Economics Discussion Papers
The Open-Access, Open-Assessment E-Journal

Discussion Paper 2008-16
April 11, 2008

How Can Voters Classify an Incumbent under Output Persistence

António Caleiro

Universidade de Évora

Abstract:

The literature on electoral cycles has developed in two distinct phases. The first one considered the existence of non-rational (naive) voters whereas the second one considered fully rational voters. In our perspective, an intermediate approach is more interesting, i.e. one that considers learning voters, which are boundedly rational. In this sense, neural networks may be considered as learning mechanisms used by voters to perform a classification of the incumbent in order to distinguish opportunistic (electorally motivated) from benevolent (non-electorally motivated) behaviour. The paper shows in which circumstances a neural network, namely a perceptron, can resolve that problem of classification. This is done by considering a model allowing for output persistence, which is a feature of aggregate supply that, indeed, may make it impossible to correctly classify the incumbent.

JEL: C45, D72, E32

Keywords: Classification; elections; incumbent; neural networks; output; persistence; perceptrons

Correspondence:
António Caleiro, Departamento de Economia, Universidade de Évora, Largo dos Colegiais 2, 7000-803 Évora, Portugal, email: caleiro@uevora.pt

The author would like to thank Steffen Hornig for an extremely useful suggestion. Obviously, the usual disclaimer applies.

1. Introduction and motivation

An electoral cycle created by incumbents is a phenomenon that seems to characterise, at least in some particular occasions and/or circumstances, the democratic economies. It is generally accepted that the short-run electorally-induced fluctuations prejudice the long-run welfare. Since the very first studies on the matter, some authors offered suggestions as to what should be done against this electorally-induced instability. For some authors, ever since the seminal paper of Nordhaus (1975), a good alternative to the obvious proposal of increasing the electoral period length is to consider that voters abandon a passive and naive behaviour and, instead, are willing to learn about incumbent's intentions.

The electoral cycle literature has developed in two clearly distinct phases. The first one, which took place in the mid-1970s, considered the existence of non-rational (naive) voters. In accordance with the rational expectations revolution, in the late 1980s the second phase of models considered fully rational voters. It is our belief that an intermediate approach is more interesting, i.e. one that considers learning voters, which are boundedly rational.

Generally speaking, learning models have been developed as a reasonable alternative to the demanding informational assumption of rational expectations models.[1] Although a number of different studies modelling learning have been presented, two main classes of models can be distinguished: rational (or Bayesian) learning and boundedly rational learning models.[2] In rational learning models, it is

[1] Moreover, through learning models it is possible to study the dynamics of adjustment between equilibria which, in most rational expectations models, is ignored. Learning models also deal with another difficulty of rational expectations models, namely the existence of multiple equilibria. The analysis of learning processes can, in fact, provide a way of selecting the 'reasonable' equilibrium or sub-set of equilibria. If the learning mechanism is chosen optimally, then a desirable rational equilibrium is selected from the set of the rational expectations equilibria (see Marcet and Sargent 1988, 1989a, 1989b). If the learning mechanism is viewed under an adaptive approach, in particular in expectational stability models, it can also act as a selection criterion in multiple equilibria models (see Evans 1986, Evans and Guesnerie 1993, and Evans and Honkapohja 1994, 1995).

[2] Westaway (1992) prefers to distinguish closed-loop learning, where agents learn about the parameters

assumed that, while the learning process is taking place, agents know the true *structural form* of the model generating the economy, but not some of the parameters of that model. In boundedly rational learning models, it is assumed that agents, during the learning process, use a 'reasonable' rule, for instance, by considering the *reduced form* of the model.

In the bounded rationality approach, various notions of expectational stability and of econometric learning procedures have been the main formulations, in accordance with the notion of time where learning takes place. While the econometric learning procedures assume real-time learning, the expectational stability principle assumes that learning takes place in notional, virtual or meta-time. Moreover, in adaptive real-time learning, agents are assumed to use an econometric procedure for estimating the perceived law of motion. The expectational stability approach considers the distinction between perceived laws and actual laws of motion of the economic system, in the sense that the actual law of motion results from the substitution of the perceived law of motion in the structural equations of the true model. We propose to use another (innovative) bounded rationality approach, namely *neural networks* as devices of *procedural learning* within a political business cycles context. Salmon (1995) is, to the best of our knowledge, one of the very few references where such learning mechanism has been applied in a policy-making problem.

In doing so, we intend to help giving an answer to a question put some time ago by Westaway (1992), i.e "*How do policymakers react to the fact that the private sector is learning*?" but, to the best of our knowledge, for a long time almost ignored by the literature. Some exceptions are Barrell et al. (1992), Başar and Salmon (1990a,1990b), Cripps (1991), Evans and Honkapohja (1994, 2003), Evans and McGough (2005), Fuhrer and Hooker (1993), Honkapohja and Mitra (2004), Marimon and Sunder (1993, 1994), Salmon (1995) and Westaway (1992). Still, the analysis of the

---

of the decision rule, from open-loop learning, where agents form an expectation of the path for a particular variable which they sequentially update.

implications of learning mechanisms in policy-making is far from being complete.

That being said, here will consider neural networks as learning mechanisms used by voters to perform a classification of the incumbent in order to distinguish opportunistic (i.e. electorally-motivated) from benevolent (i.e. non-electorally motivated) behaviour.[3] In doing so, it will be shown in which circumstances a neural network, namely a perceptron, can resolve that classification problem. To achieve this objective we will consider a quite recent version of a stylised model of economic policy, i.e. a version based on an aggregate supply curve embodying *output persistence* (see Gärtner 1996, 1997, 1999, 2000). As a matter of fact, when output persists over the mandate, it may be impossible to distinguish a benevolent incumbent from an opportunistic one.

The model under consideration fits the debate, in which the possibility of monetary policy shocks affecting aggregate output is central. Indeed the persistence of shocks to aggregate output has been, still is (and most probably it will be for some time) one of the issues predominantly subject to investigation. For instance, quite recently, it was registered an increase of interest in analyzing the persistence of output, as well as of inflation, considering its relationship with other aspects such as the degree of openness of the economies, the exchange-rate regime or the structural change on the behaviour of consumers, firms or policy-makers.

For the empirical evidence that monetary policy shocks can have permanent effects on aggregate output (or unemployment) there has been proposed some theoretical explanations, notably imperfect information about nominal fluctuations, namely about prices, and short-run nominal rigidities, such as sticky prices. For instance, considering nominal price stickiness and imperfect information, Kiley (2000) has shown that both factors allow nominal shocks to propagate in the cycle, but that only sticky prices propagate the real effects of nominal shocks. However, Wang and Wen (2006) argue that whether or not price rigidity is responsible for output

[3] The neural network methodology is to be explained below.

persistence is not a theoretical question, but an empirical one.

After the seminal work of Taylor (1980), which has shown that staggered wage-setting can lead to persistence in employment after a temporary shock, there has been proposed also staggered mechanisms to help solve the, so-called, persistence puzzle. For some time, both staggered wage-setting and staggered price-setting were considered as being similar in the process of generation of persistent real effects of monetary shocks. For instance, Rotemberg and Woodford (1997) argue that output persistence can be due to price staggering. However, some other authors, namely Huang and Liu (2002) and Christiano et al. (2005), have argued that staggered wage mechanisms are much more effective than staggered price mechanisms in generating persistence. In Ascari (2003), however, it is argued that the ability of a model to produce output persistence is not due to price or wage staggering mechanisms *per se* but, in fact, is due to the factor specificity of the model, namely the behaviour assumed by firms and by the labour force. Finally, Merkl and Snower (2007) have shown that both staggered mechanisms are complementary in generating persistent output effects in response to monetary policy shocks.

Having said that, it is important to mention, at this stage of our analysis, that the previous studies confirm the persistence of output (or unemployment) being an up-to-date relevant issue. Despite the existence of some lively debate about the causes, it is apparent the existence of output persistence. Given the existence of this fact, it thus makes sense to study also the consequences of it. In this sense, in terms of the formalization that we will use it is important to mention also Jonsson (1997), Lockwood (1997) and Svensson (1997), who analyse the consequences of output or unemployment persistence on the establishment of inflation contracts. Here it will be used the same kind of model, in our case without uncertainty, to study the consequences of output persistence on the possibility that bounded rationality voters are able to classify the observed behaviour of the incumbent as being opportunistic or benevolent. Gärtner (1996, 1997, 1999, 2000) use the same formalization to study the consequences of output persistence on the pattern of the political business cycle.

The rest of the paper is structured as follows. As a methodological tool, section 2 offers the analysis of the characteristics of the particular neural network, i.e. the perceptron that will be used to perform the task of classifying the incumbent. Section 3 explores the problem of how to classify an incumbent showing in which, if so, circumstances the perceptron can resolve that problem. Section 4 concludes.

2. The learning task of the neural network

Besides solving the task of approximating some continuous function, as in the case of a *signal extraction*, neural networks are used mainly to learn in a classification task (see Swingler 1996).[4] In this case, the input is a description of an object to be recognised and the output is an *identification* of the class to which the object belongs. The most common kind of neural network for classification purposes is the so-called *perceptron*.[5] In what follows we will consider that *bounded rationality voters* have to classify economic policies and outcomes as coming from opportunistic or from benevolent behaviour of the incumbent. So, it will be shown how perceptrons, as approximations of bounded rationality agents, would classify policies and outcomes as 'electoralist' or not, using a recent stylised model of economic policy.

In our case, a single-layer network known as *perceptron* will be used to perform the classification task or, in other words, will be used to determine the vector of weights and bias specifying a line on the space (output-inflation) such that two sub-sets of points – the opportunistic and benevolent ones – are defined. At this stage, a short explanation about how the neural network will determine the above-mentioned vector seems appropriate.

In the particular case under study, the *learning process* conducing to the above-

[4] A simple and general discussion of the neural networks methodology is given in the Annex 1.

[5] For a clear explanation of the link between perceptrons and the statistical discriminant analysis see Cho

mentioned vector of weights and bias can thus be described as follows:

1. Initial weights, w, and bias, b, are generated in an interval with enough range;[6]
2. Given some target vector y^*, with binary values associated with the two considered categories of incumbents, the error, e, is computed as the difference between y^* and the perceptron output y.
   i) If there is no error in the classification, that is $e = 0$, then $\Delta w = \Delta b = 0$;
   ii) If some pair of economic policies/states is classified as belonging to category 1, say benevolent, and should have been classified as belonging to category 0, say opportunistic, then $e = -1$. Therefore, in order to increase the chance that the input vector x will be classified correctly, the weight vector w is 'put farther away' from x by subtracting x from it; this meaning that $\Delta w = -x^{\mathrm{T}}$;
   iii) If some pair of economic policies/states is classified as belonging to category 0 and should have been classified as belonging to category 1 , then $e = 1$. Therefore, in order to increase the chance that the input vector x will be classified correctly, the weight vector w is 'put closer' to x by adding x to it; this meaning that $\Delta w = x^{\mathrm{T}}$.

To sum up, the perceptron learning rule will be based upon the following updating rules:

$$\Delta w = (y^* - y)x^{\mathrm{T}} = ex^{\mathrm{T}}, \tag{1}$$

and

$$\Delta b = (y^* - y)\mathbf{1}^{\mathrm{T}} = e. \tag{2}$$

Using (1) and (2) repeatedly – the so-called *training process* – the perceptron will

and Sargent (1996).

[6] Note that a hard limit transfer function will be used and this gives $y = 1$ when $wx + b > 0$ and $y = 1$ when $wx + b \leq 0$.

eventually find a vector of weights and bias, such that all the pairs of inflation and output are classified correctly. Indeed, it is well known that, if those pairs are linearly separable, the perceptron will always be able to perform the classification by determining a linear decision boundary.

3. The classification of the incumbent

In the electoral business cycle literature, one of the most crucial conclusions is that the short-run electorally-induced fluctuations prejudice the long-run welfare. In fact, because the electoral results depend on voters' evaluation, we can consider that if electoral business cycles do exist it is because voters, through ignorance or for some other reason, allow them to exist. This point introduces a well-known problem of *electorally-induced behaviour punishment* and its related *problem of monitoring*. In reality, voters often cannot truly judge (or classify) if an observed state or policy is the result of a *self-interested/opportunistic incumbent* or, on the contrary, results as a *social-planner/benevolent* outcome, simply because voters do not know the structure, the model or the transmission mechanism connecting policy values to state values. Moreover, a constant monitoring of incumbent behaviour seems *not* to be considered a crucial practice by the electorate.

Even so, voters do 'anticipate' the possible economic damage resulting from such *myopic* behaviour by incumbents and, especially closer to the elections, start to *classify* policies and outcomes as potentially being the result of an 'electoralist' strategy. This is done in order not to be 'fooled' by the incumbent incumbent or simply to punish the incumbent incumbent in case of clear *signals* of electorally-induced policies. In other words, a classification is made, so that for a sufficiently small sub-set of policies classified as 'electoralist', voters usually do not take that as a serious motive for punishment, but others, regarded as serious deviations, are punished.[7] In general, this

[7] Note the difference between this approach and the one considered, for instance, in Minford (1995).

classification task is made difficult by ignorance of the structural form of the model transforming policies in outcomes and also simply because information gathering costs money and time.

3.1. The model

Recently some authors have assumed an extended version of the standard aggregate supply curve $y_t = \bar{y} + \beta(\pi_t - \pi_t^e)$, where y_t denotes the level of output (measured in logarithms) that deviates from the natural level, $\bar{y}$, whenever the inflation rate, π_t, deviates from its expected level π_t^e, by considering

$$y_t = (1-\eta)\bar{y} + \eta y_{t-1} + \delta(\pi_t - \pi_t^e), \tag{3}$$

where η measures the degree of output persistence.[8] See Gärtner (1999) for an output persistence case and/or Jonsson (1997) for an unemployment persistence case.[9]

When normalizing the natural level of output such that $\bar{y} = 0$ the aggregate supply curve reduces to:

$$y_t = \phi y_{t-1} + \alpha(\pi_t - \pi_t^e), \tag{4}$$

where, following the hypothesis of adaptive expectations,

$$\pi_t^e = \gamma\pi_{t-1}, \tag{5}$$

where $0 \le \phi \le 1$ and $0 \le \gamma \le 1$.

As said before, a most common kind of neural network for classification purposes is the so-called perceptron. In order to perform the task of classifying the

---

Here, it is assumed that "*voters penalise absolutely any evidence that monetary policy has responded to anything other than news*", by 'absolutely' meaning that there is enough withdrawal of voters to ensure electoral defeat.

[8] This way of introducing persistence, which results in expression (3), is the most common in the literature (see, for instance, Gärtner 1996, Jonsson 1997, Lockwood 1997, or Svensson 1997). Also note that expression (3) could also have been determined following the 'original' Lucas supply curve (see Lucas 1973: 328).

[9] As acknowledged in Gärtner (1999), only since a few years ago authors have started to pay due attention to the consequences of considering that relevant macroeconomic variables, *in reality*, show some degree of persistence over time.

incumbent, in what concerns its behaviour during the mandate, it is required the determination of the opportunistic and benevolent solutions. These solutions differ in accordance with the way time periods are discounted: whereas for society, therefore also for a benevolent incumbent, future periods should be less important than present ones, this is not the case with an opportunistic incumbent, as future moments, i.e. those closer to the election day, are more vital than present ones, in order to explore the decay in the memory of voters.

Having said that, concerning the incumbent's objective function, we make the standard assumption that the incumbent faces a mandate divided into two periods, t $=1,2$, such that society's welfare during the mandate, *i.e.* the benevolent incumbent's objective function is given by:

$$U = U_1 + \rho U_2, \tag{6}$$

where ρ is the social rate of discount, whereas opportunistic incumbent's objective function is :

$$V = \mu V_1 + V_2, \tag{7}$$

where μ is the degree of memory of the electorate. In (6) and (7) we also admit that

$$U_t = V_t = -\tfrac{1}{2}\pi_t^2 + \beta y_t. \tag{8}$$

In these circumstances it is worth immediately noticing that, in general, excepting if $\mu\rho = 1$, the policies that maximise social welfare (6) are not the ones that maximise popularity (7). As it plausible to assume that both ρ and μ do not exceed 1, it is immediately clear that only in the case of perfect memory, *i.e.* $\mu = 1$, and both periods being equally important for society, *i.e.* $\rho = 1$, an opportunistic incumbent will behave exactly as a benevolent one. This fact allows for making it plausible to ask the question: *how to classify a incumbent?*, whose answer is supposed to be given by a neural network when separating optimal outcomes into two parts: the opportunistic and the benevolent ones. In other words, the opportunistic and benevolent solutions (policies and outcomes) will constitute the necessary inputs for the neural network application.

Given the classification task format, let us precisely define what will be called *opportunistic* or 'electoralist' *inputs*, that is policies, and *opportunistic outputs*, that is outcomes, to be compared with *benevolent inputs* and *benevolent outputs*.

Clearly, the opportunistic policy and outcomes will be, respectively, the values of inflation and output which result from the maximisation of (6) and (7) subject to (4) and (5). This immediately leads to the optimal policies:[10]

$$\pi_1^B = \alpha\beta(1-\rho(\gamma-\phi)), \tag{9}$$

$$\pi_2^B = \alpha\beta, \tag{10}$$

$$\pi_1^O = \alpha\beta\left(1-\frac{\gamma-\phi}{\mu}\right), \tag{11}$$

$$\pi_2^O = \alpha\beta. \tag{12}$$

Those policies lead to the optimal output levels:

$$y_1^B = \phi y_0 + \alpha\left(\alpha\beta(1-\rho(\gamma-\phi)) - \gamma\pi_0\right), \tag{13}$$

$$y_2^B = \phi\left(\phi y_0 + \alpha\left(\alpha\beta(1-\rho(\gamma-\phi)) - \gamma\pi_0\right)\right) + \alpha(\alpha\beta - \gamma\alpha\beta(1-\rho(\gamma-\phi))), \tag{14}$$

$$y_1^O = \phi y_0 + \alpha\left(\alpha\beta\left(1-\frac{\gamma-\phi}{\mu}\right) - \gamma\pi_0\right), \tag{15}$$

$$y_2^O = \phi\left(\phi y_0 + \alpha\left(\alpha\beta\left(1-\frac{\gamma-\phi}{\mu}\right) - \gamma\pi_0\right)\right) + \alpha\left(\alpha\beta - \gamma\alpha\beta\left(1-\frac{\gamma-\phi}{\mu}\right)\right). \tag{16}$$

Before proceeding with the classification task, it is relevant to note that there are, in fact, two possible patterns for the political business cycle: *i*) a typical one, where inflationary expansions take place immediately before the elections and *ii*) an atypical one, where the inflationary expansions take place immediately after the elections.[11] Given that:

[10] From this point onwards, the superscripts *B* and *O* identify an element as, respectively, concerning the benevolent and the opportunistic incumbent.

[11] This means that, in general, not possible to always use the observed pre-elections expansions as empirical evidence supporting the existence of an opportunistic behaviour of the incumbent as, in fact, even some experienced scholars incorrectly do.

$$\pi_2^B - \pi_1^B = \alpha\beta\rho(\gamma-\phi),\ \pi_2^O - \pi_1^O = \alpha\beta\frac{\gamma-\phi}{\mu},$$

the typical pattern will be observed when $\gamma > \phi$ and the atypical one when $\gamma < \phi$. Plainly, when $\gamma = \phi$ there will be no cycle at all. See Gärtner (1996, 1997, 1999, 2000).

Given the optimal solutions, (9) to (16), it is straightforward to verify that, because

$$\pi_1^B - \pi_1^O = \alpha\beta(\gamma-\phi)\frac{1-\mu\rho}{\mu},\ \pi_2^B - \pi_2^O = 0,$$

$$y_1^B - y_1^O = \alpha^2\beta(\gamma-\phi)\frac{1-\mu\rho}{\mu},\ y_1^B - y_1^O = -\alpha^2\beta(\gamma-\phi)^2\frac{1-\mu\rho}{\mu},$$

the typical pattern will then be characterised by:

$$\pi_2^B > \pi_1^B,\ \pi_2^O > \pi_1^O,\ \pi_1^B > \pi_1^O,\ \pi_2^B = \pi_2^O$$

and

$$y_1^B > y_1^O,\ y_2^B < y_2^O,$$

whereas the atypical pattern will be characterised by:

$$\pi_2^B < \pi_1^B,\ \pi_2^O < \pi_1^O,\ \pi_1^B < \pi_1^O,\ \pi_2^B = \pi_2^O$$

and

$$y_1^B < y_1^O,\ y_2^B < y_2^O.$$

Given that, in the previous mandate, no matter the kind of incumbent,

$$\pi_0 = \alpha\beta, \tag{17}$$

it is possible to further simplify the optimal output levels expressions, (13) to (16), to:

$$y_1^B = \phi y_0 + \alpha^2\beta(1-\rho(\gamma-\phi)-\gamma), \tag{18}$$

$$y_2^B = \phi^2 y_0 + \alpha^2\beta(\phi - 2\phi\rho\gamma + \rho\phi^2 - \phi\gamma + 1 - \gamma + \gamma^2\rho), \tag{19}$$

$$y_1^O = \phi y_0 + \alpha^2\beta\frac{\mu+\phi-\gamma-\gamma\mu}{\mu}, \tag{20}$$

$$y_2^O = \phi\left(\phi y_0 + \alpha^2\beta\frac{\mu+\phi-\gamma-\gamma\mu}{\mu}\right) + \alpha\left(\alpha\beta - \gamma\alpha\beta\left(1-\frac{\gamma-\phi}{\mu}\right)\right). \tag{21}$$

3.2. The classification task

The optimal inflation rates, (9) to (12), and output levels, (18) to (21), define the coordinates of four points in the (y, π) space. This space is to be partioned, *if possible*, in two sub-spaces by a linear decision boundary – in that consists the classification task – by the neural network. See figure 1.

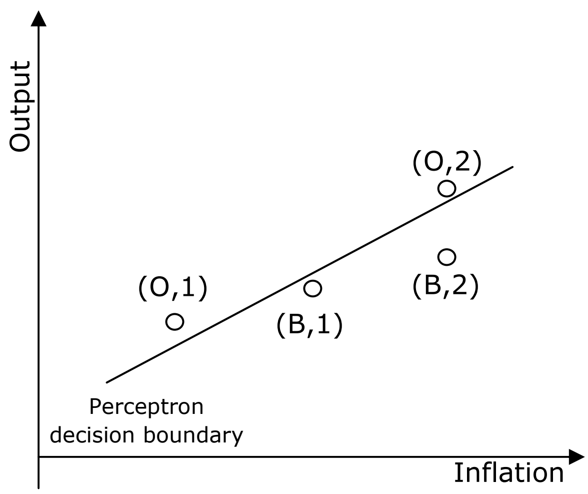

Figure 1 – The neural network classification

Figure 1 allows visualising the opportunistic and benevolent trajectories in the inflation-output, (y,π), space, showing an example where the classification of the incumbent is possible to be achieved by that kind of neural network.

There are, therefore, four points located in the (y, π) space, two of each type, *O* and *B*. This makes possible to draw two line segments connecting the two points of each kind. If these two line segments cross, it is impossible to obtain a decision boundary. This can be checked by a system of equations involving two convex combinations between these points defining the intersection between the straight line segments. They cannot be separated if the two parameters, λ_1, λ_2 in the convex combinations:

$$\lambda_1 \begin{bmatrix} y_1^B \\ \pi_1^B \end{bmatrix} + (1-\lambda_1) \begin{bmatrix} y_2^B \\ \pi_2^B \end{bmatrix} = \lambda_2 \begin{bmatrix} y_1^O \\ \pi_1^O \end{bmatrix} + (1-\lambda_2) \begin{bmatrix} y_2^O \\ \pi_2^O \end{bmatrix}, \tag{22}$$

are both between 0 and 1.

Given the optimal inflation rates, (9) to (12), and output levels, (18) to (21), the solutions for λ_1, λ_2 in (22) are:

$$\lambda_1 = \frac{\alpha^2 \beta}{\phi \mu} \frac{(\phi - \gamma)^2}{(1-\phi) y_0 + \alpha^2 \beta (\gamma - 1)}, \tag{23}$$

$$\lambda_2 = \frac{\alpha^2 \beta \rho}{\phi} \frac{(\phi - \gamma)^2}{(1-\phi) y_0 + \alpha^2 \beta (\gamma - 1)}.^{12} \tag{24}$$

Plainly, in general, the possibility to classify the incumbent depends upon the initial level of output, y_0.[13] Figure 2 thus represents those two solutions (23) and (24) as a function of y_0.

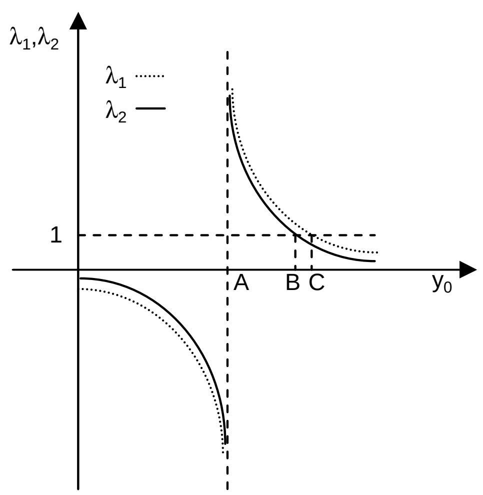

Figure 2 – The influence of initial output level

[12] Note that $\lambda_1 - \lambda_2 = \frac{1-\rho\mu}{\mu} \frac{\alpha^2 \beta}{\phi} \frac{(\gamma - \phi)^2}{(1-\phi) y_0 + \alpha^2 \beta (\gamma - 1)}$.

[13] When $\phi = \gamma$, both λ_1, λ_2 are equal to zero, meaning that both types of incumbents behave the same.

In order to have $\lambda_1 = 1$ in (23), – point C in figure 2 – the initial level of output must be:

$$y_0 = \alpha^2 \beta \frac{(\phi - \gamma)^2 + \phi(1-\gamma)\mu}{\phi(1-\phi)\mu}, \tag{25}$$

whereas, in order to have $\lambda_2 = 1$ in (24), – point B in figure 2 – the initial level of output must be:

$$y_0 = \alpha^2 \beta \frac{\rho(\phi - \gamma)^2 + \phi(1-\gamma)}{\phi(1-\phi)}. \tag{26}$$

As y_0 given by (25) is higher than y_0 given by (26),[14] this means that for

$$y_0 > \alpha^2 \beta \frac{(\phi - \gamma)^2 + \phi(1-\gamma)\mu}{\phi(1-\phi)\mu}, \tag{27}$$

$\lambda_1 < 1$ and, therefore, also that $\lambda_2 < 1$. Moreover,

$$y_0 > \alpha^2 \beta \frac{1-\gamma}{1-\phi} \tag{28}$$

guarantees that both λ_1, λ_2 are positive. See point A in figure 2. After noticing that y_0 given by (25) is higher than y_0 given by (28),[15] it is possible to consider an initial condition

$$y_0 > \alpha^2 \beta \frac{(\phi - \gamma)^2 + \phi(1-\gamma)\mu}{\phi(1-\phi)\mu}, \tag{29}$$

such that it is impossible to associate *all* the observed behaviours to the correct type of incumbent. In all the other cases, the classification task can be resolved by the perceptron. See figure 3.

---

[14] Note that $\alpha^2 \beta \frac{(\phi - \gamma)^2 + \phi(1-\gamma)\mu}{\phi(1-\phi)\mu} - \alpha^2 \beta \frac{\rho(\phi - \gamma)^2 + \phi(1-\gamma)}{\phi(1-\phi)} = \alpha^2 \beta (\phi - \gamma)^2 \frac{1-\mu\rho}{\phi\mu(1-\phi)} > 0.$

[15] Note that $\alpha^2 \beta \frac{(\phi - \gamma)^2 + \phi(1-\gamma)\mu}{\phi(1-\phi)\mu} - \alpha^2 \beta \frac{1-\gamma}{1-\phi} = \alpha^2 \beta \frac{(\phi - \gamma)^2}{\phi\mu(1-\phi)} > 0.$

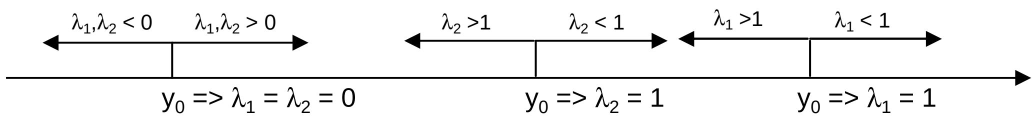

Figure 3 – The classification regions

Notwithstanding that conditionally, there is a fundamental exception. When output does not show any persistence over time, *i.e.* $\phi = 0$, which is, indeed, the most considered case in the literature, it is possible to show that a straight line with intercept between $\alpha^2\beta\gamma(\gamma p - 2)$ and $\alpha^2\beta\gamma\frac{\gamma - 2\mu}{\mu}$ and slope equal to $\alpha(\gamma + 1)$ will always divide the space in a correct way, this being eventually the result of the perceptron classification. See the Annex 2.

Plainly, in practical terms, given that a learning process takes place, from the training of the perceptron does not usually result a straight line with the above mentioned characteristics. Most importantly, given that the two straight lines connecting the two pairs of points in the output-inflation space are parallel, this guarantees that the space is linearly separable. Figure 4 shows this situation.

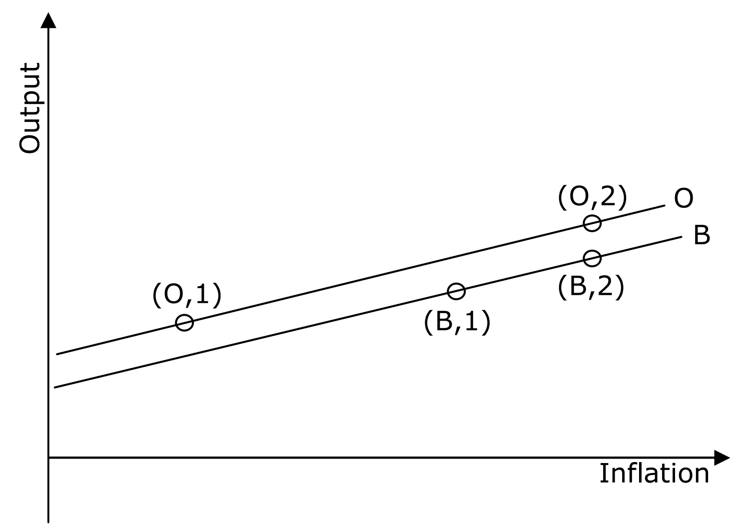

Figure 4 – A particular(ly interesting) case

As it is well known, if the space can be linearly separable, as it is the case when output does not show any persistence, the perceptron will always determine a vector of values for the weights and a bias value such that the straight line associated with these values divide the space in a correct way.

5. Concluding Remarks

The paper explores a crucial aspect in the issues of political business cycles by considering the effects of boundedly rational voters, a facet that has been basically ignored by the literature. It offers the analysis of the bounded rationality approach as a motivation for the use of neural networks as learning devices. The classification task performed by that kind of voters is done by the use a neural network in a model allowing for output persistence. It is shown that when output does not persist the classification task can always be resolved. Conversely, the resolution of the classification task, when output persists over time, depends crucially on the initial conditions.

As a direction for future improvements we would like to explore the possible dynamics of convergence for output in order to check, in the long-run, the real importance of the initial level of output. As, indeed, the steady state cycle, for each kind of incumbent are characterised by a level of output below the one identified by (29), hypothetically the resolution of the classification task may become more probable over time.

References

Ascari, G. (2003). Price/Wage Staggering and Persistence: A Unifying Framework. *Journal of Economic Surveys* 17 (4):.511-540.

Barrell, R., G.M. Caporale, S. Hall, and A. Garratt (1992). Learning about Monetary

Union: An Analysis of Boundedly Rational Learning in European Labour Markets. National Institute of Economic and Social Research Discussion Paper 22.

Başar, T., and M.H. Salmon (1990a). Inflation and the Evolution of the Credibility with Disparate Beliefs. In N.M. Christodoulakis (ed.), *Dynamic Modelling and Control of National Economies 1989: Selected papers from the 6th IFAC Symposium*. Oxford: Pergamon Press, 75-81.

Başar, T., and M. Salmon (1990b). Credibility and the Value of Information Transmission in a Model of Monetary Policy and Inflation. *Journal of Economic Dynamics and Control* 14: 97-116.

Christiano, L.J., M. Eichenbaum, and C. Evans (2005). Nominal rigidities and the dynamic effects of a shock to monetary policy. *Journal of Political Economy* 113 (1): 1-45.

Cho, I.-K., and T.J. Sargent (1996). Neural Networks for Encoding and Adapting in Dynamic Economies. In H.M. Amman, D.A. Kendrick, and J. Rust (eds.), *Handbook of Computational Economics*, *Vol. I*. Amsterdam: Elsevier Science, 441-470.

Cripps, M. (1991). Learning Rational Expectations in a Policy Game. *Journal of Economic Dynamics and Control* 15: 297-315.

Ellacott, S., and D. Rose (1996). Neural Networks: Deterministic Methods of Analysis. London: International Thomson Computer Press.

Evans, G.W. (1986). Selection Criteria for Models with Non-uniqueness. *Journal of Monetary Economics* 18: 147-157.

Evans, G.W., and R. Guesnerie (1993). Rationalizability, Strong Rationality and Expectational Stability. *Games and Economic Behavior* 5: 632-646.

Evans, G.W., and S. Honkapohja (1994). Learning, Convergence, and Stability with Multiple Rational Expectations Equilibria. *European Economic Review* 38: 1071-1098.

Evans, G.W., and S. Honkapohja (1995). Expectational Stability and Adaptive Learning: An Introduction. In A. Kirman and M. Salmon (eds.), *Learning and Rationality in Economics*. Oxford: Basil Blackwell, 102-126.

Evans, G.W., and S. Honkapohja (2003). Adaptive learning and monetary policy design. *Journal of Money Credit and Banking* 35: 1045–1072.

Evans, G.W. and B. McGough (2005). Monetary policy, indeterminacy and learning. *Journal of Economic Dynamics and Control* 29: 1809–1840.

Fuher, J.C., and M.A. Hooker (1993). Learning about Monetary Regime Shifts in an Overlapping Wage Contract Model. *Journal of Economic Dynamics and Control* 17 (4): 531-553.

Gärtner, M. (1996). Political business cycles when real activity is persistent. *Journal of Macroeconomics* 18: 679-692.

Gärtner, M. (1997). Time-consistent monetary policy under output persistence. *Public Choice* 92: 429-437.

Gärtner, M. (1999). The Election Cycle in the Inflation Bias: Evidence from the G-7 countries. *European Journal of Political Economy* 15: 705-725.

Gärtner, M. (2000). Political Macroeconomics: A Survey of Recent Developments. *Journal of Economic Surveys* 14 (5): 527-561.

Honkapohja, S., and K. Mitra (2004). Are non-fundamental equilibria learnable in models of monetary policy? *Journal of Monetary Economics* 51: 1743–1770.

Huang, K.X.D., and Z. Liu (2002). Staggered price-setting,staggered wage-setting, and business cycle persistence. *Journal of Monetary Economics* 49: 405–433.

Jonsson, G. (1997). Monetary Politics and Unemployment Persistence. *Journal of Monetary Economics* 39 (2): 303-325.

Kiley, M.T. (2000). Endogenous Price Stickiness and Business Cycle Persistence. *Journal of Money, Credit and Banking* 32 (1): 28-53.

Lockwood, B. (1997). State-contingent inflation contracts and unemployment persistence. *Journal of Money, Credit, and Banking* 29: 286-299.

Lucas, R. (1973). Some International Evidence on Output-Inflation Tradeoffs. *American Economic Review* 63: 326-334.

Marcet, A., and T.J. Sargent (1988). The Fate of Systems With "Adaptive" Expectations. *American Economic Review* 78 (2): 168-172.

Marcet, A., and T.J. Sargent (1989a). Convergence of Least Squares Learning Mechanisms in Self-Referential Linear Stochastic Models. *Journal of Economic Theory*

48 (2): 337-368.

Marcet, A., and T.J. Sargent (1989b). Convergence of Least-Squares Learning in Environments with Hidden State Variables and Private Information. *Journal of Political Economy* 97 (6): 1306-1322.

Marimon, R., and S. Sunder (1993). Indeterminacy of Equilibria in a Hyperinflationary World: Experimental Evidence. *Econometrica* 61 (5): 1073-1108.

Marimon, R., and S. Sunder (1994). Expectations and Learning under Alternative Monetary Regimes: An Experimental Approach. *Economic Theory* (4): 131-162.

Merkl, C., and D.J. Snower (2007). Monetary Persistence, Imperfect Competition, and Staggering Complementarities. IZA Discussion Paper No. 3033.

Minford, P. (1995). Time-Inconsistency, Democracy, and Optimal Contingent Rules. *Oxford Economic Papers* 47 (2): 195-210.

Nordhaus, W.D. (1975). The Political Business Cycle. *Review of Economic Studies* 42 (2): 169-190.

Rotemberg, J.J., and M. Woodford (1997). An Optimization Based Econometric Framework for the Evaluation of Monetary Policy. In J.J. Rotemberg, and B.S. Bernanke (eds.), *NBER Macroeconomics Annual 1997*, Cambridge (MA): The MIT Press, 297-346.

Salmon, M. (1995). Bounded Rationality and Learning: Procedural Learning. In A. Kirman, and M. Salmon (eds.), *Learning and Rationality in Economics*. Oxford: Basil Blackwell, 236-275.

Sargent, T.J. (1993). Bounded Rationality in Macroeconomics. Oxford: Clarendon Press.

Svensson, L. (1997). Optimal inflation targets, "conservative" central banks, and linear inflation contracts. *American Economic Review* 87: 98-114.

Swingler, K. (1996). Applying Neural Networks: A Practical Guide. London: Academic Press Limited.

Taylor, J. B. (1980). Aggregate dynamics and staggered contracts. *Journal of Political Economy* 88: 1-23.

Wall, K.D. (1993). A Model of Decision Making Under Bounded Rationality. *Journal of Economic Behavior and Organization* 20: 331-352.

Wang, P.-f., and Y. Wen (2006). Another look at sticky prices and output persistence. *Journal of Economic Dynamics and Control* 30: 2533-2552.

Westaway, P. (1992). A Forward-Looking Approach to Learning in Macroeconomic Models. *National Institute Economic Review* 2: 86-97.

White, H. (1989). Some Asymptotic Results for Learning in Single Hidden-Layer Feedforward Network Models. *Journal of the American Statistical Association* 84: 1003-1013.

Annex 1 – The neural networks methodology

Given that (artificial) neural networks are simulations of how biological neurons are supposed to work, the structure of human brains, where processing units, the so-called *neurons*, are connected by *sinapses*, is approximated by these (artificial) neural networks. As such, the interconnected network of processing units describes a model which maps a set of given *inputs* to an associated set of *outputs* values.[16] As the number of inputs does not have to be equal to the number of outputs, a neural network can, alternatively, be described as mapping one set of variables onto another set of a possibly different size.

The knowledge of the values for the input and output variables constitutes, then, the major part of the information needed to implement a neural network. Despite the minimal information requirement, this constitutes no motive for questioning the results obtained (see Salmon 1995). In fact, this characteristic makes neural networks particularly appropriate for cases where the structure connecting inputs to outputs is unknown.[17] In this sense, neural networks can be classified as 'non-structural' *procedural* models. Furthermore, they are in good agreement with a typical characteristic of bounded rationality: the *adaptive* behaviour. Indeed, the adaptation to the environment as a crucial characteristic of a neural network makes it distinct from many (standard) models of learning.[18]

Let us then clarify the *modus operandi* of neural networks by a simple formalisation as follows.[19] Given an input vector *x*, the neural network determines a

---

[16] A more formal definition would consider a neural network $<\mathbf{P}, \prec>$ to be a directed graph over the set $\mathbf{P}$ of processors (neurons), where a processor is a mapping from an input to an output space.

[17] Take, for instance, Wall (1993) which pretends to bridge the gap between *substantive rationality* and *procedural rationality*. The fact that it is considered that the exact form of the objective function is unknown is what makes this *bounded rationality* model a good example of a possible application of neural networks.

[18] In particular, neural networks relax the constant linear reduced form assumption of *least squares learning* by considering a time varying possibly non-linear stochastic approximation of that reduced form.

[19] For a sound mathematical presentation see Ellacott and Bose (1996). More advanced references include White (1989).

particular parameterisation, say β, which, in conjunction with a function g – also possibly determined by the neural network – leads to the output vector $y = g(x, \beta)$ 'closest' to some target y^*. In other words, the output units $y(k)$, $(k = 1,\dots,t)$, process, using a function g, the inputs $x(i)$, $(i = 1,\dots,r)$, previously amplified or attenuated by the connection strengths $\beta(i,k)$.[20]

The simplest neural network structure described above is usually relaxed to obtain flexibility by considering a layer of, so-called, *hidden units*. In this case, the transformation of inputs into outputs includes an intermediate processing task performed by the hidden units. Each hidden unit, then, produces, by the consideration of an *activation* or *transfer function* $f(.)$, an intermediate output $s(j)$, $(j = 1,\dots,s)$, which is finally sent to the output layer.[21]

Mathematically the network then computes:

1. The input(s) to the hidden layer, $h(j)$, as a weighted sum

$$h(j) = b(j) + \sum_{i=1}^{r} w(i,j)x(i) \qquad j = 1,\dots,s;$$

2. The output(s) of the hidden layer, which are the input(s) to the output layer, are subject to an output activation

$$s(j) = f(h(j)),$$

where f is the so-called *activation function*.

3. The output(s) of the output layer[22]

$$y(k) = \sum_{j=1}^{s} \beta(j,k)s(j) \qquad k = 1,\dots,t.$$

Plainly the two crucial elements of a neural network are the parameter set

[20] Implicitly assumed is a *feedforward model* where *signals* flow only from $x(i)$ to $y(k)$. Nevertheless, it is also possible to consider *feedback effects*.

[21] It is also (and generally) possible to consider a *bias node* 'shifting' the weighted sum of inputs by some factor $b(j)$.

[22] It is possible to consider an activation function and/or a bias before the determination of the 'final' outputs.

$\theta = (w, \beta)$ and the activation/transfer function *f*(.). The transfer function usually has the role of normalising a node's output signal strength between 0 and 1.[23] The most used are the tanh or some sigmoid function $f(h) = (1 + \exp(-h))^{-1}$ and a gaussian function or some radial basis function.

As usual, once the parameters have been set, say $\hat{\theta}$, the neural network is able to predict outputs $\hat{y} = g(\bar{x}, \hat{\theta})$ for input values $\bar{x}$ which were not included in the training data.

A.1. The learning process

As pointed out in White (1989), the output vector $y = g(x, \theta)$ can be viewed as generating a family of approximations (as θ ranges over the set Θ, say) for the *unknown* relation between inputs *x* and their corresponding outputs *y*. The best approximation can be determined by a *recursive learning* procedure known as back-propagation. The learning process – *training* – is then an iterative procedure of processing inputs through the neural network, determining the errors and back-propagating the errors through the network to adjust the parameters in order to minimise the error between the predicted and observed outputs. This method of learning is referred to as *gradient descent* as it involves an attempt to find the lowest point in the error space by a process of gradual descent along the error surface.[24]

Annex 2 – Mathematical details

In the case $\pi_0 = \alpha\beta$, the solutions are:

[23] This is why some authors designate these functions as *squashing functions*.

[24] Two factors are used to control the training algorithm's adjustment of the parameters: the *momentum factor* and the *learning rate coefficient*. The *momentum term*, which is quite useful to avoid local minima, causes the present parameter changes to be affected by the size of the previous changes. The *learning rate*

$$\pi_1^B = \alpha\beta\left(1-\rho(\gamma-\phi)\right), \qquad y_1^B = \phi y_0 + \alpha^2\beta\left(1-\rho(\gamma-\phi)-\gamma\right),$$

$$\pi_2^B = \alpha\beta, \qquad y_2^B = \phi^2 y_0 + \alpha^2\beta\left(\phi - 2\phi\rho\gamma + \rho\phi^2 - \phi\gamma + 1 - \gamma + \gamma^2\rho\right),$$

$$\pi_1^O = \alpha\beta\left(1-\frac{\gamma-\phi}{\mu}\right), \qquad y_1^O = \phi y_0 + \alpha^2\beta\frac{\mu-\gamma+\phi-\gamma\mu}{\mu},$$

$$\pi_2^O = \alpha\beta, \qquad y_2^O = \phi\left(\phi y_0 + \alpha^2\beta\frac{\mu-\gamma+\phi-\gamma\mu}{\mu}\right) + \alpha\left(\alpha\beta - \gamma\alpha\beta\left(1-\frac{\gamma-\phi}{\mu}\right)\right).$$

This means that:

$$\pi_1^B - \pi_1^O = \alpha\beta(\gamma-\phi)\ \frac{1-\mu\rho}{\mu}, \qquad \pi_2^B - \pi_2^O = 0,$$

$$y_1^B - y_1^O = \alpha^2\beta(\gamma-\phi)\ \frac{1-\mu\rho}{\mu}, \qquad y_2^B - y_2^O = -\alpha^2\beta(\gamma-\phi)^2\frac{1-\mu\rho}{\mu}.$$

Given the previous expressions, the slopes and intercepts of the straight lines, $y_t^i = a^i + b^i\pi_t^i$, for i = B,O and t = 1,2 are:

$$b^B = \frac{\phi(\phi-1)\ y_0 + \alpha^2\beta\left(-\gamma\phi+\gamma^2\rho-2\phi\rho\gamma+\rho\gamma+\rho\phi^2+\phi-\rho\phi\right)}{\beta\alpha\rho(\gamma-\phi)},$$

$$a^B = \frac{\phi\left(-1+\phi+\rho\phi^2-\phi\rho\gamma\right)\ y_0 + \alpha^2\beta\left(\phi\rho\gamma^2-\gamma\rho\phi^2+\phi-4\phi\rho\gamma+2\rho\phi^2-\gamma\phi+2\gamma^2\rho+3\phi\rho^2\gamma^2-3\phi^2\rho^2\gamma+\rho^2\phi^3-\gamma^3\rho^2\right)}{(\phi-\gamma)\ \rho},$$

$$b^O = \frac{\phi\mu(\phi-1)\ y_0 - \alpha^2\beta\left(\phi\gamma\mu-\gamma^2+2\gamma\phi-\gamma-\phi^2-\phi\mu+\phi\right)}{\beta\alpha(\gamma-\phi)},$$

$$a^O = \frac{\mu\phi\left(\phi^2+\phi\mu-\gamma\phi-\mu\right)\ y_0 - \alpha^2\beta\left(4\phi\gamma\mu-3\phi\gamma^2+3\phi^2\gamma+\gamma^3-\phi^3-2\mu\gamma^2+\mu^2\phi\gamma+\mu\phi^2\gamma-\mu\phi\gamma^2-\mu^2\phi-2\phi^2\mu\right)}{(\phi-\gamma)\ \mu}.$$

Those two lines will cross at

---

dictates the proportion of each error which is used to update parameters during learning.

$$\pi = \alpha\beta\left(1 - \frac{\alpha^2\beta\rho(\phi-\gamma)^3}{\left((\phi-1)\,y_0 + \alpha^2\beta(1-\gamma)\right)\mu\phi}\right).$$

Moreover,

$$b^B - b^O = \phi(1-\rho\mu)\ \frac{(\phi-1)\ \ y_0 - \alpha^2\beta(\gamma-1)}{\alpha\beta\rho(\gamma-\phi)},$$

$$a^B - a^O = (1-\rho\mu)\ \frac{\phi\mu(1-\phi)\ \ y_0 + \alpha^2\beta\left(-\phi\mu + \phi\gamma\mu - \rho\gamma^3 + 3\phi\rho\gamma^2 - 3\gamma\rho\phi^2 + \phi^3\rho\right)}{(\gamma-\phi)\ \ \mu\rho}.$$

In the particular case of an initial output level $y_0 = \alpha^2\beta\frac{1-\gamma}{1-\phi}$ we have

$$b^B = b^O = \alpha(1-\phi+\gamma) > 0,$$

$$a^B - a^O = -(1-\rho\mu)\ \ \alpha^2\beta\tfrac{(\gamma-\phi)^2}{\mu} < 0.$$

In case of no persistence at the output level, that is when $\phi = 0$, the solutions are:

$$\pi_1^B = \alpha\beta(1-\rho\gamma), \qquad \pi_2^B = \alpha\beta,$$

$$y_1^B = \alpha^2\beta(1-\gamma-\gamma\rho),\ y_2^B = \alpha^2\beta\left(1-\gamma+\gamma^2\rho\right),$$

$$\pi_1^O = \alpha\beta\left(1-\frac{\gamma}{\mu}\right), \qquad \pi_2^O = \alpha\beta,$$

$$y_1^O = \alpha^2\beta\frac{\mu-\gamma-\gamma\mu}{\mu}, \qquad y_2^O = \alpha^2\beta\frac{\mu+\gamma^2-\gamma\mu}{\mu}.$$

It is then straightforward to verify that:

$$\pi_1^B - \pi_1^O = \alpha\beta\gamma\frac{1-\mu\rho}{\mu} > 0, \qquad \pi_2^B - \pi_2^O = 0,$$

$$y_2^B - y_1^B = \alpha^2\beta\gamma(1+\gamma)\rho > 0, \qquad y_2^O - y_1^O = \alpha^2\beta\gamma(1+\gamma)\frac{1}{\mu} > 0,$$

$$y_1^B - y_1^O = \alpha^2\beta\gamma\frac{1-\mu\rho}{\mu} > 0, \qquad y_2^B - y_2^O = -\alpha^2\beta\gamma^2\frac{1-\mu\rho}{\mu} < 0,$$

$$y_2^O - y_1^B = \alpha^2\beta\gamma\frac{\gamma+\mu\rho}{\mu} > 0.$$

Given the previous expressions, the slopes and intercepts of the straight lines, $y_t^i = a^i + b^i\pi_t^i$, for i = B,O and t = 1,2 are:

$$b^B = \alpha(\gamma+1)$$

$$a^B = \alpha^2\beta\gamma(\gamma\rho-2)$$

$$b^O = \alpha(\gamma+1)$$

$$a^O = \alpha^2\beta\gamma\tfrac{\gamma-2\mu}{\mu}.$$

Plainly,

$$a^O - a^B = \alpha^2\beta\gamma^2\tfrac{1-\mu\rho}{\mu} > 0.$$

Economics
The Open-Access, Open-Assessment E-Journal

Please note:

You are most sincerely encouraged to participate in the open assessment of this discussion paper. You can do so by posting your comments.

Please go to:

http://www.economics-ejournal.org/economics/discussionpapers/2008-16

The Editor