ISSN 0035-6751

INTERNATIONAL REVIEW OF ECONOMICS AND BUSINESS - Volume XLII, n. 1-2 - January - February 1995

RIVISTA INTERNAZIONALE DI SCIENZE ECONOMICHE E COMMERCIALI

Anno XLII Gennaio-Febbraio 1995

Pubblicazione mensile - Sped. in abb. postale/50% - Bologna

N.ro INVENTARIO 8463

SOMMARIO

Social Security, Savings, and Retirement Behaviour in a Life-cycle Model of Economic Growth (Previdenza, risparmio e pensionamenti in un modello di crescita economica con l'ipotesi di ciclo della vita)
SIKANDAR SIDDIQUI Pag. 1

How and When Unanimity is a Superior Decision Rule (Come e quando l'unanimità è conveniente)
MARIO S. CATALANI and GIUSEPPE F. CLERICO » 19

A Model of the Audience for TV Broadcasting: Implications for Advertising Competition and Regulation (Un modello dell'audience televisivo: implicazioni per la concorrenza e la regolamentazione della pubblicità)
ALESSANDRO VAGLIO » 33

Taxation, Labor Supply, and Savings: An Intertemporal Analysis (Imposte, offerta di lavoro e risparmio: una analisi intertemporale)
THEODORE PALIVOS and CHONG K. YIP » 57

Corporate Taxation and Monopoly: The Effects of an Announced Reform (Imposte sui redditi societari e monopolio: gli effetti di una riforma annunciata)
PAOLO M. PANTEGHINI » 69

Controllo ottimo stocastico, processi browniani regolati e optimal stopping (Optimal Stochastic Control, Regulated Brownian Processes, and Optimal Stopping)
MICHELE MORETTO » 93

Verso il sistema economico diacronicamente completo: esposizione ragionata della Parte V di G. Demaria, *A New Economic Logic*, 1993 (Towards a Diachronically Complete Economic System: Exposition of Part V of G. Demaria's, *A New Economic Logic*, 1993)
ACHILLE AGNATI » 125

Recensioni (Book-reviews) » 143

SOTTO GLI AUSPICI DELLA
UNIVERSITÀ COMMERCIALE LUIGI BOCCONI
E DELLA UNIVERSITÀ DEGLI STUDI DI MILANO

CEDAM

COMITATO DI DIREZIONE - EDITORIAL BOARD

HENRI BARTOLI (Université de Paris) - WILLIAM J. BAUMOL (Princeton University) - GIOVANNI DEMARIA (Accademia Nazionale dei Lincei) - WILLIAM D. GRAMPP (Illinois University) - ARNALDO MAURI (Università di Milano) - ARIBERTO MIGNOLI (Università Bocconi) - ANTONIO MONTANER (Universität Mainz) - HISAO ONOE (Kyoto University) - ALBERTO QUADRIO CURZIO (Università Cattolica, Milano) - ROBERTO RUOZI (Università Bocconi) - ROBERT M. SOLOW (Massachusetts Institute of Technology) - SERGIO STEVE (Università di Roma) - MARIO TALAMONA (Università di Milano) - SHIGETO TSURU (Hitotsubashi University) - BASIL S. YAMEY (London School of Economics and Political Science).

DIRETTORE (EDITOR): ALDO MONTESANO (Università Bocconi)
Redazione (Editorial Office): ANNA BAGIOTTI CRAVERI
DIRETTORE (EDITOR) dal 1954 al 1983: TULLIO BAGIOTTI

RIVISTA INTERNAZIONALE DI SCIENZE ECONOMICHE E COMMERCIALI
(INTERNATIONAL REVIEW OF ECONOMICS AND BUSINESS)

Pubblicazione mensile (A monthly journal). Direzione e Redazione (Editorial Office): Via Teuliè 1, 20136 Milano (Italy), Tel. 02-58317434.

Abbonamento 1995 (Subscription 1995): Italia (Italy), Lire 200.000; estero (abroad), Lire 300.000. Annate arretrate rilegate disponibili a prezzi speciali (Bound back volumes available at special prices).

Per tutto ciò che riguarda l'AMMINISTRAZIONE, ci si rivolga unicamente alla **CEDAM** - Casa Editrice Dott. Antonio Milani - Via Jappelli 5/6 - 35121 Padova (telefono 049/65.66.77 r.a. - fax 049/87.52.900 - Casella postale 1093 - Indirizzo telegrafico: Cedam, Padova - Conto corrente postale n. 205351 nell'ufficio dei conti di Venezia).

CONDIZIONI DI ABBONAMENTO AI PERIODICI « CEDAM »

L'abbonamento è annuo e si rinnova tacitamente per l'anno successivo se non viene disdetto entro il mese di dicembre, con lettera raccomandata. Il canone di abbonamento deve essere pagato anticipatamente. In caso di mancato pagamento la Casa non effettuerà l'invio dei fascicoli. I pagamenti possono essere effettuati direttamente alla Cedam di Padova sul c/c postale n. 205351 oppure ai suoi incaricati muniti di speciale delega, che rilasceranno ricevuta sui moduli recanti il marchio Cedam e numerati progressivamente. Il rinnovo dell'abbonamento deve essere effettuato entro il 30 aprile di ogni anno. Trascorso tale termine l'amministrazione provvederà direttamente all'incasso mediante emissione di fattura e sospenderà l'invio dei fascicoli nel caso di mancato pagamento. I fascicoli non pervenuti all'abbonato devono essere reclamati prima della conclusione dell'abbonamento in corso. Decorso tale termine saranno spediti, se disponibili, contro rimessa dell'importo. L'abbonamento comporta, agli effetti legali, elezione di domicilio in Padova presso la Casa Editrice Cedam.

Direttore responsabile: Aldo Montesano - Autorizz. Tribunale di Treviso N. 113 del 22-10-54

Rivista associata all'Unione della Stampa Periodica Italiana

Tip. Leonelli - Villanova di Castenaso (Bo)

Proprietà letteraria - Stampato in Italia - Printed in Italy

Rivista Internazionale di Scienze Economiche e Commerciali
Volume 42 (1995), N. 1-2, 1-18

SOCIAL SECURITY, SAVINGS, AND RETIREMENT BEHAVIOUR IN A LIFE-CYCLE MODEL OF ECONOMIC GROWTH



by
SIKANDAR SIDDIQUI *

1. *Introduction*

In most of the industrialized countries of the West, social security systems have expanded rather steadily since the end of the Second World War. This process was accompanied by a tendency towards a lower average retirement age of elderly citizens, which was most pronounced in the 1960's and 1970's but continued during the last decade (see, e.g., Jacobs, Kohli, Rein, 1988, and Reimann, 1985, for corresponding figures on the West Germany).

Yet the post-war baby boom, which these developments can at least in part be attributed to, gave way to prolonged decline in birth rates in the mid-sixties. The resulting process of population ageing is very likely to affect the financial basis of public pension systems negatively unless appropriate policy reforms are undertaken.

In this paper, the interaction between macroeconomic variables like wages and interest rates and the modalities of an unfunded public pension system is analyzed within the framework of a discrete-time overlapping generations model in which the labour supply behaviour of the old generation is determined endogenously. This approach was introduced by Hu (1979) and taken up by Breyer (1990). The model which is set up here, however, differs from Hu's in three important aspects: Firstly, it is based on

* Lehrstuhl für Ökonometrie, Universität Konstanz, Konstanz (Germany).

I would like to thank Paul Johnson, Lex Meijdam, Bernd Raffelhüschen, Winfried Pohlmeier and seminar participants at Mannheim for helpful comments, and Monika Buhl and Klaus Kornmesser for excellent research assistance. Financial support from the Deutsche Forschungsgemeinschaft (DFG) is gratefully acknowledged.

the premise of perfect foresight (and not of stationary expectations); secondly, it assumes an actuarially fair pension system; and thirdly, it explicitly takes into account the link between wages, social security taxes and benefits, and population growth.

The approach presented here is closely linked in spirit to the recent work by Breyer and Straub (1993), who discuss the welfare effects of social security within the framework of an overlapping generations model in which the labour supply behaviour of the young generation is regarded as endogenous,whereas the amount of leisure enjoyed by the members of the older one is assumed to be constant. A major theoretical problem associated with such an approach is, however, that it is incompatible with the concept of a perfect foresight equilibrium unless the assumption of infinitely forward-looking individuals is introduced. It will become obvious in Section 3.1. that theoretical difficulty is absent in the model presented here.

The outline of the paper is as follows: In Section two, the basic assumptions about the behaviour of households and firms and about the structure of the pay-as-you-go social security system are introduced. Section 3.1. deals with the conditions for the existence of a unique macroeconomic equilibrium and addresses the question of the uniqueness and stability of possible steady states. The question of intergenerational efficiency is dealt with in Section 3.2. Section 3.3. focuses on the consequences of demographic changes with regard to capital formation and economic growth. The paper concludes with a brief summary.

2. *Basic Assumptions*

2.1. *The Household Sector.* – The basic structure of the model presented here equals the one of the standard overlapping generations model introduced by Samuelson (1958, 1975) and Diamond (1965): The population of the model economy consists of two generations of identical individuals and grows at a constant rate n. The absolute size of the young generation at time t is denoted by $N_t^{(1)}$ and the one of the old generation by $N_t^{(2)}$. Every individual is assumed to live for two periods, so that the overall size of the population adds up to

$$\bar{N}_t := N_t^{(1)} + N_t^{(2)} \tag{2.1.1}$$
$$= N_t^{(1)} + N_{t-1}^{(1)}$$
$$= (2 + n)\, N_{t-1}^{(1)}$$

An individual born at time t is assumed to supply inelastically one unit of labour during the period of his life, earning a wage ω_t, and to work a fraction $(1 - f^{(2)}_{t+1})$ of the second period before retiring and receiving a pension P_{t+1} from the government. f_t denotes the amount of leisure enjoyed by an individual in period t and is normalized to the interval [0, 1], and r_t is the interest rate prevailing at that time. Savings (s_t) are taken to be non-negative. Besides that, the assumption is made that every member of the young generation is obliged to pay a fraction b_t of his labour income as a social security tax. Under these conditions, the utility maximization problem for a person born at time t can be described as follows:

$$
\begin{aligned}
(2.1.2) \quad \max \Lambda = {} & u(c^{(1)}_t, f^{(1)}_t = 0) + \frac{1}{1+\rho} u(c^{(2)}_{t+1}, f^{(2)}_{t+1}) \\
& + \lambda_t \{\omega_t (1 - b_t) - c^{(1)}_t - s_t\} \\
& + \lambda_{t+1} \{\omega_{t+1}(1 - f^{(2)}_{t+1}) + P_{t+1} \\
& + (1 + r_{t+1}) s_t - c^{(2)}_{t+1}\} \\
& + \mu_1 \{1 - f^{(2)}_{t+1}\} + \mu_2 f^{(2)}_{t+1} + \mu_3 s_t
\end{aligned}
$$



The multipliers μ_1 to μ_3 are associated with the above-mentioned inequality constraints on $f^{(2)}_{t+1}$ and s_t:

$$
(2.1.3) \qquad \mu_1 \begin{cases} \geq 0 \text{ if } f^{(2)}_{t+1} = 1 \\ = 0 \text{ if } f^{(2)}_{t+1} < 1 \end{cases}
$$

$$
(2.1.4) \qquad \mu_2 \begin{cases} \geq 0 \text{ if } f^{(2)}_{t+1} = 0 \\ = 0 \text{ if } f^{(2)}_{t+1} > 0 \end{cases}
$$

$$
(2.1.5) \qquad \mu_3 \begin{cases} \geq 0 \text{ if } s_t = 0 \\ = 0 \text{ if } s_t > 0 \end{cases}
$$

The "instantaneous utility" function $u(c_t, f_t)$ is assumed to be increasing and quasi concave with regard to both of its arguments. Thus, differentiation of (2.1.2) with respect to $c^{(1)}_t$, $c^{(2)}_{t+1}$, s_t and $f^{(2)}_{t+1}$ yields the following optimality conditions:

$$
(2.1.6.) \qquad \frac{\partial \Lambda}{\partial c^{(1)}_t} = \frac{\partial u}{\partial c^{(1)}_t} - \lambda_t = 0
$$

$$
(2.1.7.) \qquad \frac{\partial \Lambda}{\partial c^{(2)}_{t+1}} = \frac{1}{1+\rho} \frac{\partial u}{\partial c^{(2)}_{t+1}} - \lambda_{t+1} = 0
$$

$$(2.1.8.)\quad \frac{\partial \Lambda}{\partial s_t} = -\lambda_t + (1 + r_{t+1})\lambda_{t+1} + \mu_3 = 0$$

$$(2.1.9.)\quad \frac{\partial \Lambda}{\partial f^{(2)}_{t+1}} = \frac{1}{1+\rho}\frac{\partial u}{\partial f^{(2)}_{t+1}} - \lambda_{t+1} w_{t+1} - \mu_1 + \mu_2 = 0$$

Equations (2.1.6.) to (2.1.8.) can be summarized to

$$(2.1.10.)\quad \frac{\partial u/\partial c^{(1)}_t}{\partial u/\partial c^{(2)}_{t+}} \begin{cases} = \dfrac{1 + r_{t+1}}{1+\rho} & \text{if } c_t > w_t(1-b) \\ \\ \geq \dfrac{1 + r_{t+1}}{1+\rho} & \text{if } c_t = w_t(1-b) \end{cases}$$

and combination of (2.1.7.) and (2.1.9.) yields

$$(2.1.11.)\quad \frac{\partial u/\partial f_{t+1}}{\partial u/\partial c_{t+1}} \begin{cases} \geq w_{t+1} & \text{if } \quad f_{t+1} = 1 \\ = w_{t+1} & \text{if } 0 < f_{t+1} < 1 \\ \leq w_{t+1} & \text{if } \quad f_{t+1} = 0 \end{cases}$$

Substituting for $c^{(1)}_t$, $c^{(2)}_{t+1}$ and $f^{(2)}_{t+1}$ in terms of s_t, w_t, w_{t+1} and r_{t+1} implies the following saving function:

$$(2.1.12.)\quad s_t = s_t(w_t(1-b_t),\ w_{t+1},\ r_{t+1},\ P_{t+1}) \qquad \begin{matrix} + & (<1) & - & +/- & - \end{matrix}$$

Here, as in the following equations, the signs above the particular arguments indicate whether the effect of an increase in the accompanying variable affects the respective function positively (+), negatively (–), or in an ambiguous fashion (+/–) as long as the relevant inequality constraints are inactive.

The young- and old-age consumption functions implied by equations (2.1.6.) to (2.1.9.) are

$$(2.1.13.)\quad c^{(1)}_t = c_t(w_t(1-b_t),\ w_{t+1},\ r_{t+1},\ P_{t+1}) \qquad \begin{matrix} + & (<1) & + & +/- & + \end{matrix}$$

and

$$(2.1.14.) \qquad c^{(2)}_{t+1} = c_{t+1}(w_t(1-b_t),\ w_{t+1},\ r_{t+1},\ P_{t+1}) \qquad (+,\ +/-,\ +,\ +)$$

while the duration of a representative individual's retirement period is

$$(2.1.15.) \qquad f^{(2)}_{t+1} = f_{t+1}(w_t(1-b_t),\ w_{t+1},\ r_{t+1},\ P_{t+1}) \qquad (+,\ +/-,\ +,\ +)$$

The effect of an increase in the real interest rate, r_{t+1}, on saving and young-age consumption is ambiguous because the income and substitution effects work in opposite directions. The same is true with regard to the effect of a change in w_t on the labour supply of the elderly in the same period.

The total supply of labour L_t in the model economy consists of the number of young individuals $N^{(1)}_t$ and the amount of labour supplied by the elderly:

$$(2.1.16.) \qquad \begin{aligned} L_t &= N^{(1)}_t + [1 - f^{(2)}_t(\cdot)]\, N^{(2)}_t \\ &= [2 + n - f^{(2)}_t(\cdot)]\, N^{(1)}_{t-1} \end{aligned}$$

This equation will be referred to when the equilibrium capital/labour-ratio is derived.

2.2 *The Production Sector.* – As in the conventional overlapping generations model with production, output is assumed to be supplied by a large number of firms acting under the conditions of perfect competition on goods and factor markets.

The production function of the model economy,

$$(2.2.1.) \qquad Q_t = Q(K_t, L_t)$$

is homogenous of degree one. Thus, output per unit of labour can be written as

$$(2.2.2.) \qquad \frac{Q_t}{L_t} =: q_t = Q\left(\frac{K_t}{L_t}, 1\right) \equiv q(k_t)$$

In addition to that, it is assumed that (2.2.2.) satisfies the Inada conditions,

(2.2.3.) $$q(0) = 0$$

(2.2.4.) $$q'(0) = \infty \,, \quad \text{and}$$

(2.2.5.) $$q'(\infty) = 0$$

and that depreciation is already implicitly accounted for in equation (2.2.1.).

Firms rent capital to the point where its marginal product equals its rental rate

(2.2.6.) $$r_t = q'(k_t)$$

and pay their workers a wage rate corresponding to the marginal product of labour. The production function being linear-homogeneous implies

(2.2.7.) $$\begin{aligned} w_t &= q(k_t) - r_t k_t \\ &= q(k_t) - q'(k_t) k_t \end{aligned}$$

Differentiating (2.2.6.) and (2.2.7.) with respect to k_t yields

(2.2.8.) $$\frac{dr_t}{dk_t} = q''(k_t)$$

and

(2.2.9.) $$\frac{dw_t}{dk_t} = -q''(k_t)\, k_t$$

which will be of importance for the discussion of possible steady-state equilibria.

2.3. *The Pension System.* – In the model under consideration, public old-age pensions are supposed to be financed on a pay-as-you-go basis. Social security contributions are paid only by the members of the young generation, and the possibility of government budget surpluses or deficits is assumed to be absent. If the government's only activity is to provide social security, its budget constraint can therefore be summarized as follows:

(2.3.1.) $$N_t^{(2)} P_t = w_t b_t N_t^{(1)}$$

(This follows Homburg, 1988, pp. 18-19). Under the premise which was introduced in equation (2.1.1.) about the age structure of the population, division of (2.3.1.) by $N_t^{(1)}$ yields

(2.3.2.) $$P_t = w_t b_t (1 + n)$$

In order to analyze the effect of an unfunded social security system on individual savings and on the labour supply behaviour of the elderly for given wages and interest rates, a situation without social security (namely $p_t = b_t = 0$) has to be compared to one with positive pension benefits and contribution rates that fulfil the constraint (2.3.2.).

As can be seen from equation (2.1.12.), the effect of the introduction of a pension system on individual savings is, *ceteris paribus*, unambiguously negative.

The way in which an individual's retirement behaviour is influenced by the introduction of an unfunded pension system depends on the present discounted value of the net transfers (= pension benefits minus taxes) the individual receives from the government. Seen from the perspective of a member of the young generation, this present discounted value can be expressed as follows:

$$(2.3.3.) \qquad PVT := \frac{p_{t+1}}{1 + r_{t+1}} - b_t w_t$$

$$= \frac{w_{t+1} b_{t+1} (1 + n)}{1 + r_{t+1}} - b_t w_t$$

If the wage rate w and the contribution rate b were time-invariant, equation (2.3.3.) would simplify to

$$(2.3.4.) \qquad PVT \Big|_{\substack{w_t = w_{t+1} \\ b_t = b_{t+1}}} = \frac{n - r}{1 + r}$$

Due to the positive effect that a rise transfer income has on consumption as well as on the demand for leisure, the introduction of an unfunded pension system results in earlier retirement if PVT is positive. Hence, if wages and interest rates are assumed to be time-invariant, the way in which retirement behaviour is affected by the introduction of a pension system depends on whether the real interest rate exceeds or falls short of the current growth rate of the population.

In this section, the attention has been limited to what might be called a "partial equilibrium" analysis of the effects of an unfunded pension system. Yet it would be (at least potentially) misleading to draw far-reaching conclusions as to the macroeconomic effects of social security from such an approach. Therefore, the interaction between the modalities of the pension system on the one hand and macroeconomic variables like wages and interest rates on the other will be dealt with in the following section.

3. *The Case of a Closed Economy*

3.1. *Equilibrium Behaviour.* – In a closed economy, the goods market is in equilibrium if net investment $(K_{t+1} - K_t)$ equals net saving. Within the framework of this model, however, net saving is defined as the difference between the savings of the young $(N_t^{(1)} s_t)$ and the dissaving of the old (K_t). Therefore, the aggregate capital stock is determined as follows:

$$K_t = s_{t-1} N_{t-1}^{(1)} \tag{3.1.1.}$$

Dividing both sides of (3.1.1.) by L_t yields, after combination with equation (2.1.16),

$$k_{t+1} = \frac{s_t(w_t(1-b_t),\ w_{t+1},\ r_{t+1},\ P_{t+1})}{2 + n - f_{t+1}(w_t(1-b_t),\ w_{t+1},\ P_{t+1},\ r_{t+1})} \tag{3.1.2.}$$

Together with equations (2.2.6.) and (2.2.7.), which can be understood as equilibrium conditions for the factor markets, equation (3.1.2.) describes a non-linear relationship between the current capital-labour ratio k_t and the one of the following period k_{t+1}.

Yet in spite of the inequality constraints imposed upon $f_{t+1}^{(2)}$ and s_t, it cannot be taken for granted that for every positive value of k_t there exists only one level of k_{t+1} that satisfies the condition (3.1.2.) and can thus be termed a self-fulfilling expectation. Following Galor and Ryder (1989), it can be stated that a sufficient condition for uniqueness of k_{t+1} is that the derivative of the right-hand side of equation (3.1.2.) with respect to k_{t+1} is everywhere less than unity. By virtue of equations (2.2.6.), (2.2.7.) and (2.3.2.), the first-order derivatives of s_t and $f_{t+1}^{(2)}$ with respect to k_{t+1} are as follows:

$$\frac{\partial s_t}{\partial k_{t+1}} = \left(-\frac{\partial s_t}{\partial w_{t+1}} k_{t+1} + \frac{\partial s_t}{\partial r_{t+1}}\right) q''(k_{t+1}) - \frac{\partial s_t}{\partial P_{t+1}} b_{t+1}(1+n)\, q''(k_{t+1})\, k_{t+1} \tag{3.1.3.}$$

$$\frac{\partial f_{t+1}}{\partial k_{t+1}} = \left(-\frac{\partial f_{t+1}^{(2)}}{\partial w_{t+1}} k_{t+1} + \frac{\partial f_{t+1}^{(2)}}{\partial r_{t+1}}\right) q''(k_{t+1}) + \frac{\partial f_{t+1}^{(2)}}{\partial P_{t+1}} b_{t+1} n\, q''(k_{t+1})\, k_{t+1} \tag{3.1.4.}$$

Using these two equations, one obtains the following expression for the above-mentioned condition for uniqueness:

$$(3.1.5.) \qquad \frac{(2 + n - f_{t+1})\,(\partial s_t/\partial k_{t+1}) + s_t\,(\partial f^{(2)}_{t+1}/\partial k_{t+1})}{(2 + n - f^{(2)}_{t+1})^2} < 1$$

It will be assumed throughout this chapter that this condition holds.

However, unless even further restrictions regarding the underlying utility and production functions are introduced, (3.1.2.) does not guarantee that there exists a positively-valued steady state level $\bar{k}$ of the capital-labour ratio.

Technically spoken, there is at least one steady-state equilibrium with a positive capital stock if the function

$$(3.1.6.) \qquad \Phi(\bar{k}) := \frac{s_t[(1-b_t)\,w_t(\bar{k}),\, r_{t+1}(\bar{k}),\, w_{t+1}(\bar{k}),\, p_{t+1})]}{2 + n - f_{t+1}[(1-b_t)\,w_t(\bar{k}),\, r_{t+1}(\bar{k}),\, w_{t+1}(\bar{k}),\, p_{t+1})]}$$

possesses a positively-valued *fixed point* (i.e. point $\bar{k}^* \in \mathcal{R}^+$ such that $\Phi(\bar{k}^*) = \bar{k}^*$. According to Brouwer's fixed point theorem (see, e.g., Berck, Sydsaeter, 1991, p. 24), this is the case if there exists a closed interval $[a, b] \in \mathcal{R}^+$ for which the following condition is fulfilled:

$$(3.1.7.) \qquad \forall\ \bar{k} \in [a, b] \Rightarrow \Phi(\bar{k}) \in [a, b]$$

Nevertheless, if condition (3.1.4.) holds, a given steady state value of k is not necessarily unique. Besides that, it cannot be taken for granted that k_t moves towards such a fixed point as t increases. In this context, it is useful to recur to the concept of asymptotic stability of a steady state equilibrium (this follows Luenberger, 1979, pp. 322-324):

A steady state equilibrium point $\bar{k}$ is *locally asymptotically stable* if there exists an interval $I := [c, d]$ such that whenever k_0 (= the initial value of k) lies inside that interval, k_t tends to $\bar{k}$ as t increases. It is *globally asymptotically stable* if k_t tends to $\bar{k}$ over time for any value of k_0.

A steady-state level $\bar{k}$ of the capital-labour ratio is locally asymptotically stable if the first-order derivative of k_{t+1}, with respect to k_t, evaluated at $\bar{k}$, is less than unity in absolute value. A sufficient (but not necessary) condition for a unique and globally asymptotically stable steady state to exist is that $\partial k_{t+1}/\partial k_t$ evaluated at $k_t = 0$, exceeds unity and that the condition $\partial^2 k_{t+1}/\partial k_t^2 < 0$ holds for all $k_t \in \mathcal{R}^+$.

It follows from equations (2.2.6.) and (2.2.7.) that the first-order derivatives of s_t and $f^{(2)}_{t+1}$ with respect to k_t and k_{t+1} can be expressed as follows:

$$\frac{\partial s_t}{\partial k_t} = -(1-b_t)\frac{\partial s_t}{\partial w_t^{(n)}} q''(k_t)\, k_t\, (> 0) \tag{3.1.8.}$$

$$\frac{\partial f^{(2)}_{t+1}}{\partial k_t} = -(1-b)\frac{\partial f^{(2)}_{t+1}}{\partial w_t^{(n)}} q''(k_t)\, k_t\, (\geq 0) \tag{3.1.9.}$$

(Here, $w_t^{(n)}$ denotes the net wage rate). Using these four equations, the first-order derivative of k_{t+1} with respect to k_t becomes

$$\frac{\partial k_{t+1}}{\partial k_t} = \frac{[2 + n - f^{(2)}_{t+1}(\cdot)]\frac{\partial s_t}{\partial k_t} + s_t(\cdot)\frac{\partial f^{(2)}_{t+1}}{\partial k_t}}{[2 + n - f^{(2)}_{t+1}(\cdot)]^2 - \frac{\partial s_t}{\partial k_{t+1}}[2 + n - f^{(2)}_{t+1}(\cdot)] - s_t(\cdot)\frac{\partial f^{(2)}_{t+1}}{\partial k_{t+1}}} \tag{3.1.10}$$

The numerator of equation (3.1.10) is positive. However, the sign of the denominator is indeterminate because both the effect of w_{t+1} on the retirement age and the impact of the real interest rate on saving are (at least potentially) ambiguous. This leads to the conclusion that, if the above-mentioned condition for local asymptotic stability is satisfied and the system is initiated at a point k_0 sufficiently close to a steady-state value $\bar{k}$ of the capital-labour ratio, k_t can, but need not move toward $\bar{k}$ in an oscillatory fashion as t increases. If $\partial k_{t+1}/\partial k_t$ lies between 0 and 1 when being evaluated around $\bar{k}$, the dynamic adjustment toward the steady state is non-oscillatory, whereas the opposite is true if, for values of k_t that are close to $\bar{k}$, the condition $-1 < (\partial k_{t+1}/\partial k_t) < 0$ holds.

In the remainder of this section, the assumption is made that a unique positive steady-state level of the capital-labour ratio exists and that it is globally asymptotically stable.

3.2. *Efficiency Considerations.* – Up to this point, the analysis has dealt with the question how individual decisions of households (on consumption, saving, and labour supply) and firms (on production, investment, and the demand for labour) interact in determining economic growth. The efficiency properties of such a decentralized allocation process can be assessed by comparing its outcome to the one that would be chosen by an omniscient

and benevolent central planner who maximizes a specific social welfare function with a given initial capital stock $K_0 > 0$ being at his disposal at the beginning of the planning period. In the following, it will be assumed that the social welfare function V is defined as the unweighted sum of the utility levels of the present and all future generations. Together with the resource constraint faced by the central planner and the inequality constraints on old-age leisure, the corresponding optimization problem can be formalized as follows:

$$\text{(3.2.1.)} \qquad \max.\ \bar{V} = \frac{1}{1+\rho}\, u\,(c_0^{(2)}, f_0^{(2)})$$

$$+ \sum_{t=0}^{\infty} u\,(c_t^{(1)}, 0) + \frac{1}{1+\rho}\, u\,(c_{t+1}^{(2)}, f_{t+1}^{(2)})$$

$$+ \gamma_t \{Q\,(K_t, [2 + n + f_t^{(2)}]\, N_{t-1}^{(1)})$$

$$- [(1+n)\ c_t^{(1)} + c_t^{(2)}]\, N_{t-1}^{(1)} - K_{t+1} + K_t\}$$

$$+ \eta_{1t} f_{t+1}^{(2)} + \eta_{2t}\,(1 - f_{t+1}^{(2)})$$

This equation defines an infinite-horizon optimal control problem in discrete time. Differentiation with respect to $c_t^{(1)}$, $c_t^{(2)}$, $f_t^{(2)}$ and K_t yields the following optimality conditions:

$$\text{(3.2.2.)} \qquad \frac{\partial \bar{V}}{\partial c_t^{(1)}} = \frac{\partial u^{(1)}}{\partial c_t^{(1)}} - \gamma_t\,(1+n)\, N_{t-1}^{(1)} = 0$$

$$\text{(3.2.3.)} \qquad \frac{\partial \bar{V}}{\partial c_t^{(2)}} = \frac{1}{1+\rho}\,\frac{\partial u^{(2)}}{\partial c_t^{(2)}} - \gamma_t N_{t-1}^{(1)} = 0$$

$$\text{(3.2.4.)} \qquad \frac{\partial \bar{V}}{\partial f_t^{(2)}} = \frac{1}{1+\rho}\,\frac{\partial u^{(2)}}{\partial f_t^{(2)}} - \gamma_t \frac{\partial Q}{\partial L_t}\, N_{t-1}^{(1)} + \eta_1 - \eta_2 = 0$$

and

$$\text{(3.2.5.)} \qquad \frac{\partial \bar{V}}{\partial K_t} = \gamma_t \left(1 + \frac{\partial Q}{\partial K_t}\right) - \gamma_{t-1} = 0$$

It follows from equations (3.2.2.) to (3.2.4.) that the ficticious central planner allocates consumption and leisure within an individual's lifetime exactly as the individual himself would do. This is due to the fact that the

social welfare function considered here is simply defined as the sum of individual utility levels.

Finding a unique solution on the problem of maximizing V is, however, rendered impossible by the fact that the rate at which the utility levels of future generations are discounted is implicitly assumed to be zero and that, as a consequence, the sum (3.2.1.) does not converge. Therefore, a different criterion is necessary in order to identify the optimal path of k over time. According to Seierstad and Sydsaeter (1987), the most comprehensive among the applicable optimality criteria is the one of "overtaking optimality". In the context of the model under discussion, it can be characterized as follows:

A feasible path $\{\hat{c}_t^{(1)}, \hat{c}_{t+1}^{(2)}, \hat{f}_{t+1}^{(2)}, \hat{k}_t\}$, $t \in N \cup \{0\}$, overtakes the path $\{\bar{c}_t^{(1)}, \bar{c}_{t+1}^{(2)}, \bar{f}_{t+1}^{(2)}, \bar{k}_t\}$ if there exists a time T_0 such that for all $T \geq T_0$ the inequality

$$u(\hat{c}_0^{(2)}, \hat{f}_0^{(2)}) - u(\bar{c}_0^{(2)}, \bar{f}_0^{(2)}) + \sum_{t=1}^{T} u(\hat{c}_t^{(1)}, 0) - u(\bar{c}_t^{(1)}, 0)$$

$$+ \sum_{t=1}^{T} \frac{1}{1+\rho} \{u(\hat{c}_{t+1}^{(2)}, \hat{f}_{t+1}^{(2)}) - u(\bar{c}_{t+1}^{(2)}, \bar{f}_{t+1}^{(2)})\} \geq 0$$

holds.

In order to apply the overtaking principle to the optimization problem (3.2.1.), it is useful first to find out for which steady-state level $\bar{k}$ of the capital-labour ratio the corresponding values of individual consumption $(\bar{c}^{(1)}, \bar{c}^{(2)})$ and old-age leisure $(\bar{f}^{(2)})$ yield the highest possible value of individual utility. It becomes obvious from the optimality conditions set up above that the optimal value (k^*) of the steady state capital-labour ratio is the one that satisfies the condition

$$q'(k^*) = n \tag{3.2.6.}$$

which is known as the "golden rule" of capital accumulation.

This leads to the conclusion that any sequence

$$\{\bar{k}_t\} := \{\bar{k}_0, \bar{k}_1, \bar{k}_2, \ldots\}$$

of values of the capital-labour ratio that does not converge to the "golden rule" level k^* of k as t goes to infinity cannot be optimal if assessed on the basis of the overtaking criterion.

Yet unless the initial level K_0 of the aggregate capital stock is such that the capital-labour ratio at time zero can be equated to k^* by choosing

an appropriate value of $f_0^{(2)} \in [0, 1]$, this condition does not suffice to determine the optimal path of the capital-labour ratio over time. This problem can, however, be solved by generalizing a suggestion Samuelson (1968) made for the derivation of optimal accumulation paths in the conventional overlapping generations model to fit the optimization problem considered here.

In the following, $c^{(1)*}, c^{(2)*}$ and $f^{(2)*}$ denote the values of consumption and old-age leisure associated with the "golden rule" level of the capital-labour ratio. With these definitions being introduced, maximizing social welfare as defined in equation (3.2.1.) in the sense of the overtaking criterion can be seen as equivalent to minimizing the sum

$$
\begin{aligned}
(3.2.7.) \qquad W :&= (V^* - \bar{V}) \\
&= \frac{1}{1+\rho}\left[u\left(c_0^{(2)*}, f_0^{(2)*}\right) - u\left(c_0^{(2)}, f_0^{(2)}\right)\right] \\
&+ \sum_{t=0}^{\infty} \{u\left(c_t^{(1)*}, 0\right) - u\left(c_t^{(1)}, 0\right) \\
&+ \frac{1}{1+\rho}\left[u\left(c_{t+1}^{(2)*}, f_{t+1}^{(2)*}\right) - u\left(c_{t+1}^{(2)}, f_{t+1}^{(2)}\right)\right]\} \\
&+ \gamma_t \{Q\left(K_t, [2 + n + f_t^{(2)}]\, N_{t-1}^{(1)}\right) \\
&- [(1+n)\, c_t^{(1)} + c_t^{(2)}]\, N_{t-1}^{(1)} - K_{t+1} + K_t\} \\
&+ \eta_{1t} f_{t+1}^{(2)} + \eta_{2t}\left(1 - f_{t+1}^{(2)}\right)
\end{aligned}
$$

In this case, there is a unique solution to the corresponding optimization problem in the form of a sequence of values k_t of the capital labour ratio that converges for $t \to \infty$ to the golden rule level of the capital-labour ratio.

The notion of dynamic efficiency, as it is described here by means of a hypothetical social welfare function, is not identical to the concept of intergenerational Pareto efficiency. Clearly, a time path of k that converges to a steady-state level $\bar{k} \neq k^*$ over time cannot be regarded as dynamically efficient in the former sense. It is, however, Pareto inefficient only if it is possible to find a sequence

$$\bar{\phi} := \{\bar{\phi}_0, \bar{\phi}_1, \bar{\phi}_2, \bar{\phi}_3, \ldots\}$$

of non-negative transfer payments from the young to the old that improves the utility level of at least one generation without worsening the lot of any

of the others. (Such intergenerational transfer payments can be interpreted as "lump-sum" contributions to an unfunded pension system).

In the conventional overlapping generations model, where the amount of labour supplied by the elderly is set to an exogenously determined constant, the existence of such a Pareto-improving transfer system can be ruled out if there is a unique, stable and non-oscillatory steady-state of which the corresponding capital-labour ratio falls short of the golden rule level (see Breyer, 1989). This result, however, cannot be extended to the model discussed here. The reason is that if the present discounted value of the net transfer payments a representative member of the young generation at time t_0 receives is positive (as it is the case if an unfunded pension system is to be introduced at time $t_0 + 1$), the individual will save less and retire earlier. The first of these two reactions reduces the supply of capital, whereas the second one results in a lower supply of labour. The joint effect of these two reactions on the capital-labour ratio in $t_0 + 1$ can thus be either positive or negative. If it is positive, it entails a higher wage rate in $t_0 + 1$ which, in turn, might suffice to make compensation for the transfer payments the young of that period have to pay their older contemporaries.

These considerations are also relevant for an assessment of the welfare properties of the assumed public pension system. Social security affects the development of the capital-labour ratio in two ways: on the one hand, it follows from the savings function (2.1.12.) and the equation that describes the link between social security taxes and benefits, (2.3.2.), that the effect of social security on saving is unambiguously negative. The effect of a permanent marginal increase in the contribution rate b on saving is

$$(3.2.8.) \qquad \frac{\partial s_t}{\partial b} = \frac{\partial s_t}{\partial w_t^{(n)}} [q(k_t) - q'(k_t)\, k_t] - \frac{\partial s_t}{\partial P_{t+1}} (1 + n) [q(k_{t+1}) - q'(k_{t+1})\, k_{t+1}] < 0$$

On the other hand, it depends on the present discounted value of pensions minus taxes whether such a system raises or lowers an individual's retirement age. The effect of an infinitesimally small but permanent increase in b on old-age leisure can be expressed as follows:

$$(3.2.9.) \qquad \frac{\partial f_{t+1}^{(2)}}{\partial b} = \frac{\partial f_{t+1}^{(2)}}{\partial w_t^{(n)}} [q(k_t) - q'(k_t)\, k_t] - \frac{\partial f_{t+1}^{(2)}}{\partial P_{t+1}} (1 + n) [q(k_{t+1}) - q'(k_{t+1})\, k_{t+1}] \gtrless 0$$

The present discounted value of the net benefits, in turn, does not only depend on the parameters of the pension system and the growth rate of the population, but also on wages (w_t, w_{t+1}) and on the real interest rate. It therefore varies over time even if p, b, and n are constant.

The preceding two equations indicate that a marginal increase in the contribution rate affects the capital-labour ratio as follows

$$(3.2.10.) \qquad \frac{\partial k_{t+1}}{\partial b} = \frac{[2 + n - f^{(2)}_{t+1}(\cdot)] \frac{\partial s_t}{\partial b} + s_t(\cdot) \frac{\partial f^{(2)}_{t+1}}{\partial b}}{[2 + n - f^{(2)}_{t+1}(\cdot)]^2 - \frac{\partial s_t}{\partial k_{t+1}}[2 + n - f^{(2)}_{t+1}(\cdot)] - s_t(\cdot) \frac{\partial f^{(2)}_{t+1}}{\partial k_{t+1}}}$$

It is obvious that $\partial k_{t+1}/\partial b$ is of ambiguous sign. This statement would remain true even if the sign of the denominator were known. Therefore, it cannot be told at this level of generality whether an unfunded public pension system can be efficiency-enhancing in the sense of moving the steady-state value of k closer to the golden rule level.

3.3. *Consequences of Population Ageing.* – Within the framework of this model, the macroeconomic implications of a changing age structure of the population can be assessed by evaluating the effect that a variation in the assumed rate of population growth, n, has on the steady-state value of the capital-labour ratio. Among the variety of possible policy reactions to such a change, there are two extreme cases that deserve particular attention: On the one hand, there is the possibility that the social security tax rate, b, is held constant, while the per capita pension level, p, is allowed to vary along with wages and population growth as described in equation (2.3.2.). But on the other, p can remain unchanged if the contribution rate is being adjusted accordingly over time.

If wages and interest rates could be regarded as constant over time, a decrease in population growth would, *ceteris paribus*, lead to higher per capita savings and to an increase in a representative individual's retirement age if a pay-as-you-go pension system with a constant contribution rate is in place. This conclusion can be drawn from the following two equations:

$$(3.3.1.) \qquad \left.\frac{\partial s_t}{\partial n}\right|_{(b=\bar{b}>0)} = \frac{\partial s_t}{\partial p_{t+1}} w_{t+1} b \quad (< 0)$$

$$(3.3.2.) \qquad \left.\frac{\partial f_{t+1}}{\partial n}\right|_{(b=\bar{b}>0)} = \frac{\partial f_{t+1}}{\partial p_{t+1}} w_{t+1} b \quad (>0)$$

In the case of a time-invariant pension level, a decline in n would affect savings negatively and cause individuals to retire later if w and r were assumed to be constant:

$$(3.3.3.) \qquad \left.\frac{\partial s_t}{\partial n}\right|_{(p=\bar{p}>0)} = \frac{\partial s_t}{\partial w_t^{(n)}} \frac{\bar{p}}{(1+n)^2} \quad (>0)$$

$$(3.3.4.) \qquad \left.\frac{\partial f_{t+1}}{\partial n}\right|_{(p=\bar{p}>0)} = \frac{\partial f_{t+1}}{\partial w_t^{(n)}} \frac{\bar{p}}{(1+n)^2} \quad (>0)$$

Being deduced from a "partial equilibrium" approach, these statements may be useful for an assessment of the short-run consequences of demographic changes; yet they are at least potentially misleading when its long-run effects on capital formation and the aggregate supply of labour are to be estimated. The reason is that because assets are assumed to be exclusively held by the members of the old generation, any change in its relative size ($\equiv (1+n)^{-1}$) would produce a tendency towards a higher capital-labour ratio if per capita savings and old-age leisure remained unchanged. Thus, even if the possibility of an oscillatory steady-state is ruled out by assumption, the effect of population ageing on the long-run equilibrium value of the capital-labour ratio is ambiguous as long as a pay-as-you-go pension system is present. This can be shown by evaluating the first-order derivative of k_{t+1} with respect to n:

$$(3.3.5.) \qquad \frac{\partial k_{t+1}}{\partial n} =$$

$$= \frac{[2+n-f_{t+1}^{(2)}(\cdot)]\frac{\partial s_t}{\partial n} + s_t(\cdot)\left[\frac{\partial f_{t+1}^{(2)}}{\partial n} - 1\right]}{[2+n-f_{t+1}^{(2)}(\cdot)]^2 - \frac{\partial s_t}{\partial k_{t+1}}[2+n-f_{t+1}^{(2)}(\cdot)] - s_t(\cdot)\frac{\partial f_{t+1}^{(2)}}{\partial k_{t+1}}}$$

It becomes obvious from equations (3.3.1.) to (3.3.4.) that, in the absence of social security ($b = p = 0$), a marginal decline in n raises the steady-state capital labour ratio if the denominator of (3.3.5.) is positive. If, however, an unfunded pension system is present, this purely demographic impact is accompanied (and can possibly even be, at least partially, compen-

sated) by the effects of additional policy-specific labour and saving (dis)incentives.

One of the main conclusions that can be drawn from this approach with regard to policy issues is that population ageing need not lead to exorbitant rises in social security contribution rates as long as pension benefits are not raised along with (gross) wages over time, but kept at a constant absolute level instead. In this case, the increased scarcity of labour brought about by the decline of population growth can be expected to raise gross wage rates and thus ease at least part of the additional financial burden levied upon public pension systems by the decline in population growth. If, however, the voting power of the old generation within the political decision-making process increases with its relative size, such a "solution" is not very likely to be chosen.

4. *Concluding Remarks*

Being an extension of the conventional overlapping generations model, the approach described in this paper is an attempt to explain the way in which individual decisions on consumption, saving, and labour supply interact with macroeconomic variables and the parameters of the assumed public pension system in determining economic growth. The main result is that because the introduction of an unfunded social security system affects both per capita savings and the labour supply behaviour of the old generation, its effect on the steady-state capital-labour ratio is ambiguous even if the possibility of a stable but oscillatory steady-state equilibrium is excluded by assumption. As a consequence, the verdict that a pay-as-you-go pension system cannot be welfare-improving unless the economy under consideration exhibits a tendency toward capital overaccumulation in the absence of social security need not be true under the assumptions made here.

REFERENCES

BERCK Peter and SYDSAETER Knut, *Economists' Mathematical Manual*, Berlin/Heidelberg: Springer, 1991.

BREYER Friedrich, *Ökonomische Theorie der Alterssicherung*, Munich: Vahlen, 1990.

——, "On the Intergenerational Pareto Efficiency of Pay-as-you-go Financed Pension Systems", *Journal of Institutional and Theoretical Economics*, 1989, *145*, 643-58.

—— and Straub Martin, "Welfare Effects of Unfunded Pension Systems when Labor Supply is Endogenous", *Journal of Public Economics*, No. 1, 1993, *53*, 77-91.

Diamond Peter A., "National Debt in a Neoclassical Growth Model", *American Economic Review*, No. 5, 1965, *55*, 1126-50.

Galor Oded and Ryder Harl E., "Existence, Uniqueness, and Stability of Equilibrium in an Overlapping-Generations Model with Productive Capital", *Journal of Economic Theory*, 1989, *49*, 360-75.

Homburg Stefan, *Theorie der Alterssicherung*, Berlin/Heidelberg: Springer, 1988.

Hu Sheng C., "Social Security, the Supply of Labor, and Capital Accumulation", *American Economic Review*, No. 3, 1979, *69*, 274-83.

Jacobs K., Kohli M., and Rein M., "Evolution of Early Exit: A Comparative Analysis of the Labour Force Participation of the Elderly", in M. Kohli et al., eds., *Time for Retirement. Comparative Studies of the Decreasing Age of Exit from the Labor Force*, Cambridge/New York: Cambridge University Press, 1988.

Luenberger David G., *Dynamic Systems. Theory, Models, and Applications*, New York: Wiley, 1979.

Reimann A., "Trend zur Frühverrentung noch ungebrochen", *Die Angestelltenversicherung*, 1985, *32*, 406-13.

Samuelson Paul A., "An Exact Consumption-Loan Model of Interest with or without the Social Contrivance of Money", *Journal of Political Economy*, No. 6, 1958, *66*, 467-82.

——, "Optimum Social Security in a Life-Cycle Growth Model", *International Economic Review*, No. 3, 1975, *16*, 539-44.

Seierstad Atle, and Sydsaeter Knut, *Optimal Control Theory with Economic Applications*, Amsterdam/New York: Elsevier, 1987.

PREVIDENZA, RISPARMIO E PENSIONAMENTI IN UN MODELLO DI CRESCITA ECONOMICA CON L'IPOTESI DI CICLO DELLA VITA

In questo articolo il modello standard a generazioni sovrapposte con capitale produttivo viene esteso a uno in cui l'offerta di lavoro degli anziani è determinata endogenamente. Entro questo schema, vengono analizzate le implicazioni di benessere di un sistema pensionistico a ripartizione e le conseguenze economiche di una diminuzione nella crescita demografica. Si mostra che l'aumento dei tassi di contribuzione previdenziale causati dal processo di invecchiamento della popolazione in presenza di un sistema pensionistico a ripartizione può essere limitato se le pensioni non aumentano in proporzione ai saggi salariali.

Rivista Internazionale di Scienze Economiche e Commerciali
Volume 42 (1995), N. 1-2, 19-31

HOW AND WHEN UNANIMITY IS A SUPERIOR DECISION RULE

by
MARIO S. CATALANI * and GIUSEPPE F. CLERICO **



1. *Introduction*

The kind of problems we are going to deal with can be synthetically described as follows.

a) Any body whose task is to take decisions among alternatives may follow different procedural rules: for example, simple majority, special (or qualified) majority, unanimity.

b) The decision process involves different types of uncertainty: the state of the world, the human fallibility, the asymmetry of information.

c) Hence the outcome of each decision process will be affected by a cost associated with the taking of a wrong decision (*error cost*). To this cost it must be added a cost associated with the specific procedural rule in force (*specific cost*). Then the *total cost* associated with each decision is equal to the sum of these two costs.

d) One way to formally introduce uncertainty is to probabilistically model the behavior of each member of our body. More precisely, each individual has a definite probability of making the correct decision. This probability depends on his/her skill. We must emphasize that this probabilistic setup, and even the analytical development, borrows the approach taken by Sah and Stiglitz (1985, 1986, 1988a, 1988b) and Sah (1991) to tackle a related problem.

In their paper, they "... study the decision making of committees and

* Department of Economics, University of Torino, Italy.

** Department of Economics, University of Cagliari, and Department of Economics, University of Torino, Italy.

We are grateful to an anonymous referee for helpful comments on an earlier draft. The usual caveat applies.

contrast this to the decision making of certain stylized forms of centralized versus decentralized organizations ..." (Sah and Stiglitz, 1988b, p. 451). The centralized organizations are identified with hierarchies and the decentralized ones with polyarchies.

A fairly comprehensive survey of the literature on these problems may be found in Catalani and Clerico (1994). The same authors use this approach to study two specific problems: *i.* Team cooperation *versus* independent assessment (1993a); *ii.* Hierarchies as pyramid structures (1993b).

e) If the foregoing assumptions are accepted, it seems worthwhile to ask if and how different procedural rules affect the decision costs, so as to ascertain whether conditions exist under which one procedural rule is better than the others. More precisely, the question is whether and when a procedural rule exists which minimizes the total cost of each decision.

The present paper aims at formalizing this kind of problems and to shed some light on the existence of the conditions mentioned sub *e*).

The paper is organized as follows. First of all, we use some suggestions from the literature to stylize the problem. Secondly, we present our model. Finally we discuss the main results. The proofs are shown in the Appendix.

2. *Some Insights from the Literature*

We consider a group of persons (say n) whose task is to reach a decision about a proposed project. Our stylized model assumes that the decision is of a dichotomous type: that is, either to accept or reject the project. Abstention is not allowed. The project is selected randomly from a portfolio of projects containing a proportion of $\pi\%$ of good projects. The persons, as human beings, are fallible. That means that they can incur in one of two types of errors: type *I* error consisting in rejecting a good project, and type *II* error consisting in accepting a bad project. Generally the probability of the type *I* error and the probability of the type *II* error vary across the members of the decision group: let us denote with $1-p_i$ and q_i, respectively, these two probabilities for the i-th member. The complement to one of these probabilities may be considered as a measure of individual skills. We assume that there is a cost c_1 arising from rejecting a good project and a cost c_2 arising from accepting a bad project. In the case of a correct decision we set the cost equal to zero. In this way we can construct a payoff matrix associated with this decision.

Every member of the decision group aims at minimizing the cost asso-

ciated with the decision made. Members do not pursue selfish goals; therefore we do not deal with the logrolling phenomenon.

The problem now arising is how different majority rules affect this cost. Working with a simpler framework which does not distinguish between the two types of errors and simply speaks of incorrect decisions and which deals with a symmetric payoff matrix, the literature has established a definite result: the decision rule which maximizes the probability that the group reaches the correct judgment is a majority weighted voting rule (Nitzan and Paroush, 1984; Shapley and Grofman, 1984). In the case of an asymmetric payoffs matrix, "it makes sense to slant the decision procedure ... A good corporate example may be the imposition of a significant special majority requirement in cases of 'important' decisions, such as mergers, liquidations, or charter amendments" (Nitzan and Procaccia, 1986, p. 198).

Summing up: the general framework to study this kind of problems is characterized by the following variables: the individual skills of each manager; the number (n) of members; the payoffs associated, in any *state of the world*, with the taken decision; the objective prior *likelihood* of each state of the world.

3. *Unanimity versus Simple Majority*

Let us imagine a situation in which a committee features a minority of old members and a majority of younger ones. The old members hold the committee's leadership due to their past experience. When unanimity is required we assume that the leadership will impose its will. The projects under scrutiny can be ascribed to two different types: the *conventional* and the *unconventional* ones. A conventional project is a project which is homogeneous to those already examined in the past. We are interested in analysing the relative performance of two majority rules (unanimity versus simple majority) both when the project is conventional and unconventional. The performance is measured by the expected total cost associated with the decision taken. Two items enter the total cost: *i*) the costs related to the two types of errors (as described in Section 2) which may be called the *error cost*; *ii*) the cost related to the specific majority rule in force, that is the *specific cost*.

We assume that the skills of the members of the committee are different between the two groups (old versus young members) but homogeneous within each group. In our model we will denote with the subscript *1*

the skills of the old members, and with the subscript 2 the skills of the young ones.

When the project is conventional it is reasonable to assume that the old members possess better skills: using the previous notation we can write

$$p_1 > p_2, \quad q_1 < q_2$$

On the contrary, we can imagine that younger members are better equipped to deal with a new and unconventional situation, which means

$$p_1 < p_2, \quad q_1 > q_2$$

When the committee examines a conventional project the unanimity rule entails a specific cost given by the time necessary to reach unanimity. On the other hand, the simple majority rule does not entail any specific cost.

In the case of an unconventional project the majority (young members) feels it possesses, and indeed it does possess, better skills. Consequently they feel that the leadership (represented by the old members) is just a burden and so they have an incentive to get rid of it. This implies replacing the existing decision structure with a new one. The building of this new decision structure entails a cost. The only way to avoid this situation is to constrain the decision process by the unanimity rule. In this context the simple majority rule has a specific cost (given by the collapse of the existing decision structure) which is absent when unanimity is required.

In what follows we present a model of the process through which unanimity is reached when the committee faces a conventional project. To do so we formalize the specific cost of unanimity in this situation. On the other hand, we will not formalize the specific cost of the simple majority rule in the second situation (unconventional project). Nevertheless, it is still possible to compare the dynamics of the relative performance of the two majority rules within the two contexts.

4. *The Model*

To simplify the problem we assume a decision structure (say a board) formed by three managers: so we do not consider the number of managers as parameter of our model. Consequently the wage of each manager does not enter the cost of the decision process. Let the project be a conventional one. The members are not homogeneous as there exists a member possessing a higher level of skills, while the other two are similar in this respect. Initially

each manager forms his opinion about the decision to be made independently of his peers.

When simple majority is required, since $n = 3$, a decision is immediately reached. In the case of unanimity, however, it may happen that initially the board is split. So we need to model the discussion process through which unanimity is reached.

It is reasonable to assume that the time spent in the discussion process entails a cost: this cost then characterizes this kind of majority rule. We choose to model this process with a Markov Chain with the following properties.

1) The manager possessing the highest degree of skills is able to influence the decision of the other two managers; the outcome is that the board will vote according to his opinion.

2) The members who are influenced do not influence each other and change their opinion according to a common probabilistic law represented by *transition probabilities*: these probabilities are strictly greater than zero and lower than one. Let φ be the probability to stick to the original opinion. It is reasonable to make φ dependent on the strength of the leadership.

An easy way to do this is then to write

$$\varphi = 1 - (p_1 - p_2)$$

With these assumptions certainly sooner or later unanimity is reached. But clearly the required time turns out to be a random variable. We can use the expected value of this time as a measure of the discussion cost. This cost is then the *specific cost* of unanimity. In the case of simple majority in our model the *specific cost* is equal to zero.

The *error cost* is, in both situations, equal to c_1 (if a good project is rejected) or c_2 (if a bad project is accepted). In the sequel, to make things tractable we assume that the project under discussion is a good one.

Let γ_1 and γ_2 be the *total cost* of unanimity and of simple majority, respectively. As shown in the Appendix

$$\gamma_1 = E(\tau_G) + c_1(1 - p_1)$$

where $E(\tau_G)$ is the expected time spent in reaching unanimity,

$$\gamma_2 = c_1(1 - p_2)(1 + p_2 - 2p_1p_2)$$

So we have to study the behavior of $\Delta = \gamma_1 - \gamma_2$, for changes in the parameters of the model. Clearly, $\Delta < 0$ implies that unanimity is better.

Now let $c_1^* > 0$ be the cost such that $\Delta = 0$, that is the indifference cost of the two majority rules. Then

i) if $1/2 \leq p_2 \leq 1$, and $p_1 < p_2^2/(1 - 2p_2 + 2p_2^2)$, simple majority is better;

ii) if not so, for all $c_1 > c_1^*$ unanimity is better.

This result can be summarized as follows. If $p_1, p_2 > 1/2$ and they are not very far apart, then simple majority is better. On the other hand, if it is not so, and if the error cost is sufficiently high, unanimity is preferable.

Some further insight may be gained through the following figure.

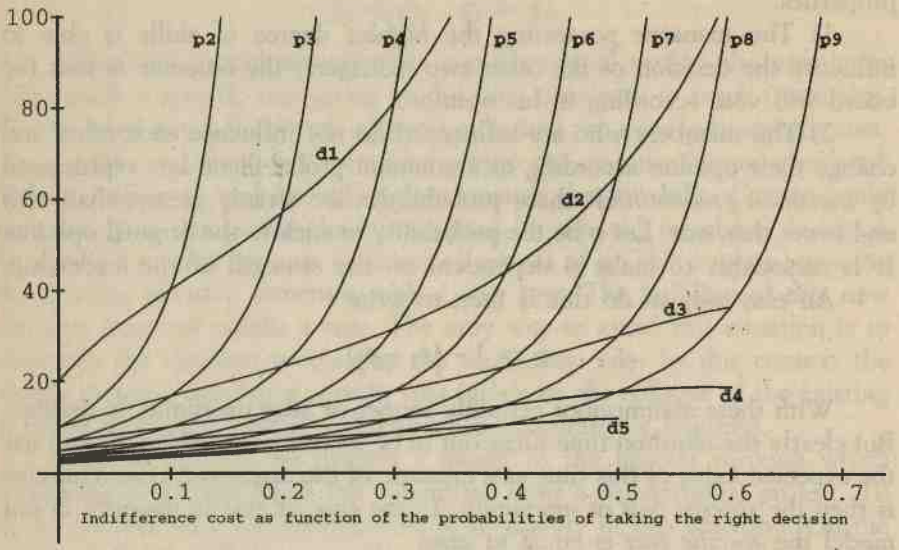

FIG. 1

On the abscissas we measure p_2, and on the ordinates the indifference value c_1^*. The curves denoted by "p*i*", $i = 2, 3, \ldots, 9$, give c_1^* as a function of p_2 for $p_1 = 0.i$. The curves denoted by "d*i*", $i = 1, 2, \ldots, 5$, represent c_1^* as a function of p_2 for

$$p_1 - p_2 = 0.i$$

Recalling that the $\{p_i\}$ are a measure of individual skills and that $(p_1 - p_2)$ is a measure of the strength of the leadership, we may conclude that

i) for a given level of leadership, c_1^* is an increasing function of p_1;

ii) for a given p_1, c_1^* is a decreasing function of the leadership level.

Now let us consider the unconventional case. We have

$$\gamma_1 = c_1 (1 - p_1)$$

$$\gamma_2 = c_1 (1 - p_2) (1 + p_2 - 2p_1 p_2) + \beta$$

where β is the cost due to the collapse of the existing committee.

Let $\hat{c}_1$ be the new indifference cost. Recalling that $p_1 < p_2$, by the same reasoning followed for the conventional project above, we get the following result:

i) if $p_2 \leq 1/2$, and $p_1 > p_2^2/(1 - 2p_2 + 2p_2^2)$, unanimity is better;

ii) if not so, for all $c_1 > \hat{c}_1$ simple majority is better.

Summing up: if p_1, $p_2 \leq 1/2$ and they are not very far apart, then unanimity is always better. Otherwise, when the error is sufficiently high, simple majority is preferable.

5. *Conclusions*

Our very simplified model confirms the intuition that decisions over relevant matters should require slanted procedures. What we have shown in particular (except for a restricted range of values of the parameters) is that the higher the *error cost*, the more likely it is that the more advantageous procedural rule is unanimity.

Let us introduce the concept of *relative leadership*, measured by $(p_1 - p_2)/p_1$. Then our model shows that the indifference value of the *error cost* (that is, the value of the error cost such that the procedural rules perform equally well) decreases as the measure of relative leadership increases.

The intuition underlying the main results of the paper in the conventional case is that, given the particular composition of the decision structure (the leadership's characteristics and role), the unanimity rule gives power to the most skillful member and therefore increases the probability of taking the right decision. When the cost of taking the wrong decision is sufficiently high, unanimity is therefore better.

In the unconventional case the results are essentially reversed. The simple majority rule gives more power to the younger members of the committee, who are the most skillful in this case, thereby increasing the probability of taking the right decision. Consequently, if the cost related to the collapse of the existing decision structure is sufficiently high, the simple majority rule turns out to be most advantageous.

APPENDIX (*by Mario S. Catalani*)

We analyze in detail the conventional case. We consider a board with three managers. When a project is scrutinized we have initially the following outcomes:

AAA
AAR
ARA
ARR
RAA
RAR
RRA
RRR

where *A* stands for acceptance of the project and *R* for its rejection. In each triple the first letter denotes the decision of the manager possessing the highest degree of skills: the remaining two letters denote the decision of the other two (homogeneous) managers. We may call each outcome a *state*.

As said in the introduction, let us label with $i = 1$ the first manager and with $i = 2, 3$, the other two. p_i is the probability that the i-th manager accepts a good project, while q_i is his probability to accept a bad project. Initially each member decides independently. Furthermore,

$$p_1 > p_2 = p_3 \qquad q_1 < q_2 = q_3$$

Depending on the project being good or bad we get an *initial distribution* $\mathbf{p}_G$ or $\mathbf{p}_B$ that we write in a vector form (each vector is a column vector):

$$\mathbf{p}_G = [p_1 p_2^2, p_1 p_2 (1-p_2), p_1 p_2 (1-p_2), p_1 (1-p_2)^2, (1-p_1) p_2^2$$

$$(1-p_1) p_2 (1-p_2), (1-p_1) p_2 (1-p_2), (1-p_1) (1-p_2)^2]'$$

$$\mathbf{p}_B = [q_1 q_2^2, q_1 q_2 (1-q_2), q_1 q_2 (1-q_2), q_1 (1-q_2)^2, (1-q_1) q_2^2$$

$$(1-q_1) q_2 (1-q_2), (1-q_1) q_2 (1-q_2), (1-q_1) (1-q_2)^2]'$$

The i-th entry in these vectors corresponds to the probability of the i-th state in the ordering given above.

If the board is initially split the process through which unanimity is reached is described by means of a Markov Chain. To construct the transition matrix of this Markov Chain we make the following assumptions.

1) The manager "1" is able to influence the other managers and, on the contrary, is not influenced by them;

2) Managers "2" and "3" have a probability φ to stick to their initial decision, either to accept or to reject the project; furthermore they act independently of each other.

Then the transition matrix $\mathbf{Q}$ has the following form:

$$\mathbf{Q} = \begin{bmatrix} 1 & \mathbf{0} & 0 \\ \mathbf{u} & \mathbf{W} & \mathbf{v} \\ 0 & \mathbf{0} & 1 \end{bmatrix}$$

where $\mathbf{u}$ is (6×1), $\mathbf{W}$ is (6×6), and $\mathbf{v}$ is (6×1). Furthermore $\mathbf{W}$ can be written as

$$\mathbf{W} = \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{EAE} \end{bmatrix}$$

where each submatrix is (3×3), $\mathbf{A}$ is given by

$$\mathbf{A} = \begin{bmatrix} \varphi^2 & (1-\varphi)^2 & \varphi(1-\varphi) \\ (1-\varphi)^2 & \varphi^2 & \varphi(1-\varphi) \\ \varphi(1-\varphi) & \varphi(1-\varphi) & \varphi^2 \end{bmatrix}$$

and $\mathbf{E}$ is the orthogonal matrix

$$\mathbf{E} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

$\mathbf{u}$ and $\mathbf{v}$ can be written as

$$\mathbf{u} = \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{0} \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} \mathbf{0} \\ \mathbf{v}_1 \end{bmatrix}$$

where each subvector is (3×1) and

$$\mathbf{u}_1 = \begin{bmatrix} \varphi(1-\varphi) \\ \varphi(1-\varphi) \\ (1-\varphi)^2 \end{bmatrix}, \quad \mathbf{v}_1 = \begin{bmatrix} (1-\varphi)^2 \\ \varphi(1-\varphi) \\ \varphi(1-\varphi) \end{bmatrix}$$

Because $\mathbf{Q}$ is a transition matrix, we have $\mathbf{u} + \mathbf{v} + \mathbf{We}_6 = \mathbf{e}_6$, and so $\mathbf{u} + \mathbf{v} = (\mathbf{I}-\mathbf{W})\,\mathbf{e}_6$, where $\mathbf{I}$ is the identity matrix and $\mathbf{e}_6$ is a (6×1)-vector of all ones.

Considering the block structure of $\mathbf{W}$, we have the simplified relationships $\mathbf{u}_1 + \mathbf{Ae}_3 = \mathbf{e}_3$, which implies $\mathbf{u}_1 = (\mathbf{I}-\mathbf{A})\,\mathbf{e}_3$, and $\mathbf{v}_1 + \mathbf{EAEe}_3 = \mathbf{e}_3$, which implies $\mathbf{v}_1 = \mathbf{E}(\mathbf{I}-\mathbf{A})\,\mathbf{Ee}_3$.

Let τ_G and τ_B be the times required to reach unanimity if the project is good or bad, respectively.

Then

$$Pr[\tau_J = 0] = \mathbf{p}_J'\mathbf{r}$$

$$Pr[\tau_J = k] = \mathbf{p}_J'(\mathbf{Q}^k - \mathbf{Q}^{k-1})\,\mathbf{r}$$

where J stands for G or B, and $\mathbf{r}$ is a (8×1)-vector with the first and the 8-th elements equal to 1 and the remaining equal to 0.

This is a proper distribution if $\lim_{i\to\infty} \mathbf{Q}^i$ exists and it is equal to (say) $\mathbf{Q}^\infty$. Since

$$\mathbf{Q}^i = \begin{bmatrix} 1 & \mathbf{0} & 0 \\ \left(\sum_{h=0}^{i-1} \mathbf{W}^h\right)\mathbf{u} & \mathbf{W}^i & \left(\sum_{h=0}^{i-1} \mathbf{W}^h\right)\mathbf{v} \\ 0 & \mathbf{0} & 1 \end{bmatrix}$$

this amounts to ascertain whether all the eigenvalues of $\mathbf{W}$ are, in absolute value, strictly less than 1.

Given the structure of $\mathbf{W}$ this, in turn, implies studying the spectrum of $\mathbf{A}$. Given our assumptions about φ, a result in Catalani (1993) allows to conclude that all the eigenvalues of $\mathbf{W}$ are indeed, in absolute value, strictly less than 1.

Now

$$\mathbf{Q}^k - \mathbf{Q}^{k-1} = \begin{bmatrix} 0 & \mathbf{0} & 0 \\ \mathbf{W}^{k-1}\mathbf{u} & \mathbf{W}^{k-1}(\mathbf{W}-\mathbf{I}) & \mathbf{W}^{k-1}\mathbf{v} \\ 0 & \mathbf{0} & 0 \end{bmatrix}$$

and so the expected value of the transition time τ_j is given by $E(\tau_j) = \mathbf{p}_j' \sum_{k=1}^{\infty} k(\mathbf{Q}^k - \mathbf{Q}^{k-1})\mathbf{r}$, where

$$\sum_{k=1}^{\infty} k(\mathbf{Q}^k - \mathbf{Q}^{k-1}) = \begin{bmatrix} 0 & \mathbf{0} & 0 \\ \sum_{k=1}^{\infty} k\mathbf{W}^{k-1}\mathbf{u} & \sum_{k=1}^{\infty} k\mathbf{W}^{k-1}(\mathbf{W}-\mathbf{I}) & \sum_{k=1}^{\infty} k\mathbf{W}^{k-1}\mathbf{v} \\ 0 & \mathbf{0} & 0 \end{bmatrix}$$

We know that the power series

$$\sum_{p=0}^{\infty} \lambda^p \text{ converges to } f(\lambda) = (1-\lambda)^{-1} \text{ for } |\lambda| < 1,$$

and so may be differentiated term by term any number of times within the circle of convergence. Consequently

$$\sum_{p=1}^{\infty} p\lambda^{p-1} = (1-\lambda)^{-2}$$

Using results in matrix theory (see, for example, Gantmacher, 1974) we know that this expansion remains valid when the scalar argument λ is replaced by a matrix whose eigenvalues lie within the circle of convergence.

This being the case for $\mathbf{W}$, we get

$$\sum_{k=1}^{\infty} k\mathbf{W}^{k-1} = (\mathbf{I}-\mathbf{W})^{-2},$$

and consequently

$$\sum_{k=1}^{\infty} k(\mathbf{Q}^k - \mathbf{Q}^{k-1})\mathbf{r} = \begin{bmatrix} 0 \\ (\mathbf{I}-\mathbf{W})^{-2}(\mathbf{u}+\mathbf{v}) \\ 0 \end{bmatrix}$$

Finally, writing π_{J1} and π_{J2} for the (3×1)-vectors made up, respectively, by the 2-nd, 3-rd, 4-th element and the 5-th, 6-th, 7-th element of $\mathbf{p}_J$, we get

$$E(\tau_J) = \pi'_{J1}(\mathbf{I}-\mathbf{A})^{-1}\mathbf{e}_3 + \pi'_{J2}\mathbf{E}(\mathbf{I}-\mathbf{A})^{-1}\mathbf{E}\mathbf{e}_3$$

What is the probability that with unanimity the project be accepted? Intuitively, it is p_1 if the project is good, and q_1 if the project is bad. That it is indeed so is confirmed by the fact that $Pr[A|$ the project is good$] = \mathbf{p}'_G\mathbf{Q}^{\infty}\mathbf{r}_1$, where $\mathbf{r}_1$ is a (8×1)-vector with the first element equal to one and all the others equal to zero. Then

$$\mathbf{p}'_G\mathbf{Q}^{\infty}\mathbf{r}_1 = p_1p_2^2 + \pi'_{G1}(\mathbf{I}-\mathbf{A})^{-1}\mathbf{u}_1$$

$$\Big| \quad = p_1p^2{}_2 + \pi'_{G1}\mathbf{e}_3 = p_1$$

Analogously it is proved that $Pr[A|$ the project is bad$] = \mathbf{p}'_B\mathbf{Q}^{\infty}\mathbf{r}_2 = q_1$, where $\mathbf{r}_1 + \mathbf{r}_2 = \mathbf{r}$.

In what follows we confine ourselves to the case when the scrutinized project is good. It seems reasonable to measure the strength of the leadership by the quantity $p_1 - p_2$: then $\varphi = 1-(p_1-p_2)$.

A little calculation shows that

$$E(\tau_G) = [p_2(1-p_2) + 2(1-p_1+p_2)(p_1+p_2-2p_1p_2)]/[(p_1-p_2)(1-p_1+p_2)]$$

If c_1 is the cost incurred by rejecting a project when it is good, then the total average cost incurred by a structure ruled by unanimity is

$$\gamma_1 = E(\tau_G) + c_1 \times Pr[R|\text{ the project is good}]$$

$$= E(\tau_G) + c_1(1-p_1)$$

that is: the cost related to the time spent in reaching unanimity plus the average cost of the outcome of the taken decision.

Considering the *states'* table given at the beginning of this Appendix, we see that, after some simplifications, the total average cost incurred by a structure ruled by simple majority is

$$\gamma_2 = c_1(1-p_2)(1+p_2-2p_1p_2)$$

Our key variable is then

$$\Delta = \gamma_1 - \gamma_2 = c_1[(1-p_1)-(1-p_2)(1+p_2-2p_1p_2)] + E(\tau_G)$$

$\Delta < 0$ means that unanimity is superior.

Let us set $\Delta = 0$ and solve the resulting equation for c_1. Let c_1^* denote the solution; then c_1^* is the indifference cost:

$$c_1^* = -E(\tau_G)/[(1-p_1)-(1-p_2)(1+p_2-2p_1p_2)]$$

Given that $E(\tau_G) > 0$, to have $c_1^* \geq 0$ (which means a meaningful solution) it is necessary that

$$(1-p_1)-(1-p_2)(1+p_2-2p_1p_2) < 0 \qquad (*)$$

If it is not so, then simple majority is always superior.

Since in the conventional case $p_1 > p_2$, we see that if $p_1 \leq 1/2$, then the above inequality (*) always holds.

Aside from this case, the inequality (*) holds if

$$p_1 > p_2^2/(1-2p_2+2p_2^2)$$

So in this situation for all $c_1 > c_1^*$ unanimity is better.

On the other hand, if

$$p_2 < p_1 < p_2^2/(1-2p_2+2p_2^2)$$

which implies $1/2 \leq p_2 < 1$, whatever c_1, then simple majority performs always better.

The dependence of the solution c_1^* on the parameters p_1 and p_2 is a troublesome problem: the best way to get around is to focus on the picture given in Section 3.

REFERENCES

CATALANI M.S., "Transition Times in a First Order Homogeneous Markov Chain: Some Matrix Form Algorithms", Dipartimento di Economia, Torino, 1993, *mimeo.*

—— and CLERICO G.F. (1993a), "Team Cooperation or Independent Assessment?", *Rivista Internazionale di Scienze Economiche e Commerciali*, No. 2, 1993, *40*, 105-18.

—— and —— (1993b), "Pyramid Decision Structures: A Preliminary Note", *Rivista Internazionale di Scienze Economiche e Commerciali*, No. 6-7, 1993, *40*, 587-97.

—— and —— (1994), "Processo decisionale e burocrazia", in corso di pubblicazione in un libro collettaneo, curato da G. Sobbrio, Milano: Giuffrè.

GANTMACHER F.R., *The Theory of Matrices*, Vol. I and II, New York: Chelsea Publishing Company, 1974.

NITZAN S. and PAROUSH J., "Are Qualified Majority Rules Special?", *Public Choice*, 1984, *42*, 257-72.

—— and PROCACCIA U., "Optimal Voting Procedures for Profit Maximizing Firms", *Public Choice*, 1986, *51*, 191-208.

SAH R.K., "Fallibility in Human Organizations and Political Systems", *Journal of Economic Perspectives*, 1991, *2*, 15-24.

—— and STIGLITZ J.E., "Human Fallibility and Economic Organizations", *American Economic Review*, Papers and Proceedings, 1985, *75*, 292-97.

—— and ——, "The Architecture of Economic Systems: Hierarchies and Polyarchies", *American Economic Review*, 1986, *76*, 716-27.

—— and —— (1988a), *Human Fallibility and Economic Organizations*, Cambridge Mass: MIT Press, 1988.

—— and —— (1988b), "Committees, Hierarchies and Polyarchies", *Economic Journal*, 1988, *98*, 451-70.

SHAPLEY L. and GROFMAN B., "Optimizing Group Judgmental Accuracy in the Presence of Interdependencies", *Public Choice*, 1984, *43*, 329-43.

COME E QUANDO L'UNANIMITÀ È CONVENIENTE

Questo lavoro mira a costruire un modello che possa render conto dell'intuizione secondo cui quando le decisioni da prendersi sono molto importanti si dovrebbe ricorrere a maggioranze qualificate. Viene proposto un modello che permette il confronto in termini di costo complessivo della decisione tra unanimità e maggioranza semplice, tanto nel caso di progetti convenzionali che non convenzionali. Nel caso di un progetto convenzionale il processo attraverso il quale si raggiunge l'unanimità è modellato con le Catene di Markov. I risultati del modello confermano l'intuizione di cui sopra. In particolare si dimostra che (con l'eccezione di un campo ristretto dei valori dei parametri) più è elevato il costo dell'errore più è verosimile che sia più conveniente l'unanimità. Un risultato sostanzialmente opposto si ottiene nel caso di progetti non convenzionali.

Rivista Internazionale di Scienze Economiche e Commerciali
Volume 42 (1995), N. 1-2, 33-56

A MODEL OF THE AUDIENCE FOR TV BROADCASTING: IMPLICATIONS FOR ADVERTISING COMPETITION AND REGULATION

by
ALESSANDRO VAGLIO *

0. INTRODUCTION

This paper uses a Hotelling-type model to analyse competition in the advertising-supported broadcasting business [1]. Firms in this industry operate on the borderline between two markets. In the first one, they "sell" to advertisers their audiences multiplied by the advertising time; on the other one, they compete for audiences. Competition in the first market *per se* has nothing of specific interest. It can be modeled as Cournot competition, with "quantities" being represented by the time devoted to advertising by each broadcasting firm multiplied by the corresponding audience; in a more complex setting, advertising time could be represented as a differentiated commodity, with its value (to the advertiser) depending on the specific broadcasting firm offering it and on the program to which advertising is attached. Since this paper does not focus on this market, I shall not go deeper in this discussion. Throughout the paper I shall assume (with some loss of understanding of the market) that the value (to the advertiser) of a unit of advertising time per viewer is constant and (with no further loss of information) that this value is 1.

What makes advertising-supported broadcasting a peculiar business are the features of the second market, the one where broadcasting firms com-

* Università degli Studi di Ancona, Dipartimento di Economia, Ancona.

While working on this essay I have been helped and encouraged by comments and suggestions by Professors Francesco Silva and Michele Polo and by Maria Rosa Battaggion. Mine is the responsibility for the article's contents.

[1] Although much of what is in the paper applies to radio broadcasting as well, for the sake of brevity here I refer only to TV broadcasting and define the audience as the "viewers".

pete for audiences. TV-watching is not a free activity: there are fixed costs (equipment and taxes), variable costs (electric power) and opportunity costs (time spent watching television could be differently used). There is however a relevant point: such costs do not differ across different TV programs. Therefore, supposing viewers perform a two-stage budgeting process, once they have determined the amount of time they want to spend watching television, the choice between two different programs depends only on their characteristics. Unlike the existing economic literature on this industry [2], in this paper the rate of advertising (i.e. the ratio of advertising time to total broadcasting time) is included among the program characteristics which are relevant to the viewer's choice [3]. This paper builds on the assumption that there exists, for any program, an "advertising-saturation level", i.e. a critical value of the advertising rate, such that any further increase in this rate reduces the viewer's utility (other things being equal). In the following sections, I shall assume that the saturation level is 0, i.e. increases in the advertising rate always reduce utility: however this is just a simplifying assumption; I do not need to deny that, in general, *some* amount of advertising is indifferent or even provides positive marginal utility to the viewer. The implication is that the viewer does pay a positive utility price other than the direct and opportunity costs mentioned above, namely, the watching of advertising. As the advertising rate increases, some viewers either shift permanently to competing programs with a lower rate, or become more prepared to shift during commercial interruptions [4]: in both cases, the program loses audience (at least, the advertiser-relevant audience, i.e. those actually exposed to advertising messages).

Since the broadcaster revenue per unit of time is the advertising rate times the audience (under the above hypothesis on the value of advertising time to advertisers), broadcasting firms then face a trade-off: either they capture large audiences by keeping the advertising rate low or they stuff

---

[2] SPENCE-OWEN (1977) represents perhaps the most rigorous theoretical treatment available. General up-to-date bibliographic references on TV economics can be found in DUNNETT (1990) and GAMBARO-SILVA (1992) (the latter, with a particular stress on the Italian case).

[3] A passing remark about this is in BARNETT-GREENBERG (1971, p. 89). An explicit Hotelling-type formalization of consumer choice in the broadcasting sector is used by BOWMAN (1975) but the advertising content is not explicitly considered among the relevant characteristics.

[4] Switching back and forth from a program to another has in general a cost in terms of utility, even if remote-control devices have strongly reduced it (think for instance of the risk of missing the first minutes of the program after the end of a commercial interruption, if the length of the latter is not known with certainty); a higher advertising rate makes it worthwhile to bear this cost.

programs with advertising interruptions, thereby losing audience. The advertising rate-setting decision thus closely resembles the price-making decision.

In this paper I shall consider two additional characteristics attached to programs other than the advertising rate.

The first one is "quality", which I shall indicate by x. If two programs are identical under every point of view, except as regards their quality levels, all consumers (independently of their tastes) prefer the program where x is highest. Firms which want to broadcast high-quality programs have to pay a cost which increases with quality itself. In this model, quality has a particular meaning. It should be more precisely described as *advertising* quality, i.e. the cost of quality is an expenditure the aim of which is to make advertising less unpleasant. This notion is formalized in subsection 1.1.

The second characteristic is the program "type". If the "type" of a given program changes then some viewers of that program will be better off, while others will be worse off. If two programs differ only as regards their "types", none will be necessarily preferred by all viewers.

Unlike the case of quality, firms which want to broadcast a program pay the same cost, independently of the type.

As we shall see, firms' decisions concerning "quality" and "type" interact with those concerning advertising rates in similar ways. If a firm increases quality, then either the audience for its program increases (other things being equal) or remains the same, but with a higher advertising rate. Thus, "quality" can be used either to "soften" competition on advertising rates (by allowing the firm to increase its rate without losing audience and without triggering "aggressive" reactions from the opponents) or to increase competition, by diverting the audience from the other firm. In the case studied here, the first effect of "quality" decisions will dominate.

The role of decisions concerning "types" is very similar. Just to make a simple example, according to standard models of advertising-supported broadcasting, if the program broadcast by CBS has an audience of 100 and the potential audience of an alternative programs is 40, NBC certainly selects the same program as CBS: supposing they share the audience in a fifty-fifty proportion, NBC is better off with 50 viewers than with 40; this is true however only if the advertising rate is constant. Suppose instead that, in the spirit of the present approach, the selection of the alternative program by NBC allowed an increase in the advertising rate by more than 25%: the advertising revenue would then be higher (for both firms) in this case than otherwise. Thus this paper suggests that broadcasting firms face an incentive to differentiate larger than otherwise expected.

In many countries, advertising rates are subjected to an upper bound, either by government will or by a restraint self-imposed by broadcasters. The consequences and the desirability of such forms of regulation are sometimes debated, although seldom within an explicit theoretical framework. Critics of advertising rate regulation stress in fact that it is advertising which allows broadcasters to supply programs of a given quality and with a certain degree of differentiation; both would be reduced by introduction (or by tightening) of regulatory constraints; the link between advertising rates and other program characteristics however is not explained. On the other hand, those who support these forms of regulation assert that advertising is in excess of some social optimum, although this optimum is never explicitly characterized.

This paper contributes to this debate in three ways.

First of all, the relationship between advertising rates, quality and program variety is explicitly modeled; advertising rate regulation is seen as reversing the sequence in which the decisions concerning advertising rates, program types and program quality are taken: while I assume that (in the absence of regulatory constraints) advertising rates are determined in the last stage of the decision process (the decisions concerning them being the most easily reversible), regulation irreversibly fixes the advertising rate before decisions concerning quality levels and program types are taken.

Secondly, I address the problem of the optimal configuration of advertising rates and quality levels. Here the main results is that, for a fairly large class of tastes and technologies, both quality and advertising rates are smaller under regulation than under market equilibrium. This is true, however, under a strongly "interventionist" assumption about the powers of the regulating authority. Imposing some additional restraints on the authority's power, the result no longer holds.

Third, I get some results concerning the positive effects of regulation as regards both quality and the degree of program differentiation. Advertising rate regulation is shown to reduce the quality of programs broadcast by the regulated firm, but to have an ambiguous effect on the opponent's quality level. On the other hand, regulation is shown to reduce the degree of program differentiation.

The paper is organised as follows. In subsection I.1. I describe preferences and technology. From that point onward I separately discuss the relation between advertising and quality and the relation between advertising and program variety. Throughout subsections I.2 and I.3 therefore the program type is kept fixed. Subsection I.2 describes unrestrained competition on advertsing rates and quality while I.3 analyses the impact of regula-

tion in a market of this type (Appendices 1 and 2 refer to these subsections). Throughout Section II instead quality is kept fixed while firms are allowed to select their program types. Subsections II.1 and II.2 respectively focus on competition and on the effects of regulation.

I. Competition in the TV broadcasting sector

I.1 *Demand and Supply*

Demand and program characteristics. – Viewers are heterogeneous; the consumer's "type" is described by a real number in the interval $[0,S]$. Viewers are uniformly distributed over types, i.e. over $[0,S]$. The type of a program is also described by a number in the interval $[0, S]$. x proxies instead the program's "quality". $x \in R_x \equiv [\underline{x}, \infty)$; $\underline{x} > 0$.

Advertising "spots" can be inserted within each program. Let the ratio A of advertising time to total broadcasting time (for a specific program) be called the program's "advertising rate" $A \in [0, A_M]$.

Therefore, a type-j viewer watching a type-z program gets a utility of:

$$U(j, z, A, x) = u(A\, g(x), |z - j|)$$

Assumption 1.

(i) $$u(\cdot) : \{[0, A_M] \times R_x \times [0, S] \to R\};$$

$$\frac{\partial u(\cdot)}{\partial (A\, g(x))} \leq 0; \qquad \frac{\partial u(\cdot)}{\partial |z - j|} \leq 0;$$

(ii) $$g(x): \bar{R}_{\underline{x}} \to \bar{R}_{+}; \qquad g(x) > 0;$$

$$g'(x) \equiv \frac{\partial g(x)}{\partial x} \leq 0; \qquad g(\underline{x}) = 1;$$

(iii) $$u(g(\underline{x})\, A_M, S) > u_0$$

where u_0 is the utility from switching off, $u(\cdot)$ and $g(\cdot)$ are everywhere continuous and differentiable.

No further constraints on the utility function are required at this stage. The relevant restrictions will be put instead on the demand functions derived from the utility function itself. The consumer maximizes utility by

choosing one program per unit of time and taking the range of available programs and the associated advertising rates as given.

Assumption 1(iii) implies that consumers never switch off, since even under the worst coinceivable conditions ($A = A_M$, $x = \underline{x}$, $|z - j| = S$) utility from TV watching exceeds utility from switching off. Suppose now that there are just two available programs, 'a' and 'b' being the corresponding value of the relevant characteristic ($b > a$). Suppose that A_1 and A_2 are the advertising rates respectively associated to programs 'a' and 'b', while x_1 and x_2 are the corresponding quality levels of the advertisement in the two programs.

Definition 1. Let:

$$h \equiv h\,(A_1 g\,(x_1),\ A_2 g\,(x_2),\ a,\ b)\colon \{[0,\ A_M]^2 \times R_{\bar{x}}^2 \times [0,\ S]^2\} \to [0,\ S]$$

be the implicit function which solves:

$$u\,(A_1 g\,(x_1),\ |h - a|) = u\,(A_2 g\,(x_2),\ |h - b|)$$

i.e. h is the type of the viewer which is indifferent between the two programs.

From assumption 1 it is immediate to conclude:

$$h_{A_1} \equiv \frac{\partial h\,(A_1 g\,(x_1),\ A_2 g\,(x_2);\ a,\ b)}{\partial\,(A_1\ g\,(x_1))} < 0$$

$$h_{A_2} \equiv \frac{\partial h\,(A_1 g\,(x_1),\ A_2 g\,(x_2);\ a,\ b)}{\partial\,(A_2 g\,(x_2))} > 0 \qquad (1.2)$$

$$h_{x_1} \equiv \frac{\partial h\,(A_1 g\,(x_1),\ A_2 g\,(x_2);\ a,\ b)}{\partial x_1} > 0$$

$$h_{x_2} \equiv \frac{\partial h\,(A_1 g\,(x_1),\ A_2 g\,(x_2);\ a,\ b)}{\partial x_2} < 0$$

$$\frac{\partial h\,(\cdot)}{\partial a} \equiv h_a > 0 \qquad \frac{\partial h\,(\cdot)}{\partial b} \equiv h_b > 0$$

Since any consumer whose type is described by a characteristic value smaller than h strictly prefers program 'a', h and $S - h$ represent the audiences for, respectively, program 'a' and program 'b'. Therefore, the audience for program 'a' increases as A_1 decreases, as the quality of advertisement on program 'a' (x_1) increases and as the value of 'a' increases.

Supply. – On the supply side, there exist two firms with no threat of entry [5], each broadcasting one program per unit of time, so that the number of available programs is always two. Firms will be henceforth labeled as "Firm 1" and "Firm 2". Firms are profit maximizers.

The stage structure of the decision process is a key issue. Throughout the paper I shall assume that the choice of the program-type is the least easily reversible, while the choice concerning the levels of x_1 and x_2 is either less easily reversible than or as easily reversible as the choice of A_1 and A_2. There are therefore two possibilities. Either:

CASE A

I) In stage I, firms select [6] irreversibly their program type, i.e. a value contained in the interval $[0, S]$; let 'a' be the value chosen by Firm 1 and 'b' the value chosen by Firm 2, and assume $b > a$;

II) in stage II, firms choose the levels (x_1, x_2) of advertising quality, the advertising rates (A_1, A_2) and broadcast the programs. Or:

CASE B

I) In stage I, firms select irreversibly their program type (as in case A);

II) in stage II, firms choose the levels (x_1, x_2) of advertising quality;

III) in stage III, firms choose advertising rates and broadcast the programs.

We shall prove in Appendix 1, however, that given our description of preferences, the two possibilities are equivalent, i.e., the equilibrium levels of advertising quality levels and of advertising rates do not differ across cases A and B. The equilibrium values would instead be different in the following case:

CASE C

I) In stage I, firms select irreversibly their program type;

II) in stage II, a third party determines (A_1, A_2);

[5] This does not seem an exceedingly unrealistic assumption, since concentration in broadcasting seems to be high and persistent. The US case, with three dominant national networks is the obvious example; in the Italian case, the end of public monopoly in broadcasting resulted in a duopoly, with the two main competitors covering a very substantial share of the audience. Since concentration seems unresponsive to changes in market size, broadcasting could probably make a good industry-test for SUTTON's (1991) exogenous/endogenous sunk cost approach.

[6] Notice that, except in unambiguous cases, firms will be described as *selecting* program types, among which viewers *choose*.

III) in stage III, firms choose the values of (x_1, x_2) and broadcast the programs.

The "third party" mentioned in stage II must obviously be interpreted as a regulating authority.

As regards costs, any program selection involves the same fixed cost; broadcasting costs are also assumed to be only of the fixed type. Since the level of these costs does not play any relevant role in the present analysis, I shall assume that it is 0 for both firms. A firm choosing a quality level of x has to bear a cost of $c(x)$; $c(x)$ is assumed to be an increasing and twice differentiable function. Some assumptions on the shape of $c(x)$ will be discussed in the context of subsection I.3.

I.2 *Competition on Advertising Rates and Quality* (fixed program type)

In this paragraph I analyse the features of competition in this market in the absence of regulation. Let us assume that the stage structure of the decision process is as described by case A above. However, we shall be interested only in stage II, since we assume program types as given. In stage II, each firm therefore fixes its own advertising rate and quality level, taking the corresponding opponent's choices as given.

The value of a unit of advertising time to an advertiser is proportional to the size of the audience exposed to that advertising message [7]. As I specified in the introduction, we shall assume that the value of one unit of advertising time per viewer is unit and it is independent of the audience size.

Definition 2

$$\Pi^1(A_1, A_2, x_1, x_2, a, b) \equiv A_1 b(A_1 g(x_1), A_2 g(x_2), a, b) - c(x_1)$$

and

$$\Pi^2(A_1 A_2, x_1, x_2, a, b) \equiv A_2[S - b(A_1 g(x_1), A_2 g(x_2), a, b)] - c(x_2)$$

Assumption 2. For any (a, b), the profit functions $\Pi^1(A_1, A_2, x_1, x_2, a, b)$ and $\Pi^2(A_1, A_2, x_1, x_2, a, b)$ are respectively strictly concave with respect to (A_1, x_1) and (A_2, x_2).

The decision problem of the firms can be described as:

[7] That the value that advertisers assign to broadcasting time is extremely sensitive not only to the size but also to the composition of the audience is documented by Evans-Fisher-McGowan (1990).

$$\max_{A_1, x_1} \Pi^1 (A_1, A_2, x_1, x_2, a, b)$$
$$\max_{A_2, x_2} \Pi^2 (A_1, A_2, x_1, x_2, a, b) \qquad \text{(I.3)}$$

The interior solutions to problem (I.3) can then be characterised by the following first-order condition:

$$\Pi^1_{A_1} \equiv \frac{\partial \Pi^1}{\partial A_1} = b + A_1 b_1 g(x_1) = 0 \qquad \text{(I.3a)}$$

$$\Pi^2_{A_2} \equiv \frac{\partial \Pi^2}{\partial A_2} = S - b - A_2 b_2 g(x_2) = 0 \qquad \text{(I.3b)}$$

$$\Pi^1_{x_1} \equiv \frac{\partial \Pi^1}{\partial x_1} = [A_1]^2 b_1 g'(x_1) - c'(x_1) = 0 \qquad \text{(I.3c)}$$

$$\Pi^2_{x_2} \equiv \frac{\partial \Pi^2}{\partial x_2} = -[A_2]^2 b_2 g'(x_2) - c'(x_2) = 0 \qquad \text{(I.3d)}$$

Assumption 3

There exists a unique (pure strategy) Nash equilibrium in advertising rates and quality levels [8], i.e. a vector:

$$(x_1^*, x_2^*, A_1^*, A_2^*) \neq 0$$

which solves the (I.3) system of first-order conditions above.

Until the end of this section, I shall omit the indication of (a, b) among the arguments of the functions so far defined, since I temporarily disregard the problem of their determination. In Appendix 1 it is shown that the equilibrium solution for case B does not differ from the one obtained for case A.

I.3 *Second Best Advertising Rate Regulation*

In this paragraph we shall study two possible versions of the decision problem facing a regulating authority which has the power to determine the level of the advertising rates.

In the first version (defined as Complete Regulation, CR), the authori-

[8] As MacLeod (1986) demonstrates, if "b" is too close to "a", equilibrium might not exist in stage II. Thus I am implicitly assuming that the difference between "a" and "b" is always large enough to rule these existence problems.

ty simultaneously determines both (A_1, A_2) and (x_1, x_2); in the second version, which I shall indicate as Partial Regulation (PR), the authority determines the level of (A_1, A_2) and firms subsequently determine the levels of (x_1, x_2). In both versions of the problem, the authority is constrained to have the firms break even and takes as given both the program types and the number of the firms (two). This is why regulation is "second-best" in both versions, although under CR the authority is less tightly constrained than under PR.

Complete regulation. – The decision problem of the regulating authority is now:

$$\max_{A_1, A_2, x_1, x_2} \int_0^h u\,(A_1 g\,(x_1),\ |a - j|)\,dj + \int_h^S u\,(A_2 g\,(x_2),\ |b - j|)\,dj$$

s.t.

$$\begin{aligned} & h = h\,(A_1 g\,(x_1),\ A_2 g\,(x_2)) \qquad \text{(I.4)} \\ & A_1 h\,(A_1 g\,(x_1),\ A_2 g\,(x_2)) - c\,(x_1) \geq 0 \\ & A_2 [S - h\,(A_1 g\,(x_1),\ A_2 g\,(x_2))] - c\,(x_2) \geq 0 \end{aligned}$$

The first-order conditions for the above problem reduce to the following expression (we report only the one related to firm 1):

$$[A_1^2 h_1 g'\,(x_1) - c'\,(x_1)] = \frac{A_1 g'\,(x_1)}{g\,(x_1)} [h + A_1 h_1 g\,(x_1)] \qquad \text{(I.4a)}$$

which (together with the 0-profit condition for firm 1) implies:

$$-\frac{c\,(x_1)}{c'\,(x_1)} = \frac{g'\,(x_1)}{g\,(x_1)} \qquad \text{(I.4b)}$$

From the first-order conditions for profit maximization (cases A and B above) we get instead:

$$-\frac{c\,(x_1)}{c'\,(x_1)} \leq \frac{g'\,(x_1)}{g\,(x_1)}$$

The inequality above can be translated into:

$$\varepsilon_c \geq |\varepsilon_a| \qquad \text{(in the equilibrium solution)}$$

$$\varepsilon_c = |\varepsilon_g| \qquad \text{(in the optimal solution)}$$

where ε_c and ε_g are respectively the elasticities of $c(x)$ and $g(x)$ with respect to x.

It is therefore possible to compare optimal and equilibrium solution under different assumptions on the shape of $c(x)$ and $g(x)$. A class of relevant and simple situations can be encompassed by the following functional forms:

$$c(x) = F + x^{\alpha} \qquad g(x) = D + x^{-\beta}$$

which imply:

$$\varepsilon_c = \frac{\alpha}{1 + F x^{-\alpha}}; \qquad \lim_{x \to \infty} \varepsilon_c = \alpha \qquad |\varepsilon_g| = \frac{\beta}{1 + D x^{\beta}}$$

If $D > 0$ or if $(D = 0, \alpha > \beta)$, then the optimal level of x is smaller than the equilibrium level. If consumer's utility must be at least as large under social optimization as under market equilibrium, then this implies that the optimal advertising rate is also smaller than its equilibrium level.

Partial regulation. – The decision problem for the authority now is:

$$\max_{A_1, A_2} \int_0^h u(A_1 g(x_1(A)), |a - j|)\, dj + \int_h^S u(A_2 g(x_2(A)), |b - j|)\, dj$$

s.t.

$$\begin{aligned} h &= h(A_1 g(x_1(A)), A_2 g(x_2(A))) \\ A_1 h(A_1 g(x_1(A)), A_2 g(x_2(A))) - c(x_1(A)) &\geq 0 \\ A_2 [S - h(A_1 g(x_1(A)), A_2 g(x_2(A)))] - c(x_2(A)) &\geq 0 \end{aligned} \qquad \text{(I.5)}$$

where: $A \equiv (A_1, A_2)$

and $(x_1(A), x_2(A))$ are the levels of (x_1, x_2) which solve the first order conditions for stage II, where firms freely determine X:

$$\frac{\partial \Pi^1}{\partial x_1} = A_1^2 h_1 g'(x_1) - c'(x_1) = 0 \qquad \text{(I.6a)}$$

$$\frac{\partial \Pi^2}{\partial x_2} = -A_1^2 h_2 g'(x_2) - c'(x_2) = 0 \qquad \text{(I.6b)}$$

The solution to the above problem cannot be easily characterized.

We are then interested here only in the positive effect of changes in $A \equiv (A_1, A_2)$ on the choice of (x_1, x_2). By differentiating system I.6a-b we can find the following:

Comparative Statics Result I

$$\frac{dx_i}{dA_i} \geq 0; \qquad \frac{dx_i}{dA_j} \gtreqless 0 \quad (i \neq j); \qquad \frac{dx_i}{dA_j} = 0 \text{ if: } b_{12} = 0$$

$$\text{if: } \frac{dx_i}{dA_j} > 0, \qquad \text{then } \frac{dx_j}{dA_i} < 0;$$

Summing up, if we assume complete regulation, as long as the cost function for quality and the "quality function" $g(x)$ have fixed components, then both the quality level and the advertising rate are smaller than their equilibrium counterparts. This conclusion lends some support to the idea that commercial television leads to too much advertising and too much *expenditure* on advertising (since $c(x)$ can be interpreted as an expenditure the purpose of which is to make advertising more attractive).

Comparative statics result I confirms that (with partial regulation) the imposition of a binding upper constraint on advertising rates does actually reduce quality *directly*, but may *indirectly* increase the quality of the competing firm (through the strategic effect). Since the sign of the strategic effect is opposite for the two firms, then if:

$$\frac{\partial x_j}{\partial A_i} > 0$$

certainly the imposition of a binding upper constraint on the advertising rate of firm '*i*' reduces the quality level in both programs. On the other hand, the characterization of the PR social optimum is not immediate unless more stringent conditions are imposed on the model.

II. Competition on Advertsing Rates and Program Types (fixed quality)

II.1 *Properties of Competition*

Let us now consider a variation on case A above, by assuming that $x_1 = x_2 = \underline{x}$.

The decision problem of firms in stage II reduces to:

$$\max_{A_1} A_1 [h(A_1, A_2, a, b)]$$
$$\max_{A_2} A_2 [S - h(A_1, A_2, a, b)] \tag{II.1}$$

Assumption 2 leads to the following first-order conditions:

$$A_1 h_1 + h = 0$$
$$S - h - A_2 h_2 = 0 \tag{II.1a}$$

In order to study the comparative statics of the system above, we need the following assumption:

Assumption 4

$$A_1 h_{1a} + h_a < 0 \tag{i}$$
$$- A_2 h_{2a} - h_a < 0 \tag{ii}$$
$$A_1 h_{12} + h_2 > 0 \tag{iii}$$
$$- A_2 h_{21} - h_1 > 0 \tag{iv}$$

Under assumption 2-3 and 4, by differentiation of system (II.1a), we get the following:

Comparative Statics Result II

$$\frac{dA_1^*(a, b)}{da} \leq 0; \quad \frac{dA_2^*(a, b)}{da} \leq 0; \quad \frac{dA_1^*(a, b)}{db} \geq 0;$$
$$\frac{dA_2^*(a, b)}{db} \geq 0$$

The result above summarizes the idea the choice of similar programs leads to lower advertising rates.

Definition 4

$$\Phi^1(A_1^*(a, b), A_2^*(a, b), a, b) \equiv A_1^*(a, b)\, h(A_1^*(a, b), A_2^*(a, b), a, b)$$

and:

$$\Phi^2(A_1^*(a, b), A_2^*(a, b), a, b) \equiv$$
$$\equiv A_2^*(a, b)[S - h(A_1^*(a, b), A_2^*(a, b), a, b)]$$

We can now state the following:

Assumption 5

The functions:

$$\Phi^1(A_1^*(a, b), A_2^*(a, b), a, b)$$

and

$$\Phi^2(A_1^*(a, b), A_2^*(a, b), a, b)$$

are concave with respect to a and b (respectively).

Therefore, the interior solution to the following problem:

$$\max_a \Phi^1(A_1^*(a, b), A_2^*(a, b), a, b)$$

$$\max_b \Phi^2(A_1^*(a, b), A_2^*(a, b), a, b)$$

is characterized by the following first-order conditions:

$$\frac{\partial \Phi^1}{\partial A_2^*} \frac{dA_2^*}{da} + \frac{\partial \Phi^1}{\partial a} = 0$$

$$\frac{\partial \Phi^2}{\partial A_1^*} \frac{dA^*}{db} + \frac{\partial \Phi^2}{\partial b} = 0$$

As it is well known, the first term in these conditions corresponds to the "strategic effect" while the second one represents the "direct effect" of a change in the program type [9]. The main effect of advertising rate regulation consists in eliminating the "strategic effect" from the decision problem of the firms. This is intuitively simple, since if each firm expects the other one to stick to the advertising rate determined by the authority, there is no reason to use differentiation in order to "soften" a competition on advertising rates which regulation has suppressed. However, by assumption I the signs of the "direct effects" are:

$$\frac{\partial \Phi^1}{\partial a} = h_a > 0 \qquad \frac{\partial \Phi^2}{\partial b} = -h_b < 0$$

[9] See Bulow-Geanakoplos-Klemperer (1985) and Fudenberg-Tirole (1984).

This implies that, once advertising rate competition disappears, the two firms tend to select very similar program types. Differentiation, however, does not entirely disappear, since when program types are too similar, regulation may cease to be binding, i.e. equilibrium advertising rates are lower than the maximum rate imposed by the authority. In the next subsection I shall expose the reasoning above in a more formal way.

II.2 *Advertising Rate Regulation*

In this subsection, I focus on the positive effects of the regulating authority fixing the advertising rates on the selection of program types made by a profit maximizing firm. Differently from subsection I.3, I shall not discuss the optimal configuration of program types and advertising rates.

Regulation takes the following form:

(i) the regulating authority fixes the level of $\bar{A}$, the upper bound on advertising rates common to both firms;

(ii) firms determine the values of 'a' and 'b', under the constraint that:

$$A_1, A_2 \leq \bar{A}$$

Obviously, this regulatory constraint might in principle be binding for none of the firms, for only one of them or for both.

Definition 3

$$A^*_1(\bar{A}, a, b) \equiv \arg\max_{A_1} A_1 b(A_1, \bar{A}, a, b)$$

$$A^*_2(\bar{A}, a, b) \equiv \arg\max_{A_2} A_2 [S - b(\bar{A}, A_2, a, b)]$$

Three outcomes of regulation can be considered:

– Non-binding regulation: $(A^*_1(a, b), A^*_2(a, b)) \leq \bar{A}$

– One-sided (j) binding regulation (OS(j)R): $A^*_j(a, b) > \bar{A}$ and $A^*_i(\bar{A}, a, b) \leq \bar{A}$

– Two-sided binding regulation (TSR): $(A^*_1(a, b), A^*_2(a, b)) \geq \bar{A}$

I now define the notion of a "two-sided binding regulated" (TSR) equilibrium. For this I need:

Definition 4

$$C^a(\bar{A}, b) \equiv \{a: A_1^*(a, b) \geq \bar{A}\} \quad \text{(i)}$$

$$UC^a(\bar{A}, b) \equiv \{a: A_1^*(a, b) \leq \bar{A}\} \quad \text{(ii)}$$

$$C^b(\bar{A}, a) \equiv \{b: A_2^*(a, b) \geq \bar{A}\} \quad \text{(iii)}$$

$$UC^b(\bar{A}, a) \equiv \{b: A_2^*(a, b) \leq \bar{A}\} \quad \text{(iv)}$$

$$OSc^a(\bar{A}, b) \equiv \{a: A_1^*(\bar{A}, a, b) \leq \bar{A}\} \quad \text{(v)}$$

$$OSc^b(\bar{A}, a) \equiv \{b: A_2^*(\bar{A}, a, b) \leq \bar{A}\} \quad \text{(vi)}$$

The correspondences in definition 4 are fairly simple to interpret. $C^j(\bar{A}, i)$ is the "constrained set" of j values (j: a, b) such that the regulatory constraint is binding for firm j for a given value of i (i: a, b; $i \neq j$). $UC^j(\bar{A}, i)$ is the "unconstrained set" of j values, for a given value of i, such that the regulatory constraint is satisfied (under the assumption that both firms are unconstrained). $OSc^j(\bar{A}, i)$ instead is the "one-sided unconstrained set" of 'j' values such that (for a given value of 'i'), the advertising rate freely chosen by the firm selecting j is smaller than or equal to $\bar{A}$, assuming that the other firm is constrained to choose $A_i = \bar{A}$.

Definition 5

A "two-sided binding regulated" (TSR) equilibrium is pair $(\tilde{a}, \tilde{b})$ such that:

$$\left.\begin{aligned} &\Phi^1(\bar{A}, \bar{A}, \tilde{a}, \tilde{b}) \geq \Phi^1(\bar{A}, \bar{A}, a, \tilde{b});\ a, \tilde{a} \in C^a(\bar{A}, \tilde{b}) \\ &\Phi^2(\bar{A}, \bar{A}, \tilde{a}, \tilde{b}) \geq \Phi^2(\bar{A}, \bar{A}, \tilde{a}, b);\ b, \tilde{b} \in C^b(\bar{A}, \tilde{a}) \end{aligned}\right| \quad (\alpha)$$

$$\Phi^1(\bar{A}, \bar{A}, \tilde{a}, \tilde{b}) \geq \Phi^1(A_1^*(\bar{A}, a, \tilde{b}), \bar{A}, a, \tilde{b});\ a \in OSc^a(\bar{A}, \tilde{b}),\ \tilde{b} \in C^b(\bar{A}, a) \quad (\beta)$$

$$\Phi^2(\bar{A}, \bar{A}, \tilde{a}, \tilde{b}) \geq \Phi^2(\bar{A}, A_2^*(\bar{A}, \tilde{a}, b), \tilde{a}, b);\ b \in OSc^b(\bar{A}, \tilde{a}),\ \tilde{a} \in C^a(\bar{A}, b)$$

$$\Phi^1(\bar{A}, \bar{A}, \tilde{a}, \tilde{b}) \geq \Phi^1(\bar{A}, A_2^*(\bar{A}, a, \tilde{b}),\ \tilde{b} \in OSc^b(\bar{A}, a),\ a \in C^a(\bar{A}, \tilde{b}) \quad (\gamma)$$

$$\Phi^2(\bar{A}, \bar{A}, \tilde{a}, \tilde{b}) \geq \Phi^2(A_1^*(\bar{A}, \tilde{a}, b)\, \bar{A}, \tilde{a}, b);\ \tilde{a} \in OSc^a(\bar{A}, b),\ b \in C^b(\bar{A}, \tilde{a})$$

$$\Phi^1(\bar{A}, \bar{A}, \tilde{a}, \bar{b}) \geq \Phi^2(A_1^*(a, \bar{b}), a, \bar{b}); \; a \in UC^a(\bar{A}, \bar{b}), \; \bar{b} \in UC^b(\bar{A}, a) \qquad (\delta)$$

$$\Phi^2(\bar{A}, \bar{A}, \tilde{a}, \bar{b}) \geq \Phi^2(A_1^*(\tilde{a}, b), A_2^*(\tilde{a}, b), \tilde{a}, b); \quad b \in UC^b(\bar{A}, \tilde{a}), \; \tilde{a} \in UC^b(\bar{A}, b)$$

I shall now show in a simple case how the above defined TSR equilibrium looks like as compared with an unrestrained equilibrium. For this purpose I need:

Assumption 6

For any 'b' and $\bar{A}$, the correspondences

$$T^a(b) \equiv C^a(\bar{A}, b) \cap UC^a(\bar{A}, b) \quad \text{and} \quad F^a(b) \equiv \{a: A_1^*(\bar{A}, a, b) = \bar{A}\}$$

are single-valued functions [10]. The two functions, $T^a(b)$ and $F^a(b)$ have therefore the following properties.

$$A_1^*(T^a(b), b) = \bar{A} \qquad A_1^*(\bar{A}, F^a(b), b) = \bar{A}.$$

Similarly, we can define $T^b(a)$ and $F^b(a)$ so that:

$$A_2^*(a, T^b(a)) = \bar{A}. \qquad A_2^*(\bar{A}, a, F^b(a)) = \bar{A}.$$

From Comparative Statics Result II we know that:

$T^i(j)$, $F^i(j)$ $\quad i, j = a, b$ are increasing functions of their arguments.

for any $\tilde{a} \leq T^a(b)$ $\quad A_1^*(\tilde{a}, b) \geq \bar{A}$

for any $\tilde{a} \leq F^a(b)$ $\quad A_1^*(\bar{A}, \tilde{a}, b) \geqq \bar{A}$

for any $\bar{b} \geq T^b(a)$ $\quad A_2^*(a, \bar{b}) \geq \bar{A}$

for any $\bar{b} \geq F^b(a)$ $\quad A_2^*(\bar{A}, a, \bar{b}) \geq \bar{A}$

I shall restrict to symmetric equilibria, i.e., situations where:

(i) $A_1^*(a, b) = A_2^*(a, b)$ iff $a = S - b$

(ii) $a^* = S - b^*$

where:

$$a^* \equiv \arg\max_a \Phi^1(A_1^*(a, b), A_2^*(a, b)\, a, b)$$

$$b^* \equiv \arg\max_b \Phi^2(A_1^*(a, b), A_2(a, b)\, a, b)$$

[10] Both functions depend also on $\bar{A}$, but such dependence will not appear in the notation for simplicity purposes.

(iii) $\bar{a} = S - \bar{b}$ where $\bar{a}$ and $\bar{b}$ are the TSR equilibrium solution.

Now look at Figure 1: since we consider a situation where $a < b$, then only the area below the 45°-line is relevant. The segment starting from S on the horizontal axis and hitting the 45°-line at $b = S/2$ represents the locus of symmetric program types (in the sense that $a = S - b$) [11].

As a consequence, the areas in the figure represent the correspondences appearing in definition 4 as follows:

$$\Gamma = [a, b: a \in UC^a(b, \bar{A}); b \in UC^b(a, \bar{A})]$$
$$\Delta = [a, b: a \in UC^a(b, \bar{A}); b \in C^b(a, \bar{A})]$$
$$\Lambda = [a, b: a \in C^a(b, \bar{A}); b \in UC^b(a, \bar{A})]$$
$$\Psi = [a, b: a \in OSc^a(b, \bar{A}); b \in C^b(a, \bar{A})]$$
$$\Theta = [a, b: a \in C^a(b, \bar{A}); b \in OSc^b(a, \bar{A})]$$
$$\Sigma = [a, b: a \in C^a(b, \bar{A}); b \in C^b(a, \bar{A})]$$

If regulation is binding for both firms, then the unregulated equilibrium (a^*, b^*) must be somewhere in the Σ area; (a limiting case is the one where (a^*, b^*) coincides with the point E) [12]. If regulation is introduced and is binding for both at the unregulated equilibrium, each firm expects the opponents' advertising rate to equal $\bar{A}$. Therefore the "strategic" terms

$$\frac{\partial \Phi^1}{\partial A_2^*} \frac{dA_2^*}{da} \qquad \frac{\partial \Phi^2}{\partial A_1^*} \frac{dA_1^*}{db}$$

disappear from the first-order conditions for problem (II.2), which reduce to:

$$\frac{\partial \Phi^1}{\partial a} > 0 \qquad \frac{\partial \Phi^2}{\partial b} < 0$$

This means that profit maximization (when both firms are regulation-constrained) induces the two firms to choose very similar program types, relative to those chosen in the unrestrained equilibrium, i.e. both firms "run away" from the "extreme" combination $(0, S)$ in the direction of point $(S/2, S/2)$. The next step is to show that, if a TSR symmetric equilibrium

[11] See Appendix 2 for a discussion of how Figure 1 has been drawn.

[12] In the special case of maximum differentiation, (Economides, 1986) $a^* = 0$, $b^* = S$. This would be the case if the utility function of the viewer $u(A, |z - j|)$ were of the type $-(A + (z-j)^2)$.

exists, then it still entails some degree of program differentiation. More precisely, this equilibrium will be point (a^E, b^E).

Take the point of view of firm 1. Given b^E, the choice of a value $a' \leqq a^E$ implies $a', b^E \in \Gamma$.

In this case, however, regulation is no longer binding for any firm and both would engage in an unrestrained advertising rate competition; the same can be said as regards a decrease in b relative to b^E. If we are considering a TSR equilibrium, then according to condition (δ) no firm finds it profitable to shift to such a region. This is why differentiation does not entirely disappear.

Under the symmetry assumptions made so far, $OSc^a(\bar{A}, b^E) = a^E$, $OSc^b(\bar{A}, a^E) = b^E$. Then in our graphic example, the unique program type which allows firm 1 to shift to a OS(2)R or to a OS(1)R configuration is still a^E (if b^E is the choice of the other firm). Since $A_j^*(\bar{A}, a^E, b^E) = \bar{A}$ (for any j), then profits accruing to both firms under OS(1)R or OS(2)R equal profit from TSR. This fulfils conditions (β) and (γ) [13].

To recapitulate on Section II, the main result here is that advertising rate regulation reduces the degree of program differentiation. This is due to the fact that if regulations is binding, each firm does not fear a reduction in the advertising rate which would be the optimal reaction of the opponent to the selection of a program type exceedingly close to its own.

III. *Concluding Comments*

The main conclusions of this paper have been summarized in the final lines of Section I and Section II. In this Section I shall simply comment on the way the results have been obtained on the general approach used in the essay. The basic idea upon which most of the paper rests is that decisions concerning advertising rates (by broadcasting firms) resemble price-making decisions: revenue from advertising equals the advertising rate times the audience; as the former increases, the latter decreases, since viewers shift to less advertising-crowded programs or to alternative uses of leisure (I ruled out this second possibility in the present essay). The relationship between advertising rates and other program characteristics can be modeled in the same way as the relationship between price and other relevant characteristics in models with differentiated products. As a consequence, even if in a "first-

[13] There remains the possibility of asymmetric equilibria in the Θ and Ψ regions, but I do not examine this development here.

best" setting, the socially optimum advertising rate were smaller than the equilibrium one, the simple imposition of a binding upper constraint on advertising rates would not necessarily be beneficial: since the choice of the advertising rate strategically interacts with other decision variables at the firms' level, such a form of regulation might adversely affect the values of these variables.

Under the assumptions made in this essay it is impossible to say anything more about second-best optimal configurations. Some positive results concerning the impact of regulation on specific program characteristics are contained in Comparative Statics Results I and II.

If, as it happens under complete regulation, the authority controls instead all of the relevant program characteristics, then not surprisingly the socially optimal solution is much easier to characterize. Some broad additional assumptions on technology and preferences are sufficient to produce specific normative results.

APPENDIX 1

Let us consider case B (see subsection I.1). Again, I shall neglect stage I choices. The appendix verifies that the Nash equilibrium for the game in stage II of case A coincides with the subgame perfect equilibrium for the stage II-stage III game of case B.

In the latter case, the decision problem for the two firms in stage III is:

$$\max_{A_1} \Pi^1(A_1, A_2, x_1, x_2)$$
$$\max_{A_2} \Pi^2(A_1, A_2, x_1, x_2) \tag{A.1}$$

First-order conditions for the problem above are:

$$b + A_1 \cdot b_1 \cdot g(x_1) = 0 \tag{A.1a}$$

$$S - b - A_2 \cdot b_2 \cdot g(x_2) = 0 \tag{A.1b}$$

Assumption A1

For any $X \equiv (x_1, x_2)$, there exists one and only one Nash equilibrium in the advertising rates, i.e. pair of values for (A_1, A_2):

$$A_1^*(X) \qquad A_2^*(X)$$

which solves the system of first-order conditions (I.4a-b).

Comparative Statics Result AI

By differentiating (I.4a-b), we find:

$$\frac{dA_i^*(X)}{dx_i} = -\frac{g'(x_i)\cdot A_i}{g(x_i)} \qquad \text{(i)}$$

$$\frac{dA_i^*(X)}{dx_j} = 0 \ (j \neq i) \qquad \text{(ii)}$$

The above result has an important implication. Equality (ii) means that the choice of, say, x_i has no strategic effect on the level of A_j in the subsequent stage. Therefore, the first-order condition on x_i is the same as if A_j were given, i.e. as if x_i and A_j were determined simultaneously.

The decision problem in stage II can be represented as follows:

$$\max_{x_1} \Pi^1[A_1^*(x_1, x_2), A_2^*(x_1, x_2), x_1, x_2] = A_1^*(x_1, x_2)$$
$$b(A_1^*(x_1, x_2)\, g(x_1), A_2^*(x_1, x_2)\, g(x_2)) - c(x_1) \qquad \text{(A.2)}$$

$$\max_{x_1} \Pi^2[A_1^*(x_1, x_2), A_2^*(x_1, x_2), x_1, x_2] = A_2^*(x_1, x_2)$$
$$[S - b(A_1^*(x_1, x_2)\, g(x_1), A_2^*(x_1, x_2)\, g(x_2))] - c(x_2)$$

And under customary concavity assumptions, an interior solution can be characterized by the following first-order conditions:

$$\frac{\partial \Pi^1}{\partial A_1}\frac{dA_1^*}{dx_1} + \frac{\partial \Pi^1}{\partial A_2}\frac{dA_2^*}{dx_1} + \frac{\partial \Pi^1}{\partial x_1} = 0 \qquad \text{(A2.a)}$$

$$\frac{\partial \Pi^2}{\partial A_2}\frac{dA_2^*}{dx_2} + \frac{\partial \Pi^2}{\partial A_1}\frac{dA_1^*}{dx_2} + \frac{\partial \Pi^2}{\partial x_2} = 0 \qquad \text{(A2.b)}$$

However:

$\frac{\partial \Pi^i}{\partial A_i} = 0$ by the envelope theorem, while: $\frac{\partial A_i^*}{\partial x_j} = 0$ by Result AI.

The two conditions above reduce however to the following:

$$\frac{\partial \Pi^1}{\partial x_1} = [A_1^*(x_1, x_2)]^2 b_1 g'(x_1) - c'(x_1) = 0 \qquad \text{(A2.c)}$$

$$\frac{\partial \Pi^2}{\partial x_2} = -[A_2^*(x_1, x_2)]^2 b_2 g'(x_2) - c'(x_2) = 0 \qquad \text{(A2.d)}$$

It therefore turns out, by comparing conditions (I.3a-d) to conditions (A1.a-b) and (A2.c-d), that the equilibrium configuration is the same in case A as in case B.

APPENDIX 2

In this appendix I partly justify the way Fig. 1 has been drawn. In fact I do not explain one feature of Fig. 1, i.e. the fact that $T^b(T^a(b)) \leq b$ only for $b \geq \frac{S}{2}$. In principle, there

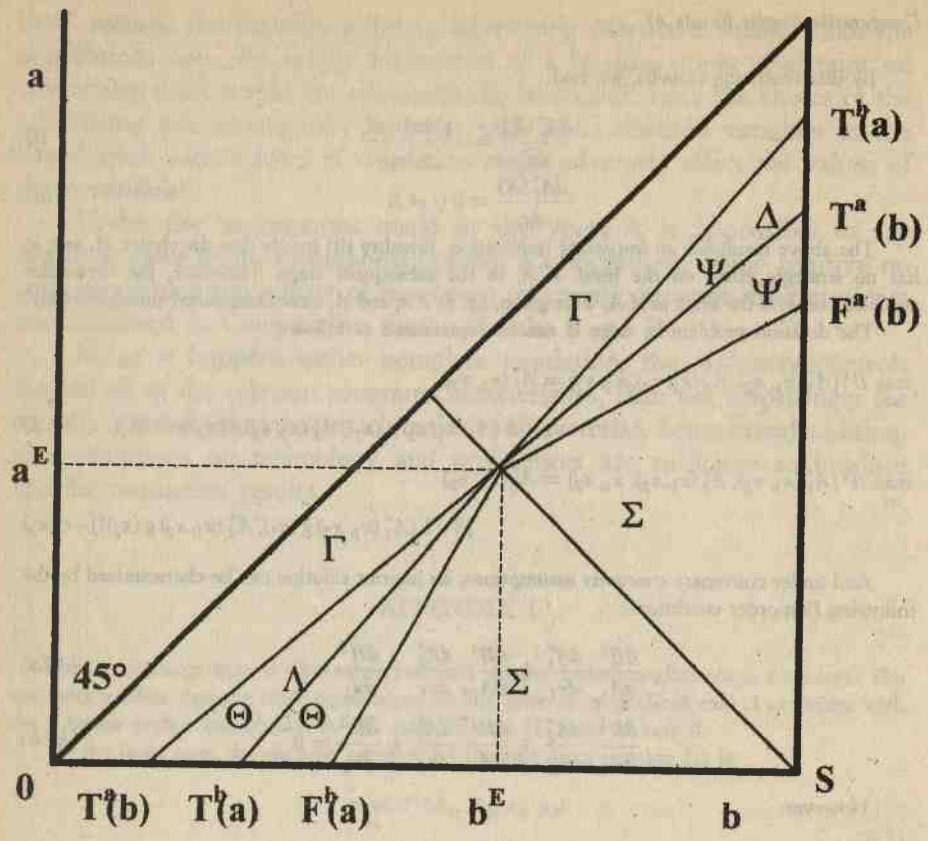

FIGURE 1

are no reasons to assume that this is the case. However, given the limitation of the discussion to symmetric equilibria, this feature of Figure 1 is not really relevant.

What I want to show is that if, for some $\bar{b}$, $A_2^*(T^a(\bar{b}), \bar{b}) \geq \bar{A}$, then $F^a(\bar{b}) < T^a(\bar{b})$. First, I shall show that if, for some a, b, $A_2^*(a, b) \geq \bar{A}$, then $A_1^*(\bar{A}; a, b) \leq A_1^*(a, b)$. In fact, $A_1^*(a, b)$, $A_2^*(a, b)$ can be regarded as the solution to the system:

$$A_1 = A_1^*(A_2, a, b) \qquad A_2 = A_2^*(A_1; a, b)$$

Under regulation, if $A_2^*(a, b) \geq \bar{A}$, then $A_2 = \bar{A}$. Since advertising rates are strategic complements (by assumptions 4 (iii) and (iv)):

$$A^*{}_1(\bar{A}; a, b) \leq A_1^*(A_2^*(a, b); a, b) = A_1^*(a, b)$$

Now suppose that for some $\bar{b}$: $A_2^*(T^a(\bar{b}), \bar{b}) \geq \bar{A}$. Then it must be $A_1^*(\bar{A}; T^a(\bar{b}), \bar{b}) \leq \bar{A}$. This implies that $A_1^*(\bar{A}; a, \bar{b}) = \bar{A}$ only if $a \leq T^a(\bar{b})$. But since: $A_1^*(\bar{A}; F^a(\bar{b}), \bar{b}) = \bar{A}$, it turns out that: $F^a(\bar{b}) \leq T^a(\bar{b})$.

Similarly, if we assume $A_2^*(T^a(\bar{b}), \bar{b}) \leq \bar{A}$ it is easy to show in the same way that $F^a(\bar{b}) \geq T^a(\bar{b})$. In this case however, it would be $A_2^*(F^a(\bar{b}), \bar{b}) \leq \bar{A}$ and therefore in general, for any $a \geq F^a(\bar{b})$, regulation would be binding for none firm. That is to say, if for some $\bar{b}$, $A_2^*(T^a(\bar{b}), \bar{b}) \leq \bar{A}$ then at $b = \bar{b}$ there exists no equilibrium where firm 2 is regulation-constrained and firm 1 is not.

REFERENCES

Barnett H.J.-Greenberg E., "TV Program Diversity - New Evidence and Old Theories", *American Economic Review*, May 1971, *61*, 89-93.

Bowman G., "Consumer Choice and Television", *Applied Economics*, 1975, *7*, 175-84.

Bulow J.I.-Geanakoplos J.D.-Klemperer P.D., "Multimarket Oligopoly: Strategic Substitutes and Complements", *Journal of Political Economy*, No. 3, 1985, *93*, 488-511.

Dunnett P., *The World Television Industry. An Economic Analysis*, London: Routledge, 1990.

Economides N., "Minimal and Maximal Product Differentiation in Hotelling's Duopoly", *Economics Letters*, 1986, *21*, 67-71.

Evans D.-Fisher F.M.-McGowan J., "The Audience-Revenue Relationship for Local Television Stations", in J. Monz, ed., *Industrial Organization Economics and the Law. Collected Papers of F.M. Fisher*, New York: Harvester-Wheatsheaf, 1990.

Fudenberg D.-Tirole J., "The Fat-Cat Effect, the Puppy Dog Ploy and the Lean and Hungry Look", *American Economic Review*, No. 2, 1984, *74*, 361-66.

Gambaro M.-Silva F., *Economia della televisione*, Bologna: Il Mulino, 1992.

MacLeod W.B., "On the Non-existence of Equilibria in Differentiated Product Models", *Regional Science and Urban Economics*, 1986, *15*, 245-62.

Spence M.-Owen B., "Television Programming, Monopolistic Competition and Welfare", *Quarterly Journal of Economics*, No. 2, 1977, 103-26.

Sutton J., *Sunk Costs and Market Structure*, Cambridge: MIT Press, 1991.

UN MODELLO DELL'AUDIENCE TELEVISIVO: IMPLICAZIONI PER LA CONCORRENZA E LA REGOLAMENTAZIONE DELLA PUBBLICITÀ

Questo saggio sviluppa la semplice idea che l'ascolto di un determinato programma possa essere negativamente influenzato dall'entità del tempo destinato alla trasmissione di pubblicità, nell'ambito del programma stesso. Se si accetta questa ipotesi, allora la decisione dell'emittente circa il tempo da destinare alla pubblicità riproduce alcune caratteristiche proprie della decisione sul prezzo da parte di imprese in mercati non perfettamente concorrenziali. Nell'articolo, que-

sta intuizione viene formalizzata ed estesa allo scopo di studiare anche le interrelazioni tra la decisione circa il tempo destinato a pubblicità e altre decisioni strategiche (in particolare, quelle riguardanti tipo e qualità dei programmi). Un modello di competizione duopolistica viene sviluppato e i suoi esiti di equilibrio vengono paragonati a quelli derivanti dall'imposizione di differenti forme di regolamentazione.

Rivista Internazionale di Scienze Economiche e Commerciali
Volume 42 (1995), N. 1-2, 57-67

TAXATION, LABOR SUPPLY, AND SAVINGS: AN INTERTEMPORAL ANALYSIS

by
THEODORE PALIVOS * and CHONG K. YIP **

I. *Introduction*

Over the last two decades, there has been a considerable amount of work on the effects of taxation on labor supply and savings [1]. This reflects the practical importance of the issue as well as the theoretical and empirical challenges in the area. For example, the effects of labor-income tax on aggregate labor supply is "at the core of debates between proponents of supply-side and Keynesian approaches to government fiscal policy" (Snow and Warren, 1989). Moreover, a strong disincentive of higher taxation on labor supply is often assumed to generate the downturn of the Laffer curve (see for example, Beenstock, 1979).

A standard framework which is often employed in the literature consists of a two-period model in which the representative individual decides upon her/his consumption profile and hours of work in one of the two periods [2]. The effects of different types of taxation (e.g., a consumption tax, a wage tax, and an interest income tax, etc.) are then examined within this integrated framework [3]. In reality, however, a tax policy is often temporary,

* Louisiana State University, Department of Economics, Baton Rouge, LA (U.S.A.).
** Georgia State University, Department of Economics, Atlanta, GA, and Chinese University of Hong Kong, Department of Economics, Hong Kong.

[1] For comprehensive surveys of the studies on labor supply and savings, see HAUSMAN (1985) and SANDMO (1985), respectively.

[2] If the individual works in the first period, then the second is regarded as her/his retirement period (SANDMO, 1985). If, on the other hand, she/he is allowed to work only in the second period, then the first is considered to be the training period (KOTLIKOFF and SUMMERS, 1979).

[3] This framework is a synthesis of two different approaches: A pure consumption approach, which examines the tax effects on savings assuming that labor income is exogenous,

i.e., it only lasts for a certain period, and/or it is anticipated, i.e., it is announced in advance. It is, therefore, important to incorporate these aspects of tax policies into the analysis. To our knowledge, this has not been considered in the existing literature.

Accordingly, in this paper, we study the effects of temporary and anticipated (labor and property) income tax policies on savings and labor supply. For this, we develop a simple two-period model in which the representative individual is provided with a labor-leisure choice in both periods [4]. Within such a framework, a temporary tax is one which is imposed just in the first period while an anticipated tax is one which is announced in the first but imposed in the second period. Such a distinction in the nature of the taxes allows us to highlight their intertemporal effects on the individual's optimal behavior. The analysis is focused on the derivation of empirical hypotheses concerning the effects of alternative tax systems. Our main finding is that the effects of a temporary tax on savings and labor supply are drastically different from those of an anticipated tax. This justifies our distinction between the two types of taxation. In general, a temporary wage tax reduces savings and has an ambiguous effect on labor supply. A temporary tax on interest income also reduces savings but enhances labor supply. An anticipated tax on labor income, on the other hand, increases savings and labor supply while an anticipated tax on interest income has ambiguous effects. Furthermore, we find that labor supply and savings are more responsive to a labor than an interest income tax. In particular, a switch from an anticipted (temporary) interest income tax to an anticipated (temporary) labor income tax will enhance (reduce) both labor supply and savings.

The remainder of the paper is organized as follows. The next section develops our theoretical framework. Section III performs some comparative statics exercises using the duality approach, while Section IV provides a comparison among various types of taxation. Finally, Section V concludes the paper with some suggestions for future research.

II. *The Model*

Consider an economy in which individuals live for two periods (0 and

---

and a pure labor supply approach which takes the savings decision to be exogenous or at least separable from the endogenous labor-leisure choice. For criticisms of both approaches see SANDMO (1985).

[4] Since we want to study the intertemporal effects of an anticipated wage tax, we allow the individual to work in both periods. This is also consistent with the life cycle model of labor supply (e.g., MACURDY, 1981).

1), consuming goods and leisure in each period. The representative agent is assumed to possess a well-behaved lifetime utility function

$$U = U(C^0, C^1, 1-L^0, 1-L^1) \tag{1}$$

where C^t and L^t, $t = 0$ or 1, denote consumption and hours of work in period t, respectively. Notice also that the time endowment in each period has been normalized to one. Furthermore, the following constraint faced by the individual describes the motion of her/his net wealth:

$$A^{t+1} - A^t = rA^t + wL^t - C^t - T^t \tag{2}$$

where A^t is the holdings of a single financial asset (in the beginning of period t) which serves as a storage of wealth; r and w denote the exogenously given real interest rate and real wage rate, respectively, which are assumed not to vary across periods [5]. Finally, T^t denotes total tax payments made in period t. We further assume that credit markets are perfect and so individuals are able to borrow and lend freely at the rate r. If we let τ^t and θ^t be the marginal tax rates on interest and labor income respectively, then

$$T^t = \tau^t rA^t + \theta^t wL^t \qquad t = 0, 1 \tag{3}$$

For simplicity, the tax schedules are assumed to be linear, i.e., the tax rates do not vary with the size of the tax base, and intertemporally separable in the sense that tax payments in each period are based on that period's income only and not on income in other periods [6]. Furthermore, the initial asset holdings (endowment) of the individual, A^0, is given exogenously. It should be noted that this endowment can be either positive or negative and that if it is positive, it will yield an interest income (i.e., property income or pure rent), rA^0, which is taxable. Finally, without loss of generality, we assume that $A^2 = 0$, i.e., the individual leaves no bequests [7].

Let $D = [1 + (1 - \tau^1) r]^{-1}$ be the (net) discount factor. Then the (present value) expenditure function corresponding to the lifetime utility in (1) can be written as

---

[5] The exogeneity assumption of factor prices is standard in the labor economics literature; for example, see MaCurdy (1981). Notice also that, although r and w are taken to be constant over time, the net interest rate and the net wage rate will vary as long as the marginal tax rates on interest and labor income are allowed to vary across periods.

[6] For analyses involving non-linear or intertemporally non-separable tax functions, see Blomquist (1988) and MaCurdy (1983), respectively.

[7] At the expense of simplicity, it is straightforward to allow for bequests; for example, see Heckman (1976).

$$E(1, D, (1-\theta^0)w, D(1-\theta^1)w, U) = \min[C^0 + DC^1 + (1-\theta^0)w(1-L^0) + D(1-\theta^1)w(1-L^1): U(\cdot) \geq U] \quad (4)$$

By standard properties of the expenditure function, the demands for consumption goods and leisure are given by the following equations:

$$C^0 = E_1,\ C^1 = E_2,\ 1-L^0 = E_3,\ 1-L^1 = E_4 \quad (5)$$

where subscripts denote partial derivatives. Let $S = A^1 - A^0$ denote savings. Then from equation (2) and (3), we have

$$S = (1-\tau^0)rA^0 + (1-\theta^0)w(1-E_3) - E_1 \quad (6)$$

Finally, using the budget constraint for each period (equation (2)), one can derive the intertemporal budget constraint, which simply states that the present value of individual's total expenditure (minimized in (4)) is equal to the present value of her/his "full wealth":

$$E(1, D(1-\theta^0)w, D(1-\theta^1)w, U) = [1 + (1-\tau^0)r]A^0 + (1-\theta^0)w + D(1-\theta^1)w \quad (7)$$

In order to keep the comparative statics tractable, we assume that consumption and leisure in both periods are normal goods. As standard in the literature, consumption and leisure across periods are assumed to be Hicksian substitutes, i.e., $E_{12} > 0$ and $E_{34} > 0$, while consumption and leisure within the same period are taken to be complements, i.e., $E_{13} < 0$ and $E_{24} < 0$ [8].

III. *Comparative Statics Analysis*

This section derives the comparative statics effects on current labor supply and savings of two types of factor taxation; namely, labor and interest income taxes. Furthermore, we distinguish taxes according to their timing. We analyze effects of temporary (unanticipated) taxes that last only for one period (period 0) and anticipated taxes that are known to take place in the next period (period 1) [9]. These types of taxes are of practical

[8] These assumptions together with the homogeneity (of degree zero) of demand functions imply that $E_{14} > 0$ and $E_{23} > 0$.

[9] The effects of a permanent tax (which lasts for both periods) can easily be derived by combining the effects of the temporary and anticipated taxes.

importance because in reality most of the tax policies can be classified in one of the two categories.

First, we totally differentiate (7) to obtain

$$E_u dU = -rA^0 d\tau^0 - (r/D^2)[E_2 - (1-\theta^1)w(1-E_4)]d\tau^1 - (1-E_3)wd\theta^0 - (1-E_4)wDd\theta^1 \quad (8)$$

Next, we differentiate the equations giving the optimal leisure and saving decisions ((5) and (6)), and then replace the terms involving dU using (8). This yields the comparative statics results which we discuss in detail below.

A. Temporary-unanticipated taxes. – Consider first the effects of a temporary tax policy on current labor supply (L^0) and savings (S) decisions. Differentiation of (5) and (6) yields the following effects of a temporary increase in the labor income tax rate (a rise in Θ^0):

$$\frac{dL^0}{d\theta^0} = wE_{33} + [(1-E_3)wE_{3u}/E_u] \gtrless 0 \quad (9)$$

$$\frac{dS}{d\theta^0} = w\left\{[(1-\theta^0)wE_{33} + E^{13}] - (1-E_3)\left[1 - \frac{(1-\theta^0)wE_{3u} + E_{1u}}{E_u}\right]\right\} < 0 \quad (10)$$

The duality approach provides a convenient way to perform the Slutsky decomposition and thus to interpret the comparative statics. The first term on the right-hand-side in (9) is the negative substitution effect of a temporary unanticipated increase in the labor income tax rate. The second term, on the other hand, is the income effect caused by an increase in θ^0, which raises labor supply (decreases consumption of leisure) provided that leisure is a normal good. Since the two effects work in opposite directions, the net effect on labor supply of a temporary increase in the wage tax is ambiguous. As far as the savings decision is concerned, equation (10) implies that an increase in θ^0 reduces savings since both the substitution (the first term) and the income effect (the second term) are negative [10].

The effect on current labor supply of a temporary increase in the interest tax rate is similar to a decrease in initial wealth since it only affects L^0 via the positive income effect (negative on leisure):

[10] The income effect (second term in (10)) is negative since (by linear homogeneity of E_u) it is equal to $-(1-E_3)(D/E_u)[E_{2u} + E_{4u}(1-\theta^1)w] < 0$.

$$\frac{dL^0}{d\tau^0} = rA^0 E_{3u}/E_u > 0$$

Finally, the effect on savings of a temporary increase in the tax rate on interest income is negative:

$$\frac{dS}{d\tau^0} = -rA^0\left[1 - \frac{(1-\theta^0)\,wE_{3u} + E_{1u}}{E_u}\right] < 0$$

which is a fairly standard conclusion (e.g., see Kotlikoff and Summers, 1979).

B. Anticipated taxes. – Next we consider anticipated changes in the marginal tax rates on interest and wage income, i.e., changes in θ^1 and τ^1. A higher labor tax in the future generates intertemporal substitution of work effort across the two periods. This together with the standard positive wealth effect implies that an increase in θ^1 raises labor supply:

$$\frac{dL^0}{d\theta^1} = E_{34} + (1 - E_4)\,wDE_{3u}/E_u > 0$$

Similarly, the effect of an increase in θ^1 on savings is also positive because of the intertemporal substitution of labor (as well as consumption), which in turn raises the labor income in the first period and hence savings:

$$\frac{dS}{d\theta^1} = wD\left[[(1-\theta^0)\,wE_{34} + E_{14}] + (1-E_4)\,\frac{(1-\theta^0)\,wE_{3u} + E_{1u}}{E_u}\right] > 0$$

The effects of an anticipated increase in the interest income tax rate on labor supply and savings are ambiguous due to the fact that income and substitution effects always work in opposite directions. The main mechanism at work is the higher net discount factor (D) resulted from the lower net interest rate because of the increase in τ^1. This creates an intertemporal substitution effect towards future labor effort. Moreover, the increase in D also lowers real wealth (since D affects the prices of future consumption and leisure positively) which generates a positive income effect on current work effort by the normality of leisure. Thus, the overall effect of an increase in τ^1 on labor supply is indeterminate:

$$\frac{dL^0}{d\tau^1} = (r/D^2)\{-[(1-\theta^1)\,wE_{34} + E_{32}]$$

$$+ [E_2 - (1-\theta^1)\,w\,(1-E_4)]\,E_{3u}/E_u\} \gtrless 0$$

Likewise, the ambiguous effect on savings is also due to the presence of a negative substitution effect and a positive income effect:

$$\frac{dS}{d\tau^1} = (r/D^2)\{-\{E_{12} + (1-\theta^1)wE_{14} + (1-\theta^0)w[E_{32} + (1-\theta^1)wE_{34}]\} + [E_2 - (1-\theta^1)w(1-E_4)][(1-\theta^0)wE_{3u} + E_{1u}]/E_u\}$$

We summarize all the comparative statics results on labor supply and savings in Table 1 [11].

Finally, our analysis suggests that, when the timing of taxation is explicitly taken into account, the relation between tax revenues and tax rates may not follow the standard bell-shaped Laffer curve. Given w and r, an anticipated wage tax, for example, will always increase the tax revenue, while a temporary interest-income tax will decrease it.

IV. *Comparison of Effects*

From the results in Table 1, we see that the only tax policy that can promote savings is an anticipated increase in the tax rate on labor income. More specifically, an anticipated wage tax unambiguously increases savings while an anticipated tax on interest income can have an effect in either direction. Among these two tax policies, we would like to investigate which

TABLE 1

DYNAMIC EFFECTS OF TAXATION ON LABOR SUPPLY AND SAVINGS

Parameter Changing	Effects on		
	L^0	L^1	S
θ^0	?	+	−
τ^0	+	+	−
θ^1	+	?	+
τ^1	?	+	?

NOTES: θ, τ, L, and S denote a wage tax, an interest tax, labor supply, and savings, respectively. Superscripts denote periods (0 and 1).

[11] For the sake of completeness, we have also derived the effects on future labor supply (L^1) (see Table 1).

one has a larger quantitative impact. As it is easily understood, this can have important implications from the policy point of view. Following the standard practice in the public finance literature, we carry out our comparison within a differential incidence framework, in which alternative sources of the same amount of government revenue are contrasted [12]. Specifically, starting from an initial zero-tax economy, we introduce a particular tax and compute the corresponding equilibrium. The equilibrium is then recalculated with the first tax replaced by another with exactly equal revenue yield. In comparing τ^1 with θ^1, this equal-revenue restriction requires [13]

$$wD\,(1 - E_4)\,d\theta^1 = (r/D^2)\,[E_2 - (1 - \theta^1)\,w\,(1 - E_4)]\,d\tau^1$$

Let $dX|_i$ be the change in the variable X resulting from an infinitesimal increase in the tax rate ($X = S,\ L^0$ and $i = \tau^t,\ \theta^t$). Using the comparative statics results derived in Section III, we obtain

$$dS|_{\theta^1} - dS|_{\tau^1} = wD\,(E_{14} + wE_{34})\,d\theta^1 + (r/D^2)\,(E_{12} + wE_{14} + wE_{32} + w^2E_{34})\,d\tau^1 > 0$$

Thus savings is more responsive to a labor than to a property income tax. Similarly, one can show

$$dS|_{\theta^1} - dS|_{\tau^0} = wD\,(E_{14} + wE_{34})\,d\theta^1 + rA^0d\tau^0 > 0$$
$$dS|_{\theta^0} - dS|_{\tau^0} = w\,(E_{13} + wE_{33})\,d\theta^0 < 0$$

Among all four taxes, we conclude that the anticipated tax on labor income has the largest positive effect, since it is the only tax that creates a positive intertemporal substitution effect on savings. On the other hand, a temporary tax on labor income has the largest negative effect on savings.

Similar results are obtained for labor supply, as well. Table 1 shows that none of the four taxes considered here reduces current labor supply unambiguously. Thus, we have to make a pairwise comparison among these taxes to assess their corresponding quantitative impacts. Following the procedure we outline above, together with the comparative statics results on labor supply (L^0) derived in the previous section, one obtains

$$dL^0|_{\theta^1} - dL^0|_{\tau^0} = E_{34}d\theta^1 > 0$$
$$dL^0|_{\theta^1} - dL^0|_{\tau^1} = E_{34}d\theta^1 + (rw/D^2)\,(E_{32} + wE_{34})\,d\tau^1 > 0$$
$$dL^0|_{\theta^0} - dL^0|_{\tau^0} = wE_{33}d\theta^0 < 0$$

[12] See, for example SUMMERS (1981) or JUDD (1987).

[13] It should be noted that this is a first-order approximation similar to the one in BROWNING (1976).

Thus, labor supply decision is more responsive to an anticipated increase in the labor income tax rate.

Before closing this section, we would like to provide some discussion on the incidence of the various taxes that we study in the paper. Notice that only the direct effects of taxation appear in (8), which is a consequence of the envelope theorem. Thus, under the equal-revenue restriction, all taxes generate the same welfare loss.

V. *Concluding Remarks*

In this paper, we have developed an intertemporal model to examine the effects of various types of taxation on labor supply and on savings. Our findings are very different from the predictions of pure consumption and pure labor supply models [14]. In pure consumption models, for example, property income taxes reduce savings due the substitution between current and future consumption. In our framework, only an unanticipated temporary tax on property income unambiguously reduces savings while an anticipated tax on future interest income has an ambiguous effect. In the case of a wage tax, on the other hand, pure labor supply models predict that work effort is reduced due to the dominant substitution effect between work and leisure. On the contrary, our model does not imply an unambiguous negative effect of wage taxation on labor supply. Thus our results lend some support to Sandmo's proposal of providing an integrated intertemporal treatment of saving and labor supply decisions (see Sandmo 1985). Moreover, we have found that a wage tax (both anticipated and temporary) has a larger (absolute) effect on savings than a property income tax.

Finally we believe that our framework can admit a number of different extensions, and for brevity we outline just two of them. First, one can examine the effects of taxation in an environment which allows for uncertainty with respect to future income. An increase in the tax rate on labor (interest) income, in that case, will reduce not only the expected value of the net wage (interest rate) but also its variance. Another interesting extension would be to allow for non-linear tax functions. This would make the value of marginal tax rates endogenous, since they would depend on individual's labor supply and savings. We leave these and other important aspects of taxation as topics for future research.

[14] See SANDMO (1985) for a summary of the results obtained in these models.

REFERENCES

BEENSTOCK Michael, "Taxation and Incentives in the UK", *Lloyds Bank Review*, 1979, *134*, 1-15.

BLOMQUIST N. Sören, "Nonlinear Taxes and Labor Supply", *European Economic Review*, 1988, *32*, 1213-26.

BROWNING Edgar K., "The Marginal Cost of Public Funds", *Journal of Political Economy*, 1976, *84*, 283-98.

HAUSMAN Jerry A., "Taxes and Labor Supply", in A.J. Auerbach and M. Feldstein, eds., *Handbook of Public Economics*, Volume 1, New York: North Holland-Elsevier Science Publishers, 1985, 213-63.

HECKMAN James J., "A Life-Cycle Model of Earnings, Learning and Consumption", *Journal of Political Economy*, No. 4, part 2, 1976, *84*, S11-S44.

JUDD Kenneth L., "The Welfare Cost of Factor Taxation in a Perfect-Foresight Model", *Journal of Political Economy*, No. 4, 1987, *95*, 675-709.

KOTLIKOFF Laurence J. and SUMMERS Lawrence H., "Tax Incidence in a Life Cycle Model with Variable Labor Supply", *Quarterly Journal of Economics*, No. 4, 1979, *93*, 705-18.

MACURDY Thomas E., "An Empirical Model of Labor Supply in a Life-Cycle Setting", *Journal of Political Economy*, No. 6, 1981, *89*, 1059-85.

——, "A Simple Scheme for Estimating an Intertemporal Model of Labor Supply and Consumption in the Presence of Taxes and Uncertainty", *International Economic Review*, 1983, *24*, 265-89.

SANDMO Agnar, "The Effects of Taxation on Savings and Risk Taking", in A.J. Auerbach and M. Feldstein, eds., *Handbook of Public Economics*, Volume 1, New York: North Holland-Elsevier Science Publishers, 1985, 265-311.

SNOW Arthur and WARREN Ronald S., Jr., "Tax Rates and Labor Supply in Fiscal Equilibrium", *Economic Inquiry*, 1989, *27*, 511-20.

SUMMERS Lawrence H., "Capital Taxation and Accumulation in a Life Cycle Growth Model", *American Economic Review*, No. 4, 1981, *71*, 533-44.

IMPOSTE, OFFERTA DI LAVORO E RISPARMIO: UNA ANALISI INTERTEMPORALE

Questo articolo sviluppa un modello intertemporale per esaminare gli effetti sull'offerta di lavoro e il risparmio dell'imposta sui fattori produttivi. Vengono qui classificate le imposte secondo la scadenza (temporanee, cioè che durano per un certo periodo, o anticipate, cioè annunciate in anticipo) e la base impositiva (reddito da lavoro o reddito da interessi). Fra queste quattro imposte, si trova che una imposta anticipata sul reddito da lavoro ha grandissimi effetti positivi

sia sull'offerta di lavoro corrente sia sul risparmio. Un'imposta temporanea sul reddito da lavoro d'altro canto ha grandissimo effetto negativo sul risparmio, ma un effetto ambiguo sull'offerta di lavoro. Così, un cambiamento da una imposta anticipata (temporanea) sul reddito da interessi a un'imposta anticipata (temporanea) sul reddito da lavoro aumenterà (ridurrà) sia l'offerta di lavoro sia il risparmio.

Rivista Internazionale di Scienze Economiche e Commerciali
Volume 42 (1995), N. 1-2, 69-91

CORPORATE TAXATION AND MONOPOLY: THE EFFECTS OF AN ANNOUNCED REFORM

by
PAOLO M. PANTEGHINI *

0. *Introduction*

Imperfect competition has been studied from many points of view, both theoretical and empirical [1]. However, only a few studies deal with the effects of a change in the corporate income tax and the investment tax credit, under this market form. Economists have preferred to deal with the technological implications of a fiscal reform. Auerbach (1989), for example, explains the effects of short-run investment after the 1986 U.S. tax reform, by introducing an investment unit cost function, depending both on investment and capital stock. This technological assumption, implemented by many empirical works, is used to explain why current investment moves "in a different direction than that suggested by the long-run effects" (p. 939), when the reform is announced.

---

* Università di Brescia.

This paper is based on Chapter 4 of my doctoral dissertation. I whish to thank Aldo Chiancone, Antonio Guccione, Marzio Galeotti and an anonymous referee for their helpful comments. I am responsible for any errors in this paper.

[1] Some authors, such as MANKIW (1985) and BLANCHARD and KIYOTAKI (1987), introduce menu costs to explain price rigidity in a monopolistically competitive context. Others test the existence of monopolistic power in different countries. GALEOTTI and SCHIANTARELLI (1990), for example, find that "monopolistic competition is a pervasive feature of U.S. industries". The result confirms those obtained by HALL (1988, 1989) and CHIRINKO and FAZZARI (1988). As regards Great Britain, SCHIANTARELLI and GEORGOUTSOS (1990) replace the assumption of perfect competition in the output market with the assumption of monopolistic competition. They find that "a monopolistic competitive structure provides [...] a possible rationale as to why output has been found empirically to be a significant determinant of investment in Q type models". Dealing with Italian data, PAPAGNI (1990) states that "monopolistic and monopsonistic behaviour is usual in many industries ...".

This paper gives an alternative explanation to the above effects, stressing the importance of a demand side determinant: the presence of some monopolistic power. As in Auerbach, we deal with a corporate income tax rate and an investment tax credit change. Contrary to Auerbach, we deal also with long run, since the assumption of imperfect competition allows for the study of interesting effects. In fact, the response of an imperfectly competitive firm to a corporate income tax rate change may depend on its size. If a fiscal reform causes a small firm to reduce investment, for example, the same reform can produce a growth in investment, when the firm under study is greater. The reason of this effect lies on the dependence of the marginal revenue of capital, *i.e.* output price times capital marginal product, on capital stock. This does not occur under perfect competition, when the production function has constant returns to scale, because neither the price of capital nor its marginal product vary.

Furthermore, we consider the short-run effects of both an unexpected and a previously announced reform. As regards the unexpected reform, we show that a change of the investment tax credit influences the shadow price of capital. This does not occur under perfect competition, when the production function shows constant returns to scale. Just as in the long run, the short-run reaction of an unexpected corporate income tax rate change depends on the firm's size.

Fiscal reforms are often announced. This causes firms to modify their choices to anticipate the reform itself, even though they operate for some time under the old regime. Therefore, they compute a linear combination between the time path of the old regime and that of the new regime, thus choosing an intermediate one.

To study the effects of the announcement, we apply the mathematics of Laplace transforms. The papers by Judd (1985, 1987) look very useful to understand this tool. Thanks to this technique we demonstrate that current investment can move in a direction opposite to the long-run one, when the corporate income tax rate reform is expected. This result confirms that obtained by Auerbach, although we use the standard assumption of quadratic installation costs of new capital.

We also prove that the same divergent short-run effect is always shown by an announced investment tax credit reform. This effect occurs since short-run investment depends only on the change in the shadow price of capital and not also on the change of this tax parameter, as the firm operates under the old fiscal regime, for some time. As maintained by Auerbach, the response of the firm depends also on "the relative tax treatment of old and new capital" (1989, p. 939).

Consequently, we find that the market structure is not fiscally neutral neither in the long run nor in the short run.

As regards the responses of labour demand and output price to the fiscal reform, we do not find any new results. However, both of them are linked to capital accumulation [2]. Therefore, a fiscal reform may be inflationary in the short run and deflationary in the long run or *viceversa*, when the reform is expected. Similarly, the same fiscal reform may reduce the short-run demand for labour and raise the long-run one or *viceversa*.

In the first section we develop the model and obtain the first order conditions for the maximization process. Then we introduce an explicit form of the model. We assume that the firm under study faces a constant returns to scale Cobb-Douglas production function and that the capital installation cost function depends only on the flow of new capital (and not also on its stock). The latter function is assumed convex. As already said, we use these simple standard assumptions to isolate the effects of the reform depending on the demand side. Therefore, all the deviations from the traditional results will be due to monopolistic power and not to particular technological assumptions. In the second section we study both the short-run and the long-run effects of an unexpected reform. In the third section we analize the effects of an announced tax reform. Fourth section concludes.

1. *The Model*

The model used is based on the study of a price-making firm, employing two inputs, capital and labour, and producing a single output. Its production function shows constant returns to scale [3]. The firm under study works in a very simple corporate tax system, represented by a corporate income tax rate, an investment tax credit and a schedule of depreciation allowances. By assumption, the fiscal depreciation rate is equal to the effective one. In addition, the firm operates in a perfect foresight context.

The firm chooses its strategy, by maximizing the present value of future net cash flows. Since, by assumption, the firm is debt-financed, the discount rate is equal to the nominal net interest rate, r, which is assumed constant for simplicity. The value function is as follows:

---

[2] As will be seen, labour demand shows a direct relation with capital, while output price shows an inverse one.

[3] As stressed by NICKELL (1978), a price-making firm, facing a constant returns to scale production function behaves just like a price-taking one with a strictly decreasing returns to scale production function. In both cases, the firm shows a bounded production plan. See also HAYASHI (1982).

$$V(0) = \int_0^{+\infty} R(t)\, e^{-rt}\, dt \tag{1}$$

and

$$R(t) = p(t)\, F(K, L; t) - w\, L(t) -$$

$$- p_k (1-\gamma) \left[I(t) + C(I; t) \right] - \gamma(t)\, \Pi_{fisc}(t) \tag{2}$$

which is the net cash flow at time t. Notice that

$$\Pi_{fisc}(t) = p(t)\, F(K, L; t) - w\, L(t) -$$

$$- \int_{-\infty}^{t} p_k \left[I(u) + C(I; u) \right] D(t, t-u)\, du \tag{3}$$

represents fiscal profit, where $p(t)$ is output price, w is nominal wage rate, p_k is the price of capital. Both p_k and w are given by assumption. Technology is given by $F(K, L)$, which is the production function for the good produced by the firm and K and L are respectively capital and labour. By hypothesis, this function is at least twice differentiable. Function $C(I)$ is the cost of installating new capital and is assumed convex in I, *i.e.* in gross investment. Fiscal parameters are represented by τ, the corporate income tax rate, and γ, the investment tax credit.

The above value function is subject to the following differential equation representing net capital accumulation:

$$\frac{dK(t)}{dt} = I(t) - \delta K(t) \tag{4}$$

where δ is the rate of depreciation of capital stock. Notice that depreciation is exponential, by assumption. Moreover, the boundary condition $K(0) = K_0$, defining the starting point of the stock of capital, holds.

The above value function can be written as follows:

$$V(0) = \int_0^{+\infty} (1-\tau(t)) \left(\left[p(t)\, F(K, L; t) - wL(t) - \right. \right.$$

$$\left. \left. - p_k (1-\gamma-\tau(t)\, \Gamma(t)) \left[I(t) + C(I; t) \right] \right) e^{-\rho t}\, dt + A_0 \right. \tag{5}$$

where

$$A_0 \equiv p_k \int_0^{+\infty} \tau(t)\, e^{-\rho t} \int_{-\infty}^{0} D(t, t-u)\,[I(u) + C(I;u)]\, du\, dt$$

is predetermined at time t, thus representing the influence of past variables on the current value function. As in Auerbach (1989), $\Gamma(t)$ is the present value of tax savings per unit of money of investment at time t. Since the depreciation of capital follows an exponential rule, that is

$$D(u) = \delta\, e^{-\delta u}$$

the term $\Gamma(t)$ becomes:

$$\Gamma(t) = \int_t^{+\infty} e^{-\rho(u-t)} D(u)\, du = \int_t^{+\infty} e^{-\rho(u-t)}\, \delta e^{-\delta(u-t)}\, du = \frac{\delta}{\rho + \delta} < 1 \qquad (6)$$

Introduce now the demand function for the good produced by the representative firm:

$$Q(t) = v p(t)^{-\eta} \qquad (7)$$

where $-\eta$ is the constant price elasticity of demand, v is a parameter. By assumption η is greater than one. Next we introduce a market-clearing condition:

$$Q(t) = F[K(t), L(t)] \qquad (8)$$

Substituting condition (8) into function (7) and solving for p yields the market price:

$$p(t) = v^{\frac{1}{\eta}} F[K(t), L(t)]^{\frac{1}{\eta}} \qquad (9)$$

Using the above equations yields the relevant Hamiltonian:

$$\mathcal{H} = e^{-\rho t}\left[(1-\tau)\left(v^{\frac{1}{\eta}} F(K, L)^{1-\frac{1}{\eta}} - w L\right) - \right.$$

$$\left. - (1-\gamma-\tau\Gamma)\, p_k [I + C(I)] + \mu [I - \delta K]\right] \qquad (10)$$

where μ is the costate variable of capital stock. The relevant first order conditions are:

$$\frac{d\mu}{dt} = (\rho + \delta)\,\mu - (1-\tau)\,\frac{\eta-1}{\eta}\,p\,F_k \tag{11}$$

$$\mu = (1-\gamma-\tau\Gamma)\,p_k\big(1 + C_I(I)\big) \tag{12}$$

$$F_L = \frac{\eta}{\eta-1}\,\frac{w}{p} \tag{13}$$

where $F_k \equiv \dfrac{\partial F(K, L)}{\partial K}$, $F_L \equiv \dfrac{\partial F(K, L)}{\partial L}$ and $C_I(I) \equiv \dfrac{\partial C(I)}{\partial I}$

Equation (11) describes the dynamics of the shadow price of capital, which coincides with the costate variable of capital. Integrating it, we obtain:

$$\mu(t) = \int_t^{+\infty} (1-\tau(s))\,\frac{\eta-1}{\eta}\,p(s)\,F_k(s)\,e^{-(\rho+\delta)(s-t)}\,ds \tag{11'}$$

which is the present discounted value of additional profits generated by an additional unit of capital good. Notice that the shadow price of capital may be different from the perfectly competitive one because of the presence of the downward sloping demand curve.

Dividing μ by p_k yields Tobin's marginal q-ratio. Since the price of capital is given, only the numerator of this ratio is relevant. Therefore, dealing with μ is the same as dealing with Tobin's marginal q.

Condition (12) shows the relation between the shadow price of capital and the marginal cost of installation. As can be seen, the structure of the market does not influence this expression. Dividing both the sides of (12) by p_k yields the relationship between the q-ratio and the marginal cost of installation.

Furthermore, equation (13) represents the optimality condition in the labour market. After easy manipulations it becomes

$$p(t) = \frac{\eta}{\eta-1}\,\frac{w}{F_L(t)} > \frac{w}{F_L(t)}$$

The above inequality holds because the competitive output price is always less than the monopolistic one. Therefore, the quantity produced by a price-making firm is less than the competitive one.

Notice that in the monopolistic case, there exists a trade-off between price and quantity. Therefore, revenue, defined as price times quantity, will show a lower tax rate elasticity. This will be useful to understand the effects of the tax reform on μ and K.

Finally, the following transversality condition must hold:

$$\lim_{t \to +\infty} \mu_t K_t e^{-\rho t} = 0 \tag{14}$$

1.1. *An explicit form of the model.* – In the following part of this paper, an explicit functional form of the firm's technology will be employed. We assume that the firm's production function is the constant returns to scale Cobb-Douglas:

$$F(K, L) = K^{\alpha} L^{1-\alpha}, \text{ with } 0 < \alpha < 1 \tag{15}$$

and that the quadratic installation cost function depends only on the flow of new capital,

$$C(I) = \frac{1}{2} I^2 \tag{16}$$

The system of differential equations, corresponding to the above first order conditions, now consists of the following equations:

$$\frac{d\mu}{dt} = (\rho + \delta)\mu - (1-\tau) g K^{l} \tag{17}$$

where $g = \alpha \frac{\eta - 1}{\eta} v^{\frac{1}{\eta\alpha + 1 - \alpha}} \left(\frac{\eta}{\eta - 1} \frac{1}{1-\alpha} w \right)^{-\frac{(1-\alpha)(\eta-1)}{\eta\alpha + 1 - \alpha}} > 0$

and $l \equiv -\frac{1}{1 + \alpha(\eta - 1)} < 0;$

$$\frac{dK}{dt} = \frac{\mu - (1 - \gamma - \tau\Gamma) p_k}{(1 - \gamma - \tau\Gamma) p_k} - \delta K \tag{18}$$

Linearize system (17)-(18):

$$\begin{bmatrix} d\mu/dt \\ dK/dt \end{bmatrix} \approx \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} \mu - \mu^* \\ K - K^* \end{bmatrix} \tag{19}$$

where

$$a_{11} \equiv \rho + \delta > 0;$$
$$a_{12} \equiv -(1-\tau) l g K^{*l-1} > 0$$
$$a_{21} \equiv \frac{1}{(1 - \gamma - \tau\Gamma) p_k} > 0$$
$$a_{22} \equiv -\delta < 0$$

As shown in Appendix 1, the roots of the characteristic equation, r_1 and r_2, show opposite signs, thus yielding a saddle point solution. The resulting saddle path is:

$$\mu(t) = \mu^* + \frac{A}{K_0 - K^*}[K(t) - K^*] \tag{20}$$

where: $\frac{A}{K_0 - K^*} \equiv \frac{a_{12}}{r_1 - a_{11}} < 0$

K_0 is the initial value of capital and the superscript "*" indicates the steady-state values of μ and K.

Consider now the behaviour of labour demand and market price. Substituting (13) into (9) and solving for L gives:

$$L(t) = v^{\frac{1}{\eta\alpha + 1 - \alpha}} \left(\frac{\eta}{\eta - 1} \frac{1}{1-\alpha} w \right)^{\frac{-\eta}{\eta\alpha + 1 - \alpha}} K(t)^{\frac{\alpha(\eta-1)}{\eta\alpha + 1 - \alpha}} \quad \forall\ t \in [0, +\infty). \tag{21}$$

As can be seen, the first derivative of L with respect to K is positive, while the second derivative

$$\frac{\partial^2 L(t)}{\partial K(t)^2} = \frac{-\alpha(\eta-1)}{(\eta\alpha + 1 - \alpha)^2} v^{\frac{1}{\eta\alpha + 1 - \alpha}} \left(\frac{\eta}{\eta - 1} \frac{1}{1-\alpha} w \right)^{\frac{-\eta}{\eta\alpha + 1 - \alpha}} K(t)^{\frac{\alpha(\eta-1)}{\eta\alpha + 1 - \alpha} - 2} < 0$$

is negative, since $\frac{-\alpha(\eta-1)}{(\eta\alpha + 1 - \alpha)^2}$ is always negative. Therefore, when K grows also L grows, but less than proportionally. This can be explained through condition (13): when capital grows, the output price p decreases, thus raising real wage and, consequently, labour marginal productivity. Conversely, capital marginal productivity decreases as K increases [4]. Therefore, the increase in the capital-labour ratio and the decrease in the output price cause the ratio between F_L and F_K to increase.

[4] Capital marginal productivity can be easily obtained through conditions (9) and (13):

$$F_K(t) = \alpha v^{\frac{1-\alpha}{\eta\alpha + 1 - \alpha}} \left(\frac{\eta}{\eta - 1} \frac{1}{1-\alpha} w \right)^{\frac{-\eta(1-\alpha)}{\eta\alpha + 1 - \alpha}} K(t)^{-\frac{1-\alpha}{\eta\alpha + 1 - \alpha}}.$$

Consider now the output price. Substituting (13) into (9) and solving for p yields:

$$p(t) = v^{\frac{\alpha}{\eta\alpha + 1 - \alpha}} \left(\frac{1}{1-\alpha} \frac{\eta}{\eta - 1} w \right)^{\frac{1-\alpha}{\eta\alpha + 1 - \alpha}} K(t)^{-\frac{\alpha}{\eta\alpha + 1 - \alpha}}$$

$$\forall\, t \in [0, +\infty) \qquad (22)$$

Differentiating expression (22) with respect to $K(t)$ gives:

$$\frac{\partial p(t)}{\partial K(t)} = -\frac{\alpha}{\eta\alpha + 1 - \alpha} v^{\frac{\alpha}{\eta\alpha + 1 - \alpha}} \left(\frac{1}{1-\alpha} \frac{\eta}{\eta - 1} w \right)^{\frac{1-\alpha}{\eta\alpha + 1 - \alpha}} K(t)^{-\frac{\alpha}{\eta\alpha + 1 - \alpha} - 1}$$

which is obviously negative. Equations (21) and (22) and their derivatives hold also in steady state.

In conclusion, price p adjusts sluggishly, while instantaneously clearing any excess of demand, because capital K moves slowly. This result is not very surprising. However equation (22) allows to study the direction of the change in price.

2. *An Unexpected Fiscal Reform*

In this section the long-run effects of changes in τ and γ are studied. Moreover, the short-run effects of the same fiscal reforms are considered, under the assumption that these changes are unexpected.

Start with the long-run results. Equating (17) and (18) to zero yields the steady-state levels of μ and K:

$$\mu^* = \frac{1}{\rho + \delta} (1 - \tau)\, g\, K^{*i} \qquad (24)$$

and

$$K^* = \frac{1}{\delta} \frac{\mu^* - (1 - \gamma - \tau\Gamma)\, p_k}{(1 - \gamma - \tau\Gamma)\, p_k} \qquad (25)$$

Using the implicit function theorem and differentiating μ^* and K^* with respect to τ yields:

$$\frac{\partial \mu^*}{\partial \tau} = -\frac{\left(\left(1-\frac{(1-\tau)\,\Gamma}{1-\gamma-\tau\Gamma}\,l\right)K^* + \frac{\Gamma}{1-\gamma-\tau\Gamma}\,\frac{1}{\delta}\right)\frac{g}{\rho+\delta}\,K^{*l-1}}{1-\frac{1/\delta}{\rho+\delta}\,g\,l\,\frac{(1-\tau)}{(1-\gamma-\tau\Gamma)\,p_k}\,K^{*l-1}} < 0 \quad (26)$$

and

$$\frac{\partial K^*}{\partial \tau} = \frac{\frac{1}{1-\gamma-\tau\Gamma}\left(\Gamma\left(K^* + \frac{z}{\delta}\right) - \frac{1/\delta}{p_k\,(\rho+\delta)}\,g\,K^{*l}\right)}{1-\frac{1/\delta}{\rho+\delta}\,g\,l\,\frac{(1-\tau)}{(1-\gamma-\tau\Gamma)\,p_k}\,K^{*l-1}} \quad (27)$$

Notice that derivative $\partial\mu^*/\partial\tau$ is always negative, while the sign of $\partial K^*/\partial\tau$ is ambiguous. As can be seen:

$$\operatorname{sgn}\left(\frac{\partial K^*}{\partial \tau}\right) = \operatorname{sgn}\left(\Gamma\left(K^* + \frac{1}{\delta}\right) - \frac{1/\delta}{p_k\,(\rho+\delta)}\,g\,K^{*l}\right)$$

i.e. $\partial K^*/\partial\tau$ is positive when $K^* > \bar{K}_1$, where

$$\bar{K}_1 = K^* \mid \Gamma\left(K^* + \frac{1}{\delta}\right) - \frac{1/\delta}{p_k\,(\rho+\delta)}\,g\,K^{*l} = 0$$

In Appendix 2 a graphic demonstration of the sign of $\dfrac{\partial K^*}{\partial \tau}$ will be given. This result differs from the perfectly competitive one, as reported in Appendix 3, where the derivatives of μ^* and K^* with respect to τ are always negative, under the assumption that $\gamma + \Gamma$ is less than 1.

The different reaction in the monopolistic case is due to the lower elasticity of revenue with respect to the corporate tax rate. When τ grows, in fact, the fiscal savings due to tax deduction allowances and the amount of the tax paid by the firm change. While the first effect makes cash flow to grow, the second one reduces it. Therefore, when the first positive effect on fiscal savings is great enough, an increase (decrease) in τ may increase (decrease) firm's production.

Notice that the effect on the fiscal savings overcomes that on revenue, only if K is great enough. Conversely, if the firm were price-taking, the second negative effect would be always greater than the positive one, thus

causing the response of a price-taking firm to depend only on the fiscal system (and not also on the firm's size).

Consider now the effect of a change in γ. Differentiating μ^* and K^* with respect to γ, yields:

$$\frac{\partial \mu^*}{\partial \gamma} = \frac{1-\tau}{\rho+\delta} g\, l\, K^{*l-1} \frac{\partial K^*}{\partial \gamma} \tag{28}$$

and

$$\frac{\partial K^*}{\partial \gamma} = \frac{1/\delta}{(\rho+\delta)\, p_k} \left(\frac{(1-\tau)}{(1-\gamma-\tau\Gamma)} \right) g\, l\, K^{*l-1} \frac{\partial K^*}{\partial \gamma} + \frac{1}{1-\gamma-\tau\Gamma} (K^* + 1/\delta) \tag{29}$$

which becomes

$$\frac{\partial K^*}{\partial \gamma} = \frac{\dfrac{1}{1-\gamma-\tau\Gamma} (K^* + 1/\delta)}{1 - \dfrac{1/\delta}{(\rho+\delta)\, p_k} \dfrac{(1-\tau)}{(1-\gamma-\tau\Gamma)} g\, l\, K^{*l-1}} > 0 \tag{29'}$$

Expression (28) captures the effect of monopolistic power, since the marginal revenue of capital, pF_K, falls as γ grows, (contrary to the perfectly competitive case where pF_K remains unchanged).

Consider now derivative (29). The first term on the right-hand side represents the effect of γ on K^* through μ^*. The second term within parentheses represents the effect of γ on fiscal deductions. Obviously, the greater is γ the greater is the fiscal benefit. The net result is shown in (29'), where a growth in the investment tax credit causes capital stock to increase, because the second effect is always greater, thus making the firm to pay a smaller amount of taxes.

The output price is now briefly considered. As can be seen in derivative (27), the effect on capital accumulation of a change in τ varies as the firm's size varies. Consequently, the direction of price adjustment depends on the firm's size. Therefore, when K is greater than K_1, an increase (decrease) in τ reduces (raises) the output price. When K is less than K_1, the contrary is true.

When γ changes, the following relation always holds: sgn $(d\gamma)$ = − − sgn (dp). Therefore, if the output demand curve is not influenced by the reform, an unexpected growth (fall) in γ always produces disinflationary (inflationary) effects.

Finally, equation (21) shows that labour demand depends on capital accumulation, thus increasing as capital increases and *viceversa*.

Consider now the short-run effects of the same fiscal reforms. Begin with the derivative of $\mu(t)$ with respect to τ:

$$\frac{\partial \mu(t)}{\partial \tau} = \frac{\partial \mu^*}{\partial \tau} - \frac{a_{12}}{r_1 - a_{11}} \frac{\partial K^*}{\partial \tau} \tag{30}$$

Since $\frac{-a_{12}}{r_1 - a_{11}}$ is positive, when $\frac{\partial \mu^*}{\partial \tau}$ and $\frac{\partial K^*}{\partial \tau}$ have the same sign – *i.e.* when $K^* < \bar{K}_1$ – we have $\operatorname{sgn}\left[\frac{\partial \mu(t)}{\partial \tau}\right] = \operatorname{sgn}\left[\frac{\partial \mu^*}{\partial \tau}\right] = \operatorname{sgn}\left[\frac{\partial K^*}{\partial \tau}\right]$. In other words $\frac{\partial \mu(t)}{\partial \tau}$ is negative and overshooting occurs because: $\left|\frac{\partial \mu(t)}{\partial \tau}\right| < \left|\frac{\partial \mu^*}{\partial \tau}\right|$.

Conversely, the short-run jump of μ can show a different direction than that suggested by the long-run shift, when $K^* > \bar{K}_1$. This ambiguity is caused by the opposite signs of $\frac{\partial \mu^*}{\partial \tau}$ and $\frac{\partial K^*}{\partial \tau}$. When the first effect $\left(\frac{\partial \mu^*}{\partial \tau}\right)$ is small and the second $\left(\frac{\partial K^*}{\partial \tau}\right)$ is great, the jump of μ must discount this counter-effect on capital and is initially shifted in the opposite direction.

Consider the effects of a change in γ. Using equation (20) yields:

$$\frac{\partial \mu(t)}{\partial \gamma} = \frac{\partial \mu^*}{\partial \gamma} - \frac{a_{12}}{r_1 - a_{11}} \frac{\partial K^*}{\partial \gamma} = -\frac{a_{12}}{a_{11}(r_1 - a_{11})} \frac{r_1}{\partial K^*} \frac{\partial K^*}{\partial \gamma}$$

with $-\frac{a_{12}}{a_{11}(r_1 - a_{11})} < 0$

thus obtaining $\operatorname{sgn}\left[\frac{\partial \mu(t)}{\partial \gamma}\right] = -\operatorname{sgn}\left[\frac{\partial K^*}{\partial \gamma}\right] = \operatorname{sgn}\left[\frac{\partial \mu^*}{\partial \gamma}\right]$. Since the effects of a change in γ on the steady state variables show opposite signs, $\left|\frac{\partial \mu(t)}{\partial \gamma}\right|$ is lower than $\left|\frac{\partial \mu^*}{\partial \gamma}\right|$ and overshooting never occurs.

3. *An Announced Fiscal Reform*

This section studies the effects of an announced fiscal reform. As Sandmo writes: "If the tax rate is expected to increase, there are tax savings from buying capital equipment now rather than later" (Sandmo, 1979, p. 150). In other words, the announcement may induce the firm to anticipate its choices and to pay *now* a fiscal penalty for buying capital before the reform, in order to save *after* the reform.

Notice that the effects of an announced reform may be strongly different than those due to an unexpected one, since short-run investment may follow a different direction from that suggested by long-run investment. As Auerbach (1989) states, an announced fiscal reform may influence current investment in two ways. First, the firm tends to "smooth the investment path to a new desired long-run capital stock" (p. 950). Second, it continues to operate for some time in the old regime, while immediately anticipating the effects of the new regime. Therefore, the firm must compare the tax treatment of old and new inputs and, consequently, current investment may be pushed "in a different direction than would be suggested by the long-run effects" (p. 939).

As already said, Auerbach explains this anomalous change by assuming that the installation costs depend (inversely) on the stock of capital too. However, in a monopolistic context, the same results can be obtained, under the assumption that the installation costs depend only on current investment.

In this section a graphic approach is employed, while its analytical explanation is developed in Appendix 4.

Consider now a change in the corporate income tax rate. As already seen, the signs of both the derivatives of μ and K with respect to τ depend on the firm's size. When K^* is less than K_1, the above derivatives have the same sign and the result is not surprising, since current investment changes according to the long-run direction.

As can be seen in Figures 1.a and 1.b, the short-run jump of vector (μ, K) (from point E to point A) is less than that occurring in the unexpected case, where (μ, K) would be directly driven upon the new saddle path (where point E' is placed). Therefore, overshooting occurs in Figures 1.a and 1.b only when point A is respectively above and below the new saddle point E'. As we know, the existence of overshooting depends on the length of the time interval between the announcement and the reform, *i.e.* the longer is the interval the shorter is the short-run jump of μ.

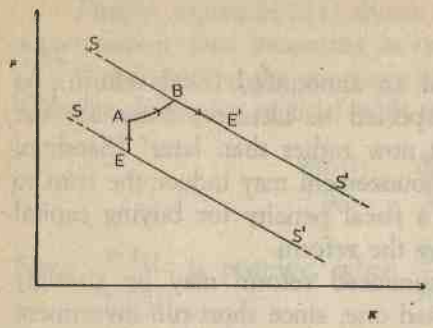

FIG. 1.a

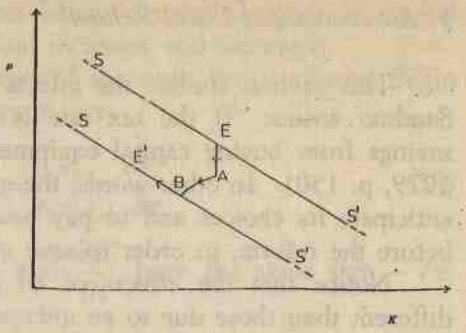

FIG. 1.b

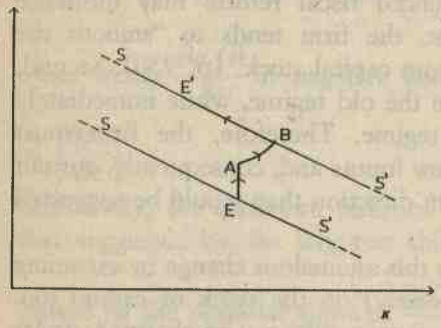

FIG. 2.a

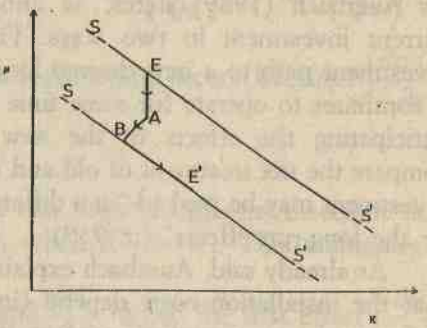

FIG. 2.b

When $K^* > \bar{K}_1$ $\left(\text{where } \frac{\partial K^*}{\partial \tau} > 0 \text{ and } \frac{\partial \mu^*}{\partial \tau} < 0\right)$ the effect is more interesting. Consider Figures 2.a and 2.b, representing respectively a decrease and an increase in τ.

As can be noted, after the jump (from E to A), current investment follows a different direction (from A to B) than would be suggested by the long-run effects. Conversely, when vector (μ, K) arrives at point B, the convergence towards the new steady state begins and investment follows the long-run direction.

To understand the short-run impact of the announced reform on investment, differentiate condition (12), thus obtaining

$$dI(t) = \frac{1}{(1-\gamma-\tau)\,p_k}\,d\mu(t) + \frac{\mu(t)}{(1-\gamma-\tau)^2}\,d\tau$$

Notice that τ does not immediately vary. Consequently, the second term in the right-hand side is null until the reform, thus yielding

$$\text{sgn}\,(dI) = \text{sgn}\,(d\mu).$$

Therefore, current investment moves according to the change in μ (or Tobin's q-ratio). The analytical demonstration of this effect is reported in proposition 1 of Appendix 4.

The same result is obtained when the change in γ is announced, although the direction of the investment change does not depend on the size of the firm. After the announcement, current investment *always* moves in the direction opposite to the long-run one [5]. Figures 2.a and 2.b can still be used to represent respectively a decrease and an increase in γ. Notice that these results can be obtained under constant returns to scale, only when the firm has some monopolistic power. If the firm were perfectly competitive – given the assumption of a constant returns to scale Cobb-Douglas production function – μ would not be influenced by the change in γ and current investment would not initially move in the opposite direction. In this case there would be only a gradual shift towards the new steady-state value of capital stock.

Consider now the dynamics of output price. Since output price moves in the same direction as capital accumulation, the expected reform may determine two different cases. Consider Figure 2.a, where capital initially increases and later decreases: in this case, output price initially increases and later decreases, while the net result is always inflationary.

When capital initially decreases and then increases (Figure 2.b), the contrary is true and the fiscal reform produces a short-run inflationary and a long-run deflationary effect.

Finally, expression (21) allows to understand the dynamics of labour demand. Consider again Figures 2.a and 2.b. Since labour demand follows the same direction as capital accumulation, these figures can be used to show respectively a decrease and an increase of the demand for this factor. Also in this case, the short-run effects are of opposite sign: Figure 2.a implies a short-run increase of labour demand, followed by a long-run decrease, while Figure 2.b implies a short-run decrease followed by a long-run increase.

4. *Conclusions*

In this work the reaction of a monopolist to a fiscal reform has been

[5] The reason is always the same. The firm does not immediately consider the effect of the cost deductions. For an analytical explanation see Proposition 2 in Appendix 3.

studied. As we have shown, the effect of a corporate income tax rate change depends on the firm's size, since marginal revenue of capital is inversely proportional to capital. Moreover, since labour and capital are the determinants of the firm's dimension, we can conclude that when the firm's size grows (falls) then pF_K falls (grows), thus making the sign of the fiscal effect to change.

When the reform is expected, the reaction of the firm depends again on its dimension. There is an interval in which the firm's size causes current investment to move in a different direction than that suggested by the long-run effects, when the reform is announced. The reason is that the change in short-run investment only depends on the change in the capital shadow price, since the firm operates under the old fiscal regime, until the reform.

When the firm is small enough, a raise of τ is followed by a decrease in vector (μ, K). In this case, when the change is unexpected, overshooting always occurs. Conversely, overshooting may or may not occur, when the reform is announced. This depends on the time interval between the announcement and the reform itself. When the interval is long enough, overshooting does not occur.

As regards the effect of a change in γ, it differs from that occurring under perfect competition. Under imperfect competition, in fact, an increase in γ raises capital accumulation, thus causing a decrease in capital marginal revenues and, consequently, in μ.

When the change in γ is announced, current investment is pushed in a different direction than that suggested by the long-run effects. The reason is the same as that explained for the change in τ: the firm continues to operate under the old regime until the reform.

In conclusion, the same interesting effects found by Auerbach, under a special technological assumption, can be obtained by simply assuming that the firm, with a constant returns to scale production function and quadratic costs of installation, has got monopolistic power.

APPENDIX 1

The roots of system (19) are the solutions of its characteristic equation:

$$(r-a_{11})(r-a_{22})-a_{12}a_{21}=0$$

These roots are:

$$r_{1,2} = \frac{\rho \pm \left(\rho^2 - 4\,(a_{11}a_{22} - a_{12}a_{21})\right)^{1/2}}{2}$$

where $r_1 < 0$ and $r_2 > 0$.

APPENDIX 2

In this appendix we use a graphic approach, to show that the following inequality:

$$\Gamma\left(K^* + \frac{1}{\delta}\right) > \frac{1/\delta}{p_k(\rho + \delta)}\, g\, K^{*l}$$

holds for $K^* > \bar{K}_1$. As shown by Figure 3, the stright line, reported on the left-hand side is greater than the hyperbole of the right-hand side for all the values of K, which are greater than the intersection value, $\bar{K}_1$.

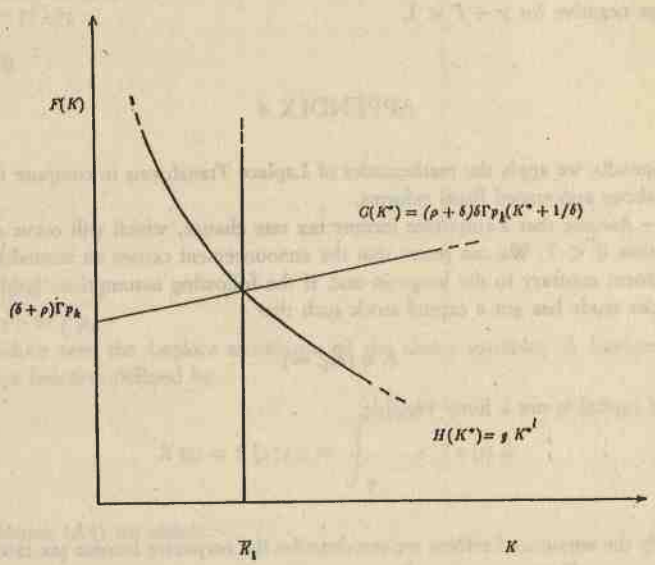

FIG. 3.

APPENDIX 3

Assume that the representative firm is perfectly competitive and that the output price is equal to 1. The system under study becomes:

$$\frac{d\mu}{dt} = (\rho + \delta)\,\mu - (1-\tau)\,\alpha \left(\frac{1}{1-\alpha}\, w\right)^{-\frac{1-\alpha}{\alpha}} \tag{17'}$$

$$\frac{dK}{dt} = \frac{\mu - (1-\gamma-\tau\Gamma)\,p_k}{(1-\gamma-\tau\Gamma)\,p_k} - \delta K \tag{18'}$$

From equation (17') and (18), we obtain the steady-state values of μ and K and compute their derivatives with respect to τ:

$$\frac{\partial \mu^*}{\partial \tau} = -\frac{\mu^*}{1-\tau} < 0$$

and

$$\frac{\partial K^*}{\partial \tau} = -\frac{1-\gamma-\Gamma}{1-\gamma-\tau\Gamma}\left(K^* + \frac{1}{\delta}\right)$$

which is always negative for $\gamma + \Gamma < 1$.

APPENDIX 4

In this appendix we apply the mathematics of Laplace Transforms to compute the short-run effects of the above announced fiscal reforms.

Proposition 1. – Assume that a corporate income tax rate change, which will occur at time T, is announced at time $0 < T$. We can prove that the announcement causes an immediate short-run jump in investment contrary to the long-run one, if the following assumptions hold:

i) the firm under study has got a capital stock such that

$$K \in (\bar{K}_1, \infty)$$

ii) the stock of capital is not a jump variable.

Proof

1) To study the announced reform we can describe the corporate income tax rate as follows:

$$\tau(t) = \tau + \varepsilon h(t)$$

where $h(t)$ is the following Heaviside function:

$$h(t) = \begin{cases} 0 \text{ for } t < T \\ 1 \text{ for } t \geq T \end{cases}$$

This means that $\tau(t)$ remains equal to τ for $0 < t < T$ and that it becomes equal to $\tau + \varepsilon$ at time T.

2) The system under study becomes:

$$\frac{d\mu}{dt} = (\rho + \delta)\,\mu - (1-\tau-\varepsilon b)\,g\,K^{l}$$

$$\frac{dK}{dt} = \left(\frac{\mu}{\left(1-\gamma-(\tau+\varepsilon b)\,\Gamma\right)p_k} - 1\right) - \delta K \tag{A3}$$

3) Linearizing system (A3) around $\varepsilon = 0$ yields

$$\begin{bmatrix} d\mu_\varepsilon/dt \\ dK_\varepsilon/dt \end{bmatrix} \approx \begin{bmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{bmatrix} \begin{bmatrix} \mu_\varepsilon \\ K_\varepsilon \end{bmatrix} + \begin{bmatrix} n_1 \\ n_2 \end{bmatrix} \tag{A4}$$

where

$J_{11} \equiv \rho + \delta > 0$

$J_{12} \equiv -(1-\tau)\,l\,g\,K^{*^{l-1}} > 0$

$J_{21} \equiv \dfrac{1}{(1-\gamma-\tau\Gamma)\,p_k} > 0$

$J_{22} \equiv -\delta < 0$

and

$\mu_\varepsilon \equiv d\mu/d\varepsilon$

$K_\varepsilon \equiv dK/d\varepsilon$

$n_1 \equiv b\,g\,K^{*l}$

$n_2 \equiv b\,\dfrac{\Gamma\,\mu^{*}}{(1-\gamma-\tau\Gamma)^2\,p_k}$

4) Introduce now the Laplace transforms of the above variables. A Laplace transform of variable x is a function defined by

$$X(s) = \mathfrak{L}[x](s) \equiv \int_0^{+\infty} e^{-st}\,x(t)\,dt$$

Using system (A4) we obtain:

$$\begin{bmatrix} \Gamma(s) \\ \Lambda(s) \end{bmatrix} = (sI-J)^{-1} \begin{bmatrix} H(s)\,(g\,K^{*l} - \delta\,p_k) + \mu_\varepsilon(0) \\ H(s)\,\dfrac{\delta K^{*}}{1-\gamma-\tau} \end{bmatrix} \tag{A5}$$

where

$$\Gamma(s) \equiv \int_0^{+\infty} e^{-st}\,\mu_\varepsilon(t)\,dt$$

$$\Lambda(s) \equiv \int_0^{+\infty} e^{-st} K_\varepsilon(t)\, dt$$

and

$$H(s) \equiv \int_0^{+\infty} e^{-st} h(t)\, dt$$

The roots of system (A5) are the same as those obtained in Appendix 1: $r_1 < 0$ and $r_2 > 0$. Notice that $K_\varepsilon(0)$, the short-run jump of capital, is equal to zero because capital is a state variable. Using system (A5) yields the short-run jump of the shadow price of capital [6]

$$\mu_\varepsilon(0) = -J_{12}\left(H(r_2)\frac{\Gamma\mu^*}{(1-\gamma-\tau\Gamma)^2 p_k}\right) / (r_2 - J_{22}) - H(r_2)\, g K^{*l} =$$

$$= H(r_2)\left(\frac{(1-\tau)\Gamma}{(1-\gamma-\tau\Gamma)^2 p_k}\frac{lg}{r_2+\delta}K^{*l-1}\mu^* - gK^{*l}\right) < 0 \qquad \text{(A6)}$$

where

$$H(r_2) \equiv \int_0^{+\infty} e^{-r_2 t} h(t)\, dt = \frac{1}{r_2} e^{-r_2 T} > 0$$

As can be seen, the higher is T the lower is $|\mu_t(0)|$. This means that the short-run jump of μ becomes greater when the date of the reform (T) is set nearer to the date of the announcement (0).

Substituting expression (A6) into system (A5), at time $t = 0$, yields:

$$dK_\varepsilon(0)/dt = \frac{1}{(1-\gamma-\tau)p_k}\mu_\varepsilon(0) \qquad \text{(A7)}$$

because both $K_\varepsilon(0)$ and $h(0)$ are null.

5) In the interval $(\bar{K}_1, \infty)$, we obtain that $\operatorname{sgn}\left(\mu_\varepsilon(0)\right) = \operatorname{sgn}\left(\frac{\partial\mu^*}{\partial\tau}\right)$. Since the short-run jump in investment has the same sign as $\mu_\varepsilon(0)$, we obtain

$$\operatorname{sgn}\left(dK_\varepsilon(0)/dt\right) = \operatorname{sgn}\left(\mu_\varepsilon(0)\right) = \operatorname{sgn}\left(\frac{\partial\mu^*}{\partial\tau}\right)$$

But, in interval $(\bar{K}_1, \infty)$ the long-run derivatives of K^* and μ^* are of opposite sign:

$$\frac{\partial K^*}{\partial\tau} > 0 \text{ and } \frac{\partial\mu^*}{\partial\tau} < 0$$

so that:

[6] Expression (A6) coincides with expression (A.6) in the Appendix of JUDD (1987).

$$\operatorname{sgn}\left(\frac{\partial K^*}{\partial \tau}\right) = -\operatorname{sgn}\left(\frac{dK_\varepsilon(0)}{dt}\right)$$

Therefore, if $K \in (\bar{K}_1, \infty)$, current investment is pushed in a different direction than that suggested by the long-run effects.

Q.E.D.

Proposition 2. – Assume that an investment tax credit change, which will occur at time T, is announced at time $0 < T$. Then the announcement always causes current investment to move in the direction opposite to the long-run one.

Proof

1) Just as in the previous proposition, the investment tax credit can be defined as follows

$$\gamma(t) = \gamma + \varepsilon h(t)$$

where $h(t)$ is the Heaviside function used in proposition 1.

2) The system under study now becomes

$$\begin{aligned} \frac{d\mu}{dt} &= (\rho + \delta)\mu - (1-\tau) g K^l \\ \frac{dK}{dt} &= \left(\frac{\mu}{(1-(\gamma+\varepsilon h)-\tau\Gamma) p_k} - 1\right) - \delta K \end{aligned} \tag{A8}$$

Linearizing the above system around $\varepsilon = 0$ yields

$$\begin{bmatrix} d\mu_\varepsilon/dt \\ dK_\varepsilon/dt \end{bmatrix} \approx \begin{bmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{bmatrix} \begin{bmatrix} \mu_\varepsilon \\ K_\varepsilon \end{bmatrix} + \begin{bmatrix} 0 \\ n_3 \end{bmatrix} \tag{A9}$$

where

$$n_3 = h \frac{\mu^*}{(1-\gamma-\tau\Gamma)^2 p_k}$$

3) Introducing Laplace transforms yields the following system:

$$\begin{bmatrix} \Gamma(s) \\ \Lambda(s) \end{bmatrix} = (sI - J)^{-1} \begin{bmatrix} \mu_\varepsilon(0) \\ H(s) \dfrac{\mu^*}{(1-\gamma-\tau\Gamma)^2 p_k} \end{bmatrix} \tag{A10}$$

4) The roots of the above system are always the same ($r_1 < 0$ and $r_2 > 0$). We consider again the Laplace transform of $h(t)$, by assuming that $s = r_2$. Under this assumption, the short-run jump of the shadow price of capital computed using system (A10) becomes

$$\mu_\varepsilon(0) = -J_{12} H(r_2) \frac{\mu^*}{(1-\gamma-\tau\Gamma)^2 p_k} / (r_2 - J_{22}) =$$

$$= H(r_2) \left(\frac{(1-\tau)}{(1-\gamma-\tau\Gamma)^2 p_k} l g K^{*l-1} \mu^* \right) / (r_2 - J_{22}) \qquad \text{(A6)}$$

Note that $\mu_\varepsilon(0)$ is always negative because l is negative. Therefore, using equations (28) and (30) yields:

$$\operatorname{sgn}\left(\mu_\varepsilon(0)\right) = \operatorname{sgn}\left(\frac{\partial \mu(t)}{\partial \gamma}\right) = \operatorname{sgn}\left(\frac{\partial \mu^*}{\partial \gamma}\right) = -\operatorname{sgn}\left(\frac{\partial K^*}{\partial \gamma}\right)$$

Hence, the announced and the unexpected change in γ drive μ in the long-run direction, while the long-run change in capital stock shows the opposite sign.

As in the previous proposition, substituting expression (A10) in system (A11) yields the short-run jump in investment

$$dK_\varepsilon(0)/dt = \frac{1}{(1-\gamma-\tau) p_k} \mu_\varepsilon(0)$$

Hence, this short-run jump shows the opposite sign than the long-run one:

$$\operatorname{sgn}\left(dK_\varepsilon(0)/dt\right) = \operatorname{sgn}\left(\mu_\varepsilon(0)\right) = -\operatorname{sgn}\left(\frac{\partial K^*}{\partial \gamma}\right)$$

Q.E.D.

REFERENCES

AUERBACH A.J., "Tax Reform and Adjustment Costs: The Impact on Investment and Market Value", *International Economic Review*, No. 4, Nov. 1989, *30*, 939-62.

BLANCHARD O. and KIYOTAKI N., "Monopolistic Competition and the Effects of Aggregate Demand", *American Economic Review*, No. 4, Sept. 1987, *77*, 647-66.

CHIRINKO R.S. and FAZZARI S.M., "Tobin's Q, Non-constant Returns to Scale and Imperfectly Competitive Product Markets", *Recherches Economiques de Louvain*, No. 3, 1988, *54*, 259-75.

GALEOTTI M. and SCHIANTARELLI F., "Monopolistic Competition in U.S. Industry in a Model with Adjustment Costs: A Euler Equation Approach", Discussion Paper n. 9011, University of Brescia, 1990.

HALL R.E., "The Relation between Price and Marginal Cost in U.S. Industry", *Journal of Political Economy*, 1988, *96*, 921-47.

——, "Invariance Properties of Solow's Productivity Residual", NBER Working Papers, 3034, July 1989.

HAYASHI F., "Tobin's Marginal q and Average q: A Neoclassical Interpretation", *Econometrica*, 1982, *50*, 213-24.

JUDD K.L., "Short-run Analysis of Fiscal Policy in a Simple Foresight Model", *Journal of Political Economy*, 1985, 298-319.

——, "Debt and Distortionary Taxation in a Simple Perfect Foresight Model", *Journal of Monetary Economics*, 1987, *20*, 51-72.

MANKIW G., "Small Menu Costs and Large Business Cycles: A Macroeconomic Model of Monopoly", *Quarterly Journal of Economics*, No. 2, May 1985, *100*, 529-39.

NICKELL S., *The Investment Decision of Firms*, Cambridge: James Nisbet & Co. and Cambridge University Press, 1978.

PAPAGNI E., "Dynamic Factor Demands and Market Structure under Rational Expectations: Some Estimates for the Italian Manufacturing System", *Applied Economics*, 1990, *22*, 1377-87.

SANDMO A., "A Note on the Neutrality of the Cash Flow Corporation Tax", *Economics Letters*, 1979, *4*, 173-76.

SCHIANTARELLI F. and GEORGOUTSOS D., "Monopolistic Competition and the Q Theory of Investment", *European Economic Review*, 1990, *34*, 1061-78.

IMPOSTE SUI REDDITI SOCIETARI E MONOPOLIO: GLI EFFETTI DI UNA RIFORMA ANNUNCIATA

Questo articolo studia gli effetti delle variazioni delle aliquote d'imposta dei redditi societari sul comportamento di un monopolista. Come si dimostra, l'effetto di lungo periodo della variazione delle aliquote dipende dalla dimensione dell'impresa. Inoltre, quando viene annunciata la riforma e la dimensione dell'impresa è sufficientemente grande, l'investimento corrente può seguire una direzione diversa da quella suggerita dagli effetti di lungo periodo. Quest'ultimo effetto si verifica anche quando varia il credito d'imposta sugli investimenti.

Rivista Internazionale di Scienze Economiche e Commerciali
Volume 42 (1995), N. 1-2, 93-124

CONTROLLO OTTIMO STOCASTICO, PROCESSI BROWNIANI REGOLATI E OPTIMAL STOPPING

di
MICHELE MORETTO *

1. *Introduzione*

Il controllo ottimo stocastico viene di recente applicato diffusamente nella soluzione di molti problemi economici in presenza di irreversibilità e incertezza, dove la dinamica della variabile di stato è opportunamente descritta da un processo Browniano regolato. Con questo termine si intende la possibilità di sovrapporre al processo che descrive la variabile di stato alcuni comportamenti-limite in corrispondenza di certi valori prefissati, definiti come barriere. Quando il livello della variabile tocca una delle barriere, il processo viene « respinto », dopodiché continua a svilupparsi liberamente fino a quando uno dei limiti è nuovamente raggiunto, ovvero viene « assorbito », cioè il processo possiede un *trap state* che, una volta raggiunto, non viene più abbandonato.

Nel primo caso, l'intervento di regolazione può essere di tipo discreto, e ricondurre con un « salto » la variabile all'interno di un dominio di ammissibilità, oppure di tipo infinitesimale, cioè quanto basta per impedire la fuoriuscita del processo dalla zona di accettazione. In letteratura il primo di questi due tipi di regolazione si definisce come *controllo ad impulso*. Se ne trovano esempi nel modello di gestione aziendale delle scorte di magazzino (S, s) di Arrow, Karlin e Scarf (1958) e di Scarf (1960) e nei più recenti lavori di Flood e Garber (1991) e di Bertola e Caballero (1992) sulla regolazione del tasso di cambio, di Grossman e Laroque (1990) sulla decisio-

* Università di Padova, Dipartimento di Scienze Economiche, Padova.

L'autore desidera ringraziare Nunzio Cappuccio, Diego Lubian e un anonimo referee per le osservazioni e i suggerimenti ricevuti, rimanendo peraltro unico responsabile di errori e omissioni. Il lavoro ha beneficiato dei fondi MURST 60%-1993.

ne di portafoglio in presenza di beni durevoli, e di Caplin e Leahy (1991) sull'aggiustamento dei prezzi con *menu costs*.

Il secondo tipo di regolazione va invece sotto il nome di *controllo istantaneo*. Fra i molti lavori che ne trattano in letteratura, basti citare quelli di Krugman (1991) e Svensson (1991) sul tasso di cambio, di Pindyck (1988), Bertola (1988), Dixit (1991a) e Kobila (1991) sulla determinazione della capacità produttiva ottimale di un'impresa in presenza di irreversibilità, di Bentolila e Bertola (1990), che affrontano la politica occupazionale di un'impresa con costi di assunzione e di licenziamento, di Dosi e Moretto (1994a) che trattano l'allocazione ottimale di una risorsa naturale fra domanda interna e livello minimo garantito per l'esportazione, e di Moretto (1994a) e Dosi e Moretto (1994b), che prendono in esame lo standard minimo ottimale di qualità del capitale ai fini del controllo dell'inquinamento.

In entrambi i casi, per i valori-limite si parla di *barriere riflettenti* (in realtà per il controllo ad impulso sarebbe più opportuno il termine di *jump boundaries*). La scelta del regolatore tra le due diverse modalità d'intervento è valutata in base alla struttura dei costi di controllo.

Se il processo viene assorbito e il momento di assorbimento è inteso come variabile di scelta, ci troviamo allora di fronte a un problema di *optimal stopping time*: in questo caso non è corretto parlare di controllo, poiché l'imposizione di una *barriera assorbente* non è dettata dalla necessità di mantenere la variabile di stato all'interno di un dato spazio di valori. Essa invece svolge il ruolo di « campanello d'allarme », che avverte che è giunto il momento di realizzare una certa operazione (ad esempio: se il prezzo di un bene raggiunge un dato limite superiore, è ottimale per un'impresa produttrice entrare nel mercato; di converso, se il prezzo di una risorsa raggiunge un limite inferiore, è opportuno abbandonare la produzione).

Anche per l'*optimal stopping time* la letteratura presenta un'elevata casistica esemplificativa: l'ottimo momento di esercizio di una *American-type option* è l'oggetto di indagine dei pionieristici lavori di Samuelson (1965), McKean (1965) e Merton (1973). Più recentemente, il momento ottimale per dare avvio a un progetto d'investimento in presenza di costi irrecuperabili è stato trattato da Bonini (1977), McDonald e Siegel (1985, 1986) e da Majd e Pindyck (1987). L'ottimo momento in cui sospendere l'estrazione di una risorsa naturale è stato considerato invece da Brennan e Schwartz (1985), Olsen e Stensland (1988), Gibson e Schwartz (1991) e Bjerksund (1991). Il valore della flessibilità legato alla possibilità di passare da un processo produttivo a un altro è stato sviluppato da Olsen e Stensland (1989) e He e Pindyck (1992).

I costi della realizzazione di un progetto o dell'estrazione di una risorsa naturale non sono gli unici considerati irrecuperabili, molti altri presentano la medesima caratteristica: le spese di entrata e uscita da un mercato – si veda, al riguardo, il lavoro di Dixit (1989a, b) e le estensioni di Bartolini (1992), Leahy (1993) e Moretto (1994b) –, nel caso di una singola impresa come in quello di un'industria competitiva, ne sono un esempio; un altro è costituito dai costi di transazione nella gestione di un portafoglio, studiati da Constantinides (1986) e Dixit (1989c). Le spese di transazione (adattamento e trasporto) impliciti nella scelta di mobilità verso un settore alternativo a quello di appartenenza da parte di un lavoratore è invece il tema trattato da Dixit e Rob (1991, 1994) [1].

Una crescita così significativa della letteratura che riguarda le applicazioni dei processi Browniani a problematiche economiche è potuta avvenire grazie a un continuo sviluppo, negli ultimi anni, della teoria matematica del controllo stocastico, e in particolare di alcuni risultati di ricerca operativa sulla regolazione ottimale dei processi Browniani.

Per il controllo ad impulso, si possono citare i lavori di Bensoussan e Lions (1975), Constantinides e Richard (1978), Harrison, Sellke e Taylor (1983). Per il controllo istantaneo, invece, una certa importanza rivestono i lavori di Harrison e Taylor (1978), Karatzas (1980), Harrison e Taksar (1983) e Harrison (1985). Una trattazione meno rigorosa di questi due tipi di controllo, che tuttavia beneficia di una più diretta interpretazione economica, è stata proposta da Dixit (1991b, 1993) e Dumas (1991).

Infine, per l'*optimal stopping time*, i riferimenti d'obbligo sono Shiryayev (1978), Krylov (1980) e Brekke e Øksendal (1991). Inoltre, particolarmente importanti per le diverse applicazioni economiche che possono avere, sono i contributi di Karatzas e Shreve (1984, 1985) e El Karoui e Karatzas (1988), in cui viene messa in evidenza la generale corrispondenza esistente tra un problema di *optimal stopping* e uno di controllo ottimo [2].

Gli strumenti metodologici per risolvere un problema di regolazione, cioè per individuare in modo ottimale le barriere, non sono gli stessi nel caso di controllo a impulso e di controllo istantaneo. Per il primo si fa riferimento alla *value matching condition* e alla *smooth pasting condition*. La

[1] Molti dei modelli di investimento con irreversibilità hanno beneficiato dell'equivalenza fra programmazione dinamica e i metodi di *option pricing*. Per un approfondimento di questa equivalenza si veda Pindyck (1991), Lund (1991), Dixit (1992) e Dixit e Pindyck (1994).

[2] Facendo leva sulla relazione fra controllo ottimo stocastico e *optimal stopping* Leahy (1993) mostra come un problema di aggiustamento dello stock di capitale di un'industria da parte di un *policy-maker* possa essere visto come una serie di *optimal stopping problems*, in relazione con il contributo di ogni singola unità di capitale al surplus del consumatore.

prima individua una condizione di indifferenza, in base alla quale il guadagno ottenuto dall'esercizio della regolazione eguaglia il costo dello stesso e, pur essendo necessaria, è valida in generale e non soltanto per i valori ottimali dei parametri di controllo. Una scelta rispondente a quest'ultimo requisito è invece garantita dal soddisfacimento della seconda condizione la quale stabilisce che, nella barriera ottima, il guadagno marginale deve assumere lo stesso valore prima e dopo il controllo. Ciò implica la continuità della derivata prima della funzione obiettivo nei punti che separano la zona di inazione da quella di intervento.

Diversamente, per un problema di controllo istantaneo la *smooth pasting condition* vale per ogni valore delle barriere, mentre, per una determinazione ottimale, deve essere soddisfatta una condizione indicata con il nome di *super contact condition*, la quale richiede la continuità della derivata seconda nei punti di regolazione. Quest'ultima, già presente nei lavori di Benes, Shepp e Witsenhausen (1980), Harrison e Taksar (1983) e Karatzas e Shreve (1984, 1985), trova il suo ruolo formale in Dumas (1991).

Infine, in virtù della corrispondenza tra *optimal stopping* e controllo ottimo stocastico, la *smooth pasting condition* diventa pure una condizione di primo ordine nella scelta del momento ottimale di assorbimento (Merton 1973, Karatzas e Shreve 1984, 1985 e Brekke e Øksendal 1991) [3].

Lo scopo del presente lavoro è duplice: in primo luogo si vuole presentare una rassegna di alcuni recenti risultati di ricerca operativa sui processi Browniani regolati che hanno consentito di risolvere buona parte dei problemi di investimento sopra ricordati. Faremo questo utilizzando un semplice esempio di controllo della produttività di un impianto industriale (o macchina).

In secondo luogo si vuole mostrare come sia possibile trattare in modo congiunto sia un problema di regolazione istantaneo, per mantenere l'efficienza dell'impianto attraverso iniezioni di spese di manutenzione (investimento), sia il problema dell'individuazione del momento ottimale in cui estromettere l'impianto dal processo produttivo. Si vedrà come, sotto alcune condizioni di regolarità, quest'ultimo sia sempre riconducibile a un caso di controllo a impulso in cui l'intervento avvenga in un unico istante temporale [4].

---

[3] Brekke e Øksendal (1991) mostrano che la soluzione per un problema di *optimal stopping* che soddisfa la *smooth pasting condition* (che loro chiamano *high contact*) è in generale oltre che necessaria anche sufficiente.

[4] Dumas (1991) presenta un caso congiunto di controllo istantaneo e a impulso, mentre Shreve, Lehoczky e Gaver (1984) studiano un modello di controllo delle scorte di magazzino con una barriera riflettente superiore e una barriera assorbente nel caso il livello delle scorte diventi zero.

La soluzione si trova quindi con l'individuazione di due barriere per il processo stocastico che descrive la produttività: una barriera riflettente inferiore e una barriera assorbente. La prima individua una regione di intervento da parte del regolatore: ogniqualvolta venga raggiunta, sono operate delle spese tali da ripristinare la produttività dell'impianto. La seconda individua invece una regione di inattività: quando questa è raggiunta, l'impianto viene dismesso e venduto al suo prezzo di realizzo. Questa barriera viene raggiunta dopo un prolungato accumulo di spese di manutenzione per cui non risulta più conveniente la tenuta in attività dell'impianto.

Le caratteristiche dell'esempio utilizzato sono introdotte nel secondo paragrafo. Il terzo è dedicato alla presentazione di alcuni risultati preliminari, mentre il quarto individua una soluzione in forma chiusa per il valore dell'impianto. Nel quinto sono ottenute le condizioni di ottimalità per le due barriere, mostrando la simmetria fra il caso riflettente e quello assorbente. Seguono, infine, le conclusioni.

2. *Un semplice esempio*

Si consideri un processo Browniano (o di Wiener) standard $\{w_t,\ t \geq 0\}$ con media nulla e varianza unitaria, e una famiglia di σ-algebre $\{\mathfrak{F}_t,\ t \geq 0\}$ crescenti tali per cui w_t è $\mathfrak{F}_t$-misurabile e $\{w_s,\ s \geq t)$ è indipendente da $\mathfrak{F}_t$. Si consideri, inoltre, il processo $\{x_t,\ t \geq 0\}$, $\mathfrak{F}_t$-misurabile, definito su $(0,\ \infty)$ il quale per ogni dato valore iniziale x_0 e in assenza di ogni controllo soddisfa la seguente equazione differenziale stocastica o equazione di stato:

$$(1) \qquad dx_t = mx_t dt + \sigma x_t dw_t \qquad x_0 = x \in (0,\ \infty)$$

In relazione al processo $\{w_t\}$, il processo $\{x_t\}$ definito dalla (1) esiste ed è unico (Gikhman e Skorohod 1972, cap. 6) e viene indicato come processo Browniano geometrico (Karlin e Taylor 1975, pag. 357), per il quale le due barriere 0 e ∞ sono inaccessibili in un tempo finito (Karlin e Taylor 1981, pag. 228-230).

Il termine m indica il tasso di variazione della variabile x_t, mentre la varianza istantanea del processo per unità di tempo è data da $\sigma^2 x_t^2$. L'equazione differenziale stocastica (1) implica che per ogni dato t il valore corrente x_t è noto con certezza mentre i valori futuri sono caratterizzati da un'incertezza crescente linearmente. Non vi è, perciò, nessun processo di apprendimento e, sebbene informazioni si rendono via via disponibili, le realizzazioni future del processo rimangono incerte.

Un semplice esempio che può essere adattato alla (1) è quello di un impianto industriale o macchina la cui capacità produttiva corrente è descritta dalla variabile x_t che, in assenza di interventi, tende a deteriorarsi nel tempo al tasso $m < 0$, e sulla quale un regolatore predisposto al controllo della produzione decide di imporre una barriera inferiore, $l \leq x_0$.

Per mantenere la capacità produttiva sopra la barriera, il regolatore applica un controllo attraverso una iniezione di manutenzione in grado di restituire alla macchina la sua efficienza primitiva. Formalmente questo equivale a definire, a partire da $\{x_t\}$, un processo stocastico regolato $\{z_t,\ t \geq 0\}$ del tipo:

$$(2) \qquad z_t = x_t L_t, \quad l \leq z_t < \infty, \quad \text{dove } L_t = \begin{cases} 1 & 0 \leq t < \tau \\ \max\limits_{\tau < s < t} (l/x_s) & t \geq \tau \end{cases}$$

Fintanto che il processo $\{x_t\}$ rimane sopra la barriera l si ha $z_t \equiv x_t$ e $L_t = 1$. La prima volta che la barriera viene toccata, per esempio all'istante τ, il controllo entra in atto per cui $L_\tau = 1$ e $z_\tau = l$.

Poiché $\{L_t,\ t \geq 0\}$ è definito a partire dal processo primitivo continuo $\{x_t\}$ anch'esso risulterà un processo continuo, con in più la caratteristica di essere non decrescente e a variazione finita, cioè cresce solamente quando $z_t = l$1.

$$dL_t = \begin{cases} > 0 & \text{quando } z_t = l \\ = 0 & \text{in tutti gli altri casi} \end{cases}$$

Le proprietà sopra descritte identificano in modo univoco il processo $\{L_t\}$ (Harrison, 1985, proposizione 6, pag. 22), il quale mantiene, attraverso un livello infinitesimo di manutenzione, il processo $\{z_t\}$ sempre sopra la barriera inferiore l. Il controllo entra in atto solamente quando $\{z_t\}$ oltrepassa l con probabilità uno.

Applicando il Lemma di Ito è possibile riscrivere il processo regolato (2) in termini differenziali:

$$(3) \qquad dz_t = L_t dx_t + x_t dL_t = mz_t dt + \sigma z_t dw_t + dL_t^o \qquad z_0 = z\,(\equiv x) \in [l,\ \infty)$$

dove $dL_t^o = x_t dL_t = z_t dL_t / L_t$.

Poiché il processo $\{L_t\}$ interpreta la regolazione imposta alla *performance* produttiva $\{x_t\}$, il termine DL_t^o rappresenta l'iniezione infinitesima di manutenzione in grado di riportare la produttività alla efficienza programmata l. Infatti, per la (2), se $z_t = l$ abbiamo che $dz_t = 0$ e il tasso al quale cresce

$\{L_t\}$ è uguale al tasso al quale decresce $\{x_t\}$ in modo che il livello di z rimanga invariato.

L'imposizione di una barriera inferiore non è generalmente esente da costi per il regolatore. Per semplicità assumiamo questo costo proporzionale al tasso di manutenzione:

$$dC_t = cdL_t^o = cz_t(dL_t/L_t) \qquad \text{con } c > 0$$

Conseguentemente, anche $\{C_t,\ t \geq 0\}$ risulta un processo continuo a variazione finita che cresce solamente quando $\{L_t\}$ cresce, cioè quando $z_t = l$.

Riassumendo, il modello fin qui descritto considera la capacità produttiva di un impianto industriale che si deteriora nel tempo e dove l'intervento di manutenzione da parte di un tecnico regolatore avviene in tempi discreti e a un costo fisso ogniqualvolta la produttività scende al di sotto di un livello prestabilito.

Inoltre, il regolatore possiede l'opzione – che può esercitare ogniqualvolta la produttività tocca la barriera inferiore – se continuare a intervenire oppure porre termine all'intervento di manutenzione. Per esempio, l'intervento potrebbe continuare fin tanto che l'accumulo di spese di manutenzione non supera un livello di guardia messo in relazione con il valore di realizzo dell'impianto una volta che ne venisse decisa la sostituzione, cioè fin tanto che il ricavo operativo complessivo che si ottiene dall'uso dell'impianto non sia uguale al suo valore di realizzo.

Questo aspetto del problema può essere inserito nel modello introducendo una barriera di assorbimento b al processo non regolato $\{x_t\}$ e quindi, attraverso la (2), al processo $\{z_t\}$. Ovviamente, perché la barriera di assorbimento sia tale, deve essere $b \leq l$. Il livello di assorbimento, quindi, equivale di fatto a introdurre uno *stopping time* per il sistema che può essere così definito:

$$(4)\quad T(b;\ l) = \begin{cases} \inf\{t \geq 0 | x_t \leq b\} = \inf\{t \geq 0 | z_t \leq l,\ L_t \leq l/b\} \\ \qquad\qquad\qquad\qquad \text{se } x_t \leq b \text{ per alcuni } t \geq 0 \\ \infty \qquad\qquad\qquad\qquad \text{se } x_t > b \text{ per ogni } t \geq 0 \end{cases}$$

Poiché nell'istante di assorbimento $T(b;\ 1)$ il processo regolato $\{z_t\}$ è giocoforza uguale a $z_T = l$, il livello di controllo risulterà pari a $L_T = l/b$. Inoltre, poiché zero non è una barriera accessibile per $\{x_t\}$, se $b = 0$ il momento di assorbimento non avverrebbe mai, almeno in tempo finito, per cui $T(0;\ l) = \infty$.

Per meglio comprendere il tipo di controllo esercitato, nella Figura 1

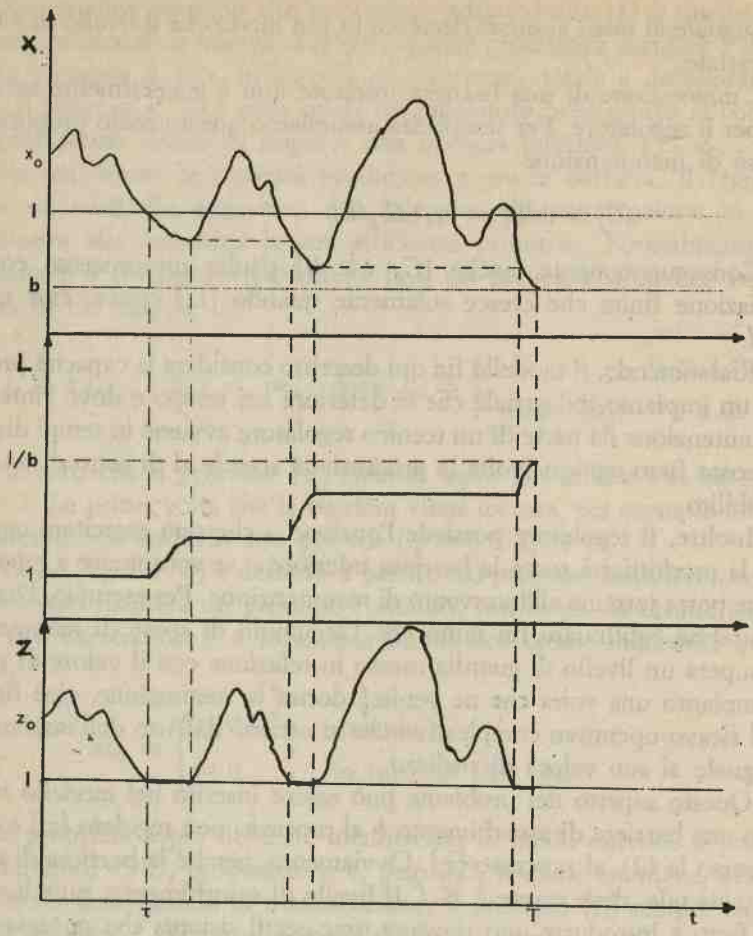

FIGURA 1.

abbiamo riportato il caso di un processo $\{x_t\}$ con barriera riflettente inferiore e *stopping time* $T(b; l)$.

Infine, se indichiamo il rendimento operativo istantaneo dell'impianto con la funzione $\phi(z_t)$, limitata e continua nell'intervallo $[l, \infty)$, con l'aggiunta delle usuali proprietà: $\phi' > 0$, $\phi'' < 0$ e $\phi'(\infty) = 0$, il valore atteso del flusso dei rendimenti netti dell'unità operativa diventa:

$$E_0\left[\int_0^T e^{-\lambda t}[\phi(z_t)\,dt - cdL_t^o] + e^{-\lambda T}\,\Psi(z_T, L_T)\right] \tag{5}$$

dove T rappresenta lo *stopping time* (4) e λ è il tasso di sconto costante nel tempo. Indichiamo la (5) con la funzione $K(z;\ l,\ T(b;\ l))$.

Poiché il processo regolato $\{z_t\}$ gode della proprietà markoviana in senso forte (Harrison, 1985, pag. 81) e $K(\cdot)$ è omogenea rispetto al tempo, il valore atteso scontato dei ricavi operativi risulta funzione delle sole condizioni iniziali $z_0 = z$. Inoltre, poiché i valori finali $z_T = l$ e $x_T = b$ ($L_T = l/b$) caratterizzano appieno lo *stopping time*, ne deriva il condizionamento rispetto a $T(b;\ l)$ [5].

La funzione $\Psi(z,\ L)$ rappresenta il valore di realizzo dell'impianto; nel momento di assorbimento risulta $K(z_T;\ l,\ T) = \Psi(z_T,\ L_T)$. Assumiamo che sia funzione crescente della produttività corrente dell'impianto come del livello totale di investimento in manutenzione messo in opera fino a quel momento. Formalmente: $\Psi_z > 0$, $\Psi_L > 0$, $\Psi_{zz} < 0$, $\Psi_{LL} < 0$, $\Psi(0,\ 0) = 0$ e $\Psi(z,\ \infty) = \bar{\Psi} > 0$. Inoltre, in virtù della relazione fra $\{x_t\}$ e $\{z_t\}$, abbiamo $\Psi_x < 0$ e $\Psi_{xx} \gtreqless 0$.

3. *Alcuni risultati*

Poiché $L_t = 1$ per tutto l'intervallo $0 \leq t \leq \tau$ e $L_T = l/b$ è possibile dividere la (5) nella somma di tre termini.

$$(6)\qquad K(z;\ l,\ T(b;\ l)) = f(z) - g(z_T,\ L_T) + E_0\left[e^{-\lambda t}\Psi(z_T,\ L_T)\right]$$

$$= E_0\left[\int_0^{\infty} e^{-\lambda t}[\phi(z_t)\,dt - c\,dL_t^o]\ |z_0 = z\right]$$

$$- E_0\left[\int_T^{\infty} e^{-\lambda t}[\phi(z_t)\,dt - c\,dL_t^o]\,[z_T = l,\ L_T = l/b\right]$$

$$+ E_0\left[e^{-\lambda T}\Psi(z_T,\ L_T)\right]$$

Il primo integrale rappresenta il rendimento che l'impresa si attende di realizzare dall'utilizzo dell'impianto quando l'orizzonte è infinito, cioè in

[5] A parte l'esempio da noi considerato molti sono gli esempi che si possono adattare alla (5). Se il problema è quello di individuare il momento ottimale per dare inizio a un progetto di investimento, allora $\Psi(x)$ è il valore presente netto del progetto quando il valore iniziale del progetto è x, e $\phi(\cdot) = c = 0$.

assenza di barriera di assorbimento. Il secondo invece valuta la perdita, in termini di rendimento atteso, in cui l'impresa incorre estromettendo la macchina dal processo produttivo all'istante T. Il terzo termine è infine il valore atteso, scontato del valore di realizzo.

Per il calcolo dei due integrali non è difficile adattare la metodologia dei processi stocastici regolati proposta da Karlin e Taylor (1975, 1981), Harrison e Taksar (1983) e Harrison (1985).

In particolare, per quanto riguarda il secondo integrale, possiamo far riferimento al seguente risultato:

Proposizione 1:

1.i) $g(z_T, L_T) = f(z_T)\, E_0(e^{-\lambda T};\ z_T = l,\ L_T = l/b)$;

1.ii) $E_0(e^{-\lambda T};\ z_T = l,\ L_T = l/b) = \left(\frac{z}{b}\right)^{-\alpha(\lambda)} \equiv \left(\frac{x}{b}\right)^{-\alpha(\lambda)}$

dove $\alpha(\lambda) = \frac{1}{\sigma^2}\left[\left(\left(m - \frac{1}{2}\sigma^2\right)^2 + 2\sigma^2\lambda\right)^{1/2} + \left(m - \frac{1}{2}\sigma^2\right)\right] > 1$

Dimostrazione: si veda l'appendice.

Poiché il rendimento atteso in assenza di barriera di assorbimento è, grazie alla markovianità del processo $\{z_t\}$, funzione delle sole condizioni iniziali, il termine $f(l)\,(x/b)^{-\alpha(\lambda)}$ è interpretabile come il valore attualizzato al tempo zero del flusso atteso dei rendimenti che l'impresa realizzerebbe se l'attività produttiva iniziasse T periodi in avanti. Il tasso di attualizzazione $(x/b)^{-\alpha(\lambda)}$ viene a dipendere, quindi, solo dall'evoluzione del processo $\{x_t\}$ fino al momento in cui per la prima volta $x_T = b$.

Ricordando che $b \leq l \leq x$, è immediato verificare che, per $b \to 0$, questo valore tende a zero, in quanto una barriera di assorbimento a zero non è mai raggiungibile dal processo $\{x_t\}$.

Lo stesso risultato, ovviamente, si può ottenere se $x \to \infty$: in questo caso il valore iniziale è caratterizzato da un livello di produttività talmente elevata che la probabilità che in un tempo finito questa scenda fino a raggiungere il livello inferiore b è nulla.

Infine, se siamo nella situazione in cui $b = l \leq x$, viene a mancare la distinzione fra barriera riflettente e barriera assorbente. Il rendimento atteso sarà dato da $f(l)\,(x/l)^{-\alpha(\lambda)}$ e lo *stopping time* coincide con il primo istante in cui il processo $\{x_t\}$ incontra il limite inferiore l, cioè $T(l;\ l) = \tau$.

Passiamo ora a valutare la funzione $f(z)$, cioè il primo integrale che compare nella (6). A tale riguardo possiamo avvalerci del seguente risultato:

Proposizione 2. Dato $\lambda > 0$, e $\phi(z_t)$ limitata superiormente su $[l, \infty)$, la variazione attesa della funzione $f(z_t)$ soddisfa la seguente equazione differenziale:

$$2.\text{i}) \quad mf_z(z_t)\, z_t + \frac{1}{2}\sigma^2 f_{zz}(z_t)\, z_t^2 - \lambda f(z_t) = -\phi(z_t)$$

con:

$$2.\text{ii}) \quad f_z(l) = c \qquad \textit{smooth pasting condition}$$

dove f_{zz} e f_z sono derivate parziali di f rispetto a z.

Dimostrazione: si veda l'appendice.

Come notato sia da Dixit (1991b) sia da Sumas (1991) ed evidenziato anche dalla dimostrazione della proposizione, la *smooth pasting condition* non riveste nessun significato di ottimalità per la regione di assimilità di $\{z_t\}$. Essa richiede solamente che la derivata prima della funzione che rappresenta il valore atteso scontato del flusso dei rendimenti futuri dell'impianto sia continua nel punto critico l. In altri termini, poiché $f(\cdot)$ è definita come il valore atteso di un integrale il cui argomento è una funzione continua e limitata $\phi(\cdot)$, la funzione $f(\cdot)$ non può risultare discontinua eccetto che nel punto critico $z_t = l$, dove si applica la regolazione. In questo punto, l'andamento di $f(\cdot)$ dipende dai costi che l'impresa sostiene, con $df = dC$.

Quale che sia la barriera l, la discussione precedente suggerisce che, quando la variabile di stato z raggiunge il valore l, un controllo infinitesimale dL^o è applicato in modo che questa salti immediatamente al livello $l + dL^o$. L'impresa paga un costo pari a cdL^o per esercitare tale controllo e ottiene un incremento del valore del flusso atteso dei rendimenti pari a $f(l + dL^o) - f(l) \cong f_z(l)\, dL^o$.

Rimane da aggiungere il valore atteso scontato di realizzo dell'impianto che abbiamo fin qui trascurato. Adottando il risultato ottenuto nella seconda parte della proposizione 1, non è difficile verificare che $E_0[e^{-\lambda T}\Psi(z_T, L_T)] \equiv \Psi(z_T, L_T)\, E_0(e^{-\lambda T};\, x_T = b) = \Psi(l, l/b)\,(x/b)^{-\alpha(\lambda)}$.

Tenendo conto delle proposizioni 1 e 2 e del valore atteso scontato di realizzo possiamo riscrivere la (6) nella forma finale:

$$(7) \quad K(z;\, l,\, T(b;\, l)) = f(z) - f(l)\left(\frac{x}{b}\right)^{-\alpha(\lambda)} + \Psi(l,\, l/b)\left(\frac{x}{b}\right)^{-\alpha(\lambda)}$$

dove $f(z)$ è ottenuta come soluzione dell'equazione differenziale 2.i). Si noti che il rendimento atteso dell'impianto al momento dello *scrapping* è uguale al valore di realizzo.

La (7) mette in luce come il valore del flusso atteso scontato dei rendimenti operativi dall'impianto, in presenza di una barriera riflettente e di una barriera assorbente legate fra loro dallo stock totale di manutenzione, sia riconducibile alla somma di due termini che dipendono dall'evoluzione congiunta dei processi $\{z_t\}$ e $\{x_t\}$. Il primo termine $f(z)$ indica il valore atteso del flusso di rendimenti in presenza di un orizzonte infinito e con barriera inferiore l, oltre la quale la produttività non può scendere. Il secondo $[\Psi(l,\, l/b) - f(l)]\,(x/b)^{-\alpha(\lambda)}$ rappresenta invece il valore atteso scontato di realizzo al netto del flusso di rendimenti che vengono perduti per la decisione di estromettere la macchina all'istante $T(b;\, l)$.

Sebbene nell'istante iniziale sia $z \equiv x$, il tasso di sconto $(x/b)^{-\alpha(\lambda)}$ dipende esclusivamente dalla dinamica del processo $\{x_t\}$, per cui mantenere distinti i processi $\{z_t\}$ e $\{x_t\}$ consente di distinguere chiaramente l'effetto della barriera riflettente dall'effetto della barriera assorbente. Inoltre, questa distinzione, assieme alla forma additiva della (7), consente di mostrare come i risultati fin qui ottenuti possano essere riassunti in un'unica proposizione indicando con $K(z,\, x;\, l,\, T(b;\, l)) = K^1(z;\, l) + K^2(x;\, T(b;\, l))$ la funzione soluzione della (6).

Proposizione 3. Dato $\lambda > 0$, e $\phi(z_t)$ limitata superiormente su $[l,\, \infty)$, la variazione attesa della funzione $K(z_t,\, x_t) = K^1(z_t) + K^2(x_t)$ soddisfa le seguenti equazioni differenziali:

$$3.\text{i}) \quad mK^1_z(z_t)\, z_t + \frac{1}{2}\sigma^2 K^1_{zz}(z_t)\, z_t^2 - \lambda K^1(z_t) = -\phi(z_t)$$

$$3.\text{ii}) \quad mK^2_x(x_t)\, x_t + \frac{1}{2}\sigma^2 K^2_{xx}(x_t)\, x_t^2 - \lambda K^2(x_t) = 0$$

con:

$$3.\text{iii}) \quad K^1_z(l) = c \qquad \textit{smooth pasting condition}$$

$$3.\text{iv}) \quad K^1(z_T) + K^2(x_T) = K^1(l) + K^2(b) = \Psi(z_T,\, z_T/x_T) = \Psi(l,\, l/b), \qquad \textit{value matching condition}$$

Dimostrazione: si veda l'appendice.

Dalla condizione 3.iv) si deduce che, nel caso in cui $b = l1$, lo *scrapping time* coincide con il primo istante in cui la produttività scende al livello l: l'impresa non interviene mai con spese di manutenzione e si limita ad avere un rendimento al momento dell'espulsione uguale al suo valore di realizzo.

Nel caso in cui $b < l$, una politica di manutenzione svolge il compito di prolungare la vita operativa dell'impianto. Da un lato l'impresa sostiene un costo aggiuntivo ogniqualvolta z_t tocca la barriera inferiore, dall'altro ottiene un flusso maggiore di rendimenti espresso da $K^1(\cdot)$. La condizione di un valore dell'impianto uguale al valore di realizzo nel momento di assorbimento rimane comunque la stessa.

Queste considerazioni implicano che la sola condizione 3.iv), la quale prende il nome di *value matching condition*, non sia sufficiente per determinare l'istante temporale ottimale in cui escludere la macchina.

Infine, come per la proposizione 2, anche in questo caso la condizione 3.iii) (*smooth pasting condition*) non assume nessun significato di ottimalità per individuare la regione di amissibilità per $\{z_t\}$; richiede solo la continuità della derivata prima della funzione $K^1(\cdot)$ nel punto l.

4. *Calcolo di* K *per dati valori di* l *e* b.

Siamo in grado ora di affrontare la ricerca di una soluzione per la (6), facendo uso di risultati presenti nella proposizione 3. Mostreremo pure come la soluzione trovata coincida con quella che si otterrebbe facendo ricorso alle proposizioni 1 e 2.

Per ottenere una soluzione in forma esplicita, assumiamo che i rendimenti operativi siano esprimibili attraverso una funzione ad elasticità costante del tipo $\phi(z_t) = pz_t^{\nu}$, con $0 \leq \nu \leq 1$ e $p > 0$.

Entrambe le equazioni differenziali 3.i) e 3.ii) sono del tipo lineare in $K^1(\cdot)$ e $K^2(\cdot)$. Inoltre la 3.ii) è già in forma omogenea. La soluzione generale è quindi formata dalle soluzioni delle equazioni differenziali omogenee associate più una soluzione particolare per la 3.i). Cioè:

$$(8) \qquad K(z, x; A_1, A_2, B_1, B_2) = K^1(z; A_1, A_2) + K^2(x; B_1, B_2)$$

con

$$K^1(z; A_1, A_2) = V(z) + A_1K^{11}(z) + A_2K^{12}(z)$$

$$K^2(x; B_1, B_2) = B_1K^{21}(x) + B_2K^{22}(x)$$

Con $V(\cdot)$ indichiamo la soluzione particolare dell'equazione differen-

ziale 3.i). $K^{11}(\cdot)$ e $K^{12}(\cdot)$, $K^{21}(\cdot)$ e $K^{22}(\cdot)$ sono le due coppie di soluzioni delle equazioni omogenee mentre A_1 e A_2, B_1 e B_2 sono delle costanti da determinare.

4.1. *Soluzione delle equazioni omogenee.* – Iniziamo con l'analizzare l'equazione differenziale 3.i). Trascurando il termine $\phi(z_t)$ otteniamo un'equazione omogenea che ricorda un'equazione differenziale di Eulero. La soluzione può essere ottenuta, per sostituzione, a partire da una forma funzionale del tipo z_t^{ϑ}.

$$A_1K^{11}(z) + A_2K^{12}(z) = A_1z^{\vartheta 1} + A_2z^{-\vartheta 2} \tag{9}$$

dove $\vartheta 1$ e $-\vartheta 2$ sono le radici dell'equazione caratteristica:

$$\frac{1}{2}\sigma^2\vartheta^2 + \left(m - \frac{1}{2}\sigma^2\right)\vartheta - \lambda = 0$$

con: $$\vartheta_1(\lambda) = -\frac{1}{\sigma^2}\left[\left(m - \frac{1}{2}\sigma^2\right) - \sqrt{\left(m - \frac{1}{2}\sigma^2\right)^2 + 2\lambda\sigma^2}\right] > 0$$

$$-\vartheta_2(\lambda) \quad \frac{1}{\sigma^2}\left[\left(m - \frac{1}{2}\sigma^2\right) + \sqrt{\left(m - \frac{1}{2}\sigma^2\right)^2 + 2\lambda\sigma^2}\right] > 1$$

Per quanto riguarda l'equazione differenziale 3.ii) si nota immediatamente che, a parte la sostituzione di variabile: x al posto di z, essa è uguale all'omogenea associata alla 3.i). La soluzione sarà pertanto, a meno delle costanti, identica a quella proposta in (9):

$$B_1K^{21}(x) + B_2K^{22}(x) = B_1x^{\vartheta 1} + B_2x^{-\vartheta 2} \tag{10}$$

4.2. *Soluzione particolare.* – Come abbiamo già accennato, la necessità di trovare una soluzione particolare riguarda solamente l'equazione 3.i). Una conveniente soluzione può essere quella di considerare il flusso atteso scontato dei rendimenti operativi trascurando sia la barriera l sia i costi della regolazione. Inoltre, poiché stiamo procedendo ignorando ogni forma di *scrapping* dell'impianto, l'orizzonte temporale è infinito.

Adottando la funzione $\phi(z) = pz^{\nu}$ e tenendo conto che in assenza di regolazione $z_t \equiv x_t$, per ogni $t \geq 0$, abbiamo:

$$(11)\quad V(z) = E_0\left[\int_0^\infty e^{-\lambda t} p z_t^\nu \, dt \,|z_0 = z\right] = \left[\int_0^\infty e^{-\lambda t} p E_0(z_t^\nu)\, dt \,|z_0 = z\right]$$

$$= \int_0^\infty p z^\nu e^{-\left(\lambda - m\nu - \frac{1}{2}\nu(\nu-1)\sigma^2\right)} dt = M z^\nu$$

$$\text{con } M = \frac{p}{\lambda - m\nu - \frac{1}{2}\nu(\nu-1)\sigma^2} > 0$$

Il secondo segno di uguaglianza discende dall'applicazione del teorema di Fubini, il terzo, invece, dalle proprietà di lognormalità del processo (1) (Harrison, 1985). Inoltre, affinché la soluzione particolare esista e abbia valore finito positivo, bisogna che $\lambda - m\nu - \frac{1}{2}\nu(\nu-1)\sigma^2 > 0$. Poiché $0 < \nu < 1$ e $m < 0$, questa condizione è sempre soddisfatta.

4.3. *Valore dell'impianto per valori fissati di* l *e* b. – Sommando (9), (10) e (11) il valore atteso dei rendimenti operativi diventa:

$$(12)\quad K(z, x; A_1, A_2, B_1, B_2) = Mz^\nu + A_1 z^{\vartheta 1} + A_2 z^{-\vartheta 2} + B_1 x^{\vartheta 1} + B_2 x^{-\vartheta 2}$$

Il primo termine che compare nel lato sinistro della (12) rappresenta il flusso atteso di rendimenti che l'impresa è in grado di ottenere dalla macchina in assenza di manutenzione e di *stopping time*. I termini rimanenti rappresentano il valore dell'opzione a introdurre una regolazione sulla produttività dell'impianto e sul livello totale della regolazione attraverso il momento di estromissione dell'impianto dal processo produttivo.

Questa interpretazione ci consente di individuare alcune condizioni di contorno. Per esempio, nel caso in cui x_t e z_t assumono valori molto alti, la probabilità che le barriere inferiori b e l vengano raggiunte, almeno in un tempo finito, tende a zero. In altri termini, l'opzione che queste barriere siano attivate dal regolatore perde di valore, per questo si pone $A_1 = B_1 = 0$.

Rimangono da determinare le costanti A_2 e B_2 e ciò può essere fatto ricorrendo alle condizioni 3.iii) e 3.iv). Riscritte tenendo conto della (12) tali condizioni diventano:

$$(13a)\qquad K_z^1(l) = V_z(l) - \vartheta_2 A_2 l^{-\vartheta 2 - 1} = c$$

(13b) $K^1(l) + K^2(b) = V(l) + A_2 l^{-\vartheta 2} + B_2 b^{-\vartheta 2} = \Psi(l,\ l/b)$

Notando che $\vartheta 2(\lambda) = \alpha(\lambda)$, risolvendo per A_2 e B_2 otteniamo:

(14a) $$A_2(l) = [V_z(l) - c]\frac{l^{1+\alpha}}{\alpha}$$

(14b) $$B_2(l,\ b) = [\Psi(l,\ l/b) - V(l) - A_2(l)\, l^{-\alpha}]\, b^{\alpha}$$

Sostituendo le constanti così individuate nella (12) abbiamo:

(15) $$K(z,\ x;\ l,\ T(b;\ l)) = V(z) + [V_z(l) - c]\frac{l^{1+\alpha}}{\alpha} z^{-\alpha}$$

$$+[\Psi(l,\ l/b) - V(l) - A_2(l)\, l^{-\alpha}]\left(\frac{x}{b}\right)^{-\alpha}$$

Con riferimento alla soluzione generale (8), i primi due termini alla destra della (15) rappresentano $K^1(z;\ l)$, mentre il terzo termine indica $K_2(x;\ T(b;\ l))$, a cui solo è associato l'effetto della barriera assorbente b.

Facendo uso della (11), la (15) diventa:

(15') $$K(z,\ x;\ l,\ T(b;\ l)) = Mz^{\nu} + \frac{1}{\alpha}[\nu M l^{\nu+\alpha} - c l^{1+\alpha}]\, z^{-\alpha}$$

$$-\frac{1}{\alpha}[(\nu+\alpha) M l^{\nu} - c l]\left(\frac{x}{b}\right)^{-\alpha} + \Psi(l,\ l/b)\left(\frac{x}{b}\right)^{-\alpha}$$

Ponendo $f(z) = Mz^{\nu} + \frac{1}{\alpha}[\nu M l^{\nu+\alpha} - c l^{1+\alpha}]\, z^{\alpha}$, da cui risulta $f(l) = \frac{1}{\alpha}[(\nu+\alpha) M l^{\nu} - c l]$, è immediato verificare che la soluzione individuata con la proposizione 3 equivale a quella ottenuta con le proposizioni 1 e 2 ed espressa dalla (7).

5. *Condizioni per un controllo ottimo*

Rimangono da individuare le condizioni attraverso le quali il regolatore identifica il livello ottimale per la barriera riflettente e per il momento di *scrapping* dell'impianto.

L'impresa ha interesse a esercitare l'opzione di introdurre la barriera

riflettente e lo *stopping time* solo nel caso in cui queste garantiscano un aumento del flusso atteso scontato dei rendimenti. Nei termini della (15) ciò si configura con la situazione in cui le costanti A_2 e B_2 siano positive.

Dalle (14) si evince che la costante A_2 risulta positiva se, nel punto critico $l > 0$, il rendimento marginale in assenza di regolazione è superiore al costo marginale della stessa. Per la barriera assorbente, B_2 risulta positiva se $\Psi(l,\ l/b) - V(l) - A_2(l)\, l^{\alpha} = \Psi(l,\ l/b) - K^1(l) > 0$ con $b > 0$: nell'istante terminale il valore di realizzo deve essere sempre superiore al valore dei rendimenti operativi continuando a spendere in manutenzione.

Inoltre, poiché la costante B_2 è funzione sia di l che di b, la scelta delle due barriere non può che avvenire in modo simultaneo. Fissare il tempo di vita dell'impianto inevitabilmente condiziona la scelta della politica di manutenzione e viceversa.

Tuttavia, invocando l'additività della funzione $K(\cdot)$ nelle due variabili z e x è possibile affrontare singolarmente la determinazione delle due barriere. Le due proposizioni che seguono definiscono le condizioni necessarie e sufficienti per la scelta ottima di una barriera, condizionatamente a un valore predeterminato per l'altra.

Iniziamo con la barriera assorbente. Poiché b compare solamente nella costante B_2, massimizzare la (15) rispetto a b equivale a trovare dei valori stazionari per B_2. Questo conduce alla seguente proposizione:

Proposizione 4. Per ogni dato valore $l > 0$, condizioni necessarie e sufficienti per la scelta ottimale della barriera di assorbimento b sono:

4.i) $K_x^2(b) = \Psi_x(l,\ l/b)$ — *smooth pasting condition* (condizione necessaria)

4.ii) $(1 + \alpha)\, \Psi_x(l,\ l/b)\, b^{-1} + \Psi_{xx}(l,\ l/b) \leq 0$ (condizione sufficiente)

Dimostrazione: si veda l'appendice.

Diversamente da quanto ottenuto nelle proposizioni 2 e 3 in questo caso la *smooth pasting condition* riveste il ruolo di condizione del primo ordine per determinare l'ottimo istante di assorbimento. Essa non è altro che la naturale estensione della *value matching condition* della proposizione 3.

Un aumento infinitesimo della barriera di assorbimento comporta, quale che sia il valore della barriera riflettente $l \geq b$, una diminuzione del tempo atteso di vita dell'impianto e quindi un decremento del valore atteso del flusso scontato dei rendimenti operativi pari a $K_x^2(b)\, db$, a cui fa fronte un

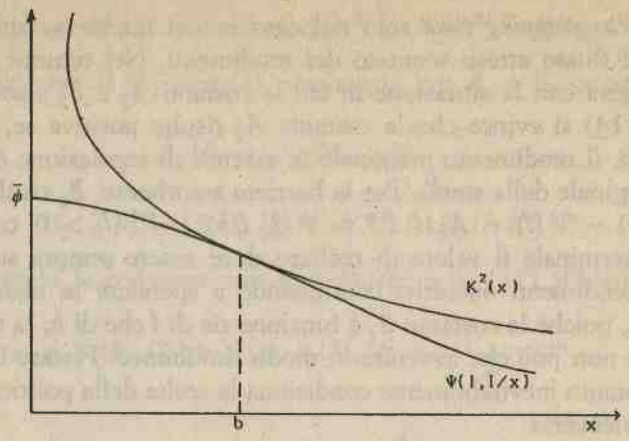

FIGURA 2.

decremento in termini di valore di realizzo pari a $\Psi_x(l, l/b)\,db$. Affinché l'impresa non trovi conveniente usare la differenza che si può formare tra questi due termini per migliorare il rendimento atteso, è necessario che la funzione $K^2(\cdot)$ sia continua e sufficientemente liscia – da cui il nome della condizione – nel punto di ottimo $x_T = b$.

Ricordando che l'impresa esercita l'opzione di introdurre uno *stopping time* se questo aumenta il valore atteso dell'impianto, anche la condizione di sufficienza impone alcune riflessioni.

Si consideri il caso in cui $\Psi \geq 0$ è costante e tale per cui $\Psi < V(l) + A_2(l)\, l^{-\alpha} = K^1(l)$. Poiché un aumento di b comporta una diminuzione del tempo atteso di vita della macchina con aumento dei rendimenti operativi perduti dall'impresa come conseguenza del riavvicinamento dello *scrapping time*, un regolatore che desiderasse rendere massima la (15) sceglierebbe sempre $b = 0$. In questo caso lascerebbe che la produttività dell'impianto venisse regolata al livello l, ogniqualvolta questo fosse necessario, senza mai deciderne l'espulsione.

D'altro canto se $\Psi > 0$, costante e $\Psi > V(l) + A_2(l)\, l^{-\alpha} = K^1(l)$, l'incremento di rendimenti operativi perduti a fronte di un aumento di b verrebbe più che compensato dalla crescita di valore atteso di realizzo che quell'aumento comporta. La scelta, in questo caso, non può che essere quella di estromettere l'impianto il più presto possibile e, nell'ipotesi di l fissato, questo comporta $b = l$.

Infine abbiamo visto il caso in cui $\Psi > 0$, con $\Psi_x < 0$. A fronte di uno spostamento verso l'alto di b, l'aumento di valore atteso di realizzo cresce meno di quanto aumenti la perdita attesa in termini di rendimenti

operativi. È possibile, condizionatamente a Ψ_{xx}, che vi sia un valore di $0 < b < l$ che eguagli al margine questi due aumenti. Le proprietà assunte per la funzione di realizzo garantiscono che ciò possa avvenire.

Un semplice diagramma è in grado di esemplificare la scelta di b. Dalla (12) è immediato verificare che $K^2(x)$ è funzione decrescente e convessa in x, mentre $\Psi(l, l/x)$ è funzione decrescente e concava in x (Figura 2). Fissato l, la costante B_2 deve aggiustarsi in modo tale che $K^2(\cdot)$ sia tangente alla funzione di realizzo $\Psi(\cdot)$; il punto di tangenza definisce il valore ottimale per b.

Passiamo ora a individuare le condizioni per determinare la barriera riflettente l.

Proposizione 5. Per ogni dato valore $0 < b < l$, condizioni necessarie e sufficienti per la scelta ottimale della barriera riflettente l sono:

5.i) $$K^1_{zz}(l) = \frac{b^\alpha}{l^\alpha - b^\alpha}\alpha l^{-1}\left(c - \Psi_z(l, l/b)\right)$$ *super contact condition* (condizione necessaria)

5.ii) $$K^1_{zzz}(l) + K^1_{zz}(l)\left(\frac{l^{-1}}{l^\alpha - b^\alpha} + \frac{\Psi_{zz}(l, l/b)}{c - \Psi_z(l, l/b)}\right) \leq 0$$ (condizione sufficiente)

Dimostrazione: si veda l'appendice.

La *super contact condition* 5.i) non è la semplice estensione della *smooth pasting condition* della proposizione 3, come per Dumas (1991) e Dixit (1991b). Essa tiene conto dell'esistenza di un vincolo sullo stock totale di manutenzione oltre il quale non è conveniente spendere per mantenere in attività l'impianto, vincolo espresso dalla condizione « terminale » (13b).

Infatti, se da un lato un aumento di l comporta un aumento del flusso atteso dei rendimenti operativi associati con l'opzione a introdurre la barriera riflettente, $K^1(\cdot)$, dall'altro richiede un aumento delle spese di manutenzione fino al momento di espulsione che diminuisce il valore dell'opzione ad estromettere la macchina, $K^2(\cdot)$. Tenendo conto della (13b), un aumento infinitesimale di l comporta una variazione netta dei rendimenti nel momento di espulsione pari a $[K^1_z(l) - \Psi_z(l, l/b)]\,dl$.

Inoltre, il valore marginale dei rendimenti in presenza di manutenzione è dato dalla funzione $K^1_z(\cdot)$ continua nel punto critico l e uguale a c,

per cui un aumento infinitesimo di l registrerà una variazione pari a $K^1_{zz}(l)\,dl$. Affinché la scelta di l risulti ottimale, è necessario che, a meno di un fattore di proporzionalità, questi due incrementi siano uguali.

Anche in questo caso un semplice diagramma adattato da quello proposto da Harrison, Selke e Taylor (1983) oppure da quello di Dixit (1991b), ci consente di mostrare come avvenga la scelta di l. Introduciamo per l'occasione la funzione:

$$H(z) = P(l,\ b)\left[cz - \Psi(z,\ z/b)\right] \tag{16}$$

$$\text{con } P(l,\ b) = \frac{b^{\alpha}}{l^{\alpha} - b^{\alpha}}\,\alpha l^{-1} > 0$$

Considerando che $\Psi(\cdot)$ è concava in z, il segno che assumono le derivate prima e seconda della (16) può essere così specificato:

$$H_z(z) = P(l,\ b)\left[c - \Psi_z(z,\ z/b)\right] = \begin{cases} < 0 & \text{se} \quad z < z^o \\ > 0 & \text{se} \quad z > z^o \end{cases} \tag{17a}$$

$$H_{zz}(z) = -\,P(l,\ b)\,\Psi_{zz}(z,\ z/b) > 0 \tag{17b}$$

Inoltre, poiché la soluzione particolare $V(z)$ è una funzione concava in z, dalla (12) è immediato verificare che l'andamento di $K^1_z(z)$ è quello indicato nella Figura 3 [6].

Fissato un valore per b, per individuare la barriera riflettente, le costanti A_2 e B_2 (o equivalentemente A_2 e P) si devono aggiustare in modo che $K^1_z(\cdot)$ risulti tangente alla funzione $H(z)$.

In riferimento alla funzione $H(\cdot)$, possiamo distinguere due possibili soluzioni: nel tratto in cui $z < z^o$ la soluzione si troverà nella parte discendente di $K^1_z(\cdot)$, mentre per il tratto in cui $z < z^o$, la soluzione si troverà nella parte ascendente di $K^1_z(\cdot)$.

Per quanto riguarda la condizione di sufficienza, merita di essere sottolineato il fatto che, mentre nel primo caso – con $K^1_{zz}(l_1)$ sempre negativa e $K^1_{zzz}(l_1)$ che lo può essere – la condizione del secondo ordine è sempre soddisfatta, nel secondo caso non si può dire altrettanto.

Infine, possiamo specializzare il risultato ottenuto nella proposizione 5 per il caso in cui la barriera assorbente assuma i valori limite $b = 0$ e $b = l$.

---

[6] Scarf (1960), Harrison e Taksar (1983) e Harrison (1985) elaborano delle condizioni generali $\phi(z)$ che garantiscono che $V(z)$ sia concava. Le condizioni da noi imposte sono comunque sufficienti affinché questo sia vero.

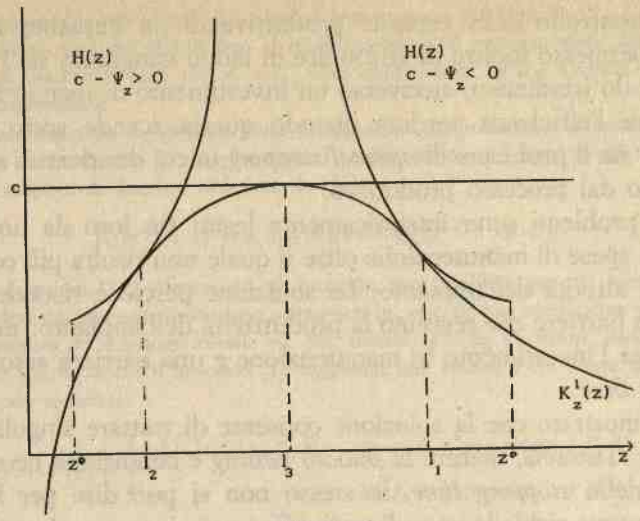

FIGURA 3.

Corollario. Nei casi limite in cui $b = 0$ e $b = l$, la condizione necessaria per la scelta di l diventa:

C.i) $K^1_{zz}(l) = 0$ se $b = 0$

C.ii) $K_x(l) = \Psi_x(l, 1)$ se $b = l$

Dimostrazione: si veda l'appendice.

Se $b = 0$, la scelta di l è riconducibile nella Figura 3 al punto di tangenza fra la funzione $K^1_z(\cdot)$ e la retta c. Solamente in questo caso la *super contact condition* C.i) si riduce ad essere la naturale estensione della *smooth pasting condition* 3.iii), come in Dumas (1991) e Dixit (1991b).

6. *Conclusioni*

Questo lavoro ha avuto il preciso intento di presentare, con un apparato matematico-probabilistico non estremamente complesso, alcuni risultati in tema di ricerca operativa sui processi Browniani regolati che hanno di recente trovato ampio impiego nella soluzione di problemi economici connessi con decisioni in condizioni di incertezza e in presenza di irreversibilità.

L'esempio che abbiamo utilizzato per presentare questi risultati è stato

quello del controllo della capacità produttiva di un impianto industriale. Questo ha permesso inoltre di affrontare in modo congiunto sia il problema di un controllo istantaneo, attraverso un investimento di manutenzione atto a ripristinare l'efficienza perduta quando questa scende sotto un livello prestabilito, sia il problema di *optimal stopping* in cui decidere di estromettere l'impianto dal processo produttivo.

I due problemi sono intrinsicamente legati fra loro da un livello di accumulo di spese di manutenzione oltre il quale non risulta più conveniente la tenuta in attività dell'impianto. La soluzione perciò è riconducibile alla scelta di due barriere che regolino la produttività dell'impianto: una barriera riflettente per l'investimento in manutenzione e una barriera assorbente per lo *scrapping time*.

Viene mostrato che la soluzione consente di trattare singolarmente le due barriere. Tuttavia, mentre la *smooth pasting* è condizione necessaria per l'ottimalità dello *stopping time*, lo stesso non si può dire per la barriera riflettente: questa richiede un ordine di differenziazione maggiore, individuata dalla *super contact condition*.

APPENDICE

Proposizione 1. Sia $T < \infty$ uno *stopping time* arbitrariamente scelto e si definisca $z_t^* \equiv z_{T+t}$, $L_t^* \equiv L_{T+t}/L_T$ per ogni $t \geq 0$. Si introduca, inoltre, il seguente funzionale:

$$f(z, L) \equiv \int_0^\infty e^{-\lambda t}[\phi(z_t)\,dt - cz_t(dL_t/L_t)] \tag{A1}$$

Adottando una sostituzione di variabile e applicando la proprietà markoviana in senso forte del processo $\{z_t\}$ (Harrison, 1985, pag. 81, proposizione 7), otteniamo:

$$\begin{aligned}
(A2)\quad E_0\left[\int_T^\infty e^{-\lambda t}[\phi(z_t)\,dt - cz_t(dL_t/L_t)]\right] &= \\
&= E_0\left[e^{-\lambda T}\int_0^\infty e^{-\lambda t}[\phi(z_t^*)\,dt - cz_t^*(dL_t^*/L_t^*)]\right] \\
&= E_0\left[e^{-\lambda T} f(z^*, L^*)\right] \\
&= E_0\left\{e^{-\lambda T} E_0\left[f(z^*, L^*)|\mathfrak{F}_T\right]\right\} \\
&= E_0\left[e^{-\lambda T} f(z_0^*, L_0^*)\right] \\
&= f(z_T)\, E_0(e^{-\lambda T};\ z_T = l,\ L_T = l/b)
\end{aligned}$$

dove $\mathfrak{F}_T$ rappresenta la più piccola σ – algebra appartenente alla famiglia $\{\mathfrak{F}_t, t \geq 0\}$ contenente tutte le informazioni disponibili fino al momento in cui il processo viene assorbito, mentre la proprietà di markov implica che z^* e L^* sono indipendenti da $\mathfrak{F}_T$. Questo prova la prima parte della proposizione.

Per quanto riguarda la seconda parte è opportuno ricordare che le condizioni $z_T = l$ e $L_T = l/b$, che caratterizzano lo *stopping time* T, corrispondono al caso in cui il processo non regolato $\{x_t\}$ colpisce la barriera inferiore b, cioè:

$$E_0(e^{-\lambda T};\ z_T = l,\ L_T = l/b) \equiv E_0(e^{-\lambda T};\ x_T = b) \tag{A3}$$

Possiamo, quindi, far riferimento al processo non regolato per il calcolo della (A3).

Si consideri un processo Browniano aritmetico $\{y_t, t \geq 0\}$ con coefficiente di trend $\mu < 0$, varianza istantanea σ^2 e valore iniziale y_0. Sia, inoltre, $q \leq y_0$ un valore costante positivo e $T(q)$ il primo istante in cui il processo $\{y_t\}$ raggiunge tale valore. $T(q)$ è uno *stopping time* definito nel modo seguente:

$$T(q) = \begin{cases} \inf(t \geq 0 \mid y_t \leq q) & \text{se } y_t \leq q \text{ per alcuni } t \geq 0 \\ \infty & \text{se } y_t > q \text{ per ogni } t \geq 0 \end{cases}$$

Si consideri ora la seguente espressione:

$$V(T \wedge t) = \exp[\beta y_{T \wedge t} - \lambda(\beta)(T \wedge t)] \tag{A4}$$

dove $T \wedge t = \min(T, t)$, $\lambda(\beta) = \beta\mu + \frac{1}{2}\beta^2\sigma^2$ è una costante da determinare. Ci si concentrerà su valori di β tali per cui $\lambda(\beta) \geq 0$.

Poiché $\{y_t\}$ è un processo di martingala anche $V(T \wedge t)$ è un processo di martingala (Karlin e Taylor, 1975, pag. 357), per cui è valida la seguente relazione:

$$E_0[V(T \wedge t)] = E_0[V(0)] = \exp(\beta y_0) \tag{A5}$$

Assumendo $\beta \leq 0$ sufficientemente piccolo in modo da assicurare che $\lambda(\beta) \geq 0$, ne discende che:

$$0 \leq V(T \wedge t) \leq \exp(\beta q)$$

In altri termini, $\{V(T \wedge t)\}$ è una famiglia di variabili casuali limitate superiormente, per cui facendo tendere $t \to \infty$ nella (A4), se lo *stopping time* T è finito, $\min(T, t) = T$ e il processo toccherà prima o poi la barriera inferiore q con probabilità uno (Karlin e Taylor, 1975, Lemma 3.3, capitolo 6, pag. 258). Cioè:

$$\lim_{t \to \infty} V(T \wedge t) = \begin{cases} \exp[\beta q - \lambda(\beta) T] & \text{se } T < \infty \\ 0 & \text{se } T = \infty \end{cases} \tag{A6}$$

Mettendo assieme la proprietà di un processo di martingala (A5) con il risultato (A6) si ottiene:

$$\exp(\beta y_0) = E_0[V(0)] = E_0[V(T)] = \lim_{t \to \infty} E_0[V(T \wedge t)] = E_0[\exp(\beta q - \lambda(\beta) T)] \tag{A7}$$

da cui riarrangiando:

$$E_0[e^{-\lambda(\beta)T}] = e^{\beta(y_0 - q)} \tag{A8}$$

La costante β è ottenuta scegliendo la radice negativa dell'equazione caratteristica $\frac{1}{2}\beta^2\sigma^2 - \beta\mu - \lambda = 0$, cioè $\beta = -\alpha(\lambda)$ dove:

$$\alpha(\lambda) = \frac{1}{\sigma^2}\left[(\mu^2 + 2\sigma^2\lambda)^{1/2} + \mu\right] > 1$$

Infine, ricordando che il processo geometrico $\{x_t\}$ è ottenuto applicando il Lemma di Ito alla trasformazione $x_t = \exp(y_t)$ con $m = \mu + \frac{1}{2}\sigma^2$, ponendo $q = \log(b)$ possiamo concludere che:

$$E_0[e^{-\lambda(\beta)T}] = \left(\frac{x_0}{b}\right)^{-\alpha(\lambda)} \tag{A8'}$$

Questo prova la seconda parte della proposizione.

Proposizione 2. Per la dimostrazione di questa proposizione verrà usata la metodologia proposta da Karlin e Taylor (1981, cap. 15), Harrison e Taksar (1985) e Harrison (1985. cap. 5).

Ancora una volta invocando la proprietà markoviana in senso forte del processo $\{z_t\}$ e in virtù della omogeneità rispetto al tempo la funzione $f(z_t)$ dipende solamente dalla condizione iniziale. Assumendone la differenziabilità e applicando il Lemma di Ito otteniamo:

$$\begin{aligned} df(z_t) &= f_z(z_t)\,dz_t + \frac{1}{2}f_{zz}(z_t)\,(dz_t)^2 \\ &= f_z(z_t)\,[L_t dx_t + x_t dL_t] + \frac{1}{2}f_{zz}(z_t)\,L_t^2\,(dx_t)^2 \\ &= f_z(z_t)\,[mx_t L_t dt + \sigma x_t L_t dw_t + dL_t^o] + \frac{1}{2}f_{zz}(z_t)\,x_t^2 L_t^2\sigma^2 dt \\ &= \left[mf_z(z_t)\,z_t + \frac{1}{2}\sigma^2 f_{zz}(z_t)\,z_t^2\right]dt + \sigma f_z(z_t)\,z_t\,dw_t + f_z(z_t)\,dL_t^o \end{aligned} \tag{B1}$$

L'equazione differenziale (B1) tiene conto del fatto che per un processo a variazione finita $(dL_t)^2 = 0$. Inoltre, poiché $dL_t = 0$ eccetto quando $z_t = l$ possiamo riscrivere la (B1) nella forma:

$$df(z_t) = \left[mf_z(z_t)\,z_t + \frac{1}{2}\sigma^2 f_{zz}(z_t)\,z_t^2\right]dt + \sigma f_z(z_t)\,z_t dw_t + f_z(l)\,dL_t^o \tag{B1'}$$

La (B1') è un'equazione differenziale stocastica in $f(\cdot)$, la quale può essere integrata ricorrendo alla formula proposta da Harrison (1985, pag. 73): $Y_t X_t = Y_0 X_0 + \int_0^t Y_s dX_s + \int_0^t X_s dY_s$. Ponendo $Y_t = f(z_t)$ e $X_t = e^{-\lambda t}$, con $dX_t = -\lambda X_t dt$ e dY_t espressa dalla (B1'), otteniamo:

$$\text{(B2)} \quad e^{-\lambda t} f(z_t) = f(z) + \int_0^t e^{-\lambda s} [m f_z(z_s) z_s + \frac{1}{2}\sigma^2 f_{zz}(z_s) z_s^2 - \lambda f(z_s)] ds + \sigma \int_0^t e^{-\lambda s} f_z(z_s) z_s dw_s + f_s(l) \int_0^t e^{-\lambda s} dL_s^o$$

Prendendo il valore atteso della (B2) e facendo tendere $t \to \infty$, se valgono le seguenti condizioni:

a) $\lim_{u \to \infty} Pr[T(u) > T(l) | z_0 = z] = 0$ per $l \leq z < u < \infty$, dove $T(l)$ e $T(u)$ hanno l'usuale significato di $\inf\{t \geq 0 \mid z_t = l\}$ e $\inf\{t \geq 0 \mid z_t = u\}$ rispettivamente,

b) $f(z_t)$ è limitata nell'intervallo $[l, \infty)$,

c) $e^{-\lambda t} f_z(z_t) z_t$ è limitata nell'intervallo $[l, \infty)$,

d) $f_z(l) = c$

e) $m f_z(z_t) z_t + \frac{1}{2}\sigma^2 f_{zz}(z_t) z_t^2 - \lambda f(z_t) = -\phi(z_t)$

si ottiene l'espressione per $f(z)$ indicata nella (6).

La condizione a) sostiene che la probabilità che il processo regolato $\{z_t\}$ raggiunga infinito prima di raggiungere un qualche altro punto interno all'insieme di amissibilità $[l, \infty)$ è nulla. Poiché il processo $\{z_t\}$ è di tipo geometrico questa condizione è sempre soddisfatta (Karlin e Taylor, 1981, pag. 228-230).

Inoltre, se vale la condizione a) e la funzione $\phi(z)$ è limitata nell'intervallo $[l, \infty)$ discende che pure le condizioni b) e c) sono sempre soddisfatte. Questo prova la proposizione.

Proposizione 3. La dimostrazione è immediata seguendo quella della proposizione 2. Invocando, ancora una volta, la proprietà markoviana in senso forte dei processi $\{x_t\}$ e $\{z_t\}$, e grazie all'omogeneità rispetto al tempo, la funzione $K(z_t, x_t)$ dipende solamente dalle condizioni iniziali. Assumendone la differenziabilità e applicando il Lemma di Ito ottteniamo:

$$\text{(C1)} \quad dK(z_t, x_t) = \left[m K_z^1(z_t) z_t + \frac{1}{2}\sigma^2 K_{zz}^1(z_t) z_t^2\right] dt + \left[m K_x^2(x_t) x_t + \frac{1}{2}\sigma^2 K_{xx}^2(x_t) x_t^2\right] dt + \sigma K_z^1(z_t) z_t dw_t + \sigma K_x^2(x_t) x_t dw_t + K_z^1(l) dL_t^o$$

L'equazione differenziale (C1) tiene conto del fatto che $dx_t dz_t = dx_t dt = dz_t dt = 0$, e che $(dL_t)^2 = 0$. Inoltre $dL_t = 0$ eccetto quando $z_t = l$.

Facendo nuovamente uso della formula di integrazione per parti proposta da Harrison (1985, pag. 73) e prendendone, poi, il valore atteso, si ottiene:

(C2) $E[e^{-\lambda T} K(z_T, x_T)] = K(z, x; l, T(b; l))$

$$+ E_0 \int_0^T e^{-\lambda s} [mK_z^1(z_s) z_s + \frac{1}{2}\sigma^2 K_{zz}^1(z_s) z_s^2 - \lambda K^1(z_s)] ds$$

$$+ E_0 \int_0^T e^{-\lambda s} [mK_x^2(x_s) x_s + \frac{1}{2}\sigma^2 K_{xx}^2(x_s) x_s^2 - \lambda K^2(x_s)] ds$$

$$+ \sigma E_0 \int_0^T e^{-\lambda s} K_z^1(z_s) z_s dw_s + \sigma E_0 \int_0^T e^{-\lambda s} K_x^2(x_s) x_s dw_s + K_z^1(l) E_0 \int_0^T e^{-\lambda s} dL_s^0$$

Se valgono le seguenti condizioni:

a) $\lim_{u \to \infty} Pr[T(u) > T(l) | z_0 = z] = 0$ per $l \leq z < u < \infty$

b) $K^1(z_t)$ e $K^2(x_t)$ sono limitate nell'intervallo $[l, \infty)$ e $[b, \infty)$ rispettivamente,

c) $e^{-\lambda t} K_z^1(z_t) z_t$ e $e^{-\lambda t} K_x^2(x_t) x_t$ sono limitate nell'intervallo $[l, \infty)$ e $[b, \infty)$ rispettivamente,

d) $K_z^1(l) = c$

e) $K(z_T, x_T) = K^1(z_T) + K^2(x_T) = \Psi(z_T, z_T/x_T)$

f) $mK_z^1(z_t) z_t + \frac{1}{2}\sigma^2 K_{zz}^1(z_t) z_t^2 - \lambda K^1(z_t) = -\phi(z_t)$

g) $mK_x^2(x_t) x_t + \frac{1}{2}\sigma^2 K_{xx}^2(x_t) x_t^2 - \lambda K^2(x_t) = 0$

si ottiene l'espressione per $K(\cdot)$ come espressa nella (6). Questo prova la proposizione.

Proposizione 4. Seguendo il suggerimento proposto da Dixit (1991b), deriviamo la (15) per un generico valore iniziale x, rispetto a b:

(D1) $$\frac{\partial K}{\partial b} = \frac{\partial B_2}{\partial b} x^{-\alpha}$$

Poiché il termine $x^{-\alpha}$ è sempre positivo per qualsiasi valore di x, il segno della (D1) sarà dato dal segno della derivata della costante B_2. Ne consegue che affinché la scelta della barriera ottimale risulti consistente temporalmente, cioè indipendente dalla condizione iniziale bisogna che $\partial B_2/\partial b = 0$. Ovvero:

(D2) $$\frac{\partial B_2}{\partial b} = \left[\alpha\left[\Psi(l,\ l/b) - V(l) - A_2(l)\, l^{-\alpha}\right] b^{-1} - \Psi_L(l,\ l/b)\, l b^{-2}\right] b^{\alpha} = 0$$

Si consideri ora la derivata della (15) rispetto a x calcolata nel punto b:

(D3) $$\frac{\partial K}{\partial x}(b) \equiv \frac{\partial K^2}{\partial x}(b) = -\ \alpha\left[\Psi(l,\ l/b) - V(l) - A_2(l)\, l^{-\alpha}\right] b^{-1}$$

Mettendo a confronto la (D2) con la (D3) e tenendo conto che $\Psi_x(l,\ l/b) = -\ \Psi_L(l,\ l/b)\, l b^{-2}$, è immediato verificare che la *smooth pasting condition* 4.i) è condizione necessaria per un massimo. La condizione sufficiente è immediata derivando la (D2) rispetto a b. Questo prova la proposizione.

Proposizione 5. Nuovamente, deriviamo la (15), per un generico valore iniziale di x e z, rispetto a l:

(E1) $$\frac{\partial K}{\partial l} = \frac{\partial A_2}{\partial l} z^{-\alpha} + \frac{\partial B_2}{\partial l} x^{-\alpha}$$

I termini $z^{-\alpha}$ e $x^{-\alpha}$ sono sempre positivi e si muovono nella stessa direzione. Affinché la decisione risulti indipendente dalle condizioni iniziali, cioè sia consistente temporalmente con $z = x$, è necessario che $\partial A_2/\partial l = -\ \partial B_2/\partial l$. Derivando le costanti (14) rispetto a l, otteniamo:

(E2a) $$\frac{\partial A_2}{\partial l} = \left[V_{zz}(l) + (l+\alpha)\left(V_z(l) - c\right) l^{-1}\right] \frac{l^{1+\alpha}}{\alpha}$$

e

(E2b) $$\frac{\partial B_2}{\partial l} = \left[\Psi_z(l,\ l/b) - c - \frac{\partial A_2}{\partial l} l^{-\alpha}\right] b^{\alpha}$$

Facendo uso delle (E2), la condizione necessaria diventa:

(E3) $$\frac{\partial A_2}{\partial l}\left[\frac{1}{b^{\alpha}} - \frac{1}{l^{\alpha}}\right] = \left(c - \Psi_z(l,\ l/b)\right)$$

Si consideri ora la derivata seconda della (15) rispetto a z calcolata nel punto l:

(E4) $$\frac{\partial^2 K}{\partial z^2}(l) = \frac{\partial^2 K^1}{\partial z^2}(l) = V_{zz}(l) + (l+\alpha)\left(V_z(l) - c\right) l^{-1}$$

Confrontando la (E3) con la (E4) si ottiene la *super contact condition* 5.i) indicata nella proposizione.

La condizione sufficiente richiede che $\partial^2 A_2/\partial l^2 + \partial^2 B_2/\partial l^2 \leq 0$. Derivando la (E3) rispetto a l e la (E4) rispetto a z, e tenendo conto della *super contact condition*, dopo alcuni passaggi algebrici si ottiene la 5.ii). Questo prova la proposizione.

Corollario. La dimostrazione della prima parte è immediata. Dalla proposizione 5 si evince che nel caso in cui $b = 0$ la *super contact condition* si riduce alla C.i).

Nel caso in cui $b = l \leq x$, viene a mancare la distinzione fra barriera riflettente e barriera assorbente. Il rendimento atteso sarà dato da $K(x, T(l, l)) = V(x) + B_2 x^{-\alpha}$, con $B_2 = [\Psi(l, l) - V(l)] l^{\alpha}$ e lo *stopping time* coincide con il primo istante in cui il processo $\{x_t\}$ incontra il limite inferiore l, cioè $T(l; l) = \tau$. Determinare questo limite inferiore coincide, quindi, con la scelta dell'*optimal stopping time* come nella proposizione 4: la C.ii) non è altro che la specializzazione della 4.i).

RIFERIMENTI BIBLIOGRAFICI

Arrow K.J., Karlin S., Scarf H., *Studies in the Mathematical Theory of Inventory and Production*, Stanford: Stanford University Press, 1958.

Bartolini L., "A Note on Entry and Exit Decisions under Uncertainty", *Economic Notes*, 1992, *21*, 157-66.

Benes V.E., Shepp L.A., Witsenhausen H.S., "Some Solvable Stochastic Control Problems", *Stochastics*, 1980, *4* 39-83.

Bensoussan A., Lions J.L., "Nouvelles Méthodes en Controle Impulsionnel", *Applied Mathematics and Optimization Quarterly*, 1975, *1*, 289-312.

Bentolila S., Bertola G., "Firing Costs and Labour Demand: How Bad is Eurosclerosis?", *Review of Economic Studies*, 1990, *57*, 381-402.

Bertola G., "Irreversible Investment", Working Paper, Princeton University, 1988.

——, Caballero R.J., "Target Zones and Realignments", *American Economic Review*, 1992, *82*, 520-36.

Bjerksund P., "The Cost of a Promise to Develop an Oil Field within a Fixed Future Date", in D. Lund, B. Øksendal, eds., *Stochastic Models and Option Values*, Amsterdam: North-Holland, 1991, 103-28.

Bonini C.P., "Capital Investment under Uncertainty with Abandonment Options", *Journal of Financial and Quantitative Analysis*, March 1977, 39-51.

Brekke K.A., Øksendal B., "The High Contact Principle as a Sufficient Condition for Optimal Stopping", in D. Lund, B. Øksendal, eds., *Stochastic Models and Option Values*, Amsterdam: North-Holland, 1991, 187-208.

Brennan M.J., Schwartz E.S., "Evaluating Natural Resource Investment", *Journal of Business*, 1985, *58*, 135-57.

Caplin A., Leahy J., "State-Dependent Pricing and the Dynamics of Money and Output", *Quarterly Journal of Economics*, 1991, *56*, 683-708.

Constantinides G., "Capital Market Equilibrium with Transaction Costs", *Journal of Political Economy*, 1986, *94*, 842-62.

——, RICHARD S., "Existence of Optimal Simple Policies for Discounted-Cost Inventory and Cash Management in Continuous Time", *Operations Research*, 1978, *26*, 620-36.

DIXIT A. (1989a), "Entry and Exit Decisions under Uncertainty", *Journal of Political Economy*, 1989, *97*, 620-38.

—— (1989b), "Hysteresis, Import Penetration and Exchange Rate Pass-Through", *Quarterly Journal of Economics*, 1989, *104*, 205-28.

—— (1989c), "Intersectoral Capital Reallocation under Price Uncertainty", *Journal of International Economics*, 1989, *26*, 309-25.

—— (1991a), "Irreversible Investment with Price Ceilings", *Journal of Political Economy*, 1991, *99*, 541-57.

—— (1991b), "A Simplified Treatment of the Theory of Optimal Regulation of Brownian Motion", *Journal of Economic Dynamic and Control*, 1991, *15*, 657-73.

——, "Investment and Hysteresis", *Journal of Economic Perspectives*, 1992, *6*, 107-32.

——, *The Art of Smooth Pasting*, Vol. 55 in *Fundamentals of Pure and Applied Economics*, Chur, Switzerland: Harwood Academic Publisher, 1993.

——, PINDYCK R.S., *Investment under Uncertainty*, Princeton: Princeton University Press, 1994.

——, ROB R., "Risk-Sharing, Adjustment and Trade", CARESS Working Paper, University of Pennsylvania, 1991.

——, ——, "Switching Costs and Sectoral Adjustments in General Equilibrium with Uninsured Risk", *Journal of Economic Theory*, 1994, *62*, 48-69.

DOSI C., MORETTO M. (1994a), "Inter-Basin Water Transfers under Uncertainty: Storage Capacity and Optimal Guaranteed Deliveries", *Journal of Environmental and Resource Economics*, 1994, *4*, 331-52.

——, —— (1994b), "Nonpoint Source Environmental Externalities and Polluter's Site Quality Standards under Incomplete Information", in C. Dosi, T. Graham-Tomasi, eds., *NPS Pollution Control: Issues and Analysis*, London: Kluwer Academic Publisher, 1994, 107-36.

DUMAS B., "Super Contact and Related Optimality Conditions", *Journal of Economic Dynamics and Control*, 1991, *15*, 675-85.

EL KAROUI N., KARATZAS I., "Integration of the Optimal Risk in a Stopping Problem with Absorption", mimeo, Columbia University, 1988.

FLOOD R.P., GARBER P.M., "The Linkage Between Speculative Attack and Target Zone Models of Exchange Rates", *Quarterly Journal of Economics*, 1991, *106*, 1367-72.

GIBSON R., SCHWARTZ E., "Valuation of Long Term Oil-Linked Assets", in D. Lund, B. Øksendal, eds., *Stochastic Models and Option Values*, Amsterdam: North-Holland, 1991, 73-101.

GIKHMAN I., SKOROHOD A.V., *Stochastic Differential Equations*, New York: Springer-Verlag, 1972.

GROSSMAN S.J., LAROQUE G., "Asset Pricing and Optimal Portafolio Choice in the Presence of Illiquid Durable Consumption Goods", *Econometrica*, 1990, *58*, 25-52.

HARRISON J.M., *Brownian Motion and Stochastic Flow Systems*, New York: John Wiley & Sons, 1985.

——, SELLKE T.M., TAYLOR A.J., "Impulse Control of Brownian Motion", *Mathematics of Operations Research*, 1983, *8*, 454-66.

——, TAYLOR A.J., "Optimal Control of a Brownian Storage System", *Stochastic Processes and Their Application*, 1978, *6*, 179-94.

——, TAKSAR M.I., "Instantaneous Control of Brownian Motion", *Mathematics of Operations Research*, 1983, *8*, 439-53.

HE H., PINDYCK R.S., "Investment in Flexible Production Capacity", *Journal of Economic Dynamics and Control*, 1992, *16*, 575-99.

KARATZAS I., "Optimal Discounted Linear Control of the Wiener Process", *Journal of Optimization Theory and Applications*, 1980, *31*, 431-40.

——, SHREVE S.E., "Connections between Optimal Stopping and Singular Stochastic Control I: Monotone Follower Problems", *SIAM Journal of Control and Optimization*, 1984, *22*, 856-77.

——, ——, "Connections between Optimal Stopping and Singular Stochastic Control II: Reflected Follower Problems", *SIAM Journal of Control and Optimization*, 1985, *23*, 433-51.

KARLIN S., TAYLOR H.M., *A First Course in Stochastic Processes*, New York: Academic Press, 1975.

——, ——, *A Second Course in Stochastic Processes*, New York: Academic Press, 1981.

KOBILA T.Ø., "Partial Investment under Uncertainty", in D. Lund, B. Øksendal, eds., *Stochastic Models and Option Values*, Amsterdam: North-Holland, 1991, 167-86.

KRUGMAN P., "Target Zones and Exchange Rates Dynamics", *Quarterly Journal of Economics*, 1991, *106*, 669-82.

KRYLOV N.V., *Controlled Diffusion Processes*, Berlin: Springer-Verlag, 1980.

LEAHY J.V., "Investment in Competitive Equilibrium: The Optimality of Myopic Behavior", *Quarterly Journal of Economics*, 1993, *58*, 1105-33.

LUND D., "Financial and Non-financial Option Valuation", in D. Lund, B. Øksendal, eds., *Stochastic Models and Option Values*, Amsterdam: North-Holland, 1991, 143-63.

MAID S., PINDYCK R.S., "Time to Build, Option Value and Investment Decisions", *Journal of Financial Economics*, 1987, *18*, 7-27.

MCDONALD R., SIEGEL D.R., "Investment and the Valuation of Firms When There Is an Option to Shut Down", *International Economic Review*, 1985, *26*, 331-49.

——, ——, "The Value of Waiting to Invest", *Quarterly Journal of Economics*, 1986, *101*, 707-27.

McKean H.P., "Appendix: A Free Boundary Problem for the Heat Equation Arising from a Problem in Mathematical Economics", *Industrial Management Review*, 1965, *6*, 32-39.

Merton R., "Theory of Rational Option Pricing", *Bell Journal of Economics and Management Science*, 1973, *4*, 141-83.

Moretto M. (1994a), "Polluter's Capital Quality Standards and Subsidy-Tax Programs in Competitive Equilibrium", in C. Carraro e J. Philar, eds., *Control and Game Theoretical Models of the Environment*, Berlin: Birkhauser, 1994, forthcoming.

—— (1994b), "Entry and Exit Decisions with Ex-Ante Unobservable Operative Costs", Quaderni del Dipartimento di Scienze Economiche, Università di Padova, 1994.

Olsen T.E., Stensland G., "Optimal Shutdown Decisions in Resource Extraction", *Economic Letters*, 1988, *26*, 215-18.

——, ——, "Optimal Sequencing of Resource Pools under Uncertainty", *Journal of Environmental Economics and Management*, 1989, *17*, 83-92.

Pindyck R.S., "Irreversible Investment, Capacity Choice and the Value of the Firm", *American Economic Review*, 1988, *78*, 969-85.

——, "Irreversibility, Uncertainty and Investment", *Journal of Economic Literature*, 1991, *29*, 1110-48.

Samuelson P.A., "Rational Theory of Warrant Pricing", *Industrial Management Review*, 1965, *6*, 13-31.

Scarf H., "The Optimality of (*s*, *S*) Policies in the Dynamic Inventory Problem", in K.J. Arrow, S. Karlin, P. Suppes, eds., *Mathematical Methods in the Social Sciences*, Stanford: Stanford University Press, 1960.

Shiryayev A.N., *Optimal Stopping Rules*, Berlin: Springer-Verlag, 1978.

Shreve S.E., Lehoczky J.P., Gaver D.P., "Optimal Consumption for General Diffusions with Absorbing and Reflecting Barriers", *SIAM Journal of Control and Optimization*, 1984, *22*, 55-75.

Svensson Lars L.O., "Target Zones and Interest Rate Variability", *Journal of International Economics*, 1991, *31*, 27-34.

OPTIMAL STOCHASTIC CONTROL, REGULATED BROWNIAN PROCESSES, AND OPTIMAL STOPPING

The stochastic control problem of regulating a state variable which follows a Brownian path is considered together with that of finding the optimal time to stop production. The objective function to be maximized is the expected

discounted value of the rewards minus a lump-sum and linear cost of adjustment plus a discounted abandonment cost for absorption. The value function is computed, and necessary and sufficient conditions for the existence of an optimal policy are given. An optimal policy turns out to be characterized by both a reflecting and an absorbing barrier.

Rivista Internazionale di Scienze Economiche e Commerciali
Volume 42 (1995), N. 1-2, 125-142

VERSO IL SISTEMA ECONOMICO DIACRONICAMENTE COMPLETO: ESPOSIZIONE RAGIONATA DELLA PARTE V DI G. DEMARIA, *A NEW ECONOMIC LOGIC*, 1993

di
ACHILLE AGNATI *

1. *Premessa*

La Parte V. *Verso il sistema economico diacronicamente completo* – che dà titolo a questa comunicazione [1] – è forse la sintesi del mai pubblicato vol. IV del *Trattato di logica economica* (I, 1962; II, 1966; III, 1974) in corso di approntamento per la stampa già nel 1983, come mi scrisse in quell'anno Demaria? Dopo avere ricostruito coi colleghi Cantarelli e Montesano il libro *A New Economic Logic* (1993) [2] ancora in dattiloscritto – da cui è tolta questa Parte V costituita in gran parte dai saggi del periodo 1977-93 – l'interrogativo è per noi retorico.

Siamo storici e « storico » sta, innanzitutto, per narratore onde diciamo che è spesso un piacevole compito attrarre l'attenzione su disegni e invenzioni del Maestro che, per noi, è stato a volta a volta e destino e cammino. Ecco, allora, l'architettura con cui Demaria predispone i pezzi sulla scacchiera – Parte V: I. *L'identificazione dei rapporti economici*; II. *Le simmetrie e il punto fisso*; III. *L'aritmetizzazione dei parametri e dei sottoparametri e dei*

* Università di Padova, Facoltà di Scienze Politiche. Dipartimento di Scienze Economiche Marco Fanno, Padova.

[1] Al 1° Convegno Nazionale dell'Associazione Italiana per la Storia del Pensiero Economico. Venezia, Ca' Dolfin 16-18 dicembre 1993.

[2] DEMARIA (1993). Per avere il quadro sistematico di *A New Economic Logic*, che comprende la Parte V di cui diciamo, è bene ricordare – accanto ad altri numerosi lavori qui nemmeno sfiorati – almeno col titolo il contenuto delle Parti precedenti: Parte I. *Il problema logico in economia*; Parte II. *Sistemi logici tradizionali del sistema produttivo*; Parte III. *Sistemi produttivi generali*; Parte IV. *Esogeneità e dinamica*.

loro effetti economici (con una applicazione alla misura degli effetti economici dello sviluppo tecnologico); IV. *Sulla teorizzazione quantitativa delle correlazioni temporali in economia reale*; V. *Il tempo come serie ininterrotta di eventi economici e non economici nuovi*; VI. *La partenogenesi economica come costante dei sistemi economici e l'istinto delle combinazioni* – scacchiera su cui Demaria dispone la partita della ricerca scientifica che propone ai successori. Qui il nostro narrare è un « gioco » implicante compartecipazione onde – in termini di ragione critica – prescinde da tutto ciò che ottunde il gusto dell'invenzione e della sorpresa: totale libertà nel dare libero corso alla ventura che così cominciamo.

2. *La via logica principale discendente dal principio d'indeterminazione*

Sulla storia della logica delle teorie e dei sistemi economici hanno inciso, tradizionalmente, le impostazioni deterministiche del razionalismo convenzionale con l'assenza – in genere – di ogni epifenomeno extraeconomico, ma la presenza – frequente – di ideologie quali palinodie astratte e concrete. Consapevole di questo, Demaria ha sempre pensato all'esistenza di una nuova via logica: quella discendente dai principi opposti che hanno trovato la prima precisazione scritta nel saggio sulla « indeterminazione in economia » del 1932, contenente radicali condizionamenti della logica avanzata da Heisenberg. Di conseguenza, è nata la logica dei fatti economici attivi, soprattutto esogeni, illustrata dall'originalità insita negli « entelechiani », quali variabili extraeconomiche esplosive, non predicibili e a ritmi prima non ricorrenti, e ciò in accordo alla verità, mai smentita, che ogni giudizio economico importa sempre un giudizio logico. Onde, per Demaria, il grande compito della logica moderna dell'Economia è giudicare e teorizzare insieme – in solidarietà più che in interdipendenza – e con la prospettiva dell'originalità e dell'incertezza ineliminabili, i tessuti economici (endogeni) ed extraeconomici (esogeni) dell'economia concreta, questi ultimi rappresentati dai parametri (« termini noti » compresi). Tessuti contenuti tutti, a titolo di esempio, nel modo indicato dal « Sistema Generale Assoluto » (SGA) notando che alla misura dei parametri attendono, con le loro teorie, gli « engineering economists » odierni dando però misure (anche con precisione) mai definitive. Questi riferimenti costituiscono il punto centrale di *A New Economic Logic* (1993).

Demaria, fin da tempo lontano, ha dialogato polemicamente con l'intellettualismo e il criticismo classico dei suoi primi maestri le cui teorie economiche erano mosse dal principio della sovranità assoluta delle forze endoge-

ne, ostili quindi al pensiero non deterministico come fosse un pensiero non verità. Molto, però, gli giovarono gli insegnamenti – quando secondo tradizione scritta quando secondo tradizione orale – di studiosi « particolari » come Maffeo Pantaleoni (1909), Gustavo Del Vecchio (a Venezia, 1926), J.A. Schumpeter (trad. it. 1932).

Giusto all'inizio della Parte V. *Verso il sistema economico diacronicamente completo*, Demaria riconosce che il SGA della Parte III non è – pur con il completamento della teoria degli entelechiani e della teoria degli antientelechiani della Parte IV – un sistema completo, cioè atto a teorizzare tutta la realtà economica. Tuttavia, di alcuni problemi che ne sono nati – aggiunge – si possono delineare certi punti fondamentali: *a*) l'identificazione dei rapporti economici; *b*) le simmetrie e il punto fisso; *c*) la valutazione dei parametri; *d*) la variabilità delle correlazioni economiche nel tempo; *e*) il tempo economico come serie ininterrotta di eventi economici nuovi; *f*) la partenogenesi con il paretiano istinto delle combinazioni.

Questa l'architettura con l'avvertimento che

« sarebbe pericoloso riduzionista chi pretendesse che il mondo economico abbia sempre per base un'unica verità. Va perciò osservato che per il sistema economico completo valgono tanto gli indirizzi nati dalla meccanica classica di Newton e Laplace e dalla stessa meccanica quantistica, quando rigorosamente deterministiche a livello sia micro che macro, quanto la dinamica più moderna con le sue vedute in tema di entropia e di irreversibilità del tempo reale, queste ultime nettamente indeterministiche. Pertanto non può aversi senz'altro un solo insieme onniscente di leggi per i movimenti economici, ma si impone una continua, mai definitiva, esplorazione del nuovo ».

3. *L'identificazione dei rapporti economici*

L'identificazione – per garantire un'unica soluzione dei parametri – ricerca se esistono altre differenti combinazioni delle equazioni investendo tutte le equazioni e tutti i sistemi di equazioni che si propongono di determinare in concreto gli ordini causali dell'Economia di un dato spazio e momento. La non totale positività della lunga discussione avutasi sull'identificazione: H. Working (1925), E.J. Working (1927), Frisch (1934), Koopmans (1950), Klein (1953), Marschak (1953), Simon (1953), Fisher (1966), Theil (1973), Malinvaud (1978), fa osservare che il punto debole è nell'identificare i parametri per via statistica introducendo ipotesi « forti » – basate sulla normalizzazione dei « disturbi », supposti sempre noti o dedotti statisticamente dalla realtà passata – ipotesi relative agli stessi parametri (sempre in numero superiore alle equazioni) onde la proposta di Demaria di un

nuovo concetto di identificazione dei rapporti economici giacché occorre integrare o sostituire i metodi statistici finora prospettati con altri metodi.

In via generale può valere il principio: con un numero abbastanza elevato di equazioni è impossibile – da un punto di vista economico – concepire un sistema di equazioni sostitutive che si accordi con gli stessi parametri delle equazioni osservate. Il problema dell'identificazione è stato, purtroppo, eccessivamente tecnicizzato introducendo, come detto, ipotesi « forti » e perciò si è scansato il vero problema economico che vi sottende – taciuto anche nella letteratura più avanzata – il problema della esogeneità. L'identificazione dei parametri e dei sistemi – detto in breve – deve essere sempre accompagnata dalla ricognizione di tutta l'esogeneità.

Il SGA è identificato sia come sistema di equazioni sia come parametri, perché non vi sono introdotti valori probabilistici, perché è – nella sua interezza – endogeno ed esogeno, perché è in forma ridotta, cioè i parametri non sono semplici insiemi di multiregressioni astratte indipendenti l'uno dall'altro. Peraltro, pur se quanto più elevato è il numero dei parametri tanto minore è la probabilità della non identificazione (anche quindi quella di nuove teorie), non basta la numerosità quando intesa non dinamicamente, giacché il divenire non è mai chiuso.

Si impone allora come ulteriore possibile riferimento la complessità, ossia dovrebbe valere la relazione: Sistema (o Insieme) endogeno + Sistema (o Insieme) esogeno + Complessità = Sistema completo, anche se la complessità non è numerabile allo stato attuale della letteratura la quale – tuttavia – fornendoci teorie sulla qualità della vita, da quelle teorie consente di trarre i parametri più espressivi dei rapporti col mondo endogeno dell'Economia (quali gli indici geografici e settoriali di criminalità, scolarità, cultura e altro), in modo da ridurre, con la aumentata complessità, le possibilità della non identificazione.

Manca, dunque, una formulazione fondamentale semplice per la identificazione dei rapporti economici: « La ragione in definitiva è una sola: l'universo economico è complesso, ed è soggetto a continue trasformazioni singolari, che hanno, almeno in parte, leggi loro proprie ».

4. *Le simmetrie e il punto fisso*

L'impossibilità pratica di strumenti teorici quali sono i vari concetti di marginalità e le moderne teorie dei giochi e delle strategie degli n partecipanti e delle relative funzioni di payoff (dato che il numero n dovrebbe essere molto elevato) – nella valutazione e nella guida della Dinamica economica –

porta a servirsi di altri strumenti logici che, pur non perdendosi nel pensiero astratto, penetrino criticamente nella realtà globale e spazio-temporale, facendo capo al concetto di simmetria.

In Economia, il concetto di simmetria è – per Demaria – inscindibile da quattro attributi: realtà, diacronia, globalità, solidarietà; ed essa non significa necessariamente ottimo, armonia, perfezione totale, proprio perché può dipendere da spazio, tempo, cambiamento di scala e altro.

Le simmetrie di cui in von Neumann-Morgenstern (1944), Samuelson (1947), Hicks (1956), Leontief (1966), Watson (1967) oltre a riferirsi a un mondo immaginario, rendono fuorvianti le generalizzazioni derivate da esse giacché non sono verifica né della realtà né vera teoria, dato il rischio – già denunciato da Cournot – di percorrere solo spazi immaginari.

Le simmetrie economiche, aritmetizzabili e riferibili quasi esclusivamente al campo macro, non sono, prese singolarmente, esprimibili da rapporti numericamente esatti. Ogni simmetria economica presenta piuttosto proprietà « di gruppo », con ampiezze « di campo » le quali, storicamente e spazialmente, sono più o meno divergenti fra loro; ma non in modo probabilistico, giacché i loro gradi di libertà si svolgono secondo direzioni e dimensioni condizionate dalle grandi forze della storia. Costruzione e concetto di simmetria non sono perciò riferibili al concetto di « struttura »: questo concetto non è « categorico » in Economia e, comunque, vi sono strutture senza simmetria.

La prima di queste simmetrie è l'uguaglianza Risparmi = Investimenti, mai in vetta alle preoccupazioni dei politici. Altra simmetria di pari importanza è la relazione – che dovrebbe mantenersi generalmente costante – tra Moneta e Prodotto nazionale oppure tra Moneta e Reddito nazionale. E poi v'è la simmetria costituita dagli stretti rapporti di stabilità nel tempo e nello spazio di certe differenze tra i saggi di interesse nei mercati finanziari, bancari, monetari; la simmetria della bilancia dei pagamenti internazionali; la simmetria delle entrate e delle uscite dello stato, il cui nocciolo è il pareggio oppure l'avanzo complessivo; la distribuzione dei salari, che in un perfetto ordinamento economico dovrebbe presentarsi secondo l'espressione sia della ereditarietà dei patrimoni sia della produttività professionale, la « piramide delle età » giacché è dalla sua forma e dal suo grado di stabilità che assume particolare rilievo ogni importante problema economico.

Le simmetrie economiche possono essere seguite da « rotture », come si esprimono i fisici, per cui si deve parlare di asimmetrie e quasi simmetrie economiche. L'inflazione è l'antisimmetria per eccellenza.

A differenza delle simmetrie economiche, la causa fondamentale delle asimmetrie economiche sta quasi sempre nel condizionamento esogeno, o –

secondo la sistematica di Demaria – nel cattivo funzionamento dei « propagatori esogeni » delle azioni economiche e nei cosiddetti "entelechiani", ossia nella presenza prolungata di asimmetrie extraeconomiche che, una volta in vita, producono altre asimmetrie extraeconomiche e poi, di conseguenza, le asimmetrie economiche, onde, infine, anche il « saccheggio dell'attività produttiva altrui », come ha scritto Schumpeter. Secondo queste catene causali si formano prima certe asimmetrie, economiche ed extraeconomiche, e poi altre; la loro solidarietà sarebbe, quindi, fortemente diacronica: e ciò ricorda la 2ª legge della termodinamica, per cui l'entropia dell'universo è sempre crescente. Quindi, i sistemi disorganizzati tenderebbero sempre più, almeno per un certo tempo, a disorganizzarsi.

Nell'analisi e nella sintesi economica – quest'ultima intesa quale schema sillogistico macroeconomico realizzato attraverso l'aggregazione-generalizzazione di una moltitudine di postulati, relazioni, lemmi, problemi – hanno dominato fino al primo trentennio di questo secolo alcuni concetti-chiave: la legge di Walras; il punto di Cournot del monopolio; il punto di Bertrand-Hotelling del duopolio; la curva dei contratti di Edgeworth; il punto di ottimo di Pareto; il punto di fuga di Amoroso; il punto di sella della teoria dei giochi. Più complete spiegazioni di statica e di dinamica economica sono state però impostate da economisti logico-deduttivi i quali hanno legato le trattazioni economiche alla tipologia combinatoria. In particolare, il maggior trionfo di questa economia matematica è costituito nell'ultimo cinquantennio dalle applicazioni di due teoremi fondamentali del punto fisso – quello di Brouwer (1910) e quello di Kakutani (1941) – che, oltre alla prova dell'esistenza o meno di soluzioni, possono liberarci, in numerosi problemi, dalle programmazioni arbitrarie.

Quanto all'ultima tipologia economica derivata dai teoremi del punto fisso, bisogna distinguere due periodi: 1) l'anteguerra con l'applicazione – nel 1937 – di von Neumann allo sviluppo economico « bilanciato »; 2) il dopoguerra quando la tipologia economica è quasi unicamente versione topologica di comunissime situazioni economiche di equilibrio e di non equilibrio, sia dello scambio, sia del problema del consumatore e, conseguentemente, anche nei problemi dell'economia o sistema economico « puro ». Altre possibili deduzioni sono ottenute coi teoremi di Walras, Pareto, Slutsky, onde il sapere economicistico ne resta avvantaggiato, consentendo alle analisi che vi ricorrono una impostazione più realistica di quella, artificiosa, dei « tâtonnements ». Soprattutto quando sono accelerate dalle tecniche del simplesso e dei calcolatori, in quanto consentono sia soluzioni in termini quantitativi sia la calcolazione di vettori sempre più approssimati, cioè con punti fissi vicini alla « immagine » del teorema. Annotazione speciale voglio-

no poi gli sviluppi, di espansione economica, di von Neumann e le dimostrazioni (circoscritte) di S. Afriat, per cui, se un bene non raggiunge il massimo saggio di sviluppo del modello, tutti i beni devono crescere a un saggio inferiore.

Vale peraltro sempre ricordare l'avvertimento di O. Morgenstern, N. Georgescu-Roegen, T.C. Koopmans i quali non danno sempre causa vinta ai teoremi del punto fisso, in quanto le odierne applicazioni trascurano che parecchie fondamentali situazioni economiche non consentono affatto di convogliarvi i teoremi relativamente meno complicati del punto fisso, né le conseguenti conclusioni. Ad esempio, con tali teoremi non si possono in alcun modo affrontare i grandi problemi del capitalismo e del collettivismo; della redistribuzione del reddito nazionale tra ceti economici; dell'inflazione e della stabilizzazione monetaria, né quelli delle disuguaglianze tra nazioni ricche e nazioni sottosviluppate. In questi ultimi tempi, parecchie impostazioni troppo irreali dei teoremi del punto fisso sono però state rivedute. Qui Demaria accenna a una preliminare osservazione, forse costruttiva, per l'avvenire della topologia economica dei teoremi del punto fisso.

Essa riguarda gli elementi, gli insiemi e le trasformazioni dei teoremi del punto fisso quando sanno artificiosamente troppo di proprietà matematiche e troppo poco di proprietà economiche, per cui le trasformazioni e le corrispondenze ottenute applicando i teoremi di Brouwer e Kakutani adducono a risultati realisticamente sterili e, conseguentemente, nulla viene precisato sulle cause necessariamente « temporali » che portano al triplice ordine di vettori – vettore dei consumi, vettore dei beni prodotti, vettore dei prezzi. In ultima analisi, vi sono ignorate le variazioni parametriche, onde gli insiemi possono essere composti da certi elementi piuttosto che da altri, e possono quindi additare situazioni che non esistono nella realtà. Questa osservazione preliminare porta Demaria a individuare due deficienze teoriche di una certa importanza.

La prima deficienza teorica è quella ricorrente nelle applicazioni dei teoremi del punto fisso ai modelli di sviluppo economico « bilanciato », *in primis* di von Neumann (1937). Infatti, anche supponendo permanenti le preferenze utilitarie; predeterminando gli input iniziali, eguagliati a ogni intervallo ai vettori dell'output; uniformando nel tempo e in tutte le unità produttive il profitto; assegnando alle unità produttive un saggio permanente di intensità tecnologica ed economica (espresso dal vettore dei saggi di intensità), la determinazione delle incognite diventa solo un problema alla Brouwer. Ma nessuna economia può concepirsi con saggi di interessi e di intensità uguali tra loro, unici nel tempo e nello spazio, oppure predeterminatamente mutevoli entro l'estremo saggio di intensità tecnologica e l'estre-

mo del saggio di intensità economica, e inoltre sempre e ovunque massimi. Questi risultati – che implicano – di regola – anche un tasso uniforme di crescita della popolazione, dei salari, dei consumi, delle produzioni; una maggiorazione periodale fissa nel tempo delle domande e delle offerte e la piena occupazione – non possono servire di base ad alcuna teoria dinamica, sia perché sarebbero in netto contrasto con gli effettivi movimenti economici, sia perché parecchi insiemi potrebbero diventare improvvisamente vuoti (come nel caso, assai frequente, degli investimenti).

La seconda deficienza teorica delle attuali applicazioni dei teoremi del punto fisso all'Economia è in due indebiti enti logici con cui esse vengono associate, onde due legami in grandissima parte inesistenti in Economia. Il primo legame è con la ottimizzazione temporale; concetto che – rappresentato dalla teoria o strategia dei giochi – costituisce – a parte il caso delle singole aziende – il prodotto di una mentalità irrealistica, lontana dalle vicende del tempo storico, e quindi probabilmente deve esserci a ogni intervallo un punto fisso differente da quelli degli altri intervalli. Il secondo legame riguarda le correnti applicazioni economiche dei teoremi del punto fisso all'equilibrio stabile; concetto la cui validità poteva costituire un mero raffinamento didascalico quando inteso come posizione da cui, se per qualunque motivo l'attività economica micro o macro si stacca, a essa ritorna sollecitamente, sempreché le stesse forze che regolano tale attività ve la riportino in modo immediato e durevole. Peraltro, appena l'economia diventa progressiva o regressiva, il concetto (naturalistico) di equilibrio stabile perde significato teorico e pratico, giacché non si possono ignorare i fattori o fatti nuovi di tale economia, per cui essa si allontana, a ogni intervallo di tempo, dalla posizione occupata nell'intervallo precedente. Quindi, i teoremi del punto fisso sono spesso male impiegati.

Demaria – detto speditamente – propone che per l'economia dinamica si debba guardare piuttosto all'ordine – datità – delle priorità causali: cosa che dovrebbe in ogni caso essere evidenziata. Se però l'ordine del sistema è elevatissimo – come nell'economia concreta moderna – non esiste possibilità di valutare tutti i relativi parametri. In breve, vi è sempre almeno un parametro per cui non è valida alcuna delle soluzioni prospettate.

5. *La valutazione dei parametri*

L'aritmetizzazione dei parametri e dei sotto-parametri e degli effetti economici che ne discendono fornita dalla dottrina contemporanea è inaccettabile, basandosi su tre punti razionali debolissimi: 1) teorie economiche

estremamente artificiose dal lato sistematico; 2) calcolo non sistematizzato delle multiregressioni; 3) substrato generico e arbitrario per l'aritmetizzazione dei sottoparametri.

La posizione di Demaria è presto esposta. Nelle 31 equazioni del SGA, le funzioni di « industria » (ossia i singoli insiemi di imprese che producono gli stessi tipi o categorie o settori di beni con le stesse tecniche e che sono indispensabili dal punto di vista logico sistematico e fattuale); le funzioni di investimento; le funzioni di risparmio; le funzioni di deficit; le funzioni di circolazione consentono – pur essendo aggregate – di comprendere tutte le particolari situazioni effettive della realtà, giacché vi coesistono, seppure circoscritte, a livello micro, funzioni omogenee di 1° grado e altre; produttività marginali costanti e variabili; rapporti ed elasticità marginali di sostituzione di un fattore di produzione all'altro, sia costanti che variabili; remunerazioni determinate dalla produttività marginale; equilibri di "industria" di breve e di lungo momento; nonché tutti i possibili tipi di interazione.

Quindi un determinato propagatore (dei 10 studiati da Demaria), per esempio il propagatore tecnologico (sul quale Demaria conduce l'esemplificazione), non viene connesso soltanto alla limitata e ingannevole distinzione corrente del progresso tecnico che muta o che lascia intatto il rapporto marginale di sostituzione, ossia la distinzione del progresso tecnico neutrale e non neutrale, ma se ne possono seguire le varie influenze su tutte le variabili in discorso.

Peraltro, mentre i parametri totali stimati del SGA possono essere determinati in modo teoricamente ineccepibile, il calcolo dei sotto-parametri si presenta con presupposti e qualificazioni che devono concepirsi in modo diverso. Solo una parte – detta sotto-parametro di ogni parametro situato – esprime la grandezza causale di un certo propagatore sulla variabile endogena considerata contenuta in una determinata equazione del SGA. L'altra eventualità è più razionale giacché la separazione dei valori dei sottoparametri di uno stesso parametro situato difficilmente può partire dalla supposizione che gli altri propagatori restino stabili.

Pertanto, considerando un certo spazio-tempo, alla domanda quali sono le variabili incognite le cui variazioni dipendono unicamente dai sottoparametri situati esprimenti la dimensione economica di un certo propagatore (tecnologico nell'esempio fatto), Demaria risponde che coi 7 teoremi del SGA da lui proposti altrove (1966) si possono risolvere numerose questioni, compresa la valutazione dei principali effetti economici macro che derivano dai singoli sotto-parametri. Quanto al problema dell'unicità delle soluzioni già risolto dal SGA in generale (con i parametri totali) la metodologia proposta coi 7 teoremi fa ricordare, allo stesso Demaria, che esso può essere

risolto senza stortura perché – in termini algebrici lineari – il SGA dà luogo sempre, nell'economia reale, a una sola soluzione.

6. *Le variabili correlazioni economiche nel tempo*

Il problema della teorizzazione quantitativa delle correlazioni temporali in economia reale è stato impostato – dalla dottrina prevalente – in un modo dal quale Demaria dissente fondamentalmente nei riguardi sia della metodologia soprattutto matematica sia della logica economica che le concerne. Il dato di fatto considerato da Hicks (1950) e da lui acquisito in tutta la sua produzione scientifica (sul quale è esemplificata la critica di Demaria) è che vi siano regolarità cicliche di 7-10 anni (per quanto non strettamente uniformi). Questa affermazione contrasta con quanto le ricerche statistiche, anche le più recenti, hanno diversamente prospettato. Una costatazione cui del resto hanno portato anche le ricerche di Demaria (1936) che così conclude:

« Se le ricerche e le vicende economiche ulteriori non modificheranno radicalmente il materiale statistico oggi disponibile, la teoria dinamica non può abbracciare con una sola teoria due o piú cicli successivi e perciò i cicli debbono essere trattati come indipendenti sistematicamente l'uno dall'altro ».

Questo premesso, nel SGA compaiono, a ogni momento del ciclo, moltissimi parametri da determinarsi sperimentalmente, ciascuno nella sua interezza e poi ciascuno nelle dieci particelle (sotto-parametri) corrispondenti ai 10 propagatori. Data questa situazione non è possibile in alcun modo teorizzare in termini matematici precisi e concisi, e con una sistematica unica, le influenze della complessa esogeneità dei propagatori sull'economia considerate nell'insieme di più cicli o anche di un solo ciclo. Pertanto, per ogni correlazione o relazione funzionale lungo un certo periodo di tempo, purché della stessa categoria economica, vi possono essere tanti valori parametrici e tanti valori funzionali ed essi secondo la loro frequenza formano inevitabilmente certi campi di variabilità con limiti sia di tempo sia per grandezze. Per la stessa definizione di campo di variabilità sono anche possibili basse e contraddittorie misure di correlazione, in accordo col fatto che il sistema economico è infinitamente flessibile nel dettaglio pur mostrando nel complesso (ma a posteriori) nessi fondamentali relativamente stabili. Infine, non esiste un solo punto di svolta o di fase per tutta l'economia, ma tanti quante sono le principali serie dinamiche.

Due i quadri o motivi fondamentali caratteristici delle relazioni dinamiche in ogni schema statistico cosí considerato: 1) quello deterministico teleo-

logico di Evans, Roos, La Volpe; 2) quello indeterministico, basato sostanzialmente sulla ovvia logica degli originatori, distributori e risultanti, logica propria anche di Demaria per il quale qualsiasi causa di dinamismo costituisce un originatore di movimento, e corrispondentemente gli effetti degli originatori sono la risultante del movimento medesimo. Tra originatori e risultante o risultanti v'è quindi un insieme di condizioni – variabile nel tempo – dalle quali dipende l'intensità dei primi sui secondi, e si dicono distributori. Gli originatori, i distributori e le risultanti possono essere economici o extraeconomici; secolari o di più limitato tempo; in particolare, le risultanti sono istantanee o durature, additive o combinative, cumulative o non cumulative, e ciò a vario grado (1°, 2°, ecc.).

Visto come si combinano questi 3 elementi, la conclusione è che nell'assoluta generalità dei casi lo sviluppo di una funzione temporale relativa a una certa variabile economica procede inevitabilmente per onde (intese, al caso, anche in modo sottilissimo) e non per la via più breve.

« Comunque » – deduce Demaria – « la teoria economica più fondata – sia limitata a un solo momento sia dinamica – deve incorporare sempre il principio di indeterminazione. Essa deve anche essere contraria a una unica causalità generale, rigida e assoluta, e lo deve essere pure perché vi sono sempre variabili nascoste inaccessibili in quanto inosservabili, che la nostra capacità di conoscenza non riesce al momento a determinare e per cui si ha solo una conoscenza statistica ».

Con questi principali riferimenti si può concludere, dal lato quantitativo, che quanto è stato svolto di critica alla concezione di Hicks (citata, come detto, quale esempio) risulta in accordo con la interpretazione statistica complessiva dei cicli di ascesa e di discesa – basata sui detti campi temporali e dimensionali (rispetto alle variabili cosí considerate) – che si presentano in economia reale ([3]).

7. *Il tempo economico come serie ininterrotta di eventi economici nuovi*

Il tempo è il nerbo dell'economia dinamica e perciò sono strettissime le sue relazioni con le causali tutte endogene ed esogene dell'economia. Per il tempo economico vale un sistema di coordinate n-dimensionali dove n è elevatissimo, e anche variabile, corrispondente alle mutabili dimensioni degli

[3] Per questa interpretazione, se ne può vedere il primo svolgimento concreto in *Correlazioni economiche nel tempo* (1936), conferenza tenuta all'Istituto fisico matematico del Politecnico e Università di Milano, nonché – in riassunto – nella Parte V, sezione IV, citata al secondo capoverso della *Premessa*.

eventi economici. « Dunque anche l'economia, come la storia, è inesplicabile senza l'originalità degli eventi economici e il loro succedere irreversibile ». Oltracciò, solo a tratti il tempo economico può essere entropico. Così, quando il sistema economico, sotto la spinta di fatti entelechiani quale la degenerazione di certi propagatori, perde i livelli tradizionali o semplicemente presenti dei suoi rapporti interni ed esterni di equilibrio (statico o dinamico).

« Pertanto il tempo da considerarsi in Economia è, sì, referenziale avuto riguardo alla datazione dei suoi eventi, ma nei 'tempi lunghi' è indispensabile il riferimento ai fatti nuovi, soprattutto ai più decisivi, e quindi gli evi sono sempre diversi. La freccia del tempo cambia direzione, grandezza e lunghezza da evo a evo, da secolo a secolo, anche da decennio a decennio, come sta provando la stessa storia contemporanea ».

« Perciò tanto il tempo storico quanto quello economico non sono solo un modo della nostra sensibilità. Hanno un'esistenza primitiva a sé che diviene, per noi, in modo sovente misterioso e improvviso dato che la sua razionalità è vista o intravvista solo a posteriori ».

Concludendo, l'interpretazione del tempo economico è passata attraverso tre logiche generali diverse. Non vale più, esclusivamente, in Economia, l'antichissima logica aristotelica, sia pure con i complementi Port Royal, che fu propria di Smith, Ricardo e seguaci, soprattutto S. Mill, W.S. Jevons, A. Marshall (con i suoi « tempi brevi e lunghi » ma assolutamente continui) e W. Mitchell. Né è più totalmente accettata quella precedente con l'aggiunta dei simbolismi aventi ritmo rigorosamente fisso di scorrimento, già « in nuce » di Leibniz, poi di B. Bolzano e dopo ancora di G. Peano, B. Russell e A.N. Whitehead e, per l'Economia, di V. Pareto e G. Debreu, quando cronologici del tutto, e perciò simile a quella discendente dal concetto matematico di tempo.

« Per fronteggiare e acquisire in tutte le scienze la scoperta realtà dei quanti è ancora diventata diversa in alcune parti fondamentali la logica, in Economia moderna, del concetto di tempo originale, così aliena dallo spirito aristotelico e dagli enti fisici e più vicina, per il passato, all'indeterminismo ontologico di Parmenide. Logica interpretata modernamente, cioè con l'abbandono del principio del terzo escluso e la teorizzazione, alla grande, dei fatti nuovi ».

8. *La partenogenesi con il paretiano istinto delle combinazioni*

Si sono superate ormai definitivamente le fasi metodologiche dell'Economia di Pareto e di Debreu e quella degli « engineering economists », prevalentemente della pura endogeneità, ossia della esclusiva sovranità dell'evento, dell'azione, dell'effetto economico, sostenuta dall'economia pura

per cui dovevano entrarvi solo i beni e i loro rapporti, con esclusione dei beni « immateriali », posizione criticata da Del Vecchio addirittura dal 1908. Si riconosce altresì, al presente, che ogni periodo di tempo esaminato eredita certe asimmetrie delle variabili economiche, soprattutto endogene, onde l'impossibilità di una assoluta precisione nei riguardi più generali, oltre il continuo problema della osservabilità (Anschaulichkeit) per cui non vi sono parametri precisi e limiti funzionali unici.

« Per ora perciò », – questo l'ultimo approdo di Demaria – « nella nostra dottrina non vi è che una non generale e non definitiva armonia tra le espressioni degli eventi che sia integrabile e ordinata ..., per cui l'essere della scienza economica è per così dire un divenire di modelli diversi in una catena né mono-tona né ordinata diversamente e solo con limitati equilibri e molti gradi di libertà mai isolabili rigorosamente nelle loro ripercussioni. Anzi, talvolta, esse si ripercuotono l'una sull'altra a molta distanza nel lontano tempo-spazio (attraverso, anche, le previsioni), e talvolta, in queste ripercussioni, con una concentrazione immediata e univoca soltanto locale. Per cui bisogna fare solo della Stocastica ... Inoltre, le ultime tendenze dottrinali ... per spiegare ... queste deviazioni non sono molto ispiratrici. Così è per la euristica della teoria della complessità, di cui si parla molto in matematica, che porta in Economia a un 'Holism' teorico insufficiente e per via della sua scarsezza tassonomica e perché sostituirebbe, come negli scritti di von Hayek, ma in modo non dimostrato, a un forse possibile principio generale dell'equilibrio o della solidarietà (limitata), un 'ordine spontaneo' tra i vari processi naturali, economici, sociali, istituzionali, morali per cui le relative deviazioni (reciproche) si annullerebbero con il volgere del tempo e le entropie, incluse nella complessità del sistema, volgerebbero, a maggioranza, nel senso delle correzioni favorevoli e non di quelle sfavorevoli ».

Per Demaria, però, c'è invece un modo generale di comportamento dottrinale di fronte alle suaccennate carenze sistematiche: almeno considerare preminente la ricerca di nuove cause-effetto « costanti », nel numero delle « variabili nascoste », numero probabilmente più elevato di quanto ipotizzato finora in via astratta. Si pensi al « costante » ciclo vitale di Modigliani nel sistema macrodinamico del risparmio, ma sotto specie di tasso di accumulazione diverso a seconda dell'età (causa-effetto « costante » tuttavia moltissimo ancora da esplorare).

Una delle « variabili nascoste » col requisito della costanza nei sistemi economici è l'energia individuale – che Demaria dice « partenogenetica » – la quale nasce e dura finché dura l'individuo e influisce su tutte le sue decisioni economiche. Essa però non deve essere confusa con l'energia biotica, di cui è solo parte, perché va messa in relazione diretta solo con certi geni del genoma individuale, causa fondamentale (secondo le vedute correnti) di determinate capacità e attività fisiche, razionali e irrazionali, anche culturali e morali che toccano da vicino la realtà economica individuale e – guardando alla mappa dei geni in un gruppo di individui, di un'impresa, di un territorio, di una categoria professionale – quella di tali gruppi.

L'energia partenogenetica – la quale non ha altra origine che sé stessa, e può quindi considerarsi primitiva, cioè senza fertilizzazione, almeno dal punto di vista della sistematica economica che interessa – potrebbe essere anche un tema chiave della crescita economica micro e macro, quindi pure uno degli strumenti della metodologia futura volta a stabilire l'origine e il campo di equilibrio della economia individuale, gruppale, nazionale nel relativo complesso.

Le linee principali del percorso degli effetti sulla realtà economica della energia partenogenetica importano di considerare sempre che detto percorso non è subito evidente giacché non procede in linea retta e differisce nei vari gruppi sistematici che compongono la realtà. I differenti geni con cui si presentano gli operatori economici nelle produzioni, sui mercati, nella distribuzione portano a distinguere il suddetto percorso a seconda che si guardi: all'endogeneità; al funzionamento dei 10 propagatori dell'esogeneità permanente; all'esogeneità non permanente con gli entelechiani distinti per tipi; ai singoli settori economici e al loro complesso; ai differenti livelli professionali; e, in via complessiva, ai gruppi sociali distinti: per classi e ceti economici nei vari popoli e quindi per razze, corrispondentemente alla loro produttività, al loro grado di risparmio e di investimento, alle fasi dinamiche dello sviluppo economico, e all'istante e nei vari periodi di tempo.

« Dunque, l'energia partenogenetica è requisito indispensabile perché si formi una struttura produttiva ». A titolo di esempio, i lavoratori sono con maggiore o minore energia partenogenetica se l'imprenditore sa o non sa scoprire e introdurre nell'impresa adatti processi produttivi. Però, il progresso economico porta a una diversa concentrazione dell'energia partenogenetica in modo da trasformare, anche per questo motivo, da generazione a generazione, e quasi come effetto meccanico, la struttura generale delle industrie, oltreché usi, abitudini, tipi di civiltà che altrimenti si dispiegherebbero per gran parte nel modo tradizionale ([4]).

« In conclusione, l'energia partenogenetica si distingue per la qualità e il numero dei geni che sono immediatamente e logicamente prima degli effetti cui dà luogo, come quelli di *excitatio*, di *novitas*, di *operositas*, prodotti dalla situazione dei genomi ».

Scrive ancora Demaria:

« L'arcipelago genomico di una società è però tutto da esplorare, anche come media, e con il pericolo di incontrare gravi situazioni di indeterminazione (logica). Non solo. Ma ciascun

[4] Per i primi svolgimenti, cfr., sull'energia partenogenetica, i due saggi di Demaria del 1982 e i due saggi del 1983.

popolo di operatori viaggia nel tempo e nello spazio con il suo bagaglio di genomi tanto che lo si può dividere per regioni che contano più o meno economicamente, onde la inanità di certe politiche, economiche, sindacali, istituzionali, costituzionali, governative in genere, quando la media dei genomi non funziona nel verso giusto ».

Particolarmente interessante, da punti di vista ideali e fattuali, che la teoria dell'energia partenogenetica – basata sulla costanza di certi geni singoli, su certi genomi individuali, sul genoma medio nazionale – collimi (nota Demaria) per vari aspetti e nel complesso dei risultati con quanto sostenuto nella teoria dei « residui » avanzata da Pareto, con cui egli cercò di descrivere « con metodi logico sperimentali » fatti individuali e sociali, non soltanto economici, dovuti agli impulsi *costanti* di varie categorie di uomini. Fra questi residui vi sono appunto le azioni – discendenti per Demaria in larga misura dai geni – che Pareto attribuisce alle tendenze individuali ma compiute senza un ragionamento vero da parte degli uomini, e ciò sia combinando cose, azioni, rapporti, anche i più disparati, onde l'« istinto delle combinazioni », sia mantenendo e conservando cose, azioni, rapporti cui nessuno più crede, onde l'« istinto degli aggregati », per cui nel complesso può avvenire – guardando all'Economia – che il reddito, le ricchezze, il benessere si portano al di qua od oltre certi limiti spiegati da altre teorie. È ovvio notare che queste stesse conseguenze e la loro categoria causale individuale sono pure quelle prese in considerazione da Schumpeter con la teoria degli imprenditori i quali, a schiera, creano del nuovo sia nelle economie capitalistiche sia nelle economie collettivistiche.

Demaria conclude scrivendo:

« Ma l'istinto delle combinazioni e degli aggregati, come nel caso dei geni perversi, può portare talvolta o sovente a combinazioni economiche non solo errate, ma penalmente esiziali alla collettività. In particolare, ciò avviene quando sull'economia gravano ordinamenti e governi pubblici che, pur di durare, consentono o promuovono addirittura politiche illusorie, mistificatrici, antiproduttrici, disastrose nel complesso, o perfino dolose perché dovute allo spirito del male che calpesta leggi economiche fondamentali e crea disfunzioni economiche su larga scala. Dunque i popoli si distinguono anche a seconda che posseggano nei loro genomi (o nei loro 'residui') certi spiriti di intrapresa e certe virtù positive e negative. E ciò non solo come catalisi o acceleramento delle nuove situazioni in cui operano l'endogeneità e la esogeneità, ma per via di fattori individuali portati costantemente alla distruzione o al suo contrario ».

RIFERIMENTI BIBLIOGRAFICI

BROUWER L.E.J., « Über eineindeutige, stetige Transformationen von Flächen in sich », *Mathematische Annalen*, 1910, *69*, 176-80.

DEL VECCHIO G., *Ricchezze immateriali e capitali immateriali*, Alessandria: Società Poligrafica Alessandrina, 1908.

DEMARIA G., « Di un principio di indeterminazione in economia dinamica », *Rivista internazionale di scienze sociali*, n. 5, 1932, *15*, 597-636.

——, « Correlazioni economiche nel tempo », *Rendiconti del Seminario matematico e fisico di Milano*, Milano: Libreria Editrice Politecnica, vol. X, 1936.

——, *Trattato di logica economica*. I. *La catallattica*, Padova: Cedam, 1962.

——, *Trattato di logica economica*. II. *Il sistema produttivo*, Padova: Cedam, 1966.

——, *Trattato di logica economica*. III. *L'esogeneità*, Padova: Cedam, 1974.

——, « Energia partenogenetica e tempo economico », in AA.VV., *Saggi in onore di Bruno Menegoni*, a cura di G. Franco in « Rendiconti del Comitato per gli studi e la programmazione economica », Vol. XX, Padova, Editrice Alceo, 1982, 115-26.

——, « Le tre componenti dell'energia partenogenetica », *Atti dell'Accademia Nazionale dei Lincei*, Serie ottava. *Rendiconti di scienze morali, storiche e filologiche*, n. 7-12, luglio-dicembre, 1982, *37*, 191-209.

——, « Teoria economica dell'energia partenogenetica », *Rivista internazionale di scienze economiche e commerciali*, n. 1, 1983, *30*, 5-40.

——, « Le cause dell'energia partenogenetica e la sua misura », *Rivista internazionale di scienze economiche e commerciali*, n. 6, 1983, *30*, 496-519.

——, *A New Economic Logic*. Indetermination, propagators, entelechians. Edited from books and articles and arranged by A. Agnati, D. Cantarelli, A. Montesano with an Introduction. Dattiloscritto in corso di revisione per la pubblicazione a stampa, 1993.

FISHER F.M., *The Identification Problem in Econometrics*, New York: McGraw-Hill, 1966.

FRISCH R., *Statistical Confluence Analysis by Means of Complete Regression Systems*, Oslo: University Institute of Economics, 1934.

HICKS J., *A Contribution to the Theory of the Trade Cycle*, Oxford: Clarendon Press, 1950.

——, *A Revision of Demand Theory*, Oxford: Clarendon Press, 1956.

KAKUTANI S., « A Generalization of Brouwer's Fixed Point Theorem », *Duke Mathematical Journal*, n. 3, 1941, *8*, 457-59.

KLEIN E., *A Textbook of Econometrics*, Evanston, Ill.: Row, Peterson and Company, 1953.

KOOPMANS T.C., ed., *Statistical Inference in Dynamic Economic Models*, Cowles Commission Monograph, n. 10, New York: Wiley, 1950.

——, REIERSOL O., « The Identification of Structural Characteristics », *The Annals of Mathematical Statistics*, 1950.

LEONTIEF W., *Input-Output Economics*, Oxford: Oxford University Press, 1966.

MALINVAUD E.L., *Méthodes statistiques de l'économétrie*, 3ème éd., Paris: Dunod, 1978.

MARSCHAK J., « Economic Measurement for Policy and Prediction », in W.C. Hood and T.C. Koopmans, eds., *Studies in Econometric Method*, Cowles Commission Monograph, n. 14, New York: John Wiley, 1953.

VON NEUMANN J.V., « Über ein ökonomischen Gleichungssystem und eine Verallgemeinerung des Brouwerschen Fixpunktsatzes », in K. Menger, ed., *Ergebnisse eines Mathematischen Kolloquium*, 8, 1935-36, Leipzig und Wien: Franz Deuticke, 1937.

——, MORGENSTERN O., *The Theory of Games and Economic Behavior*, Princeton: Princeton University Press, 1944.

PANTALEONI M., « Di alcuni fenomeni di dinamica economica », Memoria presentata all'Associazione per il progresso delle Scienze; riunione di Padova, settembre 1909, *Giornale degli economisti*, 1909-II, 211-54.

SAMUELSON P.A., *Foundations of Economic Analysis*, Cambridge, Mass.: Harvard University Press, 1947.

SCHUMPETER J.A., *La teoria dello sviluppo economico*. Riduzione italiana autorizzata dall'Autore sulla seconda edizione tedesca per G. Demaria e K. Mayer in "Nuova collana di economisti stranieri e italiani", vol. V, *Dinamica economica*, Torino: Utet, 1932, 17-182.

SIMON H.A., « Casual Ordering and Identifiability », in W.C. Hood and T.C. Koopmans, eds., *Studies in Econometric Method*, New York: John Wiley, 1953 e London: Yale University Press, 1953, 49-74.

THEIL H., *Principles of Econometrics*, New York: John Wiley, 1973.

WATSON G.S., « Linear Least Squares Regression », *The Annals of Mathematical Statistics*, December, 1967.

WORKING H., « The Statistical Determination of Demand Curves », *The Quarterly Journal of Economics*, August, 1925.

——, « What Do 'Statistical Demand Curves' Show? », *The Quarterly Journal of Economics*, February, 1927.

TOWARDS A DIACHRONICALLY COMPLETE ECONOMIC SYSTEM: EXPOSITION OF PART V OF G. DEMARIA'S, *A NEW ECONOMIC LOGIC*, 1993.

In Demaria's opinion, the great task of modern logic of Economics lies in assessing and theorising together – on the perspectives of unavoidable originality and uncertainty – all aspects of concrete economics – great and small, endogenous and exogenous – represented by parameters. These fundamentals are the core of the typescript *A New Economic Logic*. In Part V: *Towards a diachronically complete economic system*, Demaria states that his "absolute general system" is incomplete: it does not enable us to form a theory about economic reality as a

whole. Research into the remaining gaps began only in this century, and so far has barely scratched the tip of a huge iceberg. Nevertheless, several basic points that have emerged may be broadly outlined. In the present paper, these concern: *a*) the identification of economic relations; *b*) symmetries and the fixed point; *c*) evaluation of parameters; *d*) the variable economic correlations in time; *e*) economic time as an unbroken series of new economic events; *f*) parthenogenesis, with Pareto's theory of combinations.

RECENSIONI

HOUMANIDIS Lazaros: *Le salaire ouvrier depuis les classiques jusqu'à nos jours* – 1994, Sichroni Ekdotiki LTD, Atene, p. 183, s.i.p.

Quale soluzione alla problematica salariale, considerato il ruolo di estrema importanza di tale elemento nell'equilibrio socio-economico?

Il Prof. Houmanidis, per giungere a una risposta, ripercorre il pensiero economico sul tema elaborato nel passato. L'opera si segnala, quindi, per l'estrema attualità della questione in essa trattata e per la proposta ivi contenuta: l'esame delle soluzioni che storicamente si sono rivelate fallimentari consente al lettore di evitare qualunque banalizzazione sul tema.

La brillantezza espositiva dell'autore risiede nella grande capacità di sintesi delle teorie economiche, delle quali viene accentuato ogni contributo alla soluzione del problema posto, senza, però, che la stessa sintesi vada a discapito della completezza dell'analisi.

Fin dall'inizio dell'opera sono evidenziati e sottolineati gli sviluppi e le relazioni tra le varie teorie, cosicché il lettore ha sempre una visione d'insieme dell'argomento in oggetto.

Dopo una breve premessa, dedicata al pensiero greco, romano e medievale, l'excursus storico-analitico prende avvio dalla teoria salariale elaborata da Smith, Ricardo, Malthus, e dagli altri principali autori classici, esposta brevemente, dai quali emerge un quadro chiaro e completo della questione. I Classici hanno, infatti, creduto in una società nella quale ogni individuo segue i propri desideri: tali desideri, egoistici e individuali, convogliano nella legge dell'interesse personale, che è libero in relazione all'azione conomica. Essi, però, non hanno compreso l'importanza dello Stato nel controllo delle attività economiche.

L'esposizione si snoda, poi, senza alcuna soluzione di continuità, attraverso la critica al contributo apportato dai Classici da parte dei c.d. pensatori socialisti (particolarmente attenti alla questione salariale e per i quali l'intervento dello Stato è indispensabile per la formazione di una classe economica razionale) fino a trattare delle concezioni in materia, esposte dagli Edonisti divisi in due Scuole: la Scuola di Vienna e quella di Losanna. Questi ultimi fondano la teoria del salario operaio su di un sistema economico statico ideale, limitato a due soggetti: l'operaio e il suo datore di lavoro. Inoltre, essi osservano che la causa determinante il salario operaio non è unica, perché gli elementi dell'economia si influenzano l'un l'altro.

Più complessa è, invece, la concezione salariale degli autori neoclassici: essi esaminano l'individuo e il suo comportamento, per poter inserire il concetto di produttività marginale nel quadro ideale della concorrenza perfetta, cercando di prevedere le conseguenze dell'evoluzione tecnica, dell'organizzazione del mercato e del progresso delle Istituzioni.

Uno sviluppo del pensiero precedente, in chiave eclettica, è stato proposto da Cassel, Perroux e Papi.

Tutte le esposte teorie hanno, però, considerato il comportamento di ogni singolo individuo, dimenticando di esaminare la psicologia delle masse e le cause che la informano. In tale ordine di idee si pongono gli Istituzionalisti, per i quali il salario operaio, essendo un sentire

sociale, è al tempo stesso un insieme di diversi fattori, che risultano dall'evoluzione delle relazioni umane.

L'avvento della prima guerra mondiale e la possibilità, per gli economisti, di utilizzare l'elaborazione statistica per cercare, tra le altre cose, le cause che provocano la disoccupazione, rende necessaria una revisione delle teorie elaborate e comporta il sorgere di nuovi contributi dottrinali.

L'approccio alla scienza economica muta, così, radicalmente: i fatti esaminati non costituiscono una semplice verifica della teoria, ma la materia prima, dall'analisi della quale risulterà tale teoria. In questo filone scientifico si collocano Pigou e Keynes.

L'analisi statistica, però, non può prescindere da quella teorica: in tal senso si sono espressi recentissimi autori, tra cui si annoverano il prof. Jean Marchal e i componenti dell'Institut de la Science Economique Appiquée presieduto dal Prof. François Perroux, secondo i quali il salario operaio non è la ricompensa del lavoro, ma il reddito di una parte della comunità nazionale.

Alla luce di tutte le teorie esaminate nell'opera, il Prof. Houmanidis sostiene che il costo del lavoro è un costo sociale, perché risulta dalle concezioni sociali e riposa sulla considerazione sociale. Infatti, l'operaio non compara solo il rapporto tra il salario e il riposo, ma tiene conto altresì della sua sottomissione al datore di lavoro (costo di dipendenza): se il salario aumenta, il costo di dipendenza diminuisce; viceversa, se l'imprenditore abbassa i salari o aumenta le ore di lavoro, il costo di dipendenza aumenta; in tale modello economico, è necessario un intervento dello Stato perché la libertà, tanto anelata dagli operai, potrebbe condurre all'anarchia.

Su queste considerazioni si fonda la « Scuola Umanista », secondo la quale il sistema di democrazia sociale può essere rappresentato come un cilindro, la cui sommità è occupata dagli intellettuali e gli strati inferiori dagli imprenditori che lavorano socialmente, dagli agricoltori riuniti in cooperative, dai professionisti e artigiani riuniti in associazioni, e infine dagli operai riuniti in sindacati (la presenza dei sindacati impedisce il ribasso del salario sotto il livello minimo disponibile).

La scienza economica – conclude il Prof. Houmanidis – è una realtà personale, ma, poiché è contenuta nel quadro della realtà sociale, si compone di conoscenze astratte e dell'evoluzione economica reale, nella quale l'uomo, valore supremo della vita, deve portare il suo contributo, per vincere le resistenze naturali e umane.

EMANUELE MIO