UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

How to share our risks efficiently? Principles for optimal social insurance and pension provision

Teulings, C.N.

DOI
10.1007/s10645-010-9134-5

Publication date
2010

Document Version
Final published version

Published in
De Economist

Link to publication

Citation for published version (APA):
Teulings, C. N. (2010). How to share our risks efficiently? Principles for optimal social insurance and pension provision. *De Economist*, *158*(1), 1-21. https://doi.org/10.1007/s10645-010-9134-5

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:15 Oct 2026

De Economist (2010) 158:1–21
DOI 10.1007/s10645-010-9134-5
© The Author(s) 2010
This article is published with open access at Springerlink.com

DE ECONOMIST 158, NO. 1, 2010

HOW TO SHARE OUR RISKS EFFICIENTLY? PRINCIPLES FOR OPTIMAL SOCIAL INSURANCE AND PENSION PROVISION

BY

COEN N. TEULINGS*

Summary

The efficient organisation of social insurance is an important problem for modern societies. The paper discusses evidence that shocks in labour income have largely persistent effects and analyses the implications of this observation for the optimal design of institutions for wage contracting, social security, and pensions. In an optimal contract, wages reflect variations in individual productivity for incentives reasons. However, the optimal contract insures workers against firm specific shocks. These can better be born by shareholders who can diversify risks on capital markets. Progressive income taxation provides further insurance. On top of that there is scope for additional insurance based on 'verifiable' information on unemployment and health conditions. As final form of 'insurance', the paper analyzes the role of self-insurance. Income shocks can be absorbed partially by precautionary saving. The individual's saving plans for retirement and for precaution are, therefore, related issues. In an institutional setting with mandatory saving for retirement, an integration of disability and unemployment insurance on the one hand and the pension system on the other hand in a lifetime savings account allows for this interrelation. The paper analyzes how to deal with the uncertainty in the return on savings in the framework of a lifetime saving account.

Key words: social insurance, pensions, precautionary savings

JEL Code(s): H55, E21, E24

1 INTRODUCTION

In November 1942, William Beveridge published his famous report on social insurance. It provided probably the first comprehensive plan for a social insurance system. Its recommendations were based on three guiding principles. The first principle is that any proposal should benefit from all experience gathered in the past, but should not be tuned to vested interests. The second principle is that social insurance should be viewed as part only of a wider economic policy, to further social progress and to eliminate all kind of hazards. The third principle is that social security is not solely a responsibility of the

* Corresponding author: Coen N. Teulings, CPB Netherlands Bureau for Economic Policy Analysis, The Hague, The Netherlands and University of Amsterdam, Amsterdam, The Netherlands, e-mail: C.N.Teulings@cpb.nl.

state, but also of the individual. In organising social security, the state should not stifle private incentives, opportunities, and responsibilities. It should leave room and encouragement for voluntary action by each individual to provide more than social security income for himself and his family. The Beveridge report turned out to be the roadmap to the organisation of social security in the aftermath of World War II.

With benefit of the hindsight, the importance of the third principle was probably understated at that time. During the period of Eurosclerosis in the seventies and eighties of the previous century, the European social security system turned out to provide too little incentives for the economy to recover from the oil price shocks. Unemployment surged, and most countries started a reconstruction programme of the social security system to enhance incentives, in some countries with more success than in others. Though the direction of this reconstruction was clear – providing less insurance to reduce the moral hazard problem – there was no *grand design* for the type of institutions that could lead a country to the "production" frontier of insurance and incentives. It might be time to take up the first principle of the Beveridge report, and combine our experience of the past decades and the fruits of extensive scientific research in income dynamics and optimal insurance to rethink the design of the social security system.

This paper is a contribution to this project. We combine both empirical and theoretical insights, on wage formation, optimal insurance, precautionary saving, consumption planning, and portfolio choice to lay down principles for the optimal design of social security. As our starting point we choose the consumption and investment plans that are derived from the maximization of an individual's expected lifetime utility. We then ask ourselves how institutions can be helpful in the implementation of these optimal consumption and investment plans.

As our starting point we choose the empirical estimates of the process of income dynamics over the lifecycle. These estimates are remarkably similar across countries. Most of the uncertainty in labour income is caused by permanent shocks affecting the income flow for the rest of the career. This result has far reaching implications for the organization of social security, since it is much harder to insure against permanent than against transitory shocks.

A proper system of wage setting contributes to the provision of insurance in a society, by making maximum use of the ability of capital markets to diversify firm specific risks. Progressive income taxation provides further insurance. To the extent that there is verifiable information on unemployment and health conditions and that this information can be applied for offering additional unemployment and disability insurance. Beyond that, only self-insurance is helpful. An individual's optimal consumption plan allows for precautionary saving. The effect of shocks in income on consumption is smoothed be variations in the saving rate. It turns out that two statistics, ratio

of saving to current income and age, are sufficient statistics for the determination of the optimal consumption plan.

In theory, self-insurance would not require any additional institutions beyond well functioning capital markets. Individuals can lend and borrow themselves. In practice, even well trained economists are unable to do the calculations for that properly. This was probably the reason for societies to introduce pension funds in the first place. In this paper I take their existence for granted. My goal is merely to analyze how the policies of these funds can help participants to solve their lifetime consumption planning problem. An obvious idea is to extend their role to include self-insurance for unemployment and disability in their policy for setting contribution rates. Pension funds have information on the sufficient statistics to calculate optimal consumption plans.

Finally, the paper addresses the issue of how pension funds should deal optimally with the aggregate risk in a society. What should the lifetime risk profile of an individual look like? Can the aggregate risk be shared between generations beyond contemporaneous capital market transactions, by letting future generations share in today's aggregate risk? Youngsters should bear the main part of this risk, simply because their combined human and physical capital is higher than that of older generations. Current capital market institutions do not seem to allow this, because human capital is not considered appropriate as collateral.

The structure of this paper is as follows. Section 2 discusses some evidence on the stochastic dynamics of income during the lifecycle. Section 3 deals with the role of wage formation. Section 4 discusses how to optimally deal with permanent shocks to labour income. This section investigates the idea of integrating the unemployment and disability insurance with the provision of pensions for retirement. Section 5 deals with society's aggregate risk. Section 6 concludes.

2 THE EMPIRICS OF INCOME DYNAMICS

There is substantial evidence that log individual income (= wages, for the majority of the people) can be well approximated by a random walk with transitory shocks, see e.g. Abowd and Card (1989), Topel and Ward (1992), and Meghir and Pistaferri (2004) for the US, Buhai et al. (2008) for Denmark and Portugal:

$$\Delta w_t = u_t + \Delta e_t \tag{1}$$

where w_t is an individual's log wage at time t, where u_t and e_t are individual specific i.i.d. random variables with variances σ_u^2 and σ_e^2, respectively, and where Δ is the first difference operator; u_t measures the permanent and e_t

the transitory shocks. In this analysis, aggregate shocks are filtered out by including time fixed effects. Equation (1) implies that individual's wage gets two types of shocks each year, a transitory shock that dies out in a year, and a permanent shock that will affect her wage during the rest of her working career. The permanent shock of this year does not depend on the size of past shocks. From a laymen's perspective, the size of these shocks seems remarkably large. Within a job spell, both σ_u and σ_e are of the order of magnitude of 10%, implying that the combined effect of permanent and transitory shocks shifts your wage up or down on average by 17% over a period of one year, by 30% over a period 7 years, by 40% over 14 years, and 60% over 34 years. These numbers show the amount of risk workers face during their career, and hence, the large demand for insurance these risks generate. The size of these shocks is remarkably similar across countries. During the transition from one job to another, shocks are substantially larger, see Buhai and Teulings (2007) for the US. Meghir and Pistaferri (2004) show that volatility is clustered in time (ARCH). Though they do not relate these clusters of volatility to job transitions, my work with Buhai suggests that these transitions are the main source. Averaging over both job transitions and within job changes, both σ_u and σ_e are of the order of magnitude of 15–20% per year for the United States. Estimates for other countries yield a remarkably similar pattern.

The above representation of income dynamics as a combination of permanent and transitory shocks has not been remained undisputed. Guvenen (2007) reports evidence for a different pattern of shocks:

$$w_t = qt + v_t + e_t$$

$$v_t = \rho v_{t-1} + u_t \tag{2}$$

where q is a random variable, which differs between individuals but is constant over her lifetime and where ρ is a parameter; q captures heterogeneity in the return to experience between individuals, that is unobserved to the econometrician, but that may or may not be observed by the individual. When $\rho = 1$ and $\mathrm{Var}(q) = 0$, we are back to the model of Eq. (1). There is clear cut evidence for observed heterogeneity in the return to experience: it is a well known regularity that the return to experience goes up with the level of education. It turns out to be difficult to establish whether there also is unobserved heterogeneity. In other words, it is difficult to disentangle model (1) and (2) empirically, since they yield very similar patterns of dynamics. However, model (2) generates a much smaller demand for insurance, since most of the *ex post* heterogeneity in outcome is determined *ex ante*, by the return to experience u. When individuals know their return to experience, the only remaining uncertainty is the transitory shock e. The model used by Meghir and Pistaferri (2004) encompasses both model (1) and (2). They show that

TABLE 1 – PERMANENT AND TRANSITORY INCOME SHOCKS PER EDUCATIONAL GROUP

	High school drop out (%)	High school graduate (%)	College graduate (%)
Transitory shock	23	16	7
Permanent shock	18	17	21
Total shock after			
1 year	38	28	23
7 years	58	50	56
14 years	76	66	79
34 years	111	100	122

Source: calculated from Meghir and Pistaferri (2004, Table III).
The numbers in the table give the effect of on log income. The relative effect on income is then $e^x - 1$, which is approximately equal to x for small x. However, for larger x, e.g. $x = 1$ and $x = -1$, we have $e^1 - 1 = 171\%$ and $e^{-1} - 1 = -63\%$.

model (1) can explain the data well. One does not need unobserved heterogeneity in the return to education to explain the data. From the point of view of social insurance, the permanent shocks matter most. A transitory shock of 10% of your yearly income is regrettable, but fades away. A permanent shock of the same size has a much larger effect on your lifetime income. Indeed, as pointed out by Low et al. (2006), the effect of unemployment on lifetime income tends to be small compared to disability. When thinking about social security, permanent shocks to w_t are, therefore, a more important factor than a temporary spell of unemployment. The social security systems that have been build up since the publication of the Beveridge report do exactly the opposite. Hence, there is demand for a modernization of these systems.

Apart from the differences in the return to experience, Meghir and Pistaferri report another important difference between educational groups, see Table 1. For high school drop outs, transitory shocks are more important, while for college graduates, the permanent shocks matter most. Furthermore, shocks seem smallest for the middle group of high school graduates. I am unaware of similar studies for other countries than the US, but I would guess that similar patterns would emerge for other countries. A rough interpretation of these differences relates to two theoretical notions. The difference between high school drop outs and high school graduates relates to the notion of a dual labour market, where high school drop outs are confined to the secondary segment of insecure jobs with hardly any employment protection, which yields large transitory shocks in income by sudden unemployment spells. High school graduates are protected from these transitory shocks by permanent contracts. The difference between high school drop outs and

college graduates relates to notions of endogenous growth and the knowledge economy. College graduates have to specialize their human capital by investing in a particular type of innovation. Success in the one innovation provides the starting point for the next innovation, so that shocks have permanent effects on future income. Though these differences have a modest impact on the actual amount of risk, see Table 1, they suggest that college graduates need a different type of social security than high school drop outs, with a much larger focus on long run uncertainty due to the random depreciation of their human capital than on short run unemployment risks. Modernization is most needed for the college graduates.

3 THE ROLE OF WAGE FORMATION

Let us concentrate somewhat further on the permanent shocks. The permanent shock can be decomposed in three independent components:

$$u_t = i_t + \beta f_t + \delta m_t \tag{3}$$

where i_t measures shocks to the value of the worker's general human capital, f_t measures firm specific shocks affecting the well being of the firm as a whole, and m_t measures match specific shocks, or alternatively, the worker's comparative advantage at this firm; β and δ measure the share of these shocks that goes into wages. When the firm looses 10% of its market share, this translates into a negative shock f_t. When a worker has a fight with her boss or when she does not understand the specifics of the firm's production process, this translates into a negative shock m_t. Together, f_t and m_t measure the evolution of the value of the firm specific human capital. When the firm does well, but the worker does not understand the production process, the firm specific human capital is small. The same applies in the opposite case, where the worker is the right man for the job, but there is no market for the output of that job.

The question is what β and δ should be, see Buhai et al. (2008) and Bovenberg and Teulings (2008). A specific investment in the relation between a worker and her employer leads to a surplus of the continuation of the match relative to what the worker and the firm could get on the outside market after breaking up their relation. How to distribute this surplus between the worker and the firm is in principle an open question. Any value of β and δ between zero and unity is conceivable. As long as we observe the Hosios (1990) condition, that the cost of investment should be shared between the worker and the firm in the same proportion as the surplus generated by this investment, the efficient outcome is realized, provided the worker is risk neutral, which obviously she is not. Other things equal, it is, therefore, best to let the firm bear the full cost of the investment, and consequently to

attribute to it the full surplus generated of these investments, $\beta = \delta = 0$. There are two reasons for this. First, the firm's shareholders can diversify their portfolio on the capital market. Hence, from the point of view of insurance, they are much better equipped to bear firm specific risk than workers, who cannot diversify the firm specific component in wages. Second, as an implication of the first argument, the firm's cost of capital for this investment are lower, because it can charge a smaller risk premium due to the shareholders capacity to diversify the risk on the capital market. This yields a plea for a zero return to tenure (or seniority, we get to that below), and worker's wage at the job being equal to the starting wage at the best alternative job that she can get, leaving the full surplus to the firm.

Regrettably, this outcome cannot be easily implemented, for two reasons. First, it may be hard to shift the full cost of specific investment to the worker. Some of the effort spend to acquire the firm specific human capital might be hard to observe, or at least, hard to verify, so that the firm cannot directly compensate the worker for her effort. The only option then is to attribute worker's part of the match specific return, $\delta > 0$. This is the standard trade off between insurance and incentives, see Holmstrom and Milgrom (1987). As standard in this type of models, the optimal contract should provide the workers less than first best incentives for effort, due to the trade off between incentives and insurance. On top of this consideration, the worker's higher cost of capital reduces the effectiveness of these incentives, which further reduces the second best value of δ. It would be an interesting research topic as to what is the second best value of δ, but I am unaware of any study on this topic.

Note that β should be zero according to this argument, in particular in large firms: since a single worker has only a marginal effect on the performance of a large firm, making her pay contingent on the firm's performance does not contribute much to her incentives and reduces her insurance. This system is known as benchmark competition: the firm uses the average productivity of its worker to correct the individual worker's productivity for firm level effects. However, there is a caveat. Both f_t and m_t are measured relative to a market average of $f_t = m_t = 0$. Let F_t and M_t be the accumulated values of f_t and m_t since the start of the job till time t. At the moment of job start, $F_t + M_t$ must have been approximately zero, because if they were below zero, the worker would have picked another job to start working, and if they were above zero, the firm would have hired the worker before. Stated differently, competition on the labour market pushes $F_t + M_t$ to zero in new jobs. Now, as long as $F_t + M_t > 0$, there is a joint surplus for the worker and the firm for the continuation of their employment relation. However, from the point of view of the worker, continuation makes sense only as long as $\beta F_t + \delta M_t > 0$, since that is the wage she can collect in another job. This is known as the participation constraint: the worker must have incentives to

participate in the project of the firm. It is easy to see, that as long as one restrains oneself to linear contracts (Holmstrom and Milgrom 1987) provide arguments for doing that), the only option to guarantee that whenever there is a joint surplus, $F_t + M_t > 0$, then it is attractive for the worker not to quit, $\beta F_t + \delta M_t > 0$, is to set $\beta = \delta$. Hence, there is a three fold trade off between (i) the desire to safeguard workers from the firm specific risk f_t by applying benchmark competition, $\beta = 0$, (ii) the desire to provide the worker incentives for the acquisition of firm specific human capital, $\delta > 0$, and (iii) the desire to satisfy the participation constraint, $\beta = \delta$. One can expect that for an optimum, the worker should be at least to some extent an implicit shareholder of her own firm, $\beta > 0$.

Apart from the fact that the worker should have some incentives for the acquisition of firm specific human capital, there is a second reason why it is difficult to attribute the full surplus to the firm by setting $\beta = \delta = 0$. This is the inability of workers to commit *ex ante* on setting $\beta = 0$, that is, on not demanding a higher wage in case the firm's profits are high. It is hard for workers to resist this temptation when that moment comes by. Buhai et al. (2008) analyze a model where senior workers (that is: workers with higher tenure than their colleagues) have the bargaining power to capture part of the excess profits. They show empirically that $\beta > 0$ counts for both Denmark and Portugal. Other evidence summarized in Teulings and Hartog (1998) shows that wages depend more on the profitability of the firm, the more so in countries like the United States, the United Kingdom and Canada, than in Scandinavia and the Netherlands, suggesting that β is larger in the former than in the latter group of countries. It is tempting to conclude that these differences are due to the institutions for wage formation, where the latter group has some form of coordination in wage setting, while the former has not.

This discussion has revealed some complications in the trade off between incentives and insurance. The empirical evidence shows a substantial amount of heterogeneity where countries end up in this trade off. A proper "choice" of β and δ contributes to efficient social insurance. However, the word "choice" suggests an excessive amount of freedom in picking the values for β and δ. Both parameters are a reflection of the institutions for wage setting, both at the aggregate and the individual level. These institutions are not simply "chosen" by a country, they have emerged through its social economic history. The conclusion is that these institutions matter for insurance squares well with the observation that there is a high cross country correlation between the inequality in disposable income and the inequality in gross wages. Apparently, the institutions for wage setting matter a lot for the amount of insurance a country can provide.

4 RISKS, SELF-INSURANCE AND INTEGRATE SAVINGS ACCOUNTS

Though wage setting institutions can reduce the risk faced by an individual worker somewhat, the individual will face a substantial amount of income uncertainty. Hence, there is a demand for further insurance. Progressive taxation acts as a redistribution device between individuals, but also as an insurance device for the risk that individual faces in the evolution of her lifetime income. Let log disposable income d_t satisfy:

$$d_t = d_0 + \tau\, w_t \tag{4}$$

where τ is the coefficient of residual income progression (the ratio of one minus the marginal tax rate by one minus the average tax rate). In a progressive tax system, $0 < \tau < 1$, so that a one per cent increase in gross wages leads to less than a per cent increase in disposable income. Empirically, τ is around 0.85. Progressive taxation provides additional insurance, but undermines worker's incentives for investment in human capital. We take the coefficient for residual income progression as given here. One can assume that societies have experimented with a proper value of this coefficient such that it strikes a reasonable balance between their preference for insurance and redistribution on the one hand and incentives on the other hand. It is not necessary to take a stance on this issue for topic of this paper, the optimal insurance of lifetime income. Since w_t follows a random walk, so does log disposable income.

Without further information regarding the nature of shocks, insurance above what is provided by the tax system is infeasible since the insurer lacks the information to monitor the validity of a claim on income support. The Dutch experience with disability insurance during the eighties underlines the practical relevance of this conclusion. Disability was defined broadly during this period. It was a "catch all" concept, covering any cause that reduced worker's productivity below a level attained earlier on in the career. Even a declining marketability of the specific skills of the worker was counted under this heading. In that sense, disability could well be understood as a sequence of consecutive negative shocks u_t to the worker's wage. It is difficult for the insurance company to distinguish between outside shocks to a worker's productivity and lack of effort. Since the tax system already provides some insurance for the case that no further information is available, there is probably not much role for a separate disability insurance when no verifiable information on the nature of the shocks is available.

Above that, the only option that is left is to benefit from self-insurance. At first sight, this type of insurance seems to be of little value here, since the main source of uncertainty is permanent shocks. For a transitory shock, it makes sense to reduce future consumption to avoid a fall in consumption today due to a temporary low income. For a permanent shock, this does

not make sense, since the shock affects today's and future income alike. Nevertheless, self-insurance raises the individual's lifetime utility via precautionary saving. Calculations by Gourinchas and Parker (2002) suggest that the amount of precautionary saving is substantial. Suppose that an individual has zero assets and receives an income, of which the future evolution is governed by geometric random walk. The optimal consumption plan of that individual is then to save some money for tomorrow, even though this implies that expected consumption tomorrow exceeds consumption today, because the excess utility cost of low future consumption in a bad state of the world (when marginal utility of consumption is high) is larger than the utility gain of raising today's consumption (when marginal utility is of consumption is average). Hence, the optimal strategy implies an upward drift in expected consumption, see Caballero (1990).

A complete characterization of this consumption path can be derived from solving a stochastic partial differential equation. Two examples with explicit solutions of this differential equation are given in Caballero (1990) for the case with uncertainty in labour income and a Constant Absolute Risk Aversion (CARA) utility function and Teulings and De Vries (2006) in this journal for the case with uncertainty in capital income and a Constant Relative Risk Aversion (CRRA) utility function. However, in general, this equation has no closed form solution. For the case of a constant relative risk aversion utility function and evaluated at the point where the individual has not yet accumulated any savings, the optimal upward drift in consumption can be shown to be equal to:[1]

$$Drift = 1/2\,\theta\,\sigma_u{}^2, \tag{5}$$

where θ is the coefficient of relative risk aversion. Taking $\sigma_u = 17\%$, see Sect. 2, and $\theta = 2$, which is generally viewed as a reasonable value, we obtain an initial upward drift in expected consumption of 2.9% per year. As a result of this precautionary saving, the worker accumulates savings. When her career proceeds, her consumption gets increasingly paid for from the return on her savings, and her labour income becomes less relevant. Sooner or later, this makes retirement optimal, simply because the income from wealth makes working a meaningless effort.

1 The optimal consumption plan equates the expected marginal utility of consumption today and tomorrow. The marginal utility of consumption at time t is C_t^{θ}. The due to the relative risk aversion specification, consumption is proportional to the expected discounted value of life time labour income. Hence, the equality of expected marginal utilities implies:

$$1 = \exp\left(-\theta\,Drift + 1/2\,\theta^2\,\sigma_u{}^2\right)$$

Taking logs and solving for *Drift* yields the expression in the text.

The solution of the differential equation depends on current income, accumulated savings, and the age of the individual. However, given the form of the dynamic income process (a geometric Brownian, see Sect. 2) and assuming a CRRA utility function, this partial differential equation can be simplified as to depend only on the ratio of accumulated savings to current income, denoted R, see for example Teulings and De Vries (2006) for the case with uncertainty in capital income. Bodie et al. (1992) and Bovenberg et al. (2007) in this journal extend this model with transitory shocks to income, as discussed in "The Empirics of Income Dynamics". While the differential equation of the original model of Teulings and De Vries has an explicit solution, the extended model has not. However, the result that the solution depends only on the ratio of savings to current income does still apply. Then, the optimal savings rate s of an individual can be characterized as a function of the ratio of savings to income R and the individual's age t.

$$s = h\,(R, t)$$

Hence, R and t are sufficient statistics for calculating the optimal saving rate. Keeping t fixed, the higher R, the lower s, because the accumulated wealth allows a lower savings rate. Keeping R fixed, the older the worker, and the higher is s, because there is less time left to save for retirement. During the initial phase of the worker's career, people save mainly for precautionary motives. Later on in the career, people save for retirement, see Gourinchas and Parker (2002). After retirement, pensioners gradually deplete their wealth. Hence, consumption follows the income process, but with a lag, and more smoothed. On average, an individual's consumption shows on average a strong upward trend during the lifecycle, due to precaution.

When a worker faces a downward shock to his income that pushes up the value of savings to income R, the worker responds by reducing his saving rate, so that consumption tracks the fall in income only partially. In this way, the pension rights accumulated during the working life are a reflection of all shocks the worker has faced during his life. Depending on the precise way the leisure enters workers preferences, optimal retirement decisions can also be written as a function of the ratio R and the worker's age t:

$$\text{Retire if}: g\,(R, t) > 0$$

Keeping t fixed, the higher R, the higher $g(R, t)$: accumulated savings make it less attractive to continue working. Keeping R fixed, the older the worker – a higher value of t-, the higher g, since the remaining lifetime is lower, so that the accumulated wealth allows a higher level of consumption. Obviously, both function, $h(R, t)$ and $g(R, t)$ are interrelated: the greater the preference for retirement, the higher the demand for saving during the worker's career.

Bovenberg et al. (2007) provide numerical simulations that are required for the calculation of the functions $h(R,t)$ and $g(R,t)$.

This analysis has far reaching implications for the optimal institutions for social insurance, retirement and taxation. First, retirement and precautionary saving for social security reasons should be an integrated system. A downward shock in income today leads to reduced saving for retirement. This lower saving is justified because the downward shock translates into a reduction in the expected pension, while the previous accumulation of wealth was aiming at a higher pension corresponding to the income level before the downward shock. In this way, the worker smoothes the effect of the downward shock to his current income over the consumption plan for his whole remaining lifetime. Precautionary saving for self-insurance and saving for retirement can only be optimal when they are administered in an integrated saving account.

Second, savings accumulated for consumption smoothing over the lifecycle should not be subject to wealth taxation. Wealth taxation distorts, because future consumption is taxed more heavily than current consumption. Let t be the tax rate on wage income, and let r be the tax rate on wealth. Then marginal tax rate on consumption is $1-(1-t)(1-r)^T$, where T is the time lag between earning and consumption. It is easy to see that this tax rate is increasing in T. Though tax rates on wealth tend to be low, the effect can be substantial due to a high value of T. For example, for reasonable values of the marginal tax rates, e.g. $t=50\%$ and $r=1.5\%$, the marginal tax rate for current consumption is 50%, while the rate for consumption in 20 years is 63%. Optimal smoothing of consumption across the life cycle requires the marginal tax rate to be constant. This would call for a zero tax rate on wealth. From this perspective, the current tax-treatment of pension contributions and benefits is optimal: tax benefits, and not contributions, and exempt pension funds from wealth taxation. However, wealth taxation is a very effective means for containing the wealth distribution in the economy, see Pikkety (2003). Hence, one can understand why politicians prefer wealth taxation. Hence, it is useful to have an instrument to disentangle wealth held for precaution and retirement and wealth held for other reasons. Politicians want to exempt the former from taxation, but to tax the latter. Here, the function $h(R,t)$ can be helpful. It can be used to separate the wealth that one is allowed to save for consumption smoothing purposes from other wealth. This can be done in a recursive manner. At the start of the career, the person's wealth saved for precautionary saving is zero. Hence: $R=0$. The function $h(0,t)$ determines the maximum amount of savings she is allowed to add to her integrated saving account during the first year, which is exempted from taxation. To the extent that the individual realizes this saving this leads to a new value of $R>0$. Then the function $h(R,t)$ determines the savings rate for year two. And so on for later years. In this way, a country can

simultaneously exempt wealth held for precautionary motives and retirement from taxation, and tax other wealth.

All this analysis assumes that there is no objective and verifiable information about unemployment, disability or other factors that affect the evolution of labour income. To the extent that such information is available, this information can be used to provide additional benefits (or, alternatively, tax cuts) on top of the self-insurance provided by the integrated saving account.

If on top of the insurance provided by the tax system and by unemployment and disability insurance, only self-insurance makes sense, then why wouldn't we take the word "self-insurance" literally, and leave this task to the individual herself? Though this is a perfectly logical position from point of view of the standard assumptions of the neo-classical theory, my reading of the empirical evidence on hyperbolic discounting and of the recent experience with individual DC pension schemes as have been introduced in the US and the UK, is that financial planning and consumption smoothing turns out to be a far too difficult task for most of us, including myself. Moreover, we are plagued by hyperbolic discounting, postponing saving for ever since we feel that it is better to start doing so tomorrow rather than today, and we keep thinking so for ever. This asks for an institution that protects the self of tomorrow against the self of today, by imposing a mandatory saving scheme. This is exactly what pension funds do.

Whatever the reason that induced society to install pension fund, we take their existence for granted in what follows. Then, the demand for self-insurance and precautionary saving suggests that social security can best be integrated in the pension contract by means of an integrated savings account. All premiums of an individual to the pension fund are booked on this account, all capital market returns on the savings that are accumulated in this way are added to the account, and all benefits are paid from this account. An internal life insurance offered by the pension fund deals with longevity risk. The fund has information on the two sufficient statistics for the calculation of the optimal saving rate, the ratio of accumulated wealth to the labour income of the individual R, and her age t. Hence, the fund can easily calculate the individual's optimal saving rate. Pension funds can adjust the contribution rate to implement this optimal saving plan: when R goes up due to a fall in labour income, contribution rates go down to mitigate the effect of the income shock on consumption, and *vice versa*.

5 RISK TAKING, CAPITAL MARKETS, AND PENSION FUNDS

When social security is essentially an integrated savings account, we cannot avoid the question as to how these savings should be invested. Is it the best policy to invest all savings in risk free government bonds, or should some

equity[2] be included in the portfolio, and if so how much? Would labour income be risk free so that all savings are dedicated to retirement and assuming a CRRA utility function, the optimal policy would simply be to invest a fixed share of your total wealth – the sum of human and financial capital in wealth – in equity, see Teulings and De Vries (2006):

$$Share\,equity = \pi / \left(\theta \sigma^2\right), \tag{6}$$

where π is the risk premium on holding equity and where σ is the volatility of equity. The higher the risk premium, the larger the share of equity in the portfolio, the higher the volatility or the greater risk aversion, the smaller the share of equity. For reasonable parameter values, e.g. the risk premium π being equal to 4% per year and the volatility σ being equal to 20% per year, the share of equity would be about 50%. This conclusion, that workers should hold a fixed share of their total wealth in equity, has far reaching and counterintuitive implications. Value of the human capital is equal to the net discounted value of the workers expected future wage payments. Hence, this human capital is at the maximum at the beginning of the career. Gradually, this human capital is depleted. The wage the worker receives in exchange for this depreciation is partly consumed, and partly saved for retirement, thereby accumulating financial capital. However, since part of the wage income is consumed, the depreciation of human capital exceeds the accumulation of financial capital. Hence, total wealth is at the maximum at the beginning of the worker's career. Since a fixed share of this wealth should be held in equity, the investment in equity reaches a maximum at the beginning of the worker's career. During the life cycle, this wealth is gradually consumed, so that the equity position should decline gradually.

In practice, this rule is hard to implement. Workers would have to borrow large amounts at the start of their career, using their human capital as the only collateral, and invest this loan in equity. Few lenders will accept such collateral; few workers will be prepared to take so much risk at the beginning of their career. Here, pension funds can play an important role. By mandatory participation rules, pension funds have better means to extract part of the human capital of its participants as contributions. In this way, the pension fund is better able to use human capital as collateral for large loans to finance equity positions than other lenders.

The path of consumption and saving implied by this optimal investment rule is a geometric random walk. When there is a downward shock at the stock market, workers respond by raising their saving rate to make up for the loss on their equity holdings and hence by reducing their consumption. This

2 Here and beyond, we use the word equity here as acronym for the optimal risk bearing portfolio in the sense of the Modigliani-Miller model.

shift from consumption to saving is permanent. In this way the negative effect of the fall in stock prices is smoothed over the remaining life-cycle of the worker. Also from an aggregate demand perspective, this is a desirable outcome, since shocks to equity prices have only a limited effect on current consumption. To the extent that individuals are unable to make these calculations properly – e.g. due to hyperbolic discounting-, the interference of pension funds helps to stabilize the economy. Given the fact that all generations living at a particular point in time face the same stock market prices, they are all affected in the same way, and hence they adjust their consumption and saving in the same direction. However, the wealth of older generations has been affected by some early shocks that newer generations have not experienced, simply because these new generations were not alive at that moment. Hence, saving rates of different generations move parallel, but do not coincide. If older generations have faced some early negative shocks, they still have to make up for these losses, and hence they save more than younger generations who have not faced these shocks, and *vice versa* for the case the older generations have faced early positive shocks. Like in the case of uncertainty in labour income, the uncertainty in the return on equity implies that workers will invest part of the excess return on equity in extra savings for precautionary reasons. Hence, expected consumption will rise in the optimal path, see Caballero (1990).

In theory, pension funds should execute these optimal saving and investment policies on behalf of their participants, see Teulings and De Vries (2006). For younger generations, they should invest large amounts in equity, even more than the financial capital these generations have accumulated. For older generations pension funds should hold some share of the accumulated savings in bonds. For retirees, the only wealth is financial capital, so we can directly apply the rule in Eq. (6), so that about 50% of their wealth should be invested in equity. This goes against the notion applied by many pension funds that participants are exposed to investment risk till date of retirement, and then at the date of retirement the worker is supposed to sell all equity and to buy a completely risk free annuity. As a share of their financial wealth, retirees should hold less equity, but not zero.

Contribution rates of various generations at a particular point in time will be generally different, depending on investment history that a generation has experienced. Generations that started saving for the pension in bull market of the eighties face a lower contribution rate in this system than generations who that started in the beer market of the seventies. Currently, one of the cornerstones of pension funds' policy is to charge all generations the same contribution rate. From the point of view of insurance, the differences in contribution rates across generations at a particular point in time suggests that some gains from trade in investment risk remain unexploited, for if the contribution rates of various generations differ randomly (depending on the shocks that were

only experienced by the older generations), then there is a market for pooling these risks across generations. Why is this type of insurance no part of the derived optimal saving and investment policy of the pension fund? The reason is that there is a missing market for investment in equity. When younger generations could have decided on a saving and investment plan that maximizes their expected lifetime utility, they would have invested in equity even before they entered the labour market, or even before they were born, see Teulings and De Vries (2006) for an extensive discussion. In fact, the same rule in Eq. (6) applies, namely that an unborn generation should invest a fixed share of the net discounted value of its future human capital in equity. In this way, the unborn generations diversify risk optimally over time, that is, over all realizations of shocks to equity prices. This yields the optimal trade off between excess return and risk taking. Under that rule, all generations would face the same history of shocks to the stock market, and hence their contribution rates would be the same. Pension funds do actually charge all generations the same contribution rate (at least in the Netherlands). There are three interpretations of this practice. The first interpretation is that pension funds simply have no adequate policies to absorb the risk on their holdings of equity. There is some truth in this interpretation, because an optimal policy for absorbing risks would be to let the contribution rate vary with stock market fluctuations. Most pension funds prefer to keep the contribution rate constant, but there is no alternative for changing the rate. The second interpretation is more positive, namely that pension funds do actually provide this type of insurance across generations, by investing in equity on behalf of unborn generations. There is also some truth in this interpretation, in that pension funds that raise their contribution rates in response to a negative stock market shock charge higher contribution rates, also to new generations who did not participate in the fund when the shock hit the market. The third interpretation is that pension funds cannot apply this type of policies, because it would imply that after a negative shock to equity prices, new generations would enter the fund with a negative wealth. New generations could simply refuse to participate, which would lead to the collapse of the system. Obviously, there is also some truth in this interpretation.

The options for committing future generations to investment policies that were optimal *ex ante* but worked out unfavourably *ex post*, is limited. The new generation is likely to reject the inheritance of the investment done in the past on their behalf. Mandatory participation can force them to accept, but whether that works after a sequence of severe negative shocks remains to be seen. New generations will do everything they can to avoid having to bear the burden of the negative wealth, by not entering the fund. If they succeed, the incumbent generations are left with losses. Is there a policy that on the one hand benefits from this intergenerational risk sharing by investing on behalf of future generations, while on the other hand avoiding the risk that

new generations are not willing to accept the negative outcome in case of bad luck? The critical case is whether a new generation is willing to enter in the worst possible case. For a case to be the worst case, it must be true that a pension fund does not hold any equity, since by definition, holding equity implies that things can get even worse. This restriction of not holding equity immediately poses a problem for a new generation who considers whether it is worthwhile to enter, since the optimal investment rule without intergenerational risk sharing would require this generation to invest massively in equity, see our previous discussion. However, the new generation has some advantage from participating in a fund that charges all generations the same contribution, since these higher premiums will benefit them because these premiums lead to a gradual recovery of the pension fund's wealth. This recovery allows future reductions in contribution rates. Older generations also pay these higher contribution rates, but they do not share as much in the benefits because they are death by that time.

Whether or not intergenerational risk sharing – while observing the constraint that new generations must be prepared to enter the fund even in the worst possible situation – is feasible is still an open question for me. If it is feasible, it will not be easy. Therefore, it is useful to consider how a pension fund would look like that would not try to benefit from this type of intergenerational risk sharing. Then contribution rates of various generations would move parallel, but at different levels. To the extent that pension funds are going to offer integrated savings accounts that combine social security and pensions, they move away from an equal contribution rate even within a generation, since the contribution rate would depend on the history of the individual income process. Then, there is no objection at all to also allowing contribution rates to vary between generations. Hence, one would expect pension funds to move away from fixed contribution rates and intergenerational risk sharing when social security and pension get integrated.

Does the interaction of stock market and human capital risk affect the optimal holding of equity during the lifecycle? Is it still true that a worker's investment in equity should be at the maximum at the beginning of their career? As a first observation, the stock market risk and the individual specific risk in human capital are uncorrelated by construction. Hence, the rule for the share of wealth to be held in equity, see Eq. (6), still applies, see Bovenberg et al. (2007). However, the value of human capital is substantially less, since it is a risky asset. Though the risk is uncorrelated to the aggregate risk, it cannot be insured, by the incompleteness of capital markets: insuring human capital would eliminate the incentives for providing effort. Hence, one should apply some kind of risk adjusted discount rate. The risk premium drives up this discount rate, thereby reducing the net discounted value of the human capital, and hence, the amount of equity to be held at the beginning of the career. During the career, risky human capital is partly consumed, and

partly transformed in financial capital, thereby eliminating the risk discount in the human capital. This transformation might, therefore, even increase the value of the total wealth, even though part of the depreciation of the human capital is used for consumption. Hence, the amount of equity to be held might even increase during the first phase of the career. As a second observation, to the extend that aggregate risk in human capital is correlated to the stock prices, and this correlation can be expected to be substantial in the long run, human capital should be counted as a substitute for holding equity. Hence, young generations with a lot of human capital should hold less equity than discussed hitherto.

Finally, we consider two extensions to the optimal investment and consumption rules discussed till so far. First, most people buy a house at some stage in their career. This asset is a major component of the total portfolio they hold, and it is a risky asset, of which the value is correlated to equity. Likewise, the mortgage that is taken up to finance the house is also an important component in that wealth. Ideally, a pension fund should, therefore, consider the value of a worker's house when deciding how much equity to hold for that individual. The higher the value of the house, the smaller the holding of equity.

Second, the discussion till so far was based on the assumption of a CRRA utility function. This type of utility function penalizes the subject for variability in consumption; it does not penalize her for changes in consumption as such. Thus, a worker is indifferent between the consumption pattern from year 1 to 6 of *1-1-1-2-2-2* and the pattern *1-2-1-2-1-2*, since the variability in both patterns is equal, even though there are more changes in consumption in the second pattern. There is some evidence to suggest that in fact workers do prefer pattern 1 above pattern 2, due to habit formation in consumption, see Constantinides (1990). When you got used to drinking champagne, it is hard to live a day without. Changing habits is costly, and hence, workers try to stabilize consumption even more than they do in case of just standard risk aversion, by smoothing a shock over their remaining lifetime. Hence, a negative shock to either human capital or wealth is not smoothed equally over all future consumption. Smoothing goes even beyond that, in that we adjust our consumption after a e.g. 3% negative shock to our wealth by 1% the first year, an additional 1% in the second year (so, that is 2% in the second year in total), and again in the third year, and 0.5% in the fourth year and beyond (so, 3.5% in total). So, adjustment is postponed, by taking a smaller adjustment in the first year in exchange for a higher adjustment in year four and beyond. This implies that youngsters can take up more risk than elderly, because for them there are still many years left beyond year 4, while for elderly there are not that many years left. Hence, the implication of habit formation is that elderly can take less risk, since they have a shorter period left to adjust their consumption to the new level. From an aggregate demand

perspective, habit formation helps stabilizing the economy, since consumption is fairly stable in the short run.

6 CONCLUSIONS

65 Years after the publication of the Beveridge report, we might be willing to reconsider the fundamentals of our social security system. The experience of the '70 and the '80 of the previous century has taught us that incentives matter. Our income is reasonably described by a geometric random walk, where each shock has a persistent effect on your income during the rest of our life. It is tempting to interpret this phenomenon as a consequence of the knowledge economy, where the value of our human capital is not just the market wage for labour in general, but the specific knowledge that we accumulated during our lifetime, and that forms the starting point for picking up new ideas. This interpretation is supported by the fact that income looks more like a random walk for college graduates than for high school drop outs. There is an alternative interpretation of the data that states that the return to experience differs widely across individuals. Some invest early in life and would receive a steep return afterwards, others choose for a much flatter profile. This interpretation of the data generates a much smaller demand for insurance, because wide differences in income profiles are largely due to deliberate choices by individuals. Though the data provide insufficient information to rule out the possibility of heterogeneity in the return to experience, we focus on the more parsimonious model where income follows a random walk in this paper.

The persistent character of shocks asks for a rethinking of social security systems. First, firms should bear as much as possible the firm risk on the firm specific capital, since shareholders can better diversify this risk on international capital markets. Only to the extent that firms cannot adequately observe or verify the worker's specific investment, some of the cost of investment – and hence also some of its revenue – should be transferred to the worker. However, this requires that workers can commit not to bargain on future surpluses if the firm does well.

Second, the persistent nature of most income shocks makes that we should put a much greater emphasis on self-insurance and the integration of social-security and pensions. A classical pension fund had one account, to which all received contributions were added and from which all paid pension benefits were subtracted. Teulings and De Vries (2006) have raised a plea for separate accounts for each generation. This paper pushes that argument one step further, by making an argument in favour of truly individual accounts that allow a proper integration of social security and pensions. Such a pension fund can still benefit from returns to scale in asset allocation, by having two investment accounts, one with the best possible risk free portfolio, and

the other with the optimal risk bearing portfolio ("equity" in the wording of the paper), all in the spirit of the Modigliani-Miller theorem. Each individual account would then contain a proper mix of "shares" in both the risk free and in the risk bearing portfolio, where the mix depends on the characteristics of the individual, elderly participants bearing less risk than youngsters. The pension fund would set a default contribution rule that simultaneously accounts for optimal self-insurance and precaution, and for saving for a pension for retirement. Probably, the government should have a say in these rules, since they are at least partly mandatory, since pension funds are exempted from wealth taxation, and since the government has an interest in making sure that everybody will receive at least some income in the final stage of her life, to avoid that people raise claims for income support against the government. However, one can think of trade unions playing a major role in setting these rules. These issues have to be filled out on the way. However, this general line of thinking about the role of pension funds would be a major improvement in the way society deals with social security.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

REFERENCES

Abowd, J.M. and D. Card (1989), 'On The Covariance of Earnings and Hours Changes', *Econometrica*, 57, pp. 411–445.

Bodie, Z., R.C. Merton and W.F. Samuelson (1992), 'Labor Supply Flexibility and Portfolio Choice in a Life Cycle Model', *Journal of Economics Dynamics and Control*, 16, pp. 427–449.

Bovenberg, A.L., R. Koijen, T. Nijman and C.N. Teulings (2007), 'Saving and Investing Over the Life Cycle and the Role of Collective pension funds', *De Economist*, 155, pp. 347–415.

Bovenberg, A.L. and C.N. Teulings (2008), 'Rhineland Exit?', *CEPR Discussion paper*.

Buhai, S. and C.N. Teulings (2007), *Tenure Profiles and Efficient Separation in a Stochastic Productivity Model*, Tinbergen Institute.

Buhai, S., M. Portela, C.N. Teulings and A. van Vuuren (2008), *Returns to Tenure or to Seniority?,* Tinbergen Institute.

Caballero, R.J. (1990), 'Consumption Puzzles and Precautionary Savings', *Journal of Monetary Economics*, 25, pp. 113–136.

Constantinides, G.M. (1990), 'Habit Formation: A Resolution of the Equity-Premium Puzzle', *Journal of Political Economy*, 48, pp. 519–543.

Gourinchas, P. and J. Parker (2002), 'Consumption Over the Life Cycle', *Econometrica*, 70, pp. 47–89.

Guvenen, F. (2007), 'Learning Your Earning: Are Labour Income Shocks Really Very Persistent?' *American Economic Review*, 97, pp. 687–712.

Holmstrom, B. and P. Milgrom (1987), 'Aggregation and Linearity in the Provision of Intertemporal Incentives', *Econometrica*, 55, pp. 303–328.

Hosios, A.J. (1990), 'On the Efficiency of Matching and Related Models of Search and Unemployment', *Review of Economic Studies*, 57, pp. 279–298.

Low, H., C. Meghir and L. Pistaferri (2006), 'Wage Risk and Employment Risk Over the Life Cycle', *IFS Working Papers,* WO6/27.

Meghir, C. and L. Pistaferri (2004), 'Income Variance Dynamics and Heterogeneity', *Econometrica*, 72, pp. 101–132.

Pickety, T. (2003), 'Income Inequality in France, 1901–1998', *Journal of Political Economy*, 111, pp. 1004–1042.

Teulings, C.N. and J. Hartog (1998), *Corporatism or competition*, Cambridge, Cambridge University Press.

Teulings, C.N. and C.G. De Vries (2006), 'Generational Accounting, Solidarity and Pension Losses', *De Economist*, 154, pp. 63–83.

Topel, R.H. and M.P. Ward (1992), 'Job Mobility and Careers of Young Men', *Quarterly Journal of Economics*, 107(2), pp. 145–176.