NOTAS ECONÓMICAS

20

NÚMERO 20/Dezembro'04/PREÇO 7,5€/ISSN:0872-4733

REVISTA DA FACULDADE DE ECONOMIA DA UNIVERSIDADE DE COIMBRA

Nota do Director

A Feuc comemorou, em 2004, o seu 30º aniversário, e entre as iniciativas de maior vulto teve lugar a Conferência Internacional **Economic Policies in the New Millennium**. Este evento, realizado nos dias 16 e 17 de Abril, contou com a presença de mais de uma centena de conferencistas, distribuídos pelas mais variadas áreas da investigação económica contemporânea. Terminada a conferência, e de acordo com o plano inicial, os autores foram convidados a apresentar os seus trabalhos a *Notas Económicas*, tendo-se de imediato procedido à respectiva avaliação no sistema de *blind refereeing*. Deste processo de selecção resultou o presente volume de *Actas*, composto por uma conjunto de 10 artigos originais e uma contribuição especial, a cargo do Prof. Antoine d'Autume.

Este número resulta do empenho de muitos colegas de Faculdade, a quem agradeço a colaboração sempre pronta. Mas uma menção muito especial é naturalmente devida a todos os *referees*, que, numa prova de grande generosidade, tornaram possível, em tempo verdadeiramente *record* para realizações similares, a saída desta edição. A estes colegas deixo pois os meus mais vivos agradecimentos pela honra que nos deram: Ana Paula Serra, Carlos Marinheiro, Eduardo Anselmo, Eduardo Barata, Fernando Alexandre, Filipe Coelho, Graça Leão, Hélder Sebastião, Jorge Andraz, Leonida Carreira, Lígia Pinto, Linda Veiga, Luís Cruz, Luís Delfim, Luís Vasconcelos, Maria da Conceição Portela, Mário Graça Moura, Maria João Thompson, Manuel Matos, Maurizio Mistri, Miguel St. Aubyn, Nuno Valério, Óscar Lourenço, Paulo Trigo Pereira e Pedro Bação.

Economic Policy, Innovation and Growth

Antoine d'Autume
EUREQua, University of Paris I Panthéon-Sorbonne

It is appropriate to introduce a paper presented at an anniversary conference in Coimbra with a reference to the Lisbon Declaration. Meeting in March 2000, the European Council declared:

> *The Union has today set itself a new strategic goal for the next decade: to become the most competitive and dynamic knowledge-based economy in the world, capable of sustainable economic growth with more and better jobs and greater social cohesion.*

Quantitative objectives were later adopted in Barcelona in 2002. They stated that Research & Development investment in the European Union should rise to 3% of GDP by 2010, up from 1.9% in 2000. Two-thirds of this investment should be business funded, up from 56% now.

When we survey recent research on growth theory, we find it surprisingly in phase with this agenda. Endogenous growth models do stress that innovation is the main engine of growth. They give a prominent role to the innovation that takes place inside private firms and consider the share of R&D expenses as a crucial element. Human capital also appears as an essential determinant of growth. Lastly, the importance of the legal environment is increasingly recognized. Patents and intellectual property protection are part of the theoretical framework.

Our aim in this article is to provide an overview of economic policy in endogenous growth models of innovation. Where do we stand, just over twenty years after the pioneering articles of Romer (1986, 1990), Lucas (1988) and Aghion-Howitt (1992)? What should be done in order to increase the realism of the models and, in particular, to take a more subtle view of economic policies aimed at promoting growth?

We start with the comprehensive endogenous growth model of Jones-Williams (2000). This simple synthetic model allows a clear identification of the externalities and distortions which surround innovation and knowledge accumulation. We then describe corrective economic policies, which take the simple form of various subsidies.

The Jones-Williams model is a model of semi-endogenous growth, where the long run rate of growth is determined by technology alone. While it retreats from the original goal of explaining the level of the long run rate of growth, it seems to fit the facts better than previous pure endogenous growth models did. Contrary to what is sometimes asserted, it is not exempt from a strong scale effect however. We stress that this feature should induce us to analyze transitional dynamics rather than just focusing on the long run steady state.

As with more standard endogenous growth models, the Jones-Williams model provides a useful framework to assess the determinants of innovation and its macroeconomic consequences. Yet, the messages it communicates seem too simple in several respects. First the model puts too much weight on the Schumpeterian idea that monopoly on the product market is good for innovation and growth. A more balanced view of the effects of competition has been put forward by Aghion and others. Second, the specificity of knowledge transmission is overly simplified and therefore so is the role of patents. A better description and understanding of intellectual property protection is required. Growth theory may gain here from recent work in Industrial Organization.

We briefly explore these two avenues for future research.

A benchmark model of semi-endogenous growth

The Jones-Williams (2000) model is an appropriate point of departure as it captures in a simple way a number of elements of the new growth theories. Innovation takes the form of an increase in the variety of intermediate goods. Several externalities affect the innovation process.

The final good is produced using labor and a variety of intermediate goods.

$$Y = L^{\alpha} J^{\,1-\alpha}, \quad J = \left(\int_0^A x(j)^{1-1/\gamma}\, dj \right)^{\gamma/(\gamma-1)}$$

At a given time, there is a number (more precisely a mass) A of differentiated intermediate goods and $x(j)$ denotes the amount used of intermediate good j. A CES aggregator J of all intermediate goods is introduced, where $\gamma > 1$ is the elasticity of substitution between the intermediate goods. A Cobb-Douglas function links the final product Y to labor L and intermediate goods J.

The intermediate goods are made with capital. A unit of capital may be transformed into a unit of any existing intermediate good:

$$K = \int_0^A x(j)dj$$

At a symmetric equilibrium,

$$K = Ax, \quad J = A^{\frac{\gamma}{\gamma-1}}\, x = A^{\frac{1}{\gamma-1}}\, K$$

We end up with the following aggregate production function

$$Y = A^{\beta} L^{\alpha} K^{1-\alpha}, \quad \beta = \frac{1-\alpha}{\gamma-1}$$

This function displays a *taste for variety*: for given capital and labor, it is increasing in the number A of intermediate goods.

New intermediate goods are the result of a Research-Development activity. New designs are created according to the following technology

$$\dot{A} = Y_A - \psi \dot{A}$$

$$Y_A = \delta R \bar{R}^{-(1-\lambda)} A^{\phi}, \quad \textit{ex ante}$$

$$= \delta R^{\lambda} A^{\phi}, \quad \textit{ex post}$$

Y_A is the amount of new designs. Not all these designs add to the actual variety of intermediate goods, however. A proportion ψ of new designs simply replace older, inferior, intermediate goods. Equivalently, we may consider that a proportion $\psi \dot{A}/A$ of existing intermediate goods becomes obsolete per unit of time. Thus, a higher rate of innovation exacerbates this *creative destruction* effect.

R&D uses the final good as an input. Let R be the overall input in the R&D sector and δ a positive productivity coefficient. The production of new designs is subject to two kinds of externalities.

The first is the risk of duplication or a *stepping on toe* effect. The effort of other researchers has a negative effect on the actual production of a researcher. Thus, if R is the input of the representative researcher, and $\bar{R}$ the average input of other researchers, the production of new designs is proportional to $R\bar{R}^{-(1-\lambda)}$, with $0 < \lambda < 1$. Ex post, all researchers choose the same input so that $\bar{R} = R$. The amount of innovation is then proportional to R^{λ}.

This extends the original aim of Jones (1995). Aggregate R&D is subject to decreasing returns, even if it is reasonable to assume that each individual agent perceives his activity to have constant returns to scale.

The second externality is created by the existing stock of innovations. How does the size of this stock influence the difficulty to innovate? Formally, the issue is to know what value should be retained for the parameter ϕ affecting existing knowledge A in the R&D production function.

Following Romer (1990), standard models of endogenous growth assume $\phi = 1$. This describes a *standing on the shoulders* effect. Knowledge remains a public good for research, even if it is patented. This effect is strong and is responsible for the possibility of endogenous growth at a constant rate.

This over-optimistic assumption has been criticized by Jones (1995). We might alternatively consider that innovation amounts to fishing for new ideas in a pond of potential innovations. The higher the level of previous catches, the more difficult to find new fish. In such a case, ϕ should be negative.

A more realistic assumption is perhaps to admit a positive, but limited effect of previous innovation. We thus assume $0 < \phi < 1$. As we shall see, this is the case which leads to semi-endogenous growth.

We close the production side of the model with the assumption that the same final good is used alternatively for consumption, research and capital accumulation. Taking into account depreciation at rate μ, we have

$$\dot{K} = Y - \mu K - cL - R$$

Population increases at the exogenous rate η.

The demand side is standard. The representative consumer maximizes the discounted sum of instantaneous utility. The intertemporal elasticity of substitution is σ and the rate of discount ρ.

Semi-endogenous growth

As Romer (1986) discussed in his early work, endogenous growth at a constant rate results from the assumption of constant returns to scale with respect to accumulable factors. Semi--endogenous growth, on the other hand, relies on the assumptions of decreasing returns to scale with respect to accumulable factors, but increasing returns to scale with respect to all production factors. The first assumption precludes permanent growth in the absence of population growth. The second one allows positive growth of income per capita, but makes its level depend on the rate of population growth. Due to increasing returns, a strong population growth appears as a good thing for income per capita.

Assessing the presence of increasing or decreasing returns to scale in a multisectoral model such as ours is not quite clear. Eliminating the research input R between the two production function yields the overall possibility frontier

$$C + \dot{K} + \left(\frac{(1+\psi)\dot{A}}{\delta A^{\phi}} \right)^{1/\lambda} \leq A^{\beta} L^{\alpha} K^{1-\alpha} - \mu K$$

This frontier describes feasible net products for consumption C, capital accumulation $\dot{K}$ and knowledge accumulation $\dot{A}$, for given levels of A, K and L.

Let us assume that C and $\dot{K}$ are increased by a common positive factor θ, while $\dot{A}$ and A are jointly increased by a factor $\lambda\theta/(1-\phi)$. This allows the last component in the left-hand-side of the previous inequality, namely the research effort R, to increase by the factor θ.

The issue is whether the same increase in the existing stock of capital makes the increase in outputs feasible. Under our assumptions, production increases by a factor $\beta\lambda\phi/(1-\phi)+(1-\alpha)\theta$. Thus, the proportionate increase in inputs and outputs is feasible if this expression is larger than θ. In order to exclude increasing returns to scale with respect to accumulable factors, we must assume the contrary.

Turning to returns to scale with respect to all factors, we allow a proportionate increase in labor. The condition for increasing returns becomes that $\beta\lambda\phi/(1-\phi)+\alpha\phi+(1-\alpha)\theta$ should be larger than θ. This is always satisfied.

We thus are led to assume

$$\beta\lambda/(1-\phi)<\alpha$$

This ensures both decreasing returns to scale with respect to accumulable factors and increasing returns to scale with respect to all factors.

The characterization of constant rate growth paths confirms this result. Let us assume that all final good quantities grow at rate g, while knowledge grows at rate g_A and population at rate η. All these growth rates are constant.

The production functions imply

$$\alpha g=\beta g_A+\alpha\eta$$

$$(1-\phi)g_A=\lambda g$$

and therefore

$$g=\frac{\alpha(1-\phi)}{\alpha(1-\phi)-\beta\lambda}\eta,\quad g-\eta=\frac{\beta\lambda}{\alpha(1-\phi)-\beta\lambda}\eta,$$

Under the previous assumption, the long run per capita rate of growth is positive if population growth is positive. An increase in η implies an increase in $g-\eta$.

The equilibrium

The description of the equilibrium follows the usual lines.

Let ρ be the price of the final good, ρ_A the price of a patent, ν_A the rental of a patent, q the rental of intermediate goods, π the amount of profit of an intermediate good producer. All these prices are in terms of the final good. Let r be the real interest rate.

The equilibrium is described by the following relationships.

Consumer behavior:

$$\dot{c}/c = \sigma (r - \rho)$$

Demand for intermediate goods:

$$(1-\alpha)\frac{Y}{K} = q$$

Zero profit in R&D:

$$\rho_A Y_A = R$$

Mark-up pricing of the services of intermediate goods:

$$q = \frac{\gamma}{\gamma - 1}(r + \mu)$$

Profit in intermediate goods:

$$\pi = (q - r - \mu)x = \frac{q}{\gamma}\frac{K}{A} = \frac{1}{\gamma}(1-\alpha)\frac{Y}{A}$$

Arbitrage:

$$r = \frac{\nu_A}{\rho_A} + \frac{\dot{\rho}_A}{\rho_A}, \quad \frac{\nu_A}{\rho_A} = \frac{\pi}{\rho_A} - \psi g_A$$

The optimum

Using the same variables, the conditions for an optimum may be expressed as follows

$$\dot{c}/c = \sigma (r - \rho)$$

$$(1-\alpha)\frac{Y}{K} = q$$

$$\frac{1}{1+\psi}\lambda\frac{Y_A}{R} = \frac{1}{\rho_A}$$

$$q = r + \mu$$

$$r = \frac{\nu_A}{\rho_A} + \frac{\dot{\rho}_A}{\rho_A}, \quad \frac{\nu_A}{\rho_A} = \beta\frac{Y}{\rho_A A} + \frac{\phi}{1+\psi}\frac{Y_A}{A}$$

Comparing the equilibrium and the optimum

We are now able to understand why the spontaneous equilibrium is not optimal. Three equations differ between the two sets of conditions. We consider them in turn.

Firstly, research efforts depend on different notions of the marginal productivity of research.

In equilibrium, the relevant notion is the private, ex ante, marginal productivity of research. As the individual firm takes as given $\bar{R}$ as well as the amount of destructive creation $\psi\dot{A}$, it is equal to Y_A/R.

At the optimum, the relevant notion is the social, ex post, marginal productivity of research. It is equal to $[\lambda/(1 + \psi)]\ Y_A/R$. Two negative externalities add up. The first one is represented by the coefficient λ which captures a *stepping on toe* effect. There is some duplication and redundancy in research efforts. The second one is represented by $1/(1 + \psi)$. It captures the *destructive creation* or *business stealing* effect. Part of the new goods substitute the older ones.

These two externalities might induce us to think that there will be too much research at the competitive equilibrium. This is only part of the story, however.

Secondly, imperfect competition on the market for intermediate goods creates a distortion. A mark-up $\gamma/(\gamma - 1)$ appears in the pricing of patented intermediate goods. This tends to restrict the use of intermediate goods.

Lastly, the return on an innovation is very different in the two sets of conditions. In the case of the equilibrium, the rental for knowledge is simply the rate of profit. Firms, however, take into account the risk of being out of business, that is the *destructive creation* effect.

In the case of the optimum, the rental is the sum of the direct effect of knowledge on final production (*taste for variety*) and the external effect of knowledge on future research (*standing on the shoulders*).

The overall effect of all these externalities and distortions is difficult to ascertain. As was emphasized by Jones-Williams, there may be over-investment or under-investment in research, depending on the parameters of the model. The more plausible configuration, however, is one where positive externalities dominate. As agents do not take them into account, the equilibrium research effort then is lower than the optimal one.

The long run

In a semi-endogenous growth framework, the long run rates of growth are exogenous and identical at the equilibrium and the optimum. It follows that the real interest rate also takes the same value r^*. As the rate of growth of individual consumption is $g - \eta$, we have $r^* = \rho + (g - \eta)/\sigma$.

The main difference between the equilibrium and the optimum is that they are characterized by different long run levels of research effort. Letting $s = R/Y$ we get, for the optimum

$$s^* = \frac{\beta\lambda g_A}{r^* - g + (1 - \phi)g_A}$$

and, for the equilibrium,

$$s^{eq} = \frac{\frac{1-\alpha}{\gamma}\ (1 + \psi)g_A}{r^* + \psi g_A - g + g_A}$$

As just mentioned, we focus on the case $s^{eq} < s^*$.

A calibration

We use a calibration close to those considered in Jones-Williams.

We start with the following plausible values:

$$g = .02, \quad \eta = .005, \quad r^* = .045, \quad \mu = .0333, \quad \alpha = .7, \quad \gamma = 4$$

In particular, an elasticity $\gamma = 4$ yields $\gamma / (\gamma - 1)$ 1.33, that is a markup of 33% which seems satisfactory.

It follows that

$$\beta = \frac{1 - \alpha}{\gamma - 1} = .1, \quad g_A = \frac{\alpha}{\beta} (g - \eta) = .105$$

The two externality parameters λ and ϕ have to satisfy $(1 - \phi) / \lambda = g/g_A = 4/21$. We choose $\phi = \lambda = 21/25$

Lastly, we follow Jones-Williams and use the average life-time of a patent to calibrate ψ. As the replacement of a patent follows a Poisson process with parameter ψg_A, this average life-time is $1/(\psi g_A)$. Equalling this life-time to 10 years yields $\psi = 1/1.05$.

Thus

$$\lambda = \phi = .84 \qquad \psi = .95$$

This calibration yields the following results for the long run paths.

Table 1

	R / Y	C / Y	K / Y	$A^{g/gA} / Y$	$Y / L^{g/n}$
spontaneous equilibrium	0.07	0.78	2.87	0.017	0.012
optimum	0.21	0.58	3.83	0.054	0.597

The optimal growth path is both more knowledge and capital intensive than the equilibrium path. Note that the definition of knowledge intensity must take care of the fact that A and Y do not grow at the same rate, whence the presence of the exponent g/g_A. A greater research effort and a higher saving rate are required to support the optimal growth path. All these features are natural enough and the magnitudes reasonable.

The last column is more surprising, however. It reports an indicator of income per capita which varies wildly. Per capita income is fifty times higher on the optimal path than on the equilibrium path.

Let us first clarify the meaning of this observation. In the long run, income per capita may be decomposed as the following product of two terms: $Y/L = (Y/L^{g/\eta})\, L^{g/\eta - 1}$. The first term is constant on the long run path, and is reported in the table. The second term is an increasing function of L. It grows at a constant rate, and so will income per capita. The faster the rate of population growth, the higher the rate of growth of income per capita. Moreover, the higher the population size, the higher the level of per capita income.

These features are typical of semi-endogenous growth models. They clearly follow from the assumption of increasing returns with respect to all factors. A scale effect is present in semi endogenous growth models. Other things being equal, it is better to live in a country with faster population growth or, simply, a higher level of population.

Moreover, differences in savings rates and research effort have huge effects on $Y/L^{g/\eta}$ and, therefore, on the level of income per capita. This should come as no surprise. If the economy is at the same time more knowledge and capital intensive, it has to be less labor intensive, which amounts to saying that per capita income will be higher. The huge size of the level effect can be understood if we recall that this effect would be infinite in a strictly endogenous growth framework. Indeed, the optimum would then be characterized by a higher rate of growth than the equilibrium. The semi-endogenous model that we consider is not very far from this benchmark.

The conclusion is that we should not focus too much on the long run steady state levels. The initial message of semi-endogenous growth was that the focus should be on the dynamics rather than on the long run path. This message should not be forgotten.

The decentralization of the optimum

As the equilibrium is not optimal, there is scope for government intervention aimed at restoring efficiency. To this end, the government may use various tax or subsidy instruments. To keep things simple, we assume that the government is able to levy lump-sum taxes in order to balance its budget constraint. It will then be able to reach the first best optimum. The issue therefore is how to decentralize this optimum.

We introduce subsidies on patents, intermediate goods services and interest rates. Let $\hat{\rho}_A$ and $\hat{q}$ be the selling prices of patents and intermediate goods services, while ρ_A and q are the buying prices. Similarly, let $\hat{r}$ be the interest rate paid by debtors while r is the interest rate received by creditors. If θ_A and θ_q are the subsidy ratios on patents and intermediate goods, and θ_q the subsidy on interest rates, we have

$$\hat{\rho}_A = \theta_A \rho_A , \quad \hat{q} = \theta_q q , \quad \hat{r} = r - \theta_r$$

We show in the appendix that two instruments are sufficient to restore efficiency. The government should subsidize the production and use of knowledge. This can be done through different combinations of subsidies.

Long run values for these subsidy rates are, for instance,

$$\theta_r = 0 , \quad \theta_q = 1.33 , \quad \theta_A = 2.37$$

If the government chooses not to intervene on the financial market, the optimal policy is to subsidize intermediate goods, in order to compensate for the effects of the mark-up, and to subsidize R&D. This last subsidy must be very large, as it should be more than 200%.

Relying on an interest subsidy allows the government to reduce θ_q but forces an increase in θ_A. A possible combination is

$$\theta_r = 0.02 , \quad \theta_q = 1 , \quad \theta_A = 2.89$$

These numerical evaluations are purely illustrative. They do suggest, however, that strong and lasting subsidies are required to move the economy closer to the optimum.

Is competition detrimental or conducive to growth?

The endogenous growth literature has highlighted the Schumpeter argument according to which competition on the product market is bad for innovation and growth as it reduces the innovator's rent. This runs against the age-old idea that more competition should increase the pressure to innovate. Such an argument is central, for instance, in the discussion of the merits of international trade liberalization. Most people would agree that this dynamic effect of competition largely dominates the static gains of specialization.

Aghion-Harris-Vickers (1997-2001) have thus reconsidered the argument that R&D often aims at escaping competition.

They underline the fact that usual endogenous growth models assume that an innovating firm *leapfrogs* to become a leader. They point out that step by step innovation might be more realistic. A firm must catch-up before becoming a leader. The leader and the incumbent then face strong, *neck-to-neck*, competition. Both have a strong incentive to innovate.

In this scenario, the effects of the degree of competition become ambiguous.

In neck-to-neck industries, more competition on the product market induces more innovation to escape competition, that is to become a leader.

In leader-follower industries, where one firm is one step ahead, more competition on the product market means a lower rent for the winner and therefore less innovation.

Moreover a composition effect arises. More competition means a lower proportion of neck-to--neck industries as more firms escape from this situation.

Aghion-Bloom-Blundell-Griffith-Howitt (2002) provide an empirical investigation of the relationship between innovation and competition. An obvious prerequisite is how to define and measure both terms of the relationship.

The intensity of competition is characterized by the elasticity of substitution between differentiated products. It is measured, empirically, by the Lerner index in neck-to-neck industries, which should reflect this elasticity.

The level of innovation is measured by the number of patents of UK firms in the US, weighted by the number of citations in other patents.

The authors show that their theoretical model predicts the existence of an inverted-U relationship between competition and innovation. For low levels of competition the escape-competition effect dominates, while the Schumpeter effect dominates for high levels of competition.

This result is corroborated by a study of UK firms, even if the econometric estimation of such a non-linear relation may appear somewhat daring.

In any case, a more detailed discussion of competition and its role in innovation appears as a promising avenue. It has been pursued, for instance, by Encaoua-Ulph (2000) who reintroduced a role for leapfrogging.

How much should we protect intellectual property?

Another over-simplification of endogenous growth models is the representation of knowledge externalities and the way they should be dealt with institutionally.

The Romer (1990) idea was clear-cut and powerful. The patented innovation becomes a private good for production but remains a public good for future research. This assumption allows for an easy diffusion of scientific knowledge. New researchers are freely standing on the shoulders of their predecessors.

It is not clear however that this assumption takes a realistic view of the world, or that it provides an optimal arrangement.

Firms realize that the diffusion of knowledge reduces the protection offered by a patent. It makes it easier to produce substitutes.

Would a stronger protection of intellectual property be preferable? Should future researchers be paying for their use of previous ideas? In other words, should ideas themselves, rather than new goods, be patented?

On one hand, this would seem to promote upstream innovations, as they would get better compensation. On the other hand, such a scheme would most probably hinder downstream innovation. Prominent examples can be found in the fields of genetics or software research. Patenting a gene or, indeed, a computer operating system, may prevent competitors from engaging in further fruitful research. Many accounts show this to be more than a remote possibility: see, for instance, Tirole et al. (2003).

The design of intellectual property protection is difficult and may take various forms, as has first been shown by Scotchmer (1991). Followers may make business agreements with the leading firm or enter into patent pools based on the use and development of a standard. Alternatively, public authorities may enforce compulsory licenses in order to impose access, for a fee, to standard technologies.

Many recent articles examine these issues and the surveys by Tirole et al. (2003) and Encaoua-Guellec-Martinez (2003) offer useful references. Some papers, in particular, stress the implication for growth theory: see O'Donoghue-Zweimuller (2004), Grimaud (2002), Tournemaine (2004).

Appendix: The decentralization of the optimum

We examine the decentralization of the entire optimal path. This path is characterized by the following dynamic system:

$$\dot{K} = Y - \mu K - cL - R$$

$$(1+\psi)\,\dot{A} = \delta R^{\lambda} A^{\phi}$$

$$\frac{\dot{c}}{c} = \sigma\left[(1-\alpha)\,\frac{Y}{K} - \mu - \rho\right]$$

$$\frac{\dot{\rho}_A}{\rho_A} = (1-\alpha)\,\frac{Y}{K} - \mu - \beta\,\frac{Y}{\rho_A A} - \frac{1}{1+\psi}\,\phi\,\frac{Y_A}{A}$$

with

$$R = \left[\frac{\lambda}{1+\psi}\;\delta\rho_A A^{\phi}\right]^{1/(1-\lambda)}$$

On the other hand, the fiscal equilibrium is characterized by the following system:

$$\dot{c}/c = \sigma\,(r-\rho)$$

$$(1-\alpha)\,\frac{Y}{K} = q$$

$$\hat{\rho}_A Y_A = R$$

$$\hat{q} = \frac{\gamma}{\gamma - 1} (\hat{r} + \mu)$$

$$\pi = (\hat{q} - \hat{r} - \mu)x = \frac{\hat{q}}{\gamma} \frac{K}{A} = \theta_q \frac{1-\alpha}{\gamma} \frac{Y}{A}$$

$$\hat{r} = \frac{\pi}{\rho_A} - \psi g_A + \frac{\dot{\rho}_A}{\rho_A}$$

which leads to the following dynamic characterization:

$$\dot{K} = Y - \mu K - cL - R$$

$$(1 + \psi)\,\dot{A} = \delta R^{\lambda} A^{\phi}$$

$$\frac{\dot{c}}{c} = \sigma \left[\theta_r + \theta_q \frac{\gamma - 1}{1} (1 - \alpha) \frac{Y}{K} - \mu - \rho \right]$$

$$\frac{\dot{\rho}_A}{\rho_A} = \theta_q \frac{\gamma - 1}{1} (1 - \alpha) \frac{Y}{K} - \mu - \theta_q \frac{1-\alpha}{\gamma} \frac{Y}{\rho_A A} + \psi \frac{\dot{A}}{A}$$

with

$$R = [\theta_A\, \delta\, \rho_A A^{\phi}]^{1/(1-\lambda)}$$

Choosing $\theta_A = \lambda/(1+\psi)$, that is taxing the sale of patents in order to counter the two external effects of *stepping on toe* and *destructive creation,* may appear optimal. If the price ρ_A of a patent is set right through other instruments, an adequate amount of research would be provided.

This remark is misleading, however. Decentralization of the optimum does not mean making the price of patents right. Too many distortions affect this price. The true objective is to reach the right levels of research effort and capital accumulation.

The correct method is to characterize the optimum and the equilibrium in terms of quantities and, in particular, in terms of research effort.

Substituting R for ρ_A, we obtain the following equation

$$(1 - \lambda) \frac{\dot{R}}{R} = \frac{\dot{\theta}_A}{\theta_A} + \theta_q \frac{\gamma - 1}{\gamma} (1 - \alpha) \frac{Y}{K} - \mu$$

$$-\theta_q \theta_A \frac{1-\alpha}{\gamma} (1 + \psi) \frac{Y}{R} \frac{\dot{A}}{A} + (\psi + \phi) \frac{\dot{A}}{A}$$

for the case of the equilibrium and

$$(1-\lambda)\frac{\dot{R}}{R} = (1-\alpha)\frac{Y}{K} - \mu - \beta\lambda\frac{Y}{R}\frac{\dot{A}}{A}$$

for the optimum.

We are now ready to identify the two dynamic systems.

The two following conditions ensure the decentralization of the optimum.

$$\theta_r + \theta_q\frac{\gamma-1}{\gamma}(1-\alpha)\frac{Y}{K} = (1-\alpha)\frac{Y}{K}$$

$$\frac{\dot{\theta}_A}{\theta_A} + \theta_q\frac{\gamma-1}{\gamma}(1-\alpha)\frac{Y}{K} - \theta_q\theta_A\frac{1-\alpha}{\gamma}(1+\psi)\frac{Y}{R}\frac{\dot{A}}{A} + (\psi+\phi)\frac{\dot{A}}{A} = (1-\alpha)\frac{Y}{K} - \beta\lambda\frac{Y}{R}\frac{\dot{A}}{A}$$

We have three instruments for two equalities. Let us fix θ_r. Then

$$\theta_q = \frac{\gamma}{\gamma-1}\,\frac{(1-\alpha)\frac{Y}{K} - \theta_r}{(1-\alpha)\frac{Y}{K}}$$

while θ_A has to satisfy the following relationship.

$$\theta_q\theta_A\frac{1-\alpha}{\gamma}(1+\psi)\frac{Y}{R}\frac{\dot{A}}{A} = \frac{\dot{\theta}_A}{\theta_A} - \theta_r + \left(\psi+\phi+\beta\lambda\frac{Y}{R}\right)\frac{\dot{A}}{A}$$

where Y/K and the other endogenous variables take their optimal values. Note that these formulas hold for the entire trajectory and not only for the long run growth path.

Referências Bibliográficas

Aghion, P. *et al.* (2002), *Competition and Innovation: An Inverted Relationship*, mimeo.

Aghion, P.; Cohen E. (2004), *Education et croissance*, Paris, Rapport pour le Conseil d'Analyse Economique.

Aghion, P. *et al.* (1997), «Competition and growth with Step-by-Step Innovation: an example», *European Economic Review*, 41, 771-782.

Aghion, P., *et al.* (2001), «Competition, Imitation and Growth with Step-by-Step Innovation», *Review of Economic Studies*, 68, 2001, 467-492.

Aghion, P.; Howitt, P. (1992), «A Model of Growth through Creative Destruction», *Econometrica*, 60, 323-351.

D'autume, A. (1998), «Théories anciennes et théories nouvelles de la croissance», in Lassudrie--Duchêne, B. (ed.) *Connaissances économiques, approfondissements*, Paris, Economica.

Encaoua, D.; Ulph, D. (2000), «Catching-up or Leapfrogging? The Effects of Competition on Innovation and Growth», *Cahiers de la MSE*, Université Paris I.

Encaoua, D. *et al.* (2003), «The Economics of Patents: from Natural Rights to Policy Instruments», *Cahiers de la MSE*, Université Paris I.

Grimaud, A. (2002), *Non-Convexities, Imperfect Competition and Growth*, mimeo, Université de Toulouse.

Jones, C. I. (1995), «R&D-Based Models of Economic Growth», *Journal of Political Economy*, 103, 758-784.

Jones, C. I.; Williams, J. C. (2000), «Too Much of a Good Thing? The Economics of Investment», in R&D, *Journal of Economic Growth*, 5, 65-85.

Lucas, R. E. (1988), «On the Mechanics of Economic Development», *Journal of Monetary Economics*, 22, 3-42.

O'donoghue, T.; Zweimller, J. (2004), «Patents in a Model of Endogenous Growth», *Journal of Economic Growth*, 9, 81-123.

Romer, P. (1986), «Increasing Returns and Long-Run Growth», *Journal of Political Economy*, 94, 1002-1037.

Romer, P. (1990), «Endogenous Technical Change», *Journal of Political Economy*, 98, S71-S102.

Scotchmer, S. (1991), «Standing on the Shoulders of Giants», *Journal of Economic Perspectives*, 5, 1.

Tirole, J. *et al.* (2003), *Propriété intellectuelle*, Paris, Rapport pour le Conseil d'Analyse Economique, 2003.

Tournemaine, F. (2004), «Capital humain, financement de la recherche, fécondité et croissance économique», *Thèse de Doctorat*, Université de Toulouse.

Short-term Deviations from Monetary Policy Commitment in a Monetary Union: the Degrees of Freedom of an Independent Central Bank[1]

Muriel Dal-Pont / Dominique Torre / Elise Tosi
GREDEG-CNRS – University of Nice Sophia Antipolis

resumo

Este artigo apresenta um modelo que analisa o grau de autonomia de um Banco central independente, mas comprometidos com determinados objectivos, numa união monetária. Neste modelo, os resultados nominais e reais são determinados pelas interacções entre os diversos agentes, as autoridades políticas supranacionais e o Banco central da união. A informação imperfeita sobre os choques que afectam a oferta, os canais de transmissão e as expectativas de médio prazo podem levar o Banco central a desviar-se dos objectivos propostos. Essa possibilidade de desvio aplica-se a bancos centrais sem condicionantes inflacionárias e empenhados exclusivamente em cumprir um objectivo nominal. Mostra-se que, em determinadas condições, os desvios nominais relativamente aos objectivos propostos não são observáveis pelos agentes nem pela autoridade supranacional que periodicamente selecciona a constituição do Conselho para a Política Monetária do Banco central. Esses desvios aumentam a variação dos valores nominais, mas diminuem as flutuações do rendimento real. Os resultados confirmam obtidos confirmam a posição defendida por Cukierman e Metzler sobre a eficiência de um comportamento ambíguo de um Banco central, neste caso numa união monetária.

résumé / abstract

Cet article présente un modèle qui analyse le degré d'autonomie, dans une union monétaire, d'une Banque centrale indépendante mais engagée. Dans ce modèle, les résultats nominaux et réels sont déterminés par les interactions des différents agents, les autorités politiques supranationales et la Banque centrale. L'imperfection de l'information sur les chocs qui affectent l'offre, les voies de transmission et les anticipations à court terme peuvent mener la Banque à s'écarter des objectifs annoncés. C'est surtout les banques centrales sans contraintes inflationnistes et exclusivement engagées dans la réalisation d'un objectif nominal qui sont susceptibles de présenter cet écart. On montre que dans certaines conditions les écarts nominaux par rapport aux objectifs proposés ne sont pas observables par les agents ni par l'autorité supranationale qui périodiquement sélectionne la composition du Conseil pour la politique monétaire de la Banque centrale. Ces écarts augmentent la variation des valeurs nominales, mais diminuent les fluctuations du revenu réel. Les résultats confirment la position défendue par Cukierman et Metzler sur l'efficience d'un comportement ambigu d'une banque centrale, dans ce cas dans une union monétaire.

The paper presents a model analyzing the degrees of freedom of an independent but committed Central Bank within a monetary union. In the model, interactions between Agents, Supranational Political Authorities and the Central Bank of the Union determine the current nominal and real outcomes. Imperfectly distributed information on shocks affecting supply, transmission channels and short-term expectations create opportunities for a Central Bank to deviate from its announced objective. This opportunity to deviate especially applies to Central Banks free from any kind of inflationary bias and committed to a strictly nominal target. Under certain conditions we show that nominal deviations from stated targets are not observable either by Agents or the Supranational Political Authority that periodically selects the membership of the Council of Monetary Policy of the Bank. Those deviations increase the variance of nominal values but dampen fluctuations of real income. Our results confirm, within a monetary union, the position defended by Cukierman and Metzler concerning the efficiency of a Central Bank's ambiguous behaviour.

1 The authors are grateful to O. Bruno, E. Dharmon, R. Dutu, E. Farvaque, J.F. Goux, J. Hellier, D. Plihon, M.-A. Prunier, C. Bayoud and E. Saltari and an anonymous referee of *Notas Económicas* for their helpful comments on a preliminary version of this paper. The usual disclaimers apply.

Muriel Dal-Pont; Dominique Torre; Elise Tosi

1. Introdução

This paper aims to contribute towards the discussion on the degrees of freedom that an independent and committed Central Bank may benefit from. These *margins of manoeuvre* consist of the Bank's residual opportunities for action. They correspond to a category of discretionary actions that it can undertake which are not really discernable to Agents and hence do not run the risk of the Council of Monetary Policy being sanctioned. Nevertheless, the actions result from the voluntary behaviour of the issuing institution when it decides that under certain circumstances it becomes preferable not to rigidly respect its commitment, either in its own interests or in the interests of the Agents or one of the political authorities. We examine the conditions under which these initiatives emerge. We debate whether they are suited to a process in which the governors of an independent Bank are appointed by the representatives of the nations who form the monetary union. Finally, we evaluate the benefits of such margins of autonomy, from the point of view of the evolution of nominal and real values over time, this being a way to deal with their potential macroeconomic efficiency.

The remainder of this paper is organized as follows: *section 2* defines the objective of the paper in relation to the main results of the literature devoted to independence and commitment. *Section 3* proposes a model of independent Central Bank behaviour within the framework of a monetary union. *Section 4* determines the current level of inflation and output in this context and studies the degrees of freedom the Central Bank has when shocks affect fundamentals and monetary channels. It presents the opportunity for committed Central Banks to deviate from their declared monetary target when some uncertainty is attached to monetary transmission channels. The result illustrates Cukierman-Metzler's (1986) conclusions, recently discussed by Faust and Svensson (2000a and b). It is interesting to note that this incentive to deviate occurs particularly successfully when Central Banks are committed to the long-term inflationary target of the economy (free from inflationary bias).

Section 5 deals with the selection of the Council of Monetary Policy by the Supranational Political Authority and ends with the existence of a simple relationship between the respective levels of conservatism of the Supranational Political Authority and of the Central Bank. *Section 6* examines the macroeconomic impact of cases in which the Central Bank deviates from its target. Short-term deviations from monetary targets result in an increase in the variance of nominal values and in the dampening of real income fluctuations. *Section 7* concludes the analysis.

2. Central Bank independence and commitment in the making of monetary policy

Central Banks play an essential role in short-term economic stabilization. The stability of nominal values being a prerequisite for the development of long-term projects, these banks are in charge of accompanying growth by liberating fiscal and structural policies from nominal contingencies. Price stability is supposed to be best achieved by the independence of Central Banks. "The foundation of this assumption rests on the context of intertemporal games between the government and Agents, the delegation of monetary policy to an external agency more committed than Agents to price control being considered the optimal means of avoiding the time--inconsistency of announced inflationary policies" (Bagella and Becchetti, 1998). The Bank's mandate can be understood in terms of a *principal-agent* approach (M. J. Fratianni, J. von Hagen and C. Waller, 1997).

When Central Banks are primarily interested in achieving nominal targets they tend to be much more credible than in any other context. Since Barro – Gordon's works, it has been admitted that independent Central Banks are the ones most able to adopt a form of behaviour they cannot deviate from, thus ensuring their *credibility*. Exploiting and justifying their independence, the Central Banks are ready to engage in the implementation of a permanent fight against inflation.

The commitment of Central Banks to long-term or medium-term objectives has been considered for some time as "superior to discretion, which can involve a stabilization bias" (Evans and Honkapohja, 2002, p. 2). As in the case of the ECB, commitment to a strict monetary goal explicitly defined by the Bank would avoid any inflation bias or time inconsistency which would, in turn, lead to immediate and very serious sanctions on the part of the Agents. Making a commitment and respecting it have therefore been considered two essential aspects of a Central Banker's attitude whenever the Central Bank pursues a nominal objective.

This point of view, which stems from widely acknowledged theoretical results, is being increasingly challenged by literature as well as Central Bank practices. An independent Central Bank deviating temporarily, and without revealing this, from its initial commitment could benefit from greater flexibility than a government directly administering a monetary policy. This thesis seems to be admitted by the ECB itself: "the solution of an independent Central Bank allows a greater flexibility than the adoption of a strictly mechanical rule" (Issing, 2000). A. Muscatelli and C. Trecroci (2000) also deal with monetary policy rules and show that the literature generally overestimates the importance of independence and assumptions of commitment in the selection of a rule aiming at fighting inflation.

One way – amongst others – to challenge the constant meeting of Bank commitments is the discussion on the respective benefits of *transparency* and *ambiguity* in relation to Central Banks. A Central bank is transparent if it can be easily verified that it tries to achieve what it declares to be its goal. The *transparency* of Central Bank actions is often considered a guarantee of its credibility. When credibility is subordinated to transparency, the Central Bank must not only respect its final objectives but must also reveal its intermediary objectives, therefore sacrificing any short-term appreciation of the opportunity for monetary intervention. By casting a certain level of *ambiguity* over its short-term reactions, it can, on the other hand, exploit degrees of freedom and adapt its intervention to certain phases of economic activity (Cukierman and Meltzer, 1986). The theoretical literature devoted to the determination of optimal behaviour in the field of monetary policy reviews these notions of transparency, ambiguity and inertia levels applied to Central Bank behaviour (Faust et Svensson, 1999, 2000). The opacity of Central Bank behaviour may result from "the absence of clear knowledge about the 'true' model of the economy and sometimes is due to the attempt of policymakers to hedge their positions in the face of model and of political uncertainties" (Cukierman, 2002, b, p. 16). Recent advances in literature devoted to transparency and optimality in monetary policy do not provide a definitive answer in terms of the social desirability of total transparency (Cukierman, 2001, Jensen, 2001, Geraats, 2003).

Without direct connection to this strand of analysis, S.C.W. Eijffinger and M. Hoeberichts (1998) have argued that independence and conservatism are no longer joint attributes of the Central Banks, but substitutes. With greater independence, the Bank would be exempt from adopting an attitude as conservative as that of an issuing institution depending on the government. The Bank would be allowed deviations from strictly conservative principles (the achievement of a pure monetary target) if its decisions resulted from its appreciation of the current state of the economy and from its disinterested perception of monetary or real degrees of flexibility (see Berger, De Haan and Eijffinger, 2001). Whilst opportunities for deviation may be economically desirable (see Denicolò, 1998), committed and transparent Central Banks cannot support them. At this stage of the analysis, the choice seems to be between a committed, conservative, transparent and inflation-bias free Central Bank and a non-credible, accommodative and temporally inconsistent one.

The following sections attempt to explore the conditions under which it would be possible to escape, at least temporarily, from this dilemma, so that Central Banks are independent, subject to commitment inside a monetary union, openly conservative, but tacitly attempting to dampen the short-term real shocks. The degree of freedom an independent Central Bank possesses is then the joint consequence of imperfectly distributed information on the transmission channels of monetary policy and the adequate counter-cyclical short-term actions of the monetary authorities.

3. The model

The economy is composed of several identically-sized countries forming a monetary union. The current decisions of the Agents living in this area contribute towards determining the global income. Every country periodically chooses its political personnel. National elections are uniformly distributed through time and the voting figures are identical across member-states. A Supranational Political Authority composed of representatives of the different national governments periodically appoints the Monetary Policy Council. This council is actually the decision-making body of the union's Central Bank; its mandate is more extensive than that of the national governments. The choice of the monetary authority is made on the basis of its ability to accomplish the long-term objectives of the Political Authority. As a counterpart to the Bank's independence which avoids any time inconsistency in monetary policy, the Bank is required to be committed to reaching an inflation target. Non-compliance with this commitment is sanctioned by non-renewal of the mandate of the members of the Monetary Policy Council. This Council has its own objectives which it seeks to achieve within the constraint that its actions must allow the mandate of its members to be renewed.

3.1. Agents

Agents have a nominal reference $\bar{\Pi}$ known to the political and monetary authorities and representing the desirable long-term inflation rate. In the short term, this reference is corrected by a random term ε_t whose expected value is zero and whose variance is constant σ_ε^2, revealing a shock on preferences that is not observed by the monetary authorities:

$$\Pi_t^* = \bar{\Pi} + \varepsilon_t \qquad (1)$$

With this expression, we assume that the inflation reference of Agents is permanently but stationarily adjusted according to the macroeconomic context and outcomes.

During each period, the Agents determine the real income Y_t of the area, as a result of their decisions concerning production and consumption. This income has a structural component Y^N, which could be interpreted as its "natural" level. The gap between Y_t and Y^N depends on three terms. The first is a linear function of the monetary surprise $[\Pi_t - \tilde{\Pi}_t]$, where Π_t represents the effective level of inflation and $\tilde{\Pi}_t$ is the short-term expectation previously formulated by Agents. This term corresponds to the transitory monetary illusion generated by non-expected inflation. The second term integrates the discrepancy $[\Pi_t^* - \Pi_t]$ between the short-term inflation reference and the effective inflation rate. It takes into account the level of the Agents' 'aversion to inflation' and could be considered a static proxy of the cost expressed in terms of output generated by time-inconsistent expected monetary policies. Finally, a non-observed supply shock takes the form of the random variable u_t whose expected value is zero and which has a constant variance σ_u^2.

$$Y_t = Y^N + \lambda[\Pi_t - \tilde{\Pi}_t] + \mu[\Pi_t^* - \Pi_t] + u_t \qquad (2)$$

with $\lambda > 0$ and $\mu > 0$.

3.2. The Supranational Political Authority

The long-term inflationary reference of the Supranational Political Authority is the rate $\bar{\Pi}$, that is, the long-term Agents' reference. Its short-term reference Π_{st}^* is based on this same rate $\bar{\Pi}$ permanently corrected by an inflation bias A and by a random term v_t whose expected value is zero and whose variance is σ_v^2. The inflation bias introduces an element of time-inconsistency into the political choices; it comes from the pre-electoral context in which a fraction of the members of the political authority are to be found at any given time.

The random term accounts for the impact of exogenous economic or political factors affecting the union in the short term.

$$\Pi^*_{st} = \bar{\Pi} + A + v_t \qquad (3)$$

The supranational authority has mixed objectives that can take the form of a loss function whose expected value is minimized whenever a new Monetary Policy Council is chosen. This function is expressed as follows at any period:

$$L_{st} = \frac{1}{2}\left[(Y_t - Y^N)^2 + \alpha'\beta'(\Pi_t - \Pi^*_{st})^2 + \alpha'(1-\beta')\left(\Pi_t - \bar{\Pi}\right)^2\right] \qquad (4)$$

Its quadratic form is related to the increasing sensitivity of the political authority in relation to the discrepancies between current and reference values. Its real term is standard. Its nominal term is double: on one hand the political authority is sensitive to the difference between current inflation Π_t and its short-term reference Π^*_{st}; on the other hand it is also sensitive to the gap between current inflation and the long-term reference $\bar{\Pi}$. The first of both nominal terms results from the aggregation of the individual behaviour of the union's members: it accounts for the renewal objectives of each national political authority, usually inducing an inflationary bias. The second term takes into account the non–renewal threat through the future election of dissenting political authorities, should inflation diverge too far from the long-term Agent reference. Non–individualistic behaviour or the search for a common interest by the members of political authority may be relevant attitudes in relation to avoiding individual sanctions. It is an economic counterpart of the fact that the national political parties tend to be considered the accountants of the union's performance as soon as they take on responsibilities and agree to remain within the union, regardless of any reservations they have previously expressed – or they would express in opposition. The coefficient α' becomes higher as the share of the nominal target gains in importance; the coefficients β' and $(1 - \beta')$, where β' belongs to the interval [0, 1] determine the relative weight of the two components of nominal values.

The Supranational Political Authority is supposed to have no means of controlling inflation and income in the short term except by contributing to non-expected shocks in preferences and supply. Its unique role is to periodically appoint the Monetary Policy Council of the Central Bank, which is in charge of monetary policy.

3.3 The Central Bank of the Monetary Union

The Central Bank uses different monetary and non-monetary tools (adjustment of key interest rates, management of minimum reserves, open market interventions, announcements and comments on the current situation…) to carry out its monetary policy. Each of these tools has effects that are not fully predictable due to the complexity of the monetary transmission channels and the variability of opinions. The Central Bank has a short-term inflation reference Π^*_{cbt} communicated at the moment the (new) Monetary Policy Council is officially placed in charge of monetary policy. It also takes into account the long-term reference of the economy (that is the expected value of the short--term reference of the Agents which is not observed by the Bank) and the short-term reference of the Supranational Political Authority, the Bank being unable to observe the random component v_t. We assume that the Central Bank corrects the long-term reference of the economy, $\bar{\Pi}$ by a fraction of the expected difference between the Agent and the political authorities' reference:

$$\Pi^*_{cbt} = \bar{\Pi} + \theta\,(E(\Pi^*_{st}) - \bar{\Pi}) \qquad (5)$$

The particular case $\theta = 0$ corresponds to a Central Bank announcing a constant inflation rate targeting (like the early ECB).

The case $\theta = 1$ corresponds to a Central Bank smoothing the political authorities' target. It will be verified that the inflationary bias is then fully transmitted: this would be the sign of a weak of effective independence on the part of the monetary authorities.

In the intermediate case, the Central Bank transmits only a fraction of the bias.

Taking into account its observational abilities and its expertise, the Central Bank observes the supply shock u_t and it chooses Π_{cbt}, the inflation level for the period. However, the Bank does not control the transmission channels of monetary policy perfectly because of the existence of endogenous components in the money creation process (cf. Dal–Pont, Torre and Tosi, 2000) or other kinds of uncertainty regarding the determination of the price level (see Geraats, 2001, Gerdesmeier, Motto and Pill, 2002). The effective inflation is then equal to the level targeted by the Bank Π_{cbt}, incremented by a random term δ_t whose expected value is zero and whose variance is σ^2_δ.

$$\Pi_t = \Pi_{cbt} + \delta_t \qquad (6)$$

The Central Bank does not observe δ_t but knows its distribution. Agents and political authorities know of the existence of imperfections relative to the transmission channels; they know that the expected value of δ_t is zero, but they cannot observe σ^2_δ as they do not observe Π_{cbt}. Consequently, the Bank has the opportunity to periodically target an inflation level Π_{cbt} that is different from Π^*_{cbt} without revealing its transitory deviation from its long-term target. In relation to this targeted inflation rate, the Bank can then anticipate an effective inflation level given by the expected value of Π_t that is Π_{cbt}. It deduces from this an expectation $E_{cbt}(Y_t)$ of the current level of income that in particular takes into account its own observation of the level of inflation expected by the Agents, the expected value of their short-term inflationary reference and the Bank's observation of the supply shock:

$$E_{cbt}(Y_t) = Y^N + \lambda[\Pi_{cbt} - \tilde{\Pi}_t] + \mu[E(\Pi^*_t) - \Pi_{cbt}] + u_t$$

hence,

$$E_{cbt}(Y_t) = Y^N + \lambda[\Pi_{cbt} - \tilde{\Pi}_t] + \mu[\bar{\Pi} - \Pi_{cbt}] + u_t \qquad (7)$$

Like the political authorities, the Central Bank has a quadratic loss function given as (8):

$$L_{st} = \frac{1}{2}\left[(E_{cbt}(Y_t) - Y^N)^2 + \alpha(\Pi_{cbt} - \Pi^*_{cbt})^2\right] \qquad (8)$$

This loss function includes a real term, meaning that the monetary authorities are not only concerned about stabilizing inflation but also dampening the fluctuations of real output. This real component of Central Bank behaviour accounts implicitly for intertemporal smoothing gains for the Agents. We do not introduce any explicit welfare relation which would describe this preference of agents for a certain level of output stability with precision. Even if it exists, the function (stable or not) describing this kind of preference is not observed by the Central Bank which restricts itself to considering stabilizing output as one aspect of its duty. More precisely, this function considers the gap between the level of the expected income $E_{cbt}(Y_t)$ and its natural level Y^N. The Bank has a single nominal indicator: it is sensitive to the discrepancy between its short-term effective and announced targets Π_{cbt} and Π^*_{cbt}. The coefficient α ($\alpha > 0$) measures the degree of conservatism of the monetary authority.

The interpretation of this function is as follows: for the authority, deviating from the announced target may give rise to sanctions but it can do so if the aim is the reduction of a discrepancy that is too large in relation to the real target. Agents do not know the form of the loss function of the Bank, given that the Bank is committed to respecting a nominal target. Under appropriate conditions, Agents would impute the transitory deviations from nominal announced target to imperfections in transmission channels. When Π_{cbt} and Π^*_{cbt} are distinct, the difference is termed γ_t:

$$\gamma_t = \Pi_{cbt} - \Pi^*_{cbt} \qquad (9)$$

Agents do not observe the deviation of the targeted inflation from its announced level but only see the global difference between the effective rate of inflation and the announced target (10):

$$\Pi_t - \Pi^*_{cbt} = \delta_t + \gamma_t \qquad (10)$$

4. Prices and Income in the short-term

We introduce normalized notation, then present the sequence of short-term events. Finally, we determine the real and nominal outcomes for the short-term period.

4.1. Normalized variables and the reduced form of the Model

For the sake of simplicity, we define variables in normalized values using long-term references of inflation and income as normalized references. These new variables can be restated as follows:

• *Supranational Political Authority*

$\pi^*_{st} = \Pi^*_{st} - \bar{\Pi}$, as the short-term reference for the Supranational Political Authority

• *Agents*

$\pi^*_t = \Pi^*_t - \bar{\Pi}$, as the short-term inflation reference for Agents

$\tilde{\pi}_t = \tilde{\Pi}_t - \bar{\Pi}$, as the short-term inflation level expected by Agents

• *Central Bank*

$\pi^*_{cbt} = \Pi^*_{cbt} - \bar{\Pi}$, as the official (subject to commitment) level of inflation for Central Bank

$\tilde{y}_{cbt} = E(Y_{cbt}) - Y^N$, as the real current income expected by Central Bank

$\pi_{cbt} = \Pi_{cbt} - \bar{\Pi}$, as the effective short-term level of inflation targeted by Central Bank

• *Short term outcomes for the economy*

$\pi_t = \Pi_t - \bar{\Pi}$, as the current level of inflation

$y_t = Y_t - Y^N$, as the level of current real income

- *The Model in its reduced form*

In reduced form, using normalized variables, the system (1) – (10) is summed up in equations (11) to (16):

$$y_t = \lambda[\pi_t - \tilde{\pi}_t] + \mu[\epsilon_t - \pi_t] + u_t \tag{11}$$

$$L_{cbt} = \frac{1}{2}\left[(\tilde{y}_{cbt})^2 + \alpha(\pi_{cbt} - \theta A)^2\right] \tag{12}$$

$$\tilde{y}_{cbt} = \lambda[\pi_{cbt} - \tilde{\pi}_t] - \mu\pi_{cbt} + u_t \tag{13}$$

$$\gamma_t = \pi_{cbt} - \theta A \tag{14}$$

$$\pi_t = \pi_{cbt} + \delta_t \tag{15}$$

$$L_t = \frac{1}{2}\left[(y_t)^2 + \alpha'\,\beta'(\pi_t - A - \nu_t)^2 + \alpha'\,(1-\beta')(\pi_t)^2\right] \tag{16}$$

4.2. The sequence of short-term events

In the short term, the sequence of events is as follows:

1. Agents determine their short-term expectations for inflation $\tilde{\pi}_t$, which becomes public information while their short-term inflationary reference (their opinion about what would currently be the "normal rate of inflation") depends on non-revealed preference shocks ϵ_t.

2. The Central Bank observes the Agents' Inflationary Expectations (publicly recorded in sample surveys) and the supply shock u_t but not the periodic shock on Agent preferences ϵ_t. Through minimization of (12), the Central Bank determines the current rate of inflation π_{cbt} that it aims to reach using (13) to account for real retroactions. This rate of inflation may differ from the announced target π^*_{st}. From (14) the deviation γ_t can be deduced from the effective nominal target to the target subject to Central Bank commitment.

3. The impact of the shock δ_t on the transmission channels determines the current rate of inflation π_t according to (15). The current income y_t is obtained from (11) by the integration of the impact of the shocks on preferences ϵ_t. Because Supranational Political Authorities have no influence on monetary policy in the short term, the "macroeconomic shock" ν_t and the loss function (16) do not interfere with the determination of the current nominal and real values.

4.3. Short-term Outcomes and Discussion

As a result, we can evaluate π_t the deviation between the current level of inflation and the long-term inflationary reference and y_t, the deviation of the current income from its natural level:

$$\pi_t = m\alpha\theta A + m(\lambda-\mu)\lambda\tilde{\pi}_t - m(\lambda-\mu)u_t + \delta_t \tag{17}$$

$$y_t = m(\lambda-\mu)\alpha\theta A - m\alpha\lambda\tilde{\pi}_t + m\alpha u_t + (\lambda-\mu)\delta_t + \mu\epsilon_t \tag{18}$$

with $m = \frac{1}{(\lambda - \mu)^2 + \alpha} > 0$

It can be observed that the difference between the effective inflation rate and the Agent long-term reference depends on four factors: the inflationary bias of the political authorities, Agent expectations, supply-side shocks and imperfections in the transmission channels of monetary policy. The inflationary bias and the failures of the transmission mechanisms act in the expected direction. However, the impact of Agent expectations and the shock on the supply-side depend on the relative weight of the "monetary surprise" and the "aversion to inflation" effect.

The more Agent behaviour depends on the monetary illusion phenomena, the more inflation expectations will be self-fulfilling and the supply shocks will be deflationary. The inverse relationship can be observed when aversion to inflation dominates the monetary illusion phenomena.

The gap between short-term income and its natural level depends on the same factors but also on preference shocks[2]. The causality between income variations and their determinants is logical with intuition. We once again confirm the influence of the relative weight of the two components of Agent reaction to prices in determining the sign of the relations between the inflationary bias and income on the one hand, and between shocks on the propagation channels and income on the other hand.

The observed level of inflation comes from the adoption by the Central Bank of an effective target π_{cbt} generally differing from the announced target π^*_{cbt}. Then, we obtain:

$$\pi_{cbt} = m\alpha\theta A + m(\lambda - \mu)\lambda\tilde{\pi}_t - m(\lambda - \mu)u_t \qquad (19)$$

that is a target differing from π^*_{cbt} with,

$$\pi^*_{cbt} = \theta A$$

As a special case, one can determine the nominal and real magnitudes associated with rational expectations. Instead of $\tilde{\pi}_t$, we therefore consider the expected value of the short-term rate of inflation, given the observation associated with the action of the same components of the Monetary Policy Council. This rationally expected inflation rate π^*_t results from the following expression:

$$\pi^*_t = E(m\alpha\theta A + m(\lambda - \mu)\lambda\pi^*_t - m(\lambda - \mu)u_t + \delta_t)$$

that is,

$$\pi^*_t = \frac{m\alpha\theta A}{1 - m(\lambda - \mu)\lambda}$$

In this case, the current nominal and real magnitudes depend only on the inflationary bias and on the shocks on supply, preferences and monetary transmission channels. From (17), one can for instance, deduce the current rate of inflation when expectations are rational:

2 A positive preference shock is associated with a weaker Agent reaction sanctioning an inflation gap (cf. equation 7). This positive shock then results temporarily in a higher level of real output.

$$\pi_t = \frac{m\alpha\theta A}{1 - m(\lambda - \mu)\lambda} - m(\lambda - \mu)u_t + \delta_t$$

This particular case, as well as the general one given by equations (17) and (18), does not take into account the incentive constraint faced by the Central Bank. Indeed, the Bank cannot choose a target that is different to the one announced, except when the Political Authorities are not able to disclose such a deviation. In the short term, deviations cannot be observed because of the existence of transmission channels and endogenous money, which are both responsible for the imperfect control of inflation. If the long-term average level of inflation is different from its announced target, it must be concluded that the short-term deviations are not only the consequence of zero mean error terms. In this case, it can be deduced that the monetary authority targets at least periodically an unannounced level of inflation. However, the variance of the inflation rate does not provide any information, since Agents do not know it. From equation (17), this long-term average level of inflation can be expressed, which could be considered an adequate way to estimate the nature of the targeted rate of inflation by the monetary authorities since, on average, the shocks on monetary transmission vanish.

$$E(\pi_t) = m\alpha\theta A + m(\lambda - \mu)\lambda E(\tilde{\pi}_t) \qquad (20)$$

If we assume that the long-term expected rate of inflation is equal to the mean of the expected short-term rates of inflation, the expression reduces to:

$$E(\pi_t) = \frac{m\alpha\theta A}{1 - m(\lambda - \mu)\lambda}$$

This expression[3] differs from π^*_{cbt} which reduces to θA. It may be concluded that when the monetary authority transmits a part of the inflationary bias, the Central Bank cannot deviate from its announced target without these deviations being revealed in the long run. Given the value of m, it is, however, possible to determine specific cases of identity between $E(\pi_t)$ and π^*_{cbt}. Conditions (i), (ii) and (iii) define three sufficient conditions for the cancellation of any gap between $E(\pi_t)$ and π^*_{cbt}:

(i) $\lambda = \mu, \forall\, \theta$

(ii) $\mu = 0, \forall\, \theta$

(iii) $\theta = 0, \forall\, \alpha, \lambda, \mu$

In case (i), Agent actions are independent of the rate of inflation; in this extreme situation, monetary policy has no impact in real terms and monetary policy has no reason to exist anymore.

The second case (ii) corresponds to an absence of inflation aversion on the part of Agents (or to the absence of any sanction in real terms for an inflationary policy); in this case, given that the shocks are of zero mean, the only role of the Bank is to accommodate the inflation expectations

3 It should be noted that the expected long-term rate of inflation is in this case the rationally expected short--term rate.

of Agents and supply shocks, both having been previously observed. In real terms no sanctions will be incurred.

The third case (iii) is more interesting (and maybe more realistic). It represents a situation in which the monetary authority is committed to a target assimilated within the long-term reference of the economy. In this case, the Central Bank does not transmit the inflationary bias of the political authorities. With a strict short-term commitment to the target, the long-term inflation reference of Agents thus permits the Central Bank to introduce short-term transitory deviations from its nominal objective. In this case, the deviation from the target is compatible with the minimization of (13). If the autonomy of a Central Bank is defined by its capacity to deviate from the stated objectives, it can, paradoxically, be autonomous whilst committed to a target corresponding to the long-term reference of the economy and incidentally insensitive to the inflationary bias of the monetary authorities.

5. The appointment of the Council of Monetary Policy

The appointment of the Council of Monetary Policy is the outcome of a process in which several teams apply for the forthcoming mandate. Since the political authority does not know the loss function of the candidates, the only way to disclose the future behaviour of the candidates is by asking each team to reveal the expected distribution (mean-value and variance) of the rates of inflation that would result from its own actions[4].

5.1. The Submission of Candidates' Applications

For the sake of simplicity, we will assume that all the shocks on expectations, preferences, supply, inflationary bias and transmission channels are independently distributed. We will restrict the analysis to case (iii) defined above, assuming that the applications are limited to candidates whose short-term inflation target is the long-term reference of the economy. Candidate teams are thus only differentiated by their respective levels of conservatism α, which are not directly observed by the Political Authority. Each of them uses the publicly known exogenous distributions of $\tilde{\pi}_t$, u_t, δ_t, ϵ_t, to inform the Political Authority of their own distributions of nominal and real magnitudes π_t and y_t. In doing so, they implicitly solve our *section 4* problem and express π_t and y_t as functions of their respective level of conservatism α, according to (21) and (22):

$$\pi_t = m\lambda\tilde{\pi}_t - mu_t + \delta_t \qquad (21)$$

$$y_t = -\eta\lambda\tilde{\pi}_t + \eta u_t + \mu\epsilon_t + (\gamma - \mu)\delta_t \qquad (22)$$

with,

$$m = \frac{(\lambda - \mu)\lambda}{(\lambda - \mu)^2 + \alpha} \quad \text{and} \quad \eta = \frac{\alpha}{(\lambda - \mu)^2 + \alpha}$$

5.2. The Appointment of the Board

The Political Authority is only informed of the distribution of π_t and y_t for every applicant team. These magnitudes include the periodic loss function of the Authority whose expected value is minimized, as the choice of Central Banker has long-term consequences:

4 We could also imagine that a numerical simulation or a deeper technical discussion between the teams present and the political authority could enable it to apprehend the distribution of price evolution associated with each team.

$$E(L_t) = \frac{1}{2}\left[E\left((y_t)^2\right) + \alpha'\beta' E\left((\pi_t - \pi^*_{st})^2\right) + \alpha'(1-\beta')E\left((\pi_t)^2\right)\right]$$

Hence,

$$E(L_t) = \frac{1}{2}\left[Var(y_t) + \left(E(y_t)\right)^2 + \alpha'\beta' Var(\pi_t - \pi^*_{st})\right]$$

$$+ \frac{1}{2}\left[\alpha'\beta'\left(E(\pi_t - \pi^*_{st})\right)^2 + \alpha'(1-\beta')Var(\pi_t) + \alpha'(1-\beta')\left(E(\pi_t)\right)^2\right] \qquad (23)$$

The Political Authority knows the distribution of the shocks on the inflation bias v_t, which is a component of $Var(\pi_t - \pi^*_{st})$ and of $E(\pi_t - \pi^*_{st})$. From each distribution of prices and the corresponding distribution of income calculated in (12), the Political Authority is thus able to infer the expected value of its loss-function. The political authority uses this whole set of information to minimize the expected value of (16) explicitly as a function of the distribution of prices corresponding to each candidate team and implicitly as a function of the level of conservatism α of these teams. Consequently, it chooses the team best suited to form the Council of Monetary Policy by implicitly selecting the level of conservatism and minimizing its loss function.

In the case where there is a continuum of candidate teams, each being defined according to their respective level of conservatism, the selected value of α will correspond to the minimization of (23), *i.e.* in the case of non-correlated shocks (*see annex for the details of calculations*),

$$\alpha^* = \lambda^2\alpha'$$

In this case, the relationship defines non-correlated shocks and the relation of proportionality between the respective levels of conservatism of the Political Authority and of the selected team comes from intuition (see *annex 1* for a generalization of the relation between α and α'). Note that whatever the level of correlation between shocks, restricting the applications to candidate teams free from any inflationary bias will exclude all the influence of the Political Authority's inflationary bias A on the optimal level of conservatism of the selected team. Secondly, whatever their respective level of conservatism, the teams applying have no ability to stabilise the consequences of nominal target variations on prices and income, since the shocks on preferences are only observable *ex-post*. This explains why α^* does not depend on μ. Thirdly, given π^*_{st} and the nominal part of (23), the higher α' is, the higher α^* would be, as the optimal degree of conservatism of the Central Bank from the point of view of the Governments. Finally, the level of proportionality from α' to α^* depends on the need to control the effect on observed shocks, i.e. the shocks on expectations[5] and on supply, which explains the role of λ in the proportionality relationship between the two degrees of conservatism.

6. Commitment and Degrees of Freedom for Monetary Stabilization

The *margins of manœuvre* of the monetary authorities in the short term and the fact that they do not feel absolutely committed providing they deviate within a range of values that is not observable, are natural consequences of the existence of a loss function including real terms if the Central Bank's actions are not perfectly observable. Such behaviour will be socially beneficial if, and only if, gains can be identified on a macroeconomic level. From this perspective, a new

5 When expectations are rational, this first term naturally vanishes and the supply shock remains the only short--term perturbation under the Bank's control.

loss function in relation to Agents could be introduced. This function would compare the distribution of nominal and real values associated both with a rigid respect for commitment and with the deviation scenario linked to the imperfection of monetary transmission channels. Alternatively and less precisely, it is possible to compare the descriptors of inflation and income related to the different possible Central Bank actions.

If the Central Bank remains committed, the inflation rate effectively targeted is given by π^*_{st}. This rate is corrected by the monetary policy shock in order to determine the observed inflation rate (24).

$$\pi_t = \theta A + \delta_t \quad (24)$$

Then, the income of the period comes from (25):

$$y_t = (\lambda - \mu)\theta A + (\lambda - \mu)\delta_t - \lambda\tilde{\pi}_t + \mu\epsilon_t + u_t \quad (25)$$

Monetary policy introduces a long-term inflationary bias, detected by the fact that the expected value of the long-term inflation level is not equal to the long-term reference of the economy. Only case (iii) analyzed in section 5 excludes the existence of such a bias: when the Central Bank is not sensitive to the political authorities' bias, (that is $\theta = 0$, this case being assimilated to the European Central Bank's target of 2%), which is the case under analysis. When the Central Bank effectively does not deviate in the short term, we then obtain:

$$\pi_t = \delta_t \quad (26)$$

$$y_t = -\lambda\tilde{\pi}_t + u_t + \mu\epsilon_t + (\lambda - \mu)\delta_t \quad (27)$$

It can be observed that here the expected values of π_t and of y_t are actually equal to zero. Variances are given by the following expressions:

$$Var(\pi_t) = \sigma^2_\delta \quad (28)$$

$$Var(y_t) = \lambda^2 Var(\tilde{\pi}_t) + \sigma^2_u + \mu^2\sigma^2_\varepsilon + (\lambda - \mu)^2\sigma^2_\delta \quad (29)$$

Both variances should be compared to the ones that come from the short-term degree of Central Bank freedom given by equations (21) and (22) in the case where $\theta = 0$. We then obtain:

$$Var(\pi_t) = m^2\lambda^2 Var(\tilde{\pi}_t) + m^2\sigma^2_u + \sigma^2_\delta \quad (30)$$

$$Var(y_t) = \eta^2\lambda^2 Var(\tilde{\pi}_t) + \eta^2\sigma^2_u + \mu^2\sigma^2_\varepsilon + (\lambda - \mu)^2\sigma^2_\delta \quad (31)$$

Referring to the value of η, it can easily be verified that, in general, using some remaining degrees of freedom inside its binding commitment, the Central Bank increases the variance of inflation but decreases the variance of income. Compared to a central institution which rigorously respects its commitment, a Central Bank accurately deviating in the short tern will use monetary policy to partially dampen both the supply shocks and the impact of short term expectations. At the same time it will be able to improve output stabilization.

7. Concluding Comments

The aim of this paper is to examine the degrees of freedom of a Central Bank whose Council of Monetary Policy is periodically appointed by a Supranational Political Authority within the framework of a monetary union. This Council has to be committed from the moment its members are renewed and this commitment is verifiable through the results of the Bank's actions. However, the existence of imperfections that are not fully observable, associated with monetary policy channels allows the Bank to deviate periodically from its target without revealing that it does not effectively respect its commitment. The result obtained by A. Cukierman and A. H. Meltzer (1986) and recently discussed by J. Faust et L. E. O. Svensson (2000, a and b) can be extended: in certain cases that are very significant in the context examined here, maximum transparency is not desirable in the realization of the Central Bank's commitment. It is interesting to note that the degree of freedom exploited by the monetary institution may be activated when the announced objective is invariant and assimilated into the long-term reference of the economy. The Bank has all the more the opportunity to deviate from its commitment when this matches Agents' reference and neglects the inflationary bias of the political authorities. This result simultaneously justifies the independence of Central Banks and the existence of an independent Bank commitment free from any inflation bias. This situation is most likely to result in the activation of some kind of short-term "residual discretion" in order to stabilize an economy disturbed by supply shocks and characterized by volatile short-term expectations. In the absence of any commitment, there generally exists a short-term trade-off between inflation and output variability (see for instance Taylor, 1979 or Clarida, Galí and Gertler, 1999). In the case of commitment, any nominal variability in inflation generally proves to be counter-productive in real terms. Our exercise suggests that such a trade-off could equally be restated in this later case, where the actions that Central Bank can undertake concern the residual degree of freedom offered by the existence of noisy transmission channels.

References

Alesina, A.; Summers, L. J. (199, «Central Bank Independence and Macroeconomic Performance: Some Comparative Evidence», *Journal of Money, Credit and Banking*, 25, 2, 151-162.

Athey, S.; Atkeson, A.; Kehoe, P. J. (2003), «The Optimal Degree of Monetary Policy Discretion», *Federal Reserve Bank of Minneapolis, Research Staff Reports*, October.

Bagella m.; Becchetti, L. (1998), «Inflation, Central Bank Independence, Labour and Financial Governance: Some evidence from OECD Countries», in *Institutions and Economic Organization in the Advanced Economies*, edited by M. Baldassari, L. Paganetto and E. Phelps, MacMillan Press LTD.

Banaian, K.; Burdekin, R. C. K.; Willet, T. D. (1995), «On the Political Economy of Central Bank Independence», in K. D. Hoover et S. M. Sheffrin eds., *Monetarism and the Methodology of Economics: Essays in honour of Thomas Mayer*, Aldershott, Edward Elgar, 178-197.

Banaian, K.; Luksetich, W. A. (2001), «Central Bank independence, economic freedom, and inflation», *Economic Inquiry*, 39, 1, 149-161.

Begg, I; Green, D. (1998), «The political economy of the European Central Bank», in P. Arestis et M. Sawyer eds., *The Political Economy of Central Banking*, Edward Elgar Cheltenham, UK.

Berger, H.; Haan, J. de; Eijffinger, S. (2001), «Central Bank Independance: An update of Theory and evidence», *Journal of Economic Survey*, 15, 1.

Blinder, A. (1999), *Central Banking in Theory and Practice*, The MIT Press, Cambridge, Massachussets.

Brociner, A.; Chagny, O. (1996), «La Bundesbank: une orthodoxie pragmatique», *Revue de l'O.F.C.E.*, 56, janvier, 95-106.

Cukierman, A. (1992), *Central Bank Strategy and Independence*, M.I.T. Press, Cambridge Mass.

Cukierman, A. (1994), «Commitment through delegation, political influence and Central Bank independence», in J. De Beaufort Winjholds, S. Eijffinger and L. Hooghduin eds., *A Framework for Monetary Stability*, Kluwer Academic Publishers, Dordrecht, Boston, London.

Cukierman, A. (2001), «Accountability, Credibility, Transparency and Stabilization Policy in the Eurosystem», in C. Wyplosz, ed., *The Impact of EMU on Europe and the Developing Countries*, Oxford: Oxford University Press, 40-75.

Cukierman, A. (2002, a), «Does a Higher Sacrifice Ratio Mean that a Central Bank Independence is Excessive?», *Annals of Economics and Finance*, 3, 1-25.

Cukierman, A. (2002, b), «Are Contemporary Central Banks Transparent About Economic Models and Objectives and What Difference Does It Make?», *Federal Reserve Bank of St Louis, Review*, 84, 4, July-August.

Cukierman, A.; Meltzer, A. H. (1986), «A Theory of Ambiguity, Credibility, and Inflation under Discretion and Asymetric Information», *Econometrica*, 54, 5, 1099-1128.

Deane, M.; Pringle, R. (1994), *The Central Banks*, London: Hamish Hamilton.

Eijffinger, S. C. W.; Hoeberichts, M. (1998), «The trade-off between Central Bank independence and conservativeness», *Oxford Economic Papers,* 50, 397-411.

Evans, G. W.; Honkapohja, S. (2002), Monetary Policy, Expectations and Commitment, *mimeo*, May 27.

Faust, J.; Svensson, L. E. O. (2000, a), «The Equilibrium Degree of Transparency and Control in Monetary Policy», *Journal of Money, Credit and Banking*, December.

Faust, J.; Svensson, L. E. O. (2000, b), «Transparency and Credibility: Monetary Policy with Unobservable Goals», *International Economic Review*, 42, 369-397.

Fratianni, M. J.; Hagen, J. von; Waller, C. (1997), «Central Banking as a Political Principal--Agent Problem», *Economic Inquiry*, 35, 378-393.

Gerdesmeier, D.; Motto, R.; Pill, H. (2002), Paradigm Uncertainty and the role of monetary developments in monetary policy Rules, *mimeo*.

Issing, O. (2000), «Should we have faith in Central Banks», allocution at the *St. Edmund's College Millenium Year Lecture*, Cambridge, 26 October.

Lin, X. (1999), «Central-Bank Independence, Economic Behaviour, and Optimal Term Lengths: Comment», *American Economic Review,* 89, 4, 1056-1062.

Lippi, F. (1998), «On Central Bank Independence and the Stability of Policy Targets», *Scandinavian Journal of Economics*, 100, 2, 495-512.

Lohmann, S. (1992), «Optimal Commitment in Monetary Policy: Credibility versus flexibility», *American Economic Review*, 82, 273-286.

Muscatelli, A.; Trecroci, C. (2000), «Monetary policy rules, policy preferences, and uncertainty: recent empirical evidence», *Journal of Economic Surveys,* 14, 5.

O'flaherty, B. (1990), «The Care and Handling of Monetary authorities», *Economics and Politics*, March, 2, 1, 25-44.

Rogoff, K.; Sibert, A. (1988), «Elections and Macroeconomic Policy Cycles», *Review of Economic Studies*, January, 1-16.

Temple, J. (1998), «Central Bank Independence and Inflation: Good News and Bad News», *Economic Letters,* 61, 215-219.

Tietmeyer, H. (1991), «The Role of an Independent Central Bank in Europe», in P. Downes et R. Vaez-Zadeh eds., *The Evolving Role of Central Banks*, Washington: International Monetary Fund.

Waller, C. J. (1989), «Monetary policy Games and Central Bank politics», *Journal of money, credit and banking*, Nov., 21, 4, 422-432.

Waller, C. J.; Walsh, C. E. (1996), «Central Bank Independence, Economic Behaviour, and Optimal term lengths», *American Economic Review,* 86, 1, March, 150-167.

Annex

Determination of the optimum level of conservatism of the team in charge of the forthcoming Monetary Policy Council

The loss function of Political Authority is given by (23)

$$E(L_t) = \frac{1}{2}\left[Var(y_t) + \left(E(y_t)\right)^2 + \alpha'\beta' Var(\pi_t - \pi^*_{st})\right]$$
$$+ \frac{1}{2}\left[\alpha'\beta'\left(E(\pi_t - \pi^*_{st})\right)^2 + \alpha'(1-\beta')Var(\pi_t) + \alpha'(1-\beta')\left(E(\pi_t)\right)^2\right] \quad (23)$$

When shocks are not correlated, equation (23) reduces to:

$$E(L_t) = \frac{1}{2}\left[\eta^2\lambda^2 Var(\tilde{\pi}) + \eta^2\sigma_u^2 + \mu^2\sigma_\varepsilon^2 + (\lambda-\mu)^2\sigma_\delta^2 + \alpha'\beta' A^2 + \alpha' m^2\lambda^2 Var(\tilde{\pi}) + \alpha' m^2\sigma_u^2 + \alpha'\sigma_\delta^2\right]$$

In this expression, only m and η depend on α.

$$\frac{\partial\left(E(L_t)\right)}{\partial\alpha} = \frac{\partial(\eta^2 + \alpha' m^2)}{\partial\alpha}\left(\lambda^2 Var(\tilde{\pi}) + \sigma_u^2\right) = 0 \text{ i.e.,}$$

$$\alpha = \lambda^2\alpha'$$

The second order condition is:

$$\frac{\partial^2\left(E(L_t)\right)}{\partial\alpha^2}(\alpha) > 0$$

One can easily verify that this condition is satisfied for $\alpha = \lambda^2\alpha'$

When shocks are correlated, equation (23) expands into a more complex expression. Despite an increased number of terms, this expression is still linearly related to m and η, both depending in turn on α. The first order condition then results in a more complex homographic relationship between α and α', defined as following:

$$\alpha = \frac{a\alpha' + b}{c\alpha' + d}$$

with

$$a = \lambda(\lambda-\mu)^4 Cov(\tilde{\pi}, \delta) + \lambda\mu(\lambda-\mu)^3 Cov(\tilde{\pi}, \varepsilon) - (\lambda-\mu)^4 Cov(u, \delta) - \mu(\lambda-\mu)^3 Cov(u, \varepsilon)$$

$$b = 2\lambda^3(1-\lambda) Cov(\tilde{\pi}, u) + \lambda^2(\lambda-\mu)^2 Cov(\tilde{\pi}, \delta) - \lambda(\lambda-\mu)^2 Cov(u, \delta) + \lambda^5\sigma_{\tilde{\pi}}^2 - \lambda^4\sigma_{\tilde{\pi}}^2 + \lambda^3\sigma_u^2 - \lambda^3\mu\sigma_u^2$$

$$c = 2\lambda(\lambda-\mu) Cov(\tilde{\pi}, u) - \lambda(\lambda-\mu)^2 Cov(\tilde{\pi}, \delta) + \lambda\mu(\lambda-\mu) Cov(\tilde{\pi}, \varepsilon) - (\lambda-\mu)^2 Cov(u, \delta) - \mu(\lambda-\mu) Cov(u, \varepsilon) + \lambda^3\sigma_{\tilde{\pi}}^2 - \lambda^2\sigma_{\tilde{\pi}}^2 + \lambda\sigma_u^2 + \mu\sigma_u^2$$

Identification of Monetary Policy Shocks: a Graphical Causal Approach[1]

Alessio Moneta
Laboratory of Economics and Management, Sant'Anna School of Advanced Studies, Italy

resumo

Esse artigo desenvolve uma metodologia VAR estrutural baseada em modelos gráficos para identificar os choques de política monetária e medir os seus efeitos macroeconómicos. A vantagem desse procedimento é trabalhar com modelos sobre-identificados testáveis, cujas restrições são derivadas das correlações parciais entre os resíduos, adicionando-se algum conhecimento institucional. Isso permite testar algumas restrições sobre o mercado de reservas usadas em várias abordagens existentes na literatura. Os principais resultados são que nem as inovações VAR ligadas à *federal funds rate* nem as ligadas às reservas não--emprestáveis (*nonborrowed reserves*) são bons indicadores de choques de política monetária.

résumé / abstract

Cet article élabore une méthodologie des VAR structurels basée sur les modèles de graphes pour identifier les chocs de la politique monétaire et mesurer leurs effets macro-économiques. L'avantage de cette procédure est la possibilité d'utiliser des modèles suridentifiés dont les restrictions sont dérivées par des corrélations partielles des résidus, en plus des connaissances institutionnelles. Ceci permet de tester certaines restrictions relatives au marché des réserves qui ont été utilisées par de nombreuses approches dans la littérature. Les principaux résultats indiquent que ni les innovations VAR introduites sur le taux des fonds fédéraux ni celles introduites sur les réserves non empruntées (*non-borrowed reserves*) sont de bons indicateurs des chocs de la politique monétaire.

This paper develops a structural VAR methodology based on graphical models to identify the monetary policy shocks and to measure their macroeconomic effects. The advantage of this procedure is to work with testable overidentifying models, whose restrictions are derived by the partial correlations among residuals plus some institutional knowledge. This permits to test some restrictions on the reserve market used in several approaches existing in the literature. The main findings are that neither VAR innovations to federal funds rate nor innovations to nonborrowed reserves are good indicators of monetary policy shocks.

1 I am grateful to Marco Lippi and Peter Spirtes for helpful advice, and to an anonymus referee for useful comments. The usual dislaimers apply.

1. Introduction

A monetary policy shock is the portion of variation in central bank policy, that is not caused by the systematic responses to variations in the state of the economy. It is an exogenous shock, which may reflect innovations to the preferences of the members of the monetary authority (e.g. Federal Open Market Committee), measurement errors of the same members, and any other conceivable variation orthogonal to macroeconomic innovations. Vector Autoregressive (VAR) models have been extensively used to isolate and study the effects of a monetary policy shock. A VAR is given by

$$Y_t = A_1 Y_{t-1} + A_2 Y_{t-2} + \ldots + A_p Y_{t-p} + u_t \,,\; t = 1,\ldots,T \qquad (1)$$

where Y_t is a k x 1 vector of data at date $t = 1 - p,\ldots T$; A_i are coefficients matrices of size k x k; and u_t is the one-step ahead prediction error with variance-covariance matrix $\sum_u$. Equation (1) is a "reduced form" model: it merely summarizes the statistical properties of the data. To study the dynamic effects of a monetary policy innovation, one needs an "identified" model, namely a model that has an economic interpretation. The problem of identifying a VAR consists in decomposing the prediction error u_t into economically meaningful or "fundamental" innovations. Suppose that there is a vector ν_t of fundamental innovations of size k x 1, which are mutually independent. Therefore $E[\nu_t \nu'_t] = D$ is a diagonal matrix. What is needed is to find a matrix Γ such that $u_t = \Gamma \nu_t$. It turns out that:

$$\textstyle\sum_u = E[u_t u'_t] = \Gamma E[\nu_t \nu'_t]\Gamma' = \Gamma D \Gamma' \qquad (2)$$

The problem is that the $k(k + 1)/2$ non-zero elements, which can be obtained from the estimate of $\sum_u$, are not sufficient to specify Γ and the diagonal of D. Therefore, one needs further restrictions to achieve identification. In the literature, there exist three methods to impose the necessary restrictions. The first one consists in decomposing $\sum_u$ according to the Choleski factorization, so that $\sum_u = PP'$, where P is lower-triangular, defining a diagonal matrix V with the same diagonal as P and choosing $\Gamma = PV^{-1}$. This is equivalent to impose a recursive ordering of the variables, called a "Wold causal chain", as in Sims (1980). The second method consists in deriving from theoretical and institutional knowledge some "structural" relationships between the fundamental innovations $\nu_{t,i}$, $i = 1,\ldots,k$ and the one-step ahead prediction errors $u_{t,i}$, $i = 1,\ldots,k$, as in Bernanke (1986), Blanchard and Watson (1986) and Sims (1986). The third method consists in separating transitory from permanent components of the innovations, as in Blanchard and Quah (1989) and King et al. (1991).

Any of these methods deals with a high degree of arbitrariness (for a criticism see Faust and Leeper (1997)). Indeed, imposing a set of restrictions corresponds to ascribing a particular causal relation, which is often difficult to be justified, to the variables. To address this problem some authors try several identification schemes and derive stylized facts from them. Thus, Christiano et al (1999) state that "there is considerable agreement about the qualitative effects of a monetary policy shock in the sense that inference is robust across a large subset of the identification schemes that have been considered in the literature. The nature of this agreement is as follows: after a contractionary monetary policy shock, short term interest rate rise, aggregate output, employment, profits and various monetary aggregate fall, the aggregate price responds very slowly, and various measures of wages fall, albeit by very modest amounts".

These conclusions are often considered as "facts" and if a particular identification scheme does not accomplish them, it is sometimes seen as rejectable. Uhlig (1999) has persuasively argued that the way these restrictions are used may render the inference procedure circular ("we just get out what we have stuck in") and proposes to identify the effects of a monetary policy shock on output by directly imposing sign restrictions on the dynamic responses of prices, nonborrowed reserves and interest rate to the same shock. Gordon and Leeper (1994), Bernanke and Mihov (1998) and Bagliano and Favero (1998) emphasize the importance of taking account of different monetary policy regimes.

Another point of controversy is the choice of the indicator of the stance of policy. Bernanke and Blinder (1992) propose VAR innovation to the federal funds rate as measure of the monetary policy shock, basing their argument on prior information about the Fed's operating procedures. However, Christiano and Eichenbaum (1992) have made the case for using the quantity of nonborrowed reserves as indicator of monetary policy. On the other hand, Strongin (1995) argues that the central bank has to accommodate in the short run total reserves demand, therefore monetary policy shocks are the shocks to nonborrowed reserves orthogonal to shocks to total reserves.

Since there is no consensus on which of the various measures is more appropriate to capture the stance of policy, many authors check the robustness of their results using a variety of indicators (see e.g. Christiano et al.,1999). In this paper, in the spirit of Bernanke and Mihov (1998), the indicator of monetary policy stance is not assumed but rather is derived from an estimated model of the Fed's operating procedure. We employ a structural VAR model, but before imposing the restrictions derived from institutional knowledge, we narrow the number of acceptable causal structures using graphical models. The idea is that causal relationships can be inferred from the set of vanishing partial correlations among the variables that constitute such (unobserved) relationships. Graphs form a rigorous language for the "calculus" and representation of causation. This method, which is an extension of the method used in a previous paper (Moneta (2003)) associates a graph to the (unobserved) causal structure of the model and addresses the problem of identification as a problem of causal discovery from vanishing partial correlations. In particular, we infer the class of acceptable causal structures among contemporaneous variables from all the correlations and partial correlations among the residuals.

The goal of this paper is twofold. On the one hand, we want to analyze what are the effects of a monetary policy shock, when in the structural VAR the identification assumptions are derived by means of graphical models, using US macroeconomic and policy variables. The results are consistent with the stylized facts of Christiano et al. (1997) and Christiano et al. (1999), at least for the entire sample 1965-1996. However, the subsample 1979-1996 yields dynamic responses to monetary policy shocks which are qualitatively different from those stylized facts. On the other hand, we want to investigate which shock embeds better the exogenous monetary shock, in the spirit of Bernanke and Mihov (1998). The results cast some doubts on the practice of using the shock to the federal funds or the shock to nonborrowed reserves as a measure of an exogenous monetary policy shock, while they bring some support on the conjecture of Strongin (1995), that a good measure of monetary policy innovation is the shock to nonborrowed reserves orthogonal to the shock to total reserves.

The rest of the paper is structured as follows. The second section describes briefly the identification procedure of the structural VAR. The third section presents a standard model of the market for commercial bank reserves and central bank behaviors, which is a slight extension of the model used by Bernanke and Mihov (1998). The fourth section describes the data and the estimation procedure. The fifth section shows the application of the identification procedure. The sixth section summarizes the main empirical results. Conclusions are drawn in the seventh section.

2. Structural VAR and Graphical Models

Suppose that a structural model of the monetary transmission mechanism can be represented as:

$$\mathbf{A}\begin{pmatrix}\mathbf{X}_t \\ \mathbf{M}_t\end{pmatrix} = \sum_{i=1}^{p} \Gamma_i \begin{pmatrix}\mathbf{X}_{t-i} \\ \mathbf{M}_{t-i}\end{pmatrix} + \mathbf{B}\begin{pmatrix}\nu_t^x \\ \nu_t^M\end{pmatrix}, \tag{3}$$

where $\mathbf{X}_t$ is a vector of macroeconomic (non-policy) variables (e.g. output and prices) and $\mathbf{M}_t$ is a vector of variables (partially) controlled by the monetary policy-maker (e.g. interest rates and monetary aggregates). $\mathbf{Y}_t = (\mathbf{X'}_t\ \mathbf{M'}_t)'$ is a vector of length k, whose components are indicated as $(y_{1t},\ldots, y_{kt})'$ and $\nu_t = (\nu_t^{x'} \nu_t^{M'})'$.

The Structural VAR methodology suggests first to estimate the reduced form:

$$\begin{pmatrix}\mathbf{X}_t \\ \mathbf{M}_t\end{pmatrix} = \sum_{i=1}^{p} C_i \begin{pmatrix}\mathbf{X}_{t-i} \\ \mathbf{M}_{t-i}\end{pmatrix} + \begin{pmatrix}\mathbf{u}_t^x \\ \mathbf{u}_t^M\end{pmatrix}, \tag{4}$$

where $\mathbf{C}_i = \mathbf{A}^{-1}\Gamma_i$ and

$$\mathbf{u}_t = \begin{pmatrix}\mathbf{u}_t^x \\ \mathbf{u}_t^M\end{pmatrix} = \mathbf{A}^{-1}\mathbf{B}\nu_t \tag{5}$$

Then, one has to face the problem of recovering the structural shocks from the estimated residuals. If we call $\sum_u$ the covariance matrix of the estimated residuals, the identification problem in this context consists in inferring A and B from $\sum_u$. The model is overidentified if more than $k + k(k + 1)/2$ restrictions are imposed. In this case the validity of the restrictions can be tested via a likelihood ratio test statistic asymptotically distributed as a χ^2 with a number of degrees of freedom equal to the number of overidentifying restrictions (see Sims (1980), p. 17 and Doan (2000), p. 287).

The idea here is to use graphical models to strongly narrow the number of acceptable contemporaneous causal structures. Then, one can further discriminate using institutional knowledge, jointly with χ^2 tests on overidentifying restrictions. The advantage of this method with respect of the standard structural VAR approach is that eliminating the implausible causal structures significantly lowers the degree of arbitrariness. The method applied here is an extension of a search procedure that was developed in a previous paper (Moneta (2003)) and that we are going to describe briefly here, referring to the appendix for more detailed terminology.

Statistical models represented by graphs, in particular directed acyclic graphs (DAGs), have been proved to be useful to represent causal hypotheses and to encode independence and conditional independence constraints implied by those hypotheses (Pearl (2000), Spirtes et al. (2000), Lauritzen (2001), Lauritzen and Richardson (2002), see appendix for a definition of graph and directed acyclic graph). In this framework, algorithms have been developed to recover some features of the causal graph from (conditional) independence relations among the variables which constitute the unobserved causal structure. A set of algorithms starts from the assumption of direct causation, ruling out the possibility of feedbacks, loops and confounder (e.g. PC algorithm in Spirtes et al. (2000)). A more sophisticated version of it allows for latent variables (e.g. FCI algorithm in Spirtes et al. (2000)). An algorithm developed by Richardson and Spirtes (1999) deals with the problem of inferring features of the causal graph under the assumption that it may be cyclic (feedbacks and loops are allowed), but there are no latent common causes.

Here we apply a general algorithm, which is basically the first common part of the algorithms mentioned above. The algorithm has, as input, the covariance matrix among the VAR residuals and produces, as output, an undirected graph, which represents a class of possible causal relationships among the contemporaneous variables of the VAR model. The algorithm starts connecting all the contemporaneous variables $(y_{1t},\ldots, y_{kt})'$ in a complete graph and progressively eliminates most of the edges among variables which are not associated neither by a causal link, nor by a feedback link, nor by a latent variable, which we interpret as a common shock (for a definition of edges, undirected and complete graph, see appendix).[2]

The procedure is based, first, on the fact that in a VAR model partial correlations among residuals are equivalent to partial correlations among contemporaneous variables, conditioned on all the past variables (see Proposition 1 in appendix).

Second, there are two fundamental assumptions relating causes and probability. Given a graph G and any three vertices *A,B,C* belonging to G: (i) Causal Independence Assumption: if A does not cause B, and B does not cause A, and there is no third variable that causes both A and B, then A and B are independent; (ii) Causal Faithfulness Assumption: if *corr*(*A,B*|*C*) is zero then *A* and *B* are d-separated by C on the graph G. For the definition of d-separation see appendix.

Assuming normality of the error terms, the search algorithm described in appendix permits to infer an undirected graph, which represents a pattern of directed graphs (feedbacks and loops are allowed), from Wald tests on vanishing partial correlations among the residuals.

3. A Model of the Reserve Market

The undirected graph resulting from the search algorithm permits to narrow considerably the class of causal structure, but seldom this is enough for a reliable identification. Background knowledge about the way the central bank sets the monetary policy is very useful at this stage. More detailed institutional assumptions, associated with causal hypotheses, can then be tested.

Following Christiano et al (1999), a general model of the monetary authority behavior can be written as:

$$S_t = f(\Omega_t) + \nu_t^s , \qquad (6)$$

where S_t is the instrument of the monetary authority, say the federal funds rate or some monetary aggregate, f is a linear function that represents the central bank's feedback rule, Ω_t is the monetary authority's information set and ν_t^s is a monetary policy shock.

Bernanke and Mihov (1998) model the banks' total demand for reserves as:

$$TR_t = f_{TR}(\Omega_t^x) - \alpha FFR_t + \nu_t^d , \qquad (7)$$

where Ω_t^x is the information set that comprehends only current and past macroeconomic variables. According to (7), the demand for total reserves TR_t depends on Ω_t^x and is affected negatively by the federal funds rate (FFR_t). The demand for borrowed reserves is:

$$BR_t = f_{BR}(\Omega_t^x) + \beta(FFR_t - DISC_t) - \gamma NBR_t + \nu_t^b , \qquad (8)$$

where BR_t is the portion of reserves that banks choose to borrow at the discount window. According to (8), BR_t is affected positively by the federal funds rate – discount rate differential

2 There are some (recognizable) cases in which an edge in the output of the algorithm does not correspond to the presence of an edge in the causal graph representing the data generating process (for details see Moneta (2004), p. 43). So, the output graph may contain more edges than the unobserved "true" graph. In general, the algorithm gives a graph that represents a class of causal structures, not a unique causal structure. It is possible to show that if there is a "true" causal structure which has generated the data, such causal structure is included in the class of causal structures represented by the output graph of the algorithm (see Moneta (2004) for details).

and negatively by the nonborrowed reserves (NBR_t). Bernanke and Mihov (1998) assume that innovation to the discount rate is zero, which means that fluctuations in the discount rate are completely anticipated, so that $DISC_t$ does not enter in (8).

As far as the parameter γ is concerned, Christiano et al. (1999) give two reasons for including NBR_t in equation (8). The first one is that, if we would be willing to include expected values in the equation describing demand for BR_t, we should include expected values of FFR_t among the variables affecting BR_t, because of the existence of nonprice costs of borrowing at the Federal Reserve discount window. (These costs rise for banks that use too much or too frequently the discount window). Indeed, for example, when $E_t(FFR_t)_{t+1}$ is high, banks want BR_t to be low so that they can take full advantage of the high expected federal funds rate in next period without having to suffer large nonprice penalties at the discount window. Since NBR_t helps forecast future values of FFR_t, it should enter on equation (8).

The second reason is that a bank that possesses a large amount of NBR_t and is using the discount window is simply trying to profit from the spread between the federal funds rate and discount rate. In that case, the bank would suffer a higher nonprice marginal cost of borrowing. So, NBR_t should enter the equation describing demand for BR_t.

However, Bernanke; Mihov (1998) assume $\gamma = 0$, presumably in order to achieve overidentified and testable identification scheme.[3]

The following equation describes the behavior of the Federal Reserve:

$$NBR_t = f_{NBR}(\Omega_t^x) + \phi^d \nu_t^d + \phi^b \nu_t^b + \nu_t^s \qquad (9)$$

According to (9), the Fed, by means of open-market operations, can change the amount of nonborrowed reserves supplied to banks in response to shocks to the total demand for reserves and to the demand for borrowed reserves. The coefficients ϕ^d and ϕ^b denote the strength of the responses and ν_t^s represents the monetary policy shock.

Since $TR_t = NBR_t + BR_t$, we can derive from (8) (omitting the discount rate) the following equation:

$$FFR_t = -\frac{1}{\beta} f_{BR}(\Omega_t^x) + \frac{1}{\beta} TR_t + \frac{\gamma - 1}{\beta} NBR_t - \frac{1}{\beta} \nu_t^b \qquad (10)$$

From (7), (9) and (10), we can derive restrictions on the contemporaneous variables, which correspond to zero coefficients on the matrix **A** of equations (3) and (5), as we will see in section 5.

4. Data and Estimation

The data set used is the same as that of Bernanke and Mihov (1998) and consists of 6 series of monthly US data (1965:1-1996:12)[4]. We refer to the non-policy macroeconomic variables as GDP_t: real GDP (log); $PGDP_t$: GDP deflator (log); $PSCCOM_t$: Dow-Jones index of spot commodity prices (log). The policy variables are: TR_t: total bank reserves (normalized by 36-month moving average of total reserve); NBR_t: nonborrowed reserves plus extended credit (idem); FFR_t: federal funds rate.

3 This is an important limitation, as underlined by Christiano et al. (1999). Indeed, in the Bernanke and Mihov (1998) approach, one can always interpret an alleged rejection of an identification scheme as evidence against the maintained hypothesis $\gamma = 0$ and save the identification scheme. An advantage of our approach is that, thanks to the pre-selection of graphical models, we do not need to assume $\gamma = 0$ to reach overidentification and we can assess whether $\gamma = 0$ is in fact rejected or not by the data.

4 The data set was downloaded from Ben Bernanke's home page.

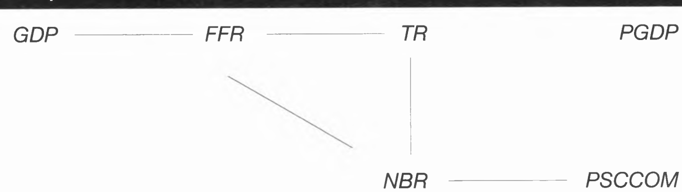
We estimate the model both in the vector error correction model parameterization and in levels (equation-by equation OLS)[5]. Since the results of the two estimations are very close, we report the results of the level estimation, in order to have a clear comparison with the results of Christiano et al. (1999) and Bernanke and Mihov (1998), who estimate the model in level. The number of lags used to estimate the VAR is 13 in the full sample (the same used by Bernanke and Mihov (1998). The covariance matrix among the residuals obtained by OLS estimation is the following:

$$\tilde{\Sigma}_u = \begin{pmatrix} 322 & 4 & -31 & -4 & 3473 & 42 \\ 4 & 26 & 16 & 15 & 650 & 26 \\ -31 & 16 & 1682 & 660 & -19341 & -388 \\ -4 & 15 & 660 & 802 & 7652 & -46 \\ 3473 & 650 & -19341 & 7652 & 2236763 & 10530 \\ 42 & 26 & -388 & -46 & 10530 & 2670 \end{pmatrix} \times 10^{-7}$$

5. Identification of the Structural Shocks

The search algorithm mentioned in section 2 and reported in appendix is employed to derive the class of admissible causal structures among the contemporaneous variables of the structural model (represented in equation (3)). The input of the search algorithm is the covariance matrix among residuals and the output is a pattern of directed graphs, which is represented by an undirected graph. Figure 1 displays the output of the algorithm for the full sample.[6]

Figure 1 – Contemporaneous structure for the full sample (time subscrips are removed for convenience)



The graph in Figure 1 has to be read according to the following criterion. An undirected edge between any two vertices *A* and *B* of the graph corresponds to one or more of the following alternatives: (i) there is a direct causal relationship from *A* to *B*; (ii) there is a direct causal relationship from *B* to *A*; (iii) there is a common shock affecting both *A* and *B*.[7] It should be emphasized that a lack of edges between any two variables does not mean that there is no correlation at all (in fact there is usually correlation through lagged variables affecting both), but just that all of the three options listed above are excluded. Thus, since in the graph of Figure 1 there is no edge starting from PGDP, prices (measured by the GDP deflator) do not affect

5 The estimation via VECM parameterization does not imply any difference in the way the identification problem is faced, since, once the covariance matrix among the residuals is estimated, the model is reconverted in levels.

6 The significance level employed in the vanishing partial correlation tests is 0.05. We remove time subscripts. A variable without time subscripts should be interpreted as a variable at time *t*.

7 Actually, there is also a fourth possibility. If there are feedbacks and common shocks in the data generating graph, the output of the algorithm may contain edges that do not correspond to any of the associations mentioned. Such configurations are however recognizable, as mentioned in footnote 1 (see Moneta (2004), p. 43 for more details).

instantaneously any other variable, prices are not affected instantaneously by any other variable, and that there is no common shock affecting contemporaneously PGDP and any other variable.

From the mere analysis of correlations and partial correlations it is very difficult to infer which structure, among (i), (ii) and (iii), holds. One needs to incorporate background knowledge, which, if it implies overidentifying constraints, can be tested.

One set of assumptions can be derived by prior knowledge about the nature of interaction between policy variables and macroeconomic non-policy variables. A common interpretation of a contemporaneous association between macroeconomic variables and policy variables is that the monetary authority monitors continuously prices and output, so that there are causal effects from non-policy macroeconomic variables to policy variables within the period (one month). A further possible identifying assumption (used e.g. in Bernanke and Mihov (1998)) is that there is no feedback from policy variables to the economy within the period.[8] This corresponds to ruling out causal relationships from any of the policy variables *NBR*, *BR* and *FFR* to any of the macroeconomic variables *GDP*, *PGDP* and *PSCCOM*. In the following we will consider and assess the hypothesis of orthogonality of the policy shock to the macroeconomic variables, which we call recursiveness assumption, against the alternative hypothesis of correlation between policy shock and these variables (non-recursiveness assumption).[9] Under the recursiveness assumption, the undirected edges of Figure 1 between *GDP* and *FFR* and between *NBR* and *PSCCOM* are interpreted as directed edges from *GDP* and *PSCCOM* to *FFR* and *NBR* respectively. Under the non-recursiveness assumption the same edges are interpreted as bi--directed.[10]

Institutional knowledge can be used to impose identifying restrictions concerning the interactions among the policy variables (*NBR*, *TR* and *FFR*). From the considerations of section 3 about the model of the reserve market, it results that the relationships between VAR residuals and structural disturbances can be represented as follows, as far as the monetary policy block is concerned:

$$u_{TR} = -\alpha u_{FFR} + \nu^d \qquad (11)$$

$$u_{FFR} = \frac{1}{\beta} u_{TR} + \frac{\gamma - 1}{\beta} u_{NBR} - \frac{1}{\beta} \nu^b \qquad (12)$$

---

8 The length of "the period" is crucial here. For example, the assumption of no feedback from policy variables to the economy is more difficult to defend at the quarterly frequency and easier to defend at the weekly frequency. The opposite occurs with the assumption of causal effects from the economy to policy variables, which is more reliable at low than at high frequencies. Notice also that we do not use, consistently with the studies quoted, real-time data, and that measurement errors, which are common in the first data releases, are embedded in the exogenous monetary shock. How the identification results would change with the use of real-time data is an interesting open research question.

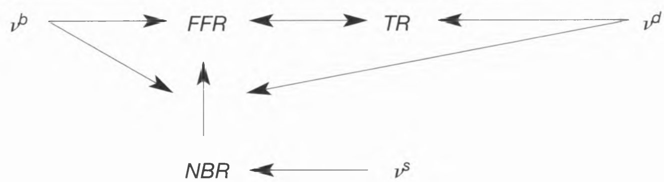
9 The scheme of identification associated with the recursiveness assumption should in general be distinguished from the recursive VAR identification scheme, which is derived by the Choleski factorization of the residuals covariance matrix and is associated with a Wold causal chain.

10 Indeed, it is possible to show that under the Faithfulness condition, the recursiveness assumption implies an absence of contemporaneous direct causes from non-policy to policy variables and an absence of a direct cause from the policy shock to non-policy variables (and an absence of any latent variable caused by the policy shock and causing non-policy variables). Moreover, under the Causal Independence condition, the non--recursiveness assumption implies that either policy variables cause non-policy variables within the period, or that the policy shock is a common shock affecting both non-policy and policy variables (or that there is a latent variable affected by policy variables and affecting non-policy variables). To put it in another way, the economic content of the recursiveness assumption is that non-policy variables do not respond within the period to realization of the policy shock, while the economic content of the non-recursiveness assumption is just the opposite.

$$u_{NBR} = \phi^d \nu^d + \phi^b \nu^b + \nu^s \quad (13)$$

The system of equations (11)-(12)-(13) corresponds to a set of causal restrictions, as illustrated below. The restrictions on the relationships among macroeconomic variables (*GDP*, *PGDP* and *PSCCOM*) and on the relationships between macroeconomic variables and policy variables, that are derived by the graph output of the search algorithm, are numerous enough, so that the system can be identified. We also consider further restrictions on the policy block, which correspond to five alternative models, the same considered by Bernanke and Mihov (1998), with the difference that we allow γ to be different from zero.

Figure 2 – Model 0. It is the model of equations (11), (12) and (13), without further restrictions



The first case we consider is the model of equations (11), (12) and (13), without further restrictions, which we call Model 0. The graph connected to this model is represented in Figure 2. In this case the monetary policy shock ν^s is related with the VAR residuals in the following way:

$$\nu^s = -(\phi^d + \phi^b)u_{TR} + (1 + \phi^b(1-\gamma))u_{NBR} - (\alpha\phi^d - \beta\phi^b)u_{FFR}\,, \quad (14)$$

which is obtained solving (11)-(12)-(13) for ν^s.

The second case is Model 0 plus the restriction $\alpha = 0$. The graph for this model, which we call model $\alpha = 0$, is represented in Figure 3. It corresponds to assuming that the demand for total reserves is inelastic in the short run. Strongin (1995) presents institutional arguments to support this assumption.

In the third case we impose to Model 0 the restrictions $\phi^d = \frac{1}{1-\gamma}$ and $\phi^b = -\frac{1}{1-\gamma}$. This corresponds to the assumption that the central bank uses NBR to neutralize borrowing and demand shocks and targets the federal funds rate. Indeed the monetary policy shock turns out to be proportional to the innovation to the federal funds rate:

$$\nu^s = -\frac{1}{1-\gamma}(\alpha + \beta)u_{FFR} \quad (15)$$

The graph related with this model, which we call model *FFR*, is represented in Figure 4. Bernanke and Blinder (1992) presented arguments in support of the federal funds rate as a measure of policy instrument.

In the next case the following restrictions are imposed: $\phi^d = 0$ and $\phi^b = 0$. In this case the monetary policy shock coincides with the VAR innovation to the nonborrowed reserves: $\nu^s = u_{NBR}$. The graph related with this model, which we call model *NBR*, is represented in Figure 5. The argument that innovations to nonborrowed reserves primarily reflect shocks to monetary policy was defended by Christiano and Eichenbaum (1995) and Christiano et al. (1996).

The fifth case we consider is the model which imposes the restrictions $\alpha = 0$ and $\phi^b = 0$ on Model 0. The implied monetary policy shock is $\nu^s = -\phi^d u_{TR} + u_{NBR}$. This corresponds to assuming that shocks to total reserves are purely demand shocks (ν^d), which the central bank has to accommodate immediately (either through open-market operations or the discount window). Therefore monetary policy shocks (ν^s) are the shocks to NBR orthogonal to ν^d. Moreover, this specification, defended by Strongin (1995), does not consider the possibility that the central bank reacts to borrowing shocks. Figure 6 represents the graph associated with this model, which is called Model *NBR/TR*.

The last case we consider is the model which corresponds to assuming that the central bank targets borrowed reserves. The restrictions imposed are $\phi^d = \frac{\beta}{\beta + \alpha\gamma}$ and $\phi^b = \frac{\beta}{\beta + \alpha\gamma}$. This implies that

$$\nu^s = -\frac{\alpha + \beta}{\beta + \alpha\gamma}(u_{TR} - u_{NBR}) = -\frac{\alpha + \beta}{\beta + \alpha\gamma} u_{BR} \qquad (16)$$

Figure 7 represents the graph associated with this model, which is called Model *BR*.

Figure 3 – Model α = 0. The demand for total reserves is inelastic in the short run, so that there is no causal effect from *FFR* to *TR*

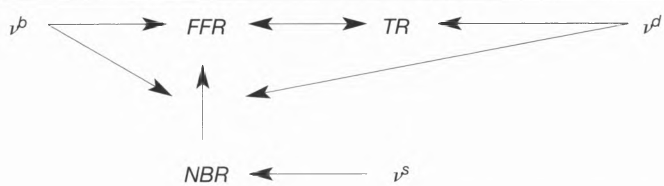


Figure 4 – Model *FFR*. The weights refer to the case γ = 0. The Fed fully offsets shocks to total reserves demand and borrowing demand and targets the federal fund rate

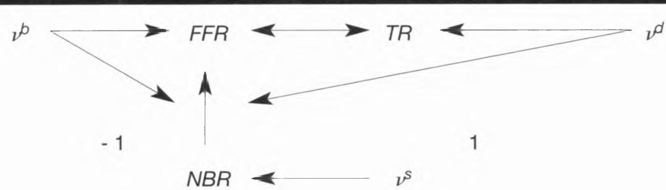


Figure 5 – Model *NBR*. Nonborrowed reserves respond only to policy shocks, so that borrowing and demand shocks do not affect *NBR*

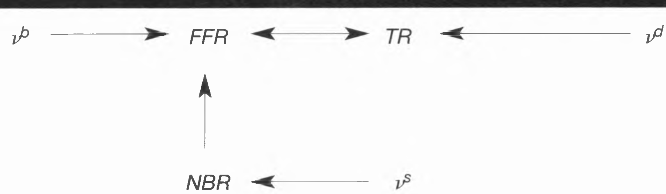


Figure 6 – Model *NBR/TR*. Monetary policy shocks are shocks to *NBR* orthogonal to demand shocks

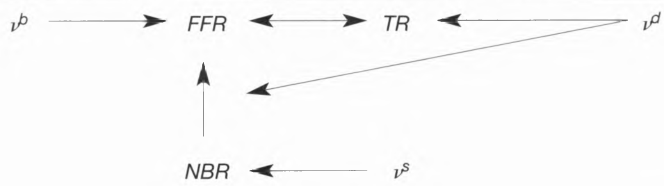
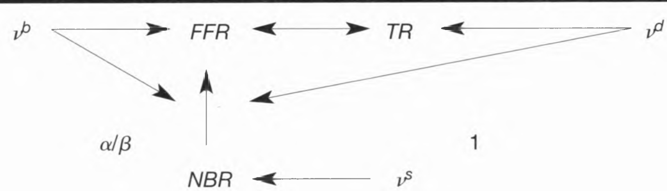


Figure 7 – Model *BR*. The weights refer to the case $\gamma = 0$. The Fed targets borrowed reserves (*TR-NBR*)



The set of restrictions implied by each model corresponds to a set of restrictions on the matrices **A** and **B** of equation (3). Equation (5) can be written as:

$$\boldsymbol{A}\boldsymbol{u}_t = \mathbf{B}\nu_t \qquad (17)$$

Imposing the restrictions of Model 0 (without recursiveness), we can write (17) as:[11]

$$\begin{pmatrix} 1 & 0 & 0 & 0 & \alpha_{15} & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & \alpha_{36} \\ 0 & 0 & 0 & 1 & \alpha_{45} & 0 \\ \alpha_{51} & 0 & \alpha_{53} & \alpha_{54} & 1 & 0 \\ 0 & 0 & \alpha_{63} & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} u_{GDP} \\ u_{PGDP} \\ u_{NBR} \\ u_{TR} \\ u_{FFR} \\ u_{PSCCOM} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & b_{34} & b_{35} & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & \alpha_{54} & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \nu_{GDP} \\ \nu_{PGDP} \\ \nu^s \\ \nu^d \\ \nu^b \\ \nu_{PSCCOM} \end{pmatrix} \qquad (18)$$

11 Notice that the theoretical restrictions that we impose do not include interactions with non-policy variables. The only restrictions about such interactions are derived from the graph-search procedure (with the exception of the general restriction embedded in the recursiveness assumption). This is because, first, we have more reliable and precise background knowledge about policy variables, than about the relations between policy and non-policy variables. Indeed, we may call the theoretical knowledge about policy variables "institutional knowledge", because is based more on assumptions about the procedures followed by the banking system, than on economic theory. Second, each set of restrictions on the policy variable comprises a precise interpretation of the exogenous monetary policy shock. Since we can easily test each set of restrictions, we can get information as to which measure represents better the exogenous monetary shock.

The restrictions on the elements of **A** and **B** for each model of the bank reserves are reported in Table 1. The relations among the parameters of equation (18) and α, β, γ, ϕ^d and ϕ^b are the following: $\phi^d = b_{34}$, $\phi^b = b_{35}$, $\alpha = \alpha_{45}$, $\beta = -\frac{1}{\alpha_{54}}$, $\gamma = 1+\frac{\alpha_{53}}{\alpha_{54}}$. The key to interpret equation (18) is looking at equation (3), where matrix **A** represents the structural relations among contemporaneous variables.

Each model is estimated by maximum likelihood estimation, using a RATS procedure based on the hill-climbing BFGS method (see Doan (2000) for details).[12]

Estimates of the parameters α, β, γ, ϕ^d and ϕ^b for each model are reported in Table 4. Each model is overidentified and produces a likelihood ratio test for the restrictions. In the same table *p* values for such tests are also reported, that indicate whether a model has been rejected or not.

We do not have space here to give specific comments on the estimates of γ, α, β, ϕ^d and ϕ^b (for a detailed analysis see Moneta (2004)). The substance of these results is reported in the next section.

We have also investigated the robustness of the results across subsamples. We do not have space here to report the results (Moneta (2004) contains a wider illustration of the results and the method to deal with small samples), whose substance is reported in the next section.

12 The results of the restrictions of Model *FFR* and Model *BR* with the non-recursiveness assumption should be taken with caution, because they do not take into account policy parameters that enter in the equation of the monetary policy shock via non-policy variables. This does not change, however, the substance of the results (see next section).I thank an anonymous referee for having raised this issue. Estimates of the parameters of model 0 are displayed in Tables 2 and 3.

Table 1 – Restrictions on A and B. Each of the six models considered implies zero restrictions on some elements of the matrices A and B of equation (17). Each model has four versions, depending on the recursiveness assumption and the assumption on γ (in one case γ is free, in the other is zero)

	recursiveness	rec. + $\gamma = 0$	non-recursiveness	non-rec. + $\gamma = 0$
Model 0	$\alpha_{15} = 0$ $\alpha_{63} = 0$	$\alpha_{15} = 0$ $\alpha_{63} = 0$ $\alpha_{53} = -\alpha_{54}$		$\alpha_{53} = -\alpha_{54}$
Model $\alpha = 0$	$\alpha_{15} = 0$ $\alpha_{63} = 0$ $\alpha_{45} = 0$	$\alpha_{15} = 0$ $\alpha_{45} = 0$ $\alpha_{63} = 0$ $\alpha_{53} = -\alpha_{54}$	$\alpha_{45} = 0$	$\alpha_{53} = -\alpha_{54}$ $\alpha_{45} = 0$
Model *FFR*	$\alpha_{15} = 0$ $\alpha_{63} = 0$ $b_{34} = -\alpha_{54} / \alpha_{53}$ $b_{35} = \alpha_{54} / \alpha_{53}$	$\alpha_{15} = 0$ $\alpha_{53} = -\alpha_{54}$ $\alpha_{63} = 0$ $b_{34} = 1$ $b_{35} = -1$	$b_{34} = -\alpha_{54} / \alpha_{53}$ $b_{35} = \alpha_{54} / \alpha_{53}$	$\alpha_{53} = -\alpha_{54}$ $b_{34} = -\alpha_{54} / \alpha_{53}$ $b_{35} = \alpha_{54} / \alpha_{53}$
Model *NBR*	$\alpha_{15} = 0$ $\alpha_{63} = 0$ $b_{34} = 0$ $b_{35} = 0$	$\alpha_{15} = 0$ $\alpha_{53} = -\alpha_{54}$ $\alpha_{63} = 0$ $b_{34} = 0$ $b_{35} = 0$	$b_{34} = 0$ $b_{35} = 0$	$\alpha_{53} = -\alpha_{54}$ $b_{34} = 0$ $b_{35} = 0$
Model *NBR/TR*	$\alpha_{15} = 0$ $\alpha_{63} = 0$ $b_{45} = 0$ $b_{35} = 0$	$\alpha_{15} = 0$ $\alpha_{53} = -\alpha_{54}$ $\alpha_{63} = 0$ $b_{45} = 0$ $b_{35} = 0$	$\alpha_{45} = 0$ $b_{35} = 0$	$\alpha_{53} = -\alpha_{54}$ $\alpha_{45} = 0$ $b_{35} = 0$
Model *BR*	$\alpha_{15} = 0$ $\alpha_{63} = 0$ $b_{34} = 1/$ $(1-\alpha_{45}(\alpha_{53}+\alpha_{54}))$ $b_{35} = (\alpha_{45}\alpha_{54})/$ $(\alpha_{45}(\alpha_{53}+\alpha_{54})-1)$	$\alpha_{15} = 0$ $\alpha_{53} = -\alpha_{54}$ $\alpha_{63} = 0$ $b_{34} = 1$ $b_{35} = \alpha_{45}\alpha_{54}$	$b_{34} = 1/$ $(1-\alpha_{45}(\alpha_{53}+\alpha_{54}))$ $b_{35} = (\alpha_{45}\alpha_{54})/$ $(\alpha_{45}(\alpha_{53}+\alpha_{54})-1)$	$\alpha_{53} = -\alpha_{54}$ $b_{34} = 1$ $b_{35} = \alpha_{45}\alpha_{54}$

Table 2 – Estimation Model 0 recursive (full sample):
ML estimation by BFGS. Convergence in 84 Iterations
Observations 371. Log Likelihood 9614.0525
Log Likelihood Unrestricted 9620.7327
Chi-Square(8) 13.605. Significance Level 0.1000

Parameter	Coeff	Std Error	T-Stat	Signif
α_{36}	0.1360	0.0386	3.5241	0.0004
α_{45}	0.0013	0.0056	0.2411	0.8094
α_{51}	-8.7989	3.6417	-3.4161	0.0156
α_{53}	30.6270	10.0566	3.0454	0.0023
α_{54}	-37.0140	13.2387	-2.7959	0.0051
b_{34}	0.7608	0.2393	3.1791	0.0014
b_{35}	-0.1810	0.2790	-0.6490	0.5163

Table 3 – Estimation Model 0 *non*-recursive (full sample):
ML estimation by BFGS. Convergence in 233 Iterations
Observations 371. Log Likelihood 9614.7285
Log Likelihood Unrestricted 9620.7327
Chi-Square(6) 12.0084. Significance Level 0.0617

Parameter	Coeff	Std Error	T-Stat	Signif
α_{15}	-0.0021	0.0007	-2.9976	0.0027
α_{36}	0.0000	0.0035	0.0017	0.9986
α_{45}	-0.0180	0.0019	-9.0292	0.0000
α_{51}	0.0007	0.4462	0.0017	0.9985
α_{53}	68.9252	2.1444	32.1408	0.0000
α_{54}	-55.4529	2.6953	-20.5736	0.0000
α_{63}	0.2305	0.0658	3.5036	0.0004
b_{34}	0.8046	0.0697	11.5274	0.0000
b_{35}	-0.8045	0.0504	-15.9625	0.0000

Table 4 – Parameters Estimates (full sample):

Model	α	β	γ	ϕ^d	ϕ^b	p-value
Recursive assmption:						
Model 0	0.0014	0.0270	0.1726	0.7609	-0.1811	0.1000
$\alpha = 0$	0	0.0290	0.1216	0.8148	-0.1939	0.1444
FFR						0.0118
NBR						0.0253
NBR/TR	0	0.0358	0.2019	0.8148	0	0.1659
BR	-0.0051	0.0532	-0.2510	0.9767	-0.0929	0.1189
Model 0, $\gamma = 0$	-0.0022	0.0326	0	0.8861	-0.2220	0.1282
$\alpha = 0$, $\gamma = 0$						0.0002
FFR, $\gamma = 0$						0.0000
NBR, $\gamma = 0$						0.0000
NBR/TR, $\gamma = 0$						0.0450
BR, $\gamma = 0$	-0.0035	0.0436	0	1.0000	-0.0794	0.1223
Non-Recursive						
Model 0	-0.0180	0.0180	-0.2429	0.8046	-0.8045	0.0617
$\alpha = 0$	0	0.0270	0.0977	0.8165	-0.2716	0.0903
FFR						0.0428
NBR						0.0435
NBR/TR	0	0.0358	0.1999	0.8165	0	0.1070
BR	-0.0055	0.0552	-0.3175	0.9696	-0.0957	0.0952
Model 0, $\gamma = 0$	-0.0269	0.0112	0	0.5927	-0.9691	0.1013
$\alpha = 0$, $\gamma = 0$	0	0.0209	0	0.8054	-0.5945	0.1109
FFR, $\gamma = 0$	-0.0036	0.0119	0	1.0000	-1.000	0.0868
NBR, $\gamma = 0$						0.0000
NBR/TR, $\gamma = 0$						0.0261
BR, $\gamma = 0$	-0.0035	0.0433	0	1.000	-0.0805	0.0890

Notes: The estimates are functions of the ML estimates of the coefficients of the matrices **A** and **B** of equation (17), obtained by the RATS procedure based on the BFGS method (see Doan (2000) for details). The last column gives *p*-values from likelihood ratio tests of overidentifying restrictions. If *p*-value > 0.05, the restrictions implied by the particular model cannot be rejected at the 5 percent level of significance. We do not report estimates of the models which have been rejected. Models that have been not rejected and the corresponding *p*-values are displayed in bold text.

6. Main Results

Our analysis permits to give some answers to the following questions.

What happens after a monetary policy shock?

If we consider the full sample 1965-1996, the qualitative responses of output, prices and interest rate are consistent with the stylized facts presented by Christiano et al. (1999) and with the results of Bernanke and Mihov (1998). After an expansionary monetary policy shock, output has an uncertain behavior in the first 2-3 months, then it increases rapidly, reaching its peak around the 15th month. The response of output in the long run is almost null, that means that money is close to being neutral in the long run.[13] Price level responds slowly in the first year, after that increases. Short term interest rate falls immediately (showing the so called "liquidity effect"), but after 10-12 months returns to its previous value. Impulse response functions of GDP, GDP deflator and federal funds rate to a monetary policy shock are displayed in Figures 8-9 for the full sample. The impulse response functions are calculated for those models which have passed the

13 This does not mean that the equation, in which *GDP* is dependent variable, is stationary (indeed it contains a unit root, according to the standard tests).

likelihood ratio test (whose results are displayed in Table 4). The results about the effects of a monetary policy shock are quite robust across different assumptions about the central banks operating procedures and are approximately repeated in the sub-sample 1965:1-1979:9. However, in the sub-sample 1979:10-1996:12 we obtain slightly different results. Output still rises after a monetary policy shock, but much more moderately. Price levels responds positively, expecially at the beginning, but very slowly. The impulse response functions for the two sub-samples are displayed in the Figures 10-11.

Which indicator most accurately captures the monetary policy shock?

Generally speaking, neither VAR innovation to the federal funds rate, nor VAR innovation to nonborrowed reserves turns out to be good indicator of monetary policy shocks. This is in the spirit of some results of Strongin (1995), Thornton (2001) and Sarno et al. (2002). Bernanke and Blinder (1992) employ innovations to interest rate as indicators of monetary policy shocks and obtain results consistent with the stylized facts (output and money rise in response to a positive monetary policy shock). Table 4, however, shows that the different specifications of *FFR* model always fail the likelihood ratio test (with one exception). Indeed there are some problems in measuring monetary policy with federal funds rate. First, as argued by Strongin (1995), "without any demonstrated empirical linkage between Federal Reserve actions and interest rate movements, it is unclear how innovations in interest rates can be reasonably be attributed to monetary policy." Second, there could be non-policy omitted variables which explain movements in interest rates. Third, Sarno et al. (2002) argue that the practice of identifying monetary policy shocks as shocks to federal funds rate should be taken with caution, because of the "information equivalence hypothesis" (all interest rates contain the same information about monetary policy and the other aggregate shocks that determine the state of the economy).

Christiano and Eichenbaum (1992) suggest that innovation in nonborrowed reserves is the correct measure of monetary policy. Analogously to what happens with the federal funds rate, Table 4 shows that the different specifications of *NBR* model are always rejected by the data. This result corroborates the argument of Strongin (1995) that a significant proportion of the movements in nonborrowed reserves is due to the Fed's accommodation of innovations in the demand for reserves, rather than policy-induced supply innovation. Indeed a good indicator of the monetary policy shock seems to be the measure suggested by Strongin (1995), which is the part of innovation to nonborrowed reserves orthogonal to innovation to total reserves. Table 4 shows that the *NBR/TR* model is rejected by the data only in the case of $\gamma = 0$.[14]

Does the recursiveness assumption hold?

The recursiveness assumption is about the orthogonality of the policy shock to the macroeconomic variables. It implies that policy variables do not influence macroeconomic non--policy variables within the period and that the monetary policy shock is not affecting simultaneously the two sets of variables (ruling out latent variables affected by the policy shock and affecting non-policy variables). We do not obtain strong results about this issue, even though the empirical evidence does not reject the hypothesis of non-recursiveness (see Table 4). The assumption of recursiveness, however, does not bring much difference for the only scope of measuring the effects of monetary policy shocks.

14 Furthermore, the model *NBR/TR* is the only model which is never rejected in the sub-samples. Model *BR* is also not rejected in the full sample, but the estimates of γ obtained are significantly negative (also in the sub--samples). This fact casts doubt on the reliability of this model. These results can be seen in detail in Moneta (2004).

Figure 8 – Responses of GDP, PGDP, and FFR to one-standard-deviation monetary shock for the sample 1965: 1-1996: 12 and with the recursiveness assumption

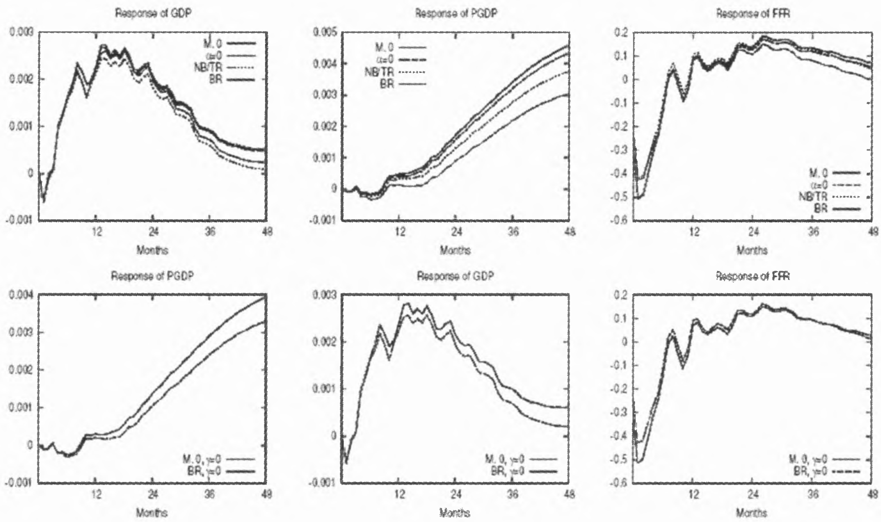


Figure 9 – Responses of GDP, PGDP, and FFR to one-standard-deviation monetary shock for the sample 1965: 1-1996: 12 and *without* the recursiveness assumption

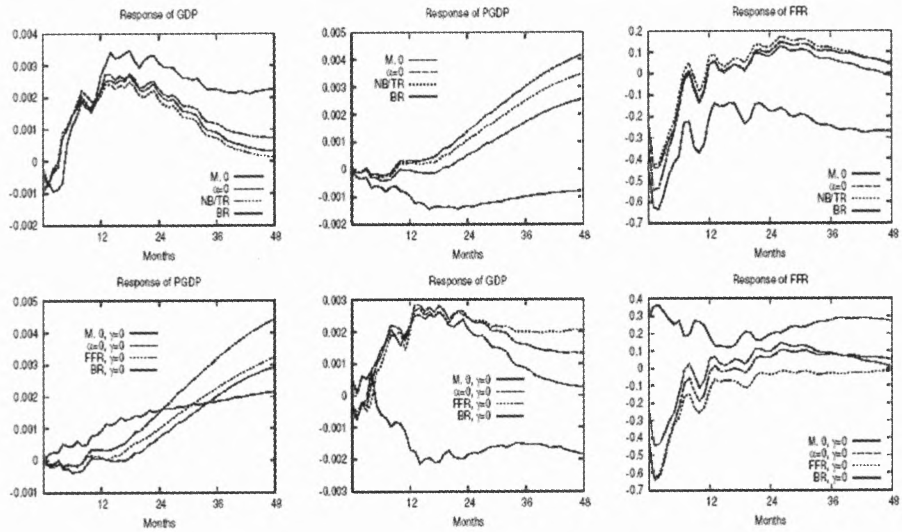


Figure 10 – Sample 1965: 1-1979: 9. The graphics on the first line refer to models identified under the recursiveness assumptions, the graphics on the second line under the non-recursiveness assumption

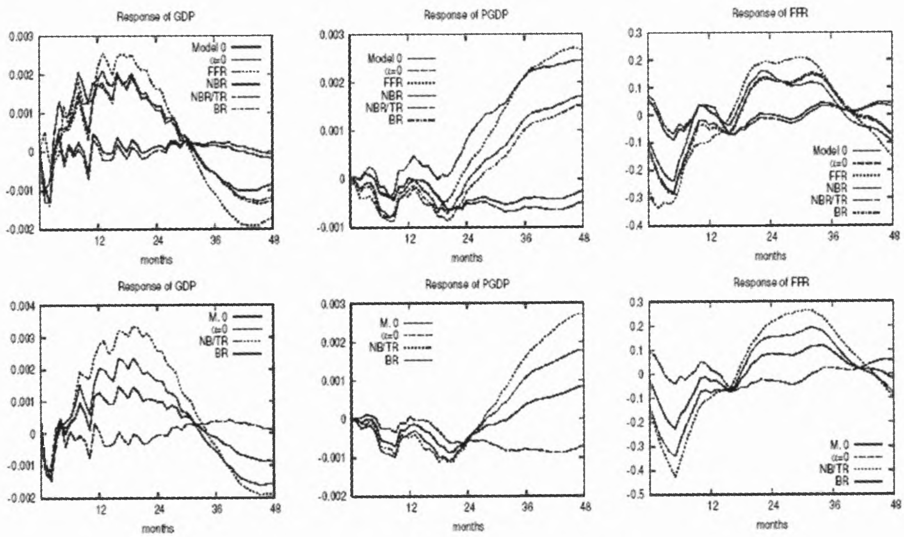
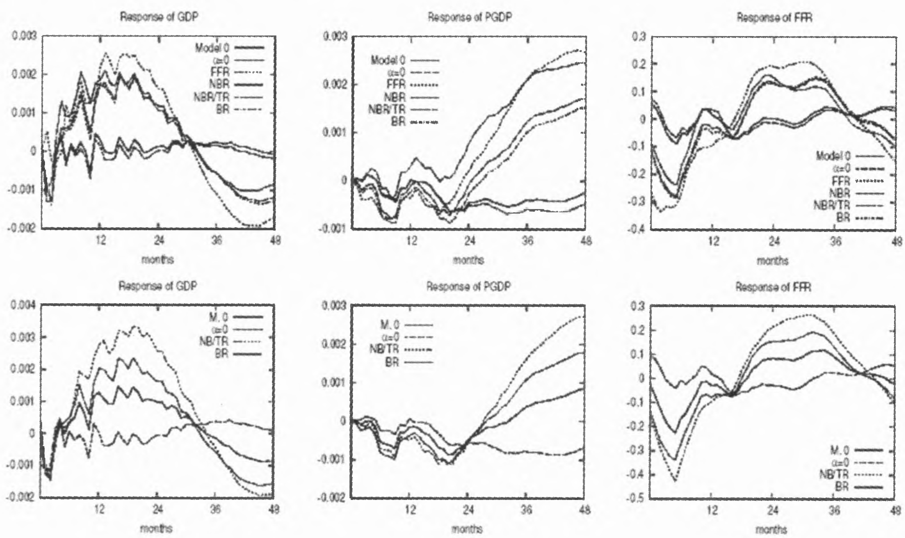


Figure 11 – Sample 1979: 10-1996: 12. The graphics on the first line refer to models identified under the recursiveness assumptions, the graphics on the second line under the non-recursiveness assumption



7. Conclusions

This paper proposed a method to identify the exogenous monetary policy disturbances in a VAR model. Since the crucial issue to identify a VAR is to differentiate between correlation and causation, graphical models permitted to impose overidentifying restrictions on the contemporaneous causal structure, in particular on the relationships among macroeconomic variables and between macroeconomic variables and policy variables. These restrictions have the advantage of being derived from the statistical properties of the data, without using auxiliary assumptions. Once we have narrowed the set of possible contemporaneous causal relationships among the variables which constitute the VAR, we have imposed restrictions derived from institutional and theoretical knowledge. Since the number of possible contemporaneous causal relationships is a finite (and relatively narrow) number, it was possible to check the robustness of our results under alternative schemes of the Fed's operating procedure and to appraise the alternative measures of monetary policy shocks used in the literature. The empirical results cast doubt on the soundness of those researches which assume that VAR innovations to federal reserve rate or nonborrowed reserves are good indicators of exogenous monetary policy shocks.

Appendix: Graphical models terminology

Graphs. A graph is an ordered pair $G = (V, E)$, where V is a nonempty set of vertices, and E is a subset of the set $V \times V$ of ordered pairs of vertices, called the edges of G. If one or both of the ordered pairs (V_1,V_2), (V_2,V_1) belong to E, V_1 and V_2 are said to be adjacent. If both ordered pairs (V_1,V_2) and (V_2,V_1) belong to E, we say that we have an undirected edges between V_1 and V_2, and write $V_1 - V_2$. We also say that V_1 and V_2 are neighbors. If all the edges of a graph are undirected, we say that it is an undirected graph. If (V_1,V_2) belongs to E, but (V_2,V_1) does not belong to E, we call the edge directed, and write $V_1 \rightarrow V_2$. We also say that V_1 is a parent of V_2 and that V_2 is a child of V_1, If all the edges of a graph are directed, we say that it is a directed graph. A path of length n from V_0 to V_n is a sequence $\{V_0,\ldots, V_n\}$ of distinct vertices such that $(V_{i-1},V_i) \in E$ for all $i = 1,\ldots, n$. A directed path is a path such that $(V_{i-1},V_i) \in E$, but $(V_i,V_{i-1}) \notin E$ for all $i = 1,\ldots, n$. A cycle is a directed path with the modification that the first and the last vertex are identical, so that $V_0 = V_n$. A graph is complete if every pair of its vertices are adjacent. A directed acyclic graph (DAG) is a directed graph which contains no cycles. Given a directed graph, the set of the vertices V_i such that there is a directed path from V_i to V_j are the ancestors of V_j and the set of vertices V_i such that there is a directed path from V_j to V_i are the descendants of V_j. The graph $G_A = (A, E_A)$ is called a subgraph of $G = (A, E)$ if $A \subseteq V$ and $E_A \subseteq E \cap (A \times A)$. Besides, if $E_A = E \cap (A \times A)$, G_A is called the subgraph of G induced by the vertex set A.

D-separation. In a directed graph G a vertex X is a collider on a path α if and only if there are two distinct edges on α both containing X and both directed on X. In a directed graph G a vertex X is =1 active on a path β relative to a set of vertices Z of G if and only if: (i) X is not a collider on β and $X \notin Z$; or (ii) X is a collider on β, and X or a descendant of X belongs to Z. A path β is active relative to Z if and only if every vertex on β is active relative to Z. In a directed graph G two vertices X and Y are d-separated by Z if and only if there is no active path between X and Y relative to Z. X and Y are d-connected by Z if and only if X and Y are not d-separated by Z.

Proposition 1: Let $u_{1t},\ldots, u_{kt}$ be the residuals of k OLS regressions of $y_{1t},\ldots y_{kt}$ on the same vector $J_{t-1} = (y_{1(t-1)},\ldots, y_{k(t-1)},\ldots, y_{1(t-p)},\ldots, y_{k(t-p)})$. Let u_{it} and u_{jt} $(i \neq j)$ be any two distinct elements of $\{u_{1t},\ldots, u_{kt}\}$, U_t any subset of $\{u_{1t},\ldots, u_{kt}\} \setminus \{u_{it}, u_{jt}\}$ and Y_t the corresponding subset of $\{y_{1t},\ldots, y_{kt}\} \setminus \{y_{it}, y_{jt}\}$, so that u_{gt} is in U_t iff y_{gt} is in Y_t, for $g = 1,\ldots, k$. Then it holds that:

$$corr(u_{it}, u_{jt} \mid U_t) = corr(y_{it}, y_{jt} \mid Y_t, J_{t-1})$$

Proof in Moneta (2003).

From Proposition 1 and Faithfulness Condition it follows that if $corr(u_{ht}, u_{it} \mid u_{jt},\ldots, u_{lt}) = 0$, then $corr(y_{ht}, y_{it} \mid y_{jt},\ldots, y_{lt}, J_{t-1}) = 0$ (where J_{t-1} is defined as in Proposition 1) and y_{ht} and y_{it} are d-separated by $y_{jt},\ldots, y_{lt}, J_{t-1}$ in the graph induced by $y_{1t},\ldots, y_{kt}, J_{t-1}$. Then it follows quite intuitively (see for a rigorous proof Proposition 2 in Moneta (2003)) that y_{ht} and y_{it} are d-separated by $y_{jt},\ldots, y_{lt}$ in the sub-graph induced by $y_{1t},\ldots, y_{kt}$. The search algorithm is displayed in Table 5.

Table 5 – Search algorithm

A.)

From the estimated covariance matrix among the VAR residuals test all the possible partial correlations among the residuals (using the Wald test procedure described in Moneta (2003)).

B.)

Form the complete undirected graph *C* on the vertex set $y_{1t},\ldots, y_{kt}$. Let *Adjacencies* (C, y_{it}) be the set of vertices adjacent to y_{it} in *C* and let *Sepset* (y_{ht}, y_{it}) be any set of vertices *S* so that y_{ht} and y_{it} are d-separated by *S*;

C.)

$n = 0$

repeat:

repeat:

select an ordered pair of variables y_{ht} and y_{it} that are adjacent in *C* such that *Adjacencies* $(C, y_{ht}) \setminus \{y_{it}\}$ has cardinality greater than or equal to *n*, and a subset *S* of *Adjacencies* $(C, y_{ht}) \setminus \{y_{it}\}$ of cardinality *n*, and if y_{ht} and y_{it} are d-separated by *S* in G_{Yt} delete edge $y_{ht} — y_{it}$ from *C*;

until all ordered pairs of adjacent variables y_{ht} and y_{it} such that Adjacencies$(C, y_{ht}) \setminus \{y_{it}\}$ has cardinality greater than or equal to *n* and all subsets *S* of *Adjacencies* $(C, y_{ht}) \setminus \{y_{it}\}$ of cardinality *n* have been tested for d-separation;

n = n + 1;

until for each ordered pair of adjacent variables y_{ht}, y_{it}, *Adjacencies* $(C, y_{ht}) \setminus \{y_{it}\}$ is of cardinality less than n;

Note: Adapted from common steps of PC-FCI-CCD algorithms of Spirtes *et al.* (2000) and Richardson and Spirtes (1999).

References

Bagliano, F.; Favero, C. (1998), «Measuring Monetary Policy with VAR Models: An Evaluation», *European Economic Review*, 42, 6, 1069-1112.

Bernanke, B. S. (1986), «Alternative Explanations of the Money-Income Correlation», *Carnegie--Rochester Conference Series on Public Policy*, 25, 49-100.

Bernanke, B. S.; Blinder, A. (1992), «The Federal Funds Rate and the Channels of Monetary Transmission», *American Economic Review*, 82, 4, 901-921.

Bernanke, B. S.; Mihov, I. (1998), «Measuring Monetary Policy», *Quarterly Journal of Economics*, 113, 3, 869-902.

Blanchard, O.; Watson, M. (1986), «Are All Business Cycles Alike?», in Gordon, R. J. (ed.), *The American Business Cycle*, Chicago, University of Chicago Press, 123-160.

Blanchard, O.; Quah, D. (1989), «The Dynamic Effects of Aggregate Demand and Supply Disturbances», *American Economic Review*, 79, 4, 655-673.

Christiano, L. J.; Eichenbaum, M. (1992), «Identification and the Liquidity Effect of a Monetary Policy Shock», in Cukierman, A. et al. (eds.), *Political Economy, Growth, and Business Cycles*, Cambridge MA, MIT Press.

Christiano, L. J.; Eichenbaum, M. (1995), «Liquidity Effects, Monetary Policy and the Business Cycle», *Journal of Money, Credit and Banking*, 27, 4.

Christiano, L. J.; Eichenbaum, M.; Evans, C. L. (1996), «The Effects of Monetary Policy Shocks: Evidence form the Flow of Funds», *Review of Economics and Statistics*, 78, 16-34.

Christiano, L. J.; Eichenbaum, M.; Evans, C. L. (1997), «Sticky Price and Limited Partecipation Models of Money: A Comparison», *European Economic Review*, 41, 1201-1249.

Christiano, L. J.; Eichenbaum, M.; Evans, C. L. (1999), «Monetary Policy Shocks: What Have We Learned and To What End?», in Taylor, J.; Woodford, M. (eds.) *Handbook of Macroeconomics*, Volume 1A, Amsterdam, Elsevier Science Ltd., pp. 65-148.

Doan, T. A. (2000), RATS Version 5, User's Guide, Evanston, IL, Estima.

Faust, J.; Leeper, E. M. (1997), «When Do Long-Run Identifying Restrictions Give Reliable Results?», *Journal of Business and Economic Statistics*, 15, 3, 345-353.

Gordon, D.; Leeper, E. M. (1994), «The Dynamic Impacts of Monetary Policy: An Exercise in Tentative Identification», *Journal of Political Economy*, 52, 6, 1228-1247.

King, R. G. *et al.* (1991), «Stochastic Trends and Economic Fluctuations», *American Economic Review*, 81, 4, 819-840.

Lauritzen, S. L. (2001), «Causal Inference from Graphical Models», in Barndorff-Nielsen, E. et al. (eds.), *Complex Stochastic Systems*, London, CRC Press.

Lauritzen, S. L.; Richardson, T. S. (2002), «Chain Graph Models and Their Causal Interpretations», *Journal of the Royal Statistical Society*, 64, 321-361.

Moneta, A. (2003), *Graphical Models for Structural Vector Autoregressions*, LEM Working Paper, Sant'Anna School of Advanced Studies, Pisa.

Moneta, A. (2004), *Graphical Causal Models and VAR-based Macroeconometrics*, unpubblished PhD thesis, Sant'Anna School of Advanced Studies, Pisa.

Pearl, J. (2000), *Causality: Models, Reasoning, and Inference*, Cambridge, Cambridge University Press.

Richardson, T.; Spirtes, P. (1999), «Automated Discovery of Linear Feedback Models», in Glymour C.; Cooper, G.F. (eds.), *Computation, Causation and Discovery*, Menlo Park, CA, AAAI Press and The MIT Press.

Sarno, L. *et al.* (2002), *The Information Equivalence Hypothesis and the Identification of Monetary Policy Shocks: A Cautionary Note*, mimeo.

Sims, C. A. (1980), «Macroeconomics and Reality», *Econometrica*, 48, 1, 1-47.

Sims, C. A. (1986), «Are Forecasting Models Usable for Policy Analysis?», *Minneapolis Federal Reserve Bank Quarterly Review*, Winter, 2-16.

Spirtes, P. *et al.* (2000), *Causation, Prediction, and Search*, Cambridge MA, The MIT Press.

Strongin, S. (1995), «The Identification of Monetary Policy Disturbances: Explaining the Liquidity Puzzle», *Journal of Monetary Economics*, 36, 463-497.

Thornton, D. L. (2001), «The Federal Reserve's Operating Procedure, Nonborrowed Reserves, Borrowed Reserves and the Liquidity Effect», *Journal of Banking and Finance*, 25, 1717-1739.

Uhlig, H. (1999), «What are the Effects of Monetary Policy on Output? Results from an Agnostic Identification Procedure?», *Journal of Monetary Economics*, 35, 463-498.

Régimes Monétaires et Théorie Quantitative du Produit Nominal au Portugal (1854-1998)[1]

João Sousa Andrade
Faculdade de Economia da Universidade de Coimbra

resumo

A economia portuguesa, como aliás outras economias ocidentais, passou por vários regimes monetários desde meados do século XIX. Procura-se saber se a teoria quantitativa da moeda aplicada ao rendimento nominal, e não aos preços, explica o comportamento do rendimento nominal nos diferentes regimes monetários. A análise foca os períodos sugeridos pelos historiadores económicos, períodos que são condensados pelo método de Bai-Perron de forma a identificar rupturas estruturais das séries temporias. Depois de seleccionados os períodos a analisar e conhecendo as características de estacionaridade da oferta monetária e do rendimento nominal, aplica-se o método de co-integração de Joahansen a essas séries. Conclui-se que a teoria quantitativa do rendimento nominal da moeda não se aplica ao regime do pardão-ouro. A confirmação empírica sugere que essa teoria só se aplica num regime de inconvertibilidade da moeda.

résumé / abstract

L'économie portugaise, comme d'ailleurs d'autres économies occidentales, a connu différents régimes monétaires depuis la moitié du XXe siècle. Nous voulons savoir si la théorie quantitative de la monnaie appliquée au revenu nominal, et non aux prix, peut expliquer le comportement du revenu nominal dans différents régimes monétaires. Nous concentrons notre analyse sur les périodes suggérées par les historiens économiques et nous les concentrons en utilisant la méthodologie de Bai-Perron pour l'identification des ruptures structurelles des séries temporelles. Après avoir pris la décision de savoir quelles périodes utiliser et connaissant les caractéristiques de stationnarité de l'offre de monnaie et du revenu nominal, nous appliquons la méthode de co-intégration de Johansen à ces deux séries. Notre conclusion est très intéressante: la théorie quantitative du revenu nominal de la monnaie ne peut être appliquée au régime de l'étalon-or. La confirmation empirique suggère que cette théorie ne peut être appliquée que dans un régime de monnaie inconvertible.

The Portuguese economy like other economies in occident has known some different monetary regimes since the middle of XIX century. We want to know if the quantity theory of money applied not to prices but to nominal income can explain the behaviour of nominal income in different monetary regimes. We concentrate on the periods used by economic historians and condensed them with the Bai-Perron methodology for identifying structural time breaks in time series. After deciding which periods to retain and knowing the stationarity characteristics of the money supply and the nominal income we applied the Johansen co-integration method to these series. Our conclusions are very interesting: the nominal income quantity theory of money can't be applied to the gold standard regime. The empirical evidence suggests its application only to an inconvertible regime.

1 Je voudrais remercier Pedro Bação, Jacques Leonard, Jean-Pierre Berdot et Maria Adelaide Duarte pour leurs commentaires et suggestions. Je tiens aussi à remercier pour ses suggestions et corrections un *referee* de la revue *Notas Económicas*.

Présentation

Par cette étude, nous allons confirmer l'hypothèse d'une élasticité unitaire du produit nominal par rapport à l'offre de monnaie dans les cas des régimes de monnaie convertible. Du point de vue logique, cette hypothèse doit être considérée avant les études sur la théorie quantitative de la monnaie appliquée aux prix. Cette préoccupation concernant la formation de la demande globale nominale, nous la retrouvons dans la tradition monétariste. Cependant, le plus souvent, les discussions sur la neutralité de la monnaie font oublier l'antécédent logique de l'hypothèse de l'élasticité unitaire du produit nominal.

Notre étude sera appliquée à l'économie portugaise depuis 1854 jusqu'en 1998. Pendant cette longue période, l'économie portugaise a connu plusieurs régimes monétaires: aussi devons-nous savoir identifier et classifier les différentes périodes. L'histoire économique ainsi que l'étude statistique de la période doivent être présentées. Ce travail étant accompli, nous devons tester notre hypothèse en tenant compte des techniques économétriques des séries non-stationnaires.

Dans une autre étude (Andrade, 2000), nous avons appliqué à l'ensemble de la période 1854-1998 une méthodologie du type suggérée par Bernanke et Mihov (1998), pour démontrer que l'élasticité du produit nominal par rapport à l'offre de monnaie n'était pas différente de l'unité pendant la période d'inconvertibilité de la monnaie portugaise. La valeur tendancielle des impulsions du produit nominal et de l'offre de monnaie qui résultent d'un choc de l'offre de monnaie, dans le contexte d'un modèle VAR, permettait la confirmation ou exclusion de la neutralité de la monnaie par rapport au revenu nominal. Mais comme l'a suggéré Mankiw (2000), il y a toujours le problème de retenir des effets qui ne sont pas différents de zéro du point de vue statistique.

Notre travail est organisé en 5 sections. Dans la première (I), nous présenterons l'importance de l'étude de la théorie quantitative du produit nominal. Nous tentons d'identifier les différents régimes que l'économie portugaise a connus de 1854 à 1998 en tenant compte de l'histoire des faits économiques portugais (II). Cette analyse sera enrichie par l'application de la méthodologie des ruptures temporelles de Bai-Perron aux données du produit nominal et de l'offre de monnaie et nous conduira à la délimitation des périodes pertinentes pour notre étude (III). Les caractéristiques du produit et de l'offre de monnaie seront analysées du point de vue de la présence d'une racine unitaire, de stationnarité, de persistance des chocs et de rupture de tendance (IV). Connaissant les périodes et les caractéristiques des variables, nous pouvons faire la recherche de modèles du produit nominal. Nous avons étudié la présence de vecteurs de co-intégration et des relations simples entre le produit nominal et l'offre de monnaie (IV). Finalement nous concluons (VI). Nous pouvons confirmer le quantitativisme du produit nominal durant les périodes de production de monnaie inconvertible. Dans une Annexe, nous indiquons les variables utilisées ainsi que leurs sources.

1. Introduction

Taylor (2000), Brainard et Perry (2000) ainsi que Cogley et Sargent (2002) ont démontré, pour les Etats-Unis, l'existence d'une réduction de la persistance du taux d'inflation depuis les années 80. Les conséquences de ce phénomène au niveau du modèle macroéconomique du taux de chômage naturel sont considérables[2]. En fait, la neutralité de la monnaie[3] ne devrait plus être retenue.

Les concepts de neutralité et, par conséquent, de non-neutralité de la monnaie, si chers aux économistes pour la compréhension du fonctionnement de l'économie se trouvent aujourd'hui

2 Voir (Hall, 1999) et (Taylor, 1998).

3 Et donc, de la politique monétaire. Mais on devrait aussi l'admettre, en général, pour les politiques de demande.

réduits à l'étude empirique d'une seule variable macroéconomique: le taux d'inflation. Nous considérons que le concept de neutralité de la monnaie est beaucoup plus profond que cela et qu'il est au cœur de nos conceptions sur le fonctionnement de l'économie. Il n'est pas légitime de le limiter à la confirmation des seules caractéristiques statistiques d'une seule variable macroéconomique. En outre, les problèmes d'équivalence observationnelle[4] ne doivent pas être ignorés à ce propos.

Après les essais de Cantillon (1755) et Bodin (1568) sur la théorie quantitative de la monnaie, nous pouvons affirmer que cette théorie a atteint une version cohérente avec Ricardo, (1810-11). L'analyse dichotomique lui permettra de ne prendre que les seuls effets de la monnaie sur les prix sans qu'elle soit affectée par les facteurs réels. Il nous a fallu attendre Wicksell, (1906) pour accepter la théorie quantitative et refuser l'ancienne dichotomie. La théorie monétaire moderne (Patinkin, 1965, 1987) fut bâtie sur les rejets de la dichotomie et du postulat de l'homogénéité (Leontief, 1936-7).

L'équilibre économique n'est que le résultat de l'égalité de l'épargne et de l'investissement, mais le taux réel nécessaire à cet équilibre n'est pas connu d'avance[5]. Que peut y faire la politique monétaire? Simplement réduire les déséquilibres entre l'épargne et l'investissement. La stabilité des prix serait non seulement un indicateur de déséquilibre mais également un expédient minimum pour l'obtention d'un tel objectif. Les transformations sociales et politiques après la Première Guerre Mondiale, ainsi que l'évolution de l'analyse économique, ont conduit à des contraintes de liquidité dans la conduite de la politique économique, lesquelles ont détruit l'étalon-or et ont commandé la conduite de la politique monétaire[6]. La politique monétaire peut réduire les fluctuations mais pas les éliminer: les différences d'opinion à propos de la crise de 1929, entre Friedman, d'un côté, et les Autrichiens, de l'autre, sont bien connues. Pour Friedman, une politique monétaire différente aurait résolu la crise. Pour les Autrichiens, il aurait été impossible de l'éliminer. La crise avait pour origine les comportements du passé, des investissements mal ajustés, mais cohérents avec les politiques monétaires suivies. Le démon de Maxwell[7] ne limite pas leur action au changement des prix, elle est plus profonde. Premièrement, les changements dans la quantité de monnaie ne peuvent jamais avoir les mêmes effets, au même moment, sur tous les biens[8]. Deuxièmement, les variations dans la quantité de monnaie introduisent un facteur dynamique dans le système économique[9]. Comment peut-on atteindre une monnaie neutre, en ce qui concerne les variables réelles de l'économie?

Beaucoup de temps s'est écoulé depuis ces affirmations de Mises jusqu'au commentaire du néo--keynésien Kaldor[10] à propos des effets de distribution associés aux variations dans la quantité de monnaie. Pour le meilleur et pour le pire[11], les conséquences des changements dans la quantité de monnaie ne s'épuisent pas dans le niveau général des prix. Il nous faut prendre les différents effets sur les individus, les biens et la production. Nous devons accepter que les distorsions dans la structure productive sont la conséquence logique de l'éloignement des valeurs du taux d'intérêt réel de sa valeur d'équilibre[12]. En conclusion, nous ne devons pas isoler

4 (Sargent, 1976).
5 (Hayek, 1978; Wicksell, 1898). Voir aussi (Cassel, 1928) et l'exposition faite dans (Humphrey, 2002).
6 (Hayek, 1936)
7 (Georgescu-Roegen, 1966). Faisant l'analogie avec le phénomène d'entropie, le démon correspond à une «nouvelle énergie» créée avec l'émission de monnaie.
8 «Changes in the quantity of money can never affect the prices of all goods and services at the same time and to the same extent». (Mises, 1949: 396).
9 «(E)very variation in the quantity of money introduces a dynamic factor into the static economic system». (Mises, 1912: 145).
10 (Kaldor, 1986).
11 N'oublions pas que si l' on rejette la neutralité de la monnaie, cela signifie qu'on pourra retourner à une conception de courte période de la courbe de Phillips. Le meilleur: la réduction temporelle du taux de chômage; le pire: le retour à des périodes inflationnistes.
12 (Mises, 1912: 21 et 556).

les conséquences sur le niveau des prix des conséquences sur le volume et la structure de production. Ainsi, la conception d'une monnaie neutre fut-elle qualifiée, par Hayek, de "caricature"!

Nous soutenons la non-neutralité de la monnaie et, en même temps, au niveau national et pour les régimes monétaires inconvertibles, nous soutenons le quantitativisme du produit nominal. L'étude de ce dernier concept doit être prise comme préalable à l'étude de la neutralité de la monnaie. Pour présenter ce concept, nous ferons référence à M. Friedman et à A. Chaîneau. On insistera sur le fait que M. Friedman, après avoir commencé par la présentation d'une théorie du produit nominal, va soumettre son idée originale à la vieille théorie quantitative des prix. Dans le cas d'A. Chaîneau, on peut voir que les concepts de déséquilibre au niveau de l'analyse de l'équilibre général, présents, ou suggérés, dans Lange (1945), Patinkin, (1965)[13]; Chaîneau (1964), furent intégrés dans son analyse macroéconomique du revenu nominal.

Friedman nous a proposé la théorie quantitative de la monnaie comme une théorie de la demande de monnaie. On connaît sa célèbre définition selon laquelle la théorie quantitative de la monnaie considère non seulement la demande de monnaie comme stable mais la considère également comme prépondérante pour la détermination du produit nominal[14]. En conséquence, la théorie quantitative est devenue ainsi une théorie du revenu nominal. C'est dans ce sens qu'on doit prendre le modèle monétariste qui s'est développé après la première moitié des années soixante.

Friedman, en cohérence avec ses idées sur la monnaie, nous a encore proposé que l'analyse macroéconomique devrait expliquer: (a) les mécanismes qui conduisent, dans une courte période, à distinguer dans les variations nominales ce qui appartient aux prix et au produit; (b) l'ajustement de courte période du revenu nominal aux variations des variables autonomes; et enfin (c) la transition des états de courte période vers l'équilibre de longue période[15]. Ce programme de recherche est cohérent dans ses objectifs. Mais, entre les mains de Friedman, il y a là un problème soulevé par le quantitativisme des prix. Cette construction d'une théorie du revenu nominal était destinée à loger la neutralité de la monnaie. La position de Friedman est très nette: les changements de la quantité de monnaie, en excès de la croissance réelle, se traduiront, dans la longue période, par des changements des prix. D'une façon synthétique, comme l'a affirmé A. Schwartz: la croissance soutenue de la monnaie, relativement à la croissance du produit, déterminera le comportement de longue période des prix[16].

Avec Friedman, l'analyse macroéconomique a commencé à s'intéresser aux variables nominales. La macroéconomie de type keynésien ne s'intéressait qu'aux seules variables réelles, et si elle admettait des changements des prix, ces variations n'étaient pas déterminées par l'offre de monnaie. Mais, pour Friedman, le quantitativisme des prix conduisait à la neutralité de la monnaie dans la longue période et, en conséquence, nous pouvons dire qu'il n'y a pas, chez Friedman, une place autonome pour un quantitativisme nominal. Le quantitativisme nominal est en même temps un quantitativisme des prix. La position monétariste d'André Chaîneau est, à ce propos, bien différente.

Dans une tradition française, qui remonte au XVIIIe siècle, A. Châineau[17] se propose de mener une analyse macroéconomique encadrée par une modélisation en termes de circuit économique. Dans une représentation sans et avec production autonome de monnaie, il en arrive à la conclusion que les comportements monétaires sont responsables de l'instabilité de la demande

13 Première édition en 1950.

14 «The quantity theorist not only regards the demand for money as stable; he also regards it as playing a vital role in determining variables that he regards as of great importance for the analysis of the economy as a whole, such as the level of money income (...)». (Friedman, 1956: 109).

15 (Friedman, 1973: 89).

16 «A sustained change in the growth rate of money relative to growth in output determines the long-run behaviour of prices». (Schwartz, 1992:19).

17 Voir surtout (Chaîneau, 1992; Chaîneau, 1995; Chaîneau, 1996).

globale et que cette instabilité va se transmettre à toute l'économie. Son objectif principal est l'étude d'une économie avec production (disons élastique) de monnaie dont le coût de production marginal est nul. Cette production est nécessaire pour deux raisons: d'une part, satisfaire le désir de thésaurisation des ménages sans créer des situations déflationnistes, d'autre part, adapter la circulation monétaire à la croissance de l'économie. Chaîneau considère également, parallèlement à la croissance réelle, la croissance des prix qui caractérise les économies industrielles (Chaîneau, 1992: 191). Les chocs provoqués par la demande globale ont des conséquences déséquilibrantes différentes dans des systèmes monétaires avec et sans contrôle de production de monnaie.

On n'insistera jamais suffisamment sur l'absence de contenu du concept de demande endogène de monnaie de la part de certains auteurs keynésiens. De la demande individuelle de monnaie à la demande globale de monnaie, on franchit le pas de géant que Patinkin (1965) a qualifié comme allant de l'expérimentation individuelle à l'expérimentation du marché. Dans le cas de cette dernière, la demande de monnaie ne peut être différente de la quantité de monnaie qui circule. Et en même temps, c'est cette quantité, (ou offre), nominale de monnaie qui détermine la valeur globale demandée dans l'économie (Chaîneau, 1992: 143)[18].

Dans sa théorie de l'équilibre macroéconomique, la demande et l'offre de monnaie sont deux concepts centraux. Pour Chaîneau, la demande et l'offre de monnaie sont des phénomènes indépendants et l'offre de monnaie précède la formation de la demande de monnaie (Chaîneau, 1992: 198 et 230). Les besoins d'endettement des entreprises auprès des banques déterminent la création monétaire tandis que la demande de monnaie est déterminée par les besoins d'encaisses de ménages (Chaîneau, 1992: 198). Les fondements microéconomiques de l'expérimentation individuelle sont les principes de l'indépendance de la demande et de l'offre de monnaie. Les conséquences de cette production de monnaie font d'elle un bien particulier: non seulement son offre est finalement indépendante de sa demande mais sa demande s'adaptera aussi à son offre (Chaîneau, 1992: 207). Aussi devons-nous accepter que ce qui se produit avec la monnaie est justement le contraire de ce qui se produit avec les autres biens. C'est justement ce type de comportement qui est à l'origine de la spécificité des conséquences des chocs monétaires.

Les deux phénomènes de demande et d'offre étant indépendants, le problème de création monétaire devient complexe en tant que résultat d'influences contradictoires. Nous pouvons penser à l'alimentation du circuit économique, soit en raison des besoins de financement bancaire des entreprises (c'est-à-dire en fonction des demandes anticipées de crédit bancaire), soit en fonction des encaisses monétaires du secteur non bancaire (Chaîneau, 1992: 199). Les conséquences sur la demande globale sont naturellement différentes dans un cas et dans l'autre. Si, dans le premier, on valorise les perspectives d'investissement, dans le deuxième, on valorise la stabilité du pouvoir d'achat de la monnaie. En tous cas, au niveau de l'expérimentation de marché, l'offre de monnaie déterminera le niveau nominal de la demande et de l'offre globale.

En résultat des décisions des entreprises, des banques et des ménages, la situation normale doit être celle d'un déséquilibre *ex ante* entre l'offre et la demande nominale de monnaie. Un programme de recherche macroéconomique, au-delà d'une théorie de détermination de la demande globale nominale, doit répondre aux questions suivantes: – comment se forme l'équilibre entre les différents flux monétaires du circuit? – quelles sont les conditions qui conduisent à la création des déséquilibres? – et, comment les déséquilibres sont-ils éliminés sachant qu'on aura toujours *ex post* l'égalité de la demande et de l'offre globale? André Chaîneau répond, plus précisément, aux deux premières questions.

Les déséquilibres nominaux sont éliminés par des ajustements des quantités et des prix. La valeur nominale de l'offre et de la demande globale ne peut être déterminée que par l'offre de

18 Voir, particulièrement, tout le chapitre 7.

monnaie (Chaîneau, 1992: 143). A supposer que, dans la longue période, les comportements de thésaurisation et de déthésaurisation soient neutres, la demande nominale globale aura une élasticité unitaire par rapport à l'offre de monnaie dans les régimes de monnaie inconvertible. C'est cette thèse que nous allons tester pour l'économie portugaise, par l'étude des différents régimes monétaires que l'économie portugaise a connus.

2. Régimes Monétaires dans l'Economie Portugaise

L'étude du comportement du produit nominal, par rapport à l'offre de monnaie, de 1854 à 1998, doit prendre en considération les différents régimes monétaires qui ont caractérisé cette période. Nous retenons la définition de régime monétaire de Bordo: "un ensemble de pratiques et d'institutions avec toute une série d'anticipations – anticipations des agents sur le comportement des décideurs et anticipations des décideurs à propos des réactions des agents à leurs décisions"[19].

D'une façon trop empiriste, Neumann (1993) propose l'idée qu'un régime monétaire peut être décrit par un taux moyen de croissance de monnaie et par une variance de ce taux. Nous pensons qu'un régime ne peut ignorer des caractéristiques empiriques d'autres variables telles que le produit nominal et réel ainsi que les prix[20]. Comme notre intérêt dans cette étude concerne le comportement du produit nominal, nous nous intéressons seulement à la manière dont ces deux variables (produit et monnaie) peuvent nous aider à délimiter les différents régimes. Pour des raisons de comparaison internationale, on étudiera aussi l'évolution des prix portugais, américains et anglais.

Normalement, l'étude des régimes monétaires commence par l'étalon-or, en passant sous silence les événements du XVIIIe siècle. Deux raisons le justifient: la dimension géographique limitée du régime et l'absence, très gênante, de statistiques disponibles. Le XIXe siècle a vu le retour à la convertibilité de la livre anglaise en 1821 et l'adoption de l'étalon-or en 1850 par l'Australie, le Canada et le Portugal. L'évolution a été telle que, dans la deuxième décennie du XXe siècle, la majorité des pays participant au commerce international étaient dans l'étalon-or. L'utilisation de ce type de changes fixes fut un facteur de croissance important du point de vue réel (Rose, 2000; Córdova and Meissner, 2000; Flandreau et Maurel, 2001; et Meissner, 2002) et financier (Bordo et Rockoff, 1996). L'idéologie de l'étalon-or dominait le XIXe siècle. La majorité des pays, à la Conférence de Paris de 1867, étaient favorables à une Union Monétaire fondée sur l'or (Russell, 1898).

Soit pour des raisons de proximité et de dépendance vis-à-vis de l'Angleterre, soit aussi à cause de l'instabilité associée à l'émission de monnaie de papier, le Portugal n'a pas connu de mouvements comme au Brésil, avec les "papelistas" (Fritsch et Franco, 2000) et au Chili, avec les "papeleros" (Subercaseaux, 1922).

Beaucoup d'asymétries furent attribuées à l'étalon-or. Citons-en quelques-unes: les conséquences des variations des prix sur la production; les centres financiers qui avaient un pouvoir de contrôle sur le taux d'intérêt; et finalement le type d'exigence faite aux pays excédentaires et déficitaires vis-à-vis de l'extérieur. En ce qui concerne le cas du Portugal, comme pays périphérique, c'est cette dernière asymétrie qui est importante (Eichengreen et Flandreau, 1997). Le Portugal a fait partie de ce que l'on considère comme le premier étalon-or (...- 1913), mais pas du deuxième (1925-1931). Ce qui veut dire que le Portugal a pu bénéficier du phénomène de crédibilité qui a caractérisé le premier étalon-or mais pas du deuxième.

L'économie portugaise a connu le cours légal des monnaies d'or françaises (rappelons que la première invasion de Napoléon eut lieu en 1807-1808) puis anglaises. La Guerre Civile a

19 «We define a regime as a set of arrangements and institutions accompanied by a set of expectations – expectations by the public with respect to policymaker's actions and expectations by policymakers about the public's reaction to their actions». (Bordo et Jonung, 1997: 5).
20 L'importance de l'emploi naît seulement, au XXe siècle, dans les années 20.

conduit, en 1832, à la reconnaissance du cours légal des monnaies d'or de l'Angleterre et d'argent de l'Espagne, du Brésil et du Mexique. A partir de 1840 jusqu'en 1851, d'autres monnaies ont bénéficié du cours légal. En 1851, seules les monnaies d'or portugaises et anglaises pouvaient bénéficier du cours légal (Mendonça, 1996).

La Lettre de Loi du 29 juillet 1854 a appliqué au Portugal le modèle anglais d'étalon-or. Pour comprendre pourquoi cette adhésion ne fut pas difficile, au-delà de la circulation des monnaies d'or (anglaises), il faut ajouter le bas prix de l'or après 1840, en conséquence de la découverte et de l'exploitation des nouvelles mines d'or de la Californie et de l'Australie (Esteves et Ferramosca, 2000). Et en fait, le Portugal fut le premier pays, après l'Angleterre, à adopter l'étalon-or au XIXe sècle.

Néanmoins le Portugal fut un très mauvais élève: déficits persistants tant au niveau externe qu'au niveau du budget de l'Etat (Mata et Valério, 1995; Reis, 2002). Pourtant le régime n'a pas mal fonctionné jusqu'en 1891, excepté l'épisode d'inconvertibilité effective d'une semaine[21] lors de la crise financière de 1876. Des raisons politiques mais aussi financières, internes et internationales, ont dicté la fin de l'étalon-or, au mois de mai 1891. Depuis 1891, la monnaie portugaise[22] n'a, en pratique, connu que le cours forcé. Le Portugal a été l'un des premiers pays à adhérer à l'étalon-or et il a été aussi le premier à éliminer, dans la Loi, tout rapport entre la monnaie (l'Escudo) et l'or en 1975[23].

Pendant la période de 1891-1931, la monnaie portugaise a connu des changes flexibles (Mata, 1991). Les années qui se succèdent, à la fin de la convertibilité, doivent être étudiées avec beaucoup d'attention. Premièrement, il faut tenir compte des pratiques d'émission des pays, lors des périodes d'inconvertibilité, qui visent un possible retour à l'étalon-or (Bordo et Redish, 1993). On dirait aujourd'hui que c'était un problème de crédibilité associé à la non convergence temporelle des politiques. Deuxièmement, la période en cause était une période d'intense internationalisation des relations économiques (Keynes et Bainville, 2002[24]; Frieden, 1997; et Hogendorn, 1998) et, en certains aspects, elle l' a été aussi pour le Portugal.

Les règles ne pouvaient pas changer beaucoup si on voulait maintenir et améliorer les relations commerciales et financières. Aussi devons-nous être très attentifs à la période post étalon-or et à celle précédant la Première Guerre Mondiale.

Après la Première Guerre Mondiale, la monétisation des déficits budgétaires et l'instabilité politique[25] ont dominé les règles d'émission de monnaie. De 1931 à 1949, l'Escudo fut ancré à la Livre anglaise[26]. Des accords conclus avec l'Angleterre ont conduit à des avoirs considérables en Livres, ce qui a maintenu cet ancrage de la monnaie portugaise au-delà de la fin de la Guerre (Bordo et Santos, 1995).

De 1949 à 1973, l'Escudo fut définitivement ancré au Dollar[27]. Les excédents extérieurs ont conduit le Portugal à posséder, à la fin des années 1960, 2,25% du stock d'or mondial. L'affaire de l'emprunt de 1927 de la Société des Nations et le refus, en 1946, de l'entrée du Portugal à l'ONU ont conduit le régime de Salazar à ne solliciter la participation au Fonds Monétaire International et à la Banque Mondiale qu'en août 1959 (Bordo et Santos, 1995).

---

21 Cependant, l'autorisation donnée par le Gouvernement à la Banque du Portugal a été de 2 mois. Je dois cette correction au referee de Notas Económicas.

22 Le Real puis l'Escudo.

23 Dans la nouvelle Loi Organique de la Banque du Portugal.

24 Nous ne pouvons pas oublier les mots de Keynes à propos de cette période: «Quel extraordinaire épisode dans l'histoire du progrès économique de l'homme, cette époque qui prit fin en août 1914!», (Keynes et Bainville, 2002: 25). Voir tout le Chapitre II, «L'Europe Avant la Guerre».

25 De 1919 jusqu'au mois de mai 1926, 30 gouvernements se sont succédé.

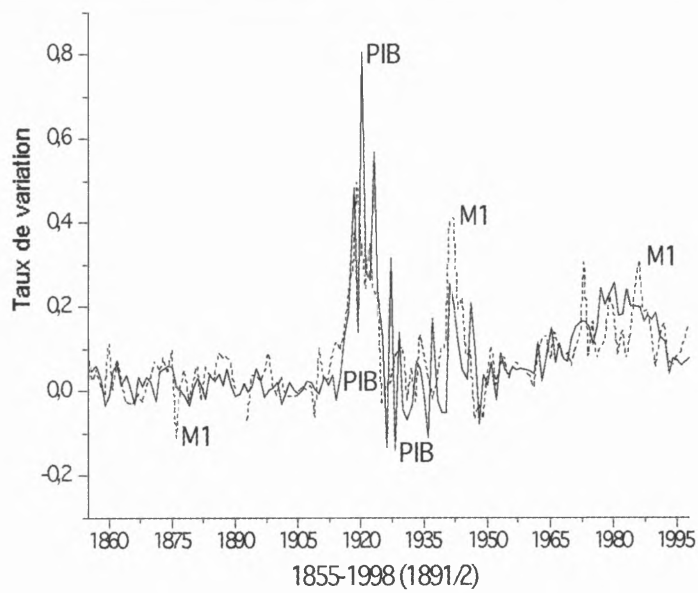
26 Sur l'éphémère expérience des 82 jours dans l'étalon-or, en 1931, on peut consulter (Santos, 1995).

27 Nous avons choisi l'année 1949 pour le début de l'après Guerre. On ne doit pas oublier que la période allant de 1939 à 1947, a été, du point de vue du taux de change, une période de transition. Voir (Valério, 1991: 9)

En 1977, en conséquence du premier choc pétrolier et de la Révolution des Œillets (avril 1974), l'accord signé avec le FMI a conduit à une politique de *crawling peg* qui s'est maintenue au-delà de l'entrée dans la CEE (janvier 1986). En 1991, les autorités monétaires portugaises ont essayé une politique de stabilisation de l'Escudo par rapport à la monnaie allemande[28]. En avril 1992, l'Escudo a commencé sa participation au mécanisme de changes du SME (dans la marge des 6%). Ce qui s'est produit depuis est commun aux autres pays de l'UE.

Dans la Figure 1, nous avons la représentation des taux de croissance du produit nominal et de l'offre de monnaie. Nous avons omis les valeurs de 1891 et 1892 pour l'offre de monnaie.

Figure 1 - Evolution du Produit et de la Monnaie



Nous pouvons constater que l'offre de monnaie, sauf pendant la période de la Première Guerre Mondiale, flotte de façon plus intensive que le produit. Tenant compte des périodes de convertibilité, de non-convertibilité et aussi d'ancrage de la monnaie portugaise, nous avons calculé, pour les taux de croissance des prix (DLP), du produit (DLQ) et de l'offre de monnaie (DLM), la moyenne (μ), l'étalon type (σ) et aussi le coefficient de variation (σ/μ). Ce coefficient mesure l'instabilité du comportement de n'importe quelle variable. Analysons le résumé des statistiques calculées dans le Tableau 1.

28 Dont le résultat était aussi la valorisation réelle de l'Escudo comme instrument pour réduire l'inflation portugaise.

Tableau 1 – Caractérisation statistique des différentes périodes									
	DPL			DLQ			DLM		
	μ	σ	σ/μ	μ	σ	σ/μ	μ	σ	σ/μ
1855-1890	0,002	0,06	23,7	0,02	0,07	3,45	0,032	0,047	1,45
1892-1914	0,006	0,03	4,54	0,005	0,03	7,19	0,026	0,041	1,57
1915-1931	0,18	0,24	1,32	0,016	0,06	4,03	0,17	0,149	0,87
1915-1945	0,12	0,2	1,65	0,011	0,06	5,84	0,15	0,143	0,94
1931-1949	0,03	0,08	2,49	0,013	0,06	4,65	0,09	0,138	1,53
1949-1972	0,02	0,02	0,98	0,049	0,03	0,64	0,07	0,046	0,67
1973-1998	0,13	0,06	0,47	0,027	0,03	1,25	0,14	0,07	0,49

La période la plus stable – quelle que soit la variable choisie – est celle après la Seconde Guerre Mondiale. Il est très intéressant de constater que la plus forte instabilité des prix a été vécue sous l'étalon-or. Dans l'après Seconde Guerre Mondiale (1949-1972), le taux de croissance du produit a été le plus élevé, et le moins instable. Les performances comparées des différentes périodes au Portugal ne s'éloignent pas beaucoup de celles d'autres pays[29]. La croissance de l'économie portugaise fut très faible de 1892 à 1949. Si l'on retient les prix et l'offre de monnaie, on constate que la période de convertibilité (1855-1890) et d'inconvertibilité avant la Première Guerre Mondiale ne diffèrent pas beaucoup.

Tableau 2 – Caractérisation du taux d'inflation pour le Portugal, EUA et RU									
	USA DPL			UK DLQ			Port. DLM		
	μ	σ	σ/μ	μ	σ	σ/μ	μ	σ	σ/μ
1855-1890	0,002	0,024	11,46	-0,008	0,015	-1,807	0,005	0,008	1,635
1892-1914	0,011	0,018	1,57	0,017	0,019	1,142	0,027	0,039	1,471
1915-1931	0,021	0,025	1,21	0,012	0,027	2,157	0,158	0,057	0,359
1915-1945	0,018	0,023	1,24	0,019	0,025	1,32	0,102	0,075	0,733
1931-1949	0,02	0,022	1,12	0,029	0,022	0,749	0,035	0,007	1,206
1949-1972	0,032	0,012	0,37	0,051	0,019	0,372	0,033	0,028	0,829
1973-1998	0,052	0,017	0,33	0,076	0,033	0,438	0,125	0,042	0,336

Dans le Tableau 2, nous avons également comparé le taux d'inflation portugais avec ceux des Etats-Unis et de l'Angleterre. Comme la réaction de Triffin (1997) à Cassel (1930), sur la stabilité des prix de 1850 à 1910, s'est fondée sur le comportement cyclique de l'économie, qui conduisait, ce dernier, à comparer des prix à différentes phases du cycle, nous avons utilisé le filtre de Hodrick-Prescott (Hodrick et Prescott, 1980), avec un $\lambda = 400$ pour obtenir des valeurs de tendance[30] (DLP*). Au-delà de la considérable instabilité de la première période pour les Etats-Unis, les résultats diffèrent d'une façon nette pour le taux d'inflation portugais dans l'Entre--deux-Guerres et après la chute de Bretton-Woods. Pour ces périodes, le taux d'inflation portugais est bien plus élevé qu'aux Etats-Unis et qu'en Angleterre. D'une façon générale, nous constatons que l'économie portugaise est plus inflationniste que les économies dominantes depuis la première moitié du XIXe siècle. Pour le seul cas de l'économie portugaise, nous pouvons confirmer les valeurs instables du taux d'inflation de 1855 à 1914. Quand on compare

29 Voir (Bordo et Santos, 1995; Bordo, 1993; Bordo et Jonung, 2001).

30 Grâce aux données disponibles, nous n'avons pas eu besoin de faire la correction des valeurs des points finaux (end points).

les valeurs de l'après Seconde Guerre Mondiale, de 1949-1972, avec les valeurs postérieures de 1973-1998, on s'aperçoit que la multiplication par 4 du taux d'inflation fut accompagnée d'une réduction de l'instabilité.

3. Périodisation de l'Histoire Monétaire Portugaise

L'étude d'une période historique déterminée suppose un découpage en sous-périodes. Ce découpage se fonde à la fois sur des connaissances historiques et économiques de ladite période et sur certaines techniques statistiques-économétriques. Ces techniques permettent la confirmation ou la révision du découpage basé sur les seules caractéristiques économiques de la période. C'est ce que nous ferons ici.

Pour l'obtention empirique des différentes phases de l'économie portugaise, nous appliquerons la méthodologie de Bai-Perron (Bai et Perron, 1998, 2001a, 2001b) aux variables qui, dans ce contexte, sont plus importantes pour nous: le produit nominal et l'offre de monnaie.

La méthode de Bai-Perron cherche à obtenir le nombre de ruptures d'une série ayant pour base le modèle linéaire suivant:

$$y_t = x'_t \cdot \beta + z'_t \cdot \delta_j + \mu_t$$

avec j=1, ..., m+1. Où m est le nombre de ruptures, x représente les variables dont les coefficients ne changent pas et z le vecteur des variables dont les coefficients changent pour chacun des m+1 régimes. Deux tests sont utilisés pour l'hypothèse nulle d'absence de rupture contre l'existence d'un certain nombre de ruptures: UDmax et WDmax. Le test $SupF_T(k)$ a comme hypothèse nulle la présence de k ruptures contre (k-1). La valeur UDmax correspond à la valeur $SupF_T(k)$ pour k maximum. WDmax, contrairement à UDmax, correspond à une moyenne pondérée des tests individuels $SupF_T(k)$.

Nous avons suivi la stratégie proposée par les deux auteurs. Le choix du nombre de ruptures a été fait sur le critère d'information LWZ (Liu et al., 1997) et les valeurs de UDmax et WDmax.

Pour l'étude du produit nominal[31] (LYC) nous supposons, au-delà d'une constante, l'offre de monnaie (LMC) comme variables dont les coefficients ne changent pas. Pour l'offre de monnaie (LMC), nous avons supposé seulement la constante. Dans les deux cas, la variable z n'est que la tendance[32].

31 Nos variables sont calculées par tête. LYC et LMC signifient logarithme du produit nominal par tête et logarithme de M1 para tête. DLYC et DLMC ne sont que les taux de variation du produit par tête et de M1 par tête.

32 Beaucoup d'autres hypothèses furent utilisées mais celles-ci nous semblent les plus adéquates face aux résultats obtenus.

Tableau 3 - Application de la méthode de Bai-Perron (1854-1998)

LYC 1854-1998				
	m = 4	Z = T	X = [1, MC	
SupFt(1)	SupFt(2)	SupFt(3)	SupFt(4)	
44,49	645,31	45,62	906,43	
(8,58)	(7,22)	(5,96)	(4,99)	(5%)
UDmax	**WDmax**			
906,43	1558,55			
(8,88)	(9,91)	(5%)		
1890	**1913**	**1941**	**1977**	
1889-91	1898-915	1938-44	1967-80 (90%)	

LMC 1854-1998				
	m = 4	Z = T	X = 1	
SupFt(1)	SupFt(2)	SupFt(3)	SupFt(4)	
0,004	206,6	2,18	451,63	
(8,58)	(7,22)	(5,96)	(4,99)	(5%)
UDmax	**WDmax**			
451,63	776,56			
(8,88)	(9,91)	(5%)		
1890	**1918**	**1940**	**1977**	
1886-00	1916-19	1918-42	1933-89 (90%)	

Comme nous pouvons le constater, que ce soit pour LYC ou pour LMC, nous avons 4 ruptures pour l'ensemble de la période (1854-1998). On pourrait avoir des doutes sur la troisième rupture dans LMC, mais la quatrième rupture annule les doutes possibles. L'identification des périodes de ruptures au niveau des 90% de probabilité est bien moins précise dans le cas de l'offre de monnaie que dans le cas du produit.

La première rupture correspond à la fin de l'étalon-or; la deuxième à la Première Guerre Mondiale, soit au début soit à la fin (1913 ou 1918); la troisième au début de la Seconde Guerre Mondiale (1940, 1941); et finalement la quatrième au début des politiques de stabilisation après la crise du pétrole et la Révolution de 1974. Ces indications de la méthode de Bai-Perron nous aident à préciser les phases, ou périodes particulières, que nous devons étudier depuis la moitié du XIXe siècle jusqu'en 1998. Nous sommes dans l'obligation d'étudier la période de l'étalon-or ainsi que la période jusqu'à la veille de la Première Guerre Mondiale (1854-1890 et 1854-1913). De même que nous devrons isoler la période de l'Entre-deux-Guerres. Un changement dans le régime du taux de change s'étant produit après 1949, nous identifierons la période de l'Entre-deux-Guerres comme allant de 1914 à 1948.

4. Les Caractéristiques Statistiques du Produit Nominal et de l'Offre de Monnaie

Prenant en considération les périodes choisies, nous procédons d'abord à l'étude des caractéristiques de stationnarité des séries du produit et de l'offre de monnaie. Notre étude ne cherche pas à comparer des régimes du point de vue théorique et empirique. Par exemple, quel doit être le comportement de l'offre de monnaie et du produit nominal pendant l'étalon-or et qu'en a-t-il été très exactement? Nous étudierons le comportement des séries pour aboutir à des modèles explicatifs du produit nominal par l'offre de monnaie.

Au-delà des tests usuels ADF, du t, $((\hat{\rho}-1)/\hat{\sigma}_{\rho})$, et du Z, $(N \cdot (\hat{\rho}-1)$, pour l'hypothèse nulle de racine unitaire, nous avons utilisé un certain nombre d'autres tests. Normalement, nos préoccupations résident dans le rejet ou l'acceptation d'une racine unitaire, mais d'autres possibilités doivent être envisagées quand on veut connaître la persistance des chocs sur les valeurs d'une série. Ainsi, parallèlement aux tests ADF, pour $\rho = 1$, nous utilisons la méthode de (Stock, 1991) pour le calcul de l'intervalle de 90% du ρ. De cette façon, on aura une idée de la valeur probable la plus élevée du ρ. Nous testons aussi l'hypothèse nulle de stationnarité en utilisant le test KPSS (Kwiatowski, et al., 1992). La valeur A(1), parfois aussi identifiée comme $\psi(1)$, nous donne, en k, l'effet retenu d'un choc unitaire k périodes avant (Campbell et Mankiw, 1987; Cochrane, 1988). Ce type d'information peut être très utile du point de vue de la connaissance de la persistance de chocs sur une variable. Nous avons également appliqué les tests proposés par Bhargava (1986)[33].

33 Nous avons utilisé le code de (Simon van Norden, 1990) actualisé par (Jeff Gable, 1993) pour la version 4 du RATS.

Le test R (R_1 et R_2; R_2 pour le cas de la présence de drift) a comme hypothèse nulle une marche aléatoire contre le cas de stationnarité, tandis que N (N_1 et N_2; N_2 pour le cas de la présence de drift) a comme hypothèse nulle une marche aléatoire contre le cas de comportement explosif ($\rho>1$). Ces deux tests doivent être lus conjointement. De l'ensemble des critères d'information sur l'analyse de persistance d'une série (Pivetta et Reis, 2002), nous préférons l'addition des coefficients des retards. L'ordre du processus autorégressif fut choisi en fonction du critère de Schwarz, qui pénalise, comme on le sait, le nombre de paramètres d'un modèle. Finalement, nous avons appliqué l'étude de (Perron, 1997) qui teste le cas de racine unitaire autour d'une tendance avec différent types de ruptures dans la tendance[34].

Les résultats sont présentés dans le Tableau 4[35].

34 Nous avons seulement retenu les cas où nous ne pouvons pas rejeter une quelconque rupture dans la série et où, en même temps, nous pouvons rejeter l'hypothèse nulle d'une tendance temporelle.

35 Notes du Tableau 4. (a) 18 observations; (b) 17 obs.; (c) sans constante; (d) 29 obs.; (e) changement de l'inclination de la tendance avec les deux segment de cette tendance sans discontinuité; (f) changement dans l'inclination; (g) 15 obs.; (h) changement dans l'inclination et dans la constante; (i) 11 obs.; (j) stationnaire aussi en niveaux; (k) 24 obs; et (l) 30 obs. Une * pour rejet de H0 à 10%, deux pour 5% et trois pour 1%. Pour le KPSS, les étoiles signifient non rejet à 2,5%, 5% et 10%.

Tableau 4 – Tests de racine unitaire et de stationnarité

			Obs	Det	L	ADF_T	ADF_Z	Stock91	KPSS_u	KPSS_t	A1	Drift	R&N	AR	Perron TB
1	LYC	1854-1890	37	T	1	3,73**	-25,93***	[0,64 0,93]	1,88	0,13**	0,66 (a)	C	0,47 / 0,21**	0,98	-3.99(1878)(a)
2	LMC	1854-1890	37	T	0	-1,35	-5,08	[0,957 1,04]	3,33	0,20	1,26(a)	C	0,22 / 0,18**	1,005	
3	DLMC	1855-1890	36	C	0	-5,43***	-32,39***	[--- 0,64]	0,22***	0,14**	0,41(b)	-	1,80** / 0,72**	0,35(c)	
4	LYC	1854-1913	60	T	1	-3,00	-13,60	[0,761 1,02]	3,00	0,56	1,14(d)	C	0,17 / 0,05	0,98	4,43(1891)(e)
5	DLYC	1854-1913	60	C	0	-5,91***	-44,83***	[--- 0,56]	0,25***	0,04***	0,37(d)	-	1,50** / 0,94**	0,07	
6	LMC	1854-1913	60	T	0	-2,98	-14,80	[0,76 1,03]	3,50	0,62	0,71(d)	-	0,21 / 0,03**	0,86	-9,05***(1889)(f)
7	DLMC	1855-1913	59	C	0	-7,08***	-54,46***		0,10***	0,09***	0,34(d)	-	1,82** / 0,60**	0,06	
8	LYC	1915-1945	31	C	0	-2,88*	-2,44	[0,76 0,995]	2,24	0,64	2,21(g)	C	0,10 / 0,04	0,92	-5,41*(1918)(h)
9	DLYC	1915-1945	31	C	0	-4,18***	-23,14***	[--- 0,83]	0,57*	0,16*	0,62(g)	-	1,50** / 1,24**	0,55	
10	LMC	1915-1945	31	T	1	-3,00	-11,90	[0,762 1,025]	1,41	0,23	3,21(g)	C	0,08 / 0,06	1,01	-3,72(1917)(e)
11	DLMC	1915-1945	31	C	0	-2,38	-10,23	[0,826 1,017]	0,38**	0,28	0,78(g)	-	0,66 / 0,57**	0,67	-3,43(1930)(e)
12	LYC	1949-1972	24	T	0	0,77	1,52	[1,02 1,05]	2,32	0,52	1,72(i)	C	0,20 / 0,04	1,007	
13	DLYC	1949-1972	24	T	0	-7,13***(j)	-31,54***(i)		1,18	0,07***	0,50(i)	-	1,30** / 0,84**	0,69	-8,31***(1969)(e)
14	LMC	1949-1972	24	T	0	-1,72	-3,37	[0,92 1,042]	2,35	0,55	1,91(i)	C	0,18 / 0,05	1,006	
15	DLMC	1949-1972	24	T	0	-3,81***	-18,76***	[0,63 0,92]	0,90	0,06***	0,51(i)	-	-1,02** / 0,10**	0,49	
16	LYC	1973-1998	26	C	1	-3,29**	-1,08	[0,72 1,019]	2,59	0,58	4,48(i)	C	0,07 / 0,03	1,007	-4,93(1988)(h)
17	DLYC	1973-1998	26	T	0	-2,29	-8,97	[0,86 1,04]	1,42	0,41	0,84(i)	-	0,40 / 0,39**	0,824	
18	LMC	1973-1998	26	C	0	-1,39	-0,44	[0,93 1,03]	2,62	0,33	1,63(i)	C	0,21 / 0,19**	1,01	-3,72(1983)(f)
19	DLMC	1973-1998	26	C	0	-3,95***	-20,48***	[--- 0,86]	0,24***	0,09***	0,55(i)	-	1,29** / 0,19**	0,21	
20	LYC	1949-1998	50	T	6	-3,68**	-22,00**	[0,65 0,94]	0,79	0,19	6,05(k)	C	0,02 / 0,01	1,003	-4,12(1975)(f)
21	DLYC	1949-1998	50	C	1	-1,90	-5,75	[0,88 1,03]	1,13	0,44	0,80(k)	-	0,43 / 0,19**	0,83	-5,45**(1989)(f)
22	LMC	1949-1998	50	T	0	-2,97	-4,04	[0,76 1,03]	4,97	1,10	2,87(k)	C	0,04 / 0,02	1,006	-3,00(1967)(e)
23	DLMC	1949-1998	50	T	0	-4,89***	-32,88***	[--- 0,76]	1,44	0,24	0,45(k)	-	1,01** / 0,14**	0,49	
24	LYC	1854-1998	145	T	2	-1,83	-6,29	[0,91 1,04]	4,61	0,81	2,82(l)	C	0,01 / 0,004	1,003	-3,29(1926)(e)
25	DLYC	1855-1998	144	C	5	-3,69***	-48,08***	[0,63 0,89]	0,56*	0,06***	0,05(l)	-	1,10** / 1,05**	0,69	-5,56**(1922)(f)
26	LMC	1854-1998	145	T	2	-1,64	-4,51	[0,91 1,03]	4,67	0,94	2,69(l)	C	0,01 / 0,004	1,004	-3,29(1904)(e)
27	DLMC	1855-1998	144	C	0	-5,19***	-46,20***	[--- 0,68]	1,23	0,11***	0,38(l)	-	0,65** / 0,64**	0,68	-5,85***(1920)(e)

Pour interpréter le Tableau 4, prenons l'exemple de la ligne 2, LMC, 1854-1890. Nous avons 37 observations. Le test ADF a été fait avec tendance (T) et on a rejeté la possibilité d'autocorrélation des erreurs par un test LM avec 0 retards. Les valeurs du t et du Z ne rejettent pas la racine unitaire. Le calcul par la méthodologie proposé par Stock nous conduit, au niveau de 90%, à une valeur limite supérieure à 1. Les tests KPSS, sans tendance et avec tendance (KPSS_u et KPSS_t), rejettent la stationnarité. La valeur d'A1 signifie que 18 ans après un choc unitaire, la variable LMC retient encore la valeur 1,26 (supérieure à la valeur du choc). Le test de Bhargava doit être fait avec drift (C). En accord avec la valeur de R2, nous ne devons pas rejeter l'hypothèse de marche aléatoire contre la stationnarité de la série et, en accord avec la valeur de N2, nous rejetons l'hypothèse de marche aléatoire contre l'hypothèse d'un processus explosif. Faisant une représentation autorégressive de la variable LMC, la somme des coefficients est égale à 1,005. Dans ce cas, l'application du test de Perron n'a aucun sens. Par contre, dans la ligne 4, il a conduit à une rupture en 1891, en supposant pour cette année une rupture du type (e)[36], mais il n'a pas conduit à rejeter la présence d'une racine unitaire.

Après les tests décrits dans le Tableau 4, nous sommes en mesure de dire quelles variables doivent être retenues comme stationnaires ou lesquelles doivent être prises comme ayant une racine unitaire. Résumons nos résultats.

– LYC, 1854-1890. Le test de Bhargava (R2) ne rejette pas l'hypothèse de marche aléatoire. Les autres tests et indicateurs nous conduisent à accepter la stationnarité de la série.

– DLMC, 1855-1890, 1855-1913, 1949-1972, 1973-1998, 1949-1998 1855-1998. DLYC, 1854--1913, 1915-1945, 1949-1972, 1855-1998. La plupart, ou même la totalité des valeurs, nous conduisent à accepter la stationnarité.

– DLMC, 1915-1945. Il est intéressant de constater que seul le test KPSS_u et la valeur des coefficients AR rejettent l'hypothèse de variable avec une racine unitaire.

– DLYC, 1973-1998. Nous n'avons aucun doute sur la présence d'une racine unitaire dans cette variable.

– DLYC, 1949-1998. Pour cette variable, nous ne pouvons pas rejeter la stationnarité autour d'une tendance avec une rupture en 1989. Toutes les autres valeurs rejettent l'hypothèse de stationnarité.

5. Les Modèles de Détermination du Produit Nominal

Une fois que nous connaissons les caractéristiques de racine unitaire du produit nominal et de l'offre de monnaie, pendant des périodes sélectionnées en raison des changements structuraux et des comportements d'émission et de taux de change, nous pouvons passer à la recherche des modèles du produit nominal en fonction de l'offre de monnaie. Notre objectif consiste donc à obtenir des modèles pour les périodes 1854-1890, 1854-1913, 1949-1998 et 1914-1998, qui recouvrent l'étalon-or portugais, son influence sur l'économie portugaise, le post Bretton Woods et la période depuis la fin de l'étalon-or international. On utilisera les caractéristiques de non--stationnarité des variables LYC et LMC.

Apparemment la recherche d'une élasticité unitaire du produit nominal est équivalente, empiriquement, à la stationnarité de la vitesse de circulation de la monnaie. Nous préférons deux autres démarches. L'obtention des modèles dynamiques pour le produit et des modèles multivariés qui résultent de l'application de la méthodologie de Johansen au produit et à la monnaie. Cette dernière méthodologie permet la détermination conjointe des deux variables retenant l'interdépendance qui peut exister entre elles et l'obtention des valeurs des paramètres de longue période qui en résultent. Avec les modèles à une équation, nous avons des résultats pour analyser les trois premières périodes et, avec les modèles de co-intégration nous pouvons analyser les deux dernières périodes.

36 Voir la note antérieure.

Les modèles dynamiques ont été le résultat des estimations de l'équation ECM

$$DLYC_t = \beta_0 + \lambda \cdot ECM_{t-1} + \sum_{i=1}^{h} \beta_i \cdot DLYC_{t-i} + \sum_{j=o}^{k} \beta_i \cdot DLMC_{t-j} + \epsilon_t$$

avec $ECM_{t-1} = (LYC_{t-1} - LMC_{t-1})$, où l'hypothèse nulle de λ ne fut jamais rejetée.

Les modèles retenus pour chaque période sont dans le Tableau 5[37]. Les étoiles ont la signification normale. Nous donnons l'exemple de la première équation (1856-1890). L'écart-type de l'estimation (σ) est de 2,8%. Les tests LM, AR(1) et ARCH(1), conduisent à accepter l'hypothèse nulle d'absence d'autocorrélation et de processus ARCH, d'ordre 1, des erreurs. L'hypothèse nulle de distribution normale[38] des erreurs ne peut, elle aussi, être rejetée. Finalement nous avons appliqué des tests DF[39] aux résidus qui confirment l'hypothèse de stationnarité. Nous avons également obtenu l'équation de longue période,

$$DLYC = C\_LP + \beta\_LP \cdot DLMC$$

et appliqué un test de Wald pour la nullité des coefficients de cette dernière équation. Le test (type Wald) de restriction $\beta\_LP = 1$, conduit à rejeter cette hypothèse.

Quelles leçons pourrons-nous retirer de ces trois équations? Les tests usuels confirment l'absence de problèmes avec les résidus. Le comportement du produit nominal pendant les périodes de l'étalon-or (Eq 1)[40] et de son influence (Eq 2) rejettent l'hypothèse d'une élasticité unitaire du produit ($\beta\_LP \neq 1$). L'étude du comportement jusqu'en 1913 (Eq 2) doit tenir compte de la stabilité du modèle, dans la mesure où il couvre deux périodes avec des règles d'émission très différentes.

---

37 DLYC{1} est équivalent à DLYCt-1. On n'a retenu aucune estimation pour la dernière période car on n'a pas pu rejeter l'hypothèse nulle des coefficients du produit et de la monnaie.

38 Test de Jarque-Bera proposé par (Doornik et Hansen, 1994).

39 Dickey-Fuller.

40 Voir le Tableau 5.

Tableau 5 – Modèles à une équation	Eq 1	Eq 2	Eq 3
	1856-1890	1856-1913	1949-1998
Const.	0,15**		
	(2,64)		
DLYC {1}		0,260**	0,430***
		(2,22)	(3,78)
DLYC {2}			0,419***
			(3,78)
DLCY {3}			
DLYC {4}			
DLMC		0,170**	
		(2,53)	
DLMC {1}	0,187**	0,189**	
	(1,86)	(2,62)	
DLMC {4}			0,155*
			(1,84)
σ	0,028	0,024	0,040
AR (1)	0,163	2,542	0,145
ARCH (1)	2,274	1,721	0,007
N (J-B)	2,054	2,431	1,47
ADF (t)	-5,24***	-7,67***	-7,17***
ADF (Z)	-31,72***	-58,43***	-50,63***
C_LP	0,015		
β_LP	0,187*	0,485***	1,027***
Wald	3,458*	21,784***	11,423***
β_LP = 1	66,874***	12,885***	0,008

Dans la Figure 2[41], nous avons l'évolution des trois types de tests de Chow proposés par (Hendry et Doornik, 2001)[42] et calculés pour la période après la chute de l'étalon-or. Comme on peut le constater, ces tests éliminent toute hypothèse d'instabilité du modèle associé à des changements dans les coefficients. Ainsi, pour cette période, 1856-1913, nous pouvons représenter le comportement du produit nominal par une seule équation qui est stable pour toute la période, et nous ne pouvons accepter l'élasticité unitaire du produit par rapport à l'offre de monnaie.

Le comportement du produit nominal pour la période 1949-1998, donné par la troisième équation (Eq 3), ne rejette pas la présence d'une élasticité unitaire par rapport à l'offre de monnaie. Comme, pendant cette période, nous avons connu des changements considérables du point de vue de la politique monétaire et de change, nous avons aussi fait l'estimation des tests de Chow (Figure 3).

41 Les valeurs sont standardisées. Ce qui signifie que, si elles dépassent l'unité, on doit refuser la stabilité des coefficients au niveau de 5%.

42 (Hendry et Doornik, 2001), p. 254: «1-Step Chow test», «Break-point Chow test» et «Forecast Chow test».

Figure 2 - Tests de Chow appliqués à l'équation 2

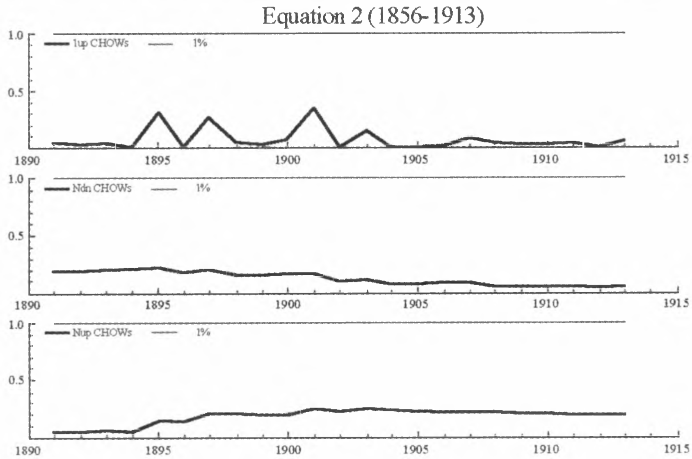
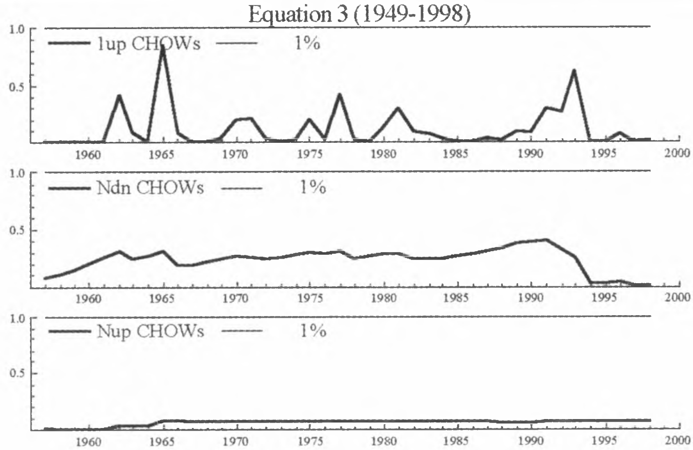


Figure 3 - Tests de Chow appliqués à l'équation 3



Nous pouvons voir que l'équation proposée est parfaitement stable depuis 1957 jusqu'à 1998.

Les conclusions que l'on peut retirer d'après ces trois équations sont nettes: la théorie quantitative du produit nominal ne peut être acceptée ni pour la période de l'étalon-or ni pour la période de forte globalisation avant la Première Guerre Mondiale. Par contre, elle s'applique à la période allant de la Seconde Guerre Mondiale à nos jours.

Nous passons maintenant à l'étude des modèles multivariés obtenus pour les variables non-stationnaires employant la méthodologie de Johansen.

Divers essais sur d'éventuelles relations de co-intégration nous ont conduit au non rejet de cette hypothèse pour les périodes de 1949-1998 ainsi que pour 1914-1998.

Notre recherche d'un vecteur de co-intégration fut conduite jusqu'à 4 retards, même si nous pensons qu'un nombre aussi élevé de retards n'est pas réaliste. Nous présenterons les résultats pour 1949-98 et pour 1914-98. Les résultats sont dans les Tableaux 6, 7 et 8, où, entre parenthèses, nous avons les valeurs de l'écart-type. On présentera les tests pour chaque période.

1949-1998

Le seul cas admis pour une relation de co-intégration inclut 4 retards et rejette la présence de constante dans le VAR associé[43] (1949-98 (RC)). Dans le Tableau 6, sont présentés les résultats de la statistique de la trace. Contrairement à ces résultats, les statistiques de la valeur propre maximale rejettent l'hypothèse d'un vecteur de co-intégration.

Tableau 6 – Co-intégration entre LYC et LMC

	L		λ	Trace	n.s. 5%
1949-98 (RC)	4	$r = 0$	0.236	20.91	19.96
		$r \leq 1$	0.139	7.47	9.24
1914-98	3	$r = 0$	0.157	15.05	12.53
		$r \leq 1$	0.006	0.55	3.84
1914-98 (RC)	4	$r = 0$	0.209	24.29	19.96
		$r \leq 1$	0.050	4.36	9.24

Tableau 7 – Coefficients des Modèles Normalisés

Min	β	Const	$\beta = 1$	α_{DLYC}	α_{DLMC}
1949-98 (RC)	-1.018	-0.406	0.033	-0.033	0.100
	(0.049)	(0.398)	0.855	(0.028)	(0.036)
1914-98	-1.009		0.033	-0.008	0.039
	(0.035)		0.855	(0.015)	(0.012)
1914-98 (RC)	-0.989	-0.344	0.043	-0.026	0.049
	(0.052)	(0.490)	0.836	(0.019)	(0.015)

43 Qu'on désignera dorénavant par VECM, dans la mesure où l'on a un VAR avec un ECM.

Tableau 8 – Equations et tests LM aux résidus avec $\beta = 1$

1949-98 (RC)	DLYC	AR 1-1 test	F (1,40) =	0.254 [0.617]
	DLMC	AR 1-1 test	F (1,40) =	1.090 [0.303]
	DLYC	Normality	$\chi_2^2 =$	2.352 [0.308]
	DLMC	Normality	$\chi_2^2 =$	11.066 [0.004]***
	DLYC	ARCH 1-1	F (1,39) =	0.035 [0.853]
	DLMC	ARCH 1-1	F (1,39) =	0.213 [0.647]
1914-98	DLYC	AR 1-1	F (1,77) =	4.121 [0.046]*
	DLMC	AR 1-1	F (1,77) =	0.996 [0.322]
	DLYC	Normality	$\chi_2^2 =$	5.046 [0.080]
	DLMC	Normality	$\chi_2^2 =$	18.095 [0.0001]**
	DLYC	ARCH 1-1	F (1,76) =	1.288 [0.260]
	DLMC	ARCH 1-1	F (1,76) =	0.017 [0.897]
1914-98 (RC)	DLYC	AR 1-1	F (1,75) =	0.094 [0.760]
	DLMC	AR 1-1	F (1,75) =	2.566 [0.113]
	DLYC	Normality	$\chi_2^2 =$	8.452 [0.014]*
	DLMC	Normality	$\chi_2^2 =$	22.880 [0.000]**
	DLYC	ARCH 1-1	F (1,74) =	0.001 [0.937]
	DLMC	ARCH 1-1	F (1,74) =	2.438e-005 [0.996]

La méthode de Johansen nous conduit à accepter la relation de longue période pour le produit nominal et l'offre de monnaie (Tableau 7):

$$LYC = 0{,}406451 + 1{,}018027 \cdot LMC$$

Le test LR de contrainte du coefficient de LMC égale à l'unité ne rejette pas cette hypothèse ($\chi_1^2 = 0.033$, pour un niveau de significativité de 85.5%, Tableau 7). Le seul problème posé par le modèle VECM avec contrainte d'élasticité unitaire (Tableau 8) concerne la normalité[44] des résidus de l'équation de l'offre de monnaie. Mais ce problème n'est pas grave (Johansen, 1995: 29).

En conclusion, pour cette période après la II GM, nous pouvons accepter l'application de la théorie du revenu nominal à l'économie portugaise. L'ajustement du produit à la situation de l'équilibre est plus lent que celui de l'offre de monnaie ($|-0.033|<0.100$, Tableau 7). Dans les deux cas, le processus d'ajustement se fait très lentement.

1914-1998

Nous avons obtenu deux relations de co-intégration pour cette période. La première était préférable si l'on tient compte des valeurs des indicateurs d'information Akaike et Schwarz. Mais l'absence de constante dans une fonction du produit nominal n'est pas satisfaisante[45]. Par ailleurs, la deuxième relation, qui inclut une constante dans l'espace de co-intégration, conduit à

44 Test de Jarque-Bera (Doornik et Hansen, 1994).

45 Cette constante traduira la croissance tendancielle.

un modèle VECM sans problèmes d'auto corrélation des résidus. Tant dans le cas de la première relation de co-intégration que dans la seconde, le test de la valeur propre maximale conduit à accepter l'existence d'un vecteur de co-intégration.

La méthode de Johansen conduit aux deux relations (Tableau 11 et 16) de longue période:

$$LYC = 1{,}009169 \cdot LMC, \text{ et}$$

$$LYC = 0{,}343706 + 0{,}989051 \cdot LMC$$

Nous avons déjà suggéré que nous préférons la deuxième relation. Plusieurs arguments lui sont favorables: le fait que la vitesse de circulation de la monnaie n'est pas égale à 1; l'absence de problèmes d'auto corrélation des erreurs dans les deux équations VECM[46] du deuxième modèle (Tableau 8); et les valeurs des coefficients d'ajustement des deux variables à l'équilibre de longue période sont bien irréalistes dans le premier modèle (Tableau 7). Pourtant, les coefficients d'ajustement à l'équilibre du deuxième modèle (Tableau 7) sont très faibles, ce qui signifie des vitesses d'ajustement très lentes, surtout pour le cas du produit nominal.

Nous devons insister sur le fait que que ces deux relations (Tableau 7) ne rejettent pas l'hypothèse d'une élasticité unitaire du produit par rapport à la monnaie.

6. Conclusion

Nous soutenons une théorie monétaire quantitativiste du revenu nominal sans que cette position nous oblige à accepter la neutralité de la monnaie. C'est-à-dire, nous acceptons l'idée de non neutralité de la monnaie et le quantitativisme du produit nominal. L'indétermination des variations réelles et des prix est bien présente dans Chaîneau et dans certaines observations de Friedman[47]. Cette même indétermination est au cœur des analyses autrichiennes.

Après l'introduction, où nous avons brièvement présenté nos idées sur la question de la neutralité de la monnaie, sur les différences entre l'expérimentation individuelle et de marché ainsi que sur l'importance du quantitativisme du produit nominal, nous avons tenté de diviser la période s'étendant de 1854 à 1998, en tenant compte des comportements des autorités monétaires portugaises et des événements mondiaux. Ensuite, nous avons essayé d'appliquer des techniques d'analyse empirique pour confirmer la division en périodes homogènes, que ce soit du point de vue du comportement du produit nominal ou de celui de l'offre de monnaie. L'analyse empirique identifie parfaitement les périodes-clés de notre histoire (voir les Tableaux 2 et 3).

Mais si le choix d'une (seule) période après la Seconde Guerre Mondiale ne peut pas être considéré comme surprenant, nous ne pouvons pas en dire autant du choix d'une période qui regroupe l'étalon-or et la période suivante d'inconvertibilité jusqu'à la Première Guerre Mondiale. Cependant, les données statistiques pour les différentes périodes nous conduisaient à ce choix et, en outre, le modèle à une équation a confirmé ce choix. Nous pensons que la période de globalisation qui a caractérisé cette période, depuis la moitié du XIXe siècle jusqu'à la Première Guerre Mondiale est responsable de l'identification de cette période comme constituant une période homogène[48]. Le taux de change de la livre n'a pratiquement pas changé de 1892 à 1913 (avec une petite appréciation de la monnaie portugaise)[49]. Pendant la même période, l'évolution

46 Le critère d'absence d'autocorrélation fut même indiqué par (Johansen, 1995: 21), pour décider du choix de la dimension du modèle.

47 Le comportement de non neutralité dans la courte période fut surtout exploité par les keynésiens. Voir l'exemple de (Weintraub et Weintraub, 1972).

48 Il est vrai que nous avons substitué les valeurs du taux de variation de l'offre de monnaie de 1891 et 1892 par des valeurs estimées avec un modèle ARIMA. Mais il ne s'agit que de deux observations.

49 (Mata et Valério, 1993: 261). Voir aussi pour l'économie internationale (Bordo et Schwartz, 1995) et les références qui y sont faites.

des prix au Portugal fut pratiquement égale à celle de l'Angleterre. Nous pouvons admettre que l'économie portugaise était intégrée dans une économie mondiale où les prix, et leur évolution, étaient connus et, par conséquent, ne dépendaient pas de l'offre de monnaie, comme le suggère la théorie quantitative du produit nominal.

La période depuis la Seconde Guerre Mondiale (plus exactement depuis 1949) avait été identifiée par une rupture dans l'après première crise du pétrole et la Révolution de 1974. Les modèles que nous avons retenus pour l'après Seconde Guerre Mondiale supportent bien cette rupture dans la mesure où les coefficients sont stables.

Les modèles estimés pour la période 1856-90 et 1856-1913 rejettent l'hypothèse d'une élasticité unitaire du produit nominal par rapport à l'offre de monnaie, tandis que le modèle estimé pour la période 1949-98 ne rejette pas cette élasticité. Pour cette dernière période, nous avons obtenu une relation de co-intégration par la méthode de Johansen. Le vecteur obtenu ne rejette pas l'hypothèse d'une élasticité unitaire. Le modèle VECM associé présente des valeurs réduites d'ajustement des variables à l'équilibre. L'ajustement du produit nominal est particulièrement lent. Ce résultat confirme l'idée des économistes monétaristes de l'existence d'ajustements lents aux variations de l'offre de monnaie. Nous avons aussi pu constater l'existence d'une relation de co-intégration pour la période de 1914-98. Pour les deux cas dans lesquels elle peut être obtenue, nous avons vu qu'on ne rejette pas l'hypothèse d'une élasticité unitaire et que les ajustements à l'équilibre sont encore plus lents, surtout pour le cas du produit nominal.

En conclusion, pour les périodes s'étendant de 1914 et 1949 à 1998, nous ne pouvons rejeter la théorie quantitative du produit nominal de la monnaie dans l'économie portugaise: une variation de l'offre de monnaie conduit à une variation (relative) identique du produit nominal. Pendant le régime d'étalon-or et celui de convertibilité et d'inconvertibilité jusqu'en 1913, la théorie quantitative du produit nominal est parfaitement rejetée: une variation de l'offre de monnaie ne conduit pas à une variation (relative) identique du produit nominal. Nous pensons que la différence des résultats réside dans le fait que, dans un régime de monnaie inconvertible, la monnaie n'a qu'une nature nationale et, par conséquent, l'offre de monnaie détermine la valeur de la demande nominale dans l'espace national. Tandis que, dans un régime de monnaie convertible, la chaîne causale doit partir de l'offre mondiale de métal vers la valeur du produit mondial. Le cas d'inconvertibilité avant la Première Guerre Mondiale correspond au cas où l'économie portugaise faisait partie intégrante d'une économie mondiale dans laquelle le produit nominal était déterminé par l'offre mondiale d'or.

Dans la logique de M. Friedman et d'A. Chaîneau, il faut alors procéder à l'étude du partage entre effets purement réels et effets prix.

Bibliographie

Andrade, J. S. (2000), Un Essai d'Application de la Théorie Quantitative de la Monnaie à l'Economie Portugaise, 1854 – 1998, *Estudos do GEMF*, 4, Faculdade de Economia da Universidade de Coimbra

Banco de Portugal (plusieurs années), Relatório do Conselho de Administração, Lisboa.

Banco de Portugal (2002), Vol. I – Séries Estatísticas: versão revista e prolongada para 1994 e 1995, http://www.bportugal.pt/publish/serlong/s1_p.htm.

Bai, J.; Perron, P. (1998), Estimating and Testing linear Models with Multiple Structural Changes, *Econometrica*, 66, 47-78.

Bai, J.; Perron, P. (2001a), *Computation and Analysis of Multiple Structural Change Models*, Working Paper, Boston University, Boston.

Bai, J.; Perron, P. (2001b), *Multiple Structural Change Models: A Simulation Analysis*, Working Paper, Boston University, Boston.

Batista, D., *et al.* (1997), New Estimates for Portugal's GDP, 1910-1958, *História Económica, Banco de Portugal*, 7.

Bernanke, B. S.; Mihov, I. (1998), The Liquidity Effect and Long-run Neutrality, *Carnegie--Rochester Conference series on Public Policy*, 49, 149-94.

Bhargava, A. (1986), On the Theory of Testing for Unit Roots in Observed Time Series, *Review of Economic Studies*, LIII, 369-84.

Bodin, J. (1568), *La Réponse de Jean Bodin à M. de Malestroit*, Paris, Édition Henri Hauser, A. Colin (1932).

Bordo, M. (1993), The Gold Standard, Bretton Woods and Other Monetary Regimes: A Historical Appraisal, *FRB of St.L. Economic Review*, 75, 2, 123-91.

Bordo, M.; Jonung, L. (1997), A Return to the Convertibility Pronciple? Monetary and Fiscal Regimes in Historical perspective, the International Evidence, in Leijonhufvud, A. (ed.), *Monetary Theory as a Basis for Monetary Policy*, London, Macmillan.

Bordo, M.; Jonung, L. (2001), A Return to the Convertibility Principle? Monetary and Fiscal Regimes in Historical Perspective, The International Evidence, *Economic Papers*, 159.

Bordo, M.; Redish, A. (1993), Maximizing Seignorage Revenue during Temporary Suspensions of Convertibility: A Note, *Oxford Economic Papers*, 45,1, 157-68.

Bordo, M.; Rockoff, H. (1996), The Gold Standard as a Good Housekeeping Seal of Approval, *Journal of Economic History*, 56, 2, 389-428.

Bordo, M.; Santos, F. T. (1995), «Portugal and the Bretton Woods International Monetary System», in Reis, J. (ed.), *International Monetary System in Historical Perspective*, New York, Macmillan, 181-208.

Bordo, M.; Schwartz, A. (1995), Taxas de Câmbio Fixas Enquanto Regra Contigente em Países Centrais e Periféricos – evidência do período 1880-1990, in Macedo, J. B. et al. (eds.), *Convertibilidade Cambial, Conferência Comemorativa do 140º Anivertsário da Adesão de Portugal ao Padrão-Ouro*, Lisboa, Banco de Portugal – F.L.A.D., 29-110.

Brainard, W.; Perry, G. (2000), Making Policy in a Changing World, in Perry, G.; Tobin, J. (eds.), *Economic Events, Ideas, and Policies: The 1960s and After*, Washington, Brookings Institution Press.

Campbell, J.; Mankiw, G. (1987), Are Output Fluctuations Transitory, *Quarterly Journal of Economics*, 102, 4, 857-80.

Cantillon, R. (1755), *Essai sur la Nature du Commerce en Général*, Paris, I.N.E.D. (1952).

Cassel, G. (1928), The Rate of Interest, the Bank Rate, and the Stabilization of Prices, *Quarterly Journal of Economics*, 42, 511-29.

Cassel, G. (1930), The Supply of Gold, *Interim Report of the Gold Delegation of the Financial Committee*, League of Nations, Geneva.

Chaîneau, A. (1964), *La Demande d'Encaisses Monétaires*, Pari, Cujas.

Chaîneau, A. (1992), *Manuel de Marcro-Économie Élémentaire*, Paris, Armand Colin.

Chaîneau, A. (1995), Le Bouclage du Circuit Économique, *Estudos do GEMF,* Faculdade de Economia da Universidade de Coimbra

Chaîneau, A. (1996), Équilibre, Déséquilibre, Monnaie, in ISEG (ed.), *Ensaios de Homenagem a Manuel Jacinto Nunes*, Lisboa, ISEG-UTL, 191-211.

Cochrane, J. (1988), How Big is the Random Walk in GNP?, *Journal of Political Economy*, 96, 5, 893-920.

Cogley, T.; Sargent, T. (2002), Evolving Post-World War II Inflation Dynamics, in Bernanke, B.; Rogoff, K. (eds.), *NBER Macroeconomics Annual 2001*, Cambridge, Ma., The M.I.T. Press.

Córdova, J. E. L.; Meissner, C. (2000), *Exchange Rate Regimes and International Trade: Evidence from the Classical Gold Standard Era*, CIDER, University of California at Berkeley.

Doornik, J. A.; Hansen, H. (1994), *A practical test for univariate and multivariate normality*, Nuffield College, University of Oxford.

Eichengreen, B.; Flandreau, M. (1997), *The Gold Standard in Theory and History*, London, Routledge.

Esteves, R. P.; Ferramosca, F. (2000), *O Mecanismo dos Câmbios em Padrão-Ouro. Estabilidade Cambial e Violações dos Pontos de Ouro, 1854-1891*, WP, FEUP, Faculdade de Economia da U. Porto.

Flandreau, M.; Maurel, M. (2001), *Monetary Union, Trade Integration and Business Fluctuations in 19th Century Europe: Just Do It*, CEPR, London.

Frieden, J. A. (1997), The Dynamics of International Monetary Systems: International and domestic Factors in the Rise, Reign, and Demise of the Classical Gold Standard (1992), in Eichengreen, B.; Flandreau, M. (ed.), *The Gold Standard in Theory and History*, London, Routledge, 207-27.

Friedman, M. (1956), The Quantity Theory of Money: a restatement, in R. Clower (ed.), *Monetary Theory, 1969*, Hardmondsworth, Penguin Books, 94-111.

Friedman, M. (1973), A Theoretical Framework for Monetary Analysis, *Journal of Political Economy*, 78, (1970), in A. Walters (ed.), *Money and Banking*, Harmondsworth, Penguin Books, 69-104.

Fritsch, W.; Franco, G. H. (2000), Brazilian Experience with the Gold Standard, in Aceña, P. M.; Reis, J. (eds.), *Monetary Standards in the Periphery, Paper, Silver and Gold, 1854-1933*, New York, St. Martin's Press.

Georgescu-Roegen, N. (1966), *Analytical Economics, issues and problems*, Cambridge, Ma., Harvard University Press.

Hall, R. (1999), Comment on “Rethinking the Role of the NAIRU in Monetary Policy: implications of model formulation and uncertainty” by Arturo Estrella and Frederic Mishkin, in Taylor, J. (ed.), *Monetary Policy Rules*, Chicago, NBER, University of Chicago Press.

Hayek, F. (1936), *La Teoria Economica y el Ciclo Economico*, Madrid, Espassa-Calpe.

Hayek, F. (1978), *Denationalisation of Money, the Argument Refined*, London, Institute of Economic Affairs.

Hendry, D.; J. Doornik, J. (2001), *Empirical Econometric Modelling Using PcGive 10*, London, Timberlake Consultants.

Hodrick, R.; Prescott, E. (1980), *Post-War U.S. Business Cycles: An Empirical Investigation*, Carnegie-Mellon.

Hogendorn, C. (1998), Capital Mobility in Historical Perspective, *Journal of Policy Modeling*, 20, 2, 141-61.

Humphrey, T. (2002), Knut Wicksell and Gustav Cassel on the Cumulative Process and the Price-Stabilizing Policy Rule, *FRB of Richmond, Economic Quarterly*, 88, 3, 59-83.

Johansen, S. (1995), *Likelihood-Based Inference in Cointegrated Vector Auto-Regressive Models*, Oxford, Oxford University Press.

Kaldor, N. (1986), *The Scourge of Monetarism (1980)*, Oxford, Oxford University Press.

Keynes, J. M.; Bainville, J. (2002), *Les Conséquences Economiques de la Paix (1920) et Les Conséquences Politiques de la Paix (1920)*, Paris, Gallimard.

Kwiatowski, D., *et al.* (1992), Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root: How Sure Are We That Economic Time Series Have a Unit Root?, *Journal of Econometrics*, 54, 1-3, 159-78.

Lange, O. (1945), *Price Flexibility and Full Employment*, Westport, Greenwood Press (1978).

Leontief, W. (1936-7), The Fundamental Assumption of Mr. Keynes' Monetary Theory of Unemployment, *Quarterly Journal of Economics*, LI.

Wu Liu, L. S.; Zidek, J. V. (1997), On Segmented Multivariate Regressions, *Statistica Sinica*, 7, 497-525.

Mankiw, N. G. (2000), *The Inexorable and Mysterious Tradeoff between Inflation and Unemployment*, Harvard University, Cambridge, MA.

Mata, E. (1991), «Exchange Rate and Exchange Policy in Portugal 1891-1931 Revisited», *Estudos de Economia*, XII, 2, 33-41

Mata, E.; Valério, N. (1993), *História Económica de Portugal*, Lisboa, Editorial Presença.

Mata, E.; Valério, N. (1995), Estabilidade monetária, Disciplina Orçamental e Desempenho Económico em Portugal desde 1854, in Macedo, J. B. et al. (eds.), *Convertibilidade Cambial, Conferência Comemorativa do 140º Anivertsário da Adesão de Portugal ao Padrão-Ouro*, Lisboa, Banco de Portugal-F.L.A.D., 259-84.

Meissner, C. (2002), A New World Order: Explaining the Emergence of the Classical Gold Standard, *WP9233*, NBER.

Mendonça, A. (1996), The Introduction of the Gold Standard in Portugal, in ISEG (ed.), *Ensaios de Homenagem a Manuel Jacinto Nunes*, Lisboa, ISEG, UTL, 73-98.

Mises, L. (1912), *The Theory of Money and Credit*, London, Jonathan Cape (1934).

Mises, L. (1949), *Human Action, a treatise on economics*, N. Haven, Yale University Press.

Neumann, M. (1993), Commentary (Michael Bordo, The Gold Standard, Bretton Woods and Other Monetary Regimes: A Historical Appraisal), *FRBStL, Economic Review*, March/April, 192-9.

Nunes, A. B. et al. (1989), «Portuguese Economic Growth, 1833-1985», *Journal of European Economic History*, 18, 2, 291-330.

Patinkin, D. (1965), *Money, Interest and Prices*, New York, Harper & Row.

Patinkin, D. (1987), Neutrality of Money, in Eatwell, J. et al. (eds.), *The New Palgrave: a Dictionary of Economics*, London, Macmillan, pp. 639-45.

Perron, P. (1997), Further Evidence on Breaking Trend Function in Macroeconomic Variables, *Journal of Econometrics*, 80, 355-85.

Pinheiro, M., et al. (1998), *Séries Longas para a Economia Portuguesa*, Lisboa, Banco de Portugal.

Pivetta, F.; Reis, R. (2002), *The Persistence of Inflation in the United States*, Harvard University Press..

Reis, J. (1991), A Evolução da Oferta Monetária Portuguesa 1854-1912, *História Económica*, 3, Banco de Portugal

Reis, J. (2002), An "Art", not a "Science"? Central Bank Management in Portugal under the Gold Standard, 1854-1891, *WP 6-02*, Banco de Portugal.

Ricardo, D. (1810-11), «The High Price of Bullion, a Proof of the Depreciation of Bank Notes», in P. Sraffa (ed.), *The Works and Correspondence of David Ricardo*, Cambridge, C.U.P., (1951), 47-127.

Rose, A. K. (2000), One Money, One Market: Estimating the Effect of Common Currencies on Trade, *Economic Policy*, 30, 7-45.

Russell, H. B. (1898), *International Monetary Conferences*, New York, Harper and Brothers.

Santos, F. T. (1995), O Último a Aderir ao Padrão-Ouro, Julho-Setembro 1931, in Macedo, J. B. et al. (eds.), *Convertibilidade Cambial, Conferência Comemorativa do 140º Aniversário da Adesão de Portugal ao Padrão-Ouro*, Lisboa, Banco de Portugal-F.L.A.D., 231-48.

Sargent, T. (1976), The Observational Equivalence of Natural and Unnatural Rate Theories of Macroeconomics, *Journal of Political Economy*, 84, 3, 631-40.

Schwartz, A. (1992), *Monetarism and Monetary Policy*, London, Institute of Economic Affairs.

Stock, J. (1991), Confidence Intervals for the Largest Autoregressive Root in U. S. Macroeconomic Time Series, *Journal of Monetary Economics*, 28, 435-59.

Subercaseaux, G. (1922), *Monetary and Banking Policy of Chile*, Oxford, Clarendon Press.

Taylor, J. (1998), Monetary Policy Guidelines for Unemployment and Inflation Stability, in Solow, R.; Taylor, J. (eds.), *Inflation, Unemployment and Monetary Policy*, Cambridge, Ma., The M.I.T. Press.

Taylor, J. (2000), Low Inflation, Pass-through, and the Pricing Power of Firms, *European Economic Review*, 44, 7, 1389-408.

Triffin, R. (1997), The Myth and Realities of the So-called Gold Standard (1968), in Eichengreen, B.; Flandreaued, M. (eds.), *The Gold Standard in Theory and History*, London: Routledge, 140-60.

Weintraub, S.; Weintraub, R. (1972), The Full Employment Model: A Critique, *Kyklos*, 25, 1, 83-100.

Valério, N. (1991), Periodozação da História Monetária de Portugal, *Estudos de Economia*, XII, 2, 3-18

Wheelock, D.; Bordo, M. (1998), Price Stability and Financial Stability: The Historical Record, *F.R.B. of St. Louis Review*, 98, 5, 41-62.

Wicksell, K. (1898), *Interest and Prices*, London, Macmillan (1965).

Wicksell, K. (1906), *Lectures on Political Economy, T. II*, London, Routledge and Sons. (1934).

Annexe: Les données utilisées dans l'analyse

Les données statistiques que nous avons utilisées ont été empruntées à plusieurs publications (Tableau 9). Les valeurs dans (Pinheiro, et al, 1998) ont été actualisées par la Banque du Portugal (Banque du Portugal, 2002). La crise provoquée par la chute de l'étalon-or en 1891 a conduit à des valeurs non normales pour l'offre de monnaie. En 1891, son taux de croissance fut inférieur à -70% et, en 1892, il atteignit presque + 20%. Nous avons corrigé ces valeurs du taux de croissance par des prévisions avec un modèle ARIMA(1,1,2) pour des données allant de 1855 à 1889. Dans certains cas, nous avons fait l'homogénéisation des séries avant 1953, par l'application des taux de croissance.

Les valeurs des indices des prix à la consommation pour l'économie américaine et anglaise sont de (Wheelock et Bordo, 1998). Nous les avons actualisées avec des données du Bureau of Labor Statistics (http://stats.bls.gov/) et de la page des National Statistics on Line (http://www.statistics.gov.uk/).

Tableau 9 – Sources des données statistiques

PIB	
1854-1909	(Nunes et al., 1989)
1910-1952	(Batista et al., 1997)
1953-1993	(Pinheiro et al., 1999)
1994-1998	(Banco de Portugal, 1996; 1998; 2000)
M1	
1854-1912	(Reis, 1991)
1913-1946	(Mata et Valério, 1993: Tableau E.1.3)
1947-1993	(Pinheiro et al., 1998)
1994-1998	(Banco de Portugal, 1996; 1998; 2000)
Population	
1854-1952	(Nunes et al., 1989)
1953-1993	(Pinheiro et al., 1998)
1994-1998	(Banco de Portugal, 1996; 1998; 2000)

Human capital, mechanisms of technological diffusion and the role of technological shocks in the speed of diffusion. Evidence from a panel of Mediterranean countries

Maria Adelaide Duarte / Marta Simões
FEUC

resumo

Este artigo procura determinar a importância do capital humano para a difusão tecnológica num conjunto de países da Bacia Mediterrânica tendo por base as previsões da teoria do crescimento endógeno. Os resultados obtidos sugerem que a especificação não linear proposta por Benhabib and Spiegel (2002), que contempla quer a existência de convergência tecnológica quer de clubes de convergência, não se aplica a estes países, mas sim a especificação linear originalmente proposta por Benhabib and Spiegel (1994), com especial relevância do capital humano para as actividades de inovação tecnológica. Testámos também a possibilidade de complementaridade entre o capital humano e o IDE e a importância do primeiro para a difusão das TIC, mas apenas esta última hipótese prevaleceu. Para terminar, analisámos ainda a importância dos choques tecnológicos em termos de difusão tecnológica utilizando um modelo VAR, verificando-se a existência de complementaridade entre a PTF, o investimento em capital físico e o capital humano na absorção de qualquer um dos choques considerados.

résumé / abstract

Cet article essaie d'établir l'importance du capital humain dans la diffusion technologique pour un ensemble de pays du bassin méditerranéen, ayant pour base la théorie de la croissance endogène. Les résultats obtenus suggèrent que ce n'est pas la spécification non-linéaire proposée par Benhabib et Spiegel (2002) – recouvrant l'existence soit de la convergence technologique soit de clubs de convergence – qui s'applique à ces pays, mais la spécification linéaire originalement proposée par Benhabib et Spiegel (1994), surtout en ce qui concerne le capital humain dans les activités d'innovation technologique. La possibilité de complémentarité du capital humain et de l'IDE ainsi que l'importance du premier dans la diffusion des TIC a été testée, mais à peine la deuxième hypothèse a été confirmée. Finalement, à l'aide d'un modèle VAR, on analyse l'importance des chocs technologiques en termes de diffusion de technologies et le résultat confirme l'existence d'une complémentarité entre la PTF, l'investissement en capital physique et le capital humain dans l'absorption des chocs considérés.

The purpose of this paper is to assess the importance of human capital as a facilitator of technological diffusion in a sample of developing Mediterranean countries based on the predictions of endogenous growth theory. The evidence does not support the Benhabib and Spiegel (2002) non-linear specification that accommodates both the hypothesis of technological convergence and convergence clubs but the linear specification originally proposed by Benhabib and Spiegel (1994), confirming a role for human capital in both innovation and imitation activities. We also tested the complementarity between FDI, a form of embodied technology diffusion, and human capital but this hypothesis was not confirmed and investigated the importance of human capital for the diffusion of ICT confirming it is fundamental to benefit from these technologies. Finally, we analysed the importance of technological shocks for technological diffusion using a VAR model finding evidence of factor complementarity between TFP, physical capital investment and human capital in the absorption of any of the shocks considered.

1. Introduction

The purpose of this paper is to empirically assess the importance of human capital as a facilitator of technological diffusion in a sample of seven developing Mediterranean countries between 1960-2000. Developing Mediterranean countries are geographically and economically[1] close to developed European countries and their economic prosperity, together with their subsequent political stability, is of major importance to European countries. Knowing the sources of growth in these countries can help us derive implications for policies that promote growth and convergence in the Southern Mediterranean countries.

Productivity differences are an important source of income differences across countries (see for instance, Easterly and Levine, 2001; Hall and Jones, 1999; Klenow and Rodriguez-Clare, 1997), so in order to understand growth the sources of productivity growth across countries have to be identified. Endogenous growth theory provides the theoretical explanations for technological progress which is responsible for sustainable long-term growth. For instance, the models of Romer (1990a), Romer (1990b), and Aghion and Howitt (1992) emphasize the importance of domestic innovation initiatives. Technological change is the result of the activity of an R&D sector that primarily uses the existing stock of knowledge and human capital as inputs. Nelson and Phelps (1966), Grossman and Helpman (1992) and Barro and Sala-i-Martin (1997) on the other hand, stress the importance of imitation activities or technology diffusion in technological change: countries further away from the technological frontier have a potential for faster growth by adopting the inventions developed elsewhere in the world. There are, however, cross-country differences in the effectiveness with which countries adopt foreign technologies, which in turn determine income differences. Nelson and Phelps (1966) stress the importance of human capital in narrowing the gap between the level of technology in practise and the theoretical level of technology[2].

The roles played by human capital, either as an input in innovation activities or as a facilitator of the diffusion of technology from abroad, have been extensively analysed and confirmed by empirical growth literature. The benchmark study in this area is probably Benhabib and Spiegel (1994) but many other examples can be found, some of which are specific to developing countries (see for instance Coe, Helpman and Hoffmaister (1997), Mayer (2001), Engelbrecht (2002), Dowrick and Rogers (2002), and Papageorgiou (2003)). In the developing Mediterranean countries that are involved in little R&D[3], technology diffusion is probably the most important source of productivity growth. It is therefore important to know whether the human capital stocks available in the developing Mediterranean countries constitute a break or promote technological convergence in these countries, in which case we would be in the presence of convergence clubs or poverty traps. The empirical growth literature either concentrates on wide samples of countries that include developed and developing countries, making it difficult to uncover specificities in certain groups of countries, or focuses on more restricted groups of countries, such as the OECD countries (see Temple (2001) for a review of empirical studies on human capital and growth focusing on OECD countries), Latin American countries (e.g., Easterly, Loyaza and Montiel

1 In November 1995, in Barcelona, there was an Euro-Mediterranean Conference of the Ministers of Foreign Affairs that marked the starting point of the Euro-Mediterranean Partnership (the Barcelona Process), a wide framework of political, economic and social relations between the Member States of the European Union and Partners of the Southern Mediterranean. This partnership led to the Euro-Mediterranean Association Agreements, whose main aim is the gradual establishment of a free-trade area.

2 “To put the hypothesis simply, educated people make good innovators, so that education speeds up the process of technology diffusion.” Nelson and Phelps (1966), p. 70.

3 According to the World Development Indicators 2000, in the 1990s Egypt spent a maximum of 0.22% of its GNP on R&D, Cyprus 0.18%, Syria 0.2%, Tunisia 0.33% and Turkey 0.53%. Only Israel spent a amount of its GNP on R&D comparable to the levels spent by the developed countries, which reached 2.35% in 1997.

(1997)), or Sub-Saharan African countries (e.g., Easterly (1997), Hoeffler (2000)). Few, however, focus specifically on developing Mediterranean countries and none to our knowledge deal with the issue of human capital and technology diffusion, a gap we would like to fill.

To assess the importance of human capital in technological change and, in greater depth, in technological diffusion in these countries we began by replicating the Benhabib and Spiegel (2002) empirical methodology that extends their Benhabib and Spiegel (1994) study to accommodate the possibility that there are convergence clubs in the process of technological growth and convergence, in the tradition of Azariadis and Drazen (1990) and Durlauf and Johnson (1995). In addition to considering a different sample of countries, we extended their empirical analysis to a panel data framework, since it would have been impossible to use cross-section econometric techniques to analyse a sample of seven/thirteen countries.

In the Benhabib and Spiegel (1994) study the authors develop a model to explain the importance of human capital in growth, inspired by the Nelson and Phelps (1966) model in which the potential speed of technology diffusion is inversely related to the degree of technological backwardness of the follower country and its absorption capability for new technologies depends on its human capital level. They assume that the technological diffusion path is exponential, i.e. the technological leader acts as a locomotive for growth in the follower country, so that the follower always converges to the leader and empirically technological growth can be written as a linear function of human capital. The Benhabib and Spiegel (2002) study extends the framework of analysis in order to accommodate the possibility of the existence of human capital thresholds in which technological convergence may occur, i.e. the existence of convergence clubs. The empirical counterpart of this model is an equation for TFP growth non-linear in human capital: if there is convergence, the diffusion path is exponential, whereas if there are convergence clubs the diffusion path is logistic.

The replication of the Benhabib and Spiegel (2002) study did not produce good results. We estimated the Total Factor Productivity (TFP) growth rate non-linear specification proposed by Benhabib and Spiegel (2002) using traditional econometric methods, Non Linear Least Squares (NLLS), but there was no evidence that human capital influences TFP growth either directly through its impact on the domestic innovation rate or indirectly by speeding up technology diffusion. These results led us to conclude that the non-linear specification is not appropriate in explaining productivity growth in our samples, which might be due to the fact that the countries in our sample possess a human capital level that is already above the threshold necessary to exploit the advantages of technological backwardness, i.e. they are not in a poverty trap.

In light of the above results, we tested the linear specification of the original Benhabib and Spiegel (1994) study since this might be the one best suited to the evidence on human capital and TFP growth in our samples. We called this linear specification the Nelson and Phelps (1966) methodology and used Ordinary Least Squares (OLS) with robust errors, fixed effects models and random effects models to estimate it. The results confirmed the two influences of human capital on productivity growth, although surprisingly the direct effect on the domestic innovation rate is much stronger that the indirect effect through technology diffusion. This is good news since, according to this specification, the technological followers will always converge to the leader.

The low impact of human capital on imitation activities, however, led us to explore the hypothesis that human capital is more important for TFP growth through embodied technology diffusion than through disembodied technology diffusion, represented before by the TFP gap to the technological leader. Technology diffuses through many channels but the availability of data for our seven developing Mediterranean countries made us restrict our analysis to the study of the complementarity between human capital and technology transfers through FDI. We also analysed the role of human capital as a facilitator of the diffusion of a particular type of technology, Information and Communication Technologies (ICTs), which play a major part in productivity growth. In both cases we took Borensztein, Gregorio and Lee (1998) and Lee (2001) as the basic framework for our estimations. The results, however, did not show any complementarity between the diffusion of technology through FDI and human capital, although again they confirmed its importance in innovation activities. On the other hand, human capital is fundamental to the diffusion of ICTs, especially human capital acquired through higher education.

Finally, in the last part of the paper, we attempt to understand the importance of technological shocks in the process of technological diffusion and the role of human capital in the absorption of these shocks. The speed of technological diffusion and consequently the evolution of cross--country differences in GDP growth rates and levels depend, to a large extent, on exogenous shocks. We modelled technological shocks through a simple Vector Autoregression (VAR) model with four variables. The main result was that almost all of the seven countries showed evidence of factor complementarity in technology, physical capital and human capital in the absorption of any of the three types of shocks considered, with this new methodology also confirming the importance of human capital in technological progress.

The paper is divided into five sections. In the next section we describe the theoretical background that explains the relationship between human capital and productivity growth and develop an empirical analysis of this relationship, based on the methodologies of Benhabib and Spiegel (2002) and Nelson and Phelps (1966). Section 3 analyses the relationship involving the complementarity between human capital and FDI as a channel of technological diffusion, on the one hand, and the importance of human capital in the diffusion of a particular type of technology, ICTs. Section 4 analyses the relationship between technological shocks and technological catch--up. In Section 5 we conclude.

2. Technological catch-up and the role of human capital: the Benhabib and Spiegel (2002) and the Nelson and Phelps (1996) methodologies

2.1. Theoretical considerations

Renewed interest in the theme of economic growth in the eighties with the seminal articles of Romer (1986) and Lucas (1988) led to research on the possible different influences of human capital on growth. Human capital plays a very important role in endogenous growth theory: it is the source of sustained long term growth due to externalities in human capital accumulation, either in production, as in the Lucas (1988) model, or in innovation, as in the Romer (1990a) model.

Mankiw, Romer and Weil (1992) (MRW) empirically challenged the predictions of the endogenous growth literature. MRW used Solow's neoclassical growth theory to show that most of the observed cross-country income differences could be explained by a neoclassical growth model augmented to include human capital as an additional input in final goods production. Since the human factor is viewed as just another input, it has two growth effects: a permanent level effect on real GDP per capita and a transitory growth effect on GDP growth rate. MRW estimated what is commonly known as a β-equation[4] and concluded that almost 80% of cross-country income differences can be explained by differences in the rates of accumulation of physical and human capital and the population growth rate. However, when the analysis is restricted to the OECD countries there is evidence of the very slight influence of human capital on growth and sometimes even the estimated coefficient has the wrong sign. This study gave rise to numerous other studies that tried to improve on it in order to ascertain the correct influence of human capital on growth. Three different methods were basically followed as well as a mixture of all three: a) improved databases and human capital proxies (e.g., Klenow and Rodriguez-Clare, 1997; Temple, 1998); b) new econometric methodologies for the estimation of growth equations

4 The β-equation or convergence regression is derived from the neoclassical growth model in the neighborhood of the steady state predicting that output growth depends negatively on initial income due to the assumption of diminishing returns to reproducible inputs and positively on the determinants of the steady state income level, among which human capital is included. The symbol β refers to the coefficient on initial income predicted to be negative for there to be convergence.

(e.g., Islam, 1995; Caselli, Esquivel and Lefort, 1996); and c) new specifications for human capital in growth models based on the predictions of endogenous growth literature (e.g., Benhabib and Spiegel, 1994; Engelbrecht, 2003; Papageorgiou, 2003).

Endogenous growth literature has focused its research agenda on the explanation of TFP, that is, on the factors and mechanisms that cause technical progress and influence TFP growth. In this new theoretical setting human capital has two new roles: it determines the domestic innovation rate and it is a facilitator of technological catch-up. It is the level of human capital that is relevant in both roles. Human capital has permanent growth effects in the first case and transitory growth effects in the second one. In fact, in a steady state growth (SSG) model of technical diffusion, the transitory growth effects will last until the follower country reaches the TFP growth rate of the leader country.

In this paper we consider that human capital influences growth through these two roles, in the spirit of the Nelson and Phelps (1966) model which was given an empirical formulation by Benhabib and Spiegel (1994). According to Nelson and Phelps (1966), the shifting of the technological frontier towards the northeast depends on the rate of invention, whilst TFP growth depends on the rate of technological diffusion, which is positively related to the technological gap, the distance between the TFP level of the leader country and that of the follower country. In order to study the technological diffusion process in two countries it is assumed that the leader country is on the technological frontier or closer to it than the follower country. The technological catch-up hypothesis means that the TFP growth rate of the follower is positively related to its technological backwardness. This is a potential economic advantage for the follower but, as the authors have pointed out, the speed at which the technological gap is closed depends on the level of human capital available in the follower country. Abramovitz (1986) designates this potential for a country to benefit from technological backwardness its 'social capability', so that "(...) a country's potential for rapid growth is strong not when it is backward without qualification, but rather when it is technologically backward but socially advanced." p. 382. The term 'social capability' has been replaced in this literature by the term "absorptive capacity" referring to the various factors that influence the ability of a country to benefit from technology developed abroad. Abramovitz (1986) and Abramovitz (1994) describe some of the factors that determine absorptive capacity, pointing out that "It includes personal attributes, notably levels of education, an attribute that is subject to measurement, however imperfectly." (Abramovitz (1994), p.88).

In their 1994 study, Benhabib and Spiegel developed a model to explain the importance of human capital for growth inspired by the Nelson and Phelps (1966) and the Romer (1990a) models, where technological progress is explained by the domestic innovation rate, which is dependent on the level of human capital and the potential speed of technology diffusion that is inversely related to the degree of technological backwardness of the follower country and its absorption capability for new technologies, which is determined by its human capital level. They assume that the technological diffusion path is exponential, i.e. the technological leader acts as a locomotive for growth in the follower so that the follower always converges to the leader. These assumptions translate into an empirical formulation in which technological growth can be written as a linear function of human capital.

Benhabib and Spiegel (2002) extended their initial model of technological diffusion in order to allow for technological diffusion following a logistic path. This extension has the advantage of reconciling the theory with some stylised facts. Technological divergence between the follower and the leader is now possible if the level of human capital of the follower is lower than a critical threshold. The introduction of a threshold of this kind reconciles the model with convergence clubs results, in the tradition of Azariadis and Drazen (1990) and Durlauf and Johnson (1995). The empirical counterpart of this model is an equation for TFP non-linear growth in human capital: if there is convergence the diffusion path is exponential, whereas if there are convergence clubs the diffusion path is logistic. In the first generation of models for technological transfer, as in Nelson and Phelps (1966), Dowrick and Nguyen (1989), and De la Fuente (1995), the micro-foundations of innovation and imitation are absent. Second generation models introduce explicit agents that respond to market incentives, as in Barro and Sala-i-Martin (1997). Although the Benhabib and Spiegel (2002) model does not deal specifically with the agent behaviour that is related to their innovation and imitation activities, they prove that their results can be derived from the Barro and Sala-i-Martin (1997) model.

2.2. Empirical analysis

We replicated the Benhabib and Spiegel (2002) and the Nelson and Phelps (1966) empirical methodologies for a sample of seven developing Mediterranean countries, Algeria, Cyprus, Egypt, Israel, Syria, Tunisia and Turkey, due to the availability of human capital data. Although small, we consider this an interesting group of countries to study since it contains some degree of data variability, with Israel and Cyprus, for instance, registering relatively high levels of income and education that allow for the identification of the relevant coefficients. Additionally, we considered a wider sample of thirteen countries by adding the European Mediterranean countries, France, Greece, Italy, Portugal, and Spain, and Ireland. These six additional countries are all Mediterranean except for Ireland which, to a certain extent, allows us to control the influence of the geographical factor on productivity growth. On the other hand, four of the six European countries, Greece, Ireland, Portugal and Spain, have undergone a process of catching-up with the initially richer European countries which may reveal similarities with the eventual process of convergence occurring in the Southern Mediterranean countries.

We extended the Benhabib and Spiegel (2002) empirical analysis to a panel data framework since it would have been impossible to use cross-section econometric techniques to analyse a sample of seven/thirteen countries. In this way we were able to explore a richer information set, with time series as well as cross section information that allowed for an even greater degree of freedom and therefore improved the efficiency of the estimators; furthermore we were able to control the omitted variable bias (see Baltagi, 2001). Adding countries to our original sample of seven developing Mediterranean countries further improves the efficiency of the estimators. In the empirical analysis we used four panel databases resulting from the use of annual data and data at 5-year intervals to reduce the impact of business cycle effects for the period 1960-2000.

Data for real GDP (rdgpl), investment shares as a ratio of the GDP (ki) and population (POP) were taken from the Penn World Tables (PWT) Mark 6.1. Human capital was proxied by the average years of schooling of the population aged 15 and over (TYR) taken from Barro and Lee (2000). Annual data for human capital, provided originally at 5-year intervals, was annualised through non-linear interpolation[5]. The data that was unavailable for Cyprus (1997-2000) and Tunisia (1960) was computed using ARIMA models for each variable.

We used TFP growth as a proxy for the technological growth rate. The TFP index is computed in logs from a Cobb-Douglas aggregate production function as the difference between real GDP and primary input use, physical capital and labour (proxied by population), weighted by their income shares, α and (1-α) respectively:

$$a_{it} = y_{it} - \alpha k_{it} - (1 - \alpha) l_{it} = y_{it} - \frac{1}{3} k_{it} - \frac{2}{3} l_{it} \qquad (1)$$

where a_{it} is the log of the country's TFP level (i) at time t, y_{it} is the log of the country's real GDP (i) at time (t), k_{it} is the log of the physical stock of the country's capital (i) at time (t) and l_{it} is the log of the population of country (i) at time (t)[6], and α is assumed to take the value 1/3. Since there was no physical capital data available we computed physical capital stocks following the

5 This was done by using the RATS' procedure DISTRIB.rsc that computes the distribution of a series changing its frequency to a higher one. We have assumed that the original series follows a random walk.
6 The Lee (2001) method was also used but the results are not considered here since they were economically meaningless.

Klenow and Rodriguez-Clare (1997) methodology through the perpetual inventory method using investment data[7]. Having obtained the series for the TFP levels at annual and 5-year intervals, we computed the TFP growth rates.

To determine the appropriate estimation procedures for our panel data analysis we studied the unit root characteristics of the TFP growth rate series using unit root panel tests. To overcome the problem of spurious regressions, that is, to be able to apply classical econometric procedures, we had to verify whether our TFP growth series was I(0), i.e. it did not have a unit root. The Appendix contains the results for the panel unit root tests proposed by Levin and Lin (1993) and Im, Pesaran and Shin (2003) which allowed us to reject the null hypothesis of the presence of a unit root in the TFP growth series. We thus estimated the Benhabib and Spiegel (2002) non-linear specification using NLLS, including a constant, a trend or individual constants. As for the Nelson and Phelps (1966) linear equation, OLS with robust errors was used, including a constant, a trend or individual constants. Fixed effects as well as random effects models were also used.

2.2.1. Empirical findings using the Benhabib and Spiegel (2002) methodology

The Benhabib and Spiegel (2002) empirical formulation for the relationship between TFP growth and human capital that we tested is given as equation (2) below:

$$g_{(TFP)it} = b + \left(g + \frac{c}{s}\right)h_{it} - \left(\frac{c}{s}\right)h_{it}\left(\frac{A_{it}}{A_{mt}}\right)^{s} + \varepsilon_{it} \qquad (2)$$

where $g_{(TFP)it}$ is the TFP growth rate of country (i) at time (t); b is the constant term; h_{it} is the stock of human capital for country (i) at time (t) in logarithms; A_{it} is the TFP level of the follower country (i) at time (t); and A_{mt} is the TFP level of the leader country (assumed to be the USA)[8].

According to equation (2), the TFP growth rate of country (i) at time (t) depends: i) on a constant term b; ii) positively on the level of human capital whose coefficient is [g + (c/s)], with the expression $[g + (c/s)]h_{it}$ capturing the contribution of the innovation process of country (i) at time (t) to its TFP growth rate; iii) negatively on the degree of technological backwardness, taking into account its interaction with the level of human capital whose coefficient is [-(c/s)], with the expression $[-(c/s)]h_{it}(A_{it}/A_{mt})$ capturing the contribution of the diffusion process of country (i) at time (t) to its TFP growth rate; and iv) on the error term that is i.i.d distributed. Equation (2) allows us to control the two types of technological diffusion paths described before – exponential (s = -1) or logistic (s = 1). In this way we can determine whether the Mediterranean countries will converge to the USA or whether they are caught in a poverty trap due to low human capital levels.

We estimated different versions of equation (2). We called each estimated version models A and B: model A uses annual data for the stock of human capital whilst model B considers the initial human capital stock average for the period 1960-1965. We estimated models A and B for both samples, considering annual data and three cases: with constant term, with trend, and with individual constants. In the case of model A, estimations were also performed for all three cases using 5-year data. The results are presented in tables 1, 2 and 3. After inspecting the results, the main conclusion is that we cannot accept the Benhabib and Spiegel (2002) specification for either of the samples. In fact, the results obtained are very weak. Let us briefly interpret the results obtained in each of the tables.

7 See the Appendix for details on the computation of the physical capital stock series.

8 Please refer to Benhabib and Spiegel (2002) for more details on the variables and equations used.

Table 1 – Seven countries (Benhabib and Spiegel, 2002)

NLLS	Model A with constant	Model A with trend	Model A with C_{is}	Model B with constant	Model B with trend	Model B with C_{is}
	TFP Annual Growth Rate					
b	-0.044 (3.84***)	-0.043 (4.14***)	–	-0.001 (0.17)	-0.042 (4.36**)	–
g	0.026 (0.36)	0.010 (1.27)	0.056 (5.24***)	0.008 (1.23)	0.007 (1.18)	-0.233 (0)
c	0.014 (0.04)	-0.00000002 (0.21)	-0.00002 (0.43)	-0.000008 (0.38)	-0.0000005 (0.27)	-0.0006 (1.03)
s	2.182 (0.06)	-12.5 (3.66***)	-7.75 (4.36***)	-8.883 (4.34***)	-10.963 (3.99***)	-5.899 (7.43***)
b_1	–	–	-0.052 (3.65***)	–	–	-0.001 (0.14)
b_2	–	–	-0.090 (4.13***)	–	–	0.396 (25.56***)
b_3	–	–	-0.049 (3.51***)	–	–	0.008 (0.78)
b_4	–	–	0.122 (4.82***)	–	–	0.427 (37.62***)
b_5	–	–	0.0025 (1.35)	–	–	0.191 (11.71***)
b_6	–	–	0.036 (2.53***)	–	–	-0.037 (3.35***)
b_7	–	–	-0.077 (4.78***)	–	–	0.0165 (13.58***)
trend	–	0.001 (3.31**)	–	–	-0.002 (5.24***)	–
see	0.071	0.068	0.067	0.074	0.031	0.073
n-k	276	275	270	276	52	270

*significant at 10% level; **significant at 5% level; *** significant at 1% level; in brackets t-student values; trend – time effect coefficient.

Table 1 presents the results of the estimation of models A and B for the sample of seven Mediterranean countries with annual data. The estimated values of the coefficients g and c are not significant and the coefficient [-(c/s)] has the wrong theoretical sign in models A and B with constant. As for s, it is significant at the 1% level only in model B but its value neither confirms a logistic path, nor an exponential path. The results for models A and B with trend improve: in model A the estimated value for coefficient s becomes significant and in model B, the estimated value for coefficient b becomes significant; in any case, however, c is not significantly different from zero, an extremely implausible result from a theoretical point of view since it dismisses any influence of technology diffusion on TFP growth. Like the previous models, [-(c/s)] has the wrong sign and again the value of s is neither equal to one or minus one. As for model A with individual constants, the results have improved compared with those obtained with the model with trend: g becomes significant at the 1% level, nonetheless c is not significantly different from zero and s is not equal to minus 1. In model B with individual constants, the results have not improved in terms of the model with trend.

Table 2 – Seven countries (Benhabib and Spiegel, 2002)

NLLS	TFP 5-year average growth rate Model A	 Model A with trend	 Model A with C_{is}
b	-0.046 (3.69***)	-0.057 (4.84***)	–
g	1.231 (0.02)	-77.05 (0.13)	0.073 (5.42***)
c	-23.052 (0.02)	2812.618 (0.13)	-0.002 (0.65)
s	19.204 (0.11)	36.50 (0)	-3.457 (2.42**)
b_1	–	–	-0.057 (3.89***)
b_2	–	–	-0.109 (4.91***)
b_3	–	–	-0.056 (3.97***)
b_4	–	–	-0.153 (5.77***)
b_5	–	–	-0.044 (2.36**)
b_6	–	–	-0.050 (3.47 ***)
b_7	–	–	-0.086 (5.47***)
trend	–	0.001 (3.09***)	–
see	0.034	0.032	0.028
n-k	52	51	46

*significant at 10% level; **significant at 5% level; *** significant at 1% level; in brackets t- student values; trend – time effect coefficient.

Table 2 presents the results of the estimation of model A for the sample of seven Mediterranean countries with data at 5-year intervals. The results have improved in comparison to those using annual data. In fact, all the individual constants are now significant and the Standard Error of the Estimate (SEE) is now 2.8% as opposed to 6.7% in the previous estimations. Nonetheless, the coefficient c is again not significantly different from zero.

Table 3 – Thirteen countries (Benhabib and Spiegel, 2002)

	TFP Annual Growth Rate			TFP 5-year average growth rate		
NLLS	**Model A**	**Model A** with trend	**Model A** with C_{is}	**Model B**	**Model B** with trend	**Model B** with C_{is}
b	-0.041 (5.15***)	-0.049 (6.25***)	–	-0.051 (4.41***)	-0.052 (5.71***)	–
g	0.074 (1.74*)	0.042 (0.85)	4.544 (1.09)	0.029 (4.59***)	0.010 (1.52)	0.003 (0.85)
c	-0.490 (0.66)	-0.410 (0.36)	-649.319 (1.07)	0.004 (0.29)	-0.000 (1.17)	-0.011 (0.30)
s	10.057 (1.23)	12.163 (0.66)	144.734 (0)	-0.956 (0.14)	54.76 (0)	9.399 (0.65)
b_1	–	–	-0.068 (5.81***)	–	–	-0.012 (0.88)
b_2	–	–	-0.100 (5.93***)	–	–	0.007 (0.51)
b_3	–	–	-0.061 (5.34 ***)	–	–	-0.008 (0.61)
b_4	–	–	-0.126 (6.88***)	–	–	-0.005 (0.37)
b_5	–	–	-0.077 (5.90***)	–	–	-0.006 (0.41)
b_6	–	–	-0.054 (4.68***)	–	–	0.0004 (0.03)
b_7	–	–	-0.083 (6.40***)	–	–	-0.013 (0.93)
b_8	–	–	-0.115 (6.95***)	–	–	0.950 (0)
b_9	–	–	-0.112 (6.72***)	–	–	0.928 (0)
b_{10}	–	–	-0.117 (6.66***)	–	–	0.902 (0)
b_{11}	–	–	-0.105 (6.57***)	–	–	4.377 (0)
b_{12}	–	–	-0.074 (5.60***)	–	–	0.910 (0)
b_{13}	–	–	-0.101 (6.50***)	–	–	0.640 (0)
trend	–	0.002 (6.22***)	–	–	-0.001 (5.35***)	–
see	0.058	0.0056	0.057	0.032	0.028	0.36
n-k	516	515	504	100	99	88

*significant at 10% level; **significant at 5% level; *** significant at 1% level; in brackets t- student values; trend – time effect coefficient.

Table 3 presents the results of the estimation of model A for the extended sample of thirteen countries with annual data and data at 5-year intervals. Considering the results with annual data for the model with constant, g is now significant at the 10% level but c and s are not significantly different from zero. For the models with trend and individual constants the results are worse. For model A with trend, s is no longer significant compared to the same model for the smaller sample. As for model A with individual constants, g and s are no longer significantly different from zero. Model A with constant is thus the model with the best results, although it is still very weak. In fact, only g and b are significant at the 10% level. Turning to the results using data at 5-year intervals, these are an improvement since g and b are now significant at 1% level. If we compare the results obtained with model A and 5-year data for the larger sample with those of the smaller sample, the results are better for the model with constant and worst for the model with individual constants.

These weak results led us to conclude that the non-linear technological diffusion specification of Benhabib and Spiegel (2002) is not suitable for our samples[9]. The fact that the three coefficients of their empirical specification, g, s and c are never significant at the same time, nor do they display simultaneously the signs predicted by theory; the fact that the coefficient c is never significantly different from zero and also the fact the value of s is not equal to one or to minus, made us believe that the non-linear specification for TFP growth as a function of human capital is not suitable for describing the technological diffusion process in our sample of countries. In particular, the logistic path for the technological diffusion process seems not to apply to our two samples. One possible explanation is that in the smaller sample the level of human capital is not constrained by a threshold, which would probably occur if we had worked with a larger sample, as the authors did, which included the poorest countries in the world.

In view of these results we decided to check whether the linear specification adopted by Benhabib and Spiegel (1994), inspired in turn by the Nelson and Phelps (1966) model, is better suited to explaining TFP growth in our sample. We call this the Nelson and Phelps (1966) methodology to distinguish it from the Benhabib and Spiegel (2002) methodology.

2.2.2. The Nelson and Phelps (1966) methodology

According to the Nelson and Phelps (1966) model, technology diffusion follows an exponential diffusion path so that the follower always converges to the leader. Assuming that this exponential path applies to all countries, Benhabib and Spiegel (1994) derive the following linear specification for the behaviour of TFP growth, adding the predictions of the Romer (1990a) model as to the influence of the level of human capital on the domestic innovation rate[10]:

$$g_{TFPit} = gH_{it} + cH_{it}\left(\frac{A_{m(t)}}{A_{it}} - 1\right) + \varepsilon_{it} \qquad (3)$$

According to equation (3), the rate of technical progress depends on the rate of innovation, which is a positive function of the stock of human capital (gH_{it}), and on a technological catch-up term, which is also a positive function of the stock of human capital. The rate of technical progress is now positively related to the degree of technological backwardness of the economy, due to the definition of technological backwardness now used ($\frac{A_{m(t)}}{A_{it}}$). The estimated equation is not equation (3) but equation (4) below, obtained after normalising the values of human capital and of the technological gap (deviations from the average value).

9 We also estimated the different models using ML methods, with one variance and individual variances, assuming an AR1 process but the results were not good.
10 Please refer to Benhabib and Spiegel (1994) for more details on the variables and equations used.

$$g_{TFPit} = gH_{it} + c_{it}Z_{it} + \varepsilon_{it}$$
$$Z_{it} = (H_{it} - \bar{H})\left[\left(\frac{A_{m(t)}}{A_{it}} - 1\right) - \left(\overline{\frac{A_{m(t)}}{A_{it}} - 1}\right)\right] \quad (4)$$

Since equation (4) is linear, we estimated it using OLS with robust errors, fixed effects and random effects models. We only estimated model A. Tables 4 to 6 present the results of the estimation of equation (4) for the sample of developing Mediterranean countries using annual data as well as data at 5-year intervals[11].

Table 4 – Seven countries (Nelson and Phelps, 1996, equation)

	TFP Annual Growth Rate			TFP 5-year average growth rate		
NLLS	**Model A** with constant	**Model A** with trend	**Model A** with C_{is}	**Model A**	**Model A** with trend	**Model A** with C_{is}
B	-0.032 (2.88***)	-0.047 (3.67***)	–	-0.035	-0.051 (4.46***)	–
g	0.009 (3.76***)	0.004 (1.51)	0.001 (1.11)	0.008	0.004 (1.36)	0.001 (0.87)
c	0.007 (2.50**)	0.004 (1.48)	0.004 (1.48)	0.006	0.003 (1.01)	-0.003 (1.36)
b_1	–	–	0.046 (2.75***)	–	–	0.009 (0.31)
b_2	–	–	-0.006 (2.06***)	–	–	-0.007 (2.29**)
b_3	–	–	0.026 (1.94*)	–	–	0.025 (2.76***)
b_4	–	–	0.009 (0.40)	–	–	0.009 (0.33)
b_5	–	–	0.029 (2.19**)	–	–	0.018 (7.11***)
b_6	–	–	0.009 (1.93*)	–	–	0.011 (2.32**)
b_7	–	–	-0.045 (2.05**)	–	–	-0.040 (1.27)
trend	–	0.001 (3.30**)	–	–	(3.30**) (3.08***)	–
see	0.071	0.069	0.067	0.034	0.031	0.033
n-k	277	276	272	53	52	48

*significant at 10% level; **significant at 5% level; *** significant at 1% level; in brackets t- student values; trend – time effect coefficient.(n-k) – degrees of freedom.

Table 4 presents the results of the estimation of model A with constant, trend and individual constants for the smaller sample, with annual data and data at 5-year intervals, using OLS. Considering the results with annual data, the best model is model A with constant. In fact all the coefficients are significant and have the sign predicted by theory. Nonetheless the values of g

11 The results of the estimation of this model for the larger sample do not change significantly either qualitatively or quantitatively so we have refrained from including them here. These results are available from the authors upon request.

and c are very small. The other two models behave very badly. In fact, g and c are never significantly different from zero in these models. As for the results with 5-year data, these models should be disregarded: g and c are never significantly different from zero.

Table 5 – Seven countries (Nelson and Phelps, 1996, equation)		
Fixed effects model	TPF annual growth rate	TFP 5-year growth rate
g	0.016 (5.19***)	0.016 (5.11***)
c	0.004 (1.69*)	0.003 (1.15)
see	0.067	0.032
n-k	310	47

*significant at 10% level; **significant at 5% level; *** significant at 1% level; in brackets t- student values; trend – time effect coefficient.(n-k) – degrees of freedom.

Table 5 presents the results of the estimation of model A for the smaller sample assuming fixed effects. The best result is obtained with annual data. In fact, all the coefficients are significant at the 1% level and at the 10% level. Notice that the value of g has increased compared with the estimation with OLS – the value of g is higher than c, whose value is very small. With 5-year data c is no longer significantly different from zero. The results are very sensitive to the frequency of the data.

Table 6 – Seven countries (Nelson and Phelps, 1996, equation)			
Random effects model	TPF annual growth rate	TFP 5-year growth rate	TFP 5-year growth rate
constant	-0.064 (3.48***)	-0.066 (3.47***)	–
g	0.014 (5.59***)	0.013 (4.89***)	0.006 (3.45***)
c	0.006 (2.67***)	0.005 (1.75)	0.006 (2.14**)
see	0.066	0.031	0.033
n-k	317	53	54

*significant at 10% level; **significant at 5% level; *** significant at 1% level; in brackets t- student values; trend – time effect coefficient. (n-k) – degrees of freedom.

Table 6 presents the results of the estimation of model A for the smaller sample, now assuming random effects. With annual data, all the coefficients are significantly different from zero at the 1% level and the same is true for the model without constant using 5-year data. The value of coefficient c is higher in both models compared with those obtained with the fixed effects model and for the last model g is no longer higher than c.

To sum up, the evidence presented in this section confirms the importance of human capital as a determinant of technological progress, based on the results of the tests of the Nelson and Phelps (1966) linear specification. A somewhat surprising result comes from the fact that the influence of human capital is felt mainly through innovation and not imitation activities. This is, however, in accordance with the results of the tests of the Benhabib and Spiegel (2002) hypothesis that did not allow us to confirm that the TFP growth rate follows a logistic function, i.e. the human capital in our sample is already higher than the threshold necessary for it to exert its influence over the technological progress growth rate. The low impact of human capital on imitation activities, however, leads us to explore in the next section the hypothesis that human capital is more important for TFP growth through embodied technology diffusion than through disembodied technology diffusion, previously represented by the TFP gap to the technological leader.

3. Human capital and channels of technology diffusion

In this section we propose to analyse a little further the role of human capital in the process of technological diffusion, focusing on the complementarity between human capital and foreign direct investment (FDI) as determinants of the technological progress growth rate, on the one hand, and on the importance of human capital as a facilitator of the diffusion of information and communication technologies (ICT). Technology diffuses through many channels, two of which have been extensively studied in the empirical literature, international trade and Foreign Direct Investment (FDI)[12]. Data availability for our seven developing Mediterranean countries made us restrict our analysis the study of the complementarity between human capital and technology transfers through FDI. We also analysed the role of human capital as a facilitator of ICT diffusion, since this plays a major role in productivity growth[13]. The diffusion of these new technologies can thus contribute to the acceleration of technology diffusion in Mediterranean countries. In both cases we took Borensztein, Gregorio and Lee (1998) and Lee (2001) as the basic framework for our estimations.

3.1. The complementarity between human capital and FDI in the process of technological diffusion

The purpose of this section is to empirically examine the complementarity between human capital and FDI in the process of technology diffusion in our sample of Mediterranean countries. FDI is one of the channels[14] through which the technology from the leader is transferred to the followers. However, the host economy needs a sufficient level of human capital in order to apply the technology of the leader, i.e. the stock of human capital of the follower country limits its absorptive capability for the technology incorporated in FDI.

We test this complementarity hypothesis in a panel data framework between 1970 and 1998 following Borensztein, Gregorio and Lee (1998) and Lee (2001) and their basic formulation:

$$D(TFP)_{it} = \alpha_0 + \alpha_1 GTFP_{it-1} + \alpha_2 TYR_{it-1} + \alpha_3 FDI_{it} + \alpha_4 FDI_{it} \times SHYR_{it} + \eta_i + \varepsilon_{it} \qquad (5)$$

where $D(TFP)_{it}$ is the log difference of the TFP level as defined in the previous section, $GTFP_{it-1}$ is the initial gap in the technology level in relation to the USA computed as $\log(\frac{TPF_{USAt-1}}{TPF_{it-1}})$, TYR_{it-1} is the initial stock of human capital measured as the average years of total schooling in the population aged 15 and over, $SHYR_{it}$ is the average years of secondary and higher education in the population aged 15 and over, FDI_{it} is the net FDI flows as a ratio to GDP, η_i represents country-specific effects, and ϵ_{it} is the error term with the usual properties.

Technological growth depends positively on the initial technological gap between the leader and the follower country – the higher the initial gap, the higher the potential for the adoption and implementation of new technologies, i.e., the higher the TFP growth rate of the follower, so we expect a positive and significant α_1 – this is the usual technological catch-up assumption of

12 See Coe and Helpman (1995), Engelbrecht (1997), Frantzen (2000), Potterie and Lichtenberg (2001), Xu (2000). See also Keller (2004) for a review of the literature on technology diffusion.

13 See Schreyer (2000), Van Ark, Inklaar and McGuckin (2003), Cette, Mairesse and Kocoglu (2004).

14 Another channel of technology diffusion from the leader to the followers is imports of machinery and transport equipment. Unfortunately, we were not able to get access to data on imports of machinery and transport equipment from the OECD countries, the countries responsible for most of the world's R&D initiatives, for our sample of Mediterranean countries.

technology diffusion models such as the Nelson and Phelps (1966) and the Barro and Sala-i--Martin (1997) models discussed before. Human capital also positively influences TFP growth, since the adoption and implementation of new technologies requires at least basic levels of skills ($\alpha_2 > 0$). On the other hand, FDI is a fundamental channel through which less developed countries gain access to the advanced technologies of the developed countries, which means that α_3 should be positive. Finally, the hypothesis that the diffusion of technology through FDI is only effective if the host economy has the necessary absorptive capability in the form of human capital is tested through the interactive term *FDI*x*SHYR* – if its coefficient is positive and significant this means that the technology spillovers coming from FDI depend on the stock of human capital. Note that now only average years of secondary and tertiary education are considered, meaning that we assume that to benefit from the FDI originating in developed countries the Mediterranean countries need more than the basic skills provided by primary education.

The TFP and human capital data are the same as those used in the previous sections. The FDI data comes from the OECD publication "Geographical distribution of financial flows to aid recipients " (OECD, 2003) and measures the net flows of FDI received by the countries in our sample from OECD countries, responsible for most of the R&D initiatives in the world.

We estimated our relationship using four different estimation procedures – the pooled ordinary least squares (OLS), the within-groups estimator, the first-differenced generalized method of moments (GMM-DIF) proposed by Arellano and Bond (1991), and the system generalized method of moments (GMM-SYS) proposed by Arellano and Bover (1995) and Blundell and Bond (1998), each corresponding to different assumptions concerning the econometric properties of the relationship we were analysing.

The pooled OLS estimator delivers unbiased and consistent estimators if there are no country--specific effects in the relationship and if the regressors are strictly exogenous. On the contrary, in the presence of country-specific effects but still strictly exogenous regressors, the within-groups estimator delivers unbiased estimators since it controls the omitted variable bias. In the presence of country-specific effects and the violation of the assumption of strict exogeneity of the regressors, however, the OLS and the within-groups estimators are biased, since equation (5) can be written as a dynamic panel with the lagged dependent variable on the right hand side. In this situation at least one of the regressors, the lagged dependent variable, is correlated with the error term and so the OLS estimate of the coefficient on initial TFP is biased upwards, whilst the within-groups estimator is biased downwards (see Nickell, 1981 and Bond, Hoeffler and Temple, 2001). If the regressors are not strictly exogenous the results from these two procedures relative to the estimated coefficient on the initial technological gap will be different, providing a clue that OLS and within-groups are not adequate estimation procedures. Arellano and Bond (1991) propose the use of the first-differenced GMM estimator to overcome this problem. This procedure consists of first differencing the TFP equation to eliminate fixed effects and then using adequate past levels of the relevant variables as instruments. However, in the presence of weak instruments the GMM-DIF estimator is biased towards the within-groups estimator and the GMM--SYS estimator proposed by Arellano and Bover (1995) and Blundell and Bond (1998), which uses both the levels and the first differences of the regressors as instruments is an adequate estimation procedure. To check the consistency of the GMM estimators used we present the results of the Sargan test of over-identifying restrictions that tests for the null of overall validity of the instruments used and the results of a test of the hypothesis that the errors are serially uncorrelated proposed by Arellano and Bond (1991). The results of the tests support the use of these estimation procedures.

In table 7 we present the results of the estimation of the different equations using annual data and the four different estimation procedures previously mentioned. To control the possibility of business cycle effects on the TFP growth rate we also estimated the different equations, averaging the data over 5-year periods[15]. The results for these estimations are presented in table 8.

---

15 For the last period, 1995-1998, we used 3-year averages.

Table 7 – Human capital and technology diffusion through FDI flows (annual data)

	Dependent variable: D(TFP) – Annual growth rate of TFP, 1970-1998											
	Pooled OLS			Within Groups			GMM-DIF[c]			GMM-SYS		
GTFP (t-1)	0.105 (4.86)**	0.106 (5.05)**	0.106 (5.12)**	0.171 (2.15)**	0.173 (2.26)**	0.175 (2.25)**	0.41 (7.59)**	0.179 (10.2)**	0.195 (9.20)**	0.181 (17.0)**	0.161 (25.8)**	0.18 (20.8)**
TYR (t-1)	0.0027 (4.18)**	0.0027 (4.31)**	0.0026 (4.13)**	0.004 (2.07)**	0.004 (2.24)**	0.004 (2.30)**	-0.01 (-0.952)	0.002 (0.802)	0.002 (0.665)	0.004 (3.72)**	0.004 (5.80)**	0.004 (5.60)**
FDI (t)		0.002 (2.66)**	0.0009 (0.549)		0.002 (1.63)*	0.0008 (0.247)**		0.001 (0.779)	-0.0024 (-0.448)		0.003 (1.42)	0.0012 (0.205)
*FDI*SHYR (t)*			0.0007 (1.21)			0.0007 (0.594)			0.002 (0.779)			0.0009 (0.393)
AR (2)[a]							0.72	0.792	0.824	0.761	0.770	0.782
Sargan Test[b]							0.029	0.921	1.000	0.000	0.000	0.000
Obs	203	203	203	203	203	203	189	189	189	196	196	196

Notes: values of the t-Student statistic in brackets. ** significant at the 5% level; * significant at the10% level.
Instruments used in GMM-DIF: lnTFPit-2, lnTYRit-3; lnFDIPit-2, lnFDI*SHYRSit-2, and lags up to the fourth lag.
Instruments used in GMM-SYS: same as for GMM-DIF and additionally instruments for the levels equations are ΔlnTFPit-1, ΔlnTYRPit-2; ΔlnFDIPit-1, ΔlnFDI*SHYRSit-1.
[a] p-values for the null hypothesis that the errors in the first-difference regression exhibit no second-order serial correlation. [b]p-values for the null hypothesis of overall validity of the instruments used. [c] Results for the one-step GMM estimator with standard errors robust to heteroskedasticity since the standard errors of the two-step GMM estimator can be seriously biased downwards.

Table 8 – Human capital and technology diffusion through FDI flows (5-year averages)

	Dependent variable: D(TFP) – Annual average growth rate of TFP, 1970-1998											
	Pooled OLS			Within Groups			GMM-DIF[c]			GMM-SYS		
GTFP (t-1)	0.067 (4.20)**	0.065 (4.34)**	0.065 (4.35)**	0.075 (3.11)**	0.094 (3.75)**	0.095 (3.42)**	0.499 (2.97)**	0.203 (4.84)**	0.201 (3.74)**	0.331 (6.53)**	0.146 (6.89)**	0.203 (5.40)**
TYR (t-1)	0.0017 (3.93)**	0.0017 (3.31)**	0.0017 (2.70)**	0.002 (2.25)**	0.002 (2.16)**	0.002 (1.81)**	-0.003 (-0.350)	-0.002 (-0.331)	0.003 (0.362)	0.009 (2.64)**	0.006 (1.77)*	0.006 (2.05)**
FDI (t)		0.0003 (0.320)	0.0005 (0.133)		-0.001 (0.631)	0.0003 (0.055)		-0.008 (-0.834)	-0.010 (-0.786)		-0.006 (-0.751)	-0.014 (-1.13)
*FDI*SHYR (t)*			-0.0009 (-0.059)			-0.001 (-0.519)			0.0019 (0.332)			0.005 (0.999)
AR (2)[a]							0.772	0.677	0.363	0.261	0.265	0.292
Sargan Test[b]							0.231	0.530	0.795	0.009	0.055	0.082
Obs	42	42	42	42	42	42	35	35	35	42	42	42

Notes: values of the t-Student statistic in brackets. ** significant at the 5% level; * significant at the10% level.
Instruments used in GMM-DIF: $\ln TFP_{it-2}$, $\ln TYR_{it-3}$; $\ln FDI_{Pit-2}$, $\ln FDI*SHYR_{Sit-2}$, and lags up to the fourth lag.
Instruments used in Sys-GMM: same as for GMM-DIF and additionally instruments for the levels equations are $\Delta \ln TFP_{it-1}$, $\Delta \ln TYR_{Pit-2}$; $\Delta \ln FDI_{Pit-1}$, $\Delta \ln FDI*SHYR_{Sit-1}$.
[a] p-values for the null hypothesis that the errors in the first-difference regression exhibit no second-order serial correlation. [b] p-values for the null hypothesis of overall validity of the instruments used. [c] Results for the one-step GMM estimator with standard errors robust to heteroskedasticity since the standard errors of the two-step GMM estimator can be seriously biased downwards.

As far as the results using annual data are concerned (Table 7), the technological catch-up hypothesis is confirmed for all the equations – the coefficient on the initial technological gap is always positive and significant, meaning that the initially more technologically backward countries were indeed the ones that exhibited faster TFP growth rates. The role of the initial level of human capital in the domestic innovation rate is also confirmed (except when we use the first--differenced GMM estimator) – its coefficient is always positive and significant. In the equation where FDI is included on its own its expected positive influence over the TFP growth is confirmed only when the pooled OLS and the within-groups estimators are used. In the case of the first--differenced GMM estimator the coefficient, although positive, is not significant and with the system GMM estimator it is positive but only significant at the 25% level. The hypothesis we are focusing on is that the technology originating in FDI flows is used effectively only if the host country has the necessary human capital, which means that in our full equation the coefficient in the interaction term between FDI and human capital should be positive and significant. From our results we see that this is not, however, the case – although the coefficient is always positive it is never significant. Furthermore, the coefficient on FDI alone always becomes non-significant and even negative when the GMM-DIF estimator is used. Our hypothesis of complementary between FDI flows and human capital is therefore not supported by this data for our seven Mediterranean countries.

Turning now to the results using 5-year averages, nothing much changes. The coefficients on the initial technological gap and human capital are still always positive and significant (except for the human capital coefficients using GMM-DIF), the coefficient on FDI when introduced on its own is never significant and it is only positive when using the pooled OLS estimator. Finally when the full equation is estimated the coefficient on FDI remains non-significant and the same happens with the coefficient on the interaction term.

To sum up, we can say that human capital on its own influences the growth rate of technological progress in our seven Mediterranean countries, due to its influence on the rate of domestic innovation but the evidence does not confirm its role in determining the TFP growth rate as a determinant of the absorptive capability of the technology embodied in FDI. Maybe a better measure for the spillovers of technology from the technological leaders to the followers would be imports of machinery and transport equipment which, unfortunately, we could not gain access to for our sample.

3.2. Human capital as a facilitator of the diffusion of IC technologies

It is widely accepted that Information and Communication technologies (ICTs) play a major role in technological progress nowadays and hence the diffusion of these new technologies contributes towards accelerating technological diffusion in our sample of Mediterranean countries. However, these new technologies require more than basic skills to be fully implemented, i.e., human capital levels are a major determinant of the absorptive capability of ICTs in the Mediterranean countries. In order to test this hypothesis we estimated the relationship between human capital and a set of ICT indicators in a panel data framework using the following equations, (6) and (7):

$$ICT_{it} = b_0 + b_1 logRGDP_{it} + b_2 TYR_{it} + \mu_i + \nu_{it} \quad (6)$$

$$ICT_{it} = c_0 + c_1 logRGDP_{it} + c_2 PYR_{it} + c_3 SYR_{it} + c_4 HYR_{it} + \mu_i + \nu_{it} \quad (7)$$

where ICT_{it} is an ICT indicator, measured alternatively as main telephone lines, number of personal computers, internet hosts, daily newspapers, and number of TV sets, all per 1,000 people[16] and $logRGDP_{it}$ is the natural logarithm of real GDP per capita from the PWT Mark 6.1 and proxies for the constraint that national financial resources represent on the investments necessary for building ICT infrastructures. In equation (6) we consider the influence of TYR_{it}, the average years of total schooling for the population aged 15 and over that proxies for the skills necessary for the implementation of ICTs. In equation (7) we consider the influence of the different levels of schooling separately, since basic literacy skills may not be sufficient to fully benefit from the ICTs and thus there might be separate effects for each level of schooling in the evolution of the different ICT indicators. PYR_{it} is the average number of years of primary education for the population aged 15 and over; SYR_{it} is average number of years of secondary education for the population aged 15 and over; and HYR_{it} is the average number of years of higher education for the population aged 15 and over. μ_i is a country-specific effect and v_{it} is the error term with the usual properties.

We present the results of the estimation of the different equations in tables 9 and 10. In the first table we ignored the presence of country-specific effects in determining the evolution of ICTs, i.e. we estimated our different equations using the pooled OLS estimator. In the second table we considered that there might have been country-specific effects governing the evolution of ICTs, so we present the results of the estimation of the different equations using the within-groups estimator.

Table 9 – Human capital and ICT diffusion (Pooled OLS)

Dependent variable	log(RGDP per capita)	TYR	PYR	SYR	HYR	$\bar{R}^2$	Obs.
Telephone lines	139.939	24.6327				0.798	172
	(2.89)**	(2.72)**					
	185.333		-92.3	53.16	745.6	0.888	172
	(5.68)**		(-4.6)**	(2.51)**	(7.80)**		
Personal computer	26.3	17.7				0.81	32
	(2.42)**	(3.58)**					
	51.9		-17.65	-25.38	272.72	0.886	32
	(1.77)**		(-0.482)	(-1.79)*	(1.36)		
Internet hosts	3.22	12.04				0.519	35
	(0.365)	(2.62)**					
	28.06		-23.18	-27.69	259.43	0.701	35
	(0.917)		(-0.554)	(-1.99)**	(1.16)		
Daily papers	66.52	16.63				0.700	63
	(2.04)**	(1.37)					
	79.40		9.34	-87.62	293.4	0.905	63
	(1.94)**		(0.266)	(-4.65)**	(1.40)		
TV sets	75.20	11.36				0.643	172
	(1.84)**	(1.20)					
	69.30		29.61	-46.91	62.38	0.703	172
	(1.79)**		(1.45)	(-1.91)**	(0.497)		

Notes: values of the t-Student statistic in brackets. ** significant at the 5% level. * significant at the10% level.

The results using the pooled OLS estimator show that real GDP and average years of schooling explain most of the development in ICTs in the Mediterranean countries with $\bar{R}^2$ higher than 50%. The availability of financial resources is an important determinant in the development of ICTs,

16 Except for the number of internet hosts, which are measured per 10,000 people. The period of coverage varies according to data availability – 1975-1998 for main telephone lines, daily newspapers and the number of TV sets, 1990-1998 for the number of personal computers, and 1994-1998 for internet hosts.

except in the case of internet hosts where only human capital is significant. As for human capital, the results confirm that the average years of total schooling influences the establishment of phone lines, personal computers and internet hosts. All variables are significant at the 5% level, while the diffusion of daily newspapers and TV sets does not depend on the years of schooling of the population – human capital is only significant at the 25% level. When we examine the influence of the different levels of schooling the results are somewhat awkward – the average years of primary education do not, in general, influence the development of any of the ICT indicators and even show a negative influence as far as phone lines are concerned; the average years of secondary education show a negative influence on all ICT indicators (negative and significant coefficients) except for the phone lines, where the influence is positive and significant as expected and, finally, the average years of higher education show a positive influence on all ICT indicators as expected, but this is only significant in the case of phone lines.

Table 10 – Human capital and ICT diffusion (Within Groups)

Dependent variable	log(RGDP per capita)	TYR	PYR	SYR	HYR	$\bar{R}^2$	Obs.
Telephone lines	409.4 (3.94)**	-19.14 (-0.951)				0.674	172
	278.37 (2.66)**		-47.4 (-0.614)	-47.01 (-0.357)	717.76 (5.22)**	0.766	172
Personal computer	200.8 (2.11)**	7.8 (0.279)				0.273	32
	44.74 (0.432)**		-362.3 (-1.9)**	428.3 (1.44)	565.65 (0.852)	0.551	32
Internet hosts	108.9 (2.14)**	7.72 (0.314)				0.114	35
	-72.29 (-0.552)		-292.6 (-1.7)*	250.9 (0.926)	1047.6 (1.56)*	0.53	35
Daily papers	66.59 (1.58)*	-14.75 (-1.48)				0.128	63
	32.45 (0.843)**		-12.34 (-0.347)	-126.85 (-2.12)**	503.59 (5.29)**	0.592	63
TV sets	55.58 (1.07)	26.04 (1.96)**				0.530	172
	75.59 (3.25)**		120.8 (1.83)*	-163.17 (-1.57)*	218.42 (1.30)	0.62	172

Notes: *values of the t-Student statistic in brackets. * significant at the 5% level; ** significant at the10% level.*

Considering that there might be country-specific effects in the development of ICTs we used the within-groups estimator, as previously mentioned, to estimate the different relationships. The fit of the equations is not as good as before, especially when the different levels of schooling are included in the regressions, although there are some small $\bar{R}^2$ such as in the case of personal computers, internet hosts and daily newspapers when the average years of total schooling is considered. Again, the availability of financial resources is an important determinant in the development of ICTs, except in the case of internet hosts and daily newspapers when the different levels of schooling are considered, and in the case of TV sets taking into account the average years of total schooling. As for human capital, the results do not confirm that the average years of total schooling influences the implementation of ICTs, with the exception of the diffusion of TV sets – where human capital is significant at the 10% level. When we examine the influence of the different levels of schooling the results are mixed – the average years of primary education only shows a positive and significant influence in the case of TV sets, the influence over personal computers and internet hosts is negative and significant, whilst the remaining influences are not significant; the average years of secondary education shows negative and significant coefficients in the case of daily newspapers and TV sets, while all the other influences are non significant; finally, the average years of higher education show a positive influence over all ICTs indicators as expected but this is only significant in the case of phone lines, daily newspapers and TV sets.

From the tests carried out in this section we can say without doubt that to fully benefit from the diffusion of ICTs that have been responsible for the acceleration of technological progress in recent years, the Mediterranean countries need the financial resources to build the necessary infrastructures and the human capital that enables people to work with these new technologies. The role of the different levels of schooling is not so clear, although one would expect that the diffusion of some ICTs like personal computers and Internet hosts require more than just the basic literacy skills provided by primary education. Some of the results that point to the negative and significant influence of primary and secondary education on the development of ICTs are puzzling.

4. Technological shocks and human capital shocks

In order to confirm or reject one of the main theses at stake in our paper, the role of human capital as a facilitator of technological progress, we have tried to apply a different econometric methodology from those used in the previous sections that would enable us to conduct our empirical research without imposing any *a priori* about exogeneity/endogeneity or substituability/complementary in production technology. The VAR methodology is the one that best suits this purpose. In addition, VAR methodology allows us to analyse the dynamics of the model resulting from different shocks (Sims (1980)). We used the Cholesky decomposition, Enders (1995) because we want to run shocks on each variable without imposing any other constraints upon the error structure of the VAR model. Nevertheless, we controlled the ordering of the variables. We think that the empirical analysis performed in this section is quite original[17] and appropriate to the study of dynamic transition paths.

4.1. The VAR model

In order to ascertain the influence of TFP growth rate shocks and human capital shocks on the economy we built a VAR model in the Sims (1980) tradition. It is a VAR model that applies to all seven economies in the smaller sample and has four variables: real GDP per capita, annual TFP growth rate, investment per capita and the stock of human capital, all expressed in logarithms.

The number of lags was chosen using the BIC criteria and the system stationarity condition: the number of lags for Algeria is two, for Cyprus five, for Egypt two, for Israel two, for Syria two, for Tunisia three and for Turkey two. The shocks simulated over the variables are unit shocks. In the case of the TFP growth rate, the impulses resulting from the shocks were accumulated so in all the figures we have plotted the level of TFP. The number of periods is twenty except for Cyprus, which is thirty, when a unit shock is simulated over TFP growth rate, in order to show that the model is stationary.

Equation (8) below presents the VAR model used in the analysis:

17 The VAR methodology applied to growth empirics is quite recent. See for instance Gali (1999), Kalaitzidakis and Korniotis (2000), Ding (2000), and Kane (2001).

$$
\begin{aligned}
y_t &= const_1 + \sum_{i=1}^{k}\alpha_{1i}y_{t-1} + \sum_{i=1}^{k}\alpha_{2i}h_{t-1} + \sum_{i=1}^{k}\alpha_{3i}inv_{t-1} + \sum_{i=1}^{k}\alpha_{4i}g_{TPF(t-1)} + \varepsilon_{1t} \\
g_{TPF(t)} &= const_4 + \sum_{i=1}^{k}\gamma_{1i}y_{t-1} + \sum_{i=1}^{k}\gamma_{2i}h_{t-1} + \sum_{i=1}^{k}\gamma_{3i}inv_{t-1} + \sum_{i=1}^{k}\gamma_{4i}g_{TPF(t-1)} + \varepsilon_{2t} \\
inv_t &= const_3 + \sum_{i=1}^{k}\delta_{1i}y_{t-1} + \sum_{i=1}^{k}\delta_{2i}h_{t-1} + \sum_{i=1}^{k}\delta_{3i}inv_{t-1} + \sum_{i=1}^{k}\delta_{4i}g_{TPF(t-1)} + \varepsilon_{3t} \\
h_t &= const_2 + \sum_{i=1}^{k}\beta_{1i}y_{t-1} + \sum_{i=1}^{k}\beta_{2i}h_{t-1} + \sum_{i=1}^{k}\beta_{3i}inv_{t-1} + \sum_{i=1}^{k}\beta_{4i}g_{TPF(t-1)} + \varepsilon_{4t}
\end{aligned}
\quad (8)
$$

4.2 Shocks Simulation

In the following we analyse briefly the main effects of the three types of shocks under consideration on the seven Mediterranean economies[18]. Table 11 contains a summary of the different types of shocks and their respective impact on the TFP level, investment, human capital and GDP.

Table 11 – Summary table of the impact of the different shocks

IMPACT / SHOCK	TPF level	Investment	Human capital	GDP
TPF growth rate	Positive and permanent	Initially strong but temporary	Positive and permanent	Positive and permanent
Human capital	Positive and permanent	Initially strong but temporary	Positive and permanent	Positive and permanent
Investment	Positive and permanent	Temporary	Positive and permanent	Positive and permanent

As far as technological shocks are concerned, there is complementarity between technology, physical capital and human capital in the absorption of this shock (see figure C. 1) in all countries except Turkey (see figure C. 2). The same conclusion applies to human capital shocks (see figure C. 3), except for Algeria (see figure C. 4) and Israel (see figure C. 5), which exhibit substitutability between physical capital and human capital in the first four and twelve years, respectively. As for investment shocks, there is complementarity between TFP, physical capital and human capital (see figure C. 6) in the absorption of this shock in all countries except Egypt (see figure C. 7) and Israel (see figure C. 8), which show substitutability between physical investment and human capital.

From the analysis above, we can conclude that the existence of factor complementary between TFP, physical capital and human capital in the absorption of any of the three types of shocks considered, for almost all the seven developing Mediterranean countries, is in accordance with the main results from the previous sections that confirm human capital as a facilitator of technological progress.

18 Syria is a representative country when faced with each of the three types of shocks considered. In the Appendix we include the figures with the response of the different variables to the shocks for Syria, as well as figures for the countries that do not follow the standard responses. Figures for the remaining countries are available from the authors upon request.

5. Concluding remarks

This paper presents evidence on the importance of human capital in technological change in a sample of seven developing Mediterranean countries using the empirical methodologies proposed by Benhabib and Spiegel (1994) and Benhabib and Spiegel (2002), both derived from the Nelson and Phelps (1966) endogenous growth model on the importance of human capital for technological diffusion. Additionally, we tested the importance of human capital in benefiting from the technology embodied in FDI and its role in speeding up the diffusion of ICTs. We also analysed the importance of technological shocks in explaining TFP growth and the importance of human capital in the absorption of these shocks.

The Benhabib and Spiegel (2002) non-linear specification for TFP has the interesting feature of accommodating both the hypothesis of technological convergence and the hypothesis of convergence clubs. The results of the estimation of this specification in a panel data framework using NLLS were, however, very weak. These results led us to conclude that this type of specification does not capture the influence of human capital as a facilitator of technological diffusion. Neither the logistic nor exponential paths for technological diffusion seem to apply to our samples. One possible explanation is the fact that for the smaller sample the level of human capital as a facilitator of technological diffusion is not constrained by a threshold, which would probably happen if we worked with a larger sample as the authors did, including the poorest countries in the world.

Due to the results above, we estimated the Benhabib and Spiegel (1994) linear formulation that we designated the Nelson and Phelps (1966) methodology to avoid confusion. This linear specification assumes an exponential diffusion path, i.e. convergence of the follower to the technological leader. The results were good, especially for the fixed effects and random effects models with annual data. Nonetheless, although the importance of human capital is confirmed by our estimations, its influence is very low when taking into account the value of the estimated coefficient for the technology diffusion term. To conclude, we could say that the Nelson and Phelps (1966) specification seems to capture the process of technological diffusion in our seven countries but the importance of human capital as a facilitator of technological imitation, though confirmed, is small.

The low impact of human capital on imitation activities led us to explore the hypothesis that human capital is more important for TFP growth through embodied technology diffusion than through disembodied technology diffusion. As in Borensztein, Gregorio and Lee (1998) and Lee (2001), we focused on FDI as the major channel of embodied technological diffusion which is only effective if the host country has the necessary human capital available. Although the results from our analysis support the technological catch-up hypothesis and the importance of initial human capital stocks for the TFP growth rate, due to its influence on the domestic innovation rate we were not able to confirm the existence of a complementarity between FDI and human capital. This may be due to the proxy used for the channel of technological diffusion – this analysis should be checked against an alternative channel such as imports of machinery and transport equipment. We also analysed the role of human capital in the diffusion of a particular kind of technology, ICTs, identified as a major source of technological progress in the world today. We considered both the aggregate influence of human capital, which revealed itself to be significant, and also the influence exerted by the human capital acquired through different levels of schooling. In this last case, the results support higher education as a main determinant of the diffusion of ICTs, a result in accordance with the idea that the diffusion of this kind of technology needs more than basic levels of literacy.

Finally, from the analysis of the importance of technological shocks for technological diffusion using a VAR model we concluded that there was factor complementary between TFP, physical capital and human capital in the absorption of any of the three types of shocks considered for almost all of our seven Mediterranean economies, in accordance with the main results of sections 2 and 3, namely the influence of human capital as a facilitator of technical progress. Note, however, that in section 4 we did not control its double role, in the innovation and imitation processes.

These results have to be considered with some care since: a) they are sensitive to the method used to compute the physical capital stock series and the TFP levels and growth rates i.e. TFP was computed based on a Cobb-Douglas aggregate production function that was imposed and not estimated; b) for these countries the usual concerns about data reliability apply, especially as far as human capital data is concerned; and c) other channels of technological diffusion should be considered, such as imports of machinery and transport equipment. These are tasks for future research on technological diffusion in this specific sample of countries.

References

Abramovitz, Moses (1986), «Catching up, Forging Ahead and Falling Behind», *Journal of Economic History*, XLVI,2, 385-406.

_____ (1994), «Catch-up and Convergence in the Postwar Growth Boom and After», in Baumol, W., et al. (eds.), *Convergence of Productivity: Cross-National Studies and Historical Evidence*, New York, Oxford University Press, 86-125.

Aghion, P.; Howitt, Peter (1992), «A Model of Growth through Creative Destruction», *Econometrica*, 60, March, 323-351.

Arellano, Manuel; Bond, Stephen (1991), «Some Tests Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations», *Review of Economic Studies*, 58, 277-297.

Arellano, Manuel; Bover, Olympia (1995), «Another Look at the Instrumental Variable Estimation of Error-Components Models», *Journal of Econometrics*, 68,1, 29-51.

Azariadis, Costas; Drazen, A. (1990), «Treshold Externalities and Economic Development», *Quarterly Journal of Economics*, 501-526.

Baltagi, Badi H. (2001), *Econometric Analysis of Panel Data*, John Wiley and Sons.

Barro, Robert J.; Lee, Jong-Wha (2000), «International Data on Educational Attainment: Updates and Implications», *CID-Harvard Working Paper*, 42.

Barro, Robert J.; Sala-I-Martin, Xavier (1997), «Technological Diffusion, Convergence, and Growth», *Journal of Economic Growth*, 2, 1-26.

Benhabib, Jess; Spiegel, Mark (2002), «Human Capital and Technology Diffusion», in Aghion, P.; S. Durlauf (eds.), *Forthcoming Handbook of Economic Growth*, Elsevier.

_____ (1994), «The Role of Human Capital in Economic Development: Evidence from Aggregate Cross-Country Data», *Journal of Monetary Economics*, 34,2, 143-173.

Blundell, Richard; Bond, Stephen (1998), «Initial Conditions and Moment Restrictions in Dynamic Panel Data Models», *Journal of Econometrics*, 87,1, 115-143.

Bond, Stephen R., *et al.* (2001), «GMM Estimation of Empirical Growth Models», *Centre for Economic Policy Research Discussion PaperNo.* 3048.

Borensztein, E., *et al.* (1998), «How Does Foreign Direct Investment Affect Economic Growth?», *Journal of International Economics*, 45, 115-135.

Caselli, Francesco, *et al.* (1996), «Reopening the Convergence Debate: A New Look at Cross--Country Growth Empirics», *Journal of Economic Growth*, 1, 363-389.

Cette, Gilbert, *et al.* (2004), «The Diffusion of ICTs and Growth of the French Economy over the Long-Term, 1980-2000», *International Productivity Monitor*, 4,Spring, 27-38.

Coe, David T.; Helpman, Elhanan (1995), «International R&D Spillovers», *European Economic Review*, 39,5, 859-887.

Coe, David T., *et al.* (1997), «North-South R&D Spillovers», *Economic Journal*, 107, 134-149.

de la Fuente, Angel (1995), «Catch-up, Growth and Convergence in the OECD», *CEPR Discussion Paper No.*1274.

Ding, Yongling (2000), «VAR Tests of Endogenous Growth Models», *University of Notre-Dame Economic Working Papers*.

Dowrick, Stephen; Nguyen, D. T. (1989), «OECD Comparative Economic Growth 1950-85: Catch-up and Convergence», *American Economic Review*, 79,5, 1010-1030.

Dowrick, Stephen; Rogers, Mark (2002), «Classical and Technological Convergence: Beyond the Solow-Swan Growth Model», *Oxford Economic Papers*, 54, 369-385.

Durlauf, Stephen; Johnson, Paul, A. (1995), «Multiple Regimes and Cross-Country Growth Behavior», *Journal of Applied Econometrics*, 10, 4, 365-384.

Easterly, William (1997), «Africa's Growth Tragedy: Policies and Ethnic Divisions», *Quarterly Journal of Economics*, 112, 1203-1250.

Easterly, William; Levine, Ross (2001), «It's Not Factor Accumulation: Stylised Facts and Growth Models», *The World Bank Economic Review*, 15,2, 177-219.

Easterly, William, *et al.* (1997), «Has Latin America's Post-Reform Growth Been Disappointing?», *Journal of International Economics*, 43, 3-4, 287-311.

Enders, Walter (1995), *Applied Econometrics Time-Series*, New-York: John Willey and Sons.

Engelbrecht, Hans-Jurgen (2003), «Human Capital and Economic Growth: Cross-Section Evidence for OECD Countries», *Economic Record*, 79,Special Issue, S40-S51.

_____ (2002), «Human Capital and International Knowledge Spillovers in TFP Growth of a Sample of Developing Countries: An Exploration of Alternative Approaches», *Applied Economics*, 34, 831-841.

_____ (1997), «International R&D Spillovers, Human Capital and Productivity in OECD Economies: An Empirical Investigation», *European Economic Review*, 41, 1479-1488.

Frantzen, Dirk (2000), «R&D, Human Capital and International Technology Spillovers: A Cross-Country Analysis», *Scandinavian Journal of Economics*, 102, 1, 57-75.

Gali, Jordi (1999), «Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?», *American Economic Review*, 89, 1, 249-271.

Grossman, Gene M.; Helpman, Elhanan (1992), *Innovation and Growth in the Global Economy*. Cambridge, Massachusetts: The MIT Press.

Hall, Robert E.; Jones, Charles I. (1999), «Why Do Some Countries Produce So Much More Output Than Others?», *Quarterly Journal of Economics*, 83-116.

Hoeffler, Anke (2000), «The Augmented Solow Model and the African Growth Debate», *CID--Harvard Working Paper No.*36.

Im, K., *et al.* (2003), «Testing for Unit Roots in Heterogeneous Panels», *Journal of Econometrics*, 115,1, 53-74.

Islam (1995), «Growth Empirics: A Panel Data Approach», *Quarterly Journal of Economics*, 110.

Kalaitzidakis, Pantelis; Korniotis, George (2000), «The Solow Growth Model: Vector Autoregression (VAR) and Cross-Section Time-Series Analysis», *Applied Economics*, 32, 739-747.

Kane, Richard (2001), «Investigating Convergence of the Us Regions: A Time-Series Analysis», *The Journal of Regional Analysis and Policy*, 31, 1, 1-22.

Keller, Wolfgang (2004), «International Technology Diffusion», *forthcoming Journal of Economic Literature*.

Klenow, Peter J.; Rodriguez-Clare, Andres (1997), «The Neoclassical Revival in Growth Economics: Has It Gone Too Far?», in Bernanke, B.; J. Rotemberg (eds.), *NBER Macroeconomics Annual*, Cambridge, MA, MIT Press, 73-102.

Lee, Jong-Wha (2001), «Education for Technology Readiness: Prospects for Developing Countries», *Journal of Human Development*, 2, 1, 115-151.

Levin, A.; Lin, C-F. (1993), «Unit Root Tests in Panel Data: New Results», *Working Paper, University of California, San Diego*.

Lucas, Robert E. (1988), «On the Mechanics of Economic Development», *Journal of Monetary Economics*, 22, July, 3-42.

Mankiw, N. Gregory, *et al.* (1992), «A Contribution to the Empirics of Economic Growth», *Quarterly Journal of Economics*, 107, 2, 407-437.

Mayer, Jörg (2001), «Technology Diffusion, Human Capital and Economic Growth in Developing Countries», *UNCTAD Discussion Paper No.* 154.

Nelson, Richard R.; Phelps, Edmund S. (1966), «Investment in Humans, Technological Diffusion and Economic Growth», *American Economic Review*, 56,1/2, 69-75.

Nickell, Stephen (1981), «Biases in Dynamic Models with Fixed Effects», *Econometrica*, 49, 6, 1417-1427.

OECD (2003), *Geographical Distribution of Financial Flows to Aid Recipients*, Paris: OECD.

Papageorgiou, Chris (2003), «Distinguishing between the Effects of Primary and Post-Primary Education on Economic Growth», *Review of Development Economics*, 7, 4.

Potterie, Bruno van Pottelsberghe de la; Lichtenberg, Frank (2001), «Does Foreign Direct Investment Transfer Technology across Borders?», *Review of Economic and Statistics*, 83, 3, 490-497.

Romer, Paul (1990a), «Endogenous Technical Change», *Journal of Monetary Economics*, 96, S71-S102.

_____ (1990b), «Human Capital and Growth: Theory and Evidence», *Carnegie-Rochester Conference Series on Public Policy*, 32, 251-286.

_____ (1986), «Increasing Returns and Long-Run Growth», *Journal of Political Economy*, 94, 5, 1002-1037.

Schreyer, Paul (2000), «The Contribution of ICTs to Output Growth: A Study of the G7 Countries», *OECD Economics Department Working Paper No.* 2.

Sims, C. (1980), «Macroeconomics and Reality», *Econometrica*, 48, 1, 1-48.

Temple, Jonathan (2001), «Growth Effects of Education and Social Capital in the OECD», *OECD Economic Studies*, 33,II, 57-101.

_____ (1998), «Robustness Tests of the Augmented Solow Model», *Journal of Applied Econometrics*, 13, 4, 361-375.

Van Ark, Bart, *et al.* (2003), «The Contribution of ICT-Producing and ICT-Using Industries to Productivity Growth: A Comparison of Canada, Europe and the United States», *International Productivity Monitor*, 6, 56-63.

Xu, Bin (2000), «Multinational Enterprises, Technology Diffusion, and Host Country Productivity Growth», *Journal of Development Economics*, 62, 2, 477-493.

Appendix

A. The physical capital stock series

We followed the Klenow and Rodriguez-Clare (1997) methodology to compute the series of physical capital stock.

The initial stock of physical capital was estimated according to the formula,

$$\text{(A.1)} \quad \left(\frac{K}{Y}\right)_{it_0} = \frac{\left(\frac{I}{Y}\right)_i}{\gamma_i + \delta_i + \eta_i}$$

where $\left(\frac{I}{Y}\right)_i$ represents the average investment rate of country i over period 1960-2000; γ_i represents the GDP per capita average growth rate of country i over period 1960-2000; andδ_i is the depreciation rate, equal to 0.03 by assumption.

Equation (A.1) can be written as,

$$\text{(A.2)} \quad K_{it_0} = \left[\frac{\left(\frac{I}{Y}\right)_i}{\gamma_i + \delta_i + \eta_i}\right]\left(\frac{Y}{POP}\right)_{it_0} POP_{it_0}$$

To apply the perpetual inventory method we considered t_0=1959. Under this assumption equation (A.2) becomes,

$$\text{(A.3)} \quad K_{i1969} = \left[\frac{\left(\frac{I}{Y}\right)_i}{\gamma_i + \delta_i + \eta_i}\right]\left(\frac{Y}{POP}\right)_{i1960}\left(\frac{1}{1+r_{y1960}}\right) POP_{i1960}\left(\frac{1}{1+r_{POP1960}}\right)$$

where $r_{pop1960}$ represents the average growth rate of the population of country (i) over 1960-2000 and r_{y1960} represents the average growth rate of real GDP per capita of country i over 1960-2000.

Real investment of country (i) at time (t), I_{it}, is computed using the formula,

$$\text{(A.4)} \quad I_{it} = \left(\frac{I}{Y}\right)_{it}\left(\frac{Y}{POP}\right)_{it} POP_{it}$$

Finally, the physical capital stock series is computed using the perpetual inventory method, according to the formula,

$$\text{(A.\%)} \quad K_{it} = \sum_{j=0}^{t}(1-\delta)^{t-j} I_{ij} + (1-\delta)^{t} K_{1959}$$

B. Panel unit root tests of the TFP growth rate series

Conventional panel data econometric estimation procedures can only be applied if the TFP growth rate series is stationary. We have used the IPS (Im, Pesaran and Shin) and the LL (Levin and Lin) panel unit root tests to check whether the TFP growth rate series is stationarity. The LL unit root test assumes heterogeneity of the coefficient on the lagged variable. As we can see from the results in table B1, for all tests we can reject the presence of a unit root, so we applied classical econometric methods to estimate the Benhabib and Spiegel (2002) and the Nelson and Phelps (1966) equations.

Table 1 – Unit-Root Panel Tests for the TFP Growth Rate Series		
	7 Countries Sample	13 Countries Sample
TFP Growth rate	t_δ	t_δ
LL_1	22.33 (0.0)	31.31 (0.0)
LL_2	32.63 (0.0)	47.19 (0.0)
LL_3	48.49 (0.0)	74.51 (0.0)
	$\bar{Z}$	$\bar{Z}$
ADF without trend	-12.25 (0.0)	-14.61 (0.0)
ADF with trend	-11.81 (0.0)	-13.78 (0.0)

Note: In square brackets we have the level of probability; ADF $\bar{Z}$ test is the test proposed by Im, Pesaran and Shin (2003) and t_δ test corresponds to the equations in Levin and Lin (1993) for the null of unit root. LL_1: $\Delta Y_{it} = \delta_i Y_{it-1} + e_{it}$; LL_2: $\Delta Y_{it} = \alpha_i + \delta_i Y_{it-1} + e_{it}$; LL_3: $\Delta Y_{it} = \alpha_{0i} + \alpha_{1i} T + \delta_i Y_{it-1} + e_{it}$.

C. Graphical analysis of shocks

C.1 Technological shocks

Figure C.1 – Syria, Unit shock on dTFP

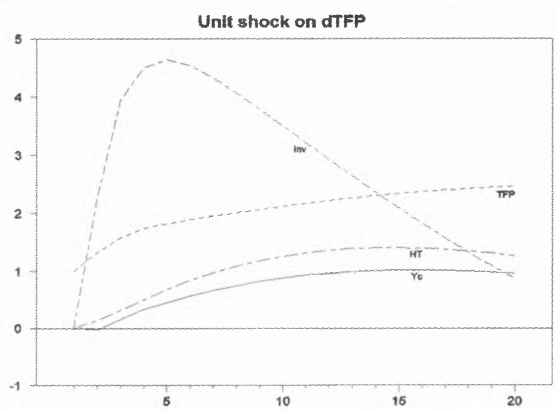
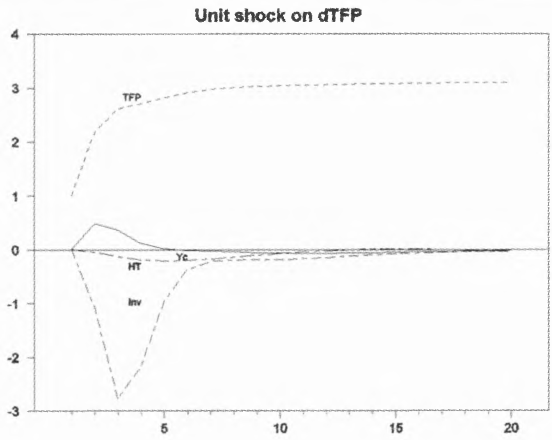


Figure C.2 – Turkey, Unit shock on dTFP



C.2. Human Capital shocks

Figure C.3 – Syria, Unit shock on HT

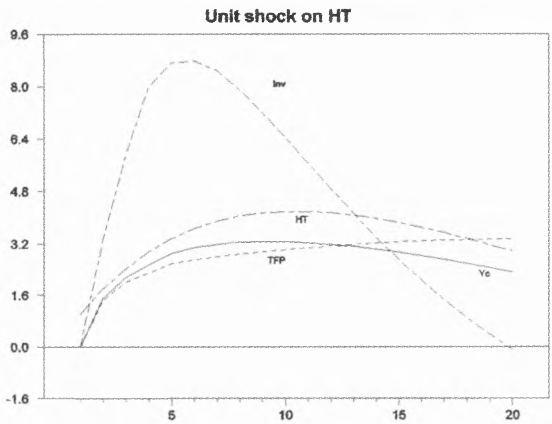


Figure C.4 – Algeria, Unit shock on HT

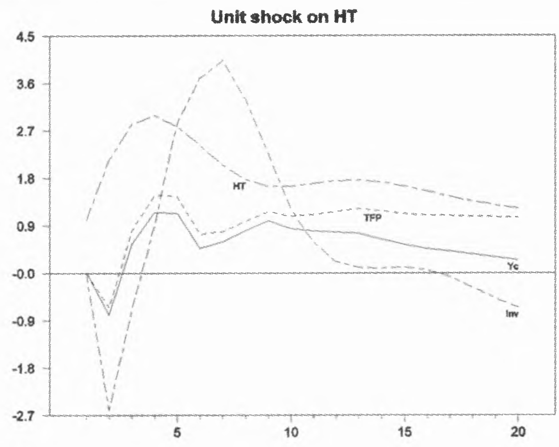
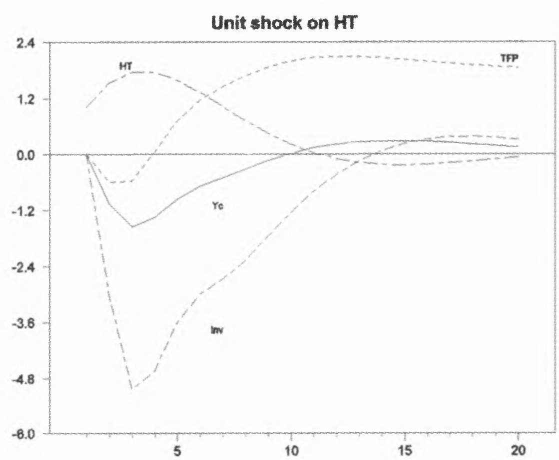


Figure C.5 – Israel, Unit shock on HT



C.3. Investment shock

Figure C.6 – Syria, Unit shock on Investment

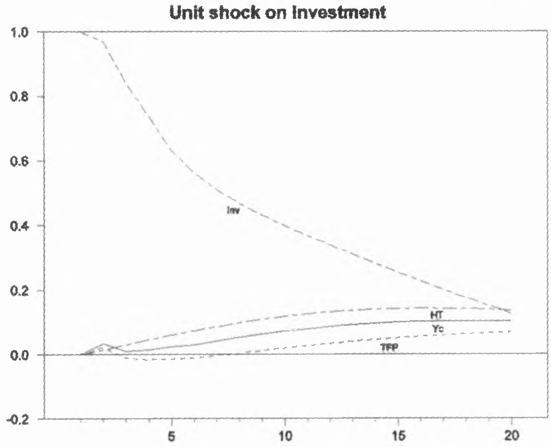


Figure C.7 – Egypt, Unit shock on Investment

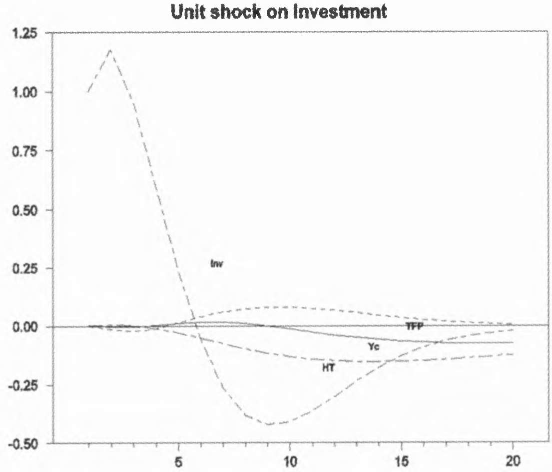
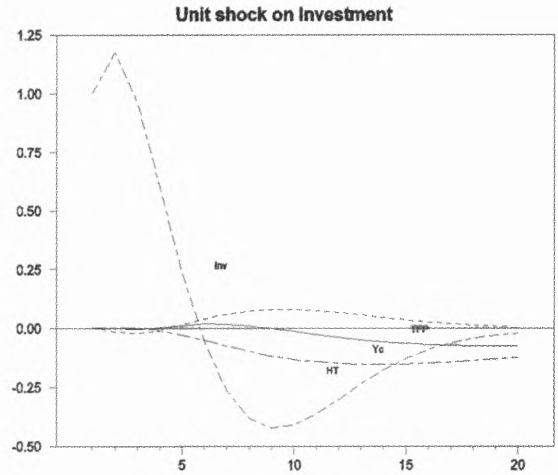


Figure C.8 – Israel, Unit shock on Investment



The German Labour Markets – The Case for More Flexibility

Peter Rühmann
University of Göttingen – Department of Economics

resumo

Com base num modelo com diversos segmentos do mercado de trabalho, este artigo tenta descrever as características do mercado de trabalho alemão. Destacam-se dois processos que têm vindo a deteriorar a situação, anteriormente favorável, desse mercado: a subcontratação (*outsourcing*) da produção à Europa central e oriental, e a deslocação de mão-de-obra para a economia subterrânea. Afirma-se que a economia subterrânea não pode substituir a economia formal quando esta enfrenta grandes pressões e se exige maior flexibilidade. O modelo utilizado parece confirmar, em termos gerais, as actuais características do mercado de trabalho alemão, revelando que o processo de ajustamento resultou mais em alterações do emprego e do PIB (oficial) do que em alterações salariais. Por último, o artigo analisa o que pode vir a acontecer se as recentes concessões salariais isoladas se transformarem numa redução geral dos níveis salariais.

résumé / abstract

Cet article utilise un modèle qui comprend plusieurs secteurs du marché du travail pour essayer de décrire les particularités du marché du travail en Allemagne. Deux caractéristiques sont mises en relief: la sous-traitance de la production aux pays de l'Europe centrale et orientale et le transfert de main-d'œuvre à l'économie parallèle. On affirme que l'économie parallèle ne peut pas remplacer l'économie formelle lorsque celle-ci subit des pressions qui demandent une plus grande flexibilité. Le modèle utilisé semble, d'une manière générale, confirmer les caractéristiques du marché allemand, montrant que le processus d'ajustement a eu comme résultat des changements de l'emploi et du PIB (officiel) plutôt que des salaires. Finalement, cet article analyse ce qui peut arriver si les récentes concessions salariales isolées deviennent une baisse générale des salaires.

A framework of different labour market segments is introduced to describe the stylised facts of the German labour market. Two processes, the outsourcing of production to Middle and Eastern Europe and the transfer of jobs to the shadow economy, are distinguished, which are eroding the formerly favourable labour market situation in Germany. It is argued that the shadow economy cannot operate as a substitute for the regular economy if the latter is put under strain and more flexibility is required. The framework used here seems to be in broad agreement with the stylised facts, showing that the process of adjustment has resulted mainly in changes to employment and the (official) GDP rather than to wages. Finally, the paper deals with the consequences that may arise if the recent emergence of selective wage concessions turns into a reduction in overall wage levels.

A framework of different labour market segments is introduced to describe the stylized facts of the German labour market. Two processes, the outsourcing of production to Middle and Eastern Europe and the transfer of jobs to the shadow economy, are distinguished, which are eroding the formerly favourable labour market situation in Germany. It is argued that the shadow economy cannot operate as a substitute for the regular economy if the latter is put under strain and more flexibility is required. The framework used here seems to be in broad agreement with the stylised facts, showing that the process of adjustment has resulted mainly in changes to employment and the (official) GDP rather than to wages. Finally, the paper deals with the consequences that may arise if the recent emergence of selective wage concessions turns into a reduction in overall wage levels.

1. The Demand dor More Flexibility

In the words of Hans-Werner Sinn, president of the Munich Ifo Institute, "Germany is the sick man of Europe, ranks lowest in growth and is unable to keep up with its neighbours" (Sinn, 2003: 1). Even if we focus our attention on the level of the GDP (per inhabitant) rather than on growth, Germany now (2003) ranks only seventh in the EU-12 (Statistisches Bundesamt, 2004a: Table 15.2). Moreover, an insidious process of outsourcing production to Middle and Eastern Europe is taking place. However, not all of Germany's economic performance criteria are looking bleak. Within the group of euro area members, Germany's share of exports has increased slightly from 28.5% (1999) to 29.6% (2002) (Sachverstaendigenrat, 2003: Table 7*, 1995 prices) and German companies seem to be gaining in intra-euro area price competitiveness (Compare ECB, 2004: Box 5). Nevertheless, doubts remain as to whether Germany will overcome its economic difficulties within the foreseeable future.

Looking at the causes of unsatisfactory performance, there is some agreement on the fact that the labour market is at the heart of Germany's economic difficulties. Unemployment is now stagnating at a comparatively high rate, with long-term unemployment representing about one third of the total (Bundesanstalt fuer Arbeit, 2003). Moreover, since the social security system is financed to a large extent by a pay-as-you-earn scheme, its sustainability is threatened. Based on these observations, many German economists have declared that the German labour markets are overregulated in all respects, including employment protection, labour standards, a rigid system of collective wage bargaining which is backed by law and a generous welfare system (see, for instance, Franz, 2002). The inability of firms to adapt, a spread of wages that is too small to effectively combat unemployment amongst lower qualified workers and wage levels that are not compatible with a sustainable position in the international product markets are consequences which the authors deplore.

These arguments concentrate on the highly regulated and relatively well-protected area of the German labour market. However, if we look more carefully at developments in Germany, it is possible to find indications of considerable flexibility in several areas. Paralleled by a drop in union membership from one third to one quarter during the nineties (Nickell, 2003: Table 9), collective bargaining agreements have declined from over 80% in the eighties to about 70% at the end of the century. Correspondingly, the percentage of employees integrated into the social security system ("Sozialversicherungspflichtig Beschaeftigte") dropped from 90% in 1993 to 84% in 2003 (calculated from Statistisches Bundesamt, 1997 and 2004b: Tables 3 and 3.10), whilst various forms of atypical work have developed. On the other hand, even within the range of collective bargaining, the so-called "opening" clauses allow for deviations from the provisions of industry-wide collective arrangements (see OECD, 2004: 145, 154-155).

After describing the main features of the highly regulated standard employment practices in Germany and contrasting them with the various forms of employment that have developed alongside them, this paper will present a graphical interpretation of the labour market situation (Section B). This interpretation distinguishes several types of labour markets in order to allow for employment relationships that range beyond the standard employment practices that remain highly regulated and fully integrated into the social security system. Subsequently, based on this analysis, the process of outsourcing, which challenges the traditional system of wage bargaining in Germany (Section C), will be investigated.

2. The German Labour Markets

2.1. Main Features

We begin our analysis of how the German labour markets function by describing the *Tarifautonomie* (the system for the free determination of wages and terms of employment by employers' associations and unions; for details see Franz, 1999: 233-237). Typically but not exclusively, such collective labour contracts are negotiated between unions and employers' associations within the industry. Despite a fall in collective bargaining and union membership, it seems safe to say that up to the present day the unions have had a considerable influence on labour contracts – at least in western Germany, where collective bargaining is considerably more widespread than in eastern Germany.

Generally, within the framework of collective bargaining, entrants' wages are bound to the wages of the incumbents. Even if some collective labour contracts contain clauses which allow lower entrants' wages, these clauses are more of an exception than a rule (Sachverstaendigenrat, 2001: 232).

The influence of the unions is backed by *labour market regulations*, especially by the institution of works councils (*Betriebsraete*) (for details, see Addison et al., 2001). At present, these must be established in any company that employs more than five people (full-time). Members are elected by the company employees and, in the bigger companies at least, a large number of them are union members.

By law, the works councils possess certain rights in relation to firms, especially with regard to both individual and collective dismissals. For instance, in case of collective dismissals the companies have to negotiate a social plan which settles severance pay and/or contains special arrangements for early retirement programmes. Moreover, the social plan may provide for programmes to help dismissed workers find new employment, by improving their qualifications, for instance. These activities work in the interests of the incumbents, either by increasing the costs of their dismissal and thereby making it less likely or by determining the financial (and other) assistance to be provided by the company if dismissals cannot be avoided. Moreover, in the latter case the social security system offers financial support, not only through the usual unemployment benefits but also in the form of certain early retirement programmes introduced in the seventies and eighties. However, these generous arrangements in favour of insiders have had to be financed, meaning that they have driven an increasing wedge between producer and consumer wages and have been detrimental to employment. Furthermore, the position of employees with regular contracts is strengthened by employment protection legislation that is considered to be comparatively strict in Germany, with only the Netherlands, Portugal and Sweden ranking higher within the EU-15 (OECD, 2004: Table 2.A2.4). Nevertheless, it should be borne in mind that – with some provisions – small companies with no more than ten employees are exempt from the most trenchant aspects of employment protection.

Recently, faced with the threat of companies shifting production abroad, employers and employees have agreed to deviations from the provisions of the industry-wide collective wage contracts. Notable examples have involved the car manufactures DaimlerChrysler and Volkswagen. Typically, employees have accepted concessions in wages and working hours in exchange for a promise from the company not to shift production to Eastern Europe and to avoid dismissals.

Besides the standard employment relationship that is classified as an unlimited full-time job with all the usual labour and social rights, different kinds of atypical work have also developed (Buch and Ruehmann, 1998). These include temporary work, fixed-term contracts, part-time work, *mini-jobs* and dependent self-employment. In 1995 more than one third of all employment relationships belonged to this category. So far, the most important of these atypical employment relationships have been the *mini-jobs* (formerly *Geringfuegige Beschaeftigung*), amounting in 2004 to about seven million jobs and representing, at present, maximum monthly earnings of 400 EUR. Obligations and rights in relation to the social security system and taxation have been changed several times but have always been limited, in comparison with regular employment where the usual labour market regulations apply (Buch, 1999: 138-147). *Mini-jobs* are attractive in combination with shadow economy activities, since it is difficult to control whether individuals who declare they are working under the *mini-job* rule earn more than the 400 EUR allowance and evade taxes and contributions for the extra they are paid. Even discounting illegal activities, it is significant that it is possible to reduce taxes and social security contributions by hiving off part of one's occupation as a *mini-job*.

Finally, the shadow economy, which has expanded in Germany during recent decades, must also be taken into account when considering the various segments of the labour market. According to estimates by Schneider and Enste (2000), its labour force amounted to more than 20% of the total labour force in the late nineties, producing goods and services totalling about 15% of the GDP. Compared with its size in the late eighties (about 12% of the GDP), shadow economy activities have increased markedly faster than the regular economy. Including the estimates of Enste and Schneider, it may be surmised that, at present, about six out of ten employment relationships are covered by collective bargaining contracts, about one quarter of all contracts belong to the regular economy but are not covered by collective bargaining agreements, and the rest (about one sixth) are located within the shadow economy.

2.2. An Interpretation of the Labour Market Situation

Referring to section 2.1., the assumption that a homogeneous labour market exists is inadequate in any proper discussion of the German labour market and the need for flexibilization. Instead, this interpretation will follow the lines of the insider-outsider approach.

The highly regulated and collectively organized area of the labour market is treated as the insider market, whilst the segments of atypical but formal employment and of shadow activities form two types of outsider market. In fact, the various segments of the labour market are heterogeneous in terms of the wages to be earned in them. For example, collectively negotiated wages can drop to less than five euros per hour (WSI Tarifarchiv, quotation from SZ, 2004). On the other hand, the jobs of high-wage earners may be fully integrated into the social security system without being covered by collective bargaining agreements. However, to simplify the following considerations it will be assumed that the insider market is the high-wage segment of the labour market, whilst the wages of outsiders are considerably lower.

We will start with the *insider* labour market segment. In principle, the bargaining position of the insiders is contestable. According to the insider-outsider approach, the upper limit for insider wages is to be found in the reservation wage of the outsiders *plus* the turnover costs due to the replacement of insiders by outsiders, i.e. in the relative profitability constraint (Lindbeck and Snower, 1988: 5). But it is doubtful whether the latent competition of outsiders is effective in those sectors within western Germany where collective wage arrangements prevail. With the exceptions mentioned above, entrants' wages in these sectors are tied to insider wages. Therefore, the effective underbidding of wages by outsiders is prevented.

After the fall of the Iron Curtain and the reintegration of the former Communist countries within the world economy, the position of western Germany's insiders has been contested by workers from both eastern Germany and Middle and Eastern Europe. But for reasons that will not be discussed here in detail, it seems that the pressure exerted by eastern Germany's labour market is not as strong as might have been expected. The labour markets of both Germanies were mainly composed of commuters from East to West Germany, with a balance of 440 thousand in 2001 (Ragnitz, 2002). In general, these commuters were bound to existing collective labour contracts in West Germany. Of course, competition by the East German labour force was perceptible when firms outsourced their production to East Germany or when East German companies could underbid their western competitors. However, wages and productivity are both, on average, lower in eastern than western Germany, with the result that unit labour costs in manufacturing are only slightly lower in eastern than western Germany (Sachverstaendigenrat, 2002: 186). Thus, it is no surprise that nowadays production outsourced not to eastern Germany but to Middle and Eastern Europe appears to be the most effective threat to the position of western German insiders.

In contrast to the replacement of insiders by domestic outsiders, outsourcing means a shift in the location of jobs and requires advance preparation in order to adapt to differences in administrative regulations and culture, as well as the building up of new plants. In general, this alternative not only requires more time but is also associated with greater imponderability in comparison with the replacement of insiders by domestic outsiders. Therefore, outsourcing production abroad is proceeding slowly. The implications of this process will be considered in section 3.

The graphical interpretation of the insider labour market contains a wage demand (WD_I) curve and an aggregate labour supply (LS_I) curve. The WD_I curve implies that individual firms are wage takers on the labour market. Referring to the arguments in section 2.1., at least until recently the German unions possessed sufficient bargaining power to influence insider wages effectively. The efforts of companies to outsource production abroad indicate that in many cases companies estimate the marginal product of labour as being lower than the wage they have to pay for it. This situation is marked by point A in Figure 1, which means that the unions have realized the maximum insider wage (W_I) that is compatible with the maintenance of insider employment at the moment.

As argued above, competition from domestic outsiders is likely to be ineffective in Germany, which means that only the *absolute* rather than the relative profitability constraint is relevant in relation to insider wages. Since dismissal costs exist, the absolute profitability constraint expressed by the APC curve, is not identical to the WD_I curve, but is shifted upwards by d which represents the (annuity of the) costs involved in dismissing insiders. Apart from other influences, d increases in relation to the demands of employment protection.

According to the interpretation of Figure 1, the reservation wage of foreign outsiders *plus* the turnover costs due to the replacement of insiders (R_f^+) lies well below the insider wage in Germany, giving rise to a steady process of outsourcing. However, until recently R_f^+ was not perceived as a strict restriction by unions when defining their wage demands in industry-level negotiations. Foreign competition seemed to be too far away geographically and outsourcing too dispersed with respect to time and companies to merit serious consideration in the process of wage bargaining at industry level. This situation, which has changed recently (See Section 2.1.), is illustrated by Figure 1, whereas the following section 3 will deal with the changes in wage bargaining that are evolving at the moment.

The starting point A represents a kind of momentary equilibrium for insider employment N_I and a level of unemployment determined by the difference LS_I minus N_I. Since human capital will depreciate gradually during unemployment, former insiders who have become unemployed loose their status after some time, and the pool of insiders will decrease gradually, shifting the LS_I curve to the left. Accordingly, the excess of supply over demand tends to be small in the insider market.

The market for *outsider* employment comprises atypical work and low-wage employment. This latter aspect of the outsider market is characterized by little union influence and rather insignificant employment protection (for instance in small companies with no more than ten employees, see Section 2.1.), but it is integrated into the social security system. Atypical work comprises several different kinds of participants on the supply side, ranging from insiders who earn an additional salary with a *mini-job,* to the registered unemployed who supplement their unemployment benefits in this way. The common feature in all these employment relationships is the lack of union influence on wage-setting and, to a certain extent, less employment protection than in the insider labour market. For the sake of simplicity, this market is specified as a competitive market. Wages are subject to the pressures of the reservation wages of those who would like to participate in this market but have not yet succeeded. In Figure 2 it is assumed that this reservation wage is the same for everyone involved in the supply side of the market (R_O). Turnover costs are neglected, since they are much lower than those in the well-protected insider market. In fact, R_O is mainly determined by the level of unemployment benefits.

As far as *shadow activities* are concerned, two types of activity may be identified, depending on the origins of the workers (domestic or foreign) involved in these markets. Both areas of the market are specified as competitive, but they differ in terms of supply. While domestic supply is restricted in this market, as it is in the other labour markets, the supply offered by foreign workers seems to be highly elastic, or even indefinitely elastic. Moreover, due to the German welfare system, the reservation wage of domestic suppliers is assumed to be higher than that of the foreign workers. On the other hand, it tends to be lower than the outsider market, which is subject to taxation and contributions, even though they are distinctly lower than those levied on the insider market. Figure 3a) reflects excess supply because at the moment it seems to be easy to hire a domestic worker for a shadow activity.

3. The Process of Erosion and its Consequences

3.1. The Process of Erosion

The system of different labour market segments is similar to a system of communicating pipes with barriers which impede any equalizing processes. As argued above, the process of outsourcing will require time. The same is true in relation to the decision to transfer an outsider job to the shadow economy, since the participants have to overcome the social norms. As a result, the equilibration processes will take full effect only after a long period of time, if at all. As long as these effects work, momentary equlibria will emerge which will be investigated here.

In the first instance, the process of erosion affects the insider segment of the labour market. In Figure 1, the initial position A will be eroded by *outsourcing* production abroad, shifting the WD_I (and with it the APC) curve inwards. This process reduces insider employment due to the assumption that the competition of foreign outsiders is too dispersed to be taken into account by unions in their wage demands. Consequently, insider unemployment will rise temporarily. It is certain that in the case of outsourcing abroad, the drop in demand for insider employment will not be compensated by an increase in the demand for domestic labour in one of the two other segments of the labour market. As argued in section 2, unemployed insiders loose their status over time and will join the supply side of the two other areas of the labour market. However, elastic supplies tend to prevent wage reductions and the rise in employment in these market segments. Consequently, within the framework presented here, the decrease in insider employment and the increase in *overall* unemployment will be the sole effects on employment and unemployment. As a result, shadow activities do not compensate for the loss of insider employment.

If an insider job is *transferred to the outsider segment* or an outsider job to the *shadow economy* these shifts of demand have no effect on wages but only on employment. As the combined result of both operations, the official GDP will be depressed while the extended GDP (including the outcome of shadow activities) will remain unchanged as long as productivity in all the labour market segments remains the same. If, on the other hand, productivity in the shadow economy is lower than in the other segments, the reduction in the GDP will not be fully compensated by shadow activities and the extended GDP will be depressed as well. To the extent that activities which have been transferred to the shadow economy were formerly undertaken by low-wage workers integrated into the welfare system, (registered), unemployment will rise.

In all cases the transfer of jobs will put the welfare state under strain. Given the way the German pay-as-you-earn system functions, the loss of jobs that had been integrated into the social security system will tend to raise contributions in the short term. As long as the increase in contributions is not fully compensated by decreasing the net wages of employees, gross wages, which are relevant to the employers, will rise and once more will provide an incentive for reducing those jobs which are integrated into the social security system. A vicious circle may emerge, which will make the labour market situation deteriorate even further.

According to the preceding analysis, the process of erosion of highly regulated employment takes place without changes to wages and affects the GDP negatively. This corresponds broadly to the stylised facts. In the nineties there was a low growth rate in the GDP and (registered) unemployment increased in general, whilst the share of employees integrated into the social security system dropped. On the other hand, wage dispersion remained almost unchanged from the eighties to the nineties of the last century (Compare OECD, 2004: 141). This seems to be compatible with the preceding analysis where the adjustment process resulted in changes in employment but not in wages. Finally, the process of outsourcing leads to a comparatively high net loss of jobs in Germany, contrary to the situation in the United States (McKinsey, quotation from Economist, 2004). This is probably due to the German labour market institutions. Doubtless the interpretation presented here is just a first approximation towards explaining this phenomenon, since it predicts that the rate of loss amounts to 100% instead of 60%, as estimated by McKinsey.

3.2. The Trend towards Flexibilization in the Labour Market

The process of erosion which has taken place since at least the nineties may have changed its character recently to a certain extent. As mentioned in Section 1, within the framework of collective bargaining opening clauses are increasingly used due to the threat of outsourcing on the part of the companies. The resulting concessions made by employees probably do not prevent outsourcing completely, but the process will continue at a lower rate. Consequently, wages are reduced selectively (and not generally), depending on the credibility of the threat to outsource production, and the impact on the overall wage level may be limited. In principle, the development which is described in the preceding section could continue for some time, at the price of steadily shrinking employment in the insider segment of the labour market.

The consequences of a permanent worsening of employment opportunities depend on the reaction of the participants in this market segment, especially the employees. In the course of time, the system of industry-wide collective bargaining could break down. Such a fundamental institutional change may result in overall (instead of selective) wage reductions. This would provoke a dilemma: It seems unlikely that reductions in the overall wage level in the insider labour market segment will prevent further outsourcing or give rise to growth in the workforce as long as they are moderate (Compare Figure 1). Only a wage reduction that is large enough to reach the lower part of the WD_I curve could contribute towards increasing insider employment, if at all, or at least towards preventing further decreases. However, such a substantial reduction in the overall wage level is likely to improve the international competitiveness of companies considerably and to increase the already existing trade surplus of Germany even more. This might provoke wage reductions in other countries and, in the final analysis; the emergence of a deflationary spiral cannot be discounted.

References

Addison, John T. *et al.* (2001), «Works Councils in Germany: Their effects on establishment performance», *Oxford Economic Papers*, 53, 659-694.

Buch, Holger (1999), *Ungeschuetzte Beschaeftigungsverhaeltnisse. Scheinselbstaendigkeit und geringfuegige Beschaeftigung auf dem deutschen Arbeitsmarkt,* Frankfurt am Main, Peter Lang.

Buch, Holger; Ruehmann, Peter (1998), «Atypical Work as a Form of Low-Wage Employment in the German Labour Markets», in Bazen, St. Et al. (eds) *Low-Wage Employment in Europe,* Cheltenham, UK; Northampton, MA., Edward Elgar, 111-126.

BUNDESANSTALT FUER ARBEIT (ed.) (2003), *Strukturanalyse: Bestand an Langzeitarbeitslosen, Arbeitsmarkt in Zahlen*, Nuernberg.

ECB (2004) *Monthly Bulletin*, May.

ECONOMIST (2004), *A survey of outsourcing*, November 13th 2004.

Franz, Wolfgang (1999), *Arbeitsmarktoekonomik,* 4., ueberarb. Aufl. Berlin, Springer-Verlag.

Franz, Wolfgang (2002), «Zur Lage auf den Arbeitsmarkt: Reformvorschlaege in der Diskussion, Flexibilitaetserfordernisse auf dem deutschenArbeitsmarkt», *Zeitschrift fuer Wirtschaftspolitik*, 51, 179-188.

Lindbeck, Assar; Snower, Dennis J. (1988), *The Insider-Outsider Theory of Employment and Unemployment*, Cambridge, Mass., London, MIT Press.

Nickell, Stephen (2003), «Labour Market Institutions and Unemployment in OECD Countries», *CESifo DICE Report,* 1, 2, 13-26.

OECD (2004), Employment Outlook.

Ragnitz, Joachim (2002), *Arbeitsangebot, Arbeitsnachfrage und ein Loesungsvorschlag fuer das ostdeutsche Arbeitsmarktproblem*, Halle, Institute for Economic Research Discussion Papers 168.

Sachverstaendigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung (2001; 2002; 2003), *Jahresgutachten 2001/02*; *2002/03*; *2003/04*, Wiesbaden.

Schneider, Friedrich; Enste, Dominik H. (2000), «Shadow Economies: Size, Causes, and Consequences», *Journal of Economic Literature*, 38, 77-114.

Sinn, Hans-Werner (2003), The Laggard of Europe, *CESifo Forum*, 4, Special Issue 1.

STATISTISCHES BUNDESAMT (2004a), *International Statistical Yearbook*, Wiesbaden.

STATISTISCHES BUNDESAMT (1997; 2004b), *Statistical Yearbook for the Federal Republic of Germany*, Wiesbaden.

SZ 2004 (Sueddeutsche Zeitung, 237, October 12th, 2004).

Appendix

Figure 1 – German Labour Markets: Insider

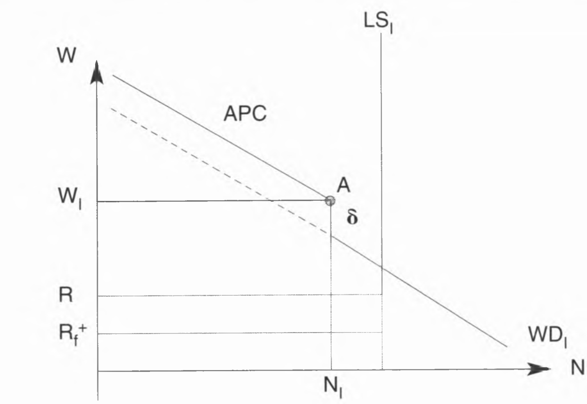


Figure 2 – German Labour Markets: Outsider

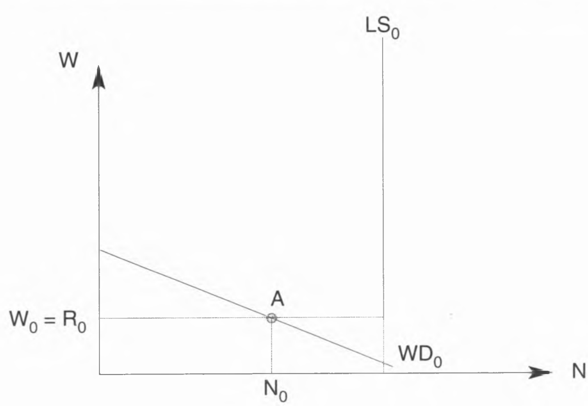


Figure 3a – Shadow Economy: Domestic Workers

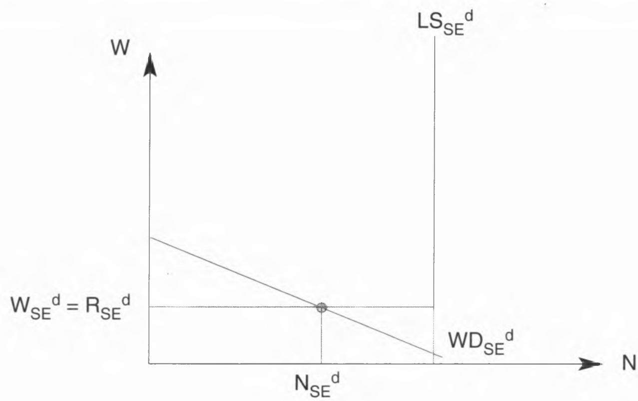
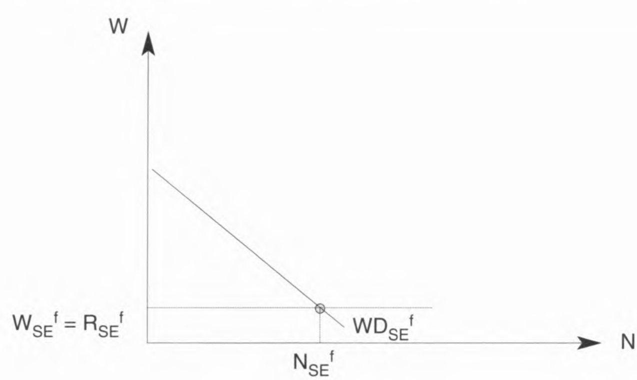


Figure 3b – Shadow Economy: Foreign Workers



The Use of Cost and Time in Project Decision Trees:
A model and an application[1]

Pedro Cortesão Godinho / João Paulo Costa
Faculty of Economics of the University of Coimbra and INESC

resumo

Neste artigo, apresenta-se um modelo bicritério para análise de projectos baseado em árvores de decisão, e uma aplicação deste modelo a um problema de planeamento de produção. Começa-se por apresentar um modelo que permite a utilização do tempo e do custo na análise de projectos. Este modelo é simples de usar, e pode ser aplicado a um grande número de situações reais em que os principais objectivos são a minimização do tempo e do custo. A construção das árvores de decisão correspondentes a este modelo pode exigir tempos de cálculo elevados, e assim tornar-se impraticável. Este problema é abordado de duas formas: primeiro, define-se um algoritmo para a geração das estratégias, e depois introduz-se um parâmetro de erro que permite evitar a geração das estratégias que estejam muito próximas de outras. Finalmente, apresenta-se uma aplicação do modelo a um problema de planeamento de produção. Define-se o problema, utiliza-se o algoritmo para gerar as estratégias eficientes, e analisam-se os resultados.

résumé / abstract

Cet article présente un modèle bi-critère d'analyse de projets ayant pour base des arbres de décision et qui est ensuite appliqué à un problème de planification de la production. On commence par présenter un modèle qui permet d'utiliser le temps et les coûts dans l'analyse de projets. Ce modèle s'emploie facilement et peut être appliqué à un grand nombre de situations réelles dont le but principal est la minimisation du temps et des coûts. La construction des arbres de décision qui correspondent à ce modèle peut exiger des temps de calcul très longs, ce qui peut le rendre impraticable. Le problème est abordé en deux étapes: on commence par définir un algorithme pour générer les stratégies et on introduit ensuite un paramètre d'erreur qui va permettre d'éviter la génération de stratégies très proches les unes des autres. Finalement, on présente l'application du modèle à un problème de planification de la production. On définit le problème, on emploie l'algorithme pour générer des stratégies efficientes et on analyse les résultats.

In this paper we present a useful bicriteria model for project analysis based on decision trees, and an application of the model to a production planning problem. We start by presenting a model that allows the use of time and cost in project analysis. This model is easy to use, and it can be applied to a large number of real-life situations when the main objectives are the minimisation of cost and the minimisation of time. The construction of the decision trees for this model may require large computational times, and thus become impracticable. We deal with this problem in two ways: first, we define an algorithm for generating the strategies, and then we introduce an error parameter, which allows us to avoid generating all the strategies that are very close to each other. Finally, we apply the model to a production planning problem. We define the problem, use the algorithm to identify the efficient strategies and analyse the results.

1 This research was partially supported by FCT, FEDER, project POCTI/EGE/58828/2004.

1. Introduction

Decision trees are often used in project analysis and appraisal (see Brealey and Myers, 2000; Magee, 1964), usually considering only the financial perspective. In many cases, there are certain factors that cannot be incorporated into the financial value of a project and that are very important in deciding whether or not it should be undertaken, so it is often useful to include multiple criteria in the analysis. The efforts to use multiple criteria in the evaluation of project decision trees have been based on multi-attribute utility theory (MAUT). Hertz and Thomas (1983) present an overview of some MAUT-based methodologies for the evaluation of multicriteria decision trees. These methodologies require either the definition of the criteria aggregation rule or the definition of some trade-offs between criteria before rolling back the tree, and they only allow the identification of the "best" alternative. Smith and Nau (1995) also use utility theory to evaluate project decision trees in incomplete markets.

Time is one important criterion that is often overlooked in the project evaluation literature. Godinho and Costa (2002) outline a new approach for the incorporation of time and financial value in project decision trees, which is detailed and analysed in depth in Godinho (2003). This approach is based on the identification of the non-dominated strategies[2]. After all the non-dominated strategies are identified, the decision-maker may use any multicriteria method to choose from them. We will present this approach in Section 2.

This approach may lead to very large trees, whose construction and evaluation may take a very long time, or even be impractical. Therefore, it is useful to build particular models that allow the automatic definition of the tree using a limited set of parameters. Godinho (2003) proposes one such model, which will be presented in Section 3. This model assumes that some different processes may be used to undertake a homogeneous task, each process having a constant cost and requiring a constant time per utilisation, and that there are switching costs and set-up times for changing the process being used. In this model, project advance follows an additive binomial process. After the processes are defined, the complete decision tree can be automatically built, and the non-dominated strategies can be identified. This model is easy to use, because it only requires a limited number of inputs: the parameters of the processes. Also, it can be used as an approximation to some continuous-time problems, and applied to a large number of real-life situations when the main objectives are the minimisation of cost and the minimisation of time, and when the considered task is homogeneous, or can be treated as such. Possible applications include the analysis of some construction projects and production planning.

The decision trees generated by this model are usually very large, and building and evaluating the trees in order to identify the non-dominated strategies may be impractical, because it takes a very long time, or even impossible due to memory limitations of the computer systems. In order to handle this problem, an algorithm for the identification of the non-dominated strategies was proposed in Godinho (2003). We will outline this algorithm in section 4. The algorithm does not require the construction of the complete decision tree, since it is able to disregard non-interesting branches as soon as they come up, and to avoid the repetition of calculations for similar branches of the tree. The algorithm also provides a more compact representation of the tree, since identical branches located in different points of the tree are represented only once.

Computational tests were performed by Godinho (2003) to assess the performance of the algorithm and some results are presented in Section 4.2. These tests have shown that the algorithm performs well when the number of non-dominated strategies is small, but it does not perform so well when that number is large. In order to overcome the problem of model usage with a large number of non-dominated strategies, we introduce a new approach: an error

2 We consider a strategy to be the complete set of decisions that will be made until the end of the project. A strategy is non-dominated if none of the other strategies is better or equal in all the criteria, and strictly better in at least one criterion.

parameter that allows us to avoid generating all the non-dominated strategies that are very close to each other. Using this approximation, we are able to reduce the number of strategies that are generated. This improves the performance of the algorithm, and we are able to apply the model to some cases that we would otherwise be unable to handle.

Finally, we describe an application of the model. We consider a production planning problem, and we calculate the model parameters for that problem. Then we use the algorithm to calculate the non-dominated strategies and we analyse the results.

This paper is structured as follows. Section 2 presents a general approach for the use of time and value (or cost) in decision trees. Section 3 presents the particular model that we use in our application. Section 4 outlines and discusses the algorithm for the identification of non-dominated strategies. Section 5 describes an application of the model. Finally, we present our conclusions in Section 6.

2. An approach for the use of bicriteria decision trees in project analysis

This section presents the general approach for the use of time and financial value in project decision trees that was proposed by Godinho and Costa (2002). We assume that we are maximising the financial value, as measured by the Net Present Value (NPV), and minimising the time. This approach focuses on the identification of the non-dominated strategies, allowing the decision-maker to choose one of these alternatives using a multicriteria method.

The evaluation of the decision tree is a two-step process. In the first step, time increments and cash flows (or other value increments) are forwarded to the leaves, in order to calculate the criteria values for each leaf. In the second step, time- and value-adjusted probabilities are calculated, and the tree is rolled back. Since we are using multiple criteria, the rolling back process is different from the one used in standard decision tree analysis.

The aggregation of criteria values across event nodes is based on the adjustment of probabilities. Since financial value and time may follow different aggregation rules, time-adjusted probabilities may differ from value-adjusted probabilities for the same branch, and they are usually different from the initial probabilities of that branch.

In an event node, the financial value is calculated as the weighted sum of the values corresponding to its branches, using the value-adjusted probabilities as weights. The adjustment of probabilities follows a discrete-time option valuation model: probabilities are adjusted to risk-neutral value-adjusted probabilities according to the pricing process of the securities on which the project value is contingent. In this paper, we use the binomial model, but other models can be used with this approach. One alternative model is the trinomial model (used, for example, by Smith and McCardle, 1998).

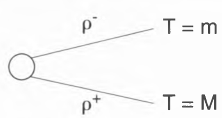
A new problem arises with the use of an option valuation model: some event nodes may not correspond to events that influence the value of the securities on which the option value is contingent – that is, private risks may exist. For example, uncertainty about whether a factory will take 18 or 20 months to be built does not influence the value of a publicly traded company with a similar factory. Such events are unique to the project, and should thus be treated as unsystematic risk. Since this risk is not relevant to the investors, the probabilities associated with such nodes are not adjusted for the calculation of financial value – only the event nodes that correspond to the passage of time require the adjustment of probabilities.

Time may be aggregated across event nodes in several different ways, depending on the situation that the decision-maker is facing. Average time to completion (or average delay relative to a specified completion time) and maximum time may, under some circumstances, be of interest. However, we believe that an uncertain time is, in most situations, considered equivalent to a fixed time between the average and the maximum of the uncertain times. So, a very general approach to aggregate time across event nodes is proposed, for which average time and maximum time are particular cases. This approach is similar to the risk-neutral valuation technique, and it relies on the use of certainty equivalents.

To start with, the decision-maker is asked for a certainty equivalent of a specified uncertain time. Then we calculate the implicit probabilities that equate the expected value of the uncertain time to that certainty equivalent. Those probabilities – time-adjusted probabilities – are used to perform the aggregation of time whenever the same pair of initial probabilities comes up (if the times are in the same order).

Assume that, in an event node, m is the shortest time, M is the longest time, and ρ^- and ρ^+ are their corresponding probabilities, as shown in Figure 1. Also assume that, faced with the uncertain time represented by that node, the decision-maker provides a certainty equivalent time CE. Then the certainty equivalent CE corresponds to an adjustment of probabilities to $\rho_T^- = (M\text{-}CE)/(M\text{-}m)$ and $\rho_T^+ = (CE\text{-}m)/(M\text{-}m)$. These time-adjusted probabilities (ρ_T^-, ρ_T^+) can be used every time the initial probabilities (ρ^-, ρ^+) come up, with the longest time corresponding to ρ^+. Notice that the choice of the longer time as the certainty equivalent leads to the maximum time to completion, and the choice of the average time as the certainty equivalent leads to the average time to completion.

Figure 1 - A generic binomial node with uncertain time

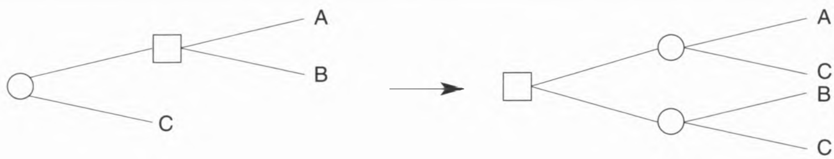


We will now consider the evaluation of decision nodes. We want the tree evaluation process to provide all the non-dominated strategies, and we use three different rules to accomplish this.

The first rule is that two consecutive decision nodes are merged. This means that, when there are consecutive decision nodes, it is the choice among all the alternatives represented by those nodes that is considered, and not consecutive choices among some of those alternatives.

The second rule is that if there is an event node before a decision node, then the decision is postponed by considering all possible combinations of decisions. Figure 2 shows the use of this rule. Note that if the lower branch were to have a decision node with two alternatives, the total number of alternatives in the resulting tree would be four. This rule may cause a large growth in the number of alternatives. So, it is important to use the next rule (the third) in order to prevent the number of alternatives from becoming too large.

Figure 2 - Example of the use of rule 2. A, B and C are different alternatives



The third rule is to eliminate all dominated alternatives in a decision node. This means that one alternative in a decision node can be eliminated if there is another alternative with a larger or equal financial value and a shorter or equal time, given that one of the inequalities is strict.

There is no need to explicitly modify the tree in order to apply these rules: they can be applied by considering that sets of alternatives are associated with the tree nodes. With these rules, the decisions involving non-dominated alternatives can be postponed until all event nodes are evaluated. Then, a multicriteria method may be used to choose from the non-dominated strategies.

This section has described the general approach for the use of time and financial value in project decision trees that was proposed by Godinho and Costa (2002). It is an approach that identifies all the non-dominated strategies, allowing the decision maker to use a multicriteria method to choose from them. In the next section we will define a particular model based in this general approach.

3. A model for the use of time and cost in project analysis

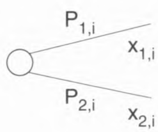
This section presents a model based on the above approach. This model assumes that time and cost are the relevant criteria, and that the decision-maker wants to minimise both criteria.

Consider a project that consists of undertaking a homogeneous task, that is, a task that can be split into several subtasks, all of them with identical characteristics. Define x^0 as the amount of work required to complete the task. It is hereafter assumed, without loss of generality, that $x^0 = 1$, and x is used to represent a fraction of the project ($0 \leq x \leq 1$).

Assume that n different processes, P_i, i = 1,..., n, may be used to undertake the project. Each process P_i is characterised by:

- a given time, $t_i \geq 0$, representing the duration of each utilisation of the process;
- a given cost per utilisation of the process, $c_i \geq 0$;
- a set of switching costs, representing the cost of switching to each of the other processes, $c_{i,j} \geq 0$, j = 1,..., n, j ≠ i ($c_{i,i} = 0$);
- a set of set-up times, representing the set-up time necessary to switch to each of the other processes, $t_{i,j} \geq 0$, j = 1,..., n, j ≠ i ($t_{i,i} = 0$);
- an initial set-up time, $t_{0,i} \geq 0$, and an initial cost, $c_{0,i} \geq 0$, representing the time and cost associated with the use of the process in the beginning of the project;
- an additive binomial process to describe project advance under each utilisation of process P_i, which can be represented by the binomial event node shown in figure 3.

Figure 3 – Binomial event node representing project advance under each utilisation of process P_i



$x_{1,i}$ and $x_{2,i}$ ($x_{1,i} > 0$, $x_{2,i} > 0$) are the possible project advance amounts under process P_i, and $P_{1,i}$ and $P_{2,i}$ are the corresponding probabilities, with $P_{1,i}+P_{2,i}=1$, $P_{1,i}>0$ and $P_{2,i}>0$. Note that this representation assumes the independence of project advances: the probability distribution of project advance under a given process is constant, and therefore independent of previous project advances. It is assumed, without loss of generality, that $x_{1,i} \leq x_{2,i}$. Since it is assumed that $x^0 = 1$, then $x_{1,i}$ and $x_{2,i}$ can be interpreted as fractions of the project. The arc corresponding to an advance $x_{1,i}$ is hereafter identified as arc 1 of the node, and the arc corresponding to an advance $x_{2,i}$ is identified as arc 2. The branches starting immediately after these arcs are identified as branches 1 and 2 of the node, respectively.

This model assumes that the use of the processes is indivisible. This means that the process being used can only be changed at the end of a complete utilisation. If, at the end of the task, the fraction of the project that has not been undertaken is smaller than the project advance in the next use of the process, the time and cost for a complete utilisation of the process will be required anyway.

Following the previously outlined approach, the aggregation of criteria values across event nodes is based on the adjustment of probabilities. For the cost, the binomial model for option valuation is used. Since, in the model being presented, the cost per process utilisation is constant for each process, the cost risk is only related to the project duration risk and to the process choices, and it is thus independent of the price of any traded asset. Therefore, the cost risk is unsystematic, meaning that no risk premium shall be required, and the value-adjusted probabilities are equal to the initial probabilities.

For the calculation of time-adjusted probabilities, the decision-maker is asked to provide two certainty equivalents for each different pair of initial probabilities occurring in the definition of the processes. One certainty equivalent corresponds to an arbitrary time T' > 0 in branch 1 and a time 0 in branch 2, and the other corresponds to a time 0 in branch 1 and a time T' > 0 in branch 2 (see Figure 4). From these certainty equivalents, two pairs of time-adjusted probabilities are calculated, $(P^{T,1}_{1,i}, P^{T,1}_{2,i})$ and $(P^{T,2}_{1,i}, P^{T,2}_{2,i})$, respectively. The first pair is used when time is longer in branch 1 than in branch 2, and the second pair is used in the opposite situation. Notice that the two different pairs of adjusted probabilities are necessary because the risk preferences of the decision-maker will usually lead her/him to adjust the same pair of probabilities differently, according to the branch that contains the longer time. As an example, if the decision-maker is averse to the time risk, she/he will adjust $P_{1,i}$ to a larger value if branch 1 contains the longer time, and she/he will adjust $P_{1,i}$ to a smaller value if branch 2 contains the longer time.

Figure 4 – Certainty equivalents elicited for the calculation of the time-adjusted probabilities

Some underlying assumptions of this model, particularly the assumptions of binomial event nodes and indivisible process utilisation may, at a first sight, seem too restrictive for the model to be useful. We do not think it is so. In a real-life problem it may sometimes be possible to define exactly during how much time a process will be used, and there may be continuous statistical distributions for the project advance. Still, models like this one may be used to approximate those problems, by defining processes such that the duration of each utilisation is small. A similar approximation is done in some continuous-time real option problems, particularly when multiple options are involved (see, for example, Trigeorgis, 1993).

An important characteristic of the model is that the decision trees are often very large. These trees may thus require very large amounts of memory space, and the corresponding calculations may take a very long time. Therefore, the approach based on the construction and evaluation of the tree is usually impractical, or even impossible to use. This problem will be considered in Section 4, where an algorithm for an efficient identification of the non-dominated strategies will be presented.

This model is easy to use, and can be applied to the analysis and planning of projects that consist of undertaking homogenous tasks, when time and cost are the relevant criteria. Possible applications of the model include the analysis of some construction projects and production planning. It is also easy to modify the model to make it more general. For example, by making small changes we would be able to drop the assumption of independent project advances, or to use functions of time and cost (for example, utilities) instead of the absolute values of the criteria. However, such modifications may make the development of efficient algorithms for the identification of non-dominated strategies more difficult, or even impossible.

4. An algorithm for the generation of the non-dominated strategies

The model described in the previous section usually leads to very large decision trees, whose construction and evaluation may take a long time and require large amounts of memory space. As an attempt to overcome this problem, an algorithm for an efficient identification of the non-dominated strategies was developed and is now described. The algorithm uses some mathematical properties of the model, which are presented in the appendix.

4.1. Main ideas of the algorithm

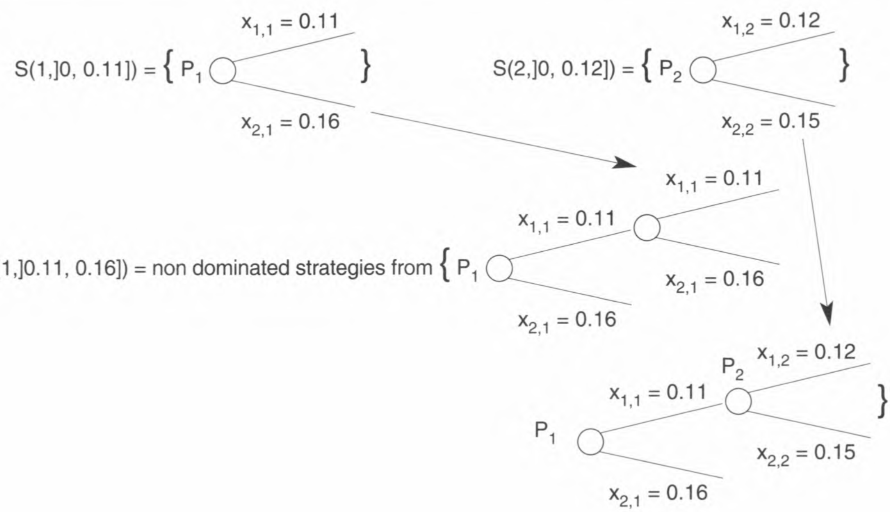
Each branch of the tree corresponds to a fraction of the project: the fraction of the project that was not undertaken in the previous arcs of the tree. Different branches of the tree, occurring in different places in the tree, may correspond to the same fraction of the project, thus leading to identical sets of non-dominated strategies. Moreover, all the branches of the tree that correspond to fractions x of the project belonging to certain intervals I lead to the same set of non-dominated strategies (see property 2 in the Appendix).

Let S(i,I) be the common set of non-dominated strategies for undertaking fractions $x \in I$ of the project starting with the use of process P_i. The algorithm calculates the sets of non-dominated strategies S(i,I) for intervals I of consecutively larger values of x, until the sets of non-dominated strategies for $x = x^0 = 1$ and for all processes are calculated.

As we explained before, the sets of non-dominated strategies that allow the completion of a given fraction x of the project by starting with a given process are the same for a given interval of values of x. To exemplify this, consider the existence of two processes, P_1 and P_2, with $x_{1,1} = 0.11$ and $x_{2,1} = 0.16$, for P_1, and $x_{1,2} = 0.12$ and $x_{2,2} = 0.15$, for P_2. If a fraction of the project that is larger or equal than 0.89 is already executed, and process P_1 is used after that, then, after using P_1 once, the project will be completed. Since this is true if the fraction of the project not previously undertaken is positive and smaller than or equal to 0.11, then S(1,]0, 0.11]) is the set that only contains the strategy of using P_1 once. So, when the fraction of the project that was not previously undertaken is $x \in]0, 0.11]$ and P_1 is used next, the only non-dominated strategy is the strategy belonging to S(1,]0, 0.11]). Similarly, for P_2, S(2,]0, 0.12]) is the set that only contains the strategy of using P_2 once. For larger values of x, the sets of non-dominated strategies are also identical for all the values of x belonging to some intervals, allowing us to avoid repeating the calculation of branches that lead to the same non-dominated strategies.

The algorithm begins with the identification of the non-dominated strategies for small project fractions. In our previous example, it would start with the calculation of S(1,]0, 0.11]) and S(2,]0, 0.12]). Then, the sets of non-dominated strategies that were previously calculated are taken into account in the identification of the non-dominated strategies for consecutively larger project fractions. In the example, the next set S(1, I) would represent the use of process P_1 first and then the use of a non-dominated strategy in branch 1. Since the non-dominated strategy to be used in branch 1 could be from either S(1,]0, 0.11]) or S(2,]0, 0.12]), two different strategies would be generated. Notice that a strategy generated in such a manner is valid for undertaking a project fraction smaller than or equal to 0.16, because the use of another strategy in branch 2 is not being considered. So, the next set to be considered would be S(1,]0.11, 0.16]), which is the set of those two strategies, if none of them is dominated, or a set containing one of those strategies, if the other one is dominated. Figure 5 represents these sets. The algorithm then proceeds until the complete project ($x = x^0 = 1$) is considered.

Figure 5 – Initial sets calculated by the algorithm for the example presented in this subsection



So, the identification of non-dominated strategies for larger fractions of the project uses the non-dominated strategies generated for smaller fractions, allowing the algorithm to discard dominated strategies as soon as they are identified. This way, it is also possible to avoid the repetition of calculations for branches with similar characteristics occurring in different places in the tree; otherwise such repetition of calculations would happen several times. Property 1 (presented in the Appendix) is also used by the algorithm, to avoid the calculation of some sets of non-dominated strategies, and to allow a more efficient calculation of other sets.

In short, we can say that the algorithm starts with tree branches corresponding to small fractions of the project (initially the fractions undertaken in the tree leaves), and then considers consecutively larger fractions until all the non-dominated strategies for the complete tree are identified. For a complete description of the algorithm, see Godinho (2003)

4.2. The performance of the algorithm

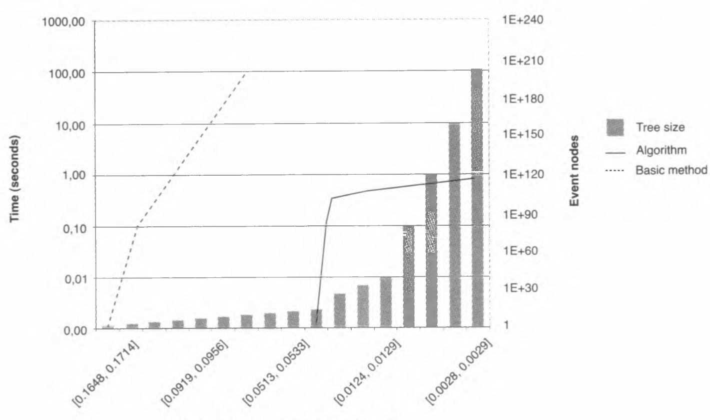
Some computational tests were conducted in order to assess the performance of the algorithm. Those tests aimed to examine the performance of the algorithm in several different situations, and to compare the performance of the algorithm with the performance of an alternative method. This alternative method, the "basic method", consists of building and evaluating the tree according to the approach defined in section 2. We now present some computational results – other results, and other details about the tests, can be found in Godinho (2003).

These tests used sequences of files. The first element of each sequence was generated by defining some parameters as constants, and the remaining parameters as samples of given uniform distributions. The other elements of each sequence were defined through sequential changes in given parameters. Twenty sequences were generated for each set of parameter distributions, and average running times were then calculated for each set. Only non-dominated strategies with different times and costs were generated (alternative strategies with identical cost and time were not generated).

The sets presented in this paper consider two processes, the first one with $t_1 = 1$ and $c_1 = 2$, and the second one with $t_2 = 2$ and $c_2 = 1$. These sets were designed to analyse the performance of the algorithm and the basic method when the size of the tree increases. In the first set, the number of non-dominated strategies was kept very small (equal to 2) and, in the second set, the number of non-dominated strategies was allowed to increase quickly with the size of the tree. In order to achieve this, we started all sequences of each set with large advance amounts ($x_{b,i}$ initially belonged to [0.25, 0.26], i = 1, 2, b = 1, 2), and decreased these amounts along each sequence. In order to keep the number of non-dominated strategies small, we defined large switching costs and set-up times in the first set ($c_{i,j} = t_{i,j} = 10000$). In the second set, we used null switching costs and set-up times, thus allowing the number of non-dominated strategies to increase quickly. The following distributions were used for the probabilities: the basic probabilities for both branches belong to [0.49, 0.51], the probability for the branch with the longest time belongs to [0.55, 0.6], and the probability for the branch with the shortest time belongs to [0.4, 0.45].

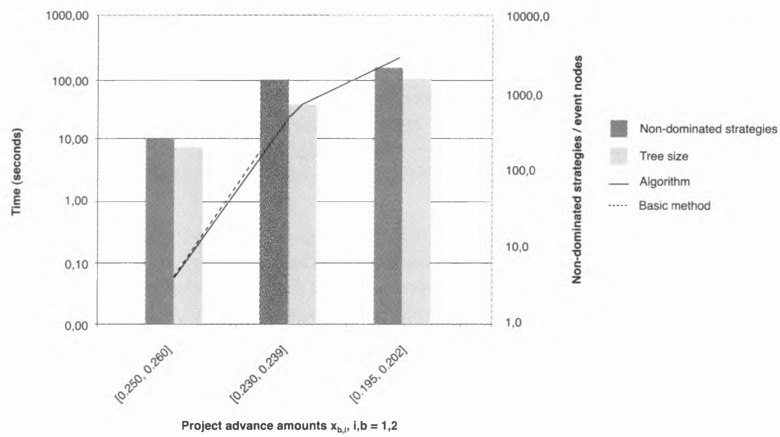
For the first set of sequences (see Chart 1), the algorithm performed much better than the basic method. In fact, it was only possible to run the basic method until the tree size reached some hundred million nodes ($x_{b,i} \in [0.0716, 0.0744]$), because disk and memory space were insufficient for larger trees. However, it was still possible to run the algorithm, with very small execution times (never much more than 1 second), for all the parameter values that were used (the tree size is estimated to be about 10^{31} event nodes for the smallest advance amounts that were considered). This shows that, when the number of non-dominated strategies is small (because of large set-up times and switching costs), the algorithm is clearly faster and much more efficient in memory usage terms than the basic method.

Chart 1 – Calculation time (scale on the left) and tree size (scale on the right) for different project advance amounts, for the tests on the first set of sequences. Logarithmic scale is used in the Y axis



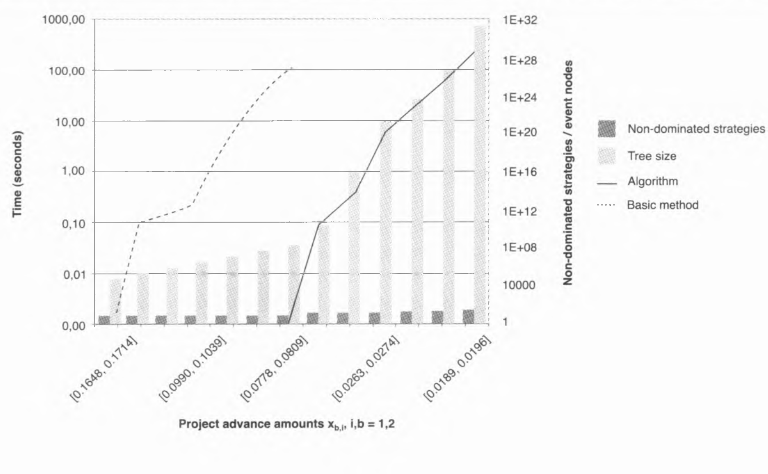
For the second set of sequences (see Chart 2), the results were quite different. In fact, the number of non-dominated strategies was very large, and the calculation times were similar for both the algorithm and the basic method. Notice that the large number of non-dominated strategies soon made their identification impracticable, both for the algorithm and for the basic method.

Chart 2 – Calculation time (scale on the left), number of non-dominated strategies and tree size (scale on the right) for different project advance amounts, for the tests on the second set of sequences. Logarithmic scale is used in the Y axis



It is also interesting to examine the results we achieved when we used extreme time-adjusted probabilities: a 100% probability for the branch with the longest time and a 0% probability for the branch with the shortest time (notice that using these probabilities is the same as using maximum time to completion to aggregate time). When these time-adjusted probabilities were used, there was no noticeable change in the results for the first set of sequences (because the number of non-dominated strategies was already very small for this set), but the results for the second set were significantly changed. In the second set of sequences (Chart 3), the number of non-dominated strategies was lower, and the algorithm performance became much better than the basic method performance.

Chart 3 – Calculation time (scale on the left), number of non-dominated strategies and tree size (scale on the right) for different project advance amounts, for the tests on the second set of sequences when extreme time adjusted probabilities were used. Logarithmic scale is used in the Y axis



The results of the tests led us to some important conclusions:

– as would be expected, the performance of both methods worsens as the number of non-dominated strategies increases and as the tree size increases;

– the performance of the algorithm is less sensitive to the increase in tree size, and more sensitive to the increase in the number of non-dominated strategies;

– the performance of the algorithm is not clearly better than the performance of the basic method when the number of non-dominated strategies is large;

– the algorithm performs much better than the basic method, both in terms of running time and in terms of memory usage, when the number of non-dominated strategies is small;

– when the maximum time to completion is used (and, in fact, whenever extreme probabilities of 0% or 100% are used), the number of non-dominated strategies is small, and the algorithm performs very well.

So, although the algorithm does not appear to be able to efficiently generate a large number of non-dominated strategies, its performance is very good when the number of non-dominated strategies is small. The use of the algorithm to identify some representative strategies, instead of all the non-dominated strategies, may thus be an efficient way to employ the model.

4.3. The use of maximum error parameters

For an effective use of the model, it is important to define a way to improve the performance of the algorithm when the number of non-dominated strategies is large. We developed a new approach, based on the use of maximum error parameters, which allows the algorithm to identify some representative strategies instead of generating all the non-dominated strategies. This way, we are able to perform a faster identification of the strategies, allowing us to use the model in some situations that we would otherwise be unable to analyse.

To understand the meaning of these parameters, let us define ϵ_t as the time error parameter and ϵ_c as the cost error parameter. These parameters mean that, for a given non-dominated strategy with time t and cost c that ceases to be generated, it can be guaranteed that there is at least one strategy that is generated with a time $t' \leq t + \epsilon_t$ and a cost $c' \leq c + \epsilon_c$. This way, it is possible to generate only one "representative" strategy for each neighbourhood of size (ϵ_t, ϵ_c). Thus, a smaller number of strategies is generated by the algorithm, and its performance is improved. The improvement in the algorithm performance depends on the values of ϵ_t and ϵ_c: the larger these values are, the better the algorithm performs, and the furthest may the generated strategies be from the true non-dominated strategies.

The practical implementation of these parameters in the algorithm required some adaptations. In fact, instead of the global time and cost error parameters considered above, we defined time and cost error parameters that are used by the algorithm every time that an elimination of dominated strategies occurs. Instead of only eliminating dominated strategies, the algorithm also eliminates some non-dominated ones, if another one exists in the neighbourhood defined by the error parameters. From these parameters, and from the maximum number of tree levels, it is possible to calculate upper bounds for the maximum errors considered before (ϵ_t and ϵ_c).

The practical use of these maximum error parameters proved to be very useful, allowing us to run the model for some parameter sets that we would otherwise be unable to handle. The application we present in the next section is one such example.

5. An application of the model

This section presents an application of the model, in order to show that it may be useful in real-life situations.

Consider the case of a textile company that received an order for 10 000 shirts from a soccer team. The company managers want to produce the shirts as quickly as possible, but they also want to have as low a production cost as possible. The company has three different sets of equipment, let us call them A, B and C, that can be used. The number and skills of the available workers, as well as other company commitments, do not allow the simultaneous use of more than one of these sets in the production of shirts. The decision about which equipment to use may be made on a daily basis.

Equipment A produces an average μ = 3000 shirts in a 5-day week, with a standard deviation of σ = 250 shirts. Its use costs €500 per day, and its initial set-up time and cost are 2 days and €1000 euros, respectively. Equipment B produces an average μ = 4800 shirts in a 5-day week, with a standard deviation of σ =500 shirts. Its use costs €1000 per day, and its initial set-up time and cost are 2 days and €2000, respectively. Equipment C produces an average μ = 2400 shirts in a 5-day week, with a standard deviation of σ = 250 shirts. Its use costs €400 per day, and its initial set-up time and cost are 1 day and €400, respectively. These characteristics of the equipment are summarized in Table 1.

Table 1 – Characteristics of the equipment

	Equipment		
	A	B	C
Average production	3000 shirts/week	4800 shirts/week	2400 shirts/week
Standard deviation of the production	250 shirts	500 shirts	250 shirts
Cost	€500/day	€1000/day	€400/day
Initial set-up time	2 days	2 days	1 days
Initial set-up cost	€1000	€2000	€400

The set-up times and switching costs for changing the equipment being used are shown in Table 2. If it can be assumed both that the number of shirts produced in a given day is independent of the number of shirts produced in the previous day and that the binomial project advance distribution provides a good enough approximation for the production distribution, then the model presented in Section 3 may be used.

Table 2 – Set-up times and switching costs for switching from one set of equipment to another

From/To		Equipment		
		A	B	C
Equipment	A		2 days / €1500	1 day / €500
	B	2 days / €1500		1 day / €1000
	C	1 day / €800	1 day / €1700	

We define three different processes, P_1, P_2 and P_3. P_1 consists of using equipment A for one day, P_2 consists of using equipment B for one day and P_3 consists of using equipment C for one day. For each of these processes, t_i is 1 day, c_i is the cost of using the equipment for one day and $t_{0,i}$ and $c_{0,i}$ are the initial cost and initial set-up time for the equipment, so $t_1 = t_2 = t_3 = 1$ day, c_1 = €500 euros, c_2 = €1000, c_3 = €400, $t_{0,1} = t_{0,2} = 2$ days, $t_{0,3} = 1$ day, $c_{0,1}$ = €1000, $c_{0,2}$ = €2000 and $c_{0,3}$ = €400. The set-up times and switching costs are $t_{1,2} = t_{2,1} = 2$ days, $t_{1,3} = t_{2,3} = t_{3,1} = t_{3,2}$ = 1 day, $c_{1,2} = c_{2,1}$ = €1500, $c_{1,3}$ = €500, $c_{2,3}$ = €1000, $c_{3,1}$ = €800 and $c_{3,2}$ = €1700.

We now define additive binomial processes whose project advance amounts are consistent with the average and standard deviation of the shirt production. For each process, the average and standard deviation of shirt production are fitted to the average and standard deviation of a binomial distribution by using the following equations:

$$\frac{\mu}{5} = P_{1,i}x_{1,i} + P_{2,i}x_{2,i}$$

$$\frac{\sigma}{\sqrt{5}} = (x_{2,i} - x_{1,i})\sqrt{P_{1,i}P_{2,i}}$$

Since $P_{2,i} = 1-P_{1,i}$, there are two equations and three unknowns. So it is possible to let one of the unknowns assume a given value, to solve the system and then to assess the validity of the values of the parameters (to assess whether the probabilities belong to]0,1[and whether the project advance amounts are positive).

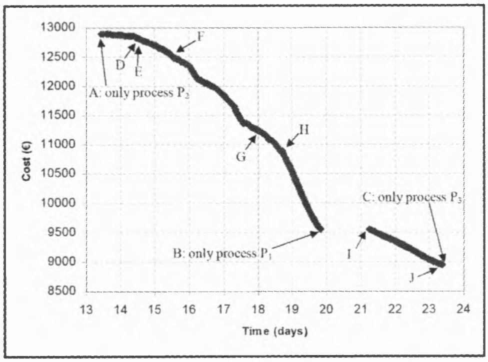
We define initial probabilities of 50% for the three processes. Solving the equation systems for all processes, we get $x_{1,1} = 488.2$ shirts and $x_{2,1} = 711.8$ shirts for process P_1, $x_{1,2} = 736.4$ shirts and $x_{2,2} = 1183.6$ shirts for process P_2 and $x_{1,3} = 368.2$ shirts and $x_{2,3} = 591.8$ shirts for process P_3. In order to keep $x^0 = 1$, the project advance amounts are normalised by dividing them by the total number of shirts to be produced (10000).

Since, for both processes, 50%/50% probabilities are used, the managers are only asked to provide one certainty equivalent. Assuming that a 50% probability of 0 days and a 50% probability of 10 days is considered equivalent to a certain time of 6 days, the time-adjusted probability for the longest time is 60%, and the time-adjusted probability for the shortest time is 40% This means that $P^{T,1}_{1,1} = P^{T,2}_{2,1} = P^{T,1}_{1,2} = P^{T,2}_{2,2} = 0.6$ and $P^{T,1}_{2,1} = P^{T,2}_{1,1} = P^{T,1}_{2,2} = P^{T,2}_{1,2} = 0.4$.

Now the model parameters are completely defined, and it seems that the model can be used to assess the efficient time/cost combinations. However, the trees that correspond to this application are very large, and the number of non-dominated strategies is also very large. This made it impossible to use either the algorithm or the basic method to generate all the non-dominated strategies. Instead, we ran the algorithm using the maximum error parameters, as described in subsection 4.3. The parameter values that we used considered a maximum time error of 0.011 days and a maximum cost error of €1 in each tree level. Since there are at most 28 levels in one such strategy tree, this leads to maximum global errors of 0.31 days and €28. As a percentage of the absolute criteria values, maximum time error is about 1.7% and maximum cost error is about 0.3%.

The model was run, allowing us to obtain 645 time/cost combinations, which are represented in Figure 6.

Figure 6 – Time/cost combinations for the representative strategies generated by the algorithm for the shirt production example



Two separated lines are clearly defined in the plot shown in Figure 6. The shortest line is on the bottom right of the plot and includes the strategies with the smallest costs. On the bottom right of this line is the strategy that consists of using only process P_3 (strategy C), which is the strategy that allows us to complete the project with the smallest cost. As we go left along this line, we get strategies with increasing costs and decreasing times. All the strategies in this line start with the use of process P_3 and, apart from strategy C, they all change to processes P_1 and P_2 in different situations. As we move left along the line, the changes from process P_3 to another process occur earlier in the tree. In strategy J, which is near strategy C, the first possible change occurs when about 33% of the project is complete, while in strategy I, at the other end of the line, the first possible change occurs when about 7% of the project is completed.

It is clear that the strategies near the left side of this shortest line (near I) do not seem attractive, since for a small additional cost it is possible to have a significant decrease in time by choosing strategy B.

The longest line is in the top left side of the plot and includes the strategies with the shortest times. This line does not look straight and, in fact, we may say that its form seems quite "odd". Notice that its form is clearly due to the non-linear effect of switching costs and set-up times in the model.

The strategy that consists of using only process P_2 (strategy A, on the top left of the line) is found at one extreme end of this line, and this is the strategy that allows us to complete the project in the shortest time. At the other extreme end of this line, we find the strategy that consists of using only process P_1 (strategy B, on the bottom right of the line). Near strategy A, we have strategies that start by using process P_2, then switch to P_3 in some situations, and sometimes switch again from P_3 to P_2 or P_1. For example, strategy D consists of starting with process P_2 and switching to P_3 in a particular situation. Near the other end of this line, the time/cost combinations correspond to strategies that start by using process P_1 and then change to process P_2. For example, strategy H starts by using P_1 and then changes to P_2 in some situations in which the project is advancing slowly.

Near the middle of this longest line, we have strategies that start by using P_3 and then switch to P_1 or P_2. It may seem odd that some strategies that start by using P_3 appear in the middle of this line, since the strategy that consists of only using P_3, and other strategies that start with P_3, are in the shortest line, and have quite different times and costs from the strategies in the middle of this longest line. There is a reason for this, though. If we want to use P_1 or P_2, depending on whether the project is advancing rapidly or slowly, it may pay to start by using P_3 and switch to one of the other processes after it is possible to know whether the project is advancing quickly or not. This is because P_3 has small switching costs and set-up times. Also, we may want to start by using P_3 and then switch to P_2 in order to speed up the project: combining the two processes gives us the intermediate times and costs. Both these situations occur in the middle of this longest line: P_3 is used at the beginning and then a change to another process occurs (eventually, when it is possible to know how quickly the project is advancing). On the other hand, in the shortest, bottom right line, we have strategies with low costs and, to achieve that, we use P_3 during most of the project. The reason for using P_3 is thus different for the two areas of the plot.

Some examples of strategies in the middle of the longest line that start by using P_3 are E, F and G. Strategy E starts with P_3 and then switches to P_2. Strategies F and G both start by using P_3 and then switch to P_2 (particularly when the project is advancing slowly) or to P_1 (particularly when the project is advancing quickly); in strategy F, some additional changes from P_1 to P_2 occur when the project advances slowly after switching to P_1.

After the strategies are generated, the decision-maker can use a multicriteria method to choose the preferred time/cost combination, and embark on the corresponding strategy.

Let us assume, as an example, that strategy G is chosen. This strategy has a cost of €11256 and a time of 17.95 days. The first process to be used in the strategy is process P_3. If the project advances quickly on the first two days, or slowly on the first day and quickly on the second, then a change to process P_1 occurs on the third day. If the project advances slowly for the first two days, then a change to process P_2 occurs on the third day. If the project advances quickly on the first day and slowly on the second, then a change to P_2 also occurs, but later. This change occurs on the fourth day if the project advances slowly on the third day, it occurs on the fifth day if the project advances quickly on the third day and slowly on the fourth, and it occurs on the sixth day if the project advances quickly on both the third and the fourth day.

6. Conclusions

In this paper we have presented a useful bicriteria model for project analysis that is based on decision trees, and described an application of the model to a production planning problem.

Following a general approach that uses bicriteria decision trees to represent investment projects, a model for the use of time and cost in project analysis was developed. This model assumes that some different processes may be used to undertake a homogeneous task. After the processes are defined, the complete decision tree can be automatically built, and the non-dominated strategies can be identified.

This model usually leads to very large decision trees, whose construction and evaluation may take a long time and require very large amounts of memory space. Therefore, the approach based on the construction and evaluation of the tree is usually impractical, or even impossible. In an attempt to overcome this problem, we have proposed an algorithm for the identification of the non-dominated strategies. The algorithm does not require the construction of the complete decision tree, since it is able to disregard non-interesting branches as soon as they come up, and to avoid the repetition of calculations for similar branches of the tree. The algorithm also provides a more compact representation of the tree, since identical branches located in different points of the tree are represented only once.

The computational tests have shown that the algorithm performs well when the number of non-dominated strategies is small, but it does not perform so well when that number is large. In order to overcome the problem of model usage with a large number of non-dominated strategies, we introduced an error parameter that allows us to avoid generating all the non-dominated strategies that are very close to each other. Using this approximation, we are able to reduce the number of strategies that are generated. This way the performance of the algorithm is improved, and we are able to apply the model to some cases that we would otherwise be unable to handle.

Finally, we applied the model to a production planning problem. We defined the problem, calculated the model parameters, and generated the strategies. We then plotted and analysed the different time/cost combinations that we obtained, and examined which types of strategies led to different combinations of criteria values. This application allowed us to conclude that the model can be easily applied to some problems, providing useful insights about the types of strategy that should be used in order to achieve different combinations of criteria values.

References

Brealey, R.; Myers, S. (2000), *Principles of Corporate Finance*, Boston, McGraw-Hill.

Godinho, P. C. (2003), *Árvores de Decisão Bicritério em Análise de Projectos (PhD Thesis)*, Coimbra, Faculty of Economics of the University of Coimbra.

Godinho, P. C.; Costa, J. P. (2002), «A Note on the Use of Bicriteria Decision Trees in Capital Budgeting», *Global Business & Economics Review*, 4, 1, 147-158.

Hertz, D. B.; Thomas, H. (1983), *Risk Analysis and its Applications*, Chichester, John Wiley and Sons.

Magee, S. (1964), «How to Use Decision Trees in Capital Investment», *Harvard Business Review*, 42, 79-96.

Smith, J. E.; Mccardle, K. F. (1998), «Valuing Oil Properties: Integrating Option Pricing and Decision Analysis Approaches», *Operations Research*, 46, 2, 198-217.

Smith, J. E.; Nau, R. F. (1995), «Valuing Risky Projects: Option Pricing Theory and Decision Analysis», *Management Science*, 41, 5, 795-816.

Trigeorgis, L. (1993), «The Nature of Option Interactions and the Valuation of Investments with Multiple Real Options», *Journal of Financial and Quantitative Analysis*, 28, 1, 1-20.

Appendix – some mathematical properties of the model

This appendix presents two mathematical properties of the model that were used in the development of the algorithm. For the proof of these properties, see Godinho (2003).

Property 1: Let S(i,x), i = 1,…, n, be the set of all strategies that begin with the use of process P_i, can be used to undertake a fraction x of the project, and are not dominated by any other strategy belonging to this same set. Then:

a) if x' > x, s ∈ S(i, x), and s can be used to undertake a fraction x' of the project, then s ∈ S(i, x');

b) if x' > x and ∀ s ∈ S(i, x), s can be used to undertake a fraction x' of the project, then S(i, x') = S(i, x).

Property 2: Let S(i, x), i = 1,…, n, be the set of all strategies that begin with the use of process P_i, can be used to undertake a fraction x of the project, and are not dominated by any other strategy belonging to this same set. Consider a process P_i and a fraction x of the project such that $x \in]x_{1,i}, 1[$. Let x' be a fraction of the project such that x' > x, and let E_j, j = 1,…, n, be the set of all values of x that are smaller than $x'-x_{1,i}$ and that cause a modification in S(j, x), that is:

$$E_j = \{e \in]0, x' - x_{1,i}[: \forall \delta > 0, S(j, e) \neq S(j, e + \delta)\}$$

Define the set A in the following way:

$$A = \begin{cases} \bigcup_{j=1}^{n} \{\alpha > 0: \alpha = e + x_{b,i}, e \in E_j, b \in \{1, 2\}\}, & \text{if } x > x_{2,i} \\ \{x_{2,i}\} \cup \bigcup_{j=1}^{n} \{\alpha > 0: \alpha = e + x_{1,i}, e \in E_j\}, & \text{if } x \leq x_{2,i} \end{cases}$$

If x ∈ A and there are no elements of A with values between x and x', then S(i, x) = S(i, x'). That is,

$$A \cap [x, x'[= \emptyset \Rightarrow S(i, x) = S(i, x').$$

The Pricing of Systematic Liquidity Risk in Stock Markets

José Luis Miralles Marcelo / Maria del Mar Miralles Quirós / José Luis Miralles Quirós
Department of Finance and Accounting; University of Extremadura, Spain

resumo

A questão de a rendabilidade estar ou não afectada pela liquidez não está ainda resolvida. A ausência de resultados concludentes na investigação empírica sugere que a relação entre a avaliação de activos e a liquidez não tem sido estudada adequadamente na literatura habitual. Considerarmos que os shocks sistemáticos da liquidez poderiam afectar o óptimo comportamento dos agentes nos mercados financeiros. De facto, as flutuações nas diferentes medições da liquidez são significativamente correlacionadas nos activos mais comuns. Em consequência, propormos a construção de um factor de liquidez baseado no rácio de Amihud (2002) e no procedimento de aproximação ortogonal de Fama e French (1993) para o incluir como mais uma variável adicional no seu modelo de três factores.

résumé / abstract

En le qui concerne la liquidité affectée ou non à la rentabilité des actifs n'est pas encore résolue. L'absence de résultats concluants dans la recherche empirique préalable suggère que la relation évaluation actifs et liquidité adéquatement n'a pas été étudiée dans la littérature standard. Nous considérons que les chocs systématiques de liquidité peuvent affecter le comportement optimal des agents sur les marchés financiers. De fait, des fluctuations dans diverses mesures de liquidité sont significativement reliées entre des actifs. Par conséquent, nous proposons la construction d'un facteur de liquidité basé le rapport d'Amihud (2002) et sur la procédure de rapprochement orthogonal Renommée et de French (1993), pour qu'il puisse être inclus comme une variable additionnelle dans son modele de trois facteurs.

The question whether liquidity affects asset returns or not remains unresolved thus far. The absence of conclusive results in previous research suggests that asset pricing and liquidity have not been properly addressed in the standard literature. We consider that systematic liquidity shocks affect the optimal behavior of agents in financial markets. Indeed, fluctuations in various measures of liquidity are significantly correlated across common stocks. Accordingly, we propose the construction of a liquidity risk factor based on the ratio of absolute stock returns on euro volume suggested by Amihud (2002) and the approximately orthogonalizing procedure of Fama and French (1993), using it as an augmenting variable in their three-factor model.

José Luis Miralles Marcelo; María del Mar
Miralles Quirós; José Luis Miralles Quirós

1. Introduction

It is generally accepted that asset liquidity influences investors' portfolio decisions because of its close relation to transaction costs. It is reasonable to think that investors who buy illiquid assets require higher expected returns on their investments because a lack of liquidity can be interpreted as an additional risk.

However, the question of whether or not liquidity affects asset returns remains unresolved. While various theoretical models have indicated that liquidity risk is an important factor in explaining returns, empirical studies have failed to find significance for this liquidity risk factor. The reason could be that liquidity-risk measures are weak or proxy total instead of systematic liquidity risk.

This paper contributes to the determination of the role of a liquidity factor in the structure of returns. Our main objective is to re-assess the role of liquidity risk, in the context of a Fama and French (1993) framework, by using the measure of systematic liquidity risk proposed by Amihud (2002).

We use the Fama and French (1993) portfolio sorting procedure to estimate the liquidity factor and then test the importance of this factor in asset returns, by contrasting the performance of four models: the standard CAPM, the Fama and French three-factor model and those two models augmented by the liquidity factor. The data includes all stocks traded on the Spanish stock market from January 1994 to December 2002. Overall, results suggest that liquidity risk is important in explaining the structure of returns in the Spanish stock market.

We consider these results to be of particular importance to professional asset managers and risk managers. Systematic risk prediction is undoubtedly of great importance when making investment decisions, when evaluating financial assets or when structuring portfolios. In short, new evidence on the importance of liquidity risk in asset pricing can be valuable for researchers and have important implications and practical value for investors.

The rest of the paper is organized as follows. Section 2 contains the literature review and provides the theoretical motivation for this analysis. Section 3 describes Amihud (2002)'s measure of liquidity risk, the portfolio formation procedure and the proposed tests to investigate the role of the liquidity factor. Section 4 indicates the data employed and highlights the empirical results obtained from the Spanish stock market. Section 5 has concluding remarks.

2. The role of liquidity in asset pricing

The question of whether liquidity determines expected returns has been widely documented in the financial literature. Using a variety of liquidity measures, studies analyze whether less liquid stocks have higher average returns than expected.

Amihud and Mendelson (1986) were among the first to examine the role of liquidity in asset pricing. They analyzed the relationship between stock returns and bid-ask spreads and found empirical evidence related to the existence of a liquidity premium. However, Eleswarapu and Reinganum (1993), who extended the sample period by ten years, examined the effect of seasonality on bid-ask spreads and returns. They found that the relationship between bid-ask spreads and asset returns is mainly limited to the month of January.

Brennan and Subrahmanyam (1996) examined whether illiquidity costs caused by adverse selection result in higher expected returns. Instead of using bid-ask spreads as a proxy for liquidity they used estimated variable and fixed transaction costs.[1] They adjusted for risk using the Fama and French (1993) three-factor model and found a concave relationship between

1 They argue that bid-ask spreads are not an appropriate proxy for liquidity as bid-ask spreads are noisy. Additionally, liquidity costs caused by asymmetric information are captured in the variable component of trading costs.

premiums and variable costs and a convex relationship between premiums and fixed costs, in contrast to the concave relationship found by Amihud and Mendelson (1986).

Previous research for the Spanish stock market tends to show quite similar distress results related to the temporal and cross-sectional behavior of bid-ask spreads and other measures of liquidity.[2] For the period 1990-1994, Tapia (1997) analyzes the seasonality of liquidity premium considering the influence of fiscal reasons on trading. The main results of this work indicate the existence of a differential behavior for the liquidity premium, but not for asset with more probability of trading by fiscal reasons. Moreover, when he includes the size variable, his results are weaker. Rubio and Tapia (1998), employing Brennan and Subrahmanyam's (1996) methodology, provide evidence on the relationship between bid-ask spreads and stock returns, analyzing the effect of seasonality. The results show a positive liquidity premium in January, although not significantly different from zero. Nevertheless, the most complete study for the Spanish stock market about bid-ask spreads was published by Blanco (1999). This work is based on the influence of minimum variations in prices on bid-ask spreads. He argues that bid-ask spreads underestimate the temporal and cross-sectional movements in liquidity.

Given the lack of robustness of empirical results, several investigators have re-examined the relationship between liquidity and asset returns using alternative measures of liquidity that allow us to approach the concept of liquidity employed by investors in their financial decisions. In this sense, a large number of papers have focused on the use of liquidity measures based on trading activity, such as trading volume (Brennan *et al.*, 1998), turnover (Datar *et al.*, 1998) or illiquidity ratio (Amihud, 2002), that allow us to obtain a larger series of observations over a longer period of time and to check the robustness of the empirical results.

Brennan, Chordia, and Subrahmanyam (1998) found that the stock volume has a significant negative effect on the cross-section of stock returns and it subsumes the negative effect of size. Spanish market evidence is reported by Miralles and Miralles (2003). However, this liquidity measure has two potential problems. First, the number of shares traded is not by itself a sufficient statistic for the liquidity of a stock since it does not take into account the differences in the number of shares outstanding or the shareholder base. Second, the use of the euro volume has a size bias.

Another related measure is turnover, i.e. the ratio of trading volume to the number of shares outstanding, which we can employ as a measure of the asset trading frequency. Datar *et al.* (1998), Rouwenhorst (1999), and Chordia *et al.* (2001) found that cross-sectionally stock returns decrease in stock turnover, which is consistent with a negative relationship between liquidity and expected return.

More recently, Amihud (2002) has proposed a new measure based on trading activity, the aggregate ratio of absolute stock returns on euro volume. In particular, Amihud's illiquidity ratio has a strong theoretical appeal. In this work, Amihud (2002) focuses on the time-series aspects of aggregate liquidity and documents for the US market: a time series relationship between liquidity and expected return on the market level.

In this regard, we have to point out those recent studies that analyze the asset pricing implications of a systematic liquidity risk factor. Chordia *et al.* (2000) find that the daily relative changes in individual asset liquidity are strongly related to changes in market and industry aggregates. However, market liquidity as a state variable in an asset pricing framework has been investigated by Pastor and Stambaugh (2003). According to earlier studies that document commonality in liquidity, they argue that changes in aggregate liquidity could be non-diversifiable and therefore a priced risk factor. They construct an aggregate monthly liquidity measure[3] and show that monthly portfolio returns command a positive risk premium for the changes in this measure, even after controlling for other systematic risk factors. Martínez *et al.* (2004), however, do not find evidence relating to this risk factor for the Spanish stock market.

---

2 A relevant example of these issues can be found in the paper by Rubio and Tapia (1996).

3 The first-order autocorrelation measure in returns conditional on signed volume.

Following these two current streams of research, the main object of this study is to construct a risk factor based on liquidity and to analyze pricing implications for the Spanish stock market over the 1994-2002 period. We generate a liquidity factor employing the orthogonal approach procedure of Fama and French (1993) and analyze whether it should be included as an augmented variable in the stochastic discount factor. Another significant contribution of this study is its use of an alternative measure of liquidity based on trading activity of common stocks, the illiquidity ratio suggested by Amihud (2002), that can be interpreted as the daily price response associated with one euro of trading volume.

3. Methodology

3.1. The illiquidity ratio

Liquidity is a broad and elusive concept that generally denotes the ability to trade large quantities quickly, at low cost, and without moving the price (Pastor and Stambaugh, 2003), but liquidity is not an observable variable. There are many proxies for liquidity, such as the relative bid-ask spread, adverse selection, depth, or probability of information-based trading. But these are based on market microstructure data, and are not available for a time series as long as is usually desirable for studying the effect on expected returns. In contrast, the illiquidity measure used in this study is calculated from daily data on returns and volume that are readily available over long periods of time for most markets.

Following previous studies for the US market reported by Amihud (2002) and Acharya and Pedersen (2003), we use the "illiquidity ratio" for our empirical analysis, as being the best proxy measure of illiquidity that computes the price response associated with one euro of trading volume.

The illiquidity ratio of stock *i* in month *t* is calculated as (1),

$$ILLIQ_{it} = \frac{1}{D_{it}} \cdot \sum_{d=1}^{D_{it}} \frac{|R_{itd}|}{V_{itd}} \qquad (1)$$

where R_{idt} and V_{idt} are, respectively, the return and dollar volume on day *d* in month *t*, and D_{it} is the number of valid observation days in month *t* for stock *i*. The reasoning behind this illiquidity measure is as follows. A stock is illiquid, that is, has a high value of $ILLIQ_{it}$, if the stock's price moves a lot in response to little volume.[4]

The advantage of using the illiquidity ratio is that it has a strong theoretical appeal. Hasbrouck (2002) finds that this measure appears to be the best of the usual proxies employed to capture Kyle's lambda.

This measure is interpreted as the daily stock price reaction to one euro of trading volume. Following Amihud (2002), $ILLIQ_{it}$ can also be interpreted as a measure of consensus belief among investors about new information. Thus, when investors agree about the implication of news, the stock price changes without trading, while disagreement induces an increase in trading volume.

Finally, it should be pointed out that this measure can be easily obtained from databases that contain daily data on stock return and volume. This makes it available to most stock markets and enables us to construct a time series of illiquidity over a long period of time, which is necessary for the study of the effects of illiquidity over time. Moreover it allows checking the robustness of the available results.

4 The estimated variable is multiplied by 10^6.

3.2. Mimicking portfolio formation

We propose the construction of an illiquidity-based risk factor, proxied by Amihud's illiquidity ratio, in the context of the Fama and French (1993) framework through the formation of mimicking portfolios. This illiquidity mimicking factor is created by obtaining the difference between the mean return on a set of illiquid stock portfolios (*I*) and the mean return on a set of very liquid (*V*) stock portfolios, named *IMV* (illiquid minus very liquid).[5] The advantage of this construction is that each factor is formed while controlling for the effect of the other Fama and French factors.

For the size and book-to-market portfolio formation procedure, we followed Fama and French (1993). At the end of December in year *t*-1, the companies were ranked by size and partitioned into small (*S*) and big (*B*) companies. Then, the sample companies were ranked by book-to-market and partitioned into three groups, high (*H*), medium (*M*) and low (*L*). Finally, the illiquidity ratio was used to rank companies into very liquid (*V*), moderately liquid (*N*) and illiquid (*I*) companies. We took the average of twelve monthly illiquidity ratios as our measure of the company's illiquidity throughout the year *t*-1 to avoid the possible effect of seasonality.

Based on the independent sorts and ranking procedure, in year *t*-1 we constructed eighteen portfolios from the intersection of the two size, three book-to-market and three illiquidity groups (*S/L/V, S/L/N, S/L/I, S/M/V, S/M/N, S/M/I, S/H/V, S/H/N, S/H/I, B/L/V, B/L/N, B/L/I, B/M/V, B/M/N, B/M/I, B/H/V, B/H/N, B/H/I*).

Following the procedure developed by Fama and French (1993), the size factor *SMB* (small minus big) was calculated each month as the difference between the simple average of the returns on the nine small company portfolios (*S/L/V, S/L/N, S/L/I, S/M/V, S/M/N, S/M/I, S/H/V, S/H/N, S/H/I*) and the simple average of the returns on the nine big company portfolios (*B/L/V, B/L/N, B/L/I, B/M/V, B/M/N, B/M/I, B/H/V, B/H/N, B/H/I*).

The book-to-market factor *HML* (high minus low) was generated each month as the difference between the simple average of the returns on the six high book-to-market company portfolios (*S/H/V, S/H/N, S/H/I, B/H/V, B/H/N, B/H/I*) and the simple average of the returns on the six low book-to-market company portfolios (*S/L/V, S/L/N, S/L/I, B/L/V, B/L/N, B/L/I*).

Also, the illiquidity factor *IMV* (illiquid minus very liquid) was created each month as the difference between the simple average of the returns on the six illiquid company portfolios (*S/L/I, S/M/I, S/H/I, B/L/I, B/M/I, B/H/I*) and the simple average of the returns on the six very liquid company portfolios (*S/L/V, S/M/V, S/H/V, B/L/V, B/H/V*).

3.3. Dependent variable portfolio formation

The next step consists of constructing our dependent variable: 10 illiquidity-based sorted portfolios according to the average illiquidity value of each security in the previous year. P1 includes the stocks with the smallest illiquidity ratio within the sample and P10 contains the stocks with the largest illiquidity ratio. Portfolio returns were also calculated giving equal weight to each asset within the portfolio. These are the portfolio returns which are employed in testing the illiquidity-based asset pricing models in the next sections.

3.4. Research method

Our approach to determining the role of an illiquidity factor in asset pricing was as follows. First of all, we analyzed the standard CAPM model within a time-series context and for each of the 10 illiquid-based portfolios using (2). However, we also analyzed the available results provided by the Fama and French three-factor model using (3). Finally, we tested the standard CAPM and Fama and French three-factor asset pricing models augmented by the illiquidity factor in (4) and (5),

5 As suggested by Chan and Faff (2004).

$$r_{jt} = \alpha_j + \beta_{jm} \cdot r_{mt} + \epsilon_{jt} \quad (2)$$

$$r_{jt} = \alpha_j + \beta_{jm} \cdot r_{mt} + \beta_{jsmb} \cdot SMB_t + \beta_{jhml} \cdot HML_t + \eta_{jt} \quad (3)$$

$$r_{jt} = \alpha_j + \beta_{jm} \cdot r_{mt} + \beta_{jsmb} \cdot IMV_t + \mu_{jt} \quad (4)$$

$$r_{jt} = \alpha_j + \beta_{jm} \cdot r_{mt} + \beta_{jsmb} \cdot SMB_t + \beta_{jhml} \cdot HML_t + \beta_{jsmb} \cdot IMV_t + \mu_{jt} \quad (5)$$

where r_{jt} is the excess return on portfolio j, r_{mt} is the excess return on the market portfolio, SMB_t is the mimicking portfolio for the size factor, HML_t is the mimicking portfolio for the book-to-market factor, and IMV_t is a mimicking portfolio for the illiquidity factor, α_j is the intercept of portfolio j, and β_{jm}, β_{jsmb}, β_{jhml} and β_{jimv} are the sensitivities to the risk factors.

We compared the overall significance of the alternative risk specifications and the statistical significance of the estimated factor exposures. And, following Ferson and Harvey (1999), we run a misspecification test for the hypothesis that the liquidity risk factor may be excluded from the regressions.[6]

We also compared the joint significance of the intercept terms. In the CAPM framework the intercept should be zero. Otherwise there are other sources of risk that are not captured by the market factor. In the multivariate framework a significant intercept implies that not all the relevant variables are included as factors or that firm specific risk is still present in the dataset.

Moreover, we observed the standard zero intercept restriction that constitutes the null hypothesis: H_0: $\alpha_j = 0$; $j = 1, 2, \ldots, 10$ using the Wald test, asymptotically distributed as a chi-square statistic with degrees of freedom equal to the number of restrictions under the null hypothesis.

While we can estimate the models using a variety of different systems-based methods, we have chosen the generalized method of moments (GMM) approach of MacKinlay and Richardson (1991). These tests may be interpreted, within the context of Grinblatt and Titman (1987), as testing that there is one risk factor that is globally mean-variance efficient. Our estimation technique employs the optimal weighting matrix, which is the inverse of the covariance matrix of the sample moments. Specifically, we impose a heteroskedasticity and autocorrelation consistent covariance matrix in all estimation, which involves a Bartlett kernel with a Newey-West fixed bandwidth and no prewhitening. And following Ferson and Foerster (1994), we use an iterated procedure. The initial weighting matrix is obtained using consistent two-stage least squares initial estimates of the parameter set.

4. Data and Empirical Results from the Spanish stock market

4.1. Data issues

The daily prices and trading volume of all stocks traded on the Spanish stock market from January 1994 through December 2002 were used in this study. This daily data is employed for the monthly calculation of firms' illiquidity ratios.

Stock return in month t is calculated as the ratio between its price in month t and in month t-1, adjusted by dividends, splits and new issues. Market return is calculated as an equally-weighted portfolio comprised of all stocks available either in a given month of the sample while the monthly Treasury Bill rate observed in the secondary market is used as the risk-free rate.

In order to construct the Fama-French risk factors, we have used the number of shares traded at the end of each sample year and the accounting information from the balance sheets of each

6 We are grateful to an anonymous referee for bringing this to our attention.

firm at the end of each sample year. The market value is calculated by multiplying the number of shares of each firm in December of the previous year by their price at the end of each month. The book value for any firm in month *t* is given by its value at the end of the previous year, remaining constant from January to December. Then, the book-to-market ratio in all months of year *t* is calculated by dividing the book value at the end of December in previous year by the market value at that date.

4.2. Empirical Results

4.2.1. Background and descriptive statistics

Table 1 reports the average characteristics of the distribution of the market return factor, the Fama-French factors, and the illiquidity-based systematic factor. The correlation coefficients between them are presented in Panel B. It is interesting to point out that the average market risk premium is positive, and hence consistent with the assumption of risk aversion. The mean return for the derived size (*SMB*) factor is negative. In this order, there is evidence in the early anomalies literature that the small firm effect may not be stable over time and may depend on factors such as business cycles. Moreover, recent evidence suggests that the size effect may have gone in reverse. The average return on the *HML* factor is positive, and the average return on the *IMV* factor is also positive. Finally, the correlations between the three last factors are low and correlation with the market factor is quite similar to previous results shown for the Spanish market (Menéndez, 2000; Nieto, 2004; Martínez *et al.*, 2004).

Table 1 – Descriptive statistics and correlation across risk factors

Panel A: Descriptive statistics				
	Mean	**Volatility**	**Skewness**	**Kurtosis**
MKT	0.8762	6.2296	-0.0023	3.9763
SMB	-0.1928	3.5307	0.7985	3.7127
HML	0.2866	3.2991	0.6307	3.4945
IMV	0.2118	11.043	0.9130	6.0562
Panel B: Correlation coefficients				
	MKT	**SMB**	**HML**	**IMV**
MKT	1.000	0.304	0.287	0.122
SMB		1.000	0.077	0.219
HML			1.000	0.155
IMV				1.000

Note: In this Table, Panel A reports the mean, volatility, skewness, and kurtosis for the excess market return (MKT) and for the mimicking portfolio factor returns of size (SMB), book-to-market (HML) and illiquidity (IMV). Panel B reports correlations between the excess market returns and the SMB, HML, and IMV factor returns. Data are monthly covering the period from January 1994 to December 2002.

4.2.2. Main asset pricing test results

For the purpose of comparison, we first run the CAPM and Fama-French model. Results are reported in Table 2, Panels A and B respectively. Several aspects of these results deserve to be mentioned. First, for the CAPM and not the Fama and French three-factor model, the risk-adjusted average return (alpha) of the P10 portfolio is significantly higher than the alpha for the P1 portfolio. Average risk-adjusted returns of stocks with high liquidity exceed those ones with low liquidity. Pastor and Stambaugh (2003) interpret the result as the average liquidity premium existing in the US market. A joint test will be performed later in the paper.

Second, the market factor is significantly related to the excess returns of the liquidity portfolios, in both the CAPM and Fama-French model. However, *SMB* and *HML* are less significant in explaining the excess returns of the liquidity portfolios.[7]

Table 2 – CAPM and Fama-French model performances of sorted by illiquidity

Panel A: Standard CAPM										
	P1	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**	**P8**	**P9**	**P10**
α	0.38 (1.08)	0.84* (2.79)	-0.71* (-2.08)	-0.29 (-0.86)	-0.53 (-1.56)	-0.19 (-0.63)	0.01 (0.02)	-0.07 (-0.25)	-0.16 (-0.40)	0.77 (1.72)
β_m	1.00* (7.74)	0.89 (10.1)	0.96 (13.0)	0.96 (6.50)	0.90 (10.0)	1.26 (10.6)	0.82 (8.31)	0.84 (10.7)	1.30 (5.59)	1.01 (4.84)
Adj. R^2	72.67	67.70	67.64	66.91	65.29	74.32	53.42	66.35	68.19	50.82
Panel B: Fama-French model										
α	0.60* (1.94)	1.06* (4.49)	0.45 (1.22)	0.35 (0.84)	-0.59 (-1.75)	-0.47 (-1.43)	0.22 (0.61)	-1.04* (-3.48)	-0.20 (-0.53)	-0.32 (-0.66)
β_m	1.03* (8.23)	0.97* (9.45)	1.04* (15.2)	1.00* (8.51)	0.98* (12.9)	1.22* (11.4)	0.87* (6.97)	0.81* (12.1)	1.14* (9.84)	0.90* (5.25)
β_{smb}	-0.12 (-0.72)	-0.45* (-3.19)	-0.45* (-2.69)	-0.21 (-1.08)	-0.46* (-4.20)	0.34* (2.53)	-0.27 (-0.98)	0.19 (1.08)	0.87* (3.93)	0.59* (2.14)
β_{hml}	-0.20 (-1.01)	-0.17 (-0.64)	-0.12 (-0.59)	-0.27 (-0.85)	-0.22 (-1.16)	-0.29 (-1.20)	-0.33* (-1.88)	-0.01 (-0.09)	0.87* (2.66)	0.77* (1.99)
Adj. R^2	72.59	72.44	71.48	67.68	70.13	76.66	55.16	66.34	81.49	59.51

* Indicates statistical significance at the 0.05 level

Note: At the beginning of each month from January 1994 to December 2002, stocks are sorted in ascending order based on their illiquidity measures, *ILLIQ*. Based on each sorting, stocks are grouped into equally-weighted decile portfolios and held for 12 months. P1 denotes the lowest *ILLIQ* decile portfolio (the most liquid decile) and P10 is the highest *ILLIQ* decile portfolio (the least liquid decile).

Panel A presents parameter estimates of the capital asset pricing model (CAPM):

$$r_{jt} = \alpha_j + \beta_{jm} r_{mt} + \epsilon_{jt}$$

And Panel B reports parameter estimates of the Fama and French three factor model:

$$r_{jt} = \alpha_j + \beta_{jm} r_{mt} + \beta_{jsmb} SMB_t + \beta_{jhml} HML_t + \eta_{jt}$$

where r_{jt} is the excess return on portfolio j, r_{mt} is the excess return on the market portfolio, SMB_t is the mimicking portfolio for the size factor and HML_t is the mimicking portfolio for the book-to-market factor, α_j is the intercept of portfolio j, and β_{jm}, β_{jsmb} and β_{jhml} are the sensitivities to the risk factors. Numbers in parentheses are t-statistics. The adjusted R-squares are reported in percentages.

7 This is consistent with previous empirical evidence from the Spanish stock market. See Nieto (2004), among others.

Table 3 reports estimates of the standard CAPM and Fama-French model augmented by the illiquidity factor. Panel A in Table 3 reveals that eight of the ten betas are statistically significant for the illiquidity (*IMV*) factor. Notably, there is a strong pattern that the illiquid portfolios have positive or at least less negative *IMV* betas, and three of these cases are significantly positive at the 5% level. In contrast, three of the very liquid portfolios have significantly negative *IMV* betas. Panel B of Table 3 shows that the illiquidity factor exhibits significant explanatory power on the time-series variation of average returns after controlling for Fama-French factors.

Table 3 also reports the adjusted *R*-squares for the time series regressions. It may safely be argued that there is a relevant improvement in the variability of portfolio returns explained by the iliquidity adjusted models. Finally, the *F*-test for the hypothesis that the iliquidity factor may be excluded from the regression is reported. In the CAPM and Fama and French models argumented by the iliquidity factor, the *F*-tests for 8 and 7 of the 10 portfolios produced p-values below 0.05.

Table 3 – CAPM and Fama-French model augmented by an iliquidity factor

Panel A: Standard CAPM augmen by an iliquidity factor										
	P1	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**	**P8**	**P9**	**P10**
α	1.52 (1.24)	-0.01 (-0.04)	-0.70 (-1.37)	0.62 (0.47)	0.74 (1.50)	-0.17 (-0.47)	0.31 (0.88)	0.08 (0.20)	0.44 (0.91)	0.23 (0.59)
β_m	1.04* (11.6)	0.92* (9.27)	1.01* (14.0)	0.99* (10.3)	0.92* (11.2)	1.25* (10.1)	0.84* (8.33)	0.80* (10.6)	1.23* (9.72)	0.93* (8.22)
β_{imv}	-0.13* (-7.18)	-0.10* (-4.49)	-0.15* (-4.12)	-0.09 (-1.62)	-0.07* (-2.55)	0.02 (0.78)	-0.07* (-2.22)	0.10* (4.57)	0.23* (3.71)	0.26* (8.40)
Adj. R^2	81.65	74.09	80.04	70.87	68.00	74.16	55.63	73.81	83.48	75.76
F-test	29.3 (0.00)	15.3 (0.00)	37.0 (0.00)	8.88 (0.00)	5.90 (0.01)	0.65 (0.42)	3.89 (0.05)	17.5 (0.00)	54.7 (0.00)	60.6 (0.00)
Panel B: Fama-French model augmented by an iliquidity factor										
α	1.46 (1.10)	-0.13 (-0.36)	-1.05* (-2.66)	0.42 (0.49)	0.71 (1.47)	-0.21 (-0.57)	0.35 (1.03)	-0.17 (-0.60)	0.31 (0.69)	0.18 (0.50)
β_m	0.99* (9.92)	0.95* (8.59)	1.00* (14.8)	0.99* (9.71)	0.98* (12.8)	1.22* (11.1)	0.86* (6.99)	0.83* (14.3)	1.17* (12.0)	0.95* (7.49)
β_{smb}	0.36* (2.31)	-0.24 (-1.36)	-0.01 (-0.07)	0.03 (0.18)	-0.40* (-3.07)	0.31 (1.75)	-0.16 (-0.40)	-0.19 (-1.10)	0.44* (2.66)	-0.13 (-0.61)
β_{hml}	0.18 (0.85)	-0.01 (-0.01)	0.22 (1.43)	-0.07 (-0.37)	-0.17 (-1.04)	-0.31 (-1.07)	-0.24 (-0.96)	-0.32* (-2.29)	0.54* (2.31)	0.19 (0.59)
β_{imv}	-0.18* (-6.95)	-0.07* (-2.53)	-0.16* (-4.45)	-0.09 (-1.58)	-0.02 (-0.69)	0.01 (0.22)	-0.04 (-0.73)	-0.14* (3.98)	0.15* (2.32)	0.27* (5.57)
Adj. R^2	83.06	74.40	80.13	69.95	69.77	76.25	55.02	74.87	85.75	75.75
F-test	35.6 (0.00)	5.30 (0.02)	25.3 (0.00)	5.21 (0.02)	0.34 (0.56)	0.04 (0.82)	0.82 (0.36)	20.0 (0.00)	17.7 (0.00)	38.5 (0.00)

* Indicates statistical significance at the 0.05 level.

Note: At the beginning of each month from January 1994 to December 2002, stocks are sorted in ascending order based on their illiquidity measures, *ILLIQ*. Based on each sorting, stocks are grouped into equally-weighted decile portfolios and held for 12 months. P1 denotes the lowest *ILLIQ* decile portfolio (the most liquid decile) and P10 is the highest *ILLIQ* decile portfolio (the least liquid decile).

Panel A presents parameter estimates of the CAPM augmented by an illiquidity factor:

$$r_{jt} = \alpha_j + \beta_{jm} r_{mt} + \beta_{jimv} IMV_t + \mu_{jt}$$

And Panel B reports parameter estimates of the Fama and French three factor model augmented by an illiquidity factor:

$$r_{jt} = \alpha_j + \beta_{jm} r_{mt} + \beta_{jsmb} SMB_t + \beta_{jhml} HML_t + \beta_{jimv} IMV_t + \nu_{jt}$$

where r_{jt} is the excess return on portfolio j, r_{mt} is the excess return on the market portfolio, SMB_t is the mimicking portfolio for the size factor, HML_t is the mimicking portfolio for the book-to-market factor, and IMV_t is a mimicking portfolio for the illiquidity factor, a_j is the intercept of portfolio j, and b_{jm}, b_{jsmb}, b_{jhml} and b_{jimv} are the sensitivities to the risk factors. Numbers in parentheses are t-statistics. The adjusted R-squares are reported in percentages. The F-test and its ρ-value for the hypothesis that the illiquidity factor may be excluded from the regression are reported.

Of course, the fact that we have found an apparent improvement in equity pricing using the previous illiquidity factor does not imply that the liquidity adjusted models are the "correct" models. We should also test whether the intercepts in the regressions above are jointly equal to zero.

Table 4 – Test for the joint significance of the intercept terms. Portfolios sorted by liquidity

Model	Wald test[1]	p-value
Standard CAPM	22.673	0.012
Fama and French model	36.765	0.000
CAPM augmented by *IMV*	15.276	0.122
Fama and French model augmented by IMV	18.581	0.045

1 Under the null hypothesis asymptotically distributed χ^2_{10}.
Note: Comparison of competing models: the standard CAPM, the Fama and French three-factor model, and both of them augmented by an illiquidity factor, named *IMV*. The joint significance of the intercept terms is analyzed employing the Wald test with ten portfolios sorted by liquidity for the period January 1994 – December 2002. Portfolios are equally weighted.

Table 4 reports the results of the Wald test that analyzes whether portfolio intercepts are jointly equal to zero. This test also indicates the risk specification suitable for the Spanish market. In other words, whether the models completely capture average returns when used as asset pricing models.

The Wald test is rejected with a significance level of 5% for all asset models considered except the third, the illiquidity-based CAPM. It is relevant to point out that we obtained the best risk specification using these portfolios, and adding the illiquidity factor to the CAPM . In addition, with a significance level of 1%, we cannot reject the null hypothesis for asset models augmented by the illiquidity risk factor.

4.2.3. Robustness check

It may not be surprising that an illiquidity factor like *IMV* that is formed on *ILLIQ* can explain the returns on (decile) portfolios formed on the same measure. In other words, the fact that the liquidity-adjusted model can account for the illiquidity-based portfolio returns that the CAPM and the Fama-French three-factor model may not be convincing proof that the liquidity-adjusted model performs better than the CAPM and the three-factor model.[8]

Therefore, before drawing some overall conclusions regarding the asset pricing role of liquidity, it is instructive to conduct a robustness check. In particular, we replicate the tests of the previous section using portfolios sorted by a random selection of stocks. Results are reported in Table 5. For comparison, the CAPM and the Fama-French model are also estimated in relation to these portfolios.

8 The authors gratefully acknowledge the suggestion of an anonymous referee.

Table 5 – Robustness check

	R1	R2	P3	P4	P5	P6	P7	P8	P9	P10
Panel A: Standard CAPM										
α	0.88	-0.10	-0.75*	0.19	0.51	0.02	0.14	-0.05	-0.05	0.03
	(1.10)	(-0.26)	(-2.12)	(0.33)	(1.30)	(0.07)	(0.46)	(-0.21)	(-0.14)	(0.13)
Adj. R^2	43.27	73.38	70.12	48.47	71.99	80.52	64.72	71.73	44.69	55.13
Panel B: Fama and French model										
α	1.26	-0.21	-1.00*	0.50	0.75	-0.16	0.38	-0.02	0.38	0.26
	(1.00)	(-0.60)	(-2.57)	(0.58)	(1.50)	(-0.45)	(1.13)	(-0.08)	(0.81)	(0.73)
Adj. R^2	41.30	86.43	86.82	47.62	71.83	80.34	78.50	69.83	42.04	50.79
Panel C: Standard CAPM augmented by an iliquidity factor										
α	1.52	-0.01	-0.70	0.62	0.74	-0.17	0.31	0.08	0.44	0.23
	(1.24)	(-0.04)	(-1.37)	(0.47)	(1.50)	(-0.47)	(0.88)	(0.20)	(0.91)	(0.59)
β_m	1.13*	1.11*	1.24*	0.93*	0.87*	1.01*	0.79*	0.91*	0.54	0.60*
	(5.25)	(21.6)	(8.56)	(6.22)	(12.2)	(11.7)	(15.2)	(5.75)	(4.92)	(7.08)
β_{imv}	0.23*	0.14*	0.10*	0.06	-0.02	-0.03	-0.05*	-0.06*	-0.03	-0.07*
	(2.63)	(4.51)	(3.19)	(0.32)	(-1.43)	(-1.09)	(-1.98)	(-1.73)	(-0.80)	(-2.12)
Adj. R^2	49.62	78.27	73.54	47.93	72.18	80.46	67.91	73.36	46.01	58.07
F-test	12.1	35.2	6.61	0.01	1.29	2.12	6.11	5.43	1.05	7.74
	(0.00)	(0.00)	(0.01)	(0.89)	(0.25)	(0.15)	(0.01)	(0.02)	(0.30)	(0.00)
Panel D: Fama-French model augmented by an iliquidity factor										
α	1.46	-0.13	-1.05*	0.42	0.71	-0.21	0.35	-0.17	0.31	0.18
	(1.10)	(-0.36)	(-2.66)	(0.49)	(1.47)	(-0.57)	(1.03)	(-0.60)	(0.69)	(0.50)
β_m	1.12*	1.05*	1.08*	0.84*	0.86*	0.98*	0.81*	0.82*	0.50*	0.59*
	(5.31)	(15.8)	(16.8)	(6.81)	(11.6)	(11.7)	(13.4)	(9.65)	(5.06)	(7.35)
β_{smb}	0.09	0.47*	1.31*	0.70*	0.11	0.25*	-0.17*	0.66*	0.30	0.07
	(0.28)	(3.35)	(6.52)	(3.44)	(0.61)	(1.99)	(-1.78)	(2.29)	(1.63)	(0.39)
β_{hml}	0.16	0.07	0.30	0.23	0.04	-0.05	-0.01	-0.50*	-0.29*	0.13
	(0.45)	(0.54)	(1.23)	(0.96)	(0.23)	(-0.41)	(-0.10)	(1.79)	(2.02)	(0.59)
β_{imv}	0.21*	0.08*	-0.05	-0.08*	-0.04	-0.06*	-0.03	-0.16*	-0.07	-0.08*
	(1.95)	(2.07)	(-1.28)	(-2.28)	(-1.44)	(-1.77)	(-1.14)	(2.49)	(-1.69)	(-1.80)
Adj. R^2	46.18	88.29	87.16	48.28	72.17	81.30	78.66	79.70	45.20	55.27
F-test	6.08	9.91	2.48	1.71	1.68	3.89	1.42	28.2	4.22	6.61
	(0.01)	(0.00)	(0.12)	(0.19)	(0.19)	(0.05)	(0.23)	(0.00)	(0.04)	(0.01)

* Indicates statistical significance at the 0.05 level.

Note: At the beginning of each month from January 1994 to December 2002, stocks are sorted randomly and grouped into equally-weighted decile portfolios and held for 12 months.

Panel A presents alpha estimates of the standard CAPM. Panel B reports alpha estimates of the Fama and French model. Panel C presents parameter estimates of the standard CAPM augmented by an illiquidity factor. And Panel D reports parameter estimates of the Fama and French model augmented by an illiquidity factor.

Numbers in parentheses are *t*-statistics. The adjusted *R*-squares are reported in percentages. The *F*-test and its *p*-value for the hypothesis that the illiquidity factor may be excluded from the regression are reported.

Table 6 – Test for the joint significance of the intercept terms. Robustness check

Model	Wald test[1]	p-value
Standard CAPM	23.832	0.021
Fama and French model	58.816	0.000
CAPM augmented by *IMV*	13.531	0.195
Fama and French model augmented	28.135	0.001

1 Under the null hypothesis asymptotically distributed χ^2_{10}.
Note: Comparison of competing models: the standard CAPM, the Fama and French three-factor model, and both of them augmented by an illiquidity factor, named *IMV*. The joint significance of the intercept terms is analyzed employing the Wald test with ten portfolios sorted randomly for the period January 1994 – December 2002. Portfolios are equally weighted.

Panels C and D of Table 5 indicate that the illiquidity factor loadings are significant for 6 of the 10 portfolios. Adjusted R-squares and F-test results also indicate that the liquidity-adjusted models perform better than the CAPM and the three-factor model. But it is interesting to point out that only the liquidity-adjusted CAPM presents insignificant risk-adjusted average returns across all portfolios.

Finally, Table 6 reports the results of the Wald test that analyzes whether portfolio intercepts are jointly equal to zero. Again, the liquidity-adjusted CAPM obtains the best risk specification.

We may then conclude that, within a time-series context, this paper presents evidence showing that an illiquidity risk factor plays a relevant role in explaining the average returns in the Spanish market.

5. Conclusions

In this paper, we have analyzed the role of liquidity as an additional factor in asset pricing. The motivation for our study was provided by the growing interest in liquidity that has emerged in the asset pricing literature over recent years.

Our empirical results support the recent evidence found in US market data and allow us to affirm that aggregate illiquidity should be a key ingredient of asset pricing models. Our results indicate that time-varying expected excess asset returns in the Spanish stock market, from January 1994 through December 2002, can be explained by an illiquidity-based CAPM model.

However, it must be recognized that our sample period is short in comparison to the available evidence on asset pricing. The results should be taken as valid just for the period being studied, and more general conclusions should be left for future research when longer series of data will be readily available. We have to point out that one feature of the methodology which reduces its appeal is the complexity surrounding the construction of the size, book-to-market and illiquidity factors. This is particularly so in smaller markets where extensive and reliable data over sufficiently long time-series are difficult to compile.

Overall, it can be stated that the main goal of the paper has been achieved. However, the observed results suggest that further empirical work would be beneficial. In particular, it would be of interest to explain the cross-sectional variation in illiquidity.

References

Acharya, V.; Pedersen, L. (2003), *Asset Pricing with Liquidity Risk*, Working Paper, London Business School.

Amihud, Y. (2002), «Illiquidity and stock returns: cross-section and time-series effects», *Journal of Financial Markets*, 5, 31-56.

Amihud, Y.; Mendelson, H. (1986), «Asset pricing and the bid-ask spread», *Journal of Financial Economics*, 17, 223-249.

Blanco, R. (1999), «Análisis de la Liquidez en el Mercado Español de Renta Variable e Impacto de las Regulaciones sobre Variaciones Mínimas de Precios», *El Mercado Español de Renta Variable. Análisis de la Liquidez e Influencia del Mercado de Derivados*, Banco de España, Servicio de Publicaciones, 66, Madrid.

Brennan, M. J.; Subrahmanyam, A. (1996), «Market microstructure and asset pricing: on the compensation for illiquidity in stock returns», *Journal of Financial Economics*, 41, 341-364.

Brennan, M. J.; Chordia, T.; Subrahmanyam, A. (1998), «Alternative factor specifications, security characteristics, and the cross-section of expected stock returns», *Journal of Financial Economics*, 49, 345-373.

Chan, H.; Faff, R. (2004), «Asset Pricing and the Illiquidity Premium», *The Financial Review*, forthcoming.

Chordia, T.; Roll, R.; Subrahmanyam, A. (2000), «Commonality in liquidity», *Journal of Financial Economics*, 56, 3-28.

Chordia, T.; Subrahmanyam, A.; Anshuman, V. (2001), «Trading Activity and Expected Stock Returns», *Journal of Financial Economics*, 59, 3-32.

Datar, V.; Naik, N.; Radcliffe, R. (1998), «Liquidity and stock returns: An alternative test», *Journal of Financial Markets*, 1, 203-219.

Eleswarapu, V.; Reinganum, M. (1993), «The Seasonal Behavior of the Liquidity Premium in Asset Pricing», *Journal of Financial Economics*, 34, 373-386.

Fama, E. F.; French, K. R. (1993), «Common risk factors in the returns on stocks and bonds», *Journal of Financial Economics*, 33, 3-56.

Ferson, W.; Foerster, S. (1994), «Finite sample properties of the generalized method of moments in tests of conditional asset pricing models», *Journal of Financial Economics*, 36, 29-55.

Ferson, W.; Harvey, C. (1999), «Conditioning Variables and the Cross Section of Stock Returns», *Journal of Finance*, 54, 1325-1360.

Grinblatt, M.; Titman, S. (1987), «The relation between mean-variance efficiency and arbitrage pricing», *Journal of Business*, 60, 97-112.

Hasbrouck, J. (2002), *Inferring Trading Costs from Daily Data: US Equities from 1962 to 2001*, Working Paper, Stern School of Business, New York University.

Mackinlay, A.; Richardson, M. (1991) «Using generalized method of moments to test mean-variance efficiency», *Journal of Finance*, 46, 511-527.

Martínez, M. A.; Nieto, B.; Rubio, G.; Tapia, M. (2004), «Asset Pricing and Systematic Liquidity Risk: An empirical investigation of the Spanish Stock Market», *International Review of Economics and Finance*, forthcoming.

Menéndez, S. (2000), «Determinantes fundamentales de la rentabilidad de las acciones», *Revista Española de Financiación y Contabilidad*, 106, 1015-1031.

Miralles, J. L.; Miralles, M. M. (2003), «Actividad negociadora y esperanza de rentabilidad en la bolsa de valores española», *Revista de Economía Financiera*, 1, 11-32.

Nieto, B. (2004), «Evaluating multi-beta pricing models: an empirical analysis with Spanish market data», *Revista de Economía Financiera*, 2, 80-108.

Pastor, L.; Stambaugh, R. (2003), «Liquidity risk and expected stock returns», *Journal of Political Economy*, 111, 642-685.

Rouwenhorst, K. (1999), «Local return factors and turnover in emerging markets», *Journal of Finance*, 54, 1439-1464.

Rubio, G.; Tapia, M. (1996), «Adverse Selection, Volume, and Transactions around Dividend Announcements in a Continuous Auction System», *European Financial Management Journal*, 2, 127-153.

Rubio, G.; Tapia, M. (1998), «The Liquidity Premium in Equity Pricing under a Continuous Auction System», *European Journal of Finance*, 4, 1-28.

Tapia, M. (1997), «Resultados Preliminares sobre la Estacionalidad de la Prima por Liquidez en España: Efectos Fiscales», *Información Comercial Española*, 764, 65-75.

Questioning rationality: the case for risk consumption

Maria Isabel Clímaco / Luís Moura Ramos
ISCA / FEUC

resumo

As hipóteses padrão de comportamento racional e de perfeita antecipação do futuro têm sido fortemente postas em causa dada a sua incapacidade para compreender certos consumos de risco. A teoria do "vício racional"de Becker e Murphy constitui um marco na medida em que desencadeou novos desenvolvimentos a partir desta teoria bem como novas e promissoras abordagens baseadas na ciência cognitiva.

Este artigo propõe uma síntese confrontado as duas principais abordagens baseadas em diferentes hipóteses quanto à natureza das preferências temporais dos consumidores.

De um lado assume-se um comportamento racional mesmo em situações limite de consumos de risco – os comportamentos viciantes. Novos desenvolvimentos na explicação dos consumos habituais e viciantes têm em consideração uma abordagem económica e também psicológica com implicações substancialmente diferentes no domínio das políticas públicas.

résumé / abstract

Les hypothèses standard de comportement rational e de perfectif anticipation de l´avenir on été fortement mise en cause en conséquence de sa incapacité pour faire comprendre certaines consumations à risque. La théorie de l´adition rational de Becker e Murphy a déchainé des nouveaux développements aussi que des nouvelles et promisseurs approches axés dans la science cognitive.

Cet article propose une synthèse qui confronte les deux principales approches à partir de différentes hypothèses sur la nature de la préférence temporelle des consommateurs. D´un coté on assume la rationalité des comportements même dans des situations de extrême risque – les comportements additives. D´autres contributions mettent en jeu des contributions économiques autant que psychologiques avec des implications fort différentes dans le domaine des politiques publiques.

The standard assumption of rational, forward looking behavior has been heavily questioned given the impossibility of understanding some risk consumption behaviors within such a framework. The Becker and Murphy theory of rational addiction made a start on this debate fostering new refinements within the original rational theory framework as well as promising approaches based on the latest developments of cognitive science.

This paper makes an overview confronting two main approaches highlighting their different time preferences assumptions. On the one hand the debate assumes rationality even in extreme situations of risk consumption – addictive behavior. On the other, new developments in the explanation of habits and addictive behaviours take an economic-psychological approach into consideration and have substantially different policy implications.

1. Introduction

Economists' debates on the effects of habits on demand and the modelling of addictive behaviour always stress the backward-looking intertemporal correlation of these consumption patterns. Moreover, in recent years, there has been a wider use of the rational choice theory approach, the dominant paradigm in economics to understanding human behaviour. One reason for this tendency seems to result from the mathematical nature of the approach, which provides opportunities for the empirical confirmation of theories. The explanation of risk consumption – a behaviour possibly leading to unwanted and harmful dependency situations – only knew substantive development after the groundbreaking[1] article by Becker and Murphy – BM – (1988). Here, the authors explored the dynamic behaviour of the consumption of addictive goods in detail, and pointed out that many phenomena previously thought to be irrational are consistent with rational optimization according to stable preferences.

This paper is a synthesis of the debate on risk consumption together with its policy implications. Section 2 presents the Becker and Murphy rational approach model. Section 3 summarizes the main extensions of the rational framework, while in Section 4 "non-rational" approaches are reviewed. Finally, in Section 5, the policy implications of the different approaches are discussed.

2. Becker and Murphy's rational model

The rational addiction model follows previous works (Becker et al., 1994) that considered the interaction of past and current consumption in a model with utility-maximizing consumers. The main features of these models are that past consumption of some goods influences their current consumption by affecting the marginal utility of current and future consumption. Therefore, past consumption is reinforcing for addictive goods.

The BM (1988) model deals with these characteristics considering a rational consumer who has a forward-looking behaviour and stable preferences.

Addictive behaviour is usually assumed to involve psychological and physiological effects: tolerance, reinforcement and withdrawal. Tolerance is associated with habit creation, implying that to obtain the same level of pleasure; the amounts consumed must be progressively augmented. Reinforcement means that greater past consumption increases the desire for present consumption. The withdrawal syndrome is linked to the intense physical and mental suffering associated with consumption cessation or decrease. These characteristics are hard to match with rational behaviour since a distinguished feature of harmful addictive behaviour is the apparent trade-off of immediate gratification or pleasure against adverse consequences in the future. While non-addicted individuals appear to recognize this trade-off and are aware of harmful consequences, addicts appear to impulsively ignore the adverse consequences of their actions in favour of immediate pleasure. It is this apparent violation of the standard assumption of rational, forward-looking maximization behaviour that is answered by Becker and Murphy's rational addiction model. Hereafter the economic approach will begin to build up a framework to explain apparently pathological aspects of some behaviours.

The BM model[2] assumes that an individual can consume two goods, an addictive, c, and a non-addictive, y, the utility function at time t can be written as:

$$U(t) = u(c(t), y(t), S(t)) \quad (1)$$

1 For a different survey of the theoretical explanations of addiction see Herrnstein and Prelec (1992).
2 Stigler and Becker (1977) was a first contribution to the BM model.

where $S(t)$ is the stock of addictive capital. At any time t, the individual's utility[3] depends on current consumption $c(t)$, current consumption of $y(t)$ and the stock of past consumption $S(t)$.

The basic definition of addiction is that a person is potentially addicted to *c* if an increase in his current consumption of *c* increases his future consumption of *c*.

According to the BM model, tolerance and reinforcement are the main characteristics of an addictive good. Tolerance implies that current utility is negatively related to past consumption ($\partial u / \partial S = u_S < 0$). Reinforcement means that past consumption leads to higher consumption today ($dc / dS > 0$) and requires that an increase in past consumption raises the marginal utility of current consumption ($\partial^2 u / \partial c \partial S = u_{cs} > 0$)[4].

The model assumes that present and future behaviour is part of a consistent maximizing plan. This intertemporal link is also expressed in the way that past consumption influences current utility through the so-called "stock of addictive capital".

The stock of the addictive capital at time t depends on the stock in time t-1 (S_{t-1}) and the consumption of the addictive good in the period t-1 (c_{t-1}).

More formally:

$$S(t) = (1 - \delta) S_{t-1} + c_{t-1}$$

δ is the rate of depreciation of the addictive stock capital. By assumption, the influence of past consumption decreases over time what is captured by the rate σ. This stock $S(t)$ decreases through the depreciation rate (δ) and increases through consumption of the addictive good (c_t).

Considering the length of life T and a constant rate of time preference δ, the utility function would be:

$$U(0) = \int_0^T e^{-\sigma t} u(t)\, dt$$

Utility is discounted exponentially which implies a constant rate of discounting[5].

2.1. Becker and Murphy model characteristics

One of the main features of the BM model is the concept of adjacent complementarity, which shows that the quantities of the addictive good consumed in different time periods are complementary – which is due to reinforcement in consumption. Thus it is possible to define addiction as a strong complementarity between past and current consumption.

Another key factor to the understanding of rational addictive behaviour is the existence of an unstable steady state[6]. An unstable steady state is one in which a small change in some factor affecting consumption (price or other) can cause a drastic change in consumption, including starting or stopping consumption.

3 They assume that U_t is a strongly concave function of *c* and *S*, and that the lifetime utility function is separable over time.

4 Although this is a sufficient condition for myopic consumers, who do not consider future consequences of their current behavior, it is not so for rational utility maximizers.

5 Exponentially discounting is *time consistent* because the relative values of consumption in any two periods remains constant. BM's model assume that time preference is exogenously determined.

6 A steady state is defined as a situation in which current consumption is just sufficient to offset the depreciation of the addictive stock ($C(t) = \delta S(t)$).

Unstable steady states are crucial to explaining why few people consume small amounts of a highly addictive good, and the majority are either abstainers or consume large quantities[7]; and to understand "cold turkey" quit and binge behaviours.

Becker and Stigler's model shows that steady-state consumption is unstable when the degree of addiction is strong (which means strong complementarity between past and current consumption). This kind of "pathological" consumption rises over time even for people who anticipate future consequences. However, the interaction between persons and addictive goods is crucial in this process. According to Becker and Murphy (1988): "a good may be addictive to some persons but not to others and a person may be addicted to some goods but not to other goods. The importance of the individual is clearest in the role of time preference in determining whether there is adjacent complementarity". It is expected that individuals who discount future more heavily, present-oriented individuals, are potentially more addicted to harmful goods than future-oriented ones.

The unstable steady states also lead to another important feature of addictions – the existence of multiple steady states. However, the utilization of a quadratic utility function[8] cannot explain multiple steady states. Instead, it implies only two steady states, one stable (addictive consumption) and one unstable (near abstinence). Consumption of the addictive good will rise over time when above unstable steady-state levels, and will fall over time (eventually until abstinence) when below the unstable steady state. In the first case, the individual consumes so much that depreciation of capital is more than compensated; in the last case, the individual starts with a positive capital stock but consumes less than depreciates. Therefore, the capital stock will depreciate until it eventually reaches zero.

Whether a consumer becomes addicted or not depends on a series of factors: the initial stock of the addictive good (S_0), the time preference rate (σ), the price of the addictive good and the depreciation rate δ. For example a drop in addictive price[9] or a lower time preference would increase steady state consumption in an addicted individual or, eventually, turn a non-addicted consumer into an addicted one.

The rational framework is also valid to justify particular behaviours of addictive consumption: "cold turkey" and binges. Cold turkey quit means that strong addictions can be stopped only with an abrupt cessation of consumption. The theory suggests that if a rational individual decides to end this consumption this is only possible by lowering the addictive stock through a significant cut in consumption decrease. Because this change in current consumption has a greater effect on future consumption when the degree of complementarity (or the degree of addiction) is stronger, then rational people are supposed to end severe addictions more rapidly than weaker ones[10].

Binge behaviour is common in certain addictions, like alcoholism, overeating etc. The BM model defines it as a cycle over time in the consumption of a good. Although seeming an "archetype of irrational behaviour", it could be consistent with rationality.

In brief, the BM model explains consumption of addictive goods in a rational-choice framework, considering a forward-looking user one who maximizes her/his utility and whose preferences are consistent over time (stable over the life cycle). To draw the consumption history of the addictive consumption, into the intertemporal optimization process, Becker and Murphy consider a stock

---

7 What Which is described by a bimodal distribution of consumption (as in smoking). This is not the case for the distribution of alcohol consumption that is, apparently, more continuous with the majority being moderate consumers.

8 Viewed only as a local approximation to of the true function near a steady state (Becker and Murphy, 1988).

9 Note that price in this context is associated with a full price concept, formed also by all those aspects that result from the money value of any adverse effects like restrictions, limits on availability and new information that raises perception relative to long-term health hazards.

10 The authors claim this behavior rational because the consumer exchanges a large short-term loss in utility (withdrawal) for an even larger long-term gain.

variable ("addictive consumption capital") serving as direct link between past and present consumption.

Rationality in this context implies a consumer who considers the future consequences of his/her current consumption decisions and discounts the future exponentially, at a constant rate.

3. Extensions of the rational addiction model

3.1. Learning by consuming – risking addiction

One of the most criticized features of the rational addictive model is the common implicit assumption of "perfect foresight". Those theories, however, fail when it comes to explaining the possibility that consumers of addictive goods regrets their past decisions and become naturally unhappy *ex post*[11].

Orphanides and Zervos' (OZ) theory of "learning and regret" (1995) intended to resolve these criticisms by proposing a new extension of BM rational addiction model. The authors recognise uncertainty and the initial inexperience of consumers as the essential features lacking in those models.

The decision to take the inevitable risk of becoming addicted is considered a rational one. Consuming an addictive good is in this context a voluntary choice, yet not an intentional one; the individual recognizes that addiction is harmful and does not wish to become addict *ex ante*. Despite his rationality he is uncertain about his predisposition to addiction, but at same time he knows it cannot be detected without the experienced gained from repeated consumption.

While experimentation is optimal for the ordinary goods it is problematic for potential addictive goods[12]. The "regret model" considers that consumption of those goods has a different addictive potential for different people, but that nobody knows its potential in advance. Nevertheless, everybody has a subjective assessment of his addictive potential and updates it up given the experiences made.

This dependence on beliefs emphasizes the key role of information and the crucial importance of individual initial beliefs in determining the risk of addiction through experimentation. Naturally those who thinkstrongly that they are non-addict individuals are more likely to risk the consumption experience; whereas if they are less confident about their judgements, they are more likely to abstain, and thus may never learn their potential.

3.1.1. The Model

The population is divided in two groups: the non-addicts and the potential addicts. Two goods are available, at any point of time *t*: an ordinary good *y* and a potentially addictive good α[13]. In this framework *a* is not necessarily addictive, particularly if the individual manifests no addictive tendencies.

As in the BM model, becoming addicted calls for the accumulation of a stock of past consumption (S_t) beyond some level – a designed "critical level" (S_c)[14]. Thus, consuming the potentially addictive good augments addictive capital, but contrary to the BM model, addictive capital only has an influence on utility for some people. Similarly the OZ model considers a constant

---

11 The prediction of the effects of post-behavioural regret or as it has been referred to in the literature, 'anticipated regret' has been based upon the regret theory (Bell, 1982). This theory assumes that the value of choosing one alternative is dependent on the alternatives simultaneously rejected and that people attempt to avoid decisions that could result in regret.

12 With some highly addictive substances, as like cocaine or heroin, the strong risk of addiction seems sufficient for the majority optimally choosinge not to experiment with the good.

13 Defined as *c* in the BM model.

14 Which corresponds to the BM unstable steady state.

depreciation rate (δ) of the consumption stock S. For isolating the differences across individuals, mentioned above, the authors introduce θ, to indicate the presence of individual addictive tendencies. Departing from an uncertain context, a person does not know, *a priori*, her value of θ. It is assumed that $\theta = 0$, if she has no addictive tendencies, and that $\theta = 1$ for potential addicts.

Formally, the momentary utility function, for any individual, appears in two separate parts:

$$U(t) = u(c_t, \alpha_t) + \theta\eta_t\nu(\alpha_t, S_t) \qquad (1)$$

$u(c_t, \alpha_t)$ represents the immediate positive rewards for current consumption of both goods, and $\nu(\alpha_t, S_t)$ represents the detrimental addictive side effects of past consumption (for example, craving, depression and illness). The more the individual has consumed in the past, the higher is the probability of such harmful effects occurring. The term η_t is a random variable with distribution:

$\eta_t = 1$ with probability $\pi(S_t)$

0 with probability $1 - \pi(S_t)$, with $\pi(0) = 0$, $\pi(S) \in \{0,1\}$

If somebody starts consuming the potentially addictive good and no shock occurs on utility, she knows that she either has no addictive potential or she has luck. If no harmful side effects are observed the individual is sure of being a non-addictive type. On the contrary, if the shock of utility happens, she instantly concludes she has addictive potential. OZ call this kind of behaviour "the learning experimentation".

Learning by consumption experimentation allows the self observation of one's addictive potential and the updating of the subjective beliefs about an individual's addictive tendencies. Contrary to the BM rational consumer, experimenters take the initial risk because they get an immediate reward, utility gain, but addiction is not certain to follow.

The chances of an individual becoming addicted are driven by the momentum of his addictive potential discovery. In general, some potential addicts discover their tendencies early through experimentation (if they are below the critical level S_c), and they manage to control their addictive tendencies by changing their consumption patterns. They eventually rapidly change their consumption ("cold turkey" effect), returning quickly to the low stable steady state ($S^s{}_1$).

Others realize their true θ too late (above the critical level S_c) and increase their consumption of the addictive good (go on a "binge") until reaching the high consumption steady state $S^s{}_2$[15]. The process of experimentation acts as a *signal* that permits individuals to continuously update their subjective beliefs (an endogenous resolution of the initial uncertainty).

Thus, the OZ "learning and regret" theory coexists with the rational framework, while makes it possible to explain the existence of addicted consumers and casual users. As they say, "without the appeal of controlled casual consumption, potential addicts would never risk addiction".

Finally, the authors also shed light on the "apparent paradox" of voluntarily being drawn into a harmful addiction and later regretting it. The model identifies "multiple motives" for the individual's regret. He may regret his *bad luck*, or that he learned his true θ too late, or even regret a wrong assessment of his probability of becoming addicted, possibly being overoptimistic.

In fact it is precisely this misinformation problem that explains how rational agents are fooled or "hooked" into an addiction.

15 In difference to BM (1988) where any positive consumption in the steady state is defined as addiction.

3.2. Consumption cycles – the dual effect of consumption capital

One of the most criticized features of the BM model is its insufficient justification of changing consumption patterns over time. The extension developed by Dockner and Feichtinger (1993) (and later by Orphanides and Zervos, 1998) presents an explanation for cyclical consumption patterns.

As mentioned, the BM model assumes that past consumption affects current consumption through a stock consumption variable. A consumer can be called addicted to a good if its consumption increases with that stock. This concept implicitly assumes that the addictive good, *c*, accumulates a single stock of consumption capital (S_t). Dockner and Feichtinger (DF) call this stock "commodity specific consumption capital".

Furthermore, they suggest that the consumption of addictive goods not only has an addictive element, but may also have reverse effects, such as the risk of severe health problems. To capture those effects a separate analysis of addiction is required. In order to achieve it they presented a model of addiction in which a single consumption good accumulates two capital stocks (S_1 and S_2). S_1 corresponds to the capital stock of the addictive good in BM model and S_2 represents the negative effects of addictive consumption.

It is the consideration of two stocks that causes the "irregular" behaviour. These counterbalancing effects (caused by S_1 and S_2) mean that optimal consumption might exhibit cycles – an addictive one and a satiating one. The addictive forces cause the increase of current consumption as past consumption accumulates (ascending part of the cycle); the satiating force causes the decrease of current consumption as habits accumulate (the descending part of the cycle).

The utility function for a representative consumer who at each instant of time *t* derives utility from consumption of good *c* and accumulates two different consumption capitals – S_1 and S_2 – will be:

$$U(t) = U(C_t, S_{1t}, S_{2t})$$

Stocks, S_1 and S_2, are measures of past consumption of c_t that affect current utility through an accumulation process:

$$S_1(t) = f_1 c(t) - \delta_1 S(t) \text{ and } S_2(t) = f_2 c(t) - \delta_2(t)$$

δ_1 and δ_2 are constant depreciation rates, assuming that $\delta_1 > \delta_2$.

Yet the consumption cycles are the result of forward-looking behaviour. Only a smoker who desires to smoke but cares about his health, and anticipates the future consequences of his current consumption, can end up in cycles of smoking and giving up smoking.

DF model demonstrates that consumption behaviour may end up in "persistent oscillations". Contrary to the smooth evolution predicted by BM model, the authors described an addictive consumption most often characterized by periods of lower consumption followed by high consumption, and so explaining why "binges" continue to cycle much through a person's lifetime.

3.3. "Myopic-rationality"

Some of the models that overcome the incompatibility of rational theory with one of the defining aspects of addictive behaviour, the apparent difficulty with delaying gratification and disregard for the future, emphasize non-rational aspects of addictive behaviour, namely myopia. Orphanides and Zervos, (1998) consider the possibility that individuals may be initially uncertain with regard to the degree the addictive good renders them myopic. Initially, individuals must weigh the momentary benefits of consumption against the negative future effects and the potential risk of severe myopia. This uncertainty provides a simple mechanism which leads the individual – with some probability – into a harmful addiction, and highlights the cause of the *ex post* regret associated with addiction.

They propose a rational model that provides a resolution for this rational approach weakness. Addicts' myopia and non-addicts' farsightedness can be easily reconciled within an optimizing expected utility framework which maintains the desirable properties of current rational models yet yields some predictions often associated with the non-rational approach. The key assumption is that as past consumption of addictive good increases, the rate of time preferences increases as well. Two sorts of individuals would then exist: one with accumulated consumption capital (in BM terms) with induced myopia, and others who choose not to experiment the addictive good maintaining a normal fixed rate of time preference.

The increased impatience from consumption of the addictive good enhances the desirability of present utility and diminishes the perceived future costs associated with current consumption. This has the effect of making consumption even more desirable and generates a reinforcement mechanism, which is precisely what may lead to addiction. The resulting myopia is a side effect of addiction and not its cause, as posited in the non-rational framework.

This model stays explicitly within the confines of dynamically consistent rational preferences, exploring the bounds of standard rational assumptions. This development of their work illustrates the importance of the existence of heterogeneous outcomes stemming from risk consumption enabling a better understanding of addiction and *ex post* regret.

Following the same trail, Vanini and Braun (2002) suggest that only by considering the degree of impatience as a determinant of the intensity of substance use will we be able to distinguish habits from addictions. Time preferences will thus depend on consumption history and it will be the accumulation of a large stock of consumption capital that generates addictive behaviour. Adopting this point of view, the assumption of exogenously fixed time preferences will only be suitable for analysing habitual consumption but would not allow addiction to be tackled.

The consideration of time inconsistent decision-makers from the outset would then be fundamental to explain consumers' non-rational behaviour namely the lack of self-control.

4. Models with time inconsistent preferences

Many researchers that have studied time-preferences have proposed formal and general models of time-inconsistent preferences[16]. Based on the seminal papers of Strotz (1956) and Pollak (1968), O'Donoghue and Rabin (1999a, 1999b) presented a time – inconsistency and self-control model of addiction.

Particularly they have coined the term "present-biased preferences" to mean that people's preference have a bias for the "present" over the "future" (where present is constantly changing)[17].

The steady state model of addiction developed by BM (1988) shows how it can be optimal for an individual to maintain a severely harmful consumption, assuming that people have naturally forward-looking behaviour and they are fully rational; they fail, however, to explain why the person chooses and maintain this harmful consumption.

---

16 In the context of dynamic inconsistency of consumption plans, Mistri (2002) recognises the fundamental importance of dynamic processes of cognitive and physiological mechanisms to explain addiction. Although they are foreign to economic analysis, economists "can and must focus on the economic effects of these factors".

17 In accordance with O'Donoghue and Rabin this is merely their term for an array of older models that went under different names. The (β,δ) preferences that they used in their papers are identical to the preferences studied by Laibson (1994) who uses the term "hyperbolic discount". Prelec (1990) uses the term "decrease impatience" for an alternative formulation of the same phenomena (O'Donoghue and Rabin, 1999a).

O'Donoghue and Rabin (1999b, 2001) introduce self-control problems – e.g., a person's awareness of future self-control addictive consumption – to determine whether it could be a credible cause of addictive over-consumption.

They present a model of addiction in which an individual has to decide in each period of time whether to consume or not (a binary choice model). This simplified version considers two of the more relevant characteristics of harmful addictive goods, negative internalities and habit formation, to explain the "trap of addiction". Negative internalities[18] say the more the product has been consumed in the past the smaller the person's current well-being (it includes health, job or personal consequences of past consumption). To generate those internalities these goods are necessarily habit-forming: the more the individual has consumed in past the more he will desire to consume now.

According to the theory, the combination of the two characteristics creates the core of the choice of becoming an addicted. The increasing consumption of the addictive good provokes less and less pleasure in the consumer, yet he may continue to consume it because refraining is increasingly painful.

To resolve this trade-off, the individual must choose between the current desire of consumption ("temptation to hit") and its future costs. The power of each, whether the current desire to consume out-weights the future costs of this consumption, depends on the individual intertemporal preferences.

Time consistency, a person's relative preference for well-being at an earlier date is the same no matter when she is asked, is assumed as a matter of course in the standard economic models. Observation and psychological research show, on the contrary, that the consistency assumption is incorrect because it ignores the human tendency to pursue immediate gratification: a person's relative preference for well-being at an earlier date over a later date gets stronger as the earlier date gets closer (O'Donoghue and Rabin, 2001).

O'Donoghue and Rabin's model assumes that people have time-inconsistent present-biased preferences, and so individuals have self-control problems because they pursue immediate gratification in a way that does not correspond to their long-run well-being.

To study the problem of self-control, the theory considers two extreme possibilities about the awareness of future self-control: *sophisticates* – those people who are fully aware of their future self-control problems, and *naifs* – those who are wholly unaware of their future self-control problems.

In addition they consider a third possibility: people with standard, time-consistent preferences (TCs) used as a reference point for methodical comparisons with *sophisticated* and *naïf* individuals. Those comparisons enable some interesting conclusions about naif behaviour, but are relatively non-conclusive about sophistication. Thus, the sophisticates can be more or less prone to consume than naifs; this uncertainty about their behaviour tendency arises because there are two ways in which their full awareness about future self-control can influence their current behaviour. First, they are pessimistic about their future, believing that they will consume more in the future than they would want (*pessimist effect*). The combination of the *pessimistic effect* with habit-forming characteristics of addictive goods may exacerbate the over-consumption due to present-biased preferences. However, a second effect (*incentive effect*) may contribute to refrain from their current consumption in order to induce themselves to resist temptation in the future. The two effects are then acting as counterbalance forces, which leads to some ambiguity about their future behaviour. Therefore, the sophisticates can be more or less prone to consume (hit) than naifs, due to their awareness and the ambiguity element by they can moderate or exacerbate over-consumption in the future.

18 The term "internalities" was earlier used earlier by Herrnstein, Lowenstein and Prelec (1993) who define an internality to be a "within-person externality" (O'Donoghue and Rabin, 2001).

With respect to the naif' behaviour it is plausible that they show always more tendency to consume than TCs do. This reflects over-consumption as a direct consequence of their present--oriented preferences.

Formally, O'Donoghue and Rabin (1999b) present a simplified version of their model, by considering consumption as a binary choice and a discrete-time model with periods 1... T. In each period, a person can either consume (or "hit", $\alpha_t = 1$) or not consume (or "refrain", $\alpha_t = 0$).

The effect of past consumption is captured by k_t, the *person's addiction level* in period *t*. It is assumed that *k* evolves according to the equation:

$$k_t = \Upsilon k_{t-1} + \alpha_{t-1}$$

where $\Upsilon \in (0, 1)$ is a parameter[19] indicating the rate at which an addiction decays. If $\Upsilon = 0$, this means that refraining for a single period gets the individual completely un-addicted. If the parameter Υ is close to one, refraining reduces the person's addiction level very little.

Applying a simple form of present-biased-preferences, and using a model originally developed by Phelps and Pollak (1968) and later by Laibson (1994, 1997), the intertemporal utility function would be:

$$U^t(u_t, ut_{t+1}, \ldots u_T) \equiv \delta^t u_t + \beta \sum_{\tau=t}^{T} \delta^{\tau-t} u_T$$

The parameter δ represents the "time-consistent" discounting, while the parameter β represents the "present bias". For $\beta = 1$ these preferences reduce (the discrete version) exponential discounting; when $\beta < 1$ these preferences will capture the time-inconsistent preference for immediate gratification.

By assuming than an optimal consumption path exists, the authors say that sophisticates and naifs trace an "optimal" consumption path and, eventually, choose the current action that is part of that consumption path. However, if TCs are always attached to the consumption path prior chosen, naifs often adjust their chosen consumption paths as their preferences change overtime.

One of the main purposes of the O'Donoghue and Rabin model is to find out whether self-control problems are a plausible source of severely harmful addictions and to draw some important conclusions on welfare grounds.

Are then people sophisticated or naïve[20]?

A person could be sophisticated[21] and knowing exactly what her futures selves will be. Or, a person could be naïve and believe her future selves' preferences will be identical to her current self's, and not realizing that her preferences will change as the timing of the decision gets closer. Finally they admit there could be intermediate cases: for instance, an individual might be aware of his present-biased preferences, but he/she may underestimate the degree of present bias.

The theory says that to the extent a person is sophisticated, he may suffer severe harm due to the "feeling of inevitability", but to the extent that a person is naïve, he may suffer due to the delay in quitting an established addiction.

---

19 The parameter Υ corresponds to $(1 - \delta)$ in the Becker and Murphy model (O'Donoghue, 2001).

20 The fact that most of economists who studied inconsistent preferences assumes sophistication is clearly justified by "rational expectations" implicit in the sophistication concept.

21 Sophisticates, like TCs, predict exactly how they will behave in the future. But, like naifs, they discount the future cost from hitting by $\beta\delta$.

Finally the authors conclude while in real-world environments lifelong feelings of inevitability seem unlikely, temptation to consume addictive goods seems universal. Their analysis suggests that for realistic environments, self-control problems are a probable source of harmful addictions only in combination with at least some degree of naiveté (naïve self-control problems). There are, probably, elements of both sophistication and naiveté in the way people anticipate their own future preferences.

O'Donoghue and Rabin (2001) point out that "...an addicted person might suffer severe harm because she procrastinates quitting – she wants to quit, and always plans to quit in near future, but never gets around to it. Moreover, naifs might develop several harmful addictions in the first place because they naively give in, temporarily, to high temptations believing they'll just quit after the temptation subsides, when in fact they end up with long-term addictions".

The inclusion of time-inconsistency[22] makes it possible to extend the role of government policy, which should depend not only on the externalities that addictive consumers eventually impose on others, but also on the "internalities" imposed by addicts on themselves (model of no-externality type[23]).

5. Policy Implications

5.1. Rational addiction and the effect of prices

In a fully rational framework (BM) there seems to be no other reason for public policy to control risk consumption except where net external costs are present. Even though any individual addict can be making optimal choices, the utility of society as a whole is inevitably reduced by addiction. All of the theoretical models presented emphasize the role of price and the importance of past consumption and future price anticipation in current consumption. Empirical applications of the rational model proved, contrary to the conventional wisdom, that addicted consumers are actually responsive to prices[24]. The key implication of the existence of a consumption stock in these models is the greater impact of any permanent price policy in the long-term than in the short--term, namely through excise taxes. Sizable long-run price elasticities were found and also they were much bigger than short-run elasticities[25]. Empirical studies even provided strong support to differential government intervention oriented towards different groups, particularly among youth. This group, along with lower income groups, displays a higher sensitivity to the monetary component of full price.

Lower income earners and young people also appear to discount the future more heavily. It can also be shown that addicts with higher discount rates respond more to changes in monetary prices whereas addicts with lower rates of discount respond more to changes in harmful consequences (Becker *et al.*, 1991).

5.2. Regret and myopia as addiction consequences – information policies

The assumption that individuals are fully aware of the consequences of addictive goods consumption is one of the most criticized aspects of rational models. The learning and regret

22 For the quasi-hyperbolic discounters, discounted utility becomes.

$$U_t + \beta \sum_{i=0}^{T-1} \delta U_{t+i}$$

23 The authors admit that one might look at the intra-personal conflicts that are generated by the hyperbolic model as intra-personal externalities.
24 Either a higher price of the good (due perhaps to a large tax) or a higher future cost (due perhaps to greater information) reduces consumption in both the short and long-run.
25 Becker, Grossman and Murphy (1991) refer cite empirical studies where the evidence from smoking, heavy drinking and hard gambling strongly supports the rational addiction model. For more recent empirical evidence see Chaloupka (1991) and Chaloupka and Wechsler (1997).

model, considering the initially uncertainty regarding the real possibility of becoming addicted, and the importance of subjective beliefs, brings a new insight to public policy goals.

Admitting that the government is in a better position to pass along better information, the theory regards as key policy elements the use of educational programs and advertising campaigns to discourage addictive consumption, especially among young people – those who are particularly exposed to the risk of addiction. This kind of argument also sustains legal restrictions like a minimum age for legal purchase.

The dissemination of information also plays an important role in curtailing the potentially harmful effects of peer pressure. Studies in social psychology have shown that peer pressure leads adolescents to systematically overestimate actual addictive consumption by their peers and thus underestimate the potential harm. Education programs could seek to rectify this misinformation problem "by showing young people that contrary to their own beliefs most of their peers do not use drugs". When we also take into account the existence of significant quitting costs[26], price changes and health information dissemination can also be seen as powerful tools to prevent people from starting to consume or as an encouragement for current consumers to quit sooner. In this way, painful *cold-turkey* quits may be avoided.

5.3. Widening public intervention – addiction as a self-control problem

Self-control problems viewed as the source of over-consumption of addictive products imply a different public policy emphasis. The distinction between two extreme kinds of consumers, the sophisticates and the naifs, means that those self-control problems, as a source of harmful addiction, may emerge only in combination with some degree of naiveté. Naif consumers may usually underestimate their future behaviour and so the theory expects that they would suffer a relatively small change in behaviour effect as a result of anticipating future prices, with regard to a time consistent consumer. In other words, while the qualitative effects of price changes are the same, smaller quantitative effects can be expected, and so less efficiency can be expected from price policies.

It is possible to perceive addiction as a market failure, in the sense that the market itself does not supply a convenient "self-control device"[27]. The voluntary use of self-control is seen as crucial, for instance in attempts at quit smoking, and so public intervention could be devised to make the teaching and dissemination of self-control more effective. However_one cannot expect this policy to be efficient with naïve consumers. So, for consumers with some level of naiveté, public intervention corrects both a self-control problem and a misperception problem, since the naïve agent is incorrect in predicting his future behaviour due to cognitive limitations[28]. Thus public interventions could also be justified because of internalities, leaving the way open for a more paternalist policy standpoint.

6. Concluding comments

In our societies market regulation policies are usually justified by market failures. Addiction, a phenomenon with negative social impacts, claims for public intervention that, assuming individuals pursuing rational decisions, could only deal with interpersonal externalities. This being so, taxation, for instance, should be designed in accordance with the size of external costs.

But people might develop harmful addictions due to rational uncertainty about the addictiveness of the product. Information policies are then necessary on the grounds of these goods' special features and full information unawareness on the part of individuals, with special emphasis on the young.

26 For a more detailed analysis see Suranovic, (1999)

27 Even if firms do have a financial incentive to provide self-control to agents with self-control, other firms have a financial incentive to break it.

28 For the results of an experiment to test addicts' rationality see Fehr, E and Zych, P. (1998).

Finally, individual choice behaviour in general and risk consumption in particular does not always conform to the fundamental premises of rational choice theory. Addictions obey certain qualitative hedonic regularities like the saliency of present benefits and distributivity of future costs. The introduction of behaviour insights stemming from cognitive science into the modelling of risk consumption allows a better explanation of individual choices. Self-control problems may then be the source of addiction and policies designed to enhance self-control may be called for, even if they threatened consumer sovereignty.

References

Becker, G.; Murphy, K. (1988), «A Theory of Rational Addiction», *Journal of Political Economy*, 96, 675-700.

Becker, G.; Grossman, M.; Murphy, K. (1991), «Rational Addiction and the Effect of Price on Consumption», *American Economic Review*, 81, 2, 237-241.

Becker, G.; Grossman, M.; Murphy, K. (1994), «An Empirical Analysis of Cigarette Addiction», *American Economic Review*, 84, 3, 396-418.

Bell, D. E. (1982), «Regret in Decision Making under Uncertainty», *Operations Research*, 30, 961-81.

Chaloupka, F. (1991), «Rational Addictive Behavior and Cigarette Smoking», *Journal of Political Economy*, 99, 4, 722-742.

Chaloupka, F.; Wechsler, H. (1997), «Price, Tobacco Control Policies and Smoking Among Young Adults», *Journal of Health Economics*, 16, 359-373.

Dockner, E. J.; Feichtinger G. (1993), «Cyclical Consumption Patterns and Rational Addiction», *American Economic Review*, 83, 1, 256-263.

Fehr, E.; Zych, P. (1998), «Do Addicts Behave Rationally?», *Scandinavian Journal of Economics*, 100, 3, 643-62.

Herrnstein, R. J.; Prelec, D. (1992), «A Theory of Addiction», in G. Lowenstein and J. Elster (eds.), *Choice Over Time,* Russell Sage Foundation, 331-360.

Herrnstein, R. J.; Lowenstein, G.; Prelec, D.; Vaughan Jr., W. (1993), «Utility Maximization and Melioration: Internalities in Individual Choice», *Journal Of Behavioral Decision Making*, 6, 149-185.

Laibson, D. (1997), «Golden Eggs and Hyperbolic Discounting», *Quarterly Journal of Economics*, 112, 2, 443-477.

Laibson, D. (1994), *Essays in Hyperbolic Discounting*, Economics, MIT.

Mistri, Maurizio (2002), «Addiction and the Dynamic Inconsistency of Consumption Plans», *International Review of Economics and Business*, 49, 2, 175-92.

O'donoghue, T.; Rabin, M. (1999a), «Doing it Now or Later», *American Economic Review*, 89, 1, 103-24.

O'donoghue, T.; Rabin, M. (1999b), *Addiction and Self-control*, Manuscript, Department of Economics, Cornell University.

O'donoghue, T.; Rabin, M. (2001), Addiction and Present-Biased Preferences, Retrieved 10 January 2004 from http:/www.people.cornell.edu/pages/edo1/papers.html.

Orphanides, A.; Zervos, D. (1995), «Rational Addiction with Learning and Regret», *Journal of Political Economy*, 103, 739-58.

Orphanides, A.; Zervos, D. (1998), «Myopia and Addictive Behavior», *The Economic Journal*, 108, 75-91.

Phelps, E. S.; Pollack, Robert A. (1968), «On Second-Best National Saving and Game Equilibrium Growth», *Review of Economic Studies*, April, 35 (2), 185-99.

Pollak, R. A. (1968), «Consistent Planning», *Review of Economic Studies*, 35, 201-208.

Prelec, D. (1990), «Decreasing Impatience: Definition and Consequences», *Working Paper* nº 90-015, Harvard University.

Stigler, G. J.; Becker, G. S. (1977), «De Gustibus Non Est Disputandum», *American Economic Review*, 67, 2, 76-90.

Strotz, R. H. (1956), «Myopia and Inconsistency in Dynamic Utility Maximization», *Review of Economic Studies*, 23, 165-180.

Suranovic, S. M.; Goldfarb, R. S.; Leonard, T. C. (1999, «An economic theory of cigarette addiction», *Journal of Health Economics*, 18, 1, 1-29.

Vanini, Paolo; Braun, Norman (2002), On Habits and Addictions, Retrieved 20 October 2003 from http//ssrn.com/abstract=362720.

Impacts des accords de libre échange Euro-tunisien: évaluation par un Modèle d'Equilibre Général Calculable en 1996

Haykel Hadj Salem
Membre du G.A.I.N.S. – Université du Maine

resumo

O objectivo deste artigo é comparar os efeitos que o comércio livre total e o comércio livre industrial entre a Tunísia e a UE podem ter na economia tunisina. Este problema foi analisado utilizando um MEGC. O presente artigo divide-se em dois grupos de simulações. O primeiro diz respeito à liberalização total do comércio e o segundo apenas a uma liberalização parcial. Depois de efectuadas as simulações, observou-se que uma liberalização do comércio que vigore durante um período razoável é uma medida eficaz para un país em vias de desenvolvimento. Além disso, considera--se que essa liberalização é preferível se for aplicada a um único produto ou a uma determinada categoria de produtos. Por outras palavras, o comércio livre progressivo – e industrial – entre a Tunísia e a UE pode ser uma boa estratégia para a criação de uma zona euro-mediterrânica de comércio livre até ao ano 2010.

résumé / abstract

L'objet de ce papier est de comparer les effets du libre échange total et les effets du libre échange industriel entre la Tunisie et l'UE sur l'économie tunisienne. Cette problématique a été évaluée à l'aide d'un MEGC. Ce papier va être divisé en deux blocs de simulations. Le premier concerne la libéralisation commerciale totale, alors que le deuxième consiste à une libéralisation commerciale partielle. Après avoir effectué ces simulations, nous remarquons qu'une libéralisation commerciale appliquée pendant une période raisonnable demeure une réforme économique efficace pour un PVD. De plus, une telle libéralisation est considérée comme préférable lorsqu'elle est appliquée à un seul ou à une certaine catégorie de produits. Autrement dit le libre échange progressif – industriel – entre la Tunisie et l'UE est considéré comme une bonne stratégie pour la constitution d'une zone de libre échange euro-méditerranéenne à l'horizon 2010.

The aim of this paper is to compare the effects of total free trade and industrial free trade between Tunisia and the EU on the Tunisian economy. The analysis of this problem uses a computable general equilibrium model (CGEM), and this paper has two simulation sections. The first concerns total trade liberalisation, and the second a partial liberalisation. After performing the simulations, the results show that a trade liberalisation during a reasonable period is an efficient policy for a developing country. Besides, the author suggests that this liberalisation will work better if applied only to a single product or a specific category of products. In other words, a gradual free industrial trade between Tunisia and the EU might be a good strategy for the creation of a free trade area up to 2010.

1. Introduction

Ces dernières années, la plupart des pays en développement se préparent à l'accès à la scène internationale qui regroupe les concurrents traditionnels et les nouveaux concurrents. Face à ce nouveau phénomène, les pays concernés ont pris des précautions au niveau interne sous forme de réformes internes (des politiques d'accompagnement, d'ajustement (plan d'ajustement structurel) et de restructuration (mise à niveau)) sous l'égide des institutions internationales pour faciliter l'application des réformes extérieures. Ces dernières réformes ont été l'objet de préoccupation de certains économistes dont le but est d'évaluer leurs impacts dans les économies en développement. Dans ce contexte, nous allons nous intéresser à l'évaluation des impacts de deux types de libéralisations commerciales entre la Tunisie et l'Union Européenne sur l'économie tunisienne. Cette comparaison empirique est effectuée à travers un modèle d'équilibre général calculable (MEGC) appliqué à la Tunisie (voir Annexe 1). Ce document illustre les résultats trouvés à partir du logiciel GAMS et il est réparti en deux paragraphes sous forme de deux blocs de simulations:

1er Bloc: Libéralisation commerciale totale (pour tous les produits[1], pour tous les partenaires).

2ème Bloc: Libéralisation commerciale partielle (seulement pour les produits industriels, pour tous les partenaires).

Le choix du 1er Bloc consiste à donner les avantages et les inconvénients d'une libéralisation commerciale qui concerne tous les produits, tous partenaires confondus.

Ce bloc renferme quatre scénarios qui mettent en évidence une baisse tarifaire progressive jusqu'à une élimination totale. Ces simulations ne correspondent pas, en réalité, à des politiques actuelles prises par la Tunisie. Mais, elles sont choisies pour les comparer à des politiques commerciales de la Tunisie dans le cadre de partenariat euro-tunisien. Ces dernières consistent à une baisse tarifaire qui touche en premier lieu les produits industriels suivant un calendrier fixé par les deux parties (à savoir la Tunisie et l'UE). Ce qui signifie que le 2ème bloc reflète d'une manière approximative le contenu de l'accord d'association euro-tunisien. Cet accord a pour pièce maîtresse l'établissement progressif d'une zone de libre échange durant une période de 12 ans. Ce libre échange concerne exclusivement les produits industriels et plus particulièrement les produits manufacturiers. Alors que les produits agricoles, de pêche et les services sont exclus dès l'entrée en vigueur de l'accord (ces produits vont être réexaminés dans les années qui viennent). Pour cette raison, la libéralisation commerciale partielle se focalise uniquement sur tous les produits industriels. Nous répartissons notre deuxième bloc en quatre scénarios: le premier consiste à une baisse de 1/12 du taux de droits de douane à l'importation des produits industriels (c'est-à-dire que la baisse tarifaire est effectuée chaque année durant une période de 12 ans). Le deuxième scénario consiste à une baisse de 1/6 du même taux (ce qui signifie que cette baisse est effectuée chaque année durant une période de 6 ans). Le troisième scénario renferme la baisse de 50% du taux de droits de douane. En d'autres termes, nous supposons que la période est de deux ans pour effectuer cette libéralisation commerciale. Enfin, nous éliminons le taux de droits de douane seulement à l'importation des produits industriels d'où l'objet du 4ème scénario. Nous avons élaboré ces quatre simulations pour montrer l'avantage et l'inconvénient du choix de la période de 12 ans par les deux parties. Le choix de deux blocs est effectué pour essayer de répondre à certaines questions posées par certains économistes, par exemple:

quel est l'impact de l'accord euro-tunisien sur l'économie tunisienne, lorsque la libéralisation commerciale concerne la totalité des produits au lieu de concerner seulement les produits industriels?.

---

1 s1: produits agricoles et de pêche; s2: produits industriels; s3: services marchands et s4: services non marchands.

quelle est la durée de la libéralisation commerciale qui est considérée comme satisfaisante et adéquate pour arriver à l'achèvement des réformes attendues par la Tunisie? (que ce soit une libéralisation totale ou partielle).

Quels sont les impacts d'une libéralisation progressive d'une part et d'autre part d'une libéralisation immédiate (c'est-à-dire une suppression des droits de douanes) dans le cadre d'un partenariat euro-tunisien?.

Après avoir répondu à ces trois questions, nous exposons des conclusions qui font l'objet du dernier paragraphe.

2. La libéralisation commerciale totale

Ce bloc renferme quatre simulations:

simulation 1: baisse tarifaire de 1/12
simulation 2: baisse tarifaire de 1/6
simulation 3: baisse tarifaire de 50%
simulation 4: éliminer les droits de douanes

Les résultats de ce blocs sont donnés en détail dans le tableau 1 (Annexe 2).

La simulation 1 (sim1) consiste en un abaissement de 1/12 du taux de droit de douane à l'importation de tous les produits. En conséquent de cet abaissement les recettes tarifaires de l'Etat (tmrev) en valeur baissent de 1328MD à 1225MD, soit 7,76%. La baisse du taux de droit de douane entraîne d'une part une baisse des prix intérieurs des produits industriels et agricoles importés (pm(s1), pm(s2)), et, d'autre part une faible augmentation des prix des services (0,2%). Nous remarquons, à partir du tableau 1, que cette baisse est plus élevée pour les prix des produits agricoles (1,68%) que pour ceux des produits industriels (0,74%). Ces prix gardent les mêmes évolutions dans les trois autres simulations (c'est-à-dire une baisse pour les prix des produits agricoles et industriels et une augmentation des services marchands).

Cette baisse tarifaire (sim1) entraîne l'accroissement attendu des importations des deux produits (agricoles et industriels de 2,21% et de 0,56% respectivement), qui se répercute sur le volume total des achats à l'étranger qui augmente (0,61%). En outre, nous pouvons remarquer que cet accroissement, des importations des produits agricoles ainsi que des produits industriels, persiste lorsqu'il s'agit d'une courte durée (6 ans (sim2), 2 ans (sim3)) ou d'une suppression des droits de douanes (sim4). Mais chaque fois que la durée de la période devient courte, nous observons qu'il y a une augmentation rapide des importations. Ceci est confirmé seulement pour les secteurs primaire et secondaire. Alors que dans le cas des services, nous observons une baisse des importations dans les trois dernières simulations, en terme de volume, cela est dû à l'augmentation de leur prix du marché.

Dans notre MEGC, nous maintenons fixe le déficit courant extérieur. Autrement dit, toute évolution des importations tunisiennes ne peut être financée que par une évolution des exportations sachant que les revenus du capital du Reste Du Monde (RDM) sont nuls. Dans nos quatre simulations, il s'agit d'un accroissement des importations des produits agricoles et industriels. Pour le cas présent les exportations des produits industriels ainsi que les produits agricoles croissent, respectivement, de 1,03% et de 0,78%. Nous constatons qu'il y a une amélioration de la balance commerciale des produits industriels (il s'agit d'une baisse du déficit) mais une aggravation du déficit commercial pour les produits agricoles (même pour les trois autres simulations). L'amélioration du déficit commercial concerne aussi le secteur tertiaire et elle est vérifiée dans les quatre simulations.

L'indice général des prix (le prix pondéré de la valeur ajoutée, *pindex*) diminue dans les quatre simulations (sim1: de 0,4%, sim2: de 0,9%, sim3: de 2,7% et sim4: de 5,6%). Ainsi chaque fois qu'il s'agit d'une courte période, il diminue de plus en plus. Cela entraîne une augmentation de plus en plus forte du taux de change réel ($e_r = (e_n / pindex)$) de 1 à 1,059 (en sim4) [1,004 (en sim1); 1,009 (en sim2); 1,028 (en sim3)].

Dans ces quatre simulations, nous concluons que la dépréciation réelle du Dinar tunisien (DT) (de 0,4% en sim1; de 0,9% en sim2; de 2,8% en sim3 et de 5,9% en sim4) est suffisamment incitative pour que la production tunisienne à destination de l'UE (son partenaire principal) atteigne le niveau requis par les besoins accrus de financement des importations.

L'augmentation des exportations tunisiennes s'effectue au détriment d'une baisse des ventes sur le marché local tunisien. Nous pouvons remarquer ce phénomène essentiellement au niveau des produits industriels, dans toutes les simulations, alors que nous le remarquons seulement à partir de la deuxième pour le cas des produits agricoles. Cela explique le fait que certains pays membres de l'UE, concurrents directs de la Tunisie au niveau agricole, profitent de l'avantage d'une PAC (Politique Agricole Commune), et que lors d'une importante baisse des droits de douanes, les exportations agricoles tunisiennes augmentent au sein du marché européen. Cela signifie que l'application d'une libéralisation totale, durant une courte période, induit à une augmentation des exportations agricoles, mais non pas à une baisse de déficit commercial agricole. Une libéralisation commerciale spécifique à chaque secteur paraît donc plus bénéfique que celle qui concerne tous les secteurs au sein d'une petite économie.

D'après la simulation 2, nous remarquons que les exportations des produits agricoles industriels et des services augmentent, respectivement de 2,06%, de 1,58% et de 0,43%; alors que les ventes sur le marché tunisien de ces trois produits diminuent de 0,09%, de 0,84% et de 0,02%. Cette relation est détectée à partir de la fonction de la demande intérieure pour un produit donné de type CET (à élasticité de transformation constante). En d'autres termes, la transformation de l'offre intérieure en une offre à l'exportation est due à une baisse du prix reçu par les producteurs sur le marché tunisien (pl(s_i), respectivement de 0,7%, de 0,6% et de 0,4% (voir sim1), avec i=1,2,3). Ce prix baisse aussi en termes relatifs, puisque le prix payé à l'exportation des trois produits est maintenu fixe (pe (s_i) =1, avec i=1,2,3).

Dans les trois dernières simulations, nous constatons que le processus de transformation des ventes locales en ventes à l'exportation était plus important dans le cas de l'industrie que dans les deux cas de l'agriculture et des services, alors que la diminution du prix local de la production industrielle était plus faible que le prix local de la production des autres secteurs. Cela s'explique essentiellement par la valeur de l'élasticité de transformation commerciale de l'industrie (qui est égale à 2), supérieure d'une part à celle de l'agriculture (égale à 1,2) et d'autre part à celle des services (égale à 0,5).

D'une manière analogue aux trois dernières simulations, nous constatons dans la première que la réorientation des trois produits vers le marché extérieur (en particulier vers le marché européen) est aussi accompagnée par une baisse de la production totale dans les secteurs agricoles et industriels et par une augmentation de la production totale dans le secteur tertiaire. Nous nous focalisons, maintenant sur la baisse de l'output agricole et industriel. Elle est moins importante lorsque la période de l'application de la libéralisation commerciale est courte (par exemple: la sim1). Nous pouvons expliquer cette baisse aussi à partir du résultat de la demande intérieure. Nous rappelons que cette demande est subdivisée en trois composantes: la consommation des ménages, l'investissement et la demande intermédiaire. D'abord, nous remarquons que la consommation des ménages n'est pas responsable d'une baisse de la demande intérieure, puisqu'elle croît en volume pour les produits agricoles et industriels à concurrence, respectivement de 0,05% et de 0,04% dans la sim1. Cette augmentation est remarquée chaque fois qu'il s'agit d'une courte période (par exemple une augmentation de 1,02% et de 0,06%, respectivement, dans la sim2). Par contre, dans le cas du secteur tertiaire, nous constatons une diminution qui se creuse pendant une courte période (sim2, sim3 et sim4). Ensuite, l'investissement total (tinv) n'arrête pas de baisser en valeur (sim1: de 0,38%, sim2: de 0,75%, sim3: de 2,28% et sim4: de 4,49%). Cette baisse est causée principalement par la réduction des recettes totales du gouvernement provenant des taxes à l'importation, qui a été de 7,76% (en sim1), et non pas par l'épargne publique. En effet, cette dernière a augmenté de 4,6% (en sim1). De plus, nous rappelons que dans notre MEGC, nous avons fixé l'épargne étrangère. Ceci explique, tout simplement que l'équilibre entre l'investissement et l'épargne est maintenu par une diminution au niveau de l'investissement, des épargnes des ménages et des entreprises et par l'augmentation de l'épargne publique.

Nous rappelons que l'investissement est constitué de trois produits (agricole, industriel et services), dans notre cas ceci explique qu'une forte baisse des ventes de produits agricoles et industriels dans le marché tunisien affecte directement une diminution de leur production. Cette dernière, à son tour, provoque la baisse de la demande intermédiaire, dans le cas des produits agricoles et industriels (en volume et en valeur) alors que celle des services augmente en terme de volume mais diminue en terme de valeur.

Nous allons nous intéresser à présent au processus de la baisse de la production au niveau des deux premiers secteurs et au processus de l'augmentation de la production au niveau du secteur des services marchands.

Dans le cas du secteur agricole, la baisse du prix intérieur du produit agricole importé (pm(*s1*)) est plus importante en valeur absolue que l'augmentation du prix du marché du produit local vendu sur le marché intérieur (pd(*s1*)). Cela provoque, sans aucun doute une substitution du produit local par le produit importé chez les consommateurs de produits agricoles. Cette substitution commerciale favorise les importations qui ont pour conséquence une chute de l'investissement agricole.

En ce qui concerne les produits industriels, nous constatons aussi une baisse du prix intérieur du produit importé (pm(*s2*)) plus important en valeur absolue que celle du prix du marché du produit local vendu sur le marché intérieur (pd(*s2*)). Cela a aussi provoqué une incitation chez les consommateurs locaux à s'approvisionner plus avec les produits importés. Ceci accélère alors la chute de l'investissement industriel. Alors que pour les secteurs des services marchands, nous remarquons qu'il s'agit d'un cas différent de ceux des deux autres secteurs. D'abord, ce secteur accueille une partie de la main d'œuvre libérée par les deux autres secteurs (en sim1: augmentation de la demande de travail par l'activité "*s3*" en volume est de 0,23%). Cet accueil est constaté aussi dans les trois autres simulations. Ensuite, l'investissement au sein des services marchands reste inchangé en volume dans les quatre simulations. Donc ce secteur n'a pas d'influence dans la chute de l'investissement en volume mais il a une faible influence en terme de valeur comparé au deux autres secteurs. Enfin, le prix intérieur des services importés (pm(*s3*)) augmente de 0,2% contrairement aux prix intérieurs des deux autres produits. Cette augmentation devient de plus en plus importante lorsqu'il s'agit d'une courte période concernant l'application de la libéralisation commerciale (0,3% dans la sim 2; 0,89% dans la sim3 et 1.98% dans la sim4). De plus, nous remarquons qu'il y a une baisse importante du prix du marché des services locaux vendus sur le marché intérieur (pd(s3)) à concurrence de 0,3% (en sim1). Ainsi ceci favorise la production locale des services marchands au détriment des services importés. Cela explique l'amélioration de ce secteur au niveau local et la satisfaction de ce type de services chez les utilisateurs locaux et les utilisateurs étrangers (à savoir les pays voisins (La Libye et l'Algérie)) bien que ce secteur soit plus développé au sein de l'UE. Donc il ne faut pas être surpris de l'augmentation de la production des services marchands qui n'est pas destinée au marché intérieur (puisqu'il y a une faible diminution de la demande locale pour les services) mais destinée au marché extérieur, puisqu'il y a un accroissement des exportations de ces services (E(s3)) de 0,24% (en sim1). Cet accroissement est plus important lorsqu'il s'agit d'une courte période (d'après sim2, sim3 et sim4).

Nous pouvons interpréter aussi la production de chacun de ces trois secteurs à partir des prix de la valeur ajoutée dans chaque activité. Decaluwé B., Martens A. et Savard L. affirment que les variations de ce prix expriment les changements dans le degré de protection effective qui est mesuré par le taux de protection effective (*TPE* mesure la variation relative de la valeur ajoutée intérieure de la branche *j* (va_j) et la valeur ajoutée internationale (vai_j))[2]:

2 (Decaluwé et al., 2001).

$$TPE_j = \frac{va_j - vai_j}{vai_j} * 100$$

Avec:

vai_j: représente la valeur ajoutée de la situation de référence puisque les prix internationaux sont égaux à 1.

A partir du tableau 1 (annexe 2), nous voyons que le prix de la valeur ajoutée du secteur agricole (pva(s1)) baisse de 0,70%, celui du secteur industriel de 0,50%, alors que celui des services marchands ne diminue que de 0,30% (dans la sim1). Ceci explique bien le dernier secteur qui reste relativement le moins déprotégé (puisque $-100 < TPE_{s1} < TPE_{s2} < TPE_{s3} < 0$)[3], d'où le fait qu'il attire la main d'œuvre laissée par les deux autres (à savoir agricole et industriel) ce qui explique évidemment l'augmentation de sa production.

Dans le cas des services non-marchands, nous rappelons que la valeur de leur production est, par définition, égale à la consommation publique. Dans notre MEGC appliqué à la Tunisie, nous avons fixé le volume des services offerts par l'Etat (x(*s4*) = 3420.9MD). C'est-à-dire que la baisse du prix à la production de ces services (px(*s4*)) entraîne automatiquement une baisse de la valeur de leur production (*g*). Cette baisse est constatée dans les quatre simulations. De plus, nous remarquons que chaque fois que la période est courte, le prix à la production baisse de plus en plus, et provoque une diminution toujours plus importante de la consommation publique en terme de valeur des services non-marchands. La production de ces services ne demande que de la main d'œuvre, cette demande en volume demeure constante dans les quatre scénarios. Cela est dû à la fixité de l'offre de ces services.

Dans le cadre des rémunérations des facteurs de production, nous constatons que la baisse des prix du marché de la production locale des trois produits (les produits agricoles, industriels et les services marchands) diminue la productivité marginale du travail en terme de valeur (c'est-à-dire le taux de salaire). Au niveau de l'économie tunisienne, nous remarquons qu'il y a de moins en moins de grandes intensités d'utilisation de main d'œuvre, ce qui provoque une diminution du salaire (*w*) de 0,4% (en sim1). Cette diminution du salaire est faible par rapport à la diminution constatée pour le rendement du capital agricole (r(*s1*)) et industriel (r(*s2*)) (respectivement de 0,70%, de 0,50% en sim1) alors qu'elle est élevée par rapport à celle constatée pour le rendement du capital des services marchands (de 0,30% en sim1). Ces résultats ont augmenté dans les trois autres simulations, tout en gardant les mêmes relations (voir sim2, sim3 et sim4). Puisque dans notre modèle, le capital est spécifique à chaque secteur il ne peut donc pas être réalloué (d'après Decaluwé B., Martens A. et Savard L.). Nous ajoutons que ces diminutions des coûts de production des facteurs primaires expliquent nécessairement la diminution quasi égale du revenu des ménages et leurs épargnes de 0,31% (sim1). Cette quasi-égalité est presque vérifiée pour les trois autres simulations. Compte tenu de la baisse des prix des deux produits composites (pq(*s1*) et pq(*s2*)), nous remarquons qu'il y a une augmentation en volume de la consommation des ménages, seulement pour les produits agricoles et industriels. D'où une réorientation des achats de ménages vers ces deux produits dont le prix composite diminue plus que celui des services marchands.

La plupart des économistes suggèrent des réformes et des politiques économiques, pour les pays en développement, qui durent pendant une période moyennement longue. Leur objectif consiste à préparer le terrain pour la résistance des petites économies en développement à ces chocs exogènes et pour pouvoir s'adapter aux conséquences de ces réformes à court terme. Cela peut être confirmé à partir de la comparaison des variations équivalentes (VE) et des variations compensatoires (VC) entre les quatre simulations.

Nous constatons à partir du tableau ci- dessous qu'il y a une augmentation du bien être seulement dans la simulation 1 et 2 alors qu'il y a diminution du bien être dans les deux dernières simulations.

---

3 taux de protection effective: TPE(s1) = -0,74%, TPE(s2) = -0,54%, TPE(s3) = -0,25%.

Tableau 1

	simulation 1		simulation 2		simulation 3		simulation 4	
	VC 1	VE 1	VC 2	VE 2	VC 3	VE 3	VC 4	VE 4
Ménages	-1,332	1,336	-1,819	1,831	5,386	-5,492	46,483	-48,4
Interprétations	nous avons VC < 0 et VE > 0: il y a une augmentation de bien-être		nous avons VC < 0 et VE > 0: il y a une augmentation de bien-être		nous avons VC > 0 et VE < 0: il y a une diminution de bien-être		nous avons VC > 0 et VE < 0: il y a une diminution de bien-être	

* Calculés par l'auteur

Ces derniers résultats confirment la préférence de l'application d'une libéralisation commerciale pendant une période raisonnable et non pas pendant une courte période (sim3) ou la suppression immédiate des droits de douanes.

A partir de ces quatre simulations qui forme le 1er bloc, nous constatons que la libéralisation commerciale touchant tous les produits apporte certains inconvénients. Les points faibles de ce type de libéralisation consistent à la divergence des résultats (comme le cas se présente d'une part du côté des produits agricoles et industriels et d'autre part du côté des services marchands), autrement dit cette libéralisation favorise certains produits parmi d'autres. Cependant nous avons appliqué les mêmes chocs pour l'ensemble des produits. Et si nous analysons en détail le cas particulier du secteur des services marchands, nous observons qu'il y a une dissociation entre la réalité et la théorie. En ce qui concerne la diminution du prix sur le marché du produit composite concerné (pq(*s3*)), elle entraîne automatiquement l'augmentation de la consommation de ces services de la part des ménages (ch (*s3*)), or ce n'est pas le cas dans nos résultats. Nous pouvons constater que lorsqu'il s'agit des réformes générales qui concernent l'ensemble des secteurs économiques, il est difficile de dissocier les effets dans chaque secteur et surtout le fait qu'il s'agisse d'une étude dans un modèle d'équilibre général calculable dont les secteurs sont interdépendants. Pour cette raison, certains économistes optent pour des réformes touchant un secteur ou une catégorie de secteur. C'est de cette façon qu'ils peuvent réussir à analyser facilement ces chocs sur l'ensemble des secteurs. Nous allons consacrer notre prochain paragraphe à ce contexte sous forme d'un bloc.

3. La libéralisation commerciale partielle

Ce bloc consiste à la libéralisation commerciale seulement pour les produits industriels et il renferme aussi quatre simulations:

simulation 5 (sim5): baisse tarifaire de 1/12
simulation 6 (sim6): baisse tarifaire de 1/6
simulation 7 (sim7): baisse tarifaire de 50%
simulation 8 (sim8): supprimer les droits de douanes

Les résultats de ce bloc sont présentés en détail dans le tableau 2 (Annexe 3).

Dans la simulation 5 (sim5), le choc extérieur consiste à la baisse du taux de droit de douane de 1/12 seulement pour les produits industriels. Ces derniers sont les premiers produits qui ont été concernés par le calendrier du démantèlement fixé par l'accord d'association euro-tunisien. Premièrement il s'agit du démantèlement total dès l'entrée en vigueur de l'accord des biens d'équipement non concurrentiels, deuxièmement, il traite le démantèlement sur cinq ans (c'est-à--dire de raison de 1/5 par an) des matières premières et des produits semi-finis non concurrentiels, troisièmement, il correspond au démantèlement des produits concurrentiels (fabriqués localement) et plus compétitifs que les produits européens sur une période de 12 ans à raison de 1/12 par an, et quatrièmement il s'agit du démantèlement des produits restants (fabriqués localement mais non compétitifs) pendant une période de 8 ans à raison de 1/8 par an. De plus, les deux signataires (Tunisie, UE) ont attribué un délai de grâce de quatre ans pour ces derniers types de produits dès l'entrée en vigueur de l'accord dont l'objet consiste à mettre à niveau les entreprises tunisiennes qui les fabriquent. Dans notre étude, nous nous sommes focalisés sur le produit industriel agrégé qui regroupe les produits industriels concurrentiels représentant une bonne partie des importations tunisiennes.

L'abaissement de 1/12 du taux de droit de douane à l'importation des produits industriels a baissé, en valeur, les recettes tarifaires de l'Etat (tmrev) de 1328 MD à 1231 MD soit de 7,3 %. Cette baisse du taux a entraîné d'une part une baisse des prix intérieurs (du marché) des produits agricoles et industriels et d'autre part une augmentation des prix intérieurs des services marchands. Mais, nous remarquons que la baisse des prix est plus importante pour les produits industriels (de 0,74% dans sim5, de 1,47% dans sim6) que pour les produits agricoles (0,08% dans sim6 et presque stable dans sim5). Dans les autres simulations (sim7 et sim8), nous observons les mêmes signes de variations. Par exemple, du fait qu'il s'agisse d'une courte période pour l'application de la libéralisation nous assistons à une baisse importante, surtout pour les produits concernés, à savoir les produits industriels (sim7: 4,49%, sim8: 8,82%).

Dans la simulation 5, ainsi que dans les autres simulations, nous remarquons que cette libéralisation a entraîné un accroissement attendu des importations des produits industriels (sim5: 0,61%, sim6: 1,23%, sim7: 3,78% et sim8: 7,91%), tandis que pour les produits agricoles et les services marchands, nous observons une baisse de plus en plus importante durant une période d'application de plus en plus longue (voir les quatre simulations). Ceci explique l'augmentation des volumes des achats à l'étranger qui se focalisent essentiellement sur l'augmentation des volumes des achats des produits industriels. A partir de ces résultats, nous déduisons que la libéralisation commerciale favorise l'importation des produits industriels au détriment des autres produits. Et ceci est vérifié pendant chaque durée de période (courte ou longue).

Dans notre MEGC appliqué à la Tunisie, nous supposons que le déficit courant extérieur est exogène. Ainsi, toute variation des importations de la Tunisie ne peut être financée que par une variation de ses exportations puisque les revenus du capital au RDM sont négligeables, voir nuls. Nous remarquons que les importations et les exportations tunisiennes évoluent dans le même sens pour les quatre simulations. L'accroissement des exportations est remarqué surtout pour les produits industriels (soit de 0,67% en sim5) et pour les services marchands (soit de 0,24%). De plus, les exportations des produits agricoles augmentent doucement dans les quatre simulations.

Concernant la balance commerciale, nous avons un déficit au niveau des produits agricoles ainsi qu'au niveau des produits industriels, par contre, nous avons un excédent au niveau de la balance commerciale des services marchands. Au cours des quatre simulations, nous constatons qu'il y a une amélioration des soldes des balances commerciales des trois produits.

L'indice général des prix (*pindex*) diminue aux cours des quatre simulations (sim5: de 0,40%; sim6: de 0,80%; sim7: de 2,30% et sim8: de 4,90%). Cela engendre une augmentation de plus en plus importante du taux de change réel ($e_r = (e_n / pindex)$) de 1 à 1,0515 (en sim8) [1,004 (en sim5); 1,008 (en sim6); 1,0235 (en sim7)] sachant que les prix internationaux sont constants. De plus, nous ajoutons que la dépréciation réelle du Dinar Tunisien (de 0,40% en sim5, de 0,80% en sim6, de 2,35% en sim7 et de 5,15% en sim8) est suffisamment incitative pour que la production tunisienne à destination de l'étranger, et plus précisément vers l'UE, atteigne le niveau requis par les besoins accrus de financement des importations.

L'augmentation des exportations tunisiennes des produits industriels est effectuée au détriment d'une baisse des ventes des mêmes produits sur le marché local tunisien. Ce qui confirme la focalisation de la politique économique du gouvernement à être de plus en plus orientée vers le marché extérieur et mondial (en terme de concurrence prix- quantité). Ce phénomène peut être concrétisé du fait qu'il s'agisse d'une période d'application de la libéralisation de plus en plus courte (sim5, sim6, sim7 et sim8).

Au cours de ces quatre simulations, nous remarquons que l'output de l'activité industrielle diminue proportionnellement à la durée de la période. Par contre l'output des deux autres activités restent constantes et/ou augmentent doucement. Cela montre que pour relancer la production industrielle locale, il est préférable d'appliquer une libéralisation progressive et appliquée à un type de produit, comme le cas se présente pour le produit industriel.

Dans la simulation 5, comme dans les trois autres simulations, nous remarquons que les exportations des produits industriels augmentent de plus en plus, alors que les ventes des mêmes produits sur le marché tunisien diminuent. Cette relation est combinée dans la fonction de la demande intérieure du produit industriel de type CET. Cela signifie que la transformation de l'offre intérieure du produit industriel en une offre à l'exportation est due à une baisse du prix reçu par les producteurs sur le marché tunisien (respectivement de 0,60% (sim5); de 1,10% (sim6); de 3,40% (sim7) et de 6,90% (sim8)), tandis que le processus de transformation des ventes locales en ventes à l'exportation était moins important dans le cas des deux autres activités. C'est-à-dire que cette libéralisation favorise la réorientation seulement des produits industriels vers le marché extérieur et en particulier vers le marché européen. La demande intérieure, quant à elle, pour les produits agricoles et les services marchands augmente doucement lorsqu'il s'agit d'une période de plus en plus courte. Nous pouvons expliquer aussi la baisse de la production des produits industriels à partir de la demande intérieure du même produit qui se compose de la consommation des ménages, de l'investissement et de la demande intermédiaire. Tout d'abord, nous constatons que la consommation des ménages des produits industriels n'est pas la cause de la baisse de la demande intérieure, puisque la demande des ménages pour ce produit croît en volume à concurrence de 0,10% en sim5 (de 0,21% en sim6; de 0,54% en sim7 et de 0,70% en sim8). Nous assistons à une faible augmentation, voire une stagnation, concernant la demande des ménages des deux autres produits, ce qui explique en partie l'augmentation de leur demande intérieure. Puis, nous évoquons la baisse importante de l'investissement total en valeur qui est due en grande partie à la baisse de la demande d'investissement totale du produit industriel (en terme de valeur) et de même qu'à la baisse du prix du marché du produit industriel composite. Nous rappelons que ce bloc consiste en une baisse des droits de douanes pour les produits industriels ce qui implique une réduction des recettes totales de l'Etat qui proviennent des taxes à l'importation (de 7,30% en sim5; de 14,68% en sim6; de 45,03% en sim7 et de 93,52% en sim8). Cette perte tarifaire est compensée automatiquement par une augmentation des recettes totales provenant de la taxation indirecte reçue par l'Etat.

Dans notre MEGC, nous avons supposé que l'épargne étrangère était fixée. De ce fait, l'équilibre entre l'investissement et l'épargne est maintenu par une diminution du niveau de l'investissement, ainsi que du niveau des épargnes du ménage et des entreprises d'une part, et d'autre part par l'augmentation de l'épargne publique. La baisse de l'investissement total est expliquée essentiellement par la baisse en volume de la demande d'investissement total du produit agricole (de 0,29% en sim6), alors que nous remarquons une stabilité en volume de la demande d'investissement total du produit industriel (en sim5 et sim6), voire une faible augmentation dans les deux dernières simulations. Par contre, en terme de valeur, nous observons que les demandes d'investissement total des trois produits diminuent dans les quatre simulations. De plus, les demandes intermédiaires totales pour les trois produits diminuent aussi en terme de valeur. Cette diminution est plus remarquée pour les produits industriels que pour les produits agricoles. Ce sont seulement les demandes intermédiaires totales des produits industriels qui diminuent, en terme de volume essentiellement, ce qui explique bien la baisse de la production des produits industriels.

Maintenant, nous nous intéressons à interpréter la production sectorielle à partir des prix de la valeur ajoutée dans chaque activité. D'après la simulation 5, nous observons que le prix de la valeur ajoutée du secteur agricole baisse de 0,20%, celui du secteur industriel baisse de 0,50% et celui des services marchands baisse de 0,30%. Ceci explique bien que le secteur agricole reste le secteur le moins déprotégé (ou le plus protégé) (puisque $-100 < TPE_{s2} < TPE_{s3} < TPE_{s1} < 0$)[4] avec le secteur des services marchands. Mais nous remarquons que c'est seulement

4 Taux de protection effective: TPE(s1)= -0.2%, TPE(s2)=-0.6%, TPE(s3)=-0.24%.

ce dernier secteur qui attire la main d'œuvre laissée par le secteur industriel (vérifié pour les quatre simulations), alors que pour les deux autres secteurs (agricole et les services non marchands), nous assistons à une demande de travail stable au niveau de la simulation 5, mais à partir des autres simulations nous remarquons qu'elle s'améliore uniquement pour le secteur agricole. Tous ces résultats expliquent la baisse de la production dans le secteur industriel et la hausse de celle du secteur agricole et du secteur des services marchands.

Nous rappelons que la production des services non-marchands est égale à la consommation publique. Dans notre MEGC, nous avons supposé que le volume des services offerts par l'Etat est exogène (x(*s4*) = 3420.9 MD). C'est-à-dire que la baisse du prix à la production de ces services (de 0,40%(sim5)) est la cause unique de la baisse de la production en valeur (*g*). Cette diminution est constatée aussi dans les trois autres simulations (de 0,90% dans sim6; de 2,60% dans sim7 et de 5,40% dans sim8).

Nous remarquons, chaque fois que la période d'application de la libéralisation commerciale est longue, une baisse de plus en plus remarquée de la consommation publique. Ce résultat peut être expliqué par l'évolution du phénomène de la privatisation en Tunisie qui est considéré comme une réforme politique dans le but de diminuer la présence de l'Etat dans le marché, en d'autres termes elle favorise la libéralisation du marché. Il faut rappeler que dans notre modèle, la production des services non-marchands ne demande que de la main d'œuvre, la demande de travail en volume demeure donc constante dans les quatre simulations. Cela bien sûr est dû à la fixité de l'offre de ces services. D'après le tableau 2, nous remarquons que les revenus des facteurs de production diminuent dans les quatre simulations. De plus, les prix du marché de la production locale de nos trois produits diminuent et provoquent une baisse de la productivité marginale du travail en valeur (c'est-à-dire le salaire). D'après la sim5, la diminution du salaire est plus faible que la diminution du taux de rendement du capital du secteur industriel dans les quatre simulations. Par contre, elle est plus grande que la diminution des taux de rendement du capital des deux autres secteurs (s1 et s3).

La diminution des coûts de production des facteurs primaires explique nécessairement la diminution quasi-égale du revenu des ménages (de 0,25%) et de ses épargnes (de 0,22%) en sim5. Cette quasi-égalité est presque maintenue pour les autres simulations.

Dans la simulation 5, nous assistons à la diminution des prix du marché des trois produits composites (pq(s1), pq(s2) et pq(s3)). Cette diminution a provoqué l'augmentation en volume de la consommation du ménage des produits industriels et des services non marchands. Dans cette simulation, nous observons la réorientation de la consommation des ménages vers les produits industriels. Cela est dû à la libéralisation appliquée dans ce type de produit. Or la consommation des biens agricoles demeurent constante au niveau des trois premières simulations et elle diminue dans la dernière.

Tableau 2

	simulation 5		simulation 6		simulation 7		simulation 8	
	VC 5	VE 5	VC 6	VE 6	VC 7	VE 7	VC 8	VE 8
Ménages	-1,033	1,035	-1,337	1,344	5,225	-5,308	41,491	-42,848
Interprétations	nous avons VC < 0 et VE > 0: il y a une augmentation de bien-être		nous avons VC < 0 et VE > 0: il y a une augmentation de bien-être		nous avons VC > 0 et VE < 0: il y a une diminution de bien-être		nous avons VC > 0 et VE < 0: il y a une diminution de bien-être	

* Calculés par l'auteur

D'après ce tableau, nous remarquons qu'il y a une augmentation de bien être dans les deux premières simulations et une diminution de bien-être dans les deux dernières. Ces résultats confirment bien l'avantage de l'application d'une libéralisation commerciale des produits industriels pendant une période raisonnable, à savoir 12 ans, comme le cas se présente dans le cadre du calendrier du démantèlement tarifaire de l'accord euro-tunisien. Ceci peut être confirmé aussi à partir de la comparaison de la simulation n° 8 avec les deux premières simulations. Nous constatons que dans cette simulation, l'évolution des agrégats est presque similaire aux deux autres (sim6 et sim7), mais elle est plus remarquée lorsque nous supprimons immédiatement les droits de douane sur les produits industriels. Ces influences de ce type (positif ou négatif) peuvent se répercuter négativement sur la structure de l'économie tunisienne, puisque cette dernière est considérée comme une petite économie en développement.

4. Conclusions

En comparant les simulations du bloc 1 avec celles du bloc 2, nous constatons de grandes différences. D'abord une libéralisation commerciale appliquée pendant une période raisonnable (progressive) demeure une bonne réforme économique pour les pays en développement. Ensuite, une telle libéralisation est considérée comme préférable lorsqu'elle est appliquée à une certaine catégorie ou a un seul produit. Enfin, nous pensons que lorsqu'il s'agit d'une libéralisation concernant certains produits en provenance d'un seul partenaire ou d'un certain groupe de partenaires, elle est plus avantageuse que celle appliquée à tous les partenaires confondus. Mais dans notre modèle, nous avons considéré que le reste du monde représente l'Union Européenne puisque cette dernière représente le principal partenaire de la Tunisie. Cependant, une étude de désagrégation du RDM paraît aussi intéressante du fait que nous mettons en évidence la notion de détournement ou la création de l'échange d'un partenaire à un autre. C'est dans ce contexte que nous pouvons mettre en évidence l'orientation des échanges d'une économie soit vers un nouveau regroupement de pays soit vers un partenaire traditionnel.

Bibliographie

Abdelkhalek, T.; Dufour, J. M. (1998), «Satistical Inference for Computable General Equilibrium Model with Application to a Model of the Maroccan Economy», *The Review of Economics and Statistics*, LXXX, 520 534.

Brook, A. *et al.* (eds.) (1998), *GAMS a User's Guide*, Washington D.C., GAMS Development Corporation.

Cororaton, Caesar B. (2000), «Construction of the 1994 SAM Of the Philippines And the Some Basic Features of PCGEM-2», *Document de travail*, MIMAP, Juillet.

Cockburn, J. *et al.* (1998), «Les leçons du mariage entre les modèles d'équilibre général calculable et la nouvelle théorie du commerce international: application à la Tunisie», *L'Actualité Economique*, 74(3), 381 413.

de Melo, J.; Tarr, D. (1992), *A General Equilibriuim Analysis of US Foreign Trade Policy*, Massachusetts, Massachusetts Institute of Technology, The MIT Press.

Decaluwé, B. *et al.* (2001), *La politique économique du développement et les modèles d'équilibre général calculable*, Montréal, Les Presses de l'Université de Montréal.

Decalwué, B.; Souissi, M. (1996), *L'accord de libre échange entre la Tunisie et l'Union Européenne: le cas du secteur agricole*, Tunis, étude non publiée faite pour le ministère de la coopération internationale et l'investissement extérieur de la Tunisie.

Dessus, S.; Suwa, A. (2000), *Régional Integration and International Reforms in the Mediterranean Area*, Paris, OECD Develpment Center.

Hamdouch, B.; Chater, M. (2000), *Impact des accords de Libre échange Euro Méditerranéens: cas du Maroc*, Rapport Final, FEMISE.

Rutherford, T. F. *et al.* (1995), *The Free Trade Agreement between Tunisia and the European Union*, document de travail non numéroté, Washington D.C., Département de recherche sur les politiques, Banque Mondiale.

Annexe 1

La structure du Modèle d'Equilibre Général Calculable (MEGC): Cas de la Tunisie

• Bloc de la production:

$$v\alpha_{td} = kt\_v\alpha_{td} * [sh\_v\alpha_{td} * k_{td}^{-rh\_v\alpha_{td}} + (1 - sh\_v\alpha_{td}) * L_{td}^{-rh\_v\alpha_{td}}]^{\frac{-1}{rh\_v\alpha_{td}}}$$

Avec $rh\_v\alpha_{td} = \frac{(1 - sig\_v\alpha_{td})}{sig\_v\alpha_{td}}$

$$v\alpha_{ntd} = L_{ntd}$$

$$x_i = Min\left[kt\_in_i * v\alpha_i, \frac{int\ \rho_i}{kt\_int\ \rho_i}\right]$$

$$x_i = v\alpha_i * kt\_in_i$$

$$int\ \rho_i = kt\_int\ \rho_i * x_i$$

$$mat_{td,i} = \alpha_{td,i} * int\ \rho_i$$

$$int\ d_{td} = \sum_i mat_{td,i}$$

$$L_{td} = v\alpha_{td} * \left[\frac{\rho\, v\alpha_{td} * (1 - sh\_v\alpha_{td})}{w * kt\_v\alpha_{td}^{-rh\_v\alpha_{td}}}\right]^{\frac{1}{1 + rh\_v\alpha_{td}}}$$

$$L_{ntd} = \frac{(\rho x_{ntd} * x_{ntd} - \sum_{td} \rho d_{td} * mat_{td,ntd})}{w}$$

• Bloc revenus -épargne: des ménages et des entreprises:

$$yl_i = \sum_i w * L_i$$

$$yk = \sum_{td} r_{td} * K_{td}$$

$$yh = yl + lambda * yk + kt\_div * div * adj + trgov * pindex + yfor$$

$$dyh = yh * (1 - dtxrh)$$

$$savh = aps * dyh$$

$yf = (1 - lambda - lambda\_for) * yk * (1 - dtxrf)$

$savh = yf - div * adj - div\_for$

• Bloc recettes -épargnes de l'état:

$yg = tmrev + dtxrev + itxrev + trrow\_g$

$tmrev = \sum_{td} tm_{td} * m_{td}$

$dtxrev = (dtxrev * yh) + (1 - lambda - lambda\_for) * yk * dtxrf$

$itxrev = \sum_{td} [itxr_{td} * (1 + adj\_ntax) * d_{td} * pl_{td} + itxr_{td} * (1 + adj\_ntax) * m_{td} * pwm_{td} * er * (1 + tm_{td})]$

$savh = yg - g - trgov * pindex - trg\_row$

• Bloc de la demande finale intérieure:

$U = \prod_{td} ch_{td}^{kt\_ch_{td}}$

$$ch_{td} = \frac{kt\_ch_{td} * dyh}{pq_{td}}$$

$g = px_{“s4”} * x_{“s4”}$

$tinv = pinv * tinv\_r$

$$inv_{td} = \frac{kt\_inv_{td} * tinv}{pq_{td}}$$

• Bloc du commerce extérieur:

$$x_{td} = kt\_x_{td} * [sh\_x_{td} * e_{td}^{rh\_e_{td}} + (1 - sh\_x_{td}) * d_{td}^{rh\_e_{td}}]^{\frac{1}{rh\_e_{td}}}$$

Avec $rh\_e_{td} = \dfrac{(1 + sig\_e_{td})}{sig\_e_{td}}$

$$\frac{e_{td}}{d_{td}} = \left[\left(\frac{pe_{td}}{pl_{td}}\right) * \left(\frac{1 - sh\_x_{td}}{sh\_x_{td}}\right)\right]^{sig\_e_{td}}$$

$$q_{td} = kt\_q_{td} * [sh\_q_{td} * m_{td}^{-rh\_m_{td}} + (1 - sh\_q_{td}) * d_{td}^{-rh\_m_{td}}]^{-\frac{1}{rh\_m_{td}}}$$

Avec: $rh\_m_{td} = \frac{1 - sig\_m_{td}}{sig\_m_{td}}$

$$\frac{m_{td}}{d_{td}} = \left[\left(\frac{pd_{td}}{pm_{td}}\right) * \left(\frac{sh\_q_{td}}{1 - sh\_q_{td}}\right)\right]^{sig\_m_{td}}$$

• Bloc des prix:

$$pv\alpha_i = \frac{px_i * x_i - (\sum_{td} pq_{td} * mat_{td,i})}{v\alpha_i}$$

$$r_{td} = \frac{pv\alpha_{td} * v\alpha_{td} - w * L_{td}}{K_{td}}$$

$$pd_{td} = (1 + itxr_{td} * (1 + adj\_ntax)) * pl_{td}$$

$$pq_{td} = \frac{(pd_{td} * d_{td} + pm_{td} * m_{td})}{q_{td}}$$

$$px_{td} = \frac{(pl_{td} * d_{td} + pe_{td} * e_{td})}{x_{td}}$$

$$pe_{td} = er * pwe_{td}$$

$$pm_{td} = (1 + itxr_{td} * (1 + adj\_ntax)) * (1 + tm_{td}) * er * pwm_{td}$$

$$pinv = \prod_{td} \left(\frac{pq_{td}}{kt\_inv_{td}}\right)^{kt\_inv_{td}}$$

$$pindex = \sum_i w\_v\alpha_i * pv\alpha_i$$

• Bloc des conditions d'équilibre:

$$cab = \sum_{td} pwm_{td} * m_{td} * er + lambda\_for * yk + trf\_row + trg\_row - \sum_{td} pwe_{td} * er - \sum_h yfor_h * er$$

$$q_{gd} = ch_{gd} + inv_{gd} + int\ d_{gd}$$

$$tinv = savh + savf + savg + cab$$

$$LS = \sum_i L_i$$

La loi de Walras: l'équilibre sur le marché des services («s3»)

$$Leon = q_{«S3»} - \sum_h (ch_{«S3»,h}) - inv_{«S3»} - int\ d_{«S3»}$$

Tableau 1 – Le modèle d'équilibre général calculable de la Tunisie
Le nombre d'équations et de variables (endogènes et exogènes)

Equations				Variables			
				Variables		Type de variable	
						endogène	exogène
Noms dans GAMS	Numéro	Indice	Nombre	Noms dans GAMS	Indices	Nombres des variables	Nombres des variables
xeq	1	i	4	x	i	4	
vaeq1	2	td	3	va	i	4	
vaeq2	3	ntd	1	va	i	4	
intpeq	4	i	4	intp	i	4	
mateq	5	td,i	3*4	mat	td,i	3*4	
leq1	6	td	3	l	i	4	
leq2	7	ntd	1	l	i	4	
ceteq	8	td	3				
eeq	9	td	3	e	td	3	
qeq	10	td	3	q	td	3	
meq	11	td	3	m	td	3	
cheq	12	td	3	ch	td	3	
geq	13		1	g		1	
inveq	14	td	3	inv	td	3	
intdeq	15	td	3	intd	td	3	
tinv_req	16		1	tinv_r			1
yleq	17		1	yl		1	
ykeq	18		1	yk		1	
yheq	19		1	yh		1	
dyheq	20		1	dyh		1	
yfeq	21		1	yf		1	
ygeq	22		1	yg			1
tmreveq	23		1	tmrev		1	
dtxreveq	24		1	dtxrev		1	
itxreveq	25		1	itxrev		1	
savheq	26		1	savh		1	
savfeq	27		1	savf		1	
savgeq	28		1	savg		1	
pmeq	29	td	3	pm	td	3	
peeq	30	td	3	pe	td	3	
pqeq	31	td	3	pq	td	3	
pxeq	32	td	3	px	i	4	
pdeq	33	td	3	pd	td	3	
pvaeq	34	i	4	pva	i	4	
pindexeq	35		1	pindex		1	
pinveq	36		1	pinv		1	
req	37	td	3	r	td	3	
eq1eq	38	gd	2				
eq2eq	39		1	tinv		1	
eq3eq	40		1	cab			1
eq4eq	41		1				
walras	42		1				

La suite:							
Equations				Variables			
				Variables		Type de variable	
						endogène	exogène
Noms dans GAMS	Numéro	Índice	Nombre	Noms dans GAMS	Indices	Nombres des variables	Nombres des variables
				w		1	
				pl	td	3	
				adj		1	
				d	td	3	
				adj_ntax		1	
				pwe	td		3
				pwm	td		3
				grant_for*			1
				er			1
				x	"s4"		1
				paygv_for			1
				trgov			1
				yfor	h		1
				tm	td		3
				div_for			1
				k	td		3
				ls			1
				dtxrf			1
				dtxrh	h		1
				itxr	td		3
				div			1
Total			**93**			**93**	**28**

* «grant_for» est le nom utilisé en GAMS pour les transferts du RDM vers le gouvernement (trrow_g).

Les variables endogènes:

x_i : output de l'activité *i* (offre totale par produit).
U : fonction d'utilité de type Cobb-Douglas
va_i : valeur ajoutée de l'activité *i*
$intp_i$: consommation intermédiaire totale de l'activité *i*
$mat_{td,i}$: matrice input output
l_i : demande de travail par activité *i*
e_{td} : exportations du produit *td*
q_{td} : demande pour le produit composite *td*
m_{td} : importations du produit *td*
ch_{td} : consommation du ménage en produit *td*
g : consommations publiques
inv_{td} : demande d'investissement totale (FBCF+variations des stocks) par produit *td*
$intd_{td}$: demande intermédiaire pour le produit *td*
yl : revenu du facteur travail
yk : revenu du facteur capital
yh : revenu du ménage

dyh : revenu disponible du ménage
yf : revenu des entreprises
yg : revenu du gouvernement
tmrev : recettes totales provenant des taxes à l'importation
dtxrev : recettes totales du gouvernement provenant des impôts directs payés par les ménages (*TRH_G*) et les entreprises (TRF_G).
itxrev : recettes totales provenant de la taxation indirecte
savh : épargne du ménage
savf : épargne des entreprises
savg : épargne du gouvernement
pm_{td} : prix intérieur (du marché) du produit importé *td*
pe_{td} : prix payé à l'exportateur du produit *td*
pq_{td} : prix du marché du produit *td*
px_i : prix à la production du produit *i*
pd_{td} : prix du marché du produit local *td* vendu sur le marché intérieur
pva_i : prix de la valeur ajoutée dans l'activité *i*
pindex : prix pondéré de la valeur ajoutée
pinv : prix de l'investissement total
r_{td} : taux de rendement du capital dans l'activité td
tinv : investissement total
w : taux de salaire moyen
pl_{td} : prix au producteur du produit *td* pour la vente sur le marché intérieur.
adj : facteur d'ajustement
d_{td} : demande pour le produit intérieur td
adj_ntax : ajustement de l'impôt indirect

Les variables exogènes:

tinv_r : investissement total réel
yg : revenu publique
cab : épargne du RDM (solde de la balance courante)
pwm_{td} : prix mondial à l'importation du produit *td*.
pwe_{td} : prix mondial à l'exportation du produit *td*
grant_for : transferts du RDM vers l'Etat
er : taux de change
paygv_for : transferts de l'Etat vers le RDM
trgov : transferts de l'Etat vers le ménage
yfor : transferts du RDM vers le ménage
tm_{td} : taux tarifaire sur le produit *td*
div_for : transferts des entreprises vers le RDM
div : transferts des entreprises vers les ménages
k_{td} : demande de capital par activité *td*
ls : offre exogène de travail
dtxrf : impôts directs payés par les entreprises
dtxrh : impôts directs payés par le ménage
$itxr_{td}$: impôts indirects par produit *td*
x_{s4} : output du produit «*s4*» (services non marchands)

Les paramètres:

1°) fonction de production:

kt_in_i : coefficient technique (fonction Leontief)
kt_intp_i : coefficient technique (fonction Leontief)
kt_va_{td} : coefficient d'échelle (fonction CES) de la valeur ajoutée
sh_va_{td} : part du capital dans la valeur ajoutée de l'activité *td*
$1\text{-}sh\_va_{td}$: part du travail dans la valeur ajoutée de l'activité *td*
w_va_i : part de la valeur ajoutée de l'activité *i* par rapport à la valeur ajoutée totale
rh_va_{td} : paramètre de substitution dans la valeur ajoutée (type CES)
sig_va_{td} : élasticité de substitution dans la valeur ajoutée (type CES)
$a_{td,i}$: coefficients entrées – sorties (volume de l'intrant intermédiaire *td* par unité de demande intermédiaire de l'activité *i*)

2°) autres paramètres:

lambda_for : part de la rémunération du capital versée au RDM (dans notre cas égal à zéro)
lambda : part de la rémunération du capital versée au ménage.
1-lambda : part de la rémunération du capital versée aux entreprises.
aps : propension marginale à épargner (égale ici à la propension moyenne) du ménage
kt_ch_{td} : part de la consommation du produit *td* par rapport au revenu disponible du ménage
kt_inv_{td} : part de l'investissement du produit *td* dans l'investissement total
kt_div : paramètre de distribution de revenu de dividende du ménage
kt_yhk : paramètre de distribution de revenu du capital du ménage

***3°) fonction à élasticité de transformation constante (CET): Concernant l'offre totale par produit* td**

kt_x_{td} : coefficient d'échelle (constante).
sh_x_{td} : paramètre distributif.
rh_e_{td} : paramètre de transformation.
sig_e_{td} : élasticité de transformation.

***4°) fonction à élasticité de substitution constante (CES):Concernant la demande intérieure pour le produit* td**

kt_q_{td} : coefficient d'échelle (constante).
sh_q_{td} : paramètre distributif
rh_m_{td} : paramètre de substitution
sig_m_{td} : élasticité de substitution

Annexe 2

Tableau 1 – Les statistiques de base et les différentes simulations (en millions de DT et en %): ***Les baisses tarifaires concernant tous les produits importés*** *("s1", "s2" et "s3")*

Les noms des variables	Les noms in GAMS	référence	simulation 1		simulation 2		simulation 3		simulation 4	
		valeur	volume	valeur	volume	valeur	volume	valeur	volume	valeur
facteur d'ajustement	*adj*	1		0,0560		0,1110		0,3160		0,5700
épargne du ménage	*savh*	2241		-0,0031		-0,0062		-0,0196		-0,0433
épargne du gouvernement	*savg*	348		0,0460		0,0891		0,2759		0,5690
épargne des entreprises	*savf*	1652		-0,0163		-0,0321		-0,0962		-0,1901
ajustement de l'impôt indirect	*adj_ntax*	0								
recettes totales provenant de la taxation indirecte	*itxrev*	1072		0,1007		0,2024		0,6250		1,3069
recettes totales du gouvernement provenant des impôts directs payés par les ménages et les entreprises	*dtxrev*	1586		-0,0038		-0,0069		-0,0214		-0,0467
recettes totales provenant des taxes à l'importation	*tmrev*	1328		-0,0776		-0,1559		-0,4789		-1,0000
revenu disponible du ménage	*dyh*	13134		-0,0030		-0,0062		-0,0198		-0,0432
revenu des entreprises	*yf*	3193		-0,0044		-0,0088		-0,0272		-0,0573
revenu du ménage	*yh*	14381		-0,0031		-0,0063		-0,0199		-0,0433
revenu du facteur capital	*yk*	9672		-0,0043		-0,0088		-0,0271		-0,0572
revenu du facteur travail	*yl*	698		-0,0043		-0,0084		-0,0262		-0,0548
importations du produit "s1"	*M s1*	452,138	0,0221	0,0049	0,0414	0,0056	0,1381	0,0215	0,3094	0,0410
importations du produit "s2"	*M s2*	8970,7	0,0056	-0,0018	0,0113	-0,0035	0,0349	-0,0090	0,0725	-0,0169
importations du produit "s3"	*M s3*	366,344	0,0000	0,0020	-0,0028	0,0002	-0,0083	0,0005	-0,0193	0,0000
exportations du produit "s1"	*E s1*	97	0,0103	0,0103	0,0206	0,0206	0,0515	0,0515	0,1031	0,1031
exportations du produit "s2"	*E s2*	5386	0,0078	0,0078	0,0158	0,0158	0,0490	0,0490	0,1029	0,1029
exportations du produit "s3"	*E s3*	2545	0,0024	0,0024	0,0043	0,0043	0,0134	0,0134	0,0279	0,0279
demande pour le produit composite "s1"	*q s1*	3675,64	0,0019	-0,0069	0,0042	-0,0125	0,0134	-0,0382	0,0295	-0,0782
demande pour le produit composite "s2"	*q s2*	22085,4	-0,0002	-0,0037	-0,0004	-0,0066	-0,0015	-0,0222	-0,0042	-0,0436
demande pour le produit composite "s3"	*q s3*	7076,92	-0,0001	-0,0011	-0,0003	-0,0052	-0,0011	-0,0179	-0,0034	-0,0399
investissement total	*tinv*	4771		-0,0038		-0,0075		-0,0228		-0,0449
demande d'investissement totale du produit "s1"	*inv s1*	347,82	0,0029	-0,0059	0,0088	-0,0079	0,0294	-0,0229	0,0647	-0,0467
demande d'investissement totale du produit "s2"	*inv s2*	4356,82	-0,0005	-0,0041	-0,0008	-0,0071	-0,0023	-0,0229	-0,0051	-0,0445
demande d'investissement totale du produit "s3"	*inv s3*	65,78	0,0000	-0,0010	0,0000	-0,0049	0,0000	-0,0168	0,0000	-0,0366
demande intermédiaire pour le produit "s1"	*intd s1*	1927,34	0,0000	-0,0093	-0,0005	-0,0177	-0,0021	-0,0528	-0,0048	-0,1089
demande intermédiaire pour le produit "s2"	*intd s2*	12330,9	-0,0004	-0,0040	-0,0008	-0,0072	-0,0023	-0,0230	-0,0044	-0,0438
demande intermédiaire pour le produit "s3"	*intd s3*	2918,61	0,0003	-0,0010	0,0003	-0,0046	0,0003	-0,0165	0,0007	-0,0359
consommation publique	*g*	3421		-0,0041		-0,0082		-0,0248		-0,0514
demande pour le produit intérieur "s1"	*d s1*	3155,71	-0,0003	0,0130	-0,0009	0,0052	-0,0031	-0,0255	-0,0065	-0,0756
demande pour le produit intérieur "s2"	*d s2*	13121	-0,0041	-0,0051	-0,0084	-0,0103	-0,0259	-0,0306	-0,0546	-0,0627
demande pour le produit intérieur "s3"	*d s3*	6710,57	0,0000	-0,0030	-0,0002	-0,0061	-0,0008	-0,0195	-0,0026	-0,0420

Tableau 1 – Les statistiques de base et les différentes simulations (en millions de DT et en %): ***Les baisses tarifaires concernant tous les produits importés*** *("s1", "s2" et "s3") la suite:*

Les noms des variables	Les noms in GAMS	Année de référence volume	Année de référence valeur	simulation 1 volume	simulation 1 valeur	simulation 2 volume	simulation 2 valeur	simulation 3 volume	simulation 3 valeur	simulation 4 volume	simulation 4 valeur
matrice input output:											
entre produit "s1" et activité "s1"	*mat s1.s1*	210	214,83	0,0000	-0,0088	0,0000	-0,0166	0,0000	-0,0508	0,0000	-0,1046
entre produit "s1" et activité "s2"	*mat s1.s2*	1525	1560,08	-0,0007	-0,0094	-0,0013	-0,0179	-0,0033	-0,0539	-0,0066	-0,1105
entre produit "s1" et activité "s3"	*mat s1.s3*	130	132,99	0,0000	-0,0088	0,0000	-0,0166	0,0077	-0,0435	0,0077	-0,0977
entre produit "s1" et activité "s4"	*mat s1.s4*	19	19,437	0,0000	-0,0088	0,0000	-0,0166	0,0000	-0,0508	0,0000	-0,1046
entre produit "s2" et activité "s1"	*mat s2.s1*	415	461,48	0,0000	-0,0036	-0,0024	-0,0087	-0,0024	-0,0230	-0,0048	-0,0442
entre produit "s2" et activité "s2"	*mat s2.s2*	9054	10068	-0,0006	-0,0041	-0,0011	-0,0074	-0,0031	-0,0237	-0,0060	-0,0453
entre produit "s2" et activité "s3"	*mat s2.s3*	1126	1252,11	0,0000	-0,0036	0,0009	-0,0054	0,0027	-0,0181	0,0062	-0,0336
entre produit "s2" et activité "s4"	*mat s2.s4*	494	549,328	0,0000	-0,0036	0,0000	-0,0063	0,0000	-0,0207	0,0000	-0,0396
entre produit "s3" et activité "s1"	*mat s3.s1*	77	77,924	0,0000	-0,0010	0,0000	-0,0049	0,0000	-0,0168	0,0000	-0,0366
entre produit "s3" et activité "s2"	*mat s3.s2*	1173	1187,08	-0,0009	-0,0018	-0,0009	-0,0058	-0,0034	-0,0202	-0,0060	-0,0423
entre produit "s3" et activité "s3"	*mat s3.s3*	1420	1437,04	0,0007	-0,0003	0,0014	-0,0035	0,0035	-0,0133	0,0063	-0,0305
entre produit "s3" et activité "s4"	*mat s3.s4*	214	216,568	0,0000	-0,0010	0,0000	-0,0049	0,0000	-0,0168	0,0000	-0,0366
entre produit "s4" et activité "s1"	*mat s4.s1*	0	0	0							
entre produit "s4" et activité "s2"	*mat s4.s2*	0	0	0							
entre produit "s4" et activité "s3"	*mat s4.s3*	0	0	0							
entre produit "s4" et activité "s4"	*mat s4.s4*	0	0	0							
taux de rendement du capital dans l'activité "s1"	*r s1*		1		-0,0070		-0,0140		-0,0440		-0,0920
taux de rendement du capital dans l'activité "s2"	*r s2*		1		-0,0050		-0,0100		-0,0300		-0,0610
taux de rendement du capital dans l'activité "s3"	*r s3*		1		-0,0030		-0,0050		-0,0170		-0,0370
prix de l'investis-sement total	*pinv*		1,54		-0,0039		-0,0078		-0,0227		-0,0448
prix pondéré de la valeur ajoutée	*pindex*		1		-0,0040		-0,0090		-0,0270		-0,0560
prix à la produc-tion du produit "s1"	*px s1*		1		-0,0060		-0,0130		-0,0390		-0,0810
prix à la produc-tion du produit "s2"	*px s2*		1		-0,0040		-0,0080		-0,0250		-0,0510
prix à la produc-tion du produit "s3"	*px s3*		1		-0,0030		-0,0060		-0,0200		-0,0420
prix à la produc-tion du produit "s4"	*px s4*		1		-0,0040		-0,0080		-0,0250		-0,0520
prix de la valeur ajoutée dans l'activité "s1"	*pva s1*		1		-0,0070		-0,0140		-0,0420		-0,0880
prix de la valeur ajoutée dans l'activité "s2"	*pva s2*		1		-0,0050		-0,0090		-0,0280		-0,0580
prix de la valeur ajoutée dans l'activité "s3"	*pva s3*		1		-0,0030		-0,0060		-0,0200		-0,0420
prix de la valeur ajoutée dans l'activité "s4"	*pva s4*		1		-0,0040		-0,0080		-0,0260		-0,0550
prix du marché du produit composite "s1"	*pq s1*		1,023		-0,0088		-0,0166		-0,0508		-0,1046
prix du marché du produit composite "s2"	*pq s2*		1,112		-0,0036		-0,0063		-0,0207		-0,0396
prix du marché du produit composite "s3"	*pq s3*		1,012		-0,0020		-0,0049		-0,0168		-0,0366

Tableau 1 – Les statistiques de base et les différentes simulations (en millions de DT et en %): ***Les baisses tarifaires concernant tous les produits importés*** *("s1", "s2" et "s3") la suite:*

Les noms des variables	Les noms in GAMS	Année de référence		simulation 1		simulation 2		simulation 3		simulation 4	
		volume	valeur	volume	valeur	volume	valeur	volume	valeur	volume	valeur
consommation du ménage en produit "s1"	*ch s1*	1368	1399,46	0,0051	-0,0037	0,0102	-0,0066	0,0322	-0,0203	0,0687	-0,0431
consommation du ménage en produit "s2"	*ch s2*	4855	5398,76	0,0004	-0,0032	0,0006	-0,0057	0,0008	-0,0199	-0,0033	-0,0427
consommation du ménage en produit "s3"	*ch s3*	4044	4092,53	-0,0002	-0,0012	-0,0007	-0,0057	-0,0022	-0,0190	-0,0064	-0,0428
demande de travail par activité "s1"	*ls1*	268	268	-0,0037	-0,0077	-0,0075	-0,0154	-0,0149	-0,0405	-0,0336	-0,0867
demande de travail par activité "s2"	*l s2*	2323	2323	-0,0013	-0,0053	-0,0026	-0,0106	-0,0069	-0,0327	-0,0133	-0,0676
demande de travail par activité "s3"	*l s3*	1764	1764	0,0023	-0,0017	0,0040	-0,0041	0,0113	-0,0150	0,0221	-0,0341
demande de travail par activité "s4"	*l s4*	2634	2634	0,0000	-0,0040	0,0000	-0,0080	0,0000	-0,0260	0,0000	-0,9071
consommation intermédiaire totale de l'activité "s1"	*intp s1*	702	754,234	0,0000	-0,0048	-0,0014	-0,0106	-0,0014	-0,0303	-0,0043	-0,0606
consommation intermédiaire totale de l'activité "s2"	*intp s2*	11752	12815,2	-0,0006	-0,0046	-0,0011	-0,0085	-0,0031	-0,0271	-0,0060	-0,0530
consommation intermédiaire totale de l'activité "s3"	*intp s3*	2676	2822,14	0,0007	-0,0022	0,0011	-0,0050	0,0034	-0,0169	0,0064	-0,0350
consommation intermédiaire totale de l'activité "s4"	*intp s4*	727	785,333	0,0014	-0,0030	0,0014	-0,0062	0,0014	-0,0204	0,0014	-0,0403
valeur ajoutée de l'activité "s1"	*va s1*	2574	2574	-0,0004	-0,0074	-0,0008	-0,0148	-0,0016	-0,0435	-0,0035	-0,0912
valeur ajoutée de l'activité "s2"	*va s2*	5100	5100	-0,0004	-0,0054	-0,0010	-0,0100	-0,0031	-0,0310	-0,0059	-0,0635
valeur ajoutée de l'activité "s3"	*va s3*	6353	6353	0,0005	-0,0025	0,0009	-0,0051	0,0030	-0,0171	0,0060	-0,0363
valeur ajoutée de l'activité "s4"	*va s4*	2634	2634	0,0000	-0,0040	0,0000	-0,0080	0,0000	-0,0260	0,0000	-0,0550
output de l'activité "s1"	*x s1*	3328	3328	-0,0003	-0,0063	-0,0006	-0,0136	-0,0015	-0,0404	-0,0033	-0,0840
output de l'activité "s2"	*x s2*	17918	17918	-0,0005	-0,0045	-0,0011	-0,0091	-0,0031	-0,0280	-0,0060	-0,0567
output de l'activité "s3"	*x s3*	9176	9176	0,0005	-0,0025	0,0011	-0,0049	0,0032	-0,0169	0,0061	-0,0362
taux de salaire moyen	*w*		1		-0,0040		-0,0080		-0,0260		-0,0550

Tableau 1 – Les statistiques de base et les différentes simulations (en millions de DT et en %): ***Les baisses tarifaires concernant tous les produits importés*** *("s1", "s2" et "s3")*

Les noms des variables	Les noms in GAMS	Année de référence valeur	simulation 1 valeur en (%)	simulation 2 valeur en (%)	simulation 3 valeur en (%)	simulation 4 valeur en (%)
prix intérieur(du marché) du produit importé "s1"	*pm s1*	1,249	-0,0168	-0,0344	-0,1025	-0,2050
prix intérieur(du marché) du produit importé "s2"	*pm s2*	1,224	-0,0074	-0,0147	-0,0425	-0,0833
prix intérieur(du marché) du produit importé "s3"	*pm s3*	1,012	0,0020	0,0030	0,0089	0,0198
prix au producteur du produit "s1" pour la vente sur le marché intérieur	*pl s1*	1	-0,0070	-0,0130	-0,0400	-0,0840
prix au producteur du produit "s2" pour la vente sur le marché intérieur	*pl s2*	1	-0,0060	-0,0120	-0,0360	-0,0740
prix au producteur du produit "s3" pour la vente sur le marché intérieur	*pl s3*	1	-0,0040	-0,0090	-0,0280	-0,0580
prix payé à l'exportation du produit "s1"	*pe s1*	1	0,0000	0,0000	0,0000	0,0000
prix payé à l'exportation du produit "s2"	*pe s2*	1	0,0000	0,0000	0,0000	0,0000
prix payé à l'exportation du produit "s3"	*pe s3*	1	0,0000	0,0000	0,0000	0,0000
prix du marché du produit local "s1" vendu sur le marché intérieur	*pd s1*	0,977	0,0133	0,0061	-0,0225	-0,0696
prix du marché du produit local "s2" vendu sur le marché intérieur	*pd s2*	1,047	-0,0010	-0,0019	-0,0048	-0,0086
prix du marché du produit local "s3" vendu sur le marché intérieur	*pd s3*	1,012	-0,0030	-0,0059	-0,0188	-0,0395

Légende: les activités et les produits:
"s1": agriculture et pêche
"s2": industries
"s3": services marchands
"s4": services non marchands

Annexe 3

Tableau 2 – Les statistiques de base et les différentes simulations (en millions de DT et en %): ***Les baisses tarifaires concernant seulement les produits industriels*** *("s2")*

Les noms des variables	Les noms in GAMS	Année de référence	simulation 5		simulation 6		simulation 7		simulation 8	
		valeur	volume	valeur	volume	valeur	volume	valeur	volume	valeur
facteur d'ajustement	*adj*	1		0,0730		0,580		0,4200		0,7970
épargne du ménage	*savh*	2241		-0,0022		-0,0049		-0,0161		-0,0348
épargne du gouvernement	*savg*	348		0,0460		0,0948		0,2845		0,5833
épargne des entreprises	*savf*	1652		-0,0163		-0,0327		-0,0975		-0,1937
ajustement de l'impôt indirect	*adj_ntax*	0								
recettes totales provenant de la taxation indirecte	*itxrev*	1072		0,0942		0,1894		0,5821		1,2127
recettes totales du gouvernement provenant des impôts directs payés par les ménages et les entreprises	*dtxrev*	1586		-0,0032		-0,0057		-0,0170		-0,0366
recettes totales provenant des taxes à l'importation	*tmrev*	1328		-0,0730		-0,1468		-0,4503		-0,9352
revenu disponible du ménage	*dyh*	13134		-0,0024		-0,0050		-0,0160		-0,0348
revenu des entreprises	*yf*	3193		-0,0034		-0,0066		-0,0204		-0,0426
revenu du ménage	*yh*	14381		-0,0025		-0,0050		-0,0161		-0,0348
revenu du facteur capital	*yk*	9672		-0,0032		-0,0065		-0,0202		-0,0424
revenu du facteur travail	*yl*	698		-0,0046		-0,0092		-0,0280		-0,0582
importations du produit "s1"	*M s1*	452,138	-0,0028	-0,0028	0,0083	-0,0091	0,0304	0,0319	-0,0608	-0,0645
importations du produit "s2"	*M s2*	8970,7	0,0061	-0,0013	0,0123	-0,0026	0,0378	-0,0088	0,0791	-0,0161
importations du produit "s3"	*M s3*	366,344	0,0000	0,0020	-0,0028	0,0002	-0,0055	0,0033	-0,0166	0,0009
exportations du produit "s1"	*E s1*	97	0,0103	0,0103	0,0103	0,0103	0,0206	0,0206	0,0515	0,0515
exportations du produit "s2"	*E s2*	5386	0,0067	0,0067	0,0136	0,0136	0,0420	0,0420	0,0876	0,0876
exportations du produit "s3"	*E s3*	2545	0,0024	0,0024	0,0047	0,0047	0,0141	0,0141	0,0295	0,0295
demande pour le produit composite "s1"	*q s1*	3675,64	-0,0006	-0,0035	0,0011	-0,0070	-0,0031	-0,0187	-0,0067	-0,0397
demande pour le produit composite "s2"	*q s2*	22085,4	-0,0001	-0,0037	-0,0003	-0,0074	-0,0011	-0,0218	-0,0033	-0,0445
demande pour le produit composite "s3"	*q s3*	7076,92	0,0001	-0,0018	0,0001	-0,0048	0,0003	-0,0165	-0,0001	-0,0347
investissement total	*tinv*	4771		-0,0034		-0,0069		-0,0208		-0,0409
demande d'investissement totale du produit "s1"	*inv s1*	347,82	0,0000	-0,0029	0,0029	-0,0088	-0,0059	-0,0214	-0,0088	-0,0418
demande d'investissement totale du produit "s2"	*inv s2*	4356,82	0,0000	-0,0036	0,0000	-0,0072	0,0003	-0,0204	0,0008	-0,0406
demande d'investissement totale du produit "s3"	*inv s3*	65,78	0,0000	-0,0020	0,0000	-0,0049	0,0000	-0,0168	0,0000	-0,0346
demande intermédiaire pour le produit "s1"	*intd s1*	1927,34	-0,0005	-0,0040	-0,0011	-0,0075	-0,0042	-0,0198	-0,0090	-0,0420
demande intermédiaire pour le produit "s2"	*intd s2*	12330,9	-0,0007	-0,0043	-0,0015	-0,0086	-0,0045	-0,0252	-0,0093	-0,0503
demande intermédiaire pour le produit "s3"	*intd s3*	2918,61	0,0000	-0,0020	0,0000	-0,0049	0,0000	-0,0168	-0,0003	-0,0353
consommation publique	*g*	3421		-0,0044		-0,0085		-0,0263		-0,0541
demande pour le produit intérieur "s1"	*d s1*	3155,71	0,0003	0,0177	0,0003	0,0146	0,0009	0,0040	-0,0015	-0,0149
demande pour le produit intérieur "s2"	*d s2*	13121	-0,0044	-0,0053	-0,0088	-0,0107	-0,0272	-0,0319	-0,0572	-0,0644
demande pour le produit intérieur "s3"	*d s3*	6710,57	0,0002	-0,0028	0,0003	-0,0056	0,0006	-0,0172	0,0008	-0,0368

Tableau 2 – Les statistiques de base et les différentes simulations (en millions de DT et en %): ***Les baisses tarifaires concernant seulement les produits industriels*** *("s2") la suite:*

Les noms des variables	Les noms in GAMS	Année de référence		simulation 5		simulation 6		simulation 7		simulation 8	
		volume	valeur	volume	valeur	volume	valeur	volume	valeur	volume	valeur
matrice input output:											
entre produit "s1" et activité "s1"	*mat s1.s1*	210	214,83	0	-0,0029	0,0000	-0,0059	0,0048	-0,0110	0,0048	-0,0286
entre produit "s1" et activité "s2"	*mat s1.s2*	1525	1560,08	-0,0013	-0,0042	-0,0026	-0,0085	-0,0066	-0,0221	-0,0131	-0,0459
entre produit "s1" et activité "s3"	*mat s1.s3*	130	132,99	0,0000	-0,0029	0,0077	0,0018	0,0077	-0,0081	0,0154	-0,0184
entre produit "s1" et activité "s4"	*mat s1.s4*	19	19,437	0,0000	-0,0029	0,0000	-0,0059	0,0000	-0,0156	0,0000	-0,0332
entre produit "s2" et activité "s1"	*mat s2.s1*	415	461,48	0,0000	-0,0036	0,0000	-0,0072	0,0000	-0,0207	0,0024	-0,0391
entre produit "s2" et activité "s2"	*mat s2.s2*	9054	10068	-0,0010	-0,0046	-0,0020	-0,0092	-0,0062	-0,0267	-0,0126	-0,0534
entre produit "s2" et activité "s3"	*mat s2.s3*	1126	1252,11	0,0009	-0,0027	0,0018	-0,0054	0,0044	-0,0163	0,0089	-0,0329
entre produit "s2" et activité "s4"	*mat s2.s4*	494	549,328	0,0000	-0,0036	0,0000	-0,0072	0,0000	-0,0207	0,0000	-0,0414
entre produit "s3" et activité "s1"	*mat s3.s1*	77	77,924	0,0000	-0,0020	0,0000	-0,0049	0,0000	-0,0168	0,0000	-0,0346
entre produit "s3" et activité "s2"	*mat s3.s2*	1173	1187,08	-0,0009	-0,0028	-0,0026	-0,0075	-0,0060	-0,0227	-0,0128	-0,0469
entre produit "s3" et activité "s3"	*mat s3.s3*	1420	1437,04	0,0007	-0,0013	0,0021	-0,0028	0,0049	-0,0120	0,0092	-0,0257
entre produit "s3" et activité "s4"	*mat s3.s4*	214	216,568	0,0000	-0,0020	0,0000	-0,0049	0,0000	-0,0168	0,0000	-0,0346
entre produit "s4" et activité "s1"	*mat s4.s1*	0	0								
entre produit "s4" et activité "s2"	*mat s4.s2*	0	0								
entre produit "s4" et activité "s3"	*mat s4.s3*	0	0								
entre produit "s4" et activité "s4"	*mat s4.s4*	0	0								
taux de rendement du capital dans l'activité "s1"	*r s1*		1		-0,0020		-0,0040		-0,0120		-0,0260
taux de rendement du capital dans l'activité "s2"	*r s2*		1		-0,0060		-0,0110		-0,0350		-0,0710
taux de rendement du capital dans l'activité "s3"	*r s3*		1		-0,0020		-0,0050		-0,0150		-0,0330
prix de l'investis-sement total	*pinv*		1,54		-0,0032		-0,0071		-0,0208		-0,0409
prix pondéré de la valeur ajoutée	*pindex*		1		-0,0040		-0,0080		-0,0230		-0,0490
prix à la produc-tion du produit "s1"	*px s1*		1		-0,0020		-0,0050		-0,0150		-0,0320
prix à la produc-tion du produit "s2"	*px s2*		1		-0,0040		-0,0080		-0,0230		-0,0470
prix à la produc-tion du produit "s3"	*px s3*		1		-0,0030		-0,0060		-0,0190		-0,0390
prix à la produc-tion du produit "s4"	*px s4*		1		-0,0040		-0,0090		-0,0260		-0,0540
prix de la valeur ajoutée dans l'activité "s1"	*pva s1*		1		-0,0020		-0,0040		-0,0140		-0,0300
prix de la valeur ajoutée dans l'activité "s2"	*pva s2*		1		-0,0050		-0,0100		-0,0320		-0,0650
prix de la valeur ajoutée dans l'activité "s3"	*pva s3*		1		-0,0030		-0,0060		-0,0190		-0,0400
prix de la valeur ajoutée dans l'activité "s4"	*pva s4*		1		-0,0050		-0,0090		-0,0280		-0,0580
prix du marché du produit composite "s1"	*pq s1*		1,023		-0,0029		-0,0059		-0,0156		-0,0332
prix du marché du produit composite "s2"	*pq s2*		1,112		-0,0036		-0,0072		-0,0207		-0,0414
prix du marché du produit composite "s3"	*pq s3*		1,012		-0,0020		-0,0049		-0,0168		-0,0346

Tableau 2 – Les statistiques de base et les différentes simulations (en millions de DT et en %) *la suite:*

Les noms des variables	Les noms in GAMS	Année de référence		simulation 5		simulation 6		simulation 7		simulation 8	
		volume	valeur	volume	valeur	volume	valeur	volume	valeur	volume	valeur
consommation du ménage en produit "s1"	*ch s1*	1368	1399,46	0,0000	-0,0029	0,0000	-0,0059	0,0000	-0,0156	-0,0022	-0,0354
consommation du ménage en produit "s2"	*ch s2*	4855	5398,76	0,0010	-0,0026	0,0021	-0,0051	0,0054	-0,0154	0,0070	-0,0347
consommation du ménage en produit "s3"	*ch s3*	4044	4092,53	0,0000	-0,0020	0,0002	-0,0047	0,0007	-0,0161	0,0002	-0,0343
demande de travail par activité "s1"	*ls1*	268	268	0,0000	-0,0050	-0,0037	-0,0053	0,0112	-0,0171	0,0261	-0,0334
demande de travail par activité "s2"	*l s2*	2323	2323	-0,0022	-0,0071	-0,0043	-0,0133	-0,0138	-0,0414	-0,0276	-0,0840
demande de travail par activité "s3"	*l s3*	1764	1764	0,0028	-0,0022	0,0057	-0,0034	0,0159	-0,0126	0,0323	-0,0276
demande de travail par activité "s4"	*l s4*	2634	2634	0,0000	-0,0050	0,0000	-0,0090	0,0000	-0,0280	0,0000	-0,0580
consommation intermédiaire totale de l'activité "s1"	*intp s1*	702	754,234	0,0000	-0,0032	0,0000	-0,0066	0,0014	-0,0175	0,0028	-0,0356
consommation intermédiaire totale de l'activité "s2"	*intp s2*	11752	12815,2	-0,0010	-0,0044	-0,0020	-0,0089	-0,0062	-0,0258	-0,0126	-0,0519
consommation intermédiaire totale de l'activité "s3"	*intp s3*	2676	2822,14	0,0011	-0,0020	0,0019	-0,0038	0,0045	-0,0137	0,0090	-0,0286
consommation intermédiaire totale de l'activité "s4"	*intp s4*	727	785,333	0,0014	-0,0031	0,0014	-0,0065	0,0014	-0,0195	0,0014	-0,0393
valeur ajoutée de l'activité "s1"	*va s1*	2574	2574	0,0000	-0,0020	0,0004	-0,0036	0,0012	-0,0129	0,0027	-0,0274
valeur ajoutée de l'activité "s2"	*va s2*	5100	5100	-0,0010	-0,0060	-0,0020	-0,0119	-0,0061	-0,0379	-0,0125	-0,0767
valeur ajoutée de l'activité "s3"	*va s3*	6353	6353	0,0006	-0,0024	0,0014	-0,0046	0,0042	-0,0148	0,0088	-0,0315
valeur ajoutée de l'activité "s4"	*va s4*	2634	2634	0,0000	-0,0050	0,0000	-0,0090	0,0000	-0,0280	0,0000	-0,0580
output de l'activité "s1"	*x s1*	3328	3328	0,0000	-0,0020	0,0003	-0,0047	0,0012	-0,0138	0,0027	-0,0294
output de l'activité "s2"	*x s2*	17918	17918	-0,0010	-0,0050	-0,0020	-0,0100	-0,0061	-0,0290	-0,0126	-0,0590
output de l'activité "s3"	*x s3*	9176	9176	0,0008	-0,0022	0,0015	-0,0045	0,0044	-0,0147	0,0088	-0,0305
taux de salaire moyen	*w*		1		-0,0050		-0,0090		-0,0280		-0,0580

Tableau 2 – Les statistiques de base et les différentes simulations (en millions de DT et en %) *la suite:*

Les noms des variables	Les noms in GAMS	Année de référence valeur	simulation 5 valeur en (%)	simulation 6 valeur en (%)	simulation 7 valeur en (%)	simulation 8 valeur en (%)
prix intérieur(du marché) du produit importé "s1"	*pm s1*	1,249	0,0000	-0,0008	-0,0016	-0,0040
prix intérieur(du marché) du produit importé "s2"	*pm s2*	1,224	-0,0074	-0,0147	-0,0449	-0,0882
prix intérieur(du marché) du produit importé "s3"	*pm s3*	1,012	0,0020	0,0030	0,0089	0,0178
prix au producteur du produit "s1" pour la vente sur le marché intérieur	*pl s1*	1	-0,0230	-0,0050	-0,0160	-0,0330
prix au producteur du produit "s2" pour la vente sur le marché intérieur	*pl s2*	1	-0,0060	-0,0110	-0,0340	-0,0690
prix au producteur du produit "s3" pour la vente sur le marché intérieur	*pl s3*	1	-0,0040	-0,0080	-0,0260	-0,0550
prix payé à l'exportation du produit "s1"	*pe s1*	1	0,0000	0,0000	0,0000	0,0000
prix payé à l'exportation du produit "s2"	*pe s2*	1	0,0000	0,0000	0,0000	0,0000
prix payé à l'exportation du produit "s3"	*pe s3*	1	0,0000	0,0000	0,0000	0,0000
prix du marché du produit local "s1" vendu sur le marché intérieur	*pd s1*	0,977	0,0174	0,0143	0,0031	-0,0164
prix du marché du produit local "s2" vendu sur le marché intérieur	*pd s2*	1,047	-0,0010	-0,0019	-0,0048	-0,0076
prix du marché du produit local "s3" vendu sur le marché intérieur	*pd s3*	1,012	-0,0030	-0,0059	-0,0178	-0,0375

Légende: les activités et les produits:
"s1": agriculture et pêche
"s2": industries
"s3": services marchands
"s4": services non marchands