replace $\overline{u}_t$ by its value, we find without difficulty:

$$\overline{\mathbf{I}}_t^{(1)} = \mid \det(I_p + \varphi_{t,1}^{(1)} \Psi H_{t,1}) \mid^{-\frac{1}{2}} \exp \omega_4(I_{t-1}, s_{t,e}^{(2)}, s_t^{(1)}, z_t, \beta_t, K_{t,1}, d_{t,1}, y_t^a)$$

where I_p denotes the $p \times p$ identity matrix.

By consequence, we have:

$$V_t^{(1)} \overset{not.}{=} E_{t-1}[\exp(-\frac{\varphi_{t,1}^{(1)}}{2} W_t^{(1)}(y_t))] = \exp \omega_1(I_{t-1}, s_{t,e}^{(2)}, s_t^{(1)}, z_t, \beta_t, K_{t,1}, d_{t,1}, y_t^a) \cdot \overline{\mathbf{I}}_t^{(1)}$$

$$= \mid \det(I_p + \varphi_{t,1}^{(1)} \Psi H_{t,1}) \mid^{-\frac{1}{2}} \exp \omega_5(I_{t-1}, s_{t,e}^{(2)}, s_t^{(1)}, z_t, \beta_t, K_{t,1}, d_{t,1}, y_t^a)$$

where

$$\omega_5 \overset{def.}{=} \omega_1 + \omega_4$$

After several algebraic manipulations, one obtains:

$$\omega_5(I_{t-1}, s_{t,e}^{(2)}, s_t^{(1)}, z_t, \beta_t, K_{t,1}, d_{t,1}, y_t^a) = -\frac{\varphi_{t,1}^{(1)}}{2}[y_{t-1}' A_t' \overline{H}_t^{(1)} (C_t^{(1)} s_t^{(1)} + C_t^{(2)} s_{t,e}^{(2)})$$

$$+(s_{t,e}^{(2)\prime} C_t^{(2)\prime} + s_t^{(1)\prime} C_t^{(1)\prime}) \overline{H}_t^{(1)} (A_t y_{t-1} + B_t z_t + D_t) + (s_{t,e}^{(2)\prime} C_t^{(2)\prime} + s_t^{(1)\prime} C_t^{(1)\prime}) \overline{H}_t^{(1)} (C_t^{(1)} s_t^{(1)} + C_t^{(2)} s_{t,e}^{(2)})$$

$$+(z_t' B_t' + D_t') \overline{H}_t^{(1)} (C_t^{(1)} s_t^{(1)} + C_t^{(2)} s_{t,e}^{(2)})] + \varphi_{t,1}^{(1)} (s_{t,e}^{(2)\prime} C_t^{(2)\prime} + s_t^{(1)\prime} C_t^{(1)\prime})[I_p - \varphi_{t,1}^{(1)} K_{t,1} (\Psi^{-1} + \varphi_{t,1}^{(1)} H_{t,1})^{-1}] h_{t,2}$$

$$+\text{independent of } s_t^{(1)}$$

where:

$$\overline{H}_t^{(1)} \overset{not.}{=} K_{t,1} - \varphi_{t,1}^{(1)} H_{t,1} M_{t,1}^{-1}(\varphi_{t,1}^{(1)}) H_{t,1}; \quad M_{t,1}(\varphi_{t,1}^{(1)}) \overset{not.}{=} \Psi^{-1} + \varphi_{t,1}^{(1)} H_{t,1}$$

The first order condition in $s_t^{(1)}$ writes:

$$-\frac{\varphi_{t,1}^{(1)}}{2} C_t^{(1)\prime} \overline{H}_t^{(1)} A_t y_{t-1} - \frac{\varphi_{t,1}^{(1)}}{2} C_t^{(1)\prime} \overline{H}_t^{(1)} (A_t y_{t-1} + B_t z_t + D_t) - \varphi_{t,1}^{(1)} C_t^{(1)\prime} \overline{H}_t^{(1)} C_t^{(1)} s_t^{(1)}$$

$$-\frac{\varphi_{t,1}^{(1)}}{2} C_t^{(1)\prime} \overline{H}_t^{(1)\prime} C_t^{(2)} s_{t,e}^{(2)} - \frac{\varphi_{t,1}^{(1)}}{2} C_t^{(1)\prime} \overline{H}_t^{(1)} C_t^{(2)} s_{t,e}^{(2)} - \frac{\varphi_{t,1}^{(1)}}{2} C_t^{(1)\prime} \overline{H}_t^{(1)} (B_t z_t + D_t)$$

$$+\varphi_{t,1}^{(1)} C_t^{(1)\prime} [I_p - \varphi_{t,1}^{(1)} K_{t,1} (\Psi^{-1} + \varphi_{t,1}^{(1)} H_{t,1})^{-1}] h_{t,1} = 0 \overset{(\varphi_{t,1}^{(1)} \neq 0)}{\Leftrightarrow} -C_t^{(1)\prime} \overline{H}_t^{(1)} A_t y_{t-1} - C_t^{(1)\prime} \overline{H}_t^{(1)} (B_t z_t + D_t)$$

$$-C_t^{(1)\prime} \overline{H}_t^{(1)} C_t^{(2)} s_{t,e}^{(2)} + C_t^{(1)\prime} [I_p - \varphi_{t,1}^{(1)} K_{t,1} (\Psi^{-1} + \varphi_{t,1}^{(1)} H_{t,1})^{-1}] h_{t,1} = C_t^{(1)\prime} \overline{H}_t^{(1)} C_t^{(1)} s_t^{(1)}$$

It follows that:

$$s_t^{g(1)}(I_{t-1}, z_t, \beta_t, K_{t,1}, d_{t,1}, y_t^a, s_{t,e}^{(2)}) = G_t^{(1)} y_{t-1} + g_t^{(1)}$$

with

$$G_t^{(1)} \overset{not.}{=} -(C_t^{(1)\prime} \overline{H}_t^{(1)} C_t^{(1)})^{-1} (C_t^{(1)\prime} \overline{H}_t^{(1)} A_t)$$

$$g_t^{(1)} \overset{not.}{=} -(C_t^{(1)\prime}\overline{H}_t^{(1)}C_t^{(1)})^{-1}C_t^{(1)\prime}[\overline{H}_t^{(1)}(B_tz_t+D_t+C_t^{(2)}s_{t,e}^{(2)})-(I_p-\varphi_{t,1}^{(1)}K_{t,1}(\Psi^{-1}+\varphi_{t,1}^{(1)}H_{t,1})^{-1})h_{t,1}]$$

Let us now analyze the stochastic optimization programme for the follower. At each period t, the reaction strategy of the follower is described by the asymmetric information problem:

$$s_t^{g(2)} = \underset{s_t^{(2)}\in\Lambda_t^{(2)}}{\arg\max} E_{t-1}U_t^{(2)}(W_{[1,t]}^{(2)}(s_t^{(2)}, s_t^{g(1)}) \mid s_t^{g(1)})$$

$$s.t.: \begin{cases} y_t = A_ty_{t-1} + C_t^{(1)}s_t^{g(1)} + C_t^{(2)}s_t^{(2)} + B_tz_t + D_t + u_t \\ y_0,\ y_t > 0, \quad t = 1, ..., T_2 \quad \text{(economic constraints)} \end{cases}$$

Using the same reasoning as above, we find without difficulty the following linear control equations:

$$s_t^{g(2)}(I_{t-1}, z_t, \beta_t, K_{t,2}, d_{t,2}, y_t^g, s_t^{g(1)}) = G_t^{(2)}y_{t-1} + g_t^{(2)},\ t = 1, ..., T_2$$

with:

$$G_t^{(2)} \overset{not.}{=} -(C_t^{(2)\prime}\overline{H}_t^{(2)}C_t^{(2)})^{-1}(C_t^{(2)\prime}\overline{H}_t^{(2)}A_t)$$

$$g_t^{(2)} \overset{not.}{=} -(C_t^{(2)\prime}\overline{H}_t^{(2)}C_t^{(2)})^{-1}C_t^{(2)\prime}[\overline{H}_t^{(2)}(B_tz_t + D_t + C_t^{(1)}s_t^{g(1)}) - (I_p - \varphi_{t,1}^{(2)}K_{t,2}M_{t,2}^{-1}(\varphi_{t,1}^{(2)}))h_{t,2}]$$

and

$$\overline{H}_t^{(2)} \overset{not.}{=} K_{t,2} - \varphi_{t,1}^{(2)}H_{t,2}M_{t,2}^{-1}(\varphi_{t,1}^{(2)})H_{t,2}$$

$$M_{t,2}(\varphi_{t,1}^{(2)}) \overset{not.}{=} \Psi^{-1} + \varphi_{t,1}^{(2)}H_{t,2}$$

$$H_{t,2} \overset{def.}{=} K_{t,2}, \quad h_{t,2} \overset{not.}{=} K_{t,2}y_t^g - d_{t,2}$$

It is important to note that the strategies adopted depend on the players' types. The present approach improves the traditional solution concept of Stackelberg strategy with feedback /closed-loop information structure for discrete-time linear-quadratic games in the case when the players exhibit endogenous risk (Chen and Cruz Jr., 1972; Simaan and Cruz Jr. 1973a, 1973b; Gardner Jr. and Cruz Jr., 1977; Medanic 1978; Basar and Selbuz 1979; Tolwinski 1981; Jibbe et al., 1984; Zadrozny 1988; Ambler and Paquet 1997; Li et al. 2002; Chen and Zadrozny 2002; Jungers et al. 2006; Nie et al., 2008, amongst others).

8. Concluding Remarks

In this paper, the risk-sharing contract between a leader and a follower is studied in the case of a discrete-time finite-horizon environmental game with coupling constraints. The proposed approach is appropriate for a wide class of economic problems related to incentive Stackelberg games with closed-loop information structure and endogenously risk-averse players. The traditional hypothesis of a risk-neutral leader is a potential borderline case in the present framework. This possibility must be envisaged with prudence when considering dynamic strategic interactions in highly fluctuating environments. The study allows to differentiate between the players' risk profiles and shows that there exist situations where the leader is more risk-averse compared to the follower. The effects of potential taxes and recompenses on the follower's risk behavior are explored. Several realistic scenarios revealing the potential dominant effect of taxes on recompenses, and vice versa, are proposed in this direction. Taxes generally amplify the follower's degree of risk-aversion, while recompenses have a compensation effect. A good topic for further research is to develop a behavioural game-theoretic approach which examines the government-private sector relationship from this new perspective. The gouvernment is the natural leader (it has the power to tax or to grant behavior) while the private sector is the natural follower that optimizes its stochastic objective function for given governmental decisions. For this type of relationship, the private sector's ability plays an important role in making rational expectations about the government decisions. The case where governmental free-riding exists is another interesting topic of research that is worth exploring. We leave this challenge to future studies.

References

[**1**] Abreu, D., 1988. "On theTheory of Infinitely Repeated Games with Discounting", *Econometrica*, **56**, 384-396.

[**2**] Ambler, S., and A. Paquet, 1997. "Recursive methods for computing equilibria of general equilibrium dynamic Stackelberg games", *Economic Modelling*, **14**, 155–173.

[**3**] Basar, T. and Selbuz, H., 1979. "Closed-loop Stackelberg strategies with applications in the optimal control of multilevel systems", *IEEE Transactions on Automatic Control*, **24**, 166-179.

[**4**] Basar, T. and Olsder, G. J., 1980. "Team-optimal closed-loop Stackelberg strategies in hierarchical control problems", *Automatica,* **16**, 409-414.

[**5**] Basar, T. and Olsder, G. J., 1995. Dynamic Non-cooperative Game Theory. 2 nd ed., *Academic Press*, London.

[**6**] Borch, K. H., 1962. "Equilibrium in a Reinsurance Market", *Econometrica*, **30**, 424-444.

[**7**] Breton, M., Sokri, A. and Zaccour, G., 2008. "Incentive equilibrium in an overlapping-generations environmental game", *European Journal of Operational Research*, **2**, 687-699.

[**8**] Bystrom, O. and Bromley, D. W., 1998. "Contracting for Nonpoint-Source Pollution Abatement", *Journal of Agricultural and Ressource Economics*, **23** (1), 39-54.

[**9**] Cabe, R. and J. Herriges, 1992. "The Regulation of Nonpoint-Source Pollution under Imperfect and Asymmetric Information", *Journal of Environmental Economics and Management*, **22** (2), 134-146.

[**10**] Caussade, P. and M. Prat, 1990. "Transport Modelling in Watershed", *Ecological Modelling*, **52**, 135-179.

[**11**] Chen, C. I. and J. B. Cruz Jr., 1972. "Stackelberg solution for two-person games with biased information patterns", *IEEE Transactions on Automatic Control*, **17**, 791-798.

[**12**] Chen, P. and P. A. Zadrozny, 2002. "An anticipative feedback solution for the infinite-horizon linear-quadratic dynamic Stackelebrg game", *Journal of Economic Dynamics and Control*, **26**, 1397-1416.

[**13**] Cohen, D. and P. Michel, 1988. "How Should Control Theory be Used to Calculate a Time Consistent Government Policy", *Review of Economic Studies*, **55**, 263-374.

[**14**] Corner, J. L., and P. D. Corner, 1995. "Characteristics of decisions in decision analysis practice", *The Journal of Operational Research Society*, **46**, 304-314.

[**15**] Dinar, A., Albiac, J., and Sanchez-Soriano, J., 2008. Game theory and policymaking in natural resources and the environment, Routledge (London - GBR), 345 p.

[**16**] Evans, R. and J. P. Thomas, 2001. "Cooperation and Punishment", *Econometrica*, **4**, 1061-1075.

[**17**] Filar, J. and Carraro, C., 1995. Control and Game-Theoretic Models of the Environment, Vol. **2**, Springer-Verlag New York, LLC, 388 p.

[**18**] Gardner, B. F. Jr. and J. B. Cruz, Jr., 1977. "Feedback Stackelberg Strategy for a Two Player Game", *IEEE Trans. on Automatic Control*, **22**, 270-271.

[**19**] Jehiel, P. and B. Moldovanu., 1995. "Negative externalities may cause delay in negotiation", *Econometrica*, **6**, 1321-1335.

[**20**] Jibbe, M., Sawan, M., and Tran, M., 1984. "Stackelberg strategies in discrete linear quadratic problems", *International Journal of Systems Science*, **15**, 1279-1282.

[**21**] Jungers, M., Trelat, E., and Abou-Kandil, H., 2006. "Stackelberg strategy with closed-loop information structure for linear-quadratic games", Preprint Hal, 23 pages.

[**22**] Krawczyk, J. B., 1985. "Management of effluent discharges: a dynamic game model", in Control and Game-Theoretic Models of Environment, Annals of the International Society of Dynamic Games, eds Carraro. C & A. Filar.

[**23**] Krawczyk, J. B., 1995. "Controlling a dam Environmentally acceptable standards through the use of a decision support tool", *Environmental and Resources Economics*, **5** (3), 287-304.

[**24**] Kydland, F., 1975. "Non-cooperative and Dominant Player Solutions in Discrete Dynamic Games", *International Economic Review*, **16**, 321-335.

[**25**] Kydland, F., 1977. "Equilibrium solutions in dynamic dominant-player models", *Journal of Economic Theory*, **15**, 307-324.

[**26**] Kydland, F., and E. Prescott., 1977. "Rules Rather than Discretion: The Inconsistency of Optimal Plans", *Journal of Political Economy*, **85**, 473-493.

[**27**] Laffont, J. J., 2003. The Principal-Agent Model: The Economic Theory of Incentives, Edward Edger Publishing limited, Cheltenham, UK.

[**28**] Laffont, J. J. and Martimort, D., 2002. "The Theory of Incentives: The Principal-Agent Model", *Princeton University Press*, Princeton, New Jersey.

[**29**] Laffont, J. J. and Tirole, J., 1993. "A Theory of Incentives in Procurement and Regulation". *The MIT Press*, Cambridge, Massachusetts.

[**30**] Li, M., Cruz Jr., J. B., and Simaan, M. A., 2002. "An approach to discrete-time incentive feedback Stackelberg games", *IEEE Transactions on Systems, Man and Cybernetics*, Part A: System and Humans, **32**, 472–481.

[**31**] Medanic, J., 1978. "Closed-loop Stackelberg strategies in linear-quadratic problems", *IEEE Transactions on Automatic Control*, **23**, 632-637.

[**32**] Meran, G. and U. Schwalbe, 1987. "Pollution Control and Collective Penalties", *Journal of Institutional and Theoretical Economics*, **143** (4), 616-629.

[**33**] Miller, M. and M. Salmon, 1985. "Dynamic Games and the Time Inconsistency of Optimal Policy in Open Economis", *The Economic Journal*, **95**, 124–137.

[**34**] Murphy, K. J., 1992. "Determinants of contract durations in collective bargaining agreements", *Industrial and Labor Relations Review*, **45**, 352-365.

[**35**] Myerson, R. B., 1997. Game Theory: Analysis of Conflicts, Harvard University Press, Cambridge, Massachusetts.

[**36**] Nash, J., 1950. "Equilibrium points in n-person games", *Proceedings of the National Academy of Sciences*, **36**, 48-49.

[**37**] Nash, J., 1951. "Non-cooperative games", *Annals of Mathematics*, **54**, 286-295.

[**38**] Nie, P. Y., Lai M. Y., and Zhu, S. J., 2008. "Dynamic feedback Stackelberg games with non-unique solution", *Nonlinear Analysis: Theory, Methods & Applications*, **69**, 1904-1913.

[**39**] Pareto, V. 1909. Manuel d'ecoonomie politique, Qiard, Paris.

[**40**] Pindyck R. S., 2001. "Optimal timing problems in environmental economics", Massachusetts Institute of Technology, Cambridge, MA 02139.

[**41**] Protopopescu, D., 2007. "Improving the Risk Concept: A Revision of Arrow-Pratt Theory in the Context of Controlled Dynamic Stochastic Environments", *Working Paper* No. **727.08**, CODE, Universitat Autònoma de Barcelona, Spain.

[**42**] Roth, J., Schulz, J. and White, A., 1989. "Taxpayer Compliance", Philadelphia, *University of Pensylvania Press.*

[**43**] Sandblom, C. L. and J. L. Banasik, 1985. "Economic policy with bounded controls", *Economic Modelling,* **2**, 135-148.

[**44**] Segerson, K., 1988: "Uncertainty and Incentives for Nonpoint Pollution Control", *Journal of Environmental Economics and Management*, **15** (1), 87-98.

[**45**] Shortle, J. S. and J. W. Dunn, 1986. "The Relative Efficiency of Agricultural Source Water Pollution Control Policies", *American Journal of Agricultural Economics*, **68**, 668-677.

[**46**] Shortle, J. S., Horan, R. D. and Abler, D. G., 1998. "Research Issues in Nonpoint Pollution Control", *Environmental and Ressource Economics*, **11**, 571-585.

[**47**] Simaan, M. and J. B. Cruz Jr., 1973. "On the Stackelberg strategy in nonzero-sum games", *Journal of Optimization Theory and Applications,* **5**, 533-555.

[**48**] Simaan, M. and J. B. Cruz Jr., 1973. "Additional aspects of the Stackelberg strategy in nonzero-sum games", *Journal of Optimization Theory and Applications,* **11**, 613-626.

[**49**] Stackelberg, Von H., 1934. Marketform und Gleichgewicht, Springer, Viena. An english translation appeared in 1952 entitled "The Theory of Market Economy", Published by Oxford University Press, Oxford.

[**50**] Tidball, M. and G. Zaccour, 2005. "An environmental game with coupling constraints", *Environmental Modeling and Assessment*, **2**, 153-158.

[**51**] Tolwinski, B., 1981. "Closed-loop Stackelberg solution to a multistage linear-quadratic game", *Journal of Optimization Theory and Applications*, **34**, 485-501.

[**52**] Van Der Ploeg, L., 1984. "Risk and Linear-Quadratic Stabilization", *Economic Letters*, **15**, 73-78.

[**53**] Vroman, S. B., 1989. "Inflation uncertainty and contract duration", *Review of Economics and Statistics*, **71**, 677-681.

[**54**] Wu, J., Teague, M. L., Mapp, H. P. and D. J. Bernardo, 1995. "An empirical analysis of the relative efficiency of policy instruments to reduce nitrate water pollution in the U. S. southern high plains", *Canadian Journal of Agricultural Economics*, **43** (3), 403-420.

[**55**] Xepapadeas, A. P., 1991. "Environmental Policy under Imperfect Information", *Journal of Environmental Economics and Management*, **20** (2), 113-126.

[**56**] Xepapadeas, A. P., 1992. "Environmental Policy Design and Dynamic Non-point Source Pollution", *Journal of Environmental Economics and Management*, **23** (1), 22-39.

[**57**] Xepapadeas, A. P., 1994. "Controlling environmental externality: observability and optimal policy rules", Dortrecht, Kluwer Academic Publishers.

[**58**] Zadrozny, P. A., 1988. "A consistent closed-loop solution for infinite-horizon linear-quadratic dynamic Stackelberg games", *Journal of Economic Dynamics and Control*, **12**, 155-159.