Artigos

Multinational Corporations, Foreign Investment, and Royalties and License Fees: Effects on Host-Country Total Factor Productivity

Argentino Pessoa Faculdade de Economia do Porto

resumo

Neste artigo examinamos, do ponto de vista teórico e empírico, a relação entre IDE (Investimento Directo Estrangeiro) recebido e PTF (produtividade total dos factores), à luz do paradigma OLI (organização, localização e internalização). Utilizamos técnicas de regressão em painel para analisar os efeitos do IDE recebido e do pagamento de royalties e Licenças sobre a PTF agregada duma amostra de 16 países da OCDE, entre 1985 e 2002. Os testes empíricos mostram que o IDE recebido e o pagamento de royalties e licenças tem um impacto positivo na PTF dos países e, também, sugerem que a dimensão dos efeitos positivos está dependente do nível de desenvolvimento desses países. Por outro lado, a nossa investigação mostra que, quando outros factores permanecem constantes, o IDE recebido e as royalties pagas são substitutos no impacto positivo que exercem sobre a PTF.

résumé / abstract

Cet article analyse, du point de vue théorique et empirique, la relation entre l'investissement direct étranger (IDE) reçu et la PGF (productivité globale des facteurs). Avec un encadrement théorique inspiré du paradigme OLI (organisation, localisation et internalisation), nous utilisons des techniques de régression sur un panel de donnés pour mesurer les impacts des entrées d'IDE et des redevances payées à l'étranger sur la PGF à partir d'un échantillon de 16 pays de l'OCDE, entre 1985-2002. Les tests empiriques démontrent que l'IDE entré et le paiement de redevances ont un impact positif sur la PGF. Ils montrent aussi que la dimension de l'effet positif dépend du niveau de développement de ces pays. Par ailleurs, notre recherche montre que, tous les autres facteurs restant inchangés, l'IDE reçu et les redevances payées sont des substituts réciproques par rapport à l'impact positif sur la PGF.

In this paper we examine the relationship between inward FDI and total factor productivity (TFP) in a framework motivated by the OLI paradigm. A panel data approach is used to study the effects of FDI and payments of royalties and license fees (R&L) on aggregate TFP in a sample of 16 OECD countries, between 1985 and 2002. Our empirical tests show that FDI and R&L have a positive impact on host-country TFP, and also suggest that the amount of positive effects of FDI and R&L is dependent on the level of development of the receiving country. Additionally, our data show that, when other factors remain constant, inward FDI and R&L payments are substitutes in positively influencing TFP of the host country.

Classificação JEL: C33, F21, F23, F43.

1. Introduction

In recent years improved attention has been paid to the deep analysis of the effects of Foreign Direct Investment (FDI) (see, for example, Blomström and Kokko, 2003; and Görg and Greenaway, 2004). The main reason is that FDI often involves the transfer of knowledge from one country to another (*e.g.*, Carr *et al*., 2001), making it a potentially important vehicle for international diffusion of technology[1]. As a matter of fact, theoretical models of foreign investment suggest that there should be a positive relationship between FDI and international diffusion of technology (Caves, 1974, 1996; Markusen, 1995). Knowledge will move through demonstration effects, labour turnover, or reverse engineering.

The positive effects predicted by theoretical models, and documented by some empirical studies, have been driving a considerable change in the attitude towards inward FDI over the last couple of decades, as most countries have liberalized their policies to attract investments from foreign multinational corporations (MNCs). Additionally, Mullen and Williams (2005) have studied the role of FDI in regional economic performance over the period 1977 to 1997 and have found that inward foreign investment has played a 'strong, vital role in regional economic activity' with significant effects on 'regional productivity and growth' (2005, p. 279). Accordingly, in the expectation that some of the knowledge brought by foreign companies may spill over to the receiving country's domestic firms, governments across the world have lowered various entry barriers and opened up new sectors to foreign investment. An increasing number of national governments also provide a variety of forms of investment incentives to encourage foreign owned companies to invest in their countries.

The economic rationale for offering special incentives to attract FDI frequently derives from the conviction that foreign investment produces externalities in the form of technology transfers and spillovers[2]. Romer (1993), for example, argues that there are significant 'idea gaps' between rich and poor countries, foreign investment being an important instrument in transferring technological and business know-how to poorer countries, with substantial spillover effects for the poorer economy as a whole.

But, in spite of the predictions of theoretical models, the empirical microeconomic and microeconometric literature on the occurrence of positive effects of inward FDI is, at best, controversial. If, on the one hand, early studies using industry-level data, such as Blomström and Persson (1983), find that foreign presence in an industry positively influences domestic labour productivity, and the earliest statistical analyses of inter-industry effects of FDI claim that technical progress did not only take place in the FDI own industries, but also in other sectors (Katz, 1969), on the other hand, some more recent studies using firm-level data are less supportive of the existence of positive spillovers[3]. Furthermore, some authors as Aitken *et al.* (1997) and Haddad and Harrison (1993) find that foreign investment has a negative effect on the performance of domestically owned firms. In addition, at a macroeconomic level, the evidence is not more conclusive. While some studies – using aggregate FDI flows for a broad cross-section of countries – have generally suggested a positive role for FDI in generating economic growth, particularly in environments where the capability of absorption matters (De Gregorio, 1992), other works using more sophisticated econometric methods (Carkovic and Levine, 2002; Alfaro *et al*., 2004; Chowdhury and Mavrotas, 2006) cast doubts about the optimistic conclusions of the first analyses[4].

1 Of course, there are many other reasons why FDI has become a much-discussed topic. One of them is the dramatic increase in the global flow and the resulting rise in its importance as a source of investment funds for an increasing number of countries.

2 As it is well known, in absence of spillovers, there is no reason for policy to differentiate between FDI and other forms of investment, including domestic investment.

3 Of course there are some exceptions. Among them, we have the evidence from China's manufacturing sector provided by Chuang and Hsu (2004).

4 In addition, Lichtenberg and van Pottelsberghe de la Potterie (1996) have analysed the importance of FDI for

As it is well known, some problems have affected the microeconomic empirical studies of inward FDI spillover effects. Firstly, because data problems are particularly acute with regard to service industries, most research on FDI at the firm level focuses exclusively on goods[5]. Secondly, empirical work on FDI is generally overwhelmed by the limited availability and quality of the data. As a result, empirical research on FDI at firm level is largely limited to firms from just a few countries. Furthermore if, as Cantwell (1989) and Aitken and Harrison (1999) argue, positive technology spillovers do not occur in all industries, the estimates of empirical studies at the firm level may be more or less biased, according to the type of industry included in the sample of firms. On the other hand, macroeconomic empirical studies had been criticized too. A great lot of the criticisms have been directed to their econometric grounds (Carkovic and Levine, 2002)[6] and to the direction of causality in the relationship between FDI and economic growth (Kholdy, 1995; Chowdhury and Mavrotas, 2006).

However, the difficulty in finding robust empirical evidence to support the existence of positive spillovers can be associated to a more fundamental problem: researchers can be looking in the wrong place and with erroneous instruments. As a matter of fact, the macroeconomic evidence follows usually the framework of cross-country growth regressions, trying to explain the effect of FDI on per capita economic growth rate, this rate being typically measured in five-year periods. Our work uses a different framework. On the one hand, we follow Solow (2001) in considering that analysis shall focus more directly on TFP as the proper left-hand-side variable; on the other hand we use a descriptive framework motivated by the OLI paradigm, developed by Dunning (1981, 1988), and by the knowledge capital model (Markusen and Maskus, 1999) which require the inclusion within explanatory variables of one that controls the payment of royalties and license fees (R&L).

Consequently, in face of the above-mentioned change in the attitude towards inward FDI over the two recent decades, given the limited guidance to the empirical work provided by the theory and the inconclusive results of the empirical literature on the relationship between FDI and economic growth, it is significant to examine whether the positive effects of FDI are strong and systematic enough to justify supporting foreign investment with fiscal and financial incentives. Particularly, if we are in face of both positive and negative spillovers, as some microeconomic empirical evidence seems to point, it is crucial to find out the net effect of inward FDI at country level.

Our work uses a descriptive framework and, based on it, investigates how foreign investment has affected the aggregate total factor productivity (TFP) of OECD countries. We use a panel data approach, which has the advantages mentioned by Görg and Strobl (2003), to estimate the elasticity of TFP in relation to inward FDI, to royalties and license fees (R&L), and to the country's level of development. We also explore the interaction between inward FDI and R&L.

The remainder of the paper is structured as follows: Section 2 discusses the relationship between FDI and productivity; section 3 reviews the FDI evidence on host-country growth, spillovers and learning. Section 4 introduces the conceptual framework that drives our analysis and presents a statistical model that helps to rationalize some possible linkages between FDI and TFP. Section 5 presents estimates on elasticities of TFP with respect to both FDI and R&L. Section 6 deals with human capital as a possible channel of impact of level of development on TFP. Finally, section 7 concludes.

international diffusion of technology in thirteen OECD countries, with the same R&D weighting approach that Coe and Helpman (1995) and Keller (1998) use for imports, and they do not find significant effects from *inward* FDI.

5 This lack of empirical research on FDI in the services sector is increasingly troublesome, owing to the growing importance of services in production, in trade, and in investment.

6 For example, Carkovic and Levine (2002) argue that the macroeconomic findings on growth and FDI must be viewed suspiciously, because existing studies do not fully control for simultaneity bias, country-specific effects, and the routine use of lagged dependent variables in growth regressions.

2. The relationship between FDI and productivity

It is possible to identify three main channels through which inward FDI is thought to improve productivity of a host-country: a) *direct* improvement in efficiency through the redirection of local resources towards more productive uses, including within purchased firms; b) *indirect* impact via spillovers and other externalities associated with interactions between the foreign affiliate and the host country economy; and c) increase in domestic market competition. All the three channels can contribute to reduce the substantial gaps existing between countries.

The most obvious direct impact on productivity is through the reallocation of resources, including within firms, towards more productive activity. Since domestic firms are likely to have better knowledge and access to domestic markets, a foreign firm that decides to enter the market must have compensating advantages (Graham and Krugman, 1991). It seems likely that the foreign firms will benefit from advantages over domestic competitors deriving from some assets of the former like superior management skills, more advanced technology, and better access to international markets. Given the limited level of resources within the host economy, the entry of a foreign firm, whether by acquisition or physical investment, is likely to be associated with a shift of resources and effort away from a less productive activity. In this way, FDI is expected to make an increase in overall productivity within the economy possible. The dimension of the host country's benefits depends on the action of a large bulk of linkages between foreign owner(s) and domestic local firms and customers as well as on the effectiveness of the foreign firm to prevent the gains to be totally transferred to the host country's firms and consumers. For example, a foreign-owned firm may contribute directly to the productivity of domestic firms by lowering input costs, or by increasing the demand for inputs produced by local suppliers.

From the interactions with foreign affiliates, local firms can capture part of the benefits of the advantages of the former, this *indirect* impact being usually rationalized as positive externalities. Productivity spillovers from FDI take place when the entry or presence of multinational corporations increases the productivity of domestic firms in a host country, and the MNCs do not fully internalize the value of these benefits[7]. For example, local firms may learn with the practice of MNCs or MNCs may transfer cost-free technology and know-how to local suppliers in order to improve the quality of inputs (Rodriguez-Clare, 1996). Alternatively, local competitors might benefit by attracting employees with firm-specific knowledge from the foreign affiliate (Fosfuri *et al*., 2001). In case of vertical spillovers, that is, productivity spillovers that take place due to linkages between foreign firms and their local suppliers[8] it can be impossible to prevent the action of positive externalities.

But the absorptive capacity of the host country matters. In essence, the idea is that the magnitude of spillover benefits is dependent on the ability and motivation of local firms to interact with, learn from, and invest in, the technology and ideas that are spilt by the foreign firm. It is usually alleged that spillovers will only accrue if the technology, knowledge or any other 'gap' between foreign and local firms is not too large. However, the importance of relative backwardness is theoretically controversial. While Findlay (1978) suggests that the greater the distance between two economies in terms of development, the greater the accumulation of available opportunities to take advantage of in the less advanced economy, and so the greater the pressure for change and the more fast the assimilation of new technology, Glass and Saggi (1998) argue that the larger the gap, the lower the quality of the technology transferred and the lower the potential for spillovers.

The entrance of foreign affiliates is usually seen as a way of strengthening domestic market competition, thereby leading eventually to higher productivity, lower prices and more efficient resource allocation, but it can also lead to greater market concentration and reducing domestic

7 For a recent overview on spillover studies, see Castellani and Zanfei (2006).
8 For the several forms of operation of vertical spillovers see Smarzynska (2004)

market competition. The risk of the latter is exacerbated not only if the entrant has an important international market position, if the barriers to entry into the industry are high, but also because of host country characteristics: either a small market or markets unconnected geographically, competition laws weak or weakly enforced (OECD, 2002). So, the increase in domestic market competition is not always an assured outcome of the FDI entry, and where that outcome is achieved it can be either reduce the productivity of domestic firms as suggested by Aitken and Harrison (1999) and Konings (2001), or have positive effects as noted by Kokko (1996) and Driffield (2001). In order to assess the dimension of direct and indirect effects of FDI on productivity, and in face of possible contradictory effects of market competition, there is a large ground for empirical investigation, both at the micro and at the macro level.

3. Empirical evidence

As was clarified in the previous section there could be effects of FDI on productivity without the presence of spillovers. However, the empirical literature on growth effects of FDI, both at a macroeconomic level and at a microeconomic level, relies almost exclusively on the occurrence of spillovers from foreign owned to domestically owned firms. A complete review of empirical literature on FDI spillovers is outside the scope of this paper, which aims both at emphasizing some useful aspects to support the descriptive framework that we shall use as a basis of empirical work and to show the utility of a rationale like ours. Readers interested in a critical assessment of the role of FDI should see Blomström and Kokko (1998), or Görg and Greenaway (2004) and Pessoa (2008) if interested in more recent reviews.

Empirical microeconomic literature on productivity spillovers of FDI is abundant and varied (Castellani and Zanfei, 2006). For simplicity we can separate them in two types: case studies and statistical analyses. Case studies are specifically useful in describing the actual linkages between multinationals and host country players such as the particular forms assumed by technical assistance, control of quality, management training, and organization of the production process that are key aspects of the interaction of MNCs with their local suppliers (Moran 2001). The cases studied by Mansfield and Romeo (1980) and Rhee and Belot (1990), for example, illustrate various channels through which spillovers are assumed to occur, with a particular emphasis on technology transfer to domestic firms. Case studies have also documented the importance of local skills and in-house technological capacity to adapt and use techniques developed elsewhere (Lall, 1992; Evenson and Westphal, 1995; Ariffin and Figueiredo, 2004).

In general, case studies have shown significant positive FDI spillovers. For example, Larrain *et al*. (2000) show substantial spillover benefits for the local economy generated by Intel's investment in Costa Rica in 1997 (specifically, creating new training programmes in higher education institutions; producing important signalling effects on other investors; and attracting new suppliers to Costa Rica)[9]. However, the evidence from case studies is inconclusive. On the one hand it is, at best, mixed. On the other hand, case studies do not always offer quantitative information and do not easily generalize.

The mixed nature of evidence can frequently be seen. For example, Rhee and Belot (1990) suggest that foreign entrants led to the creation of booming domestic textiles industries in Mauritius and Bangladesh. In contrast, Mansfield and Romeo (1980) surveyed 26 US-based MNCs and reported that in only a few cases had FDI accelerated the access of local competitors to new technology. Likewise, the conclusion about the Intel's investment in Costa Rica made by Hanson (2001) is very different from the one of Larrain *et al*. (2000). In face of the drawbacks of case studies, the research is usually based on industry-level and micro-level studies and is done

9 Other examples include a case study of three electronics investors in Singapore (Lim and Fong, 1982); a case study of the investments of General Motors and Ford in Brazil and the investment of Intel in Costa Rica (Hanson, 2001); and a survey of 72 senior managers in Kenyan manufacturing firms (Gershenberg, 1987).

through statistic analyses searching horizontal and vertical productivity spillovers, using either cross-section or panel data.

The literature on horizontal and vertical spillovers was recently surveyed by Görg and Greenaway (2004). These authors analyzed 42 studies (16 with cross-sectional data and 26 with panel data) on horizontal productivity spillovers in manufacturing industries in developed, developing and transition economies. In the 26 studies which employ panel data, the most appropriate estimating framework as was argued by Görg and Strobl (2003), surveyed by these authors only eight studies find unambiguously positive evidence and are almost all of them for developed countries, seven show evidence of negative effects of MNCs on host-country firms and the remainder present mixed or statistically insignificant results. Görg and Greenaway (2004) also surveyed studies on vertical spillovers and among the five studies using panel data only two indicate positive and statistically significant results: the one of Smarzynska (2004) with evidence for backward spillovers and the other made by Driffield *et al*. (2002) searching for forward spillovers.

As firm-level studies of particular countries (*e.g*., Germidis, 1977; Aitken and Harrison, 1999) fail in finding evidence of positive technology spillovers from foreign firms to domestically owned ones, several authors (*e.g.*, Wheeler and Mody, 1992; Harrison, 1996, Aitken *et al*., 1997; De Mello, 1997; Carkovic and Levine 2002) are driven to conclude that there are no reasons to believe that FDI accelerates overall economic growth. However, the absence of spillovers verified at a micro level is not sufficient to discard the positive influence of FDI on TFP. This non-appearance may be due to several reasons, among them, for example, it can be the result of a selection bias or the influence of FDI on TFP can be the resulting effect of reallocation of resources and not the result of spillovers measured at the firm level. In these cases a macroeconomic analysis is more adequate to show the global effect of FDI.

At a macroeconomic level, the influence of FDI on the host country growth has been studied through wide-ranging cross-country studies in which the rate of growth of real GDP or GDP per capita is related to the inward flow (or stock) of FDI. A pioneer study in using aggregate FDI flows for a broad cross-section of countries was done in the context of the convergence controversy and has put forward a positive role for FDI in generating economic growth (Blomström, *et al*., 1994). These authors have found that, among developing countries, from 1960 to 1985, ratios of FDI inflow to GDP in a five-year period were positively related to growth in the subsequent five-year period. However, when the authors divide the sample in two groups this positive effect only remains for countries with the higher income within the group of the developing countries.

Other empirical investigations have shown that FDI promoted growth only in some groups of countries, and have put forward some possible explanations for that fact: the growth effects of FDI are associated to contexts where there is higher capability of absorption, where some specific type of trade policy was implemented, or where education has attained some minimum threshold (de Mello, 1997; Kottaridi, 2005). In relation to the host country trade policies Bhagwati (1978), suggested that the growth effects of inward FDI could be positive or adverse, depending on the type of trade policy used: the effectiveness of FDI in encouraging growth would be improved by an export promotion policy and diminished by an import substitution policy[10]. Other authors emphasize the capability of absorption (De Gregorio, 1992). In some studies, as in the extensively quoted paper of Borensztein *et al*. (1998), which use a dataset of FDI flows from industrialized countries to sixty-nine developing countries, it is found that FDI is an important vehicle for transferring technology and higher growth only when the host country has a minimum threshold of highly educated workforce that allows it to exploit FDI spillovers[11]. This explanation

10 Testing this hypothesis Balasubramanyam *et al*. (1996) have concluded that only in developing countries that have used an export promotion policy, higher inward FDI flows were associated with faster growth.
11 Equally, Xu (2000) using data on U.S. MNCs finds that a country needs to reach a minimum human capital threshold in order to take advantage of the technology transfer from MNCs.

is partly endorsed by Alfaro *et al*. (2004), in a regression analysis for the period 1975-1995 as a whole, who have found little support to the idea that FDI has an exogenous positive effect on economic growth, but they show that host country conditions, such as the level of education, play an important role in allowing the positive effects of FDI to materialize. But these authors add another factor, which allegedly play a key role: the development of local financial markets.

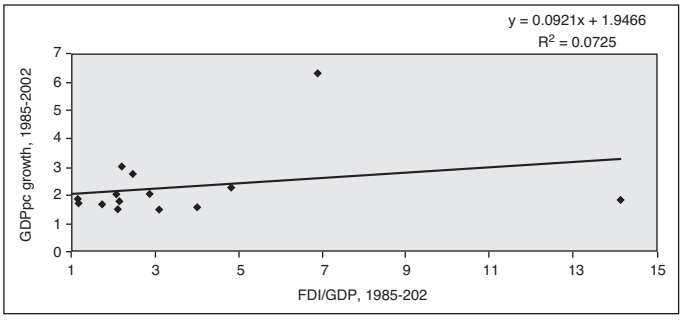
Recently Carkovic and Levine (2002), using new statistical techniques and two new databases to re-examine the relationship between economic growth and FDI, find that FDI inflows do not exert an independent influence on economic growth over the entire period, 1960-1995. Hence, these authors put forth that FDI inflows show no more than sporadically significant effects in five-year periods. They also find that none of the variables used in precedent studies consistently determine the effect of FDI on growth, even if some of them are statistically significant in some recipe of controlling variables.

To sum up, as with the microeconomic studies of productivity spillovers, those of the effects of FDI inflow on economic growth are inconclusive. Almost all find positive effects in some periods, or among some groups of countries[12], in some specifications, but one cannot say from these studies that there are universal impacts of FDI on growth (Lipsey, 2002). So, it is time to try to identify the FDI effects on productivity with a different framework. One empirical finding is particularly important in the context of our present work. De Mello (1999), in a panel data study of aggregate country effects, found that FDI inflows raised growth through different channels, according the country's level of development. In developed countries, FDI inflows increased TFP growth, but not fixed investment, while in developing countries it raised fixed investment, but not TFP growth. Our paper searches the relationship between FDI and TFP in a sample of developed countries.

4. Descriptive framework

How does one come to know whether the amount of inward Foreign Direct Investment measures anything interesting concerning diffusion of technology? One possibility is to search the effects of FDI on economic growth, adopting a cross-section view as in figure 1. In this figure we scatter the FDI/GDP ratio against GDP per capita growth. This is the most commonly used approach to carry out macroeconomic FDI empirical tests (De Gregorio, 1992; Balasubramanyam, *et. al.,* 1996; Borensztein, *et al*., 1998; Alfaro, *et al*., 2004). The main reasons alleged for this procedure are that FDI can offer an incentive to competition, bring with it savings and innovation, increase capital formation, and through these effects lead to job creation and economic growth.

Figure 1 – FDI and growth of GDP per capita



Source: World Bank (2004).

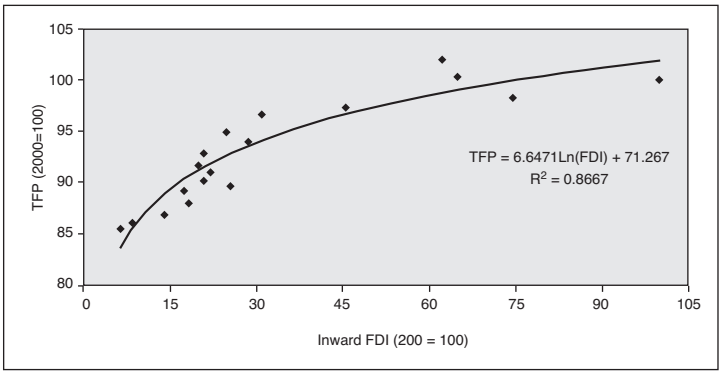
12 For example. Campos and Kinoshita (2002) find positive growth effects of FDI among 25 Central and Eastern European and former Soviet transition countries.

However, this is not a consensual outcome of theory. For example, Brecher and Diaz-Alejandro (1977), Brecher (1983), Boyd and Smith (1992), predict that in the presence of pre-existing trade, price, financial, and other distortions FDI will harm resource allocation and decrease growth. Accordingly, theory produces ambiguous predictions about the growth effects of FDI and therefore an econometric assessment of growth effects must control all the above-mentioned, and possibly other, distortions. So, figure 1 shows that the correlation between FDI and per capita growth is near zero. This is not an unexpected result, given the empirical evidence reviewed in the previous section.

The path we follow in this work is not to search correlations between FDI and GDP per capita growth rate (which depends on a lot of factors, with some of them demand driven, if rates are calculated in five years periods as is usual in empirical studies) but to look for correlations between FDI and TFP. Figure 2 depicts data on average of both inward FDI and TFP in a sample of 17 OECD countries from 1985 to 2002[13]. Like figure 1, Figure 2 also shows the equation of regression and the adjusted trend line.

As is apparent from the comparison between figure 1 and figure 2 the correlation is much more significant between FDI and TFP than between FDI and the growth rate of GDP per capita. Furthermore, figure 2 shows that the association between FDI and TFP is likely to be not linear: in general an increase in the index of FDI is associated to an increase in the index of TFP, though the latter increases with a decreasing rate. Of course figures 1 and 2 are not directly comparable, because while in the former each point corresponds to a pair of FDI and GDP per capita growth for a given country, in the latter each point corresponds to the sample average of FDI and TFP in a given year, and there is a lot of heterogeneity among countries. But our data show that the association between FDI and TFP is stronger than the association between FDI and the growth rate of GDP per capita at the country level, too[14].

Figure 2 – Association between TFP and Inward FDI



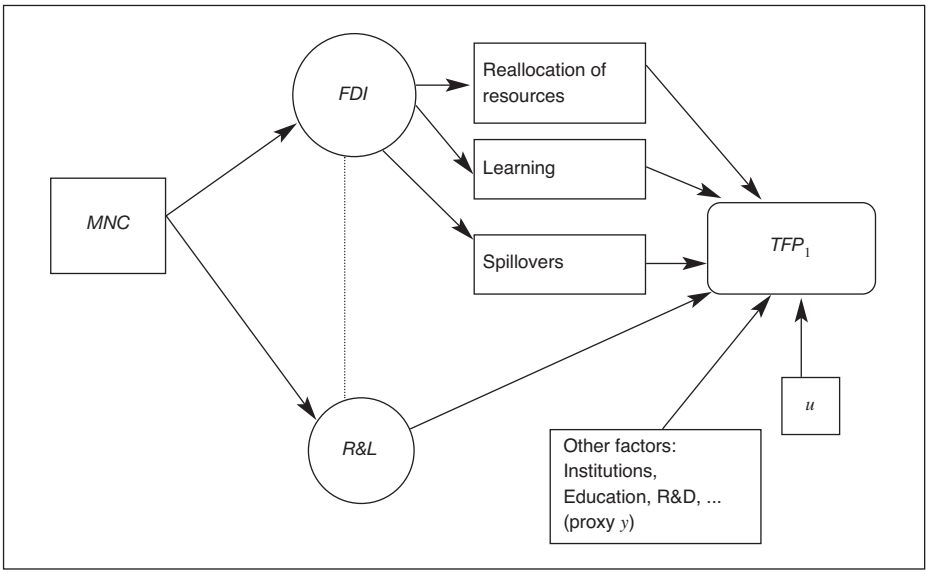
Source: World Bank (2004) for FDI and OECD (2004) for TFP

What are the reasons to hypothesize that an increase in inward FDI is associated to an increase in TFP, and to think that this relationship must be stronger than the one between FDI and GDP per capita growth? In order to answer this question, it is useful to introduce figure 3, which basically allows a more detailed discussion of the underlying assumptions of correlations between FDI and TFP.

13 The economies used in our empirical tests (see section 5) plus Denmark.
14 In effect, the simple regression of FDI on TFP at a country level for the 1985-2002 period shows positive and statistically significant estimates in all countries of our sample only with an exception: Spain.

Figure 3 illustrates the main influences in TFP associated to foreign multinational corporations (MNCs), which we shall review. We assume that through FDI and royalties and license fees (R&L) MNCs contribute to international diffusion of technology and this affects Total Factor Productivity (TFP_t) of a host country in t period because it promotes learning[15] and generates spillovers. But figure 3 depicts another FDI impact on TFP: the reallocation of resources, a channel to which scarce attention has been paid in empirical literature.

Figure 3 – FDI and TFP



As we have seen in section 2, the most common theoretical justification for the existence of spillovers is that overseas investors possess certain advantages that local firms can capture through their interactions with foreign affiliates. For example, local firms may learn with the practice of MNCs or MNCs may transfer technology and know-how to local suppliers in order to improve the quality of inputs (Rodriguez-Clare, 1996). Alternatively, local competitors might benefit by attracting employees with firm-specific knowledge from the foreign affiliate (Fosfuri *et al*., 2001). Local firms that are customers of the foreign affiliate may also benefit from the supply of more sophisticated intermediate inputs. For the host country, these benefits derive from the intentional development of forward and backward linkages between local and foreign firms, as well as from the accidental leakage of knowledge and technology to competing and unrelated firms.

There is a huge literature highlighting the central role of MNCs as producers and disseminators of knowledge and technological innovations (*e. g.,* Teece, 1977; Cantwell, 1989; Narula, 2003)[16]. A vast literature also deals with ownership, location and internalization — the OLI paradigm developed by Dunning (1981, 1988) — and with the knowledge capital model (Markusen and

15 Domestic firms can learn the MNCs' superior production or management techniques through several ways including observation (demonstration effects).
16 This knowledge-asset view is also supported empirically. For example, multinationals are much more R&D-intensive than purely domestic firms (Griffith, 1999).

Maskus, 1999) and horizontal multinational firms. This literature explains the internalization decision between FDI and licensing. In some cases MNCs transfer technologies of new vintages through direct investment and they license or transfer their older technologies through joint ventures (Mansfield and Romeo, 1980). But, because MNCs often choose to internalize some part of their transactions, their importance for an economy can be underestimated if we rely only on the figures of FDI flows as carriers of international diffusion of technology, as is usually done by the empirical literature on the growth effects of FDI. In other words, MNCs may influence the TFP of a host country through FDI, trough licensing or both.

The most significant part of the literature on MNCs emphasizes technology as a driving force for the internationalization of the operations of such firms. As powerful as technology might be in driving the internationalization of firms, it is not the only intangible asset that firms may seek to exploit worldwide. Patents and copyrights can convey obvious competitive advantages to the firm that holds them. In some industries, the assets are in the form of brand names for which worldwide consumers are willing to pay a premium (for example, cola beverages). Firms owning such assets can, of course, license country-specific production rights, rather than choose to invest in foreign production facilities. Accordingly, it is because FDI flows alone do not tell the whole story about the transfer of technology and other intangible assets that we introduce, among the determinants of TFP in figure 3, the royalties and license fees (R&L) paid by the authorized use of intangible, non-financial, non-produced assets and proprietary rights, such as patents, copyrights, trademarks, franchises and industrial processes. The inclusion of R&L is also adequate to show that the possible ways in which the advantages possessed by MNCs can be internationally transferred fall between the two extremes represented by licensing to an independent firm, an example of arms length transaction, and intra-firm transfer.

From an MNC viewpoint, both modes of transfer have their advantages and disadvantages. For example licensing to an independent firm will be convenient when the licensor lacks some assets other than the intangibles which are required for FDI, such as capital, a widespread trade and support network, and so on. Also, licensing decreases the risks deriving from political changes in the host-country (for example expropriation), and it has a shorter lead-time than starting a subsidiary from the beginning. On the other hand intra-firm transfer avoids any outflow of technology to other firms and it is more favourable than licensing when arms length transactions are complex and difficult to enforce. Another situation in which intra-firm transfer is convenient with respect to licensing arises as a consequence of the transfer process itself, namely in the presence of no negligible transfer costs[17]. A policy designed to attract FDI must have these points in consideration, because they condition the amount of incentives offered and the resulting effect on TFP.

Productivity spillovers may take place when local firms improve their efficiency by copying technologies of foreign affiliates operating in the local market either based on observation or by hiring workers trained by the affiliates. These are knowledge spillovers in nature. Another kind of spillovers occurs if multinational entry leads to more severe competition in the host country market and forces local firms to use their existing resources more efficiently or to search for new technologies (Blomström and Kokko, 1998). Although this distinction must be done for policy purposes because knowledge spillovers present a rationale for government action to subsidise FDI inflows, and such rationale is absent when the improved productivity of local firms is due to increased competition, for our general endeavour such distinction may be negligible. Equally negligible, for our present purposes, is the division between horizontal and vertical spillovers, although MNCs behaviour may be different in face of each one of them, if MNCs try to prevent technology leakage and spillovers from taking place. In general, this goal can be achieved through diverse ways: formal protection of their intellectual property, trade secrecy, paying higher wages or locating in countries characterized by limited imitative capacities of their domestic firms.

17 Teece (1977) showed that transfer costs can be considerable and can in particular situations account for more than 20 per cent of the cost of developing the technology.

However, to prevent the occurrence of spillovers through these instruments is much easier in the case of horizontal spillovers.

Of course that other than the above mentioned factors affect TFP_t: institutions (Narula and Dunning, 2000; Mudambi and Navarra; 2002), skilled labour (Coughlin and Segev, 2000; Keller, 1996), technological capacity (Glass and Saggi, 1998), infrastructures (Coughlin *et al*., 1991) developed financial markets (Alfaro *et al*., 2004), and so on. All these factors are associated to the level of development that country i enjoys. So, we assemble all these factors in a single variable, y_{it}, termed level of comparative development. The inclusion of these factors in a single *proxy* is also justified because there is not an unambiguous theory of the comparative importance of each one of those factors. One the other hand, assembling all those factors in a single proxy allows us to focus our attention on the FDI and R&L variables, which are the main endeavour of our investigation. Furthermore, the importance of comparative development is highlighted in several analyses of individual host countries and in various statistical analyses. Although that importance had been more stressed in the context of developing countries (e.g., Blomström *et al.,* 1994; Balasubramanyam, 1998) than for advanced countries, a lot of other evidence points to the importance of comparative level of development. For instance, among the studies employing panel data that report unambiguous positive FDI effects the vast majority is related to developed countries: United States (Keller and Yeaple, 2003); United Kingdom (Liu *et al*., 2000; Haskel *et al*., 2002); Ireland (Görg, and Strobl, 2003). In contrast, studies using firm level panel data find evidence of negative effects only in transition economies and developing countries: Bulgaria (Konings, 2001); Czech Republic (Djankov and Hoekman, 2000); Venezuela (Aitken and Harrison, 1999), for instance.

A number of extreme simplifications were made in drawing figure 3 and in defining the various terms. In the centre of figure 3 there are three unobservable variables: reallocation of resources, learning and (other) spillovers that are affected by *FDI* and that we presume affect TFP_t. But, as reallocation, learning and spillovers are very difficult to measure we shall compute the effect of the activities behind such variables. Figure 3 is a statistical descriptive framework rather than a 'theory' of FDI. It indicates that fact by adding an error u to the determinants of TFP, in this way making the explanatory variables imperfect measures of TFP[18]. For example, FDI and *R&L* are taken as exogenous but, if as it is likely, *FDI* is correlated with *R&L*, then one might expect feed back in subsequent periods, making the relationship between TFP and FDI much more complex. So, what is depicted in figure 3 is, at best, a very crude reduced-form-type relation whose theoretical underpinnings have still to be worked out. But one has to start someplace. Nevertheless, figure 3 does provide a scheme to discuss much of the research in the effects of FDI and to estimate some relevant elasticities. It is also a first step towards the test of the Internalization thesis, a subject that has received scant attention in empirical literature. To our knowledge, ours is the first work to search simultaneously the empirical effects of FDI and R&L on TFP.

5. Empirical tests: FDI and R&L

Our empirical work estimates the impact of FDI and R&L on TFP, for the period between 1985 and 2002, in a panel data of 16 OECD economies: Australia, Belgium, Canada, Finland, France, Germany, Greece, Ireland, Italy, Netherlands, Japan, Portugal, Spain, Sweden, United Kingdom, and United States. We are conscious about the reduced number of countries, but the lack of TFP data prevents the use of a more extended sample. In the empirical tests we use the following variables and data, for the country i and the time t.

18 A 'theory' would have to be explicit about the conditions (economic, technological, and legal) under which the benefits of MNCs are transformed in TFP. Such a theory would start with the underlying notions of learning and spillovers and with the more precise mechanism driving the effects of FDI on TFP and likely feedbacks. Furthermore, a theory would have to be able to give an unambiguous explanation to the patent change in attitude towards FDI over the last couple of decades, as most countries have liberalized their policies to attract foreign investments.

Table 1 – Variables and data sources

	Variables	Data sources
TFP_{it}	Total factor productivity;	OECD (2004)
FDI_{it}	Foreign Direct Investment, net annual inflows[19];	World Bank (2004)
RL_{it}	Annual payment to the exterior of R&L[20];	World Bank (2004)
Y_{it}	GDP per capita of country i over USA GDP per capita (PPP at constant 1995 international $).	World Bank (2004)
RRD_{it}	Number of Researchers in R&D per million of active people, expressed in full-time equivalent (FTE).	World Bank (2007)

Notes: TFP was calculated by the OECD for the purpose of international comparisons and it is based on harmonized prices for ICT capital goods (OECD, 2004). The annual values of FDI and RL, are from World Development Indicators and are collected from Balance of Payments at current US$ (World Bank, 2004). TFP, FDI and RL are converted by us to index numbers (base year = 2000).

In order to begin, it is helpful to write down the simplest possible model that might connect the first three variables of table 1, in natural logarithm form, in the spirit of figure 3:

$$LnTEP_{it} = \beta_{0it} + \beta_1 LnFDI_{it} + \beta_2 LnRL_{it} + u_{it} \qquad (1)$$

Equation (1) is formalized assuming that TFP is independent of the level of comparative development of country i. The equation is estimated in a panel of data (an unbalanced panel because of some missing values for the variables), using some distinct techniques. As it is well acknowledged, the key advantage of the panel data models is its increased precision of the estimators, in confront with cross-section models. However, in panel estimation, the number of underlying assumptions grows correspondingly. In fact, equation (1) can be estimated by several methods, according to assumptions about: i) the intercept (β_{0it}); and ii) the relationship between the independent variables and the error term (u_{it}).

Respecting to i) and apart from the case of no intercept, that is $\beta_{0it} = 0$, three alternative specifications can be built, based on three key assumptions: a) identical intercept for all countries ($\beta_{0it} = \beta_0$), which results from an estimation technique with a common constant; b) a different intercept estimated for each country — $\beta_{0it} = \beta_{0i}$, with $E(\beta_{0i} u_{it}) \neq 0$ — the well-known fixed effects estimator; c) intercepts estimated as random variables across countries.— $\beta_{0it} = \beta_0 = v_i$ with $E(v_i u_{it}) = 0$, that is, assuming that and errors are uncorrelated) — the so called random effects model.

Usually, it is assumed the superiority of models with Fixed Effects (FE) and models with Random Effects (RE) over pooled models with a common constant. On the other hand, researchers typically apply some tests as a means of disentangling, between FE and RE, the most accurate method to deal with the considered data. In this paper, we shall follow a different path owing to three main reasons.

19 Foreign direct investment is net inflows of investment to acquire a lasting management interest (10 per cent or more of voting stock) in an enterprise operating in an economy other than that of the investor. It is the sum of equity capital, reinvestment of earnings, other long-term capital, and short-term capital as shown in the Balance of Payments.

20 *Royalties* and License Fees are payments between residents and non-residents for the authorized use of intangible, non-produced, non-financial assets and proprietary rights (such as patents, copyrights, trademarks, industrial processes, and franchises) and for the use, through licensing agreements, of produced originals of prototypes (such as manuscripts and films).

First, we need to know if there is some non-observable country heterogeneity, that is, heterogeneity in the explanation of *TFP* variation not controlled by the independent variables. This implies the need of admitting the absence of such heterogeneity, which is made in estimation models with a common constant. This means to discard the *a priori* assumption of FE and RE superiority, and to choose between a common constant's model for all economies and a model based on specific fixed effects for the cross-section of countries. Accordingly, we will test these two hypotheses.

Second, there is a lot of evidence pointing to the fact that unobservable and time constant factors are, at least partially, associated to the locational characteristics of the country (Wheeler and Mody, 1992; Brainard, 1997). Such locational factors can determine a geographical dimension of positive vertical spillovers (Sjöholm, 1999; Kugler, 2000), and can also be associated to other less time-variant characteristics of the host country like policy preferences[21]. So, this evidence points to the fact that country heterogeneity is not random.

Third, our panel is unbalanced, what means that some observations are missing so that T_i varies among cross-sections. In this case, RE model has an important drawback relatively FE: typically, computer econometric packages (as is our case with program Eviews 3.1) use the largest value of T_i in computing the variance estimates. This procedure is consistent only if the number of missing observations is asymptotically negligible. Obviously, this is not our case, where are missing 47, and 115 when human capital is introduced, in a balanced panel number of 288. So, for all the above reasons, we do not consider the RE hypothesis[22]. In the present paper, we will begin by estimating a model with a common constant and another with fixed effects and, after that, we will apply a test in order to decide which one provides more reliable estimates.

Respecting to ii), that is the relationship between the dependent variables and the error term (u_{it}), we will calculate estimates using three different methods, according some assumptions of the classical least squares method (homocedasticity, no serial or contemporaneous correlation) are respected or violated. Consequently, in table 2 we present the estimates calculated by system OLS (Ordinary Least Squares) method, whereas the subsequent tables show estimates that result from GLS (Generalized Least Squares) and SUR (Seemingly Unrelated Regression).

As above-mentioned, in table 2 we show estimates that are calculated by system OLS. Some of them are based on a regression with a common constant (columns 1, 2, 3 and 4), whereas others (columns 1', 2', 3' and 4') are resulting from a fixed effects model, that is, with a specific constant for each one country.

With a common constant, the estimates presented in column 1 show statistically significant positive elasticities of *TFP* in order to both *FDI* and *R&L*: 1 percent increase in FDI is associated with an increase in TFP of 0.029 per cent, and 1 percent increase in R&L implies a percent increase in TFP of 0.054. The t tests indicate that the coefficients are significant at the 1 per cent level and the adjusted coefficient of determination ($\bar{R}^2$) shows that the equation (1) explains more than fifty percent of the TFP variation.

However, the estimates shown in column (1) may be biased owing to the assumption implicit on the estimation method that the behaviour of the economy is time and cross-section invariant. But if the behaviour of the OECD economies varies in both dimensions, one form of getting away the total homogeneity of time and country behaviour assumed is to admit that elasticities are equal in every sample economy, but that there is some heterogeneity embraced by the constant in the regression, which becomes specific to each one of the countries. In this procedure, known as

21 For instance some countries, like Japan, have a long history of aggressively encouraged licensing of foreign technology at the same time that they restrict the operation of wholly owned subsidiaries of MNCs while others have had a much more non-discriminating behaviour.
22 The random effects approach has other considerable shortcoming. For instance, there is little justification for treating the individual effects as uncorrelated with the other regressors.

Table 2 – Impacts of FDI and R&L on TFP, OLS estimation

	(1)	(1')	(2)	(2')	(3)	(3')	(4)	(4')
Common constant	4.22* (123.4)	---	4.07* (187.47)	---	4.132* (41.89)	---	2.993* (14.581)	---
F. Effects	---	a)	---	a)	---	a)	---	a)
LnFDI	0.029* (7.93)	0.024* (7.60)	0.095* (21.79)	0.063* (9.37)	0.0299* (7.84)	0.023* (8.68)	0.460* (8.998)	0.115** (2.16)
LnRL	0.054* (7.38)	0.056* (8.72)	0.095* (18.12)	0.076* (12.45)	0.052* (7.71)	0.041* (9.47)	-0.0144† (-0.2294)	-0.117** (-2.33)
LnFDI*LnRL	---	---	-0.017* (-14.66)	-0.009* (-6.26)	---	---	---	---
Lny	---	---	---	---	0.022† (1.108)	0.325* (9.18)	0.295* (6.0592)	0.234* (4.49)
LnFDI*Lny	---	---	---	---	---	---	-0.1016* (-8.446)	-0.021*** (-1.74)
LnRL*Lny	---	---	---	---	---	---	0.0147† (0.9805)	0.038* (3.10)
T	18	18	18	18	18	18	18	18
N	16	16	16	16	16	16	16	16
Obs	241	241	241	241	241	241	241	241
F test	163.27	1078.02	161.47	605.91	109.206	714.022	94.93	362.95
$\bar{R}^2$	0.58	0.83	0.67	0.85	0.58	0.87	0.67	0.87
Joint significance F	$F_{15,223} = 21.699$		$F_{15,222} = 16.603$		$F_{15,222} = 31.374$		$F_{15,220} = 22.241$	

Source: Calculations based on OECD (2004) and World Bank (2004).

Notes: a) For saving space, the country specific effects are not reported in the table, but they are available from the author on request. t tests are shown in brackets: *significant at the 1 percent level; **significant at the 5 percent level; ***significant at the 10 percent level; †not significant. Standard errors and covariance matrix are White (1980) heteroscedastic corrected.

fixed effects model, individual country effects result from several unobservable and time-constant factors. So, table 2 presents estimates using the fixed effects model for equation (1) also. For space-saving reasons, the country specific effects are not reported in the table, but they are available from the author on request. As predicted, allowing for country specific fixed effects has turned the estimates of the elasticity in order to *FDI* lower and the elasticity in order to *R&L* higher, but both coefficients remain significantly positive. On the other hand the consideration of the fixed effects increase the explicative power of the equation, as measured by $\bar{R}^2$.

An important question, which emerges from figure 3, is the nature of the relationship between FDI and R&L. Are they complements or substitutes in influencing TFP? In order to answer to question we run the regression (1) with an interaction term — *LnFDI*LnRL*. The estimates, included in column 2 e 2', show that all the coefficients are statistically significant and have the predicted sign. Particularly the negative sign of the interaction term shows that FDI and R&L are substitutes in influencing TFP.

However, the estimates shown in columns 1 and 2 and 1' and 2', based on equation (1), may omit some relevant variables. The most obvious candidate is the level of development of the country *i* relatively to the technological frontier, as was depicted in figure 3. This is consistent with models that have addressed the hypothesis that the host country's level of technical development may matter as a starting point (Findlay, 1978; Glass and Saggi, 1998). Furthermore, if spillovers should not be expected in all kinds of industries, the level of technological development of the host country matters. In particular, in countries with a low level of development, foreign MNCs may sometimes operate in 'enclaves', where neither products nor technologies have much in common with those of local firms. In such circumstances, there may be little scope for learning, and for spillovers to occur.

With these considerations in mind, it is now the time to introduce other relevant variables in the framework, as depicted in equation (2):

$$LnTFP_{it} = \beta_{0it} + \beta_1 LnFDI_{it} + \beta_2 LnRL_{it} + \beta_3 Lny_{it} * LnFDI_{it} + \beta_4 Lny_{it} * LnRL_{it} + \beta_5 Lny_{it} + u_{it} \quad (2)$$

Equation (2) adds to equation (1) the level of comparative development of the country *i* in log scale (Lny_{it}) and two interaction terms Lny_{it}*$LnFDI_{it}$ — interaction between the level of development and *FDI*, and Lny_{it}*$LnRL_{it}$ — interaction between the level of development and *RL*. If FDI is a *carrier* to the coming in of technology, we expect that the lower the country's technological level is the larger the positive effects of FDI would be. Hence, resulting a negative sign for coefficient on interaction variable — Lny_{it}*$LnFDI_{it}$. On the other hand, we expect a positive sign of the interaction term between development level and *RL* indicating that the increase of the development level will lead to a larger benefit of technology use licenses, perhaps as a consequence of the improved benefit of complementarity among technologies.

So, table 2 shows also in columns 3 and 3' and 4 and 4' the elasticity's behaviour in face of the level of comparative development and the interaction between this level and the basis variables. The introduced modifications have implied some alterations in the estimated coefficients. Let us begin by the estimation without interaction terms (columns 3 and 3'). Estimates show the expected signs independently of the method of estimation. The elasticity of *TFP* in order to *FDI* and the elasticity of *TFP* in relation to *RL* are both, in general, lower than in previous specifications, but their signs remain positive. The estimates show, whatever the estimation method was, that everything else constant, and without interaction terms 1 percent increase in *FDI* is associated to a percent increase in *TFP*, included in the interval [0.02, 0.03] and 1 percent increase in *RL* is associated to a percent increase in *TFP* ranging between 0.04 and 0.05. However, the level of comparative development elasticity is only statistically significant in the fixed effects model (column 3'). When equation is estimated with a common constant this coefficient does not appears to be significantly different from zero. On the other hand, with the fixed effects model it is significantly positive: *ceteris paribus*, 1 percent increase in the level of development implies an increase in *TFP* of 0.325 percent.

Considering the interaction terms (columns 4 and 4'), the sign of the coefficient of *LnRL* turns out to be negative. The FDI effect decreases with the increase of the level of development, but it is always positive; The *RL* effect is negative for countries with a low level of development (below 21 percent of US level). But overall, the *RL* have an increasing effect over *TFP* as the level of development increases. In our view this means that the level of development is crucial to determine the amount of profit that a country appropriates from the assets that originate the payment of royalties and license fees. On the other hand, the level of development is also important to assess the total effect of FDI on TFP at the aggregate level.

However, there is an important difference between the pooled LS with a common constant and the fixed effects model. In the estimation with a common constant the coefficient of *LnRL* together with the coefficient of the interaction of this variable with the level of development (*LnRL*Lny*) does not appear significantly different from zero. We think that this lack of

significance is due to the restriction imposed by the common constant that limits the possibility of accounting the heterogeneity of the different countries. The fixed effects model can accommodate such heterogeneity in the $\beta_{0it} = \beta_{0i}$ term.

So, in order to come to a sound decision on whether fixed effects or pooled OLS model with a common constant is appropriate, we calculate the Joint significance F test for the null and alternative following hypotheses:

H_0: $\beta_{0it} = \beta_0$, for all $i = 1,\ldots.$, 16 and for all t (no cross-section fixed effects);

$H_1 \beta_{0it} \neq \beta_0$, for all $i = 1,\ldots.$, 16 and for all t (cross-section fixed effects).

As is visible from the bottom line of table 2 the joint significance F test largely exceeds the critical value from the F table for 15 degrees of freedom (df) of the numerator and for 220, or slightly higher, df of the denominator, which is approximately 2.13, at 1 per cent level of significance. So, the results imply that the null hypothesis (that there are no cross-section fixed effects) is rejected.

But, in table 2, the comparison between the columns (3) and (4) and the columns (3') and (4'), respectively is illustrative of a possible problem that can affect the OLS estimates: the existence of heterocedasticity. The heterogeneity not controlled by the fixed effects can make evident that the conditional variance of *TFP* (which is equal to that of u_i) increases as comparative level of development, or *RL*, increases.

In fact, the estimates of Table 2, obtained by system OLS, have implicit the verification of the assumptions of the classic linear regression model. A critical assumption of this model is that the disturbances have the same variance. If this assumption is not satisfied we have heteroscedasticity. Although the heterocesdaticity does not destroy the unbiasedness and consistency properties of the OLS estimators make that these estimators are no longer efficient (have not the minimum variance). In such case the unbiased, consistent and efficient estimators are provided by the method of Generalized Least Squares (GLS). Thus, table 3 reports GLS estimates for the same sample and for the same regression equations previously estimated by OLS.

Table 3 – Impacts of FDI and R&L, GLS estimates (cross section weights)

	(5)	(6)	(7)	(8)
F. Effects	a)	a)	a)	a)
LnFDI	0.021*	0.048*	0.019*	0.129*
	(15.02)	(4.83)	(15.59)	(5.87)
LnRL	0.047*	0.062*	0.040*	-0.097*
	(12.07)	(8.07)	(17.96)	(-4.89)
LnFDI* LnRL	---	-0.0065*	---	---
		(-2.87)		
Lny	---	---	0.406*	0.345*
			(25.69)	(15.80)
LnFDI*Lny	---	---	---	-0.026*
				(-5.05)

to be continued in the nex page

Table 3 – Impacts of FDI and R&L, GLS estimates (cross section weights) ***continued***

	(5)	(6)	(7)	(8)
LnRL*Lny	---	---	---	0.032* (6.73)
T	18	18	18	18
N	16	16	16	16
Obs	241	241	241	241
$\bar{R}^2$	0.99	0.99	0.99	0.99

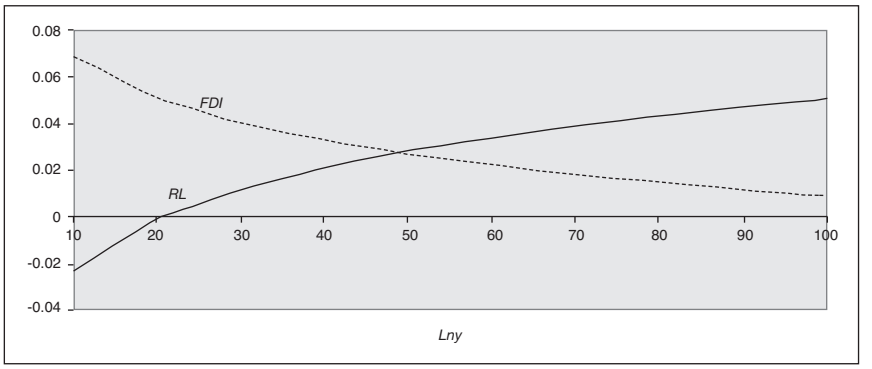
Source: Calculations based on OECD (2004) and World Bank (2004).

Notes: a) For saving space, the country specific effects are not reported in the table, but they are available from the author on request. t tests are shown in brackets: *significant at the 1 percent level. Standard errors and covariance matrix are White (1980) heteroscedastic corrected.

As is visible from table 3 the estimates obtained by GLS have the same signs as the coefficients estimated by OLS, but have a clear advantage in terms of statistical significance: all coefficients estimated by GLS are significant at the 1 percent level.

Based on GLS estimates, Figure 4 shows the association between level of development, FDI inflows and payment of R&L. The *FDI* line and the *RL* line are designed using the estimates of column (8) in table 3. When the level of development increases from 10 percent to 100 percent of US level, the *RL* effect on *TFP* increases from –0.02 to 0.05 and the *FDI* effect on *TFP* decreases from 0.07 to 0.009. So, GLS estimation confirms the above said idea that the level of comparative development must be controlled if we need to access the real impact of *FDI* and *RL* on *TFP*.

Figure 4 – Effects of FDI and RL on TFP, given the level of development



Source: The same as table 3.

The GLS estimation accounts for cross-equation heteroscedasticity by minimizing the weighted sum-of-squared residuals. The equation weights are the inverses of the estimated equation variances. However, this type of GLS (with cross section weights) is not a warranty of the inexistence of autocorrelation, when we are dealing with panels of data. In these cases, autocorrelation can be not only a correlation between members of series ordered in time but also

a correlation associated to observations not spontaneously distributed in space. So, if the possibility of autocorrelation is real, it is more prudent to use another estimation method: the one known as Seemingly Unrelated Regression (SUR)[23]. The SUR method, also called multivariate regression, or Zellner's method, is the feasible GLS estimator when the residuals are both cross-section heteroscedastic and contemporaneously correlated.

So, table 4 reports SUR estimates, that is, estimates that take in account heteroscedasticity, and contemporaneous correlation in the errors across equations. The results are very similar to the ones provided by GLS, both in signs and in absolute value. All the coefficients are statistically significant at 1 percent level and the level of development goes on appearing as crucial to the precise estimation of impacts of *FDI* and *RL* on *TFP*.

Table 4 – Interaction between FDI and R&L, OCDE, 1985-2002. SUR estimation

	(9)	(10)	(11)	(12)
F. Effects	a)	a)	a)	a)
LnFDI	0.022*	0.062*	0.021*	0.127*
	(23.81)	(17.81)	(22.43)	(5.41)
LnRL	0.057*	0.074*	0.041*	-0.124*
	(30.42)	(32.65)	(24.74)	(-6.62)
LnFDI* LnRL	---	-0.0095*	---	---
		(-12.24)		
Lny	---	---	0.333*	0.242*
			(27.10)	(13.77)
LnFDI*Lny	---	---	---	-0.025*
				(-4.58)
LnRL*Lny	---	---	---	0.040*
				(8.94)
T	18	18	18	18
N	16	16	16	16
Obs	241	241	241	241
$\bar{R}^2$	0.81	0.84	0.85	0.86

Source: Calculations based on OECD (2004) and World Bank (2004).

Notes: a) All the regressions are based on a fixed effects model. For simplicity the country specific effects are not reported in the table, but they are available from the author on request. t tests are shown in brackets: *significant at the 1 percent

6. Level of development and human capital

We have previously argued that absorptive capacity of the host country matters. In essence, the idea is that the magnitude of FDI benefits is dependent on the ability and motivation of local firms to learn from the technology and ideas that are spilt by foreign firms. This absorptive capacity is directly related with the human capital stock, as was emphasized by the seminal paper of Nelson

23 See Zellner (1962).

and Phelps (1966) and many others (v.g , Borensztein *et al*., 1998; de Mello, 1997; Ariffin and Figueiredo, 2004; Kottaridi, 2005).

Building on Nelson and Phelps (1966), several authors (for example, Cohen and Levinthal, 1989) have argued that human capital plays a dual role in promoting TFP growth: first, it enables a country to directly influence productivity by determining the capacity to innovate; second, a higher level of human capital increases the capacity of an economy to absorb foreign technology allowing a country to close the gap, between its level of productivity and that of the leading country, faster. So, a more qualified workforce can better not only render FDI more fruitful for foreign investors, but also take advantage of foreign R&D-induced ideas and use capital goods imports (embodying advanced foreign technologies) more effectively (Pessoa, 2008).

Consequently, the positive impact of level of development on *TFP* and its significant interaction with *FDI* and *RL*, reported in the previous section, can be side-products of the effect of human capital. In fact, the human capital in general and particularly the human capital that is more apt to take profit from R&D (Research and Development) increase with the level of development, and so the latter can be acting in the regressions as a *proxy* of the formers. To test this hypothesis, we add to the equation (2) a variable that controls the effect of human capital on *TFP*.

We think that the most adequate variable for controlling the above mentioned dual role in promoting *TFP* is the number of Technicians in R&D, because they are people whose main tasks require technical knowledge and experience in engineering. But, because data on Technicians in R&D are very scarce, we have chosen the number of Researchers in R&D (RRD) per million of active people, expressed in full-time equivalent (FTE), as the feasible *proxy* of the human capital. Researchers in R&D are professionals engaged in the conception or creation of new knowledge, products, processes, methods, or systems and in the management of the projects concerned, and so are an important factor in render the existent technologies more fruitful.

According to the above lines, we estimated regressions based on three methods: OLS, GLS and SUR. The results are presented in table 5, where is also presented , one of the statistical tests introduced by Hausman in his seminal paper (1978). The time span is the same of the previous tables, but the inclusion of the number of researchers, due to the scarcity of data forced to reduce the number of observations (from 241 to 173) of the panel.

Whereas OLS estimation shows that the *RRD* coefficient is not significantly different from 0, along with absence of statistical significance of *FDI* and of the interaction between *FDI* and the level of development, the regressions based on GLS and SUR show that all coefficients have the predict signs and are significant at a 1 percent level. However, the OLS estimates shall be unreliable. They are likely to be affected by the troubles previously detected (heteroscedasticity and autocorrelation), now aggravated by the reduction of the number of observations.

So, discarding OLS estimates, and based only on GLS and SUR methods, we can say that *ceteris paribus* 1 percent increase in *FDI* has a positive effect on *TFP* that can be either 0.12 percent or 0.14 percent, according to the method of estimation. On the other hand, 1 percent increase in *RL* reduces TFP by 0.14 percent. However, this negative effect can be compensated by the interaction with the level of development, which has a positive impact on *TFP*. As predicted, the *RRD* variable has a positive impact on *TFP*, as well as the level of comparative development.

Both GLS and SUR estimates reject the idea that the level of comparative development is only a *proxy* of *RDD* and that the comparative level of development has no direct effect on *TFP*. On the contrary, the inclusion of *RRD* has maintained the significance of the coefficient of the level of comparative development and, what is more, the value of *TFP* elasticity relatively to the level of development only has slightly varied (a small retreating in GLS estimation and slight increase in SUR estimation).

Although the absolute value of the coefficients does not present a large variation form GLS to SUR estimation, we have tried to identify, between the two, the most reliable method, using the

Table 5 – FDI, R&L, level of development and Human capital

	OLS	GLS (cross section weights)	SUR	q
	(13)	(14)	(15)	(16)
LnFDI	0.1070†	0.141*	0.118*	0.02302*
	(1.486)	(4.339)	(3.754)	(2.776)
LnRL	-0.138***	-0.142*	-0.138*	-0.00377†
	(-1.719)	(-3.794)	(-4.608)	(-0.169)
Lny	0.241*	0.309*	0.255*	0.053751*
	(4.000)	(10.22)	(10.586)	(2.949)
LnFDI*Lny	-0.0202†	-0.0290*	-0.0230*	-0.00593*
	(-1.197)	(-3.854)	(-3.165)	(-3.161)
LnRL*Lny	0.0412**	0.0417*	0.0410*	0.000653†
	(2.221)	(4.921)	(5.849)	(0.137)
Ln RRD	0.0358†	0.0455*	0.0393*	0.006155†
	(1.398)	(3.745)	(4.855)	(0.680)
T	18	18	18	18
N	16	16	16	16
Obs	173	173	173	173
$\bar{R}^2$	0.87	0.99	0.87	

Source: Calculations based on OECD (2004) and World Bank (2004, 2007).

Notes: t tests are shown in brackets: *significant at the 1 percent level; **significant at the 5 percent level; ***significant at the 10 percent level; †not significant. In OLS and GLS estimations, standard errors and covariance matrix are White (1980) heteroscedastic corrected. All the regressions are based on a fixed effects model. For saving space the country specific effects are not reported in the table, but they are available from the author on request.

statistic that results from the procedure described in Hausman (1978). Looking at this statistic we saw mixed results. The results showed in the last column are obtained making the subtraction between the coefficients estimated according to GLS and according to SUR methods, and the tests *t* below them are calculated from the difference between the respective variances, as in Hausman (1978). As is visible, the difference between the coefficients estimated by GLS and by SUR estimation is only significant for 3 in 6 variables. Whereas the difference in *FDI*, in the level of comparative development and in the interaction between these two variables is significant at a 1 percent level, the differences respecting to *RL*, *RRD* and interaction between *RL* and level of development are not statistically significant.

7. Concluding remarks

In this study, we have tested the effects of FDI on the aggregate TFP in a panel data of 16 OECD countries in the 1985-2002 period. Our empirical tests show that inward FDI has a positive impact on host country TFP, possibly because FDI is a channel through which technologies are transferred internationally. This conclusion corroborates both the ones of some microeconomic studies that use panel data, like the works of Liu *et al*. (2000), Haskel *et al*. (2002), Keller and Yeaple (2003) and Görg and Strobl (2003), and also the general conclusion of

Mullen and Williams (2005), at the regional level. Accordingly our work provides support to the countries that aggressively use policy to attract investments from foreign multinational corporations. Our work also sheds some light on the consequences of the internalization decision, which is in the core of the OLI paradigm. Our data show that, when other factors remain constant, inward FDI and R&L payments are substitutes in influencing positively TFP of the host country.

The inclusion among explanatory variables of one that controls the level of comparative development permits to illustrate the fact that the amount of positive effects of both FDI and R&L is dependent on the level of development of the receiving country, and to confirm the suspicion of de Mello (1999) and the idea of Grossman and Helpman (1991) that the impact of FDI is lower in technological leaders than in laggard countries. Our results also demonstrate that the studies that do not control the level of development through one or several variables (education, R&D outlays, infrastructures) — which are strongly and positively correlated with level of development — may present biased estimates.

When we consider the interaction between the level of development and R&L, the negative effect of R&L on TFP can help to explain the scarce use of patents and other 'intangible' assets as brand names and copyrights by the technological laggards and consequently the scarce technological content of production and exports of these countries. The negative impact of R&L for low levels of development provides some rationality to the behaviour of firms of those countries that invest more heavily in machines and equipment than in paying for ideas. But, dynamically, given the complementarity between development level and R&L, corroborated by positive sign of respective interaction term, the licensed use of foreign technologies can represent a means of technical renovation even in countries that are not near the technological frontier.

Finally, our study shows that the positive effect of the level of development on *TFP* is not exclusively due to the human capital. Even though the human capital has a direct effect on TFP, it is not a mere *proxy* of the country's level of development, and so studies that aim to assess the precise impact of FDI on TFP must be aware of its inclusion in the analysis.

References

Aitken, B.; Harrison, A. (1999) Do Domestic Firms Benefit from Foreign Investment? Evidence from Venezuela, *American Economic Review*, 89, 605-618.

Aitken, B.; Hanson, G.; Harrison, A. (1997) Spillovers, Foreign Investment, and Export Behavior, *Journal of International Economics*, 43, 1-2, 103-32.

Alfaro, L. *et al.* (2004) FDI and economic growth: the role of local financial markets, *Journal of International Economics* 64, 89-112.

Ariffin, Norlela; Figueiredo, Paulo N. (2004) Internationalization of Innovative Capabilities: Counter-Evidence from the Electronics Industry in Malaysia and Brazil, *Oxford Development Studies*, 32, 4, 559-583.

Balasubramanyam, V. N. (1998) The MAI and Foreign Direct Investment in Developing Countries, *Discussion Paper EC10/98*, Lancaster University.

Balasubramanyam, V. N.; Salisu, M.; Dapsoford, D. (1996) Foreign Direct Investment and Growth in EP and IS Countries, *Economic Journal* 106, 92-105.

Blomström, M.; Kokko, A. (1998) Multinational Corporations and Spillovers, *Journal of Economic Surveys*, 12, 2, 1-31.

Blomström, M.; Kokko, A. (2003) The Economics Of Foreign Direct Investment Incentives, *Working Paper n. 9489*, National Bureau Of Economic Research.

Blomström, M.; Persson, H. (1983) Foreign Investment and Spillover Efficiency in an Underdeveloped Economy: Evidence from the Mexican Manufacturing Industry, *World Development,* 11(6): 493-501.

Blomström, M.; Lipsey, R.E.; Zejan, M. (1994) What explains the Growth of Developing Countries? In: William J. Baumol, Richard R. Nelson and Edward N. Wolff (eds.), *Convergence of Productivity: Cross-National Studies and Historical Evidence*, New York, Oxford University Press, 243-259.

Borensztein, E.; De Gregorio, J.; Lee, J.W. (1998) How does foreign investment affect growth?, *Journal of International Economics*, 45, 115-35.

Boyd, J. H.; Smith, B.D. (1992) Intermediation and the Equilibrium Allocation of Investment Capital: Implications for Economic Development, *Journal of Monetary Economics*, 30, 409-432.

Brainard, S. L.; (1997) An Empirical Assessment of the Proximity-Concentration Trade-Off between Multinational Sales and Trade, *American Economic Review*, 87, 4, 520-44.

Brecher, R.; Diaz-Alejandro, C. (1977) Tariffs, Foreign Capital and Immiserizing Growth, *Journal of International Economics*, 7, 317-322.

Brecher, R. (1983) Second-Best Policy for International Trade and Investment, *Journal of International Economics*, 14, 313-320.

Campos, N.F.; Kinoshita, Y. (2002) FDI as technology transferred: some panel evidence from transition economies, *The Manchester School*, 70, 3, 398-419.

Cantwell, J. (1989) *Technological Innovation and Multinational Corporations*, Oxford: Basil Blackwell.

Carkovic, M.; Levine, R. (2002) Does Foreign Direct Investment Accelerate Economic Growth?, University of Minnesota, *Working Paper*.

Carr, D.L.; Markusen, J.R.; Maskus, K.E. (2001) Estimating the Knowledge-Capital Model of the Multinational Enterprise, *American Economic Revie*w, 91, 3, 693-708.

Castellani, D.; Zanfei, A. (2006) *Multinational Firms, Innovation and Productivity*, Cheltenham, Edward Elgar.

Caves, R. E. (1996) *Multinational Enterprise and Economic Analysis*, 2nd edition, Chapter 1, Cambridge, Cambridge University Press.

Caves, R. E. (1974) Multinational Firms, Competition and Productivity in Host-Country Markets, *Economica* 41, 162, 176-93.

Chowdhury, A.; Mavrotas, G. (2006) FDI and Growth: What Causes What?, *World Economy*, 29, 1, 9-19.

Chuang, Y.-C.; Hsu, P.-F. (2004) FDI, trade, and spillover efficiency: evidence from China's manufacturing sector, *Applied Economics*, 36, 1103-1115.

Coe, D.; Helpman, E. (1995) International R&D Spillovers, *European Economic Review*, 39, 5, 214-42.

Coughlin, C. C.; Segev, E. (2000) Location Determinants of New Foreign-Owned Manufacturing Plants, *Journal of Regional Science,* 40, 2, 323-51.

Coughlin, C. C.; Terza, J. V.; Arromdee, V. (1991) State Characteristics and the Location of Foreign Direct Investment within the United States. Review *of Economics and Statistics*, 73, 4, 675-83.

De Gregorio, J. (1992) Economic growth in Latin America, *Journal of Development Economics,* 39, 58-84.

De Mello, L. R. (1997) Foreign Direct Investment in Developing Countries and Growth: A Selective Survey, *Journal of Development Studies,* 34, l, l-34.

De Mello, L. R. (1999) Foreign Direct Investment-led Growth: Evidence from Time Series and Panel Data, *Oxford Economic Papers*, 51, 133-151.

Driffield, N. (2001) The Impact on Domestic Productivity of Inward Investment into the UK, *Manchester School,* 69, 1, 103-119.

Driffield, N.; Munday, M.; Roberts, A. (2002) Foreign Direct Investment, Transactions Linkages, and the Performance of the Domestic Sector, *International Journal of the Economics of Business,* 9, 3, 335-51.

Dunning, J. H. (1981) *International Production and the Multinational Enterprise*, London, George Allen, and Unwin.

Dunning, John H. (1988) The Eclectic Paradigm of International Production: A Restatement and Some Possible Extensions, *Journal of International Business Studies*, 19, 1, 1-31.

Evenson, R. E.; Westphal, L. (1995) Technological Change and Technology Strategy, in: Jere Behrman and T. N. Srinivasan (eds.), *Handbook of Development Economics*, vol. 3, Amsterdam, North-Holland.

Findlay, R. (1978) Relative Backwardness, Direct Foreign Investment, and the Transfer of Technology: A Simple Dynamic Model, *Quarterly Journal of Economics,* 92, 1, 1-16.

Fosfuri, A.; Motta, M.; Ronde, T. (2001) Foreign Direct Investment and Spillovers through Workers' Mobility, *Journal of International Economics,* 53, 1, 205-22.

Germidis, D. (1977) *Transfer of Technology by Multinational Corporations*, OECD, Paris.

Gershenberg, I. (1987) The Training and Spread of Managerial Know-how, a Comparative Analysis of Multinational and Other Firms in Kenya, *World Development,* 15, 7, 931-39.

Glass, A.; Saggi, K. (1998) International Technology Transfer and the Technology Gap. *Journal of Development Economics,* 55, 2, 369-98.

Görg, H.; Greenaway, D. (2004) Much Ado about Nothing: Do Domestic Firms Really Benefit from Foreign Direct Investment, *The World Bank Researcher Observer*, 19, 2, 171-197.

Görg, H.; Strobl, E. (2003) Multinational Companies, Technology Spillovers and Plant Survival, *Scandinavian Journal of Economics,* 105, 4, 581-95.

Graham, E.; Krugman, P. (1991) *Foreign Direct Investment in the United States*, Institute for International Economics, Washington D.C.

Griffith, R. (1999) Using the ARD Establishment Level Data to Look at Foreign Ownership and Productivity in the UK. *Economic Journal*, 109, F416-F442.

Grossman, G.; Helpman, E. (1991) *Innovation and Growth in the Global Economy.* Cambridge, MA: MIT Press.

Haddad, M.; Harrison, A. (1993) Are there positive spillovers from direct foreign investment? Evidence from panel data for Morocco, *Journal of Development Economics,* 42, 1, 51-74.

Hanson, G. H. (2001). Should Countries Promote Foreign Direct Investment?, *G-24 Discussion Paper Series*, UNCTAD.

Harrison, A. (1996) Determinants and Consequences of Foreign Investment in Three Developing Countries, in Mark, Roberts; Tybout, James (eds.), *Industrial Evolution in Developing Countries: Micro Patterns of Turnover, Productivity, and Market Structur*e, Oxford: Oxford University Press.

Haskel, J. E.; Pereira, S. C.; Slaughter, M. J. (2002). Does Inward Foreign Direct Investment Boost the Productivity of Domestic Firms?, *NBER Working Paper n. 8724*, National Bureau Of Economic Research.

Katz, J. M. (1969) *Production Functions, Foreign Investment and Growth*, Amsterdam, North Holland.

Keller, W. (1998) Are international R&D spillovers trade related? Analyzing spillovers among randomly matched trade partners, *European Economic Review,* 42, 1469-1481.

Keller, W.; Yeaple, S. R. (2003) Multinational Enterprises, International Trade, And Productivity Growth: Firm Level Evidence From The United States, *NBER Working Paper n. 9504*, National Bureau Of Economic Research.

Keller, W. (1996) Absorptive Capacity: Understanding the Creation and Acquisition of Technology in Development, *Journal of Development Economics,* 49, 199-227.

Kholdy, S. (1995) Causality between Foreign Investment and Spillover Efficiency, *Applied Economics,* 27, 745-49.

Kokko, A. (1996) Productivity Spillovers from Competition between Local Firms and Foreign Affiliates, *Journal of International Development,* 8, 517-530.

Konings, J. (2001) The Effects of Foreign Direct Investment on Domestic Firms: Evidence from Firm Level Panel Data in Emerging Economies, *Economics of Transition,* 9, 3, 619-33.

Kottaridi, C. (2005) The 'core-periphery' pattern of FDI-led growth and production structure in the EU, *Applied Economics,* 37, 99-113.

Kugler, M. (2000) The Diffusion of Externalities from Foreign Direct Investment: Theory Ahead of Measurement, *Discussion Papers in Economics and Econometrics*, University of Southampton, U.K.

Lall, S. (1992). Technological Capabilities and Industrialization, *World Development,* 20, 2, 165-86.

Larrain, F. B.; Lopez-Calva, L. F.; Rodriguez-Clare, A. (2000) Intel: A Case Study of Foreign Direct Investment in Central America, *Center for International Development Working Paper 58*, Harvard University.

Lichtenberg, F.; van Pottelsberghe de la Potterie, B. (1996) International R&D spillovers: A re-examination, *NBER working paper n. 5668*, National Bureau Of Economic Research.

Lim, L.; Fong, P. E. (1982) Vertical Linkages and Multinational Enterprises in Developing Countries, *World Development,* 10, 7, 585-95.

Lipsey, R.E. (2002) Home and Host Country Effects of FDI, *NBER Working Paper n. 9293*, National Bureau Of Economic Research.

Liu, X. *et al.* (2000) Productivity Spillovers from Foreign Direct Investment: Evidence from UK Industry Level Panel Data, *Journal of International Business Studies*, 31, 3, 407-425.

Mansfield, E.; Romeo, A. (1980) Technology Transfer to Overseas Subsidiaries by US-based Firms, *Quarterly Journal of Economics,* 95, 4, 737-50.

Markusen, James (1995) The Boundaries of Multinational Enterprises and the Theory of International Trade, *Journal of Economic Perspectives*, 9, 1, 169-89.

Markusen, James; Maskus, K. (1999) Discrimination among Alternative Theories of FDI, *NBER Working Paper 7164*, National Bureau Of Economic Research.

Moran, T. (2001) *Parental Supervision: the new paradigm for foreign direct investment and development*, Washington, DC, Institute for International Economics.

Mudambi, R.; Navarra, R. (2002). Institutions and international business: A theoretical overview. *International Business Review*, 11, 6, 635-646.

Mullen, J. K.; Williams, M. (2005) Foreign Direct Investment and Regional Economic Performance, *Kyklos*, 58, 2, 265-282.

Narula, R. (2003) *Globalization & Technology*, Cambridge (UK), Polity Press.

Narula, R.; Dunning, John H. (2000). Industrial development, globalization and multinational enterprises: New realities for developing countries. *Oxford Development Studies*, 28, 2, 141-167.

OECD (2002) *Foreign Direct Investment for Development: Maximising Benefits, Minimising Costs*, OECD, Paris.

OECD (2004) Multifactor productivity based on harmonized prices for ICT capital goods-data: indices and percentage changes, *OECD Productivity database*, 17 December.

Pessoa, A. (2008) MNCs, FDI and Host Country Productivity: A Theoretical and Empirical Appraisal, *The ICFAI University Journal of Managerial Economics*, 6, 4, 25-66.

Rhee Y. W.; Belot, T. (1990) Export Catalysts in Developing Countries: A Review of Eleven Success Stories, Washington, DC, World Bank, *Discussion Paper n. 72*.

Rivera-Batiz, F.; Rivera-Batiz, L. (1990) The effects of direct foreign investment in the presence of increasing returns due to specialization, *Journal of Economic Development*, 34, 2, 287-307.

Rodriguez-Clare, A. (1996) Multinationals, Linkages, and Economic Development, *American Economic Review,* 86, 4, 852-73.

Romer, P. (1993) Idea gaps and object gaps in economic development, *Journal of Monetary Economics,* 32, 3, 543-573.

Sjöholm, F. (1999) Productivity Growth in Indonesia: The Role of Regional Characteristics and Direct Foreign Investment, *Economic Development and Cultural Change*, 47, 3, 559-584.

Smarzynska, B. K. (2004) Does Foreign Direct Investment Increase the Productivity of Domestic Firms? In Search of Spillovers through Backward Linkages, *American Economic Review*, 94, 3, 605-627.

Solow, R. M. (2001) What have we learned from a decade of empirical research on growth? Applying growth theory across countries, *The World Bank Economic Review*, 15, 2, 283-288.

Teece, D. J. (1977) Technology Transfer by Multinational Firms: The Resource Cost of Transferring Technological Know-How, *Economic Journal*, 87, 242-261.

Wheeler, D.; Mody, A. (1992) International Investment Location Decisions: The Case of U.S. Firms, *Journal of International Economics,* 33, I-2, 57-76.

White, H. (1980) A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity, *Econometrica*, 48, 817-838.

World Bank (2004, 2007) *World Development Indicators* (WDI-CDRom).

Xu, B. (2000) Multinational Enterprises, Technology Diffusion, and Host Country Productivity Growth, *Journal of Development Economics,* 62, 477-493.

Zellner, A. (1962) An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests of Aggregation Bias, *Journal of the American Statistical Association,* 57, 348-368.

Time Preference and Cyclical Endogenous Growth in an AK Growth Model

Orlando Gomes* Escola Superior de Comunicação Social, IPL; UNIDE/ISCTE – ERC

resumo

O artigo desenvolve um modelo de crescimento endógeno tipo AK com uma taxa de preferência intertemporal endogenamente determinada. Seguindo a literatura relacionada com este tema, assume-se que o grau de impaciência revelado pelo agente representativo, no que respeita ao consumo futuro, depende do rendimento. Para ser preciso, o modelo proposto estabelece uma relação entre o hiato do produto e a taxa de desconto associada à sequência futura de funções de utilidade. São analisadas quer a dinâmica local quer a dinâmica global. Do ponto de vista da análise local, vários resultados de estabilidade podem ser obtidos, dependendo dos valores de parâmetros. O estudo de dinâmica global permite encontrar ciclos económicos endógenos na circunstância em que o agente representativo falha na consideração de um requisito fundamental de optimalidade. Numa segunda etapa, o modelo é alargado ao papel do lazer e, neste caso, as flutuações endógenas já serão compatíveis com um cenário de completa racionalidade.

* I would like to acknowledge the important comments of an anonymous referee, which helped to substantially improve the contents of the paper. I am also grateful to Luís Cavique for helpful assistance. The usual disclaimer applies.

résumé / abstract

Cet article développe un modèle de croissance endogène de type AK avec un certain niveau de préférence intertemporelle déterminée endogènement. Dans la littérature se rapportant à cette question, en ce qui concerne la consommation future, il est supposé que le degré d'impatience manifesté par l'agent représentatif dépend du revenu. Pour être précis, le modèle propose d'établir une relation entre l'intervalle du produit et le taux d'escompte associés à la séquence future de fonctions d'utilité. Sont analysées à la fois et la dynamique locale et la dynamique globale. Du point de vue de l'analyse locale, plusieurs résultats de la stabilité peuvent être atteints, selon les valeurs des paramètres. L'étude de la dynamique globale permet de trouver des cycles économiques endogènes, quand l'agent représentatif rate une exigence fondamentale de l'optimalité. Dans une deuxième étape, le modèle comprendra aussi le rôle des loisirs et, dans ce cas, les fluctuations endogènes sont déjà compatibles avec un scénario complet de la rationalité.

The paper develops an AK endogenous growth model with an endogenously determined rate of intertemporal preference. Following some of the related literature, we assume that the degree of impatience that is revealed by the representative agent, regarding future consumption, depends on income. To be precise, the proposed framework establishes a link between the output gap and the discount rate attached to the sequence of future utility functions. We analyze both local and global dynamics. From a local analysis point of view, a variety of stability results is possible to obtain, depending on parameter values. The study of global dynamics allows finding endogenous business cycles in the circumstance in which the representative agent overlooks one essential requirement for optimality. On a second stage, the model is extended to include the role of leisure and, in this case, endogenous fluctuations are compatible with full rationality.

Classificação JEL: O41, C61, E32.

1. Introduction

The paper studies the implications of assuming an endogenous rate of time preference when assessing the dynamics of a simple AK endogenous growth model. The proposed setup is developed in discrete time and takes, as the central assumption, the dependence of the utility discount rate on the economy's output gap. We consider that a representative agent chooses her rate of time preference by evaluating how the economy performs relatively to a potential output time trend. Typically, one should expect a low discount rate when the economy performs well (the representative agent becomes more patient) and a high discount rate when the economic performance falls short of its potential (impatience rises).

The model has its inspiration in earlier studies concerning endogenous time preference and growth, in the tradition of Uzawa (1968) and Epstein (1987). Three features distinguish our model from other approaches. First, we focus solely on endogenous growth, i.e., relevant variables (output, capital and consumption) will grow at a same constant rate in the long term (see Dolmas, 1996, Drugeon, 1996, and Palivos *et al.*, 1997 for alternative approaches to the endogenous growth – endogenous time preference analysis).

Second, we concentrate solely on the role of output as an influence over the way the future is discounted. Typically, consumption is the central variable, i.e., the discount rate is influenced by the level of consumption; this relation tends to have a positive sign, that is, individuals tend to become increasingly impatient (higher discount rate) with a rise in the present level of consumption. Here, we neglect consumption and take income as the single determinant of the discount rate. In this respect, we follow relevant empirical work by Hausman (1979), Lawrance (1991) and Samwick (1998), who study the relation between utility discounting and income levels; they unanimously agree that the evidence strongly points to a rate of time preference that varies inversely with the agent's income. Building on this evidence, Becker and Mulligan (1997) develop a model where the positive relation between wealth and patience is explored.

Third, we are not directly concerned with the level of income, but with a relative measure: we consider that the discount rate is a function of the output gap. Thus, our focus is not on the statement 'the wealthier are more patient' but with the idea that 'if my present level of wealth is above the expected / potential level, then I am more patient'. This seems a reasonable perspective, in the sense that the representative agent reacts to business cycles and formulates a subjective appreciation of the value of future consumption on the grounds of a more or less optimistic view of the future, which is given by a comparison between effective and potential output. Therefore, a two step procedure is adopted: first, the agent evaluates how the economy performs and this has impact over her sentiment or confidence to the future; second, optimistic sentiment tends to be translated on a more patient attitude towards consumption, while a pessimistic sentiment will lead to a more pronounced impatience. All things considered, we might say that our assumption is basically that patience is procyclical.

The idea that patience is procyclical is also present in Meng (2006), who develops a model of a socially determined discount rate. Under his analysis, indeterminacy is found when, in a single agent intertemporal utility maximization problem, the agent becomes more willing to defer consumption as a result of an increasingly wealthier economy. The indeterminacy result arises if along with the previous living standard – patience relation, it is also established a positive relation between the value of the discount rate and the economy wide level of consumption. Besides the indeterminacy result, the referred relations also allow to measure how lifetime utility is influenced by society. If one takes as reasonable the idea that individual patience rises with aggregate income and falls with aggregate consumption, then it is straightforward to perceive that the agent's lifetime utility rises when the economy becomes wealthier and falls when average consumption rises. In the words of Meng (2006),

"the two conditions imply that the society's living standard is like a public good that yields a positive externality to the agent's utility, and the society's consumption is like a public good that delivers a negative externality" (page 2677).

The effect produced by social consumption can be thought as a jealousy effect. Individuals become eager to consume when the economy's levels of consumption rise.

In the following sections, we concentrate on the role of output fluctuations over time preference and of time preference over long term growth. Consumption effects are overlooked, and there is a coincidence between individual and social income effects, since we are working with a single representative agent. We will consider a generic relation between the output gap and the discount rate, in the sense that we leave open the possibility of positive output gaps producing a positive or a negative effect over such rate. Nevertheless, we will find that interesting dynamic results arise eminently on the reasonable circumstance where the discount rate is countercyclical (i.e., patience is procyclical).

The undertaken analysis is both local and global. On a local perspective, we find that instability tends to persist for most of the admissible parameter values, although saddle-path stability can prevail as well. Local indeterminacy or fixed-point stability (the case in which the two eigenvalues of the Jacobian matrix of the system are inside the unit circle) is absent under the selected parameterization.

From a global dynamics viewpoint, we discover an interesting endogenous business cycles result, which occurs in the presence of the relation of opposite sign between the output gap and the discount rate. At this level, we might argue that the endogenous time preference version of the standard growth model becomes one additional framework of growth in which endogenous fluctuations are likely to arise, as it happens on other growth analyses, namely the ones that consider increasing returns / externalities on production (e.g., Christiano and Harrison, 1999 and Guo and Lansing, 2002), learning (Cellarier, 2006), financial development (Caballé *et al.*, 2006) or competitive environments under extreme conditions regarding the shape of the production function or the constant value of the intertemporal discount rate (e.g., Nishimura *et al.*, 1994, Nishimura and Yano, 1995 and Boldrin *et al.*, 2001).

The relation between time preference and cyclical motion is explored in the literature by Balasko and Ghiglino (1995) and Drugeon (1998). The first authors study the presence of endogenous business cycles on an overlapping generations model. Conventionally, this type of model generates cycles if unrealistically high rates of time preference are assumed. The argument of the authors is that if preferences are not homothetic, it is possible to prove that endogenous fluctuations arise under reasonable levels of impatience. Drugeon (1998), in turn, assumes an endogenous rate of time preference, which depends on the individual level of consumption (negatively) and on the consumption standards of the society as a whole (positively). Combining endogenous impatience with the idea that higher consumption standards of the society imply a greater productivity of its members, Drugeon finds conditions for local indeterminacies and sustained oscillation motion, without the need of assuming increasing returns or any unconventional form for the production function.

The result on endogenous cyclical motion requires a limitation on the capacity of the representative agent to act in a fully optimal way. In particular, one has to assume that the agent is unable to avoid a negative steady state growth rate (consumption and capital fall at a constant rate approaching zero asymptotically). This bounded rationality scenario is the straightforward outcome of the inability of the representative agent to take, alongside with other optimality conditions, a transversality condition able of ruling out undesirable long term results.

On a second stage, we extend the proposed model by considering that the representative agent attributes utility to leisure. Assuming decreasing marginal utility concerning the leisure argument, alongside with decreasing marginal utility of consumption, one encounters an extended version of the benchmark model, where cycles persist for positive leisure utility, although as one increases the relative relevance of leisure the amplitude of the cycles becomes less pronounced.

This extension is relevant because it allows to find cyclical motion in a fully optimal scenario, i.e., when the transversality condition is met and thus long term growth is strictly non negative.

The remainder of the paper is organized as follows. Section 2 describes the model's features. Sections 3 and 4 are dedicated to the study of local and global dynamics. In section 5, we introduce leisure in the utility function and study the dynamic properties of this version of the model. Finally, section 6 concludes.

2. The AK Growth Model with Time Preference Depending on the Output Gap

Consider a standard endogenous growth economy. Variables y_t, k_t and c_t respect to effective output or income, stock of capital and level of consumption in moment t. These variables may be understood as aggregate levels or per capita values, because no population growth is assumed. The capital accumulation constraint has its habitual form, $k_{t+1} = y_t - c_t + (1 - \delta) \cdot k_t$, k_0 given, with $\delta \geq 0$ the depreciation rate. The representative household maximizes consumption utility over time under an infinite horizon, i.e., she maximizes $V_0 = \sum_{t=0}^{+\infty} (\beta_t)^t \cdot U(c_t)$. Function U is a conventional continuous, positive and concave utility function and $\beta_t < 1$ is the discount factor. Concerning utility, we just assume a simple logarithmic function, $U(c_t) = \ln c_t$, and the production process in our endogenous growth framework is just given by an AK function, $y_t = Ak_t$, with $A > 0$ the index of technology.

The single difference between the standard AK model and our framework is that we let the discount factor vary as a function of the economy's output gap, x_t. The output gap is defined as the difference in logs between effective output, y_t, and potential output, y_t^*. This last notion of output coincides with a long run trend, that is, y_t^* is the steady state value of y_t, and therefore the steady state must be characterized by a null output gap, $x^* = 0$.

Let $\rho > 0$ be the discount rate in the absence of deviations of the output relatively to its potential level and consider the following discount factor: $\beta(x_t) = 1/[1 + \rho \cdot (1 + f(x_t))]$, with $f(x_t) = a \cdot x_t$ and a a parameter representing the extent in which the output gap influences the rate of time preference. According to the discussion in the introduction, we might expect a to be negative, i.e., there is, eventually, a relation of opposite sign between the output gap and the discount rate, meaning that the better is the performance of the economy relatively to its potential, the less the representative consumer discounts future utility, that is, the more patient the agent will be. To study the model's dynamics, we do not impose a priori any constraint on the value of this parameter.

An additional assumption is that the representative agent takes decisions today concerning the discount of future utility based on the available information about the output gap, which we consider to be the information of the last period, i.e., $\beta_t = \beta(x_{t-1})$. With the previous problem's specification we may compute, on a straightforward manner, a two endogenous variables – two equations system able to characterize the movement over time of the aggregates capital and consumption. This system is obtained by building a Hamiltonian function and, from this, by computing the first order conditions,

$$H(k_t, c_t, q_t) = U(c_t) + q_{t+1} \cdot [(A - \delta) \cdot k_t - c_t] \tag{1}$$

In equation (1), q_t is the present-value co-state variable of k_t. We may define $q_t \equiv \beta(x_{t-1}) \cdot p_t$ the current-value co-state variable of k_t. Determining the first order conditions, one gets,

$$H_c = 0 \Longrightarrow \beta(x_t) \cdot p_{t+1} = {c_t}^{-1} \tag{2}$$

$$\beta(x_t) \cdot p_{t+1} - p_t = -H_k \Longrightarrow$$
$$(1 + A - \delta) \cdot \beta(x_t) \cdot p_{t+1} + \beta_k \cdot p_{t+1} \cdot [(A - \delta) \cdot k_t - c_t] = p_t \tag{3}$$

with $\beta_k = -\rho \cdot a \cdot \left(\beta(x_t)\right)^2 / k_t$ the derivative of the discount factor in order to k_t.

Resorting to relation (2), one may transform (3) into an equation of motion that characterizes the evolution of consumption over time, i.e.,

$$c_{t+1} = \frac{1+A-\delta-\rho \cdot a \cdot (A-\delta) \cdot \beta(x_{t+1})}{1/(\beta(x_t) \cdot c_t)-\rho \cdot a \cdot \beta(x_{t+1})/k_{t+1}} \tag{4}$$

Equation (4) is derived in appendix A.

Let parameter γ represent the long term growth rate of the various endogenous variables (from the production function, the resource constraint and difference equation (4), it is straightforward to perceive that output, capital and consumption all grow at a same steady state growth rate). It will be useful to work with variables that do not grow in the long term, and thus we define $\hat{y}_t \equiv \frac{y_t}{(1+\gamma)^t}$, $\hat{k}_t \equiv \frac{k_t}{(1+\gamma)^t}$, and $\hat{c}_t \equiv \frac{c_t}{(1+\gamma)^t}$; note that $\hat{y}^* \equiv \frac{y_t^*}{(1+\gamma)^t}$, is a constant. The dynamic system we propose to analyze is composed by the capital accumulation constraint and the consumption equation in (4), after replacing the original variables by the constant steady state variables,

$$\hat{k}_{t+1} = \frac{1+A-\delta}{1+\gamma} \cdot \hat{k}_t - \frac{1}{1+\gamma}, \cdot \hat{c}_t \tag{5}$$

$$\hat{c}_{t+1} = \frac{1}{1+\gamma} \cdot \frac{1+A-\delta-\rho \cdot a\,(A-\delta) \cdot \beta(\hat{x}_{t+1})}{1/(\beta(\hat{x}_t) \cdot \hat{c}_t)-\rho \cdot a \cdot \beta(\hat{x}_{t+1})/((1+A-\delta) \cdot \hat{k}_t-\hat{c}_t)} \tag{6}$$

Note that $\beta(\hat{x}_t)=1/\left[1+\rho \cdot (1+a \cdot (\ln\hat{y}_t - \ln\hat{y}_t^*))\right]$, which is equivalent to $\beta(\hat{x}_t)=1/\left[1+\rho \cdot (1+a \cdot (\ln\hat{k}_t - \ln\hat{k}^*))\right]$, with $\hat{k}^*$ the steady state value of $\hat{k}_t$.

The computation of steady state conditions from (5) and (6), allows finding a unique and constant consumption-capital ratio and to determine a unique rate of growth for the main economic aggregates. By solving $(\hat{k}_{t+1}, \hat{c}_{t+1}) = (\hat{k}_t, \hat{c}_t) \equiv (\hat{k}^*, \hat{c}^*)$ we obtain the ratio $\hat{c}^*/\hat{k}^* = A - \delta - \gamma$, with the growth rate given by $\gamma = \frac{1+A+\delta-1/\beta}{1/\beta+(1-\beta) \cdot a}$ [note that we define $\beta \equiv 1/(1+\rho)$]. This is the growth rate of effective output, capital and consumption in the long run and of the potential values of the variables regardless from the assumed time moment. Proposition 1 refers to the growth rate result.

Proposition 1. *In the endogenous growth model with endogenous time preference, the following conditions characterize long term economic growth:*

(i) if $1+A-\delta > 1/\beta$ *and* $a < -1/(\beta \cdot (1-\beta))$, *then the economy's steady state growth rate is negative and it declines with an increase in impatience;*

(ii) if $1+A-\delta > 1/\beta$ *and* $a > -1/(\beta \cdot (1-\beta))$, *then the economy's steady state growth rate is positive and it declines with an increase in impatience;*

(iii) if $1+A-\delta < 1/\beta$ *and* $a < -1/(\beta \cdot (1-\beta))$, *then the economy's steady state growth rate is positive and its value rises with an increase in impatience;*

(iv) finally, if $1+A-\delta < 1/\beta$ *and* $a > -1/(\beta \cdot (1-\beta))$, *then the economy's steady state growth rate is negative and its value rises with an increase in impatience.*

Proof: It is straightforward to separate four cases, regarding the sign of γ. In cases $1+A-\delta > 1/\beta \wedge a > -1/(\beta \cdot (1-\beta))$ and $1+A-\delta < 1/\beta \wedge a < -1/(\beta \cdot (1-\beta))$, the growth rate is positive; when $1+A-\delta > 1/\beta \wedge a < -1/(\beta \cdot (1-\beta))$ or $1+A-\delta < 1/\beta \wedge a > -1/(\beta \cdot (1-\beta))$, the growth rate is negative. Obviously, $1+A-\delta = 1/\beta$ implies $\gamma = 0$, while $a = -1/(\beta \cdot (1-\beta))$

corresponds to an infinite growth rate. Moreover, computing the derivative $\frac{\partial \gamma}{\partial a} = -(1-\beta) \cdot \frac{1+A-\delta-1/\beta}{[1/\beta + (1-\beta)\cdot a]^2}$ we find a negative value for the derivative if $1+A-\delta > 1/\beta$ and a positive value when $1+A-\delta < 1/\beta$. Thus, when the first one of these relations holds, γ falls with increases in a; since a is a measure of impatience (a higher a means that for a same output gap, the representative agent increases her degree of impatience by rising the time preference rate), it is true that the referred condition implies a relation of the same sign between growth and patience. If the second relation holds, γ and a evolve in the same direction, that is, less patience or a higher discount rate are in this case synonymous of an increasing growth rate.

Note that under condition $a = 0$, we are back on the trivial AK model, where $\gamma = \beta \cdot (1+A-\delta) - 1$. Note, as well, that possibilities (*i*) and (*iv*) of proposition 1 are ruled out once we add to the optimality conditions a transversality condition $\lim_{t\to+\infty} k_t \cdot \beta(x_{t-1})^t \cdot p_t = 0$, or, equivalently, $\lim_{t\to+\infty} \frac{k_t}{c_{t-1}} \cdot \beta(x_{t-1})^t = 0$. This makes sense as a terminal condition if capital and consumption grow in time at a same positive (or zero) rate. Its purpose is, in fact, to drive away the possibility of undesirable or senseless long term outcomes. Because the inability of the representative agent in taking into account the transversality condition (and therefore the possibility of negative growth rates) is a basic requisite for the persistence of business cycles in this environment, as one will realize in section 4, we maintain this as an open possibility: the failure to be fully rational and to optimize accordingly implies a disaccumulation of capital and a loss of consumption that occur under a pattern of irregular cycles rather than linearly. Endogenous volatility and limited rationality arise, in this way, as the two faces of a same coin.

One may argue that even a boundedly rational agent is capable of perceiving that the chosen growth path will lead to an everlasting decline on consumption and capital levels and, therefore, that the agent should be able to select an alternative growth path. Although having this in mind, we will keep this possibility in the analysis that follows, since it constitutes a first step to build the analysis in section 5, where the introduction of leisure in the utility function allows encountering cyclical motion for positive growth rates obtained under a strict optimality behaviour.

Let us return to the basic dynamic properties of the system in consideration.

Replacing the expression of the growth rate in the consumption – capital ratio one obtains $\frac{\hat{c}^*}{\hat{k}^*} = (1-\beta) \cdot \frac{a\cdot(A-\delta) + (1+A-\delta)/\beta}{(1/\beta + (1-\beta)\cdot a}$. Thus, the following result holds,

Proposition 2. *In the endogenous growth model with endogenous time preference:*

(i) if $1+A-\delta > 1/\beta$*, then the steady-state consumption – capital ratio rises with an increase in impatience;*

(ii) if $1+A-\delta < 1/\beta$*, then the steady-state consumption – capital ratio falls with an increase in impatience.*

Proof: Just compute $\frac{\partial(\hat{c}^*/\hat{k}^*)}{\partial a} = \frac{(1-\beta)\cdot\gamma}{1/\beta + (1-\beta)\cdot a}$. This is a positive value under $1+A-\delta > 1/\beta$ and a negative value in the opposite circumstance, that is, $1+A-\delta < 1/\beta$, independently of the sign of a. Thus, when the first of the conditions holds, the ratio $\hat{c}^*/\hat{k}^*$ and a move in the same direction, that is, an increasing impatience translated on a higher discount rate (higher a) is synonymous of a higher relative level of consumption in the long term. Symmetrically, the second condition implies a movement of opposite direction, that is, increased patience (lower a) leads to a lower relative level of consumption in the long term■

The result in proposition 2 should be carefully evaluated. For instance, assuming that $1+A-\delta > 1/\beta$, we can have a higher relative level of consumption as the representative consumer becomes increasingly impatient, but nevertheless the absolute level of consumption can be in fact lower as

the discounting becomes stronger. This is clear if we look at the result in proposition 1. In the considered case, the economy grows less as the impatience rises; thus, we may have a rise in $\hat{c}^*/\hat{k}^*$, but with a decline in the growth rate of both aggregates. Therefore, impatience favours consumption relatively to capital accumulation in the long term but penalizes both, as the economy becomes less capable of growing.

3. Local Dynamic Behaviour

The linearization of system (6)-(7) around $(\hat{k}^*, \hat{c}^*)$ yields,

$$\begin{bmatrix} \hat{k}_{t+1} - \hat{k}^* \\ \hat{c}_{t+1} - \hat{c}^* \end{bmatrix} = \begin{bmatrix} \left.\frac{\partial \hat{k}_{t+1}}{\partial \hat{k}_t}\right|_{(\hat{k}^*,\hat{c}^*)} & \left.\frac{\partial \hat{k}_{t+1}}{\partial \hat{c}_t}\right|_{(\hat{k}^*,\hat{c}^*)} \\ \left.\frac{\partial \hat{c}_{t+1}}{\partial \hat{k}_t}\right|_{(\hat{k}^*,\hat{c}^*)} & \left.\frac{\partial \hat{c}_{t+1}}{\partial \hat{c}_t}\right|_{(\hat{k}^*,\hat{c}^*)} \end{bmatrix} \cdot \begin{bmatrix} \hat{k}_t - \hat{k}^* \\ \hat{c}_t - \hat{c}^* \end{bmatrix} \tag{7}$$

The elements of the Jacobian matrix are

$$\left.\frac{\partial \hat{k}_{t+1}}{\partial \hat{k}_t}\right|_{(\hat{k}^*,\hat{c}^*)} = \frac{1 + A - \delta}{1 + \gamma};$$

$$\left.\frac{\partial \hat{k}_{t+1}}{\partial \hat{c}_t}\right|_{(\hat{k}^*,\hat{c}^*)} = -\frac{1}{1 + \gamma};$$

$$\left.\frac{\partial \hat{c}_{t+1}}{\partial \hat{k}_t}\right|_{(\hat{k}^*,\hat{c}^*)} = \frac{A - \delta - \gamma}{(1 + \gamma) \cdot [1 + A - \delta - (1 - \beta) \cdot a \cdot (A - \delta)]} \cdot$$

$$\left[(1-\beta)^2 \cdot a^2 - (1 + \mathrm{A} - \delta) \cdot \gamma - \frac{1-\beta}{\beta} \cdot a \cdot (1 + \gamma)^2 - (1 - \beta) \cdot a \cdot (1 + A - \delta) \cdot (\mathrm{A} - \delta - y)\right];$$

$$\left.\frac{\partial \hat{c}_{t+1}}{\partial \hat{c}_t}\right|_{(\hat{k}^*,\hat{c}^*)} = \frac{\frac{(1 + \gamma)^2}{\beta} + (1 - \beta) \cdot a \cdot [(A - \delta - \gamma] - (1 - \beta) \cdot a \cdot \gamma] \cdot (A - \delta - \gamma)}{(1 + \gamma) \cdot [1 + A - \delta - (1 + \beta) \cdot a \cdot (A - \delta)]}$$

The evaluation of the signs of the eigenvalues of the Jacobian matrix in (7) does not produce intelligible results; even though we have only four parameters in the system, no meaningful relation is found when using the trace and the determinant to explore stability properties. Thus, a numerical example is developed, letting parameter a vary, while attributing reasonable values to the other parameters. We follow Guo and Lansing (2002) in choosing $\beta = 0.962$ and $\delta = 0.067$, and we consider $A = 0.148$ [this is a value for the technology parameter that allows a 4% equilibrium growth rate ($\gamma = 0.04$), when $a = 0$].

The chosen parameterization produces the following results (for $-100 \leq a \leq 100$ and letting λ_1 and λ_2 represent the eigenvalues of the Jacobian matrix in (7)):

- $a \in$ [-100;-28.459)∪(-28.366;-27.494)∪(-13.873;-0.25)∪(27.932;100] ⇒ Eigenvalues are complex roots;
- $a \in$ [-28.459;-28.366]∪[-27.494;-27.336]∪[-0.25;0]∪[26.372;27.932] ⇒ Eigenvalues are real roots ($|\lambda_1|, |\lambda_2| > 1$);
- $a \in$ (-27.336;-13.873)∪(0;26.372) ⇒ Eigenvalues are real roots ($|\lambda_1| < 1, |\lambda_2| > 1$);

Looking at the values of the eigenvalues, one observes that local indeterminacy does not occur; saddle-path stability can occur for some negative and positive values of a and, as a becomes increasingly higher in absolute value we will have instability with real or complex eigenvalues. We conclude that under reasonable economic conditions (a depreciation rate of 6.7%, an equilibrium discount rate of 3.95% and an equilibrium growth rate of 4%), the only possibility regarding a stable outcome (saddle-path stability) arises for values of a immediately above zero and for a small interval of negative values. Various stability outcomes are possible and these are very sensitive to variations in the parameter value.

In appendix B, we present a table with the trace and the determinant of the Jacobian matrix, alongside with the two eigenvalues, for different values of a. According to the table, a strong negative impact of the output gap over the discount rate generates a non-optimal negative growth rate (for $a < -1/(\beta \cdot (1\text{-}\beta)) = -27.356$). The results in the table confirm, for integer values of a, the conditions presented above, i.e., that instability and saddle-path stability outcomes alternate as we change the value of the parameter giving the impact of the output gap over the value of the utility discount rate. In the mentioned table, complex eigenvalues are not displayed but they all possess, with exception of the case $a=-28$, positive real parts, given that the trace is above zero; these positive real parts are lower than one for cases $a \leq 36$, and above one otherwise. In the last column, where instability and saddle-path stability arise in an alternate manner, the instability result is confirmed by presenting the value of the square root of $Det(J)$. This square root corresponds to the modulus of the complex eigenvalues pair. It is well known that if $\sqrt{Det(J)} < 1$, then solutions converge to equilibrium (stability holds) and the equilibrium point is a stable focus; if $\sqrt{Det(J)} > 1$, then solutions will diverge and the equilibrium will be an unstable focus (trajectories will depart from the equilibrium point in a divergent oscillatory way). In the situations presented in the table of appendix B, all the cases involving complex roots are cases in which $\sqrt{Det(J)} > 1$, i.e., cases that display local oscillatory divergence.

4. Endogenous Business Cycles

What should we expect to find when assessing the long run behaviour of the endogenous variables under the parameterization selected in the previous section? Since we have not chosen to stay over any eventual saddle-path, no long term stable trajectories are likely to arise; fixed point stability is absent. However, one observes through numerical experimentation that for an interval of values of a, invariant cycles appear. The values of a for which such kind of long term motion exists are all negative, meaning that endogenous cycles are compatible with the intuitive idea that less discounting / more patience arises for a higher output gap.

For the selected set of parameter values, figure 1 displays the bifurcation diagram[1]. We regard that a region of cycles exists and that this stops as we leave the instability area and enter into the region of saddle-path stability, according to the local dynamics characterization. Selecting one of the values of a for which the cycles are present ($a=-75$), figures 2 to 4 draw an attractor and the long term time series of both variables. The attractor corresponds to an invariant cycle, i.e., to a result of complete a-periodicity but where regularity features imply talking about quasi-periodicity rather than chaos.

***** Figures 1 to 4 *****

In figures 1 to 4, we have chosen $\hat{k}_0 = 1$, $\hat{c}_0 = 0.1$ and $\hat{k}^* = 1$.

The graphical analysis allows to observe that endogenous business cycles arise in the endogenous time preference model when a positive output gap produces a lower discount rate and a negative output gap leads to a higher discount rate, relatively to the reference level ρ. In

1 All the figures in the paper (excluding figure 5) are drawn using IDMC software (interactive Dynamical Model Calculator). This is a free software program available at www.dss.uniud.it/nonlinear, and copyright of M. Lines and A. Medio.

this way, cycles are self-reinforcing: the persistence of business fluctuations is the result of a process where deviations from the effective output relatively to the potential level imply a change in time preference which induces the output to fluctuate; these fluctuations will imply a permanent lack of coincidence between effective and potential output that in turn triggers once again successive shifts in the time preference, and so on. Hence, the steady state equilibrium is never achieved; cycles are perpetuated.

The values of a for which endogenous cycles arise correspond to values implying a negative long term growth, because they are all below the threshold condition $a = -1/(\beta \cdot (1\text{-}\beta))$. Thus, cycles are the straightforward outcome of the absence of complete rationality. Agents are unable to compute a transversality condition and therefore to act optimally, allowing for the possibility of long run negative growth rates. In this case, the cyclical motion displayed in figures 3 and 4 that is presented for stationary variables $\hat{k}_t$ and $\hat{c}_t$, characterizes a process through which endogenous cycles define the pattern of diminishing stock of capital and consumption levels in time.

5. Leisure in the Utility Function

We now develop an extension of the benchmark model of the previous sections. This extension assumes that leisure is an argument of the utility function. The new setting allows to evidence that cyclical motion is compatible with positive growth rates and, consequently, with an integral optimization scenario that makes good use of the transversality condition. The relevant conclusion is that an endogenous rate of time preference where impatience is inversely related to the output gap is capable of generating aggregate fluctuations on output, capital and consumption that are compatible with a full rationality setup.

Consider that the representative agent is endowed with a unit of time, which can be split in working time and leisure time. Thus, generated output will correspond to $y_t = A \cdot k_t \cdot u_t$, with u_t the share of the agent's time allocated to the production of goods. The capital accumulation constraint becomes now $k_{t+1} = (1 + A \cdot u_t - \delta) \cdot k_t - c_t$, k_0 given.

The agent attributes utility to leisure and, thus, the utility function gains an additional argument. The following functional form is adopted, $U(c_t,1-u_t) = \ln c_t + m \cdot \ln(1-u_t)$, with $m>0$. Under this specification, consumption and leisure generate complementary utility: it produces more utility an intermediate level of both consumption and leisure than a great quantity of one and a low availability of the other. Diminishing marginal utility is assumed for both arguments of the utility function.

We set up the Hamiltonian function,

$$H(k_t,c_t,q_t) = U(c_t,1-u_t) + q_{t+1} \cdot \left[(A \cdot u_t - \delta) \cdot k_t - c_t\right] \tag{8}$$

As in section 2, $q_t \equiv \beta(x_{t-1}) \cdot p_t$, with $\beta(x_t) = 1/\left[1 + \rho \cdot (1 + a \cdot x_t)\right]$ and $x_t = \ln k_t - \ln k_t^*$. First order condition (2) continues to hold; condition (3) gives place, in the present context, to

$$(1 + A \cdot u_t - \delta) \cdot \beta(x_t) \cdot p_{t+1} + \beta_k \cdot p_{t+1} \cdot \left[(A \cdot u_t - \delta) \cdot k_t - c_t\right] = p_t \tag{9}$$

A new first-order condition is obtained,

$$H_u = 0 \Rightarrow \frac{m}{1-u_t} = \beta(x_t) \cdot p_{t+1} \cdot A \cdot k_t \tag{10}$$

From conditions (2) and (10), we can establish that

$$u_t = 1 - \frac{m}{A} \cdot \frac{c_t}{k_t} \tag{11}$$

Because $u_t \in (0,1)$, the following constraint applies to this version of the model: $\frac{c_t}{k_t} < \frac{A}{m}$.

Taking in consideration (2) and (11), we transform motion equation (9) into a difference equation for the consumption variable that is similar to (4) (this equation is derived in appendix C). The

transversality condition continues to be $\lim_{t\to+\infty} k_t \cdot \beta(x_{t-1})^t \cdot p_t = 0$ and, as before, we begin by assuming that a lack of capacity of the representative agent may lead her to overlook such condition, meaning that eventually the economy might begin growing at a negative rate; however, as stated earlier, cyclical patterns of growth only found for negative growth rates in the original scenario without leisure, now arise in situations where the transversality condition is accounted for.

Considering, once again, variables that do not grow in the steady state, $\hat{k}_t$ and $\hat{c}_t$, one reaches the dynamic system to be subject to analysis; this is composed by equations (12) and (13),

$$\hat{k}_{t+1} = \frac{1+A+\delta}{1+\gamma} \cdot \hat{k}_t - \frac{1+m}{1+\gamma} \cdot \hat{c}_t \tag{12}$$

$$\hat{c}_{t+1} = \frac{1}{1+\gamma} \cdot \frac{1+A-\delta-\rho \cdot a \cdot (A-\delta) \cdot \beta(\hat{x}_{t+1})}{1/(\beta(\hat{x}_t) \cdot \hat{c}_t) + (m-(1+m) \cdot \rho \cdot a \cdot \beta(\hat{x}_{t+1})) / ((1+A-\delta) \cdot \hat{k}_t - (1+m) \cdot \hat{c}_t)} \tag{13}$$

Comparing system (5)-(6) with (12)-(13), we confirm that the first is a particular case of the second, for $m = 0$. Therefore, it will be interesting to study the dynamics of the model as one increases the relevance of leisure in the utility function (i.e., as one increases the value of m). As before, given the complexity of the expressions of the derived difference equations, this is only achievable under a concrete numerical example.

For now, let us characterize generic results concerning the steady state.

Proposition 3. *In the endogenous time preference model with leisure in the utility function, the main economic aggregates will grow, in the steady state, at rate*

$$\gamma = \frac{1+A-\delta-1/\beta-\frac{m}{1+m}\cdot(A-\delta)}{1/\beta+(1-\beta)\cdot a-\frac{m}{1+m}}$$

Proof: By applying condition $\hat{k}_{t+1} = \hat{k}_t \equiv \hat{k}^*$ to (12), we find a unique consumption-capital ratio: $\hat{c}^* / \hat{k}^* = (A-\delta-\gamma)/(1+m)$. The information provided by this ratio may be used when evaluating (13) under the steady state condition $\hat{c}_{t+1} = \hat{c}_t \equiv \hat{c}^*$; such evaluation, allows to find the growth rate in the proposition■

Observe that under $m = 0$, we are back on the growth rate of the model where leisure utility is absent. The impact of increasing the relative relevance of leisure in terms of utility will depend on the value of the other parameters. Proposition 4 states an important result.

Proposition 4. *The impact of the relative weight of leisure utility over long term growth will depend on the value of the parameter of the discount function:*

If $a < -\frac{1+A-\delta}{(A-\delta)\cdot\beta}$, *then the potential growth rate increases along with* m*;*

If $a > -\frac{1+A-\delta}{(A-\delta)\cdot\beta}$, *then the potential growth rate falls with an increase in* m*.*

Therefore, as long as the potential growth rate increases with m, we assure that patience rises with an increasingly positive (or a decreasingly negative) output gap.

Proof: To prove the proposition we just have to compute the derivative

$$\frac{\partial\gamma}{\partial m} = \frac{1}{(1+m)^2} \cdot \frac{\left(1+A-\delta-1/\beta-\frac{m}{(1+m)}\cdot(A-\delta)\right) - (A-\delta)\cdot\left(1/\beta+(1-\beta)\cdot a-\frac{m}{(1+m)}\right)}{\left(1/\beta+(1-\beta)\cdot a-\frac{m}{1+m}\right)^2}$$

A negative derivative implies that as m increases, γ falls, and the opposite for a derivative with a positive sign. The above expression has a negative sign if $\left(1 + A - \delta - 1/\beta - \frac{m}{(1+m)} \cdot (A - \delta)\right) < (A - \delta) \cdot \left(1/\beta + (1 - \beta) \cdot a - \frac{m}{(1+m)}\right)$, which is equivalent to the second condition in the proposition; if we want to obtain a positive value for the derivative, the symmetric condition applies (this is the first condition in the proposition).

Applying the benchmark parameter values of previous sections, we have $\gamma = \frac{0.042 - 0.081 \cdot \frac{m}{1+m}}{1.04 + 0.038 \cdot a - \frac{m}{1+m}}$; this value increases with an increase in m if $a < -13.873$, and it decreases with an increase in m if $a > -13.873$.

Concerning steady state results, one can also present the following,

Proposition 5. *For an interior steady state solution, the following constraints on the value of m must be satisfied:*

i) $m < \frac{\frac{1}{\beta} + (1 - \beta) \cdot a}{\frac{1-\beta}{\beta} + (1 - \beta) \cdot a - \frac{1-\beta}{A} \cdot \left[a \cdot (A - \delta) + \frac{1 + A - \delta}{\beta}\right]}$;

ii) $m: \frac{a \cdot (A - \delta) + (1 + A - \delta)/\beta}{(1+m) \cdot \left(\frac{1}{\beta} + (1 + \beta) \cdot a\right) - m} > 0.$

The second condition can be presented in more detail if one takes two cases,

a) $a < -\frac{1 + A - \delta}{(A - \delta) \cdot \beta} \Longrightarrow m < -\frac{1/\beta + (1 - \beta) \cdot a}{(1 - \beta)/\beta + (1 - \beta) \cdot a}$;

b) $a > -\frac{1 + A - \delta}{(A - \delta) \cdot \beta} \Longrightarrow m > -\frac{1/\beta + (1 - \beta) \cdot a}{(1 - \beta)/\beta + (1 - \beta) \cdot a}$.

Proof: Replacing the steady state growth rate in the steady state consumption-capital ratio, which was presented in the proof of proposition 3, one obtains an expression for this long term ratio that is a function of our several parameters: $\frac{\hat{c}^*}{\hat{k}^*} = (1 - \beta) \cdot \frac{a \cdot (A - \delta) + (1 + A - \delta)/\beta}{(1+m) \cdot (1/\beta (1 - \beta) \cdot a) - m.}$.

This ratio has to be a positive value, and this requirement corresponds precisely to the second condition in the proposition. Noticing that u_t is given by (11), the steady state value of the share of time allocated to working hours is

$u^* = 1 - \frac{m}{A}(1 - \beta) \cdot \frac{a \cdot (A - \delta) + (1 + A - \delta)/\beta}{(1+m) \cdot (1/\beta + (1 - \beta) \cdot a) - m}$. To guarantee an interior solution, the value of u^* must rest between 0 and 1; $u^* < 1$ is guaranteed under the same condition that allows for assuring that the consumption-capital ratio is positive; the condition $u^* > 0$ implies a value of m bounded by the first constraint in the proposition■

Once more, steady state results may be illustrated for the specific numerical case under consideration. Take the chosen parameter values, including $a = -75$. In this case, we have:

$\gamma = \frac{0.042 - 0.04 \cdot m}{0.04\, m - 1.811}$; $\frac{\hat{c}^*}{\hat{k}^*} = \frac{0.184}{1.811 + 2.811 \cdot m}$; $u^* = \frac{1.811 + 1.569 \cdot m}{1.811 + 2.811 \cdot m}$. Our example obeys the

Our example obeys the constraints that are necessary to impose to parameters. First, note that relatively to the growth rate this is negative for $m = 0$ (γ = -0.023), as one remarked in the analysis of the benchmark case; note too that as m increases the growth rate also increases, a result that is in accordance with proposition 4. The growth rate becomes positive for $m > 1.051$. The chosen parameter values allow for an admissible interior solution, as characterized in proposition 5, since the consumption-capital ratio is positive and the share u^* is clearly bounded between 0 and 1.

Regard that under this example, a higher relative utility of leisure reduces the level of the steady state consumption-capital ratio. A higher relative utility of leisure implies, as well, that the share of work time declines. This second result is, indeed, intuitive, reflecting in this way that our parameters are able to characterize reasonable economic conditions.

To understand the dynamics of the model with leisure, we proceed as with the benchmark framework, first by looking at local dynamics and, on a second stage, by analyzing global dynamic properties.

In terms of local dynamics, we compute the elements of the Jacobian matrix,

$$\left.\frac{\partial \hat{k}_{t+1}}{\partial \hat{k}_t}\right|_{(\hat{k}^*,\hat{c}^*)} = \frac{1 + A - \delta}{1 + \gamma};$$

$$\left.\frac{\partial \hat{k}_{t+1}}{\partial \hat{c}_t}\right|_{(\hat{k}^*,\hat{c}^*)} = -\frac{1+ m}{1 + \gamma};$$

$$\left.\frac{\partial \hat{c}_{t+1}}{\partial \hat{k}_t}\right|_{(\hat{k}^*,\hat{c}^*)} = \frac{A - \delta - \gamma}{(1 + \gamma) \cdot [1 + A - \delta - (1 - \beta) \cdot a \cdot (A - \delta)] \cdot (1 + m)};$$

$$\left[(1-\beta)^2 \cdot a^2 \cdot (1 + \mathrm{A} - \delta) \cdot \gamma - \frac{1-\beta}{\beta} \cdot a \cdot (1 + \gamma)^2 + \left(\frac{m}{1 + m} - (1 - \beta) \cdot a\right) \cdot (1 + A - \delta) \cdot (\mathrm{A} - \delta - \gamma)\right];$$

$$\left.\frac{\partial \hat{c}_{t+1}}{\partial \hat{c}_t}\right|_{(\hat{k}^*,\hat{c}^*)} = \frac{1}{(1 + \gamma) \cdot [1 + A - \delta - (1 - \beta) \cdot a \cdot (A - \delta)]} \cdot$$

$$\left[\frac{(1 + \gamma)^2}{\beta} + (1-\beta) \cdot a \cdot ((A - \delta - \gamma) - (1-\beta) \cdot a \cdot \gamma) \cdot (A - \delta - \gamma) - \frac{m}{1 + m} \cdot (A - \delta - \gamma)^2\right]$$

The previous expressions are not easy to work with; from a generic point of view, one can just confirm that imposing $m = 0$, we are back on the non leisure case. To address the present model's local dynamics, we proceed by assuming all the previous parameter values plus a variable m; in this way, we can explore how a change in the leisure parameter changes dynamic results. The numerical example allows to reach the following outcome: independently of the value of m, the eigenvalues of the Jacobian matrix are always complex values and instability prevails; for $m = 0$ the trace and the determinant of the Jacobian matrix are $Tr(J) = 1.872$ and $Det(J) = 1.091$, and these will decrease as one increases the value of m, such that $\lim_{t\to+\infty} Tr(J) = 1.851$ and $\lim_{t\to+\infty} Det(J) = 1.01$.

Since the determinant stays always above unity, we confirm the presence of instability, independently of the value of the parameter. This result can be confirmed by looking at the upper left panel of figure 5. The several panels in this figure represent the relation between the trace (on the horizontal axis) and the determinant (on the vertical axis) of the Jacobian matrix of the system under analysis. This relation is drawn, for each one of the considered values of parameter a, for any possible positive value of m. One verifies that in most of the circumstances instability will prevail independently of the value of m; however, for some values of a (e.g. $a = -10$ or $a = -5$) saddle-path stability is found.

***** Figure 5 *****

The table in appendix D presents the values of the trace and determinant for different values of a, in the two extremes, $m = 0$ and $m \to +\infty$. The correctness of these values can be checked by looking at the extremes of the lines in the various graphics of figure 5.

In terms of global dynamics, we already know that, for the considered parameterization (including $a = -75$), cycles exist when $m = 0$. The graphical representation of a bifurcation diagram, for m as the bifurcation parameter, allows for extending this result. Figure 6 displays the bifurcation diagram and figure 7 presents a long run attractor for a value of m different from 0. As in the benchmark case, we consider $\hat{k}_0 = 1$, $\hat{c}_0 = 1$ and $\hat{k}^* = 1$ to draw the figures.

***** Figures 6 and 7 *****

Figure 6 allows for observing that only for values of m from 0 to around 2 to 3, it is possible to find cycles; afterwards, instability will prevail. We also see that cycles will slightly reduce their amplitude as one increases m. The figure refers to a bifurcation diagram, as figure 1, but we are unable to identify any bifurcation point: there is an abrupt jump from cycles to instability; however, the kind of cycles that we obtain are characteristic of the fluctuations produced by a Neimark-Sacker bifurcation (or Hopf bifurcation in discrete time), since such fluctuations correspond to quasi-periodic orbits (see Medio and Lines, 2001, for details on bifurcations and classification of types of cyclical motion). A Neimark-Sacker bifurcation occurs when the determinant of the Jacobian matrix crosses unity, and cycles will arise for values of the determinant above unity; this is compatible with the results found on the numerical investigation.

The attractor in figure 7 reveals one important point that was absent in the original model; in the model without leisure in the utility function, cycles only arose for negative growth rates; now, as we have referred, we have a positive growth rate for $m > 1.0506$, and cycles continue to exist for values of m above this level. In the example in figure 6, we have $m = 2$, which implies a steady state growth rate around 0.5%, and in this case we observe that quasi-periodic cycles continue to exist. Hence, the particular example contains two relevant ideas for the interval of values of m for which cycles are observed: first, as the relevance of leisure in the utility function increases, cycles decrease their amplitude; second, as the relevance of leisure in the utility function increases, the economy's long term / potential growth rate increases as well.

The second of the above ideas must be re-emphasized as the main result of the analysis. While we have begun by stating that persistent cyclical motion is possible in the standard Ramsey growth model with an endogenous time preference, but only if taking an assumption of lack of rationality that leads agents to be unable to rule out negative growth rates from their decision plans, the introduction of leisure into the precise same setting allows for aggregate fluctuations that are fully compatible with an optimization scenario.

Other cases, concerning other values of a (as the ones in table 1), do not produce significantly different results. In fact, numerical exploration allows to identify only two global dynamics results: instability (i.e., impossibility of finding any long term path for capital and consumption) and quasi-periodic cycles, that are mostly found for significantly negative values of parameter a.

6. Conclusions

A discrete time version of the simple AK endogenous growth model was modified by considering endogenous time preference. Intertemporal preference depends on the output gap and the main assumption corresponds to the logical argument that the higher is the effective output relatively to its potential level, the more patient a representative agent tends to be. This modification of the growth model immediately produces complicated dynamics as the relations between variables in the difference equations that characterize the model become nonlinear.

Two versions of the model were developed. The first was the simple Ramsey model of utility maximization, with a constant marginal returns production function. In this version of the model, one can only address the relation between capital and consumption. Instability was dominant (although for some values of the parameter that measures the relation between the output gap and the discount rate, saddle-path stability was observed as well), but below a given negative value of such parameter endogenous business cycles arise. Thus, one concludes that endogenous fluctuations may exist in the simple AK growth model as long as an endogenous intertemporal discount rate is assumed and the relation between the output gap and the discount rate is largely negative. Nevertheless, this endogenous cycles outcome is attained only if a deficiency on optimal behaviour is assumed: negative growth rates arising from neglecting the transversality condition will determine the eventuality of a fluctuations result.

The second version of the model has introduced leisure in the utility function. The scenario in which cycles have arisen in the first model becomes a particular case of a wider framework, where no leisure or a low share of leisure implies cycles, but a relatively high share of leisure leads to global instability. The introduction of leisure allows realizing that a negative growth rate is not essential to observe the presence of cycles, meaning that an endogenous discount rate can effectively generate endogenous fluctuations under a strict optimality scenario.

Appendix A – Derivation of Equation (4)

Rewrite equation (3) as

$$(1 + A - \delta) - \rho \cdot a \cdot \beta(x_t) \cdot \left[(A - \delta) - \frac{c_t}{k_t} \right] = \frac{p_t}{\beta(x_t) \cdot p_{t+1}} \tag{A1}$$

Relation (2) allows replacing the co-state variable by consumption in (A1), as follows,

$$(1 + A - \delta) - \rho \cdot a \cdot \beta(x_t) \cdot \left[(A - \delta) - \frac{c_t}{k_t} \right] = \frac{c_t}{\beta(x_{t-1}) \cdot c_{t-1}} \tag{A2}$$

Equation (A2) is equivalent to

$$c_t = \frac{(1 + A - \delta) - \rho \cdot a \cdot (A - \delta)\,\beta(x_t)}{1/(\beta(x_{t-1}) \cdot c_{t-1}) - \rho \cdot a \cdot \beta(x_t)/k_t} \tag{A3}$$

Considering (A3) one time period ahead, we have dynamic equation (4).

Appendix B – Local Dynamic Properties for Different Values of a ($a \in \{-50,-49,\ldots,0,\ldots,49,50\}$)

a	γ	$Tr(J)$	$Det(J)$	λ_1	λ_2	Stability result
-50	-0.0482	1.9291	1.5849	Complex	Complex	Instability ($\sqrt{Det(J)} = 1.2589$)
-49	-0.0505	1.9319	1.5803	Complex	Complex	Instability ($\sqrt{Det(J)} = 1.2571$)
-48	-0.0529	1.9348	1.5766	Complex	Complex	Instability ($\sqrt{Det(J)} = 1.2556$)
-47	-0.0556	1.9377	1.5739	Complex	Complex	Instability ($\sqrt{Det(J)} = 1.2545$)
-46	-0.0586	1.9408	1.5722	Complex	Complex	Instability ($\sqrt{Det(J)} = 1.2539$)
-45	-0.0619	1.944	1.5717	Complex	Complex	Instability ($\sqrt{Det(J)} = 1.2537$)
-44	-0.0656	1.9474	1.5727	Complex	Complex	Instability ($\sqrt{Det(J)} = 1.2541$)
-43	-0.0698	1.951	1.5753	Complex	Complex	Instability ($\sqrt{Det(J)} = 1.2551$)
-42	-0.0746	1.9549	1.5801	Complex	Complex	Instability ($\sqrt{Det(J)} = 1.257$)
-41	-0.08	1.959	1.5875	Complex	Complex	Instability ($\sqrt{Det(J)} = 1.26$)
-40	-0.0864	1.9636	1.5981	Complex	Complex	Instability ($\sqrt{Det(J)} = 1.2642$)
-39	-0.0938	1.9687	1.6129	Complex	Complex	Instability ($\sqrt{Det(J)} = 1.27$)
-38	-0.1026	1.9745	1.6331	Complex	Complex	Instability ($\sqrt{Det(J)} = 1.2779$)
-37	-0.1132	1.9815	1.6609	Complex	Complex	Instability ($\sqrt{Det(J)} = 1.2887$)
-36	-0.1263	1.9899	1.6992	Complex	Complex	Instability ($\sqrt{Det(J)} = 1.3035$)
-35	-0.1429	2.0008	1.7534	Complex	Complex	Instability ($\sqrt{Det(J)} = 1.3242$)
-34	-0.1644	2.0157	1.8327	Complex	Complex	Instability ($\sqrt{Det(J)} = 1.3538$)
-33	-0.1935	2.0378	1.9556	Complex	Complex	Instability ($\sqrt{Det(J)} = 1.3984$)

Appendix B – Local Dynamic Properties for Different Values of a ($a \in \{-50,-49,\ldots,0,\ldots,49,50\}$) *(Cont.)*

a	γ	$Tr(J)$	$Det(J)$	λ_1	λ_2	Stability result
-32	-0.2351	2.0745	2.1643	Complex	Complex	Instability ($\sqrt{Det(J)} = 1.4712$)
-31	-0.2996	2.1473	2.5802	Complex	Complex	Instability ($\sqrt{Det(J)} = 1.6063$)
-30	-0.4129	2.3425	3.7258	Complex	Complex	Instability ($\sqrt{Det(J)} = 1.9302$)
-29	-0.664	3.488	12.286	Complex	Complex	Instability ($\sqrt{Det(J)} = 3.5052$)
-28	-1.6939	-2.2398	15.694	Complex	Complex	Instability ($\sqrt{Det(J)} = 3.9615$)
-27	3.0738	4.0073	-1.5479	-0.3549	4.3621	Saddle-path
-26	0.8058	2.2398	0.2654	0.1255	2.1142	Saddle-path
-25	0.4637	2.0649	0.5544	0.3172	1.7477	Saddle-path
-24	0.3255	2.0162	0.6948	0.4411	1.5752	Saddle-path
-23	0.2507	1.9982	0.781	0.533	1.4651	Saddle-path
-22	0.2039	1.9909	0.8394	0.6062	1.3847	Saddle-path
-21	0.1718	1.9883	0.8813	0.6671	1.3212	Saddle-path
-20	0.1485	1.9879	0.9125	0.7193	1.2686	Saddle-path
-19	0.1307	1.9888	0.9364	0.7654	1.2233	Saddle-path
-18	0.1167	1.9904	0.955	0.8071	1.1832	Saddle-path
-17	0.1055	1.9924	0.9698	0.8459	1.1465	Saddle-path
-16	0.0962	1.9946	0.9816	0.8833	1.1114	Saddle-path
-15	0.0884	1.9971	0.9912	0.9221	1.075	Saddle-path
-14	0.0818	1.9997	0.9991	0.9762	1.0235	Saddle-path
-13	0.0761	2.0023	1.0056	Complex	Complex	Instability ($\sqrt{Det(J)} = 1.0028$)
-12	0.0711	2.005	1.011	Complex	Complex	Instability ($\sqrt{Det(J)} = 1.0055$)
-11	0.0668	2.0078	1.0155	Complex	Complex	Instability ($\sqrt{Det(J)} = 1.0077$)
-10	0.0629	2.0106	1.0193	Complex	Complex	Instability ($\sqrt{Det(J)} = 1.0096$)
-9	0.0595	2.0134	1.0225	Complex	Complex	Instability ($\sqrt{Det(J)} = 1.0112$)
-8	0.0564	2.0162	1.0253	Complex	Complex	Instability ($\sqrt{Det(J)} = 1.0126$)
-7	0.0537	2.0191	1.0277	Complex	Complex	Instability ($\sqrt{Det(J)} = 1.0137$)
-6	0.0511	2.022	1.0298	Complex	Complex	Instability ($\sqrt{Det(J)} = 1.0148$)
-5	0.0489	2.0249	1.0317	Complex	Complex	Instability ($\sqrt{Det(J)} = 1.0157$)
-4	0.0468	2.0278	1.0334	Complex	Complex	Instability ($\sqrt{Det(J)} = 1.0166$)
-3	0.0448	2.0307	1.035	Complex	Complex	Instability ($\sqrt{Det(J)} = 1.0174$)
-2	0.0431	2.0336	1.0366	Complex	Complex	Instability ($\sqrt{Det(J)} = 1.0181$)
-1	0.0414	2.0365	1.038	Complex	Complex	Instability ($\sqrt{Det(J)} = 1.0188$)
0	0.0399	2.0395	1.0395	1	1.0395	Instability ($\sqrt{Det(J)} = 1.0196$)
1	0.0385	2.0425	1.041	0.9771	1.0654	Saddle-path

Appendix B – Local Dynamic Properties for Different Values of a
($a \in \{-50,-49,\ldots,0,\ldots,49,50\}$) *(Cont.)*

a	γ	$Tr(J)$	$Det(J)$	λ_1	λ_2	Stability result
2	0.0372	2.0454	1.0425	0.9636	1.0819	Saddle-path
3	0.036	2.0484	1.044	0.9534	1.095	Saddle-path
4	0.0348	2.0514	1.0456	0.9452	1.1062	Saddle-path
5	0.0338	2.0545	1.0473	0.9385	1.116	Saddle-path
6	0.0327	2.0575	1.0491	0.9328	1.1246	Saddle-path
7	0.0318	2.0605	1.051	0.9282	1.1324	Saddle-path
8	0.0309	2.0636	1.0531	0.9244	1.1392	Saddle-path
9	0.03	2.0667	1.0552	0.9213	1.1453	Saddle-path
10	0.0292	2.0698	1.0576	0.919	1.1507	Saddle-path
11	0.0285	2.0729	1.06	0.9174	1.1555	Saddle-path
12	0.0277	2.076	1.0627	0.9165	1.1595	Saddle-path
13	0.0271	2.0791	1.0655	0.9162	1.1629	Saddle-path
14	0.0264	2.0823	1.0685	0.9166	1.1657	Saddle-path
15	0.0258	2.0854	1.0716	0.9177	1.1678	Saddle-path
16	0.0252	2.0886	1.075	0.9194	1.1692	Saddle-path
17	0.0246	2.0918	1.0786	0.9219	1.1699	Saddle-path
18	0.0241	2.095	1.0823	0.9252	1.1699	Saddle-path
19	0.0236	2.0983	1.0863	0.9292	1.1691	Saddle-path
20	0.0231	2.1015	1.0905	0.9342	1.1674	Saddle-path
21	0.0226	2.1048	1.0949	0.9401	1.1647	Saddle-path
22	0.0221	2.1081	1.0996	0.9471	1.161	Saddle-path
23	0.0217	2.1114	1.1045	0.9555	1.1559	Saddle-path
24	0.0213	2.1147	1.1096	0.9655	1.1492	Saddle-path
25	0.0209	2.1181	1.1149	0.9777	1.1404	Saddle-path
26	0.0205	2.1214	1.1205	0.993	1.1284	Saddle-path
27	0.0201	2.1248	1.1264	1.0143	1.1105	Instability ($\sqrt{Det(J)} = 1.0613$)
28	0.0197	2.1282	1.1325	Complex	Complex	Instability ($\sqrt{Det(J)} = 1.0642$)
29	0.0194	2.1316	1.1389	Complex	Complex	Instability ($\sqrt{Det(J)} = 1.0672$)
30	0.019	2.1351	1.1455	Complex	Complex	Instability ($\sqrt{Det(J)} = 1.0703$)
31	0.0187	2.1385	1.1524	Complex	Complex	Instability ($\sqrt{Det(J)} = 1.0735$)
32	0.0184	2.142	1.1595	Complex	Complex	Instability ($\sqrt{Det(J)} = 1.0768$)
33	0.0181	2.1455	1.167	Complex	Complex	Instability ($\sqrt{Det(J)} = 1.0803$)
34	0.0178	2.149	1.1747	Complex	Complex	Instability ($\sqrt{Det(J)} = 1.0838$)
35	0.0175	2.1526	1.1827	Complex	Complex	Instability ($\sqrt{Det(J)} = 1.0875$)

Appendix B – Local Dynamic Properties for Different Values of a ($a \in \{-50,-49,\ldots,0,\ldots,49,50\}$) *(Cont.)*

a	γ	$Tr(J)$	$Det(J)$	λ_1	λ_2	Stability result
36	0.0172	2.1561	1.191	Complex	Complex	Instability ($\sqrt{Det(J)} = 1.0913$)
37	0.017	2.1597	1.1996	Complex	Complex	Instability ($\sqrt{Det(J)} = 1.0952$)
38	0.0167	2.1633	1.2084	Complex	Complex	Instability ($\sqrt{Det(J)} = 1.0993$)
39	0.0165	2.167	1.2176	Complex	Complex	Instability ($\sqrt{Det(J)} = 1.1034$)
40	0.0162	2.1706	1.2271	Complex	Complex	Instability ($\sqrt{Det(J)} = 1.1077$)
41	0.016	2.1743	1.2368	Complex	Complex	Instability ($\sqrt{Det(J)} = 1.1121$)
42	0.0157	2.178	1.2469	Complex	Complex	Instability ($\sqrt{Det(J)} = 1.1167$)
43	0.0155	2.1817	1.2573	Complex	Complex	Instability ($\sqrt{Det(J)} = 1.1213$)
44	0.0153	2.1855	1.268	Complex	Complex	Instability ($\sqrt{Det(J)} = 1.1261$)
45	0.0151	2.1892	1.2791	Complex	Complex	Instability ($\sqrt{Det(J)} = 1.131$)
46	0.0149	2.193	1.2904	Complex	Complex	Instability ($\sqrt{Det(J)} = 1.136$)
47	0.0147	2.1968	1.3021	Complex	Complex	Instability ($\sqrt{Det(J)} = 1.1411$)
48	0.0145	2.2007	1.3141	Complex	Complex	Instability ($\sqrt{Det(J)} = 1.1464$)
49	0.0143	2.2045	1.3265	Complex	Complex	Instability ($\sqrt{Det(J)} = 1.1517$)
50	0.0141	2.2084	1.3392	Complex	Complex	Instability ($\sqrt{Det(J)} = 1.1572$)

Appendix C – Derivation of the Consumption Difference Equation in the Model with Leisure

Given the expression of the derivative β_k that one has computed in section 2, we rewrite (9),

$$(1 + A \cdot u_t - \delta) - \rho \cdot a \cdot \beta(x_t) \cdot \left[(A \cdot u_t - \delta) - \frac{c_t}{k_t} \right] = \frac{p_t}{\beta(x_t) \cdot p_{t+1}} \tag{C1}$$

Condition (2) holds under optimality, and thus (C1) is equivalent to

$$(1 + A \cdot u_t - \delta) - \rho \cdot a \cdot \beta(x_t) \cdot \left[(A \cdot u_t - \delta) - \frac{c_t}{k_t} \right] = \frac{c_t}{\beta(x_{t-1}) \cdot p_{t-1}} \tag{C2}$$

Now, we replace u_t by the equivalent expression in (11) to obtain

$$\left(1 + A - \delta - m \cdot \frac{c_t}{k_t}\right) - \rho \cdot a \cdot \beta(x_t) \cdot \left[A - \delta - (1 + m) \cdot \frac{c_t}{k_t} \right] = \frac{c_t}{\beta(x_{t-1}) \cdot c_{t-1}} \tag{C3}$$

Solving (C3) in order to c_t, one will have

$$c_t = \frac{1 + A - \delta - \rho \cdot a \cdot (A - \delta) \cdot \beta(x_t)}{1 / (\beta(x_{t-1}) \cdot c_{t-1}) + (m - (1 + m) \cdot \rho \cdot a \cdot \beta(x_t)) / k_t} \tag{C4}$$

Difference equation (C4) is easily transformed in (13) through the consideration of the definitions of $\hat{k}_t$ and $\hat{c}_t$.

Appendix D – Trace and determinant of the Jacobian matrix of the linearized system for the model with leisure in the utility function

a	$m = 0$		$m \to +\infty$	
	$Tr(J)$	$Det(J)$	$Tr(J)$	$Det(J)$
-75	1.872	1.091	1.851	1.01
-50	1.929	1.117	1.906	1.006
-30	2.343	1.457	1.956	1.003
-25	2.065	0.644	1.969	1.003
-20	1.988	0.937	1.983	1.002
-15	1.997	0.993	1.997	1
-10	2.011	1.017	2.011	0.998
-5	2.025	1.031	2.026	0.992
0	2.04	1.04	2.04	1.04
5	2.055	1.046	2.056	1.008
10	2.07	1.051	2.073	1.004

References

Balasko, Y.; Ghiglino, C. (1995) On the existence of endogenous cycles, *Journal of Economic Theory*, 67, 566-577.

Becker, G. S.; Mulligan, C. B. (1997) The endogenous determination of time preference, *Quarterly Journal of Economics*, 107, 729-758.

Boldrin, M.; Nishimura, K.; Shigoka T.; Yano M. (2001) Chaotic equilibrium dynamics in endogenous growth models, *Journal of Economic Theory*, 96, 97-132.

Caballé, J.; Jarque X.; Michetti, E. (2006) Chaotic dynamics in credit constrained emerging economies, *Journal of Economic Dynamics and Control*, 30, 1261-1275.

Cellarier, L. (2006) Constant gain learning and business cycles, *Journal of Macroeconomics*, 28, 51-85.

Christiano, L.; Harrison, S. (1999) Chaos, sunspots and automatic stabilizers, *Journal of Monetary Economics*, 44, 3-31.

Dolmas, J. (1996) Recursive utility and balanced growth, *Journal of Economic Dynamics and Control*, 20, 657-680.

Drugeon, J. P. (1996) Impatience and long-run growth, *Journal of Economic Dynamics and Control*, 20, 281-313.

Drugeon, J. P. (1998) A model with endogenously determined cycles, discounting and growth, *Economic Theory*, 12, 349-369.

Epstein, L. G. (1987) A simple dynamic general equilibrium model, *Journal of Economic Theory*, 41, 68-95.

Guo, J. T.; Lansing, K. J. (2002) Fiscal policy, increasing returns and endogenous fluctuations, *Macroeconomic Dynamics*, 6, 633-664.

Hausman, J. A. (1979) Individual discount rates and the purchase and utilization of energy-using durables, *Bell Journal of Economics*, 10, 33-54.

Lawrance, E. C. (1991) Poverty and the rate of time preference: evidence from the panel data, *Journal of Political Economy*, 99, 54-77.

Medio, A.; Lines M. (2001) *Nonlinear dynamics: a primer*, Cambridge UK, Cambridge University Press.

Meng, Q. (2006) Impatience and equilibrium indeterminacy, *Journal of Economic Dynamics and Control*, 30, 2671-2692.

Nishimura, K.; Sorger, G.; Yano, M. (1994) Ergodic chaos in optimal growth models with low discount rates, *Economic Theory*, 4, 705-717.

Nishimura, K.; Yano, M. (1995) "Nonlinear dynamics and chaos in optimal growth: an example, *Econometrica*, 63, 981-1001.

Palivos, T.; Wang, P.; Zhang, J. (1997) On the existence of balanced growth equilibrium, *International Economic Review*, 38, 205-224.

Samwick, A. A. (1998) Discount rate homogeneity and social security reform, *Journal of Development Economics*, 57, 117-146.

Uzawa, H. (1968) Time preference, the consumption function, and the optimum asset holdings, in Wolfe, J. N. (ed.) *Value, Capital and Growth: Papers in Honour of Sir John Hicks*, Edinburgh, University of Edinburgh Press.

Figures

Figure 1 – Bifurcation diagram (k_t, a)

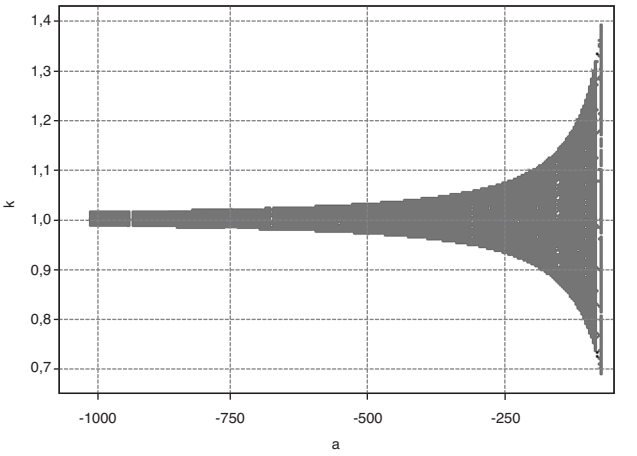


Figure 2 – Attractor (a = -75)

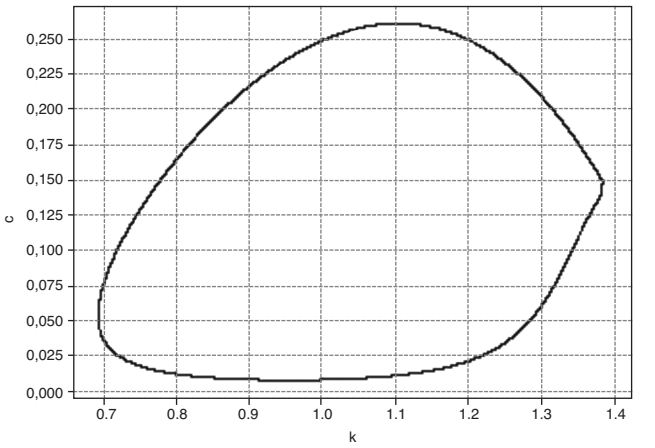


Figure 3 – Long term time series (capital; $a = -75$)

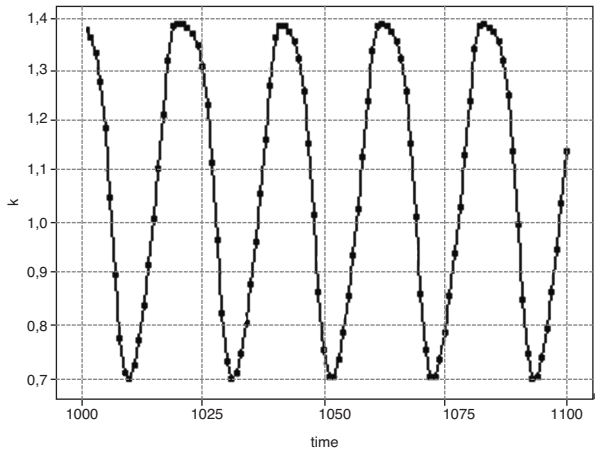


Figure 4 – Long term time series (consumption; $a = -75$)

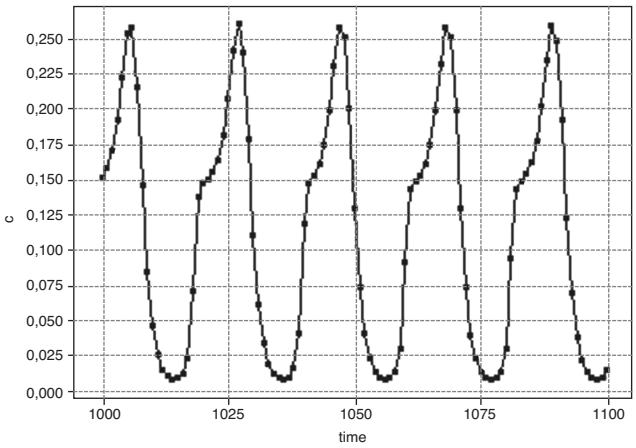
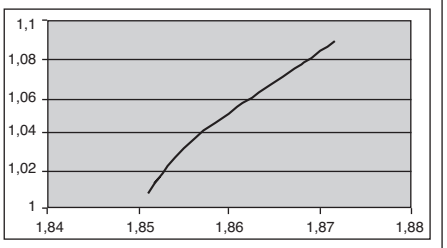
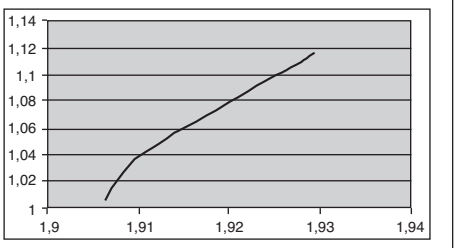


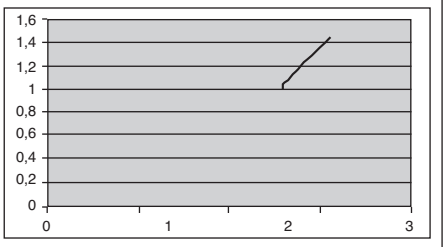
Figure 5 – The trace-determinant relation, for $m \geq 0$ and different values of a
(the horizontal axis respects to the trace and the vertical axis to the determinant)



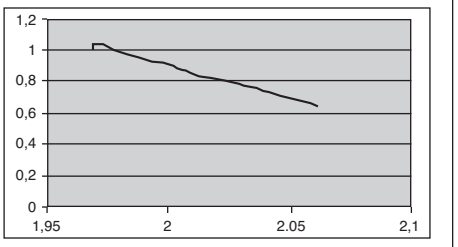
a = -75



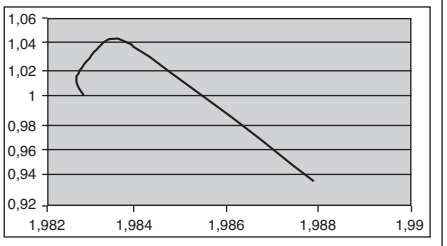
a = -50



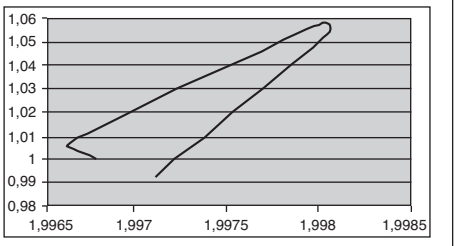
a = -30



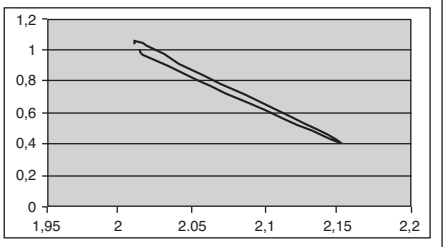
a = -25



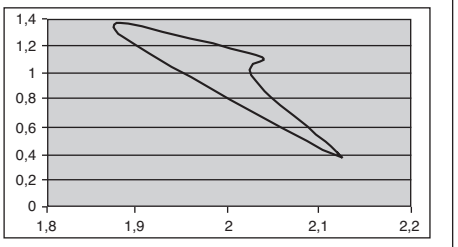
a = -20



a = -15



a = -10



a = -5

Figure 5 – The trace-determinant relation, for $m \geq 0$ and different values of a
(the horizontal axis respects to the trace and the vertical axis to the determinant) *(Cont.)*

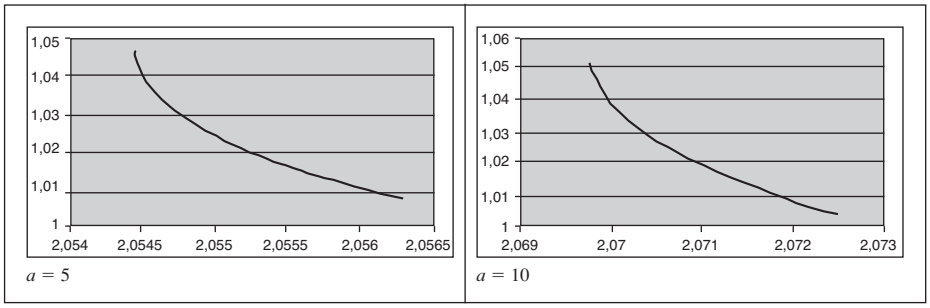


Figure 6 – Bifurcation diagram (k_t, m) (model with leisure in the utility function).

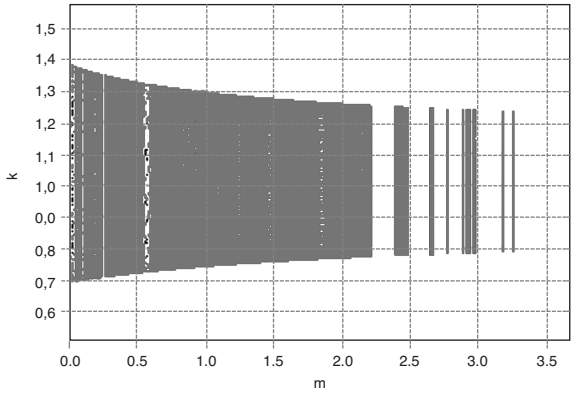
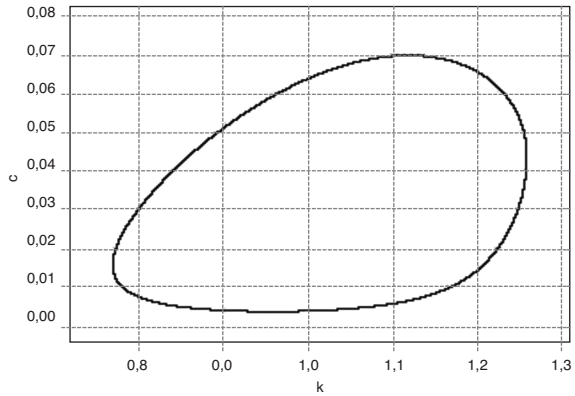


Figure 7 – Attractor (model with leisure in the utility function; $m = 2$).



A Reforma de 2007 do Sistema Público de Pensões em Portugal – Uma Análise Crítica das Escolhas Normativas Implícitas

Maria Clara Murteira FEUC

resumo

No presente artigo defende-se a necessidade de atribuir um papel central à função social dos sistemas de pensões no processo de formulação de políticas neste domínio, avaliando os efeitos das medidas adoptadas sobre os níveis de rendimento garantidos. Seguindo esta linha argumentativa, efectua-se uma avaliação crítica da reforma de 2007 do sistema de pensões em Portugal. Desenvolvida com o intuito de assegurar o equilíbrio financeiro do sistema, a reforma é objecto de crítica por duas razões fundamentais: pelo seu impacto negativo no bem-estar dos pensionistas; e por se afigurar metodologicamente inconsistente na óptica da política social, pois realiza o ajustamento através da diminuição da garantia de rendimento, negligenciando os objectivos sociais do sistema. Finalmente, propõe-se uma metodologia alternativa de formulação de políticas que principia pela definição de objectivos para a garantia de rendimento na reforma.

résumé / abstract

Un processus cohérent de réforme des retraites doit considérer la fonction sociale des systèmes des pensions comme étant fondamental. En conséquence, il doit évaluer les effets des politiques sur les revenus assurés aux retraités. Ainsi, l'article examine la réforme des retraites mise en place au Portugal en 2007. La réforme a eu pour objectif principal d'assurer l'équilibre financier à long terme du système des pensions et peut être critiquée pour deux raisons fondamentales: les effets négatifs des mesures introduites sur le bien-être des retraités; et son incohérence méthodologique du point de vue de la politique sociale, puisque le niveau des prestations a été pris comme variable d'ajustement et, donc, les objectifs sociaux du système ont été négligés. Ensuite, l'article propose une autre méthodologie qui formule, au départ, les objectifs à poursuivre en matière de revenus à assurer aux retraités.

This paper aims to emphasize the idea that the social function of pension systems should be a matter of primary concern when pension policies are designed. Thus, it is fundamental to assess the effects of policies on the income support guaranteed. Taking this view into consideration, the paper develops a critical assessment of the reform of the Portuguese pension system introduced in 2007. Designed to achieve financial sustainability, the reform is criticized for two main reasons: its negative effects on pensioners' welfare; and its incoherent methodology, because the income guarantee was taken as the adjustment variable and, as a result, the social objectives of the system were neglected. Finally, an alternative methodology for policy formulation that takes income security in retirement as the starting point is proposed.

Classificação JEL: D6, H55

Introdução

O nível de rendimentos assegurado aos pensionistas na velhice não tem sido objecto da devida atenção na literatura económica e no discurso político. As análises de cariz económico e político sobre o futuro das pensões têm consagrado como tema central o previsível acréscimo da despesa decorrente do envelhecimento populacional. Esta preocupação tem motivado diversas reformas introduzidas nos sistemas de protecção social, em países da Europa e de outras partes do mundo. Nem sempre têm sido salientadas, porém, as consequências negativas, de um ponto de vista de bem-estar, da tentativa de conter os custos dos sistemas. Os efeitos da contenção da despesa em pensões podem-se revelar particularmente gravosos em países que não garantem ainda uma considerável segurança de rendimento à maior parte da população reformada. Tal é o caso português, em que o âmbito material dos direitos garantidos é ainda limitado.

Este artigo desenvolve uma reflexão crítica sobre as reformas recentemente introduzidas no sistema público de pensões em Portugal. No ponto 1, descrevem-se as alterações regulamentares recentes, referindo as características gerais do sistema em vigor antes da introdução das reformas e as linhas gerais do novo esquema. No ponto 2, efectua-se uma análise crítica das reformas mencionadas destacando dois aspectos: o seu impacto negativo no bem-estar dos pensionistas; e a sua formulação, metodologicamente inconsistente na perspectiva da política social, por negligenciar os objectivos sociais do sistema. No ponto 3, efectua-se uma breve alusão a estratégias alternativas para lidar com o envelhecimento demográfico. No ponto 4, propõe-se uma metodologia para a formulação de políticas que se concentra na função social do sistema e, consequentemente, toma como ponto de partida a garantia de rendimento na reforma. Finalmente, no ponto 5, apresentam-se algumas notas conclusivas.

1. As Alterações Regulamentares Introduzidas

Neste ponto analisam-se os traços fundamentais da regulamentação do regime geral de segurança social dos trabalhadores dependentes do sector privado, no que se refere às condições de atribuição da pensão e ao respectivo método de cálculo, comparando o esquema anterior com o que foi recentemente introduzido.

O esquema que vigorava anteriormente reconhecia o direito à pensão de velhice quando o beneficiário atingia a idade legal da reforma, de sessenta e cinco anos, e um período mínimo de registo de remunerações ou equivalente[1] (períodos de doença, maternidade, assistência à família, serviço militar, etc.), de quinze anos, com, pelo menos, cento e vinte dias de densidade contributiva.

Com a entrada em vigor do Decreto-Lei n.º 329/93, a regra de cálculo da pensão passou a ser a seguinte: $PE = 2\% \times n \times RR$, onde a sigla "PE" designa a pensão estatutária, "n" o número de anos de contribuições e "RR" a remuneração de referência. A última determinava-se tendo em conta as remunerações dos dez anos de melhores remunerações dos últimos quinze anos de actividade. As remunerações consideradas passam então a ser actualizadas de acordo com coeficientes definidos por Portaria, tendo em conta o índice de preços no consumidor sem

1 Ao longo do tempo as condições de atribuição das pensões foram sofrendo alterações. O Decreto-Lei n.º 329/93, de 25 de Setembro, introduziu alterações significativas a este nível. É de salientar que a idade legal da reforma das mulheres, até 1994, era mais baixa, sessenta e dois anos, tendo aumentado gradualmente (seis meses em cada ano) para sessenta e cinco anos. No que se refere ao prazo de garantia, os requisitos modificam-se ao longo do tempo. Sendo inicialmente exigido um prazo de garantia de cinco anos, entre 1988 e 1994, a lei passa a exigir um mínimo de dez anos de contribuições, independentemente do número de meses com registo de remunerações.

habitação[2]. O montante estatutário da pensão deveria necessariamente estar compreendido entre 30% e 80% da remuneração de referência. A lei, porém, vem garantir um nível mínimo de pensão. Em consequência, o montante da primeira pensão pode ultrapassar o valor definido pela fórmula de cálculo, devido às regras de garantia de níveis mínimos. A partir de 1999[3], os níveis mínimos passam a ser escalonados em função do número de anos de contribuições, correspondendo a diferentes percentagens da remuneração mínima nacional.

A lei garantia a indexação regular das prestações. Os acréscimos periódicos ou extraordinários das pensões seguiram o princípio da discriminação positiva, que envolve um acréscimo relativo superior das pensões mais baixas. Esta política permitiu um acréscimo de poder de compra das prestações de nível mínimo. Assim, entre 1986 e 2005, o crescimento nominal do nível mínimo de pensão foi superior à taxa de inflação (ver Quadro 1, em Anexo).

Os níveis mínimos de pensões cresceram com o propósito explícito de indexar o nível mínimo da pensão à remuneração mínima nacional. A Portaria n.º 800/98, de 22 de Setembro, consagrou o direito a um valor mínimo de pensão de velhice equiparado ao valor da remuneração mínima nacional (deduzido da taxa contributiva do trabalhador dependente) para os titulares que tenham cumprido uma carreira contributiva completa. Para os pensionistas com carreira incompleta, são definidos escalões de níveis mínimos que representam uma percentagem da remuneração mínima nacional (líquida da contribuições). A Lei n.º 32/2002, de 20 de Dezembro, definiu um período de transição para alcançar o objectivo de convergência, findo o qual o nível mínimo das prestações deveria igualar a remuneração mínima (líquida de contribuições). Em 2005[4], os beneficiários que apresentavam mais de trinta anos de contribuições atingem efectivamente tal objectivo e os que apresentavam carreiras mais curtas passam a receber prestações que correspondem a diferentes percentagens desse valor (entre 65% e 100%).

No ano de 2002, ocorre uma importante alteração regulamentar que modifica a fórmula de cálculo das prestações[5]. As novas regras estabelecem que o nível da pensão passe a ser calculado com base numa remuneração de referência dependente das remunerações de toda a carreira contributiva. Contudo, a lei definia um período de transição, entre 2002 e 2016, ao longo do qual o método de cálculo mais favorável poderia ser usado para determinar o valor das pensões. Assim, os efeitos da introdução de novas regras não estavam a ser plenamente sentidos em 2007, ano em que é introduzida uma nova lei. O Decreto-Lei n.º 187/2007, de 10 de Maio, vem modificar a regra de cálculo das pensões de velhice e acelerar a transição para a fórmula de cálculo que faz depender a remuneração de referência e, assim, o nível da pensão, das remunerações da toda a carreira contributiva.

No que diz respeito às alterações legislativas recentes, entre as medidas introduzidas[6], salientam-se três que se sumariam de seguida.

Uma das medidas consistiu na introdução de um "factor de sustentabilidade", que provoca a redução do valor da pensão à medida que a esperança de vida aumenta. Tal factor é definido pelo quociente entre a esperança de vida em 2006 e a esperança de vida no ano anterior à reforma. Este coeficiente, com valor inferior à unidade e decrescente à medida que a esperança de vida aumenta, será multiplicado pelo valor estatutário da pensão. A operação referida irá

2 Antes da entrada em vigor do Decreto-Lei n.º 329/93, de 25 de Setembro, não era efectuada a actualização das remunerações anteriores.
3 Portaria n.º 1018/98, de 4 de Dezembro.
4 Portaria n.º 1316/2005, de 22 de Dezembro.
5 Decreto-Lei n.º 35/2002, de 19 de Fevereiro, de acordo com a Lei n.º 32/2000, de 8 de Agosto.
6 Na análise que se segue refere-se apenas ao regime geral de segurança social. É de salientar que foram tomadas medidas de relevo tendentes a harmonizar o esquema que abrange os beneficiários da Caixa Geral de Aposentações com o esquema público. Ver Lei n.º 60/2005 de 29 de Dezembro. É de salientar que o Decreto-Lei n.º 187/2007 também se aplica ao cálculo das pensões dos trabalhadores da Administração Pública admitidos depois de 1 de Setembro de 1993, nos termos do Decreto-Lei n.º 286/93.

produzir uma redução gradual do valor da pensão em relação ao seu nível estatutário. A lei introduz, porém, mecanismos que podem compensar o decréscimo tendencial do valor das prestações. Para tal, será necessário que os trabalhadores prolonguem o período activo ou que aceitem aumentar as suas contribuições.

A referida mudança da regra de cálculo das pensões é a medida que suscita uma maior reflexão. O direito à pensão de velhice é reconhecido se o beneficiário apresentar um período de contribuições mínimo de quinze anos. A pensão é determinada a partir de uma remuneração de referência que passa a depender das remunerações auferidas ao longo de toda a carreira contributiva. Desde 1994, os níveis de pensões são calculados com base no valor médio das remunerações auferidas nos melhores dez anos de remunerações dos últimos quinze anos de actividade. Tal regra tinha sido já modificada pelo Decreto-Lei n.º 17/2000, de 8 de Agosto, que estabelecia que o cálculo da pensão considerasse, gradualmente, as remunerações de toda a carreira contributiva (art. 57.º-3). Esta alteração regulamentar relativamente ao esquema introduzido pelo Decreto-Lei n.º 329/93, de 25 de Setembro, é significativa. O novo método de cálculo só foi adoptado com a entrada em vigor do Decreto-Lei n.º 35/2002, de 19 de Fevereiro. Como, porém, a lei estabelecia um período de transição, durante o qual os dois métodos, ou uma combinação de ambos, poderiam ser aplicados para o cálculo da pensão[7], de forma a garantir aos beneficiários o regime mais favorável, o verdadeiro impacto de tal medida foi reduzido[8]. A pensão podia ser calculada de três formas: de acordo com as regras de cálculo anteriores, sendo a remuneração de referência equivalente à média dos melhores dez anos de remunerações dos últimos quinze; de acordo com a nova regra, dependendo das remunerações de toda a carreira contributiva; ou com base numa média ponderada dos dois valores anteriores.

A reforma introduzida pelo Decreto-Lei n.º 187/2007, de 10 de Maio, vem acelerar a transição para a nova regra de cálculo. Doravante, os beneficiários, inscritos até ao final do ano de 2001, terão a respectiva pensão calculada com base numa média ponderada de duas componentes: a primeira determina-se tendo em conta os melhores dez anos de remunerações dos últimos quinze anos de actividade; a segunda considera as remunerações de toda a carreira contributiva. A ponderação de cada componente depende do momento da reforma (anterior ou posterior a 31/12/2016) e do número de anos de contribuições anterior e posterior a um momento de referência (31/12/2006, para os que se reformem até 2016, e 31/12/2001, para os que se reformem depois dessa data). Para os beneficiários inscritos depois de 31/12/2001, a pensão será calculada com base nas remunerações de toda a carreira contributiva.

O terceiro aspecto que merece destaque relaciona-se com o método de indexação das prestações. Em 2006, foram adoptadas novas regras neste domínio[9], com efeitos a partir do ano seguinte. Por um lado, a indexação das prestações deixa de estar dependente apenas da inflação, passando a relacionar-se, também, com o ritmo de crescimento económico. Por outro lado, a remuneração mínima nacional deixa de ser tomada como referência para a indexação das prestações de nível mínimo. Um novo termo de referência, o "Indexante de Apoios Sociais" (IAS), passa a ser usado para calcular as prestações e a sua actualização. O valor do IAS é definido anualmente pelo governo. Em 2007, o IAS foi determinado com base no valor da remuneração mínima nacional e do índice de preços no consumidor[10]. Nos anos seguintes, o indexante será ajustado tendo em conta as taxas de crescimento económico e de inflação.

Importa salientar a consequência da substituição da remuneração mínima nacional pelo IAS como termo de referência para a indexação das prestações de nível mínimo: representa a renúncia ao princípio da indexação salarial destas prestações, tratando-se, por isso, de uma alteração fundamental. No que se refere às restantes prestações, o novo método de indexação

7 Ver art. 12.º e 13.º do Decreto-Lei n.º 35/2002, de 19 de Fevereiro.
8 Uma referência à experiência do período entre 2002 e 2004, demonstrativa do impacto reduzido dessa medida, é efectuada adiante. Ver ponto 2.
9 Lei n.º 53-B/2006, de 29 de Dezembro.
10 É de assinalar que o nível do IAS se situa, presentemente, abaixo da remuneração mínima nacional.

não irá garantir a todos os reformados a manutenção do valor real das pensões. Só as pensões de valor inferior a 1,5 IAS terão o seu valor real garantido. O crescimento nominal das outras pensões poderá não acompanhar a evolução dos preços no consumidor. As pensões superiores a 1,5 IAS e inferiores a 6 IAS só irão manter o seu valor real se o crescimento económico ultrapassar 2%. As pensões superiores a 6 IAS só irão manter o seu valor real se o crescimento económico ultrapassar 3%.

É de referir, para finalizar, que o Decreto-Lei n.º 187/2007, de 10 de Maio, introduz novas regras destinadas a encorajar a manutenção em actividade dos idosos. Por um lado, a antecipação da idade da reforma passa a estar sujeita a uma penalização maior, correspondente a 6% de redução da pensão estatutária por cada ano de antecipação (em vez de 4,5%, como anteriormente). Por outro lado, o diploma estabelece incentivos ao prolongamento da vida activa.

2. Análise Crítica das Medidas Introduzidas

2.1. O Efeito das Medidas sobre os Níveis de Rendimentos dos Pensionistas

As três medidas referidas no ponto anterior convergem no sentido de reduzir o montante das pensões, tendo efeitos a dois níveis: na taxa de substituição, definida no momento de passagem à reforma, e na evolução subsequente do rendimento dos pensionistas.

Em primeiro lugar, a adopção de novas regras para o cálculo da pensão – na sequência da introdução do "factor de sustentabilidade" e da nova fórmula de cálculo – irá reduzir a taxa de substituição do rendimento no momento de passagem à reforma. Importa referir, por um lado, que o "factor de sustentabilidade" é, por definição, um coeficiente de redução da pensão. Por outro lado, a alteração da fórmula de cálculo, fazendo depender o nível das prestações não das remunerações finais mas das remunerações auferidas ao longo de toda a carreira contributiva, produzirá também a redução das pensões. Tendo em conta a tendência para o crescimento real das remunerações ao longo do tempo, a remuneração média dos últimos anos ultrapassa, em geral, a remuneração média de toda a carreira contributiva. Em consequência, o nível das pensões, quando determinado pela nova regra, torna-se inferior para a generalidade dos pensionistas de velhice. A experiência recente é demonstrativa: estima-se que "no período entre 2002 e 2004, cerca de 238060 pensões foram atribuídas, das quais 41713 (aproximadamente 17,52%) foram definidas com base na totalidade da carreira contributiva […] garantindo assim aos beneficiários o montante de pensão mais favorável" (European Commission, 2005: 3). De acordo com o texto da lei, acelerar a transição para a nova fórmula de cálculo é desejável porque reforça a natureza contributiva do sistema. A transição é vista como um requisito de justiça. Tal conclusão deriva de uma perspectiva específica sobre a finalidade e os objectivos do sistema de pensões, perfilhada pelos proponentes da reforma, fundamentada numa concepção de justiça particular. É uma concepção baseada num princípio de justiça comutativa, do qual decorre um critério de equidade individual, traduzido no postulado de que cada indivíduo deve receber o equivalente ao seu contributo. Outras perspectivas sobre a finalidade e os objectivos do sistema, fundamentadas em concepções de justiça distributiva, remeterão para critérios de equidade distintos.

Em segundo lugar, os novos mecanismos de ajustamento das prestações, sendo também menos favoráveis que os anteriores, irão afectar negativamente as trajectórias de rendimento no período da reforma. As novas regras não garantem nem a indexação dos níveis mínimos de pensão à remuneração mínima nacional – não estando assegurado que estes cresçam de acordo com evolução geral de rendimentos na sociedade – nem o crescimento nominal das pensões de nível superior a 1,5 IAS ao ritmo da inflação, podendo ocorrer a erosão do respectivo poder de compra.

As novas regras de cálculo da pensão e o novo método de indexação terão inevitavelmente consequências negativas na distribuição do rendimento dos idosos. Assim, a recente reforma pode ser criticada, em primeiro lugar, pelas suas consequências sobre o bem-estar dos

pensionistas. Desenvolvida com o propósito de assegurar a sustentabilidade financeira do sistema público de pensões, a reforma vai envolver, de facto, a contenção da despesa em pensões. Negligenciaram-se, porém, os efeitos negativos das medidas introduzidas sobre os níveis de rendimentos dos pensionistas. Tal aspecto torna-se particularmente preocupante no caso português, pois os direitos garantidos têm ainda um âmbito material limitado, ao contrário do que se passa em muitos países da Europa. A maioria dos pensionistas não atingiu ainda uma significativa segurança de rendimento na reforma. Na transição para a reforma, ocorrem perdas de rendimento consideráveis, devido à integração passada no mercado de trabalho, caracterizada por baixos níveis remuneratórios e carreiras contributivas incompletas. Ainda assim, quer as regras de cálculo das pensões que antes vigoravam – que definiam a taxa de substituição do rendimento na passagem à reforma – quer as regras de indexação das prestações – que delineavam as dinâmicas de rendimento no período da reforma – eram mais favoráveis do que as agora introduzidas.

O reforço do laço entre contribuições e prestações: um requisito de justiça?

Antes de prosseguir, merece a pena discutir a ideia de que o reforço da natureza contributiva do sistema constitui um requisito de justiça. Esta ideia, explicitada no texto da lei, associa-se a uma perspectiva particular sobre a finalidade e os objectivos do sistema de pensões, fundamentada numa concepção de justiça específica. Perspectivas diferentes sobre a finalidade e os objectivos do sistema, fundamentadas em concepções de justiça alternativas, apresentam visões distintas dos requisitos de justiça. Importa, pois, evidenciar os pressupostos normativos que estão subjacentes a estas perspectivas.

As pensões de reforma têm uma natureza contributiva pelo facto de se relacionarem com as remunerações anteriormente recebidas e de requererem contribuições prévias. Os esquemas contributivos podem, no entanto, apresentar diferentes configurações no que se refere ao laço estabelecido entre as contribuições individuais de ciclo de vida e as prestações esperadas. As diferentes opções que podem ser tomadas neste domínio traduzem basicamente diferentes perspectivas sobre a desejável presença nos esquemas de dois tipos de redistribuição do rendimento: a redistribuição individual intertemporal e a redistribuição interindividual. Note-se que a extensão da redistribuição incorporada nos esquemas de pensões reflecte, essencialmente, os limites aceites para a esfera da responsabilidade social. Trata-se, pois, de uma escolha normativa, relacionada com os princípios de solidariedade que podem ser realizados na esfera das pensões.

Reconhecem-se diferentes concepções sobre a extensão desejável da redistribuição interindividual nestes esquemas. Um laço forte entre contribuições e prestações permite aos indivíduos receber na reforma um montante de prestações intimamente relacionado com o valor total das contribuições pagas no período de actividade. Num esquema com estas características a redistribuição interindividual do rendimento é praticamente inexistente, pois é excluída *ex ante*. Pode, porém, ocorrer *ex post*, em resultado da diferente longevidade dos indivíduos.

Muitos autores argumentam a favor de um laço forte entre contribuições e prestações. Os esquemas de pensões deveriam operar como um mecanismo de transferências individuais intertemporais: a redistribuição do rendimento processar-se-ia entre o período de actividade e o período da reforma. Assim sendo, este mecanismo de transferências deveria ser avaliado segundo um critério de equidade individual. O valor actual da esperança das prestações totais a receber deveria manter um laço estreito com o valor actual das contribuições efectuadas ao longo da carreira contributiva. A presença de mecanismos de redistribuição interindividual deveria ser excluída destes esquemas.

Apesar de os defensores deste modelo reconhecerem que o sistema desempenha duas funções, a redistributiva e a seguradora, argumentam que a fronteira entre ambas pode ser delimitada com clareza. A redistribuição interpessoal deveria ser realizada apenas através de prestações de natureza não contributiva (universais ou sujeitas a condição de recursos). A

esfera da responsabilidade social deveria restringir-se a este domínio. As prestações contributivas, por seu turno, situar-se-iam na esfera da responsabilidade individual. Desta forma, as prestações recebidas por um qualquer indivíduo deveriam estar estritamente relacionadas com as suas contribuições prévias. Esta perspectiva fundamenta-se numa concepção ética que atribui valor fundamental ao respeito pelos direitos de propriedade: as contribuições equivalem à constituição de uma poupança individual prévia que legitima a atribuição das prestações futuras.

Há, porém, visões alternativas sobre a natureza do laço entre contribuições e prestações[11]. Outros autores sublinham que os esquemas contributivos não necessitam de manter uma proporcionalidade estreita entre contribuições e prestações. Os esquemas de pensões não são encarados como um mecanismo de poupança forçada que permite aos indivíduos transferir recursos para si mesmos, entre o período de actividade e o período da reforma. Apesar de as prestações serem definidas numa base comutativa – pois relacionam-se com o estatuto laboral, o seu propósito é substituir remunerações cessantes e têm como requisito de atribuição a existência de contribuições prévias – o seu valor total esperado não terá de manter uma proporcionalidade estrita com o montante das contribuições efectuadas. Num esquema público podem estar presentes mecanismos de redistribuição interindividual. A função redistributiva pode ser desempenhada não apenas pela componente não contributiva da estrutura das prestações mas também pela sua componente contributiva. Assim, pode estar envolvido um certo grau de redistribuição nos esquemas contributivos. A redistribuição pode ocorrer *ex post*, em resultado das diferenças de longevidade individuais ou ser especificada *ex ante*. De facto, estes mecanismos incluem frequentemente regras que estabelecem níveis mínimos de pensões ou normas que garantem taxas de substituição superiores para os rendimentos mais baixos. É de referir, também, que a regra de cálculo, por vezes, enfraquece deliberadamente o laço entre as contribuições pagas ao longo de toda a carreira e as prestações recebidas. Quando o objectivo é garantir um certo nível de bem-estar, o valor das pensões pode ser calculado com base nas remunerações recebidas nos últimos anos de actividade e não a partir das remunerações da totalidade da carreira contributiva.

A configuração de um esquema contributivo, em particular no que diz respeito à natureza do laço estabelecido entre contribuições e prestações, pode, pois, ser diferenciada. Pressupostos normativos distintos irão condicionar a escolha de diferentes configurações para estes esquemas. Se a primeira concepção de justiça analisada confere um valor fundamental ao respeito pelos direitos de propriedade (pois, segundo ela, a atribuição das prestações futuras é legitimada pela constituição de uma poupança individual prévia), uma concepção ética de inspiração igualitarista que privilegie outros valores – como a liberdade, entendida não apenas na sua formulação negativa mas na sua faceta positiva[12], ou a realização das potencialidades humanas, etc. – pode ser mais favorável à aceitação de mecanismos de redistribuição interindividuais. Privilegiar outros princípios e valores pode conduzir à formulação de objectivos alternativos – como, por exemplo, realizar uma certa paridade de níveis de vida entre pensionistas e activos – e outras configurações para o sistema de pensões.

2.2. Uma Crítica Metodológica

A reforma em análise pode também ser objecto de crítica metodológica por não definir objectivos para o nível das prestações, que se torna a variável de ajustamento. Uma política

11 Ver Thompson (1983). Para uma discussão da base normativa das pensões, ver Gillion *et al*. (2000).
12 O conceito de liberdade pode ter duas formulações, uma negativa – entendida como ausência de restrição – e outra positiva. Na sua faceta positiva, a liberdade é entendida como capacidade de agir e de realizar os seus propósitos fundamentais ou, na terminologia de Sen, como capacidade de escolha do modo de vida que os indivíduos valorizam. O pensamento liberal tem subjacente uma definição negativa de liberdade. O papel do Estado deve ser reduzido na medida em que impõe limites à acção individual. Os críticos do liberalismo, por seu turno, argumentam que a liberdade positiva é um produto da organização social e que a sua promoção requer a intervenção pública.

social coerente neste domínio requer uma especificação clara dos objectivos sociais a atingir. Ora, as pensões de reforma (de natureza contributiva) são mecanismos de garantia de substituição do rendimento de actividade. Por isso, importará formular objectivos em dois domínios: a taxa de substituição do rendimento, no momento de passagem à reforma, e a indexação das prestações ao longo do tempo.

No texto da lei[13], porém, não há qualquer referência aos objectivos a atingir em matéria de rendimento garantido aos pensionistas nem às consequências, a esse nível, das medidas adoptadas. Em 2006, o Ministério da Solidariedade e Segurança Social publicou um documento em que revela as linhas estratégicas da reforma[14], onde, também, nenhuma referência é efectuada aos objectivos referidos. No final desse ano, o Ministério das Finanças e da Administração Pública, no Relatório do Orçamento de Estado para 2007[15], irá divulgar as estimativas dos efeitos da reforma no que se refere à redução de prestações. No Anexo III desse documento, é efectuada uma análise da sustentabilidade da segurança social. As reformas programadas são consideradas necessárias para promover a redução futura da despesa em pensões e apresentadas como solução para garantir a sustentabilidade financeira do sistema público.

De acordo com as projecções do governo, a despesa no conjunto do subsistema previdencial deveria crescer cerca de 2.4% do PIB nos próximos vinte anos. Em 2006, o subsistema apresentava ainda um excedente mas, em 2025, iria apresentar um défice de -1,9% do PIB. A despesa em pensões (velhice, invalidez, sobrevivência) iria crescer de 5,5% para 7,9% do PIB. As pensões de velhice deveriam crescer de 4% do PIB, em 2006, para 6,3%, em 2025. De acordo com as projecções oficiais, as novas regras permitem a redução futura da despesa em pensões e do crescimento real da pensão média (Gráfico 1, em Anexo). Estes efeitos são considerados positivos na perspectiva do Orçamento de Estado.

Esta metodologia, porém, na perspectiva da política social, revela-se inconsistente. O objectivo explícito da reforma é equilibrar financeiramente o subsistema previdencial. De modo a garantir a realização de tal propósito, o nível das prestações torna-se a variável de ajustamento.

Por um lado, é de salientar que, na óptica da política social, o equilíbrio financeiro a longo prazo do subsistema previdencial deve ser perspectivado como "a restrição" a respeitar e não como "o objectivo primordial a atingir". Por outro lado, uma metodologia coerente de definição de medidas de política neste domínio exige a especificação clara dos objectivos a atingir no que se refere à substituição de rendimento na reforma. A função social do sistema consiste em fornecer tal garantia. Por isso, na perspectiva da política social, o nível de rendimento assegurado na reforma deve ser considerado variável objectivo. Consequentemente, não deverá ser tomado como um aspecto residual numa política traçada com o propósito de assegurar a sustentabilidade financeira do sistema. Os objectivos a atingir no que se refere à garantia de rendimento na reforma deverão ser tratados, sempre, como princípios fundamentais orientadores de qualquer reforma. Tais objectivos devem ser claramente formulados *a priori* e, do meu ponto de vista, tomados como ponto de partida.

Importa referir ainda que, conforme se verá no ponto seguinte, a política de redução da despesa em pensões não constitui a única via possível para assegurar o equilíbrio financeiro do subsistema previdencial. É possível realizar tal propósito, recorrendo a outras variáveis de ajustamento.

3. Estratégias Alternativas para Lidar com o Envelhecimento Demográfico

Neste ponto efectua-se uma referência breve às direcções alternativas que as políticas de pensões podem seguir para enfrentar a longevidade crescente. A discussão das alternativas é

13 Decreto-Lei n.º 187/2007, de 10 de Maio.
14 Ministério do Trabalho e da Solidariedade Social (2006).
15 Ministério das Finanças e da Administração Pública (2006).

efectuada a partir de uma óptica macroeconómica[16]. Para desenvolver a análise, importa efectuar uma alusão prévia à função económica das pensões, distinguindo a perspectiva dos indivíduos da perspectiva agregada. A partir da reflexão subsequente conclui-se que o rumo seguido em Portugal não constituía a única solução possível, havendo diversas formas de ajustar os esquemas de pensões neste contexto.

3.1. A Função Económica das Pensões

A Perspectiva dos Indivíduos

Na perspectiva dos indivíduos, a função económica das pensões consiste em assegurar a afectação intertemporal do consumo, ou seja, transferir consumo do período activo para o período da reforma. A obrigação de contribuir para um esquema de pensões reduz o consumo no período de actividade, com o propósito de garantir o consumo no período da reforma.

Só há duas formas de transferir consumo no tempo: armazenar produção corrente para utilização futura ou trocar a produção corrente por um direito à produção futura (Barr, 2004). Só a segunda é exequível. Para tal, podem ser seguidos dois métodos: poupando uma parte do salário, o trabalhador pode adquirir activos que irão ser trocados, no período da reforma, por bens produzidos pelas gerações activas; ou obtendo uma promessa (dos filhos, do empregador, do Estado) de atribuição futura (no seu período da reforma) de bens produzidos pela população activa. Os dois modos essenciais de organizar as pensões aproximam-se destes métodos: a capitalização baseia-se na acumulação de activos financeiros e a repartição numa promessa.

A Perspectiva Agregada

a) O custo económico da população reformada na perspectiva macroeconómica

Na perspectiva da sociedade no seu conjunto, o custo económico da população reformada equivale à fracção do rendimento total utilizada para seu consumo (Thompson, 1998). Os bens e serviços consumidos pela população reformada são, assim, a medida do seu custo económico.

Os rendimentos gerados no processo produtivo remuneram os factores de produção, sendo esses rendimentos utilizados, em parte ou na totalidade, para consumir os bens e serviços produzidos no período. Tendo em consideração que os reformados se encontram em situação de inactividade, importa questionar como pode o seu consumo ser assegurado. Os bens e serviços por estes consumidos são produzidos em períodos próximos daquele em que são consumidos. Contudo, o consumo dos reformados não se pode basear no rendimento de trabalho, tendo de assentar noutros recursos. Por um lado, se os reformados forem detentores de capital, uma parte do seu consumo pode ser resultante da remuneração do capital. Por outro lado, o seu consumo pode resultar de transferências de recursos realizadas por aqueles que continuam a trabalhar. Este tipo de transferências representa a parte mais expressiva do conjunto dos recursos dos reformados[17].

As sociedades utilizam combinações diferentes de três mecanismos para realizar as transferências de poder de compra da população em idade activa para os reformados: transferências informais intrafamiliares; regimes de contribuições obrigatórias, que incidem sobre o rendimento dos activos, destinando-se a financiar as prestações pagas aos reformados; transacções de activos físicos ou financeiros entre a população reformada (vendedora) e a

16 Neste ponto segue-se de perto a análise desenvolvida por Barr (2004). Sobre a economia das pensões, ver Thompson (1998), Barr (2004), Barr (2006), Barr e Diamond (2006).
17 Os rendimentos de capital suportam apenas minoritariamente o consumo da população reformada. Dois factores explicam este facto: o reinvestimento dos rendimentos de capital, por um lado, e a forma como a propriedade do capital se encontra distribuída entre os membros da sociedade, por outro. Ver Thompson (1998: 38).

população activa (compradora). Note-se que um sistema de capitalização pode ser visto como mecanismo do último tipo. A população em idade activa concebe tal mecanismo como um modo de formação de poupança para a reforma quando, de facto, corresponde à aquisição de activos.

Os três mecanismos suscitam uma redução voluntária ou involuntária do consumo da população activa (inferior à totalidade do rendimento gerado no processo produtivo) para possibilitar o consumo dos reformados (superior ao respectivo rendimento com origem no processo de produção).

b) Os métodos de organização de direitos sobre o produto futuro

A capitalização é um método de acumulação de activos financeiros cujo rendimento permite adquirir bens numa data posterior. As contribuições para estes esquemas são investidas em activos financeiros, sendo o respectivo retorno creditado no fundo de pensões. As reservas constituídas nos fundos de capitalização são utilizadas para fazer face às responsabilidades, isto é, para efectuar o pagamento das prestações.

Os esquemas de repartição são desenvolvidos pelo Estado e têm natureza legal. Tais esquemas operam transferências instantâneas entre trabalhadores e pensionistas, pois as prestações pagas aos reformados, em cada período, provêm das contribuições efectuadas pelos trabalhadores no mesmo período. Na perspectiva individual, o direito à pensão baseia-se num contrato implícito: o pagamento de cotizações no presente confere o direito ao recebimento de prestações no futuro. Porém, numa perspectiva macroeconómica, o Estado está apenas a promover transferências correntes de um grupo da população para outro.

O argumento de que os esquemas de capitalização protegem melhor os indivíduos do que os esquemas de repartição em relação às alterações demográficas não foi comprovado[18]. Essa ideia constitui, segundo Barr (2004: 194), um exemplo ilustrativo do sofisma da composição. Na perspectiva individual, um esquema de pensões efectua uma transferência intertemporal de consumo. Porém, tal transferência não é possível à escala social: o consumo dos reformados provém sempre, e em última instância, dos bens produzidos pela futura geração de trabalhadores. De um ponto de vista agregado, a função económica dos esquemas de pensões consiste em repartir o produto total de um período entre trabalhadores e reformados – o consumo dos primeiros é reduzido para permitir o consumo dos segundos. A repartição, porém, evidencia de forma inequívoca que as pensões envolvem recursos correntes.

Numa visão macroeconómica, a variável fundamental é o produto. O objectivo dos sistemas de pensões é permitir aos reformados continuar a consumir no período de inactividade. Como os bens e serviços por estes consumidos são produzidos nesse período pelos activos, o produto futuro é variável essencial, pois dele provirá o consumo dos reformados futuros. A repartição e a capitalização são apenas mecanismos financeiros de organização de direitos sobre o produto futuro[19].

3.2. Políticas para Enfrentar o Envelhecimento Populacional

De um ponto de vista macroeconómico, o custo económico da população reformada é equivalente à fracção do rendimento total utilizada para seu consumo, C_r/Y. Deste modo, a longevidade crescente tende a elevar este quociente. Para enfrentar o aumento da esperança de vida, identificam-se dois tipos de políticas: políticas do lado da procura (com efeitos no numerador da expressão) e políticas do lado da oferta de bens e serviços (com efeitos no denominador).

18 Para uma síntese do debate entre capitalização e repartição, ver, por exemplo, Blanchet e Villeneuve (1997).
19 Assim sendo, a superioridade da capitalização, como método de financiamento das pensões, poderia ser comprovada caso se demonstrasse a sua maior capacidade de suscitar o crescimento do produto. O argumento será adiante referido.

No que se refere às políticas do lado da procura, a repartição da despesa em bens de consumo, entre activos e reformados, pode ser modificada na sequência do aumento das contribuições ou da redução das prestações. No primeiro caso, o ónus do ajustamento recai exclusivamente sobre os trabalhadores. No segundo caso, o ónus do ajustamento incide na íntegra sobre os reformados. É ainda possível recorrer a uma terceira solução: o aumento da idade da reforma. Esta medida pode potenciar dois efeitos simultâneos: o aumento do número de trabalhadores e a redução do número de reformados, dependendo a sua viabilidade do contexto do mercado de trabalho. A ser concretizada, tal via permite a contenção da despesa e, ao mesmo tempo, evita a redução dos níveis de vida dos pensionistas. O prolongamento da vida activa pode apresentar-se como uma alternativa à contenção do consumo. Assim, do lado da procura, configuram-se diversas soluções: reduzir as pensões, aumentar as contribuições, ajustar a idade da reforma ou uma combinação das três medidas. Como Barr (2006: 13) observa, "a um dado nível de produto nacional, estas três políticas – reduzir pensões, elevar contribuições ou aumentar a idade da reforma – são apenas mecanismos para dividir o *output* de diferentes formas entre trabalhadores e pensionistas".

É também possível desenvolver medidas do lado da oferta. As políticas que promovem o crescimento económico contribuem para atenuar a pressão decorrente do envelhecimento populacional, pois permitem o crescimento do consumo *per capita*. O acréscimo do produto pode ser atingido através de duas estratégias fundamentais: o aumento do produto por trabalhador ou o aumento do número de trabalhadores. A elevação do produto por trabalhador pode resultar de acréscimos na quantidade e qualidade do capital e da melhoria da qualidade do trabalho. O acréscimo do número de trabalhadores pode ser atingido através da redução do desemprego, elevando a idade da reforma, aumentando a participação feminina no mercado de trabalho ou através da imigração.

No tocante às estratégias para enfrentar o envelhecimento demográfico, dado que o produto é uma variável central, as políticas que promovem o crescimento são decisivas na resposta à questão (Barr, 2004). Uma regulação macroeconómica mais favorável ao crescimento e ao emprego terá um importante papel nessa resposta.

Face ao exposto, torna-se pertinente o confronto das modalidades de financiamento das pensões, na medida em que estas podem, eventualmente, influenciar o nível de produção. Os partidários da capitalização argumentam que esta seria superior à repartição por potenciar o crescimento económico. A vantagem da capitalização adviria do facto de suscitar acréscimo de poupança, que se iria traduzir em acréscimo do investimento produtivo e, por essa via, do produto. No entanto, os estudos teóricos e empíricos desenvolvidos sobre a matéria são inconclusivos e controversos. Carece de demonstração que a capitalização seja um método eficiente para promover o investimento produtivo e, por essa via, o crescimento económico[20]. Em primeiro lugar, não é certo que a capitalização eleve a poupança, pois um acréscimo da poupança obrigatória pode ser contrabalançado pela redução da poupança voluntária. Em segundo lugar, mesmo que se verifique um aumento da poupança, não está comprovado que este envolva um aumento do investimento produtivo. Por um lado, não é certo que a poupança se converta automaticamente em investimento produtivo. Por outro lado, se o ritmo de crescimento económico for lento, a elevação da poupança poderá ter efeitos depressivos sobre o nível de actividade económica. Finalmente, o aumento do investimento produtivo pode não ter um efeito significativo sobre o *output*.

A capitalização, enquanto mecanismo destinado a promover o crescimento económico, opera de forma indirecta e com resultados indeterminados. Assim sendo, Barr (2004) recomenda que se considerem prioritariamente as políticas que promovem directamente o crescimento do produto.

Conclui-se que podem ser desenvolvidas diversas soluções para enfrentar os problemas futuros suscitados pelo envelhecimento demográfico. Havendo a possibilidade de recorrer a políticas do

20 O argumento é examinado, por exemplo, em Barr e Diamond (2006).

lado da procura e a políticas do lado da oferta, na discussão sobre as reformas não se efectuou qualquer referência às segundas. No que respeita às políticas do lado da procura, a possibilidade de elevar as contribuições não foi sequer considerada[21] e excluiu-se o aumento da idade da reforma[22].

O governo português decidiu seguir uma única via: reduzir as pensões, fazendo com que os pensionistas suportem na íntegra o encargo do ajustamento. O rumo político seguido afigura-se criticável, do meu ponto de vista, dada a especificidade da situação dos pensionistas no país. Como é sabido, o seu nível de vida relativo[23] é significativamente inferior ao da população em geral e o seu risco de pobreza bastante superior[24]. Neste contexto, a opção anterior afigura-se particularmente gravosa por colocar o encargo do ajustamento exclusivamente sobre os pensionistas.

A estratégia de reduzir pensões

É de mencionar que a política de contenção da despesa em pensões ocorre em simultâneo com a introdução de novos mecanismos dirigidos aos pensionistas mais pobres, como uma prestação especial, sujeita a condição de recursos, designada "complemento solidário para idosos"[25], destinada a atenuar situações de pobreza.

Importa salientar que a estratégia de reduzir pensões, em simultâneo com o reconhecimento explícito de que a prevenção da pobreza entre os idosos é prioridade política, representa um ponto de viragem na filosofia do sistema. Reflecte uma mudança radical de perspectiva, pois a intervenção pública fica focalizada num único objectivo, o de prevenir a pobreza, negligenciando o outro objectivo fundamental dos sistemas de pensões, i.e. a salvaguarda dos níveis de vida no período da reforma.

21 A possibilidade de elevar as contribuições para o sistema foi liminarmente excluída. Dois argumentos essenciais, não validados teórica ou empiricamente, sustentam o dogma, hoje generalizado, da impossibilidade de aumento das contribuições. Por um lado, a elevação das contribuições representa um aumento do custo do trabalho, o que conduz à perda de competitividade e desencadeia uma redução da rendibilidade das empresas. Desta forma, tal solução seria prejudicial ao emprego e, em última análise, aos próprios sistemas de pensões. Este argumento é discutido em Thompson (1998). Por outro lado, a elevação das contribuições iria reduzir a poupança e, em consequência, o investimento produtivo e a produção. A estes dois argumentos, adiciona-se, por vezes, um terceiro, o da resistência política da população ao aumento das contribuições. Sobre o dogma da impossibilidade de aumentar as contribuições ver Math (2001).

22 É de observar que, no actual contexto de crescimento e emprego, seria dificilmente defensável que a idade da reforma fosse tomada como variável de ajustamento.

23 Ver, por exemplo, European Commission (2006). De acordo com os dados aí publicados, em muitos países, a posição relativa de rendimento dos idosos é próxima da média da população em geral, ou mesmo superior. Em Portugal, a relação entre o rendimento mediano equivalente dos idosos e das *coortes* mais jovens é apenas de 76%.

24 O risco de pobreza é medido em termos relativos e definido como a percentagem de indivíduos com rendimento disponível equivalente inferior a um limiar de 60% da mediana nacional. A incidência da pobreza entre os indivíduos com mais de 65 anos ultrapassa os 29%, número substancialmente superior ao da população em geral, que ronda 19%.

25 Decreto-Lei n.º 232/2005, de 29 de Dezembro. No preâmbulo deste diploma, lê-se: "[o] complemento solidário para idosos traduz uma verdadeira ruptura com a anterior política de mínimos sociais para idosos, através de uma aposta na concentração de recursos disponíveis nos estratos da população idosa com menores rendimentos, na atenuação das situações de maior carência […] e na solidariedade familiar". O texto da lei afirma estas prestações como um elemento de ruptura que vem reconfigurar a política de mínimos sociais. Discordo profundamente da lógica subjacente a esta medida. As prestações sujeitas a condições de recursos deveriam, na minha perspectiva, ter um carácter residual, pela sua natureza estigmatizante, no âmbito de um sistema de garantia de níveis mínimos de rendimento. Ora, de acordo com o texto do diploma, a introdução desta prestação poderá vir a pôr em causa o aumento generalizado do valor das pensões mínimas. A concretizar-se este intuito, operar-se-á, a meu ver, um processo de regressão social inaceitável.

Este é um objectivo fundamental do sistema, ou seja, assegurar que as pessoas possam manter na reforma, na medida do possível, os níveis de vida alcançados no período de actividade. Porém, o facto de se admitir ao mesmo tempo o decréscimo programado das pensões médias, por via da redução das taxas de substituição para a maioria dos reformados, e a desindexação dos rendimentos dos pensionistas da evolução geral de rendimentos e preços, com a consequente redução do seu nível de vida relativo, comprova que este objectivo foi negligenciado.

A reforma é inspirada por uma visão dominante, porém contestável. "A protecção oferecida pelo Estado de bem-estar é frequentemente discutida em termos de redução de pobreza, mas esta é uma visão muito limitada das suas funções. A redução da pobreza é um importante objectivo, mas é só um dos propósitos de programas como as pensões de reforma […]" (Atkinson, 1999: 5). No mesmo sentido, a propósito do título, aparentemente paradoxal, do seu artigo "*The Welfare State versus the Relief of Poverty"*, afirma Barry (1990: 73): "[N]um Estado de bem-estar bem organizado quase toda a tarefa de reduzir a pobreza será efectuada através de políticas cujos objectivos e racionalidades são bastante diferentes. […] Se o Estado de bem-estar tiver de ser identificado com um objectivo, será o da manutenção de rendimento e não o da redução da pobreza".

4. Uma Metodologia Alternativa

Concluiu-se que os objectivos a atingir em matéria de garantia de rendimento na reforma deverão ser tratados, sempre, como princípios fundamentais orientadores da arquitectura dos esquemas de pensões e que tais objectivos devem ser claramente formulados *a priori*. Em conformidade com este princípio, propõe-se, no presente tópico, uma metodologia alternativa para configurar os esquemas de pensões que principia pela definição de objectivos para os níveis de rendimento garantidos na reforma. Nesta matéria, o objectivo social estabelecido pode ser muito diferenciado: desde a garantia de níveis mínimos de pensões até à garantia de manutenção dos níveis de vida anteriormente atingidos.

A definição dos objectivos é uma escolha fundamental que remete, evidentemente, para o domínio normativo. Importa, por isso, explicitar *a priori* um conjunto de pressupostos normativos assumidos. Seguidamente, discute-se a forma como os objectivos gerais se podem traduzir em critérios mais específicos, que representam uma tradução concreta dos objectivos sociais a atingir no domínio dos rendimentos garantidos.

4.1. Os Objectivos das Políticas e a Garantia de Rendimento na Reforma

As pensões de reforma podem ter natureza contributiva ou não contributiva. Em geral, as pensões não contributivas são garantidas a todos os cidadãos, na idade da reforma (prestações universais ou sujeitas a condições de recursos), com o objectivo de evitar ou reduzir situações de pobreza. As pensões de natureza contributiva são garantidas a todos os trabalhadores, a partir do momento em que estes atingem a idade legal da reforma, com o objectivo de substituir o rendimento de actividade. O direito às prestações decorre da realização de contribuições prévias, durante um período relevante, estando o montante da pensão dependente das remunerações anteriores e do número de anos de contribuições. No presente artigo examina-se apenas este tipo de prestações.

Os objectivos das pensões de natureza contributiva

No tocante às prestações de natureza contributiva, garantidas a todos os trabalhadores no momento da reforma, os objectivos das políticas públicas podem variar desde a garantia de níveis mínimos até à salvaguarda dos níveis de vida previamente atingidos. Em consequência, estes esquemas podem ter configurações distintas relacionadas com diferentes finalidades das políticas. Frequentemente, garantem níveis mínimos de pensões aos trabalhadores que apresentam um período mínimo de contribuições, com o propósito de prevenir situações de

pobreza. Este, porém, pode não ser o único objectivo. Em muitos casos, os sistemas têm um âmbito mais alargado, garantindo níveis de pensões calculados com base nos níveis de remunerações anteriores e que são significativamente superiores aos níveis mínimos de rendimento destinados a evitar a pobreza. Tais níveis de prestações são garantidos para permitir preservar, na medida do possível, os níveis de vida anteriormente atingidos[26].

4.2. Uma Visão Normativa sobre a Garantia de Rendimento na Reforma

Antes de apresentar a visão defendida no tocante à garantia de rendimento na reforma, importa clarificar a perspectiva normativa assumida.

Em primeiro lugar, é conveniente explicitar a visão perfilhada no que se refere à configuração dos esquemas contributivos. No ponto 2.1 efectuou-se uma alusão breve às diferentes configurações que estes esquemas podem assumir, em particular no que diz respeito à natureza do laço estabelecido entre contribuições e prestações. A opção entre modelos alternativos é condicionada pelos princípios e valores privilegiados. Perfilho uma visão que se afasta da concepção que advoga o estabelecimento de um laço forte entre contribuições e prestações. Segundo essa perspectiva, o sistema de pensões desempenha duas funções, a redistributiva e a seguradora, podendo a fronteira entre ambas pode ser delimitada com clareza. A redistribuição interpessoal deveria ser realizada apenas através de prestações não contributivas, que caberiam na esfera da responsabilidade social e seriam legitimadas por uma ideia de necessidade. Nos esquemas contributivos, por seu turno, a redistribuição deveria estar ausente. Estes são encarados como mecanismos de transferência de recursos do indivíduo para si mesmo, entre duas fases do "ciclo de vida", cabendo no âmbito da responsabilidade individual. A visão descrita aproxima-se do pensamento liberal. A atribuição das prestações futuras seria legitimada pela constituição de uma poupança individual prévia (as contribuições realizadas). Para respeitar direitos de propriedade, importa manter um laço estreito entre as prestações e as contribuições.

A concepção ética que perfilho atribui um valor fundamental à igualdade de liberdade positiva. Em consequência, o meu pensamento inscreve-se numa corrente mais favorável à inclusão nestes esquemas de mecanismos de redistribuição interindividual. A redistribuição pode ser desempenhada pela componente não contributiva da estrutura das prestações ou pela componente contributiva. As prestações de natureza não contributiva (universais ou sujeitas a condições de recursos) destinam-se a fazer face a situações de necessidade. Os esquemas contributivos, por seu turno, podem também ter objectivos redistributivos, integrando mecanismos como, por exemplo, as regras que definem níveis mínimos de pensões ou a garantia de taxas de substituição superiores para os rendimentos mais baixos. Nesta lógica, não é necessário que o laço estabelecido entre as contribuições realizadas e as prestações recebidas, a nível individual seja estrito. Quando o objectivo das políticas consiste em assegurar um certo nível de bem-estar, a regra de cálculo pode enfraquecer deliberadamente esse laço. Tal acontece, por exemplo, quando o valor das pensões é calculado com base nas remunerações recebidas nos últimos anos de actividade e não a partir das remunerações da totalidade da carreira contributiva.

Em segundo lugar, na minha perspectiva, a justiça entre contemporâneos deve ser considerada uma preocupação ética fundamental, sendo fulcral promover a realização de uma certa paridade de níveis de vida entre pensionistas e activos. Assim, a configuração dos esquemas

26 Verifica-se hoje uma reversão de tendência na fixação de objectivos para as políticas de pensões em relação ao movimento observado, a partir dos anos sessenta do século XX, em muitos países europeus. Na década referida, operou-se uma mudança radical nos objectivos das políticas de pensões: passou-se da garantia de níveis de vida mínimos para a salvaguarda dos níveis de vida prévios. Este movimento ocorreu primeiro na Alemanha e Suécia sendo seguido, nos anos sessenta e setenta, por outros países desenvolvidos. Foi o resultado de uma política deliberada destinada a melhorar os níveis de vida dos pensionistas. Na sequência dessa política, em muitos países europeus e também no Canadá, nos Estados Unidos e no Japão, a diferença entre os níveis de vida médios dos pensionistas e da restante população reduziu-se significativamente. Ver Rein e Turner (1999) e Myles (2002).

de pensões deverá ser delineada em conformidade. No que respeita às finalidades e objectivos das políticas, considero essencial não renunciar à prossecução dos dois objectivos primordiais das pensões: a prevenção da pobreza e a salvaguarda dos níveis de vida no período da reforma. Como se referiu anteriormente, a reforma recente focaliza a intervenção do Estado quase exclusivamente no primeiro objectivo, negligenciando o segundo. A concretizar-se tal tendência, reduzir-se-á significativamente o âmbito da acção pública neste domínio. Com efeito, renuncia-se ao objectivo fundamental de assegurar a manutenção dos níveis de vida no período da reforma, de cuja prossecução depende a possibilidade de realizar a paridade entre os níveis médios de vida dos reformados e dos seus contemporâneos.

Explicitada a posição assumida relativamente aos aspectos normativos referidos, importará analisar as consequências desta visão no tocante à formulação dos objectivos a atingir em matéria de garantia de rendimento na reforma. Dois domínios fundamentais requerem ponderação: a substituição de rendimento que ocorre no momento de cessação de actividade e a indexação do rendimento durante o período da reforma.

Em consonância com perspectiva explicitada, os rendimentos dos pensionistas, a cada momento, devem ser avaliados em termos relativos, tendo em consideração os níveis de rendimento correntes na sociedade a cada momento. Na transição para a reforma, os pensionistas podem sofrer uma perda de rendimento significativa se as pensões médias forem muito inferiores aos níveis remuneratórios anteriores. Consequentemente, o rendimento da pensão pode vir a situar-se significativamente abaixo do nível remuneratório anteriormente atingido e dos níveis médios de rendimento correntes na sociedade nesse momento. A dinâmica de rendimento no período da reforma, por seu turno, irá depender fundamentalmente dos mecanismos de indexação das prestações. O rendimento relativo dos pensionistas pode continuar a diminuir neste período. Assim, a redução da posição relativa de rendimento observada na transição para a reforma pode manter-se ou intensificar-se no período da reforma[27].

Devido à redução do rendimento que ocorre na transição para a reforma ou à forma como as pensões são indexadas posteriormente, os pensionistas podem ter um nível de vida relativamente baixo. Neste caso, a sociedade pode enfrentar um problema de justiça entre contemporâneos. Tal situação virá a ocorrer quando as prestações não são suficientes para evitar a pobreza na velhice ou quando os rendimentos dos pensionistas se situarem significativamente abaixo dos níveis correntes na sociedade. Uma maior incidência de pobreza entre os idosos ou uma significativa disparidade de rendimento dos pensionistas relativamente à população no seu conjunto são, do meu ponto de vista, eticamente relevantes. A justiça entre contemporâneos[28] deverá ser considerada uma preocupação ética fundamental.

Na sequência da adopção desta perspectiva, a definição de objectivos a atingir em matéria de garantia de rendimento – quer para a substituição de rendimento que ocorre no momento de cessação de actividade, quer para a indexação de rendimento ao longo do período da reforma – deverá ser orientada por critérios de natureza relativa.

Na transição para a reforma, importa ter em conta, em primeiro lugar, que o objectivo das pensões consiste em substituir o rendimento de actividade cessante[29]. Assim, um primeiro ponto

27 É de sublinhar que o nível de vida relativo dos pensionistas, em cada momento, reflecte, em geral, não uma situação transitória mas permanente. Por um lado, os rendimentos dos pensionistas são determinados pelas remunerações prévias. A distribuição das pensões reproduz em regra o padrão de desigualdade da distribuição das remunerações. Por outro lado, no período da reforma, os níveis das pensões dependem das regras de indexação mas têm, em geral, uma correlação estreita com os níveis anteriores de pensões.
28 Sobre desigualdades entre contemporâneos, ver Fleurbaey e Michel (1992) e Fleurbaey (2002).
29 Para os pensionistas de rendimentos mais baixos, as prestações garantidas ao longo do tempo deverão permitir, no mínimo, assegurar a sua participação social. O rendimento que permite a participação social é, obviamente, dinâmico e relativo, pois condicionado pelos níveis de rendimento correntes na sociedade, ao longo do tempo. Esta matéria exige análise conceptual específica, que não cabe no âmbito do presente artigo. O tópico é desenvolvido, por exemplo, em Atkinson (1995).

de referência deverá ser constituído pelas remunerações anteriores. Admitindo que o objectivo da política é permitir aos pensionistas manter, na medida do possível, os níveis de vida anteriores, as últimas remunerações devem ser tomadas como referência. Em segundo lugar, de acordo com a perspectiva normativa defendida, os objectivos a atingir deverão ser formulados com base num critério relativo e, em consequência, o segundo ponto de referência deverá corresponder ao nível médio de rendimento corrente na sociedade, no momento. O objectivo a atingir na transição para a reforma, segundo a perspectiva agora explicitada, deveria consistir em evitar uma queda substancial no nível de vida. Em termos concretos, a substituição do rendimento individual, que ocorre nesse momento, deveria garantir pensões médias não muito inferiores ao valor médio das últimas remunerações. Desta forma, os pensionistas conseguiriam manter, na medida do possível, os níveis de vida anteriores. Se a primeira pensão, em média, não se situar muito abaixo do nível remuneratório anterior, cumprir-se-á a condição necessária para assegurar também que o rendimento médio dos pensionistas não se torna significativamente inferior ao rendimento médio corrente na sociedade.

No período da reforma, como se referiu, a dinâmica de rendimento dos pensionistas depende fundamentalmente dos mecanismos de indexação. Estes são o seu principal factor explicativo. Assim, as regras de ajustamento dos níveis de prestações constituem um mecanismo fundamental para evitar uma incidência crescente da pobreza dos idosos e, também, para garantir que o rendimento médio dos pensionistas não se distancia significativamente do rendimento médio corrente na sociedade a cada momento. Para tal, importará que a revalorização das prestações se alinhe não só com a evolução dos preços mas com a evolução geral dos rendimentos.

4.3. Breve Referência à Condicionante Financeira

Não estando no âmbito do presente artigo proceder a uma discussão aprofundada da sustentabilidade financeira futura do sistema público de pensões, justifica-se, porém, uma alusão breve à condicionante financeira.

Uma das principais críticas dirigidas à reforma relaciona-se com o facto de realizar o equilíbrio financeiro futuro do subsistema previdencial através da contenção da despesa em pensões. Ora, a opção tomada não constituía, porém, a única solução para realizar o equilíbrio financeiro do sistema. Se a parte das despesas públicas em pensões tende a aumentar mais rapidamente que o PIB, há diversas formas de promover o equilíbrio financeiro do sistema. Esse propósito pode ser assegurado por via da contenção das despesas, por via do aumento das receitas ou através de uma combinação dos dois tipos de medidas.

Há um conjunto de políticas que podem ser promovidas no lado da despesa. Uma solução possível foi a escolhida pelos mentores da reforma: reduzir a despesa, presente ou futura, em pensões directamente. Esta via coloca o ónus do ajustamento sobre os reformados. Uma outra solução passa por permitir uma despesa futura em pensões superior, sendo esta compensada por uma redução de outras componentes da despesa pública[30]. Uma terceira hipótese consistiria em acumular recursos no presente para fazer face ao previsível aumento da despesa futura em pensões e, assim, reduzir o esforço contributivo futuro[31].

O equilíbrio financeiro presente e futuro, porém, pode realizar-se não apenas por via da contenção das despesas mas também através do aumento das receitas potenciado, por exemplo, pela elevação das contribuições sociais. No entanto, a hipótese de tomar as contribuições como variável de ajustamento não foi sequer ventilada. Generaliza-se, hoje, a posição dogmática segundo a qual é impossível aumentar as contribuições.

30 Esta hipótese é referida por Barr (2001), podendo ser concretizada, por exemplo, através da redução da dívida pública no presente, medida que permitiria reduzir o pagamento dos juros da dívida no futuro e, assim, reduzir o esforço contributivo futuro necessário para suportar uma despesa superior em pensões.
31 Sobre a matéria ver, por exemplo, Barr (2001: 125-126).

De igual modo, o aumento da idade da reforma tem efeitos positivos numa perspectiva orçamental: contribui, em simultâneo, para elevar as receitas e diminuir as despesas do subsistema. Para que a adopção desta medida seja viável, porém, é necessário reunir um conjunto de condições no mercado de trabalho. Em particular, a idade da reforma só deverá ser tomada como variável de ajustamento se o rendimento estiver próximo do nível de pleno emprego.

Também uma regulação macroeconómica mais favorável ao crescimento e ao emprego tem impacto positivo no equilíbrio financeiro do sistema. Em particular, o aumento da população activa empregada (resultante da redução do desemprego, do aumento da participação feminina no mercado de trabalho ou da imigração) contribui directamente para o equilíbrio financeiro dos esquemas da repartição, por via da elevação das receitas de contribuições.

Assim, o ponto de vista defendido pode ser sumariado como se segue. A reforma teve por objectivo primordial assegurar o equilíbrio financeiro do sistema público de pensões, realizado através da contenção da despesa em pensões, mas esta não era a única alternativa possível. O nível das prestações foi considerado como "variável de ajustamento" e não como "variável objectivo". Por seu turno, o equilíbrio financeiro do sistema foi tomado como "objectivo primordial" e não como "restrição". Na óptica da política social, uma metodologia consistente tem como ponto de partida um conjunto de objectivos sociais a atingir, sendo a prossecução desses objectivos condicionada pela restrição financeira. De acordo com a perspectiva normativa defendida no presente artigo, é fundamental sustentar não só o objectivo de prevenção de pobreza entre os idosos mas também o objectivo de promover uma maior paridade de níveis de vida entre os pensionistas e os seus contemporâneos, por razões de justiça social. Num contexto em que os problemas de financiamento do subsistema se tornam candentes, a realização progressiva de tais objectivos sociais exige que o ajustamento se concretize preferencialmente através das vias que não conflituam com esses objectivos, nomeadamente as que não colocam o ónus do ajustamento sobre os pensionistas. São de referir: outras políticas do lado da despesa acima mencionadas; a elevação das contribuições; ou do aumento da idade da reforma, se o contexto do mercado de trabalho permitir. É de destacar, ainda, a importância de uma regulação macroeconómica mais favorável ao crescimento e ao emprego, pelos seus efeitos positivos no equilíbrio financeiro do sistema.

5. Observações Finais

Como foi referido, a teoria económica demonstra que há diversas formas de ajustar os esquemas de pensões para fazer face à longevidade crescente. O governo português decidiu seguir uma via única: reduzir as pensões, colocando o encargo do ajustamento exclusivamente sobre os reformados. Tal opção afigura-se particularmente inapropriada, dado que, em Portugal, os pensionistas auferem um nível médio de prestações reduzido, facto que determina um nível de vida relativamente baixo e um risco de pobreza significativamente superior ao da população total.

É de destacar que a reforma recentemente introduzida em Portugal tem grande alcance, pois representa indubitavelmente um ponto de viragem na filosofia do sistema. Como se viu anteriormente, as pensões de reforma apresentam dois objectivos primordiais, a prevenção da pobreza e a salvaguarda dos níveis de vida no período da reforma. A alteração filosófica de fundo agora introduzida consiste na renúncia à prossecução do segundo objectivo. A intervenção do Estado passa a estar concentrada exclusivamente no primeiro objectivo, sendo a prevenção da pobreza explicitamente considerada uma prioridade política. O nível médio das pensões irá ser reduzido, tendo sido tomado como variável de ajustamento no âmbito de uma política que tem por objectivo principal assegurar o equilíbrio financeiro do sistema.

Ao mesmo tempo, generaliza-se a ideia de que os esquemas contributivos se devem configurar por analogia com os mecanismos seguradores privados. Por um lado, afirma-se a tendência para manter um laço estreito entre contribuições e prestações a nível individual, reflectida na

definição de uma nova regra de cálculo, que faz depender o valor da pensão das remunerações da totalidade da carreira contributiva. Por outro lado, o valor da pensão, já que o seu cálculo depende do "factor de sustentabilidade", passa a relacionar-se com a esperança média de vida no momento de cessação de actividade. A equidade individual torna-se a norma implícita de justiça.

Na lógica da reforma, a esfera da responsabilidade social pode abranger prestações de natureza não contributiva (do regime não contributivo ou prestações sujeitas a condições de recursos, como o complemento solidário para idosos). Porém, os esquemas de natureza contributiva passam a adquirir um cunho próximo de um mecanismo segurador, sendo relegados para a esfera da responsabilidade individual[32].

As medidas agora introduzidas no esquema contributivo irão traduzir-se num declínio das taxas de substituição, na sequência da introdução do "factor de sustentabilidade" e da nova fórmula de cálculo das pensões, dependentes, doravante, das remunerações da totalidade da carreira contributiva. Determinar o valor da pensão com base no valor das últimas remunerações ou das remunerações de toda a carreira contributiva tem impacto diferente na taxa de substituição. Como, em geral, as carreiras apresentam níveis remuneratórios superiores nos últimos anos de actividade, quando as prestações dependem das remunerações de toda a carreira, a taxa de substituição é inferior. Em consequência, os rendimentos dos pensionistas, tendo uma relação mais estreita com as remunerações de toda a carreira, distanciar-se-ão do rendimento corrente na sociedade, no momento da reforma.

As novas regras de indexação irão operar no mesmo sentido, moldando trajectórias de rendimento na reforma que se distanciam da evolução geral de rendimentos na sociedade. Caminha-se para uma protecção mínima de rendimento em que a própria revalorização das pensões se desliga da evolução salarial (caso das prestações de nível mínimo) e dos preços (caso das prestações superiores a 1,5 IAS).

As medidas introduzidas conduzem inevitavelmente a uma redução do nível de vida relativo dos pensionistas. Daqui resultará um problema de justiça entre contemporâneos que, na minha perspectiva, é um problema ético relevante.

32 Este movimento, corrente na Europa, representa segundo Math (2001: 40), "virar as costas à história do desenvolvimento da segurança social, garantia colectiva baseada na solidariedade pelo trabalho. Ela conduz [...] a basear os rendimentos das pessoas idosas [...] seja em prestações financiadas pelo Estado e justificadas pela solidariedade ou a necessidade, seja em rendimentos de propriedade assegurados por uma poupança prévia".

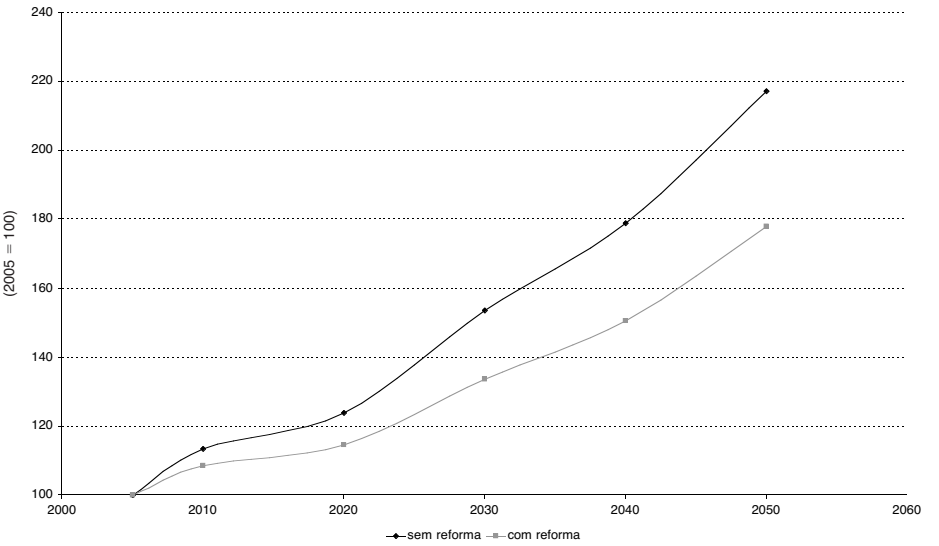
Anexo

Quadro 1 – Taxa de crescimento real do "valor mínimo de pensão" (vmp)

	Crescimento real do vmp (%)		Crescimento real do vmp (%)
1986	12,3	1997	1,6
1987	52,3	1998	1,3
1988	3,1	1999	1,8
1989	-0,3	2000	1,5
1990	2,7	2001	1,5
1991	5,6	2002	1,9
1992	4,7	2003	0,7
1993	1,7	2004	3,1
1994	0,8	2005	2,0
1995	1,2		

Fontes: Cálculos efectuados pela autora com base nos "valores mínimos de pensão" estabelecidos em Portaria e nos valores do IPC sem habitação.

Gráfico 1 - Crescimento da pensão média em termos reais



References

Atkinson, A. B. (1995) *Incomes and the Welfare State – Essays on Britain and Europe*, Cambridge, Cambridge University Press.

Atkinson, A. B. (1999) *The Economic Consequences of Rolling Back the Welfare State*, Cambridge Mass., The MIT Press.

Barr, N. (2001) *The Welfare State as Piggy Bank – Information, Risk, Uncertainty, and the Role of the State*, Oxford, Oxford University Press.

Barr, N. (2004) *Economics of the Welfare State*, 4.ª Ed., Oxford University Press, Oxford.

Barr, N. (2006) Pensions: Overview of the Issues, *Oxford Review of Economic Policy*, 22, 1, 1-14.

Barr, N.; Diamond, P. (2006) The Economics of Pensions, *Oxford Review of Economic Policy*, 22, 1, 15-39.

Barr, B. (1990) , *Ethics*, 100, 503-529; reimpresso *in* Goodin, R. and Ware, A. (eds.), *Needs and Welfare*, London, Sage, 73-103.

Blanchet, D.; Villeneuve, B. (1997) Que reste-t-il du débat répartition-capitalisation?, *Revue d'Économie Financière*, 40, 157-174.

European Commission (2005) *National Strategy Report: adequate and sustainable pension systems*, Relatório submetido em Julho de 2005 à Comissão Europeia. [http://ec.europa.eu/employment_social/social_protection/docs/2005/pt_sintese_en.pdf]

European Commission (2006) *Adequate and Sustainable Pensions: Synthesis Report,* Luxembourg, Office for Official Publications of the European Communities.

Fleurbaey, M. (2002) Retraites, Générations et Catégories Sociales: de l'équité comme contrainte à l'équité comme objectif, *Revue d'Économie Financière*, 68, 91-111.

Fleurbaey, M.; Michel, P. (1992) Quelle justice pour les retraites?, *Revue d'Économie Financière*, 23, 47-64.

Gillion, C., Turner, J., Bailey, C.; Latulippe, D. (eds.) (2000) *Social Security Pensions*, *Development and reform*, Geneva, International Labour Office.

Math, A. (2001) Quel avenir pour les retraites para répartition en Europe?, *Revue de l'IRES*, 36, 2, 3-59.

Ministério das Finanças e da Administração Pública (2006) *Orçamento de Estado para 2007*: *Relatório*, Outubro, Lisboa.

Ministério do Trabalho e da Solidariedade Social (2006) *Linhas Estratégicas da Reforma da Segurança Social*, 3 de Maio, Lisboa.

Myles, J. (2002) A New Social Contract for the Elderly? *in* Esping-Andersen, G. with Gallie, D. *et al.*, *Why we need a new Welfare State*, Oxford, Oxford University Press, 130-172.

Rein, M.; Turner, J. (1999) Le travail, la famille, l'État et le marché: des sources de revenus composites pour les ménages âgés*, Revue Internationale de Sécurité Sociale*, 52, 3, 109-125.

Thompson, L. H. (1983) The Social Security Reform Debate, *Journal of Economic Literature* 21, 4, 1425-1467.

Thompson, L. H. (1998) *Older and Wiser: The Economics of Public Pensions*, Washington, D.C., The Urban Institute Press.