PEF, **11** (1): 21–52, January, 2012.
doi:10.1017/S1474747210000284 First published online 2 November 2010

Optimal onset and exhaustion of retirement savings in a life-cycle model

MARIE-EVE LACHANCE*
Department of Finance, College of Business Administration, San Diego State University, San Diego, CA, USA
(*e-mail:* Marie.Lachance@sdsu.edu)

Abstract

This paper facilitates the exploration of optimal individual retirement savings strategies within a life-cycle framework by providing a convenient tool to implement a model suggested by Yaari (1965) with an uncertain lifetime and borrowing constraints. The solution is given both for the general case and for cases leading to closed-form equations such as power utility and Gompertz mortality. Illustrations for a wide range of parameters indicate that starting to save for retirement in the first phase of one's career is rarely optimal. Of course, this is not to say that young workers should not save for other motives – a limitation of this model is that risks besides mortality are not considered. The conclusion should also be interpreted cautiously as it is difficult to represent every possible individual circumstance and saving incentive in a single model. The intuition behind the result is that an efficient strategy allocates the burden of financing retirement first to periods with higher income (i.e. lower opportunity costs), creating the potential for an initial period without savings when income grows.

1 Introduction

With Social Security's upcoming underfunding issues and employers' reluctance to offer defined benefit pension plans, Americans will have to rely increasingly on their personal savings to provide a secure retirement. In that context, there is a growing need to understand the individual component of retirement savings. While Social Security and defined benefit pensions are typically analyzed by complex benefit formulas based on service and salary, individual savings require a different approach. They are the result of a series of economic decisions where the individual evaluates the relative costs and benefits of foregoing current consumption for future income. Life-cycle models can capture these economic tradeoffs and lay the foundation for the theoretical analysis of individual retirement savings. The objective of this paper is to fill a gap in the literature by providing an explicit formulation for the optimal retirement savings strategy in a life-cycle model that includes borrowing constraints.

The starting point for this analysis is a versatile life-cycle model proposed by Yaari (1965) featuring intertemporal consumption decisions, time-varying income,

* The author would like to thank the anonymous referees and Swaminathan Badrinath for helpful suggestions.

an uncertain lifetime, and borrowing constraints. In a retirement planning context, a flexible specification for the income process is necessary to represent realistic changes over the career and at retirement. Longevity risk must also be part of the model because it is a key determinant of savings toward retirement. Imposing borrowing constraints (also called liquidity or non-negative wealth constraints) avoids two types of distortions. First, workers are likely to wish to consume more than they earn at the beginning of their career. Second, when mortality is added to a life-cycle model, borrowing against pensions becomes more attractive because there is a low likelihood that the loan will have to be repaid.

Yaari (1965) recognizes that the borrowing constraint can be binding over a number of segments of the solution and, as a result, the optimal consumption process takes two different forms in the binding and non-binding phases. However, his solution is incomplete in the sense that it does not provide an algorithm to identify the binding segments: at any given point in time, it is not known which one of the two possible solutions for the optimal consumption process applies. Leung (1994, 2000, 2001, 2007) adjusts Yaari's results with partial success by focusing solely on the terminal wealth depletion time. Davies (1981) provides numerical illustrations showing that the solution starts in a binding period, but only offers a necessary condition for the solution, not a sufficient one. This paper is able to address these limitations by adopting a different approach that exploits the dual version of the optimization problem.[1] Accordingly, the first contribution of this paper is to provide a solution to Yaari's model for the general case, which includes an algorithm that identifies the number and timing of the binding phases. This solution does not require numerical optimization and can be directly implemented with a simple application of the bisection method. Whether this solution is based on closed-form equations depends on the model's specific inputs. In that regard, a second contribution of this paper is to show that closed-form equations can be obtained for some realistic sets of assumptions. A full solution for the case with power utility and Gompertz mortality is presented and other cases are also discussed.

The solution to Yaari's model then permits an examination of the optimal retirement savings strategy. The analysis assumes a typical hump-shape income profile, where earnings grow at the beginning of the career and ultimately drop at retirement. Note that to simplify the analysis, the following factors are not modeled explicitly: investment risk, life insurance and pensions purchases, real estate dynamics, ability to borrow against future income with credit cards, and differential taxation of retirement savings. The solution with the hump-shaped income profile generally has three segments: (1) there is an initial period where the worker does not save at all for retirement, (2) later, when his income increases sufficiently, the individual starts saving at a time τ, and (3) eventually retirement savings are exhausted by a time T. The third contribution of this paper is to derive the comparative statics for the times τ and T.[2]

[1] With dual methods, a primary optimization problem is rewritten as an equivalent dual optimization problem. While the solution to both problems is the same, the alternative dual formulation sometimes makes it easier to obtain a solution. Applications of dual methods are commonly encountered in the finance literature, e.g. see He and Pages (1993) and Karatzas and Shreve (1998) for a description.

[2] Leung (2000) previously derived similar comparative statics for T, but not for τ.

These results indicate that the following factors are associated with both an earlier onset and a later exhaustion of savings: low time discounts, longer life expectancy, greater returns on assets, higher incomes, and lower pensions. While it is possible that starting to save right away on the first day of work is optimal, a wide range of numerical illustrations is considered in this paper and $\tau = 0$ only when the rate of return is very high. In other words, younger workers should not be expected to save for retirement unless they are given substantial incentives to do so.

The parsimony of this paper's model also leads to an explicit equation for the saving rate in the interval (τ, T). With a power utility function, the saving rate is one minus a multiple of the pension-to-income ratio. The result implies that the individual saves up to a point where the marginal benefit of doing so equals its marginal cost. The formula provides an intuitive interpretation for the marginal benefit: it is computed as if the dollar saved now was accumulated with interest and consumed at time T, and then discounted for mortality and time preferences. This solution highlights two of the major obstacles standing in the way of greater retirement savings: (1) opportunity costs are high when earnings are low, which is often the case early in the career and in some socio-economic groups, and (2) the horizon for discounting is time T, a time well past retirement with consequently very high discount rates. These results suggest that financial incentives intended to promote retirement savings can meet with a limited response if their appeal is mostly based on future benefits and fail to address the fundamental issue of high opportunity costs inhibiting saving.

The results can be related to several strands of literature besides Yaari (1965), Davies (1981), and Leung (2007). For instance, it can be shown that the solution in the interval (τ,T) is identical to that obtained from a life-cycle model with an horizon of T and no borrowing constraint. The solution shows that optimal savings increase gradually with income, which is in line with the automatic increase feature recently adopted by some defined contribution plans and the Save More Tomorrow strategy proposed by Thaler and Benartzi (2004). From an empirical perspective, the model can offer an explanation for low take-up rates in matching programs requiring individual contributions (e.g. see Mitchell *et al*., 2007; Duflo *et al*., 2007). Furthermore, this paper's conclusion that retirement savings may arise only later in one's career agrees with empirical findings from the precautionary savings literature. Notably, Davies (1981), Hubbard *et al*. (1995), Carroll (1997), Gourinchas and Parker (2002), and Cocco *et al*. (2005) provide various estimates of the age of onset of retirement savings ranging from 33 to 50. With a set of baseline parameters, this paper places that age at 36, which is within the reported range. However, this paper differs in that it does not rely on a specific set of parameters, it illustrates the sensitivity of the results to the widest possible range of parameters. This analysis permits the reconciliation of the various ages reported in the literature by examining their choice of assumptions.

The rest of this paper is as follows. Section 2 presents the general solution to Yaari's model and Section 3 provides closed-form solutions for certain cases. Section 4 analyzes the optimal retirement savings strategy in the context of typical hump-shaped income profiles. It also performs various robustness tests and compares the results to those obtained with related models. Section 5 concludes by suggesting directions for future research.

2 General case

The choice of Yaari's (1965) model as a basis for our analysis merits discussion, given the common use of more complex models today. Adding risks to the model typically lead to one of three scenarios. The first approach is to choose a model with a known closed-form solution such as the one in Merton (1971). This solution is for an optimization problem with consumption and portfolio choices for the case with Hyperbolic Absolute Risk Aversion (HARA) utility, without mortality, and without borrowing constraints. Merton gives the solution to that problem both for the cases with and without labor income. Omitting the borrowing constraint is not an issue in the version of the model without income because wealth has to remain greater than zero to avoid the infinite disutility of zero consumption. When income is included in the model, however, this argument does not apply anymore because consumption can be financed out of income. As a result, illustrations based on that model tend to show periods with negative wealth. The second approach (e.g. He and Pages, 1993) includes a borrowing constraint and retains an explicit solution, however it has to assume that the model's parameters are stationary. This implies constant income growth, which is particularly problematic when studying retirement savings because a decline in income at retirement is needed to motivate these savings. The third approach (e.g. Cocco *et al.*, 2005) does not rely on stationary assumptions, but requires that the problem be solved by numerical optimization. While more flexible, this approach also has its disadvantages. Generating the solution can be time-intensive, given the long-term nature of the problem and the large number of nodes to consider for the state variables. This argument is often invoked when restrictive assumptions are used to model pension income. Another issue with the numerical optimization process is that it can obscure the solution, and conclusions are sensitive to the parameters chosen for numerical illustrations.

This paper suggests a different approach to preserve the closed-form nature of the solution: instead of sacrificing realism by using stationary parameters or allowing negative wealth, the sources of risk other than mortality are left out. When analyzing the retirement savings decision only, including portfolio choice and income risk is not as critical as when the object of the analysis is portfolio choice or precautionary savings. Once these risks are removed, many models with lifetime uncertainty and borrowing constraints boil down to a special case of Yaari's (1965) model, which is why this model was selected as a basis for this paper. The model presented in this paper is offered as an analytical complement, rather than a substitute for more elaborated numerical models.

2.1 Model and assumptions

The optimization problem consists of finding the stream of consumption that maximizes lifetime utility, subject to a budget constraint and a borrowing constraint, i.e.

$$\max_{c>0} \int_0^{\omega} f(t)u(c_t)dt, \tag{1}$$

subject to

$$dW_t = [W_t r - c_t + y_t]dt, \quad W_0 = w_0, \tag{2}$$

and

$$W_t \geqslant 0 \quad \text{for all } t, \tag{3}$$

where:

- $f(t)$ is a discount function that can be used to represent time preferences, mortality, or any other time-varying component of utility;
- $\omega \leqslant \infty$ is the last possible age in the problem;
- $u(c)$ is a continuous utility function of consumption with $u'(c)>0$, $u''(c)<0$, $\lim_{c\to 0} u'(c)=\infty$, and $\lim_{c\to\infty} u'(c)=0$;
- $r>0$ is the after-tax real risk-free rate of return;
- W_t is the wealth process and $w_0 \geqslant 0$ is the initial wealth;
- $y_t>0$ is a deterministic, finite, after-tax, real income stream (including pensions or annuities) that the individual expects to receive in the future.

The optimal consumption and wealth processes are denoted respectively by c_t^* and W_t^*; the corresponding value function is given by

$$V(t, W_t^*) = \int_t^{\omega} f(t)u(c_t^*)dt. \tag{4}$$

Retirement problems are a special case of the model in which the income process and the utility function may change upon retirement. These retirement-specific issues are discussed in Section 4. For now, the life-cycle model is kept general with only a mild technical restriction (which is relaxed in Appendix A): the inputs $u'(y_t)$, $f(t)$, and r are assumed to be such that $\lim_{t\to\omega} \lambda(t)=0$ and $\lambda(t)$ is continuous, where $\lambda(t)$ is a function defined by

$$\lambda(t) = u'(y_t)f(t)e^{rt}. \tag{5}$$

2.2 Solution

Appendix A presents a formal version of the solution along with a proof. With technical details set aside, the solution can be exposed more intuitively in this section. As noted by Yaari (1965), the solution is composed of two types of segments: (1) those where the borrowing constraint is binding ($W_t^*=0$ and $c_t^*=y_t$), and (2) those where the borrowing constraint is not binding ($W_t^*>0$ and $c_t^*>0$).[3] Yaari characterizes the optimal consumption growth when there is an interior solution, but does not delimitate the periods where the borrowing constraint is binding.

To complete Yaari's solution, let $T_1, T_2, \ldots, T_M$ denote the times when the solution alternates between the binding and non-binding periods. The relationship between

[3] While Yaari also mentions the case where $c_t^*=0$, this case is ruled out by the assumption that $\lim_{c\to 0} u'(c)=\infty$.

these times and the periods with $W_t^* > 0$ or $W_t^* = 0$ is illustrated below:

	$T_0 = 0$ – T_1	T_1 – T_2	T_2 – T_3	T_3 – T_4	…	T_M – $T_{M+1} = \omega$
Case 1	$W_t^* > 0$	$W_t^* = 0$	$W_t^* > 0$	$W_t^* = 0$	⋯	$W_t^* = 0$
Case 2	$W_t^* = 0$	$W_t^* > 0$	$W_t^* = 0$	$W_t^* > 0$	⋯	$W_t^* = 0$

At inception, the solution may start either in a non-binding period (Case 1) or in a binding period (Case 2). Afterwards, the solution can alternate a number of times between the two types of periods. As noted by Leung (1994), there exists a time T_M after which the borrowing constraint is binding forever. In this model, this stems from the assumption that $\lambda(t)$ converges to zero. A priori, M can be any number if there are no restrictions on the model's parameters. M should typically be one, two, or three with a smooth income profile, but can easily be higher when income fluctuates.

Appendix A gives an algorithm that identifies M and the times $T_1, T_2, \ldots, T_M$. As a starting point, these times must satisfy the budget constraint and consequently the equations

$$W(0, T_1) = w_0 \tag{6}$$

and

$$W(T_m, T_{m+1}) = 0 \tag{7}$$

for all non-binding phases (T_m, T_{m+1}) with $m > 0$ (see equation (11) for definition of $W(t, T)$). As noted in Davies (1981), the boundaries of a non-binding period must also satisfy $\lambda(T_m) = \lambda(T_{m+1})$. These conditions are necessary, but they are not sufficient to determine the times $T_1, T_2, \ldots, T_M$ because (1) the solution to these equations is not necessarily unique and (2) satisfying these equations does not guarantee that the borrowing constraint is satisfied everywhere. To address these issues, Appendix A suggests the following approach. First, identify the intervals D_i, $i = 1, \ldots, N$, where the function $\lambda(t)$ is decreasing. This notation allows us to introduce an inverse function $\mathscr{T}_i(t)$ such that

$$\lambda(t) = \lambda(\mathscr{T}_i(t)) \quad \text{with } t \in D_k \text{ and } \mathscr{T}_i(t) \in D_i. \tag{8}$$

When defined, this inverse function makes it possible to rewrite equation (7) as a single-variable problem

$$W(T_m, \mathscr{T}_i(T_m)) = 0. \tag{9}$$

Within each period D_i, the properties of $W(t, T)$ guarantee that there can be at most one solution t_i to (6) or (9). If there is more than one period i where a solution t_i to (6) or (9) exists, then the optimal one is the one with the highest $\lambda(t_i)$. The dual approach is used in Appendix A to show that this criterion also ensures that the borrowing constraint is satisfied everywhere.

With the times $T_1, T_2, \ldots, T_M$ defined, the solution at time t in any non-binding period (T_m, T_{m+1}) is given by

$$c_t^* = u'^{-1}\left(\frac{\lambda(T_{m+1})e^{-rt}}{f(t)}\right), \tag{10}$$

$$W_t^* = W(t, T_{m+1}) = \int_t^{T_{m+1}} e^{-r(s-t)}(c_s^* - y_s)ds, \tag{11}$$

and

$$V(t, W_t^*) = B(t, T_{m+1}) + C(T_{m+1}, T_{m+2}) + \ldots + C(T_M, \omega), \tag{12}$$

where

$$B(t, T) = \int_t^T f(s)u(c_s^*)ds \quad \text{and} \quad C(t, T) = \int_t^T f(s)u(y_s)ds. \tag{13}$$

By contrast, the solution when time t is in any binding period $[T_m, T_{m+1}]$ takes the simpler form

$$c_t^* = y_t, \quad W_t^* = 0, \tag{14}$$

and

$$V(t, W_t^*) = C(t, T_{m+1}) + B(T_{m+1}, T_{m+2}) + \ldots + C(T_M, \omega). \tag{15}$$

2.3 Relation with previous literature

The solution presented in this section can be compared to others found in the literature. First, our solution can be contrasted with the one in Davies (1981) who also uses Yaari's (1965) model as a basis. Davies (1981) provides a characterization of the solution in the non-binding periods which is equivalent to equations (9) and (10), but does not mention that this equation can have multiple solutions.[4] Thus, he gives a necessary rather than a sufficient condition for the solution. Without a $\lambda(t_i)$ test like the one suggested in Section 2.2, there is no guarantee that a solution that satisfies (9) is optimal and implies a non-negative wealth process. It is not clear how Davies (1981) addresses this issue in his numerical illustrations because he does not specify the methodology used to produce them. Mariger (1987) also notes the limitation in Davies (1981), and proposes a test for the case where the solution starts in a non-binding phase at time 0.[5]

Second, parallels can be drawn with models featuring income risk and non-negative wealth such as Deaton (1991), Carroll (1997), Gourinchas and Parker (2002),

[4] Davies (1981) also differs by finding a case where the solution ends in a non-binding period, however this can simply be attributed to the choice of an inverse logistic distribution for mortality (see discussion in Appendix A for the case where the force of mortality converges to a positive constant over an infinite horizon).

[5] The test selects the horizon T_1 such that c_0^* computed with that horizon is minimized. This is similar to the criteria of choosing the solution with the highest $\lambda(T_1)$ because c_0^* decreases in $\lambda(T_1)$. However, it differs because it requires testing all possible values $T_1 \in [0, \omega]$ and does not specify how to determine the onset of a non-binding period.

and Cocco *et al.* (2005). They analyze optimal consumption by graphing it as a function of cash-on-hand $X_t = W_t + y_t$, which is equal to wealth plus income. For the case with power utility, the policy functions display a kinked pattern with consumption tracking income closely at low levels of cash-on-hand and becoming almost linear in cash-on-hand at higher levels. Carroll (1997) describes impatient individuals as engaging in buffer-stock behavior: they save when cash-on-hand is below a target and dissave otherwise. The concept of a target level of wealth motivating saving or dissaving is also present in Deaton (1991) and Gourinchas and Parker (2002).

These results reinforce the notion that consumption behavior is explained by wealth and that consumption increases with wealth. Our solution offers a different perspective where wealth plays a more incidental role. To explain this, it must first be observed that not all wealth is created equal in the model and this affects the relationship with consumption. Specifically, there are two types of wealth: exogenous wealth (the amount w_0 endowed at time 0) and endogenous wealth, which arise when the individual sets aside a portion of his income ($W_t^* - w_0 e^{rt}$ at time t). For variations in exogenous wealth, we obtain similar findings as the ones mentioned above: consumption tracks income when wealth is low ($w_0 = 0$), it increases with wealth afterwards (when $w_0 > 0$), and at time 0 the individual saves only if his wealth is below a given target. However, this relationship does not hold when endogenous wealth is contemplated. When savings represent a reserve that the individual builds up to smooth future income declines, a greater wealth W_t^* at time t does not imply higher consumption at that time. For instance, when the individual saves for retirement, a greater level of wealth can coincide with a period of greater savings. For the case where $w_0 = 0$ and all wealth in the model is endogenous, equation (10) shows that c_t^* can be expressed without any direct reference to wealth.[6]

In the endogenous case, it is the shape of the function $\lambda(t)$ that determines the individual consumption/saving strategy. Letting $T \equiv T_{m+1}$, c_t^* can be rewritten as $u'^{-1}\,(u'(y_t)\lambda(T)/\lambda(t))$, and this equation offers a simple rule to establish whether the individual is saving or not at any time t within a non-binding period

$$\begin{aligned} c_t^* < y_t \text{ (saving)} \quad & \text{if } \lambda(t) < \lambda(T), \\ c_t^* > y_t \text{ (dissaving)} \quad & \text{if } \lambda(t) > \lambda(T). \end{aligned} \tag{16}$$

The intuition behind this result will be discussed later in Section 4.1. The solution for c_t^* shows that in a model with certainty, an individual who accumulates his optimal level of wealth W_t^* at time t does not need to consider this level of wealth directly when establishing how much to consume. The decision is forward-looking, and income profiles play a key role in the cost–benefit utility comparison. Anticipating a decline in income in the future is a motivation for saving at some point, although savings may be postponed to periods with higher incomes and lower opportunity costs. After engaging in a saving phase, the individual will eventually initiate a decumulation phase when his income declines sufficiently.

6 There might be some confusion because both c_t^* and W_t^* are a function of time and decreasing functions of $\lambda(T)$. Technically, c_t^* can be written as an increasing function in W_t^*. However, this relationship does not predict how a change in W_t^* affects c_t^*. There is a unique W_t^* in the solution and to obtain a different one, the model's assumptions must be changed. This change would also affect the relationship between c_t^* and W_t^*.

The analytic nature of our results also allows us to derive an analogue to Carroll's (1997) criteria for an impatient individual, but without requiring stationary assumptions.[7] An individual will not start saving at any point in his lifetime if $\lambda'(t) = \lambda(t)(u''(y_t)y_t'/u'(y_t) + f'(t)/f(t)) \leqslant 0$ for all t. The scenario with $\lambda'(t) > 0$ is really what triggers having any savings in our model since the solution requires that $W_t^* > 0$ at time t.[8] If $w_0 = 0$, this implies that the individual must start saving out of his income at some point before time t, although this may not be immediately at time 0. Savings start at time 0 only when there exists a period $i > 1$ with $W(0, \mathcal{T}_i(0)) \geqslant 0$. According to the condition, savings typically arise when retirement and its associated drop in income is included in a life-cycle model. Savings can also emerge when income risk is such that the income growth rate g_t is lower than a given threshold in some scenarios. Thus, both precautionary and retirement (life-cycle) savings are motivated by the same desire to smooth future income declines. Retirement can be interpreted as a limit case of precautionary savings where income declines with certainty. The difference between the two cases is then a matter of horizon. If income can decline in the next period, the individual does not have the luxury to postpone savings because the condition $W_t^* > 0$ must be satisfied immediately. This may explain why typically early career savings are observed in models with income risk, but not in those with retirement planning considerations alone.

3 Explicit solutions

The application of the solution described in the previous section is facilitated when there is a closed-form equation for $W(t, T)$. The challenge when trying to obtain this type of solution is to avoid sacrificing realism. This section shows that it is possible to achieve this objective with a set of assumptions which is flexible enough to reflect many credible scenarios.

3.1 Assumptions

While other functions will be discussed in Appendix B, this section starts with the commonly used power utility function

$$u(c_t) = \frac{c_t^{1-\gamma}}{1-\gamma}, \quad \gamma \neq 1, \tag{17}$$

where $\gamma > 0$ is the coefficient of relative risk aversion. This utility of consumption is multiplied by the usual exponential time-discount function $e^{-\beta t}$ with $\beta > 0$. The uncertain time of death is represented by a Gompertz distribution. This is a two-parameter mortality distribution, which is frequently used because it has the double advantage of fitting actual mortality rates fairly well and leading to closed-form

[7] Carroll (1997) defines an impatient individual as one who would not save if income was certain. In an infinite horizon model, the condition is $-g\gamma + r - \beta + \sigma^2(\gamma+1)/2 < 0$, where g is the constant income growth rate, γ is the coefficient of risk aversion in the power utility function, β is the time discount rate, and σ^2 is the variance in the change in permanent income. The criteria $\lambda'(t) \leqslant 0$ is similar, except that $\sigma^2 = 0$ and $g = g_t$.

[8] In Appendix A, one of the necessary conditions is that $\lambda'(t) \leqslant 0$ for all $t \in [T_m, T_{m+1}]$ within a binding period.

solutions for many problems. This distribution posits that the propensity to die increases exponentially with age and its force of mortality is given by $\delta_t = e^{(t-m)/b}/b$. Lifetime uncertainty is incorporated in the model by using the standard equivalent approach of multiplying the utility at time t by the probability $p_{0,t}$ of surviving from time 0 to time t. In other words, $f(t) = e^{-\beta t} p_{0,t}$ in the objective function in equation (1). For the Gompertz distribution, the probability of surviving from time t to time T is given by

$$p_{t,T} = \exp(e^{-m/b}(e^{t/b} - e^{T/b})). \tag{18}$$

To obtain a flexible specification for the income process, the interval $[0, \omega]$ is first divided into J subintervals $[t_j, t_{j+1}]$ with $j = 1, \ldots, J$. The function $J(t)$ is then introduced to identify the period j corresponding to a time t. The level of income at the beginning of each interval is denoted by y_{t_j}. At any of these times, the income process is allowed to jump downwards, which permits a representation of the typical drop in income at retirement. Over a period $[t_j, t_{j+1})$, the income process is assumed to grow exponentially at a rate g_j and is given by

$$y_t = y_{t_j} \exp(g_j(t - t_j)) \quad \text{for} \quad t \in [t_j, t_{j+1}). \tag{19}$$

Since the rate of income growth can be varied over the J periods and these periods can be made arbitrarily small, this model can be used to replicate most income patterns. In particular, the level of pension income can be modeled as any function of past earnings or any arbitrary amount.

3.2 Solution

By substituting the assumptions given above in the equations in Section 2, the following closed-form expressions are obtained for the wealth process $W(t, T)$ and the components $B(t, T)$ and $C(t, T)$ of the value function

$$W(t, T) = \lambda(T)^{-1/\gamma} A_{t,T} - Y_{t,T}, \tag{20}$$

$$B(t, T) = \frac{\lambda(T)^{1-1/\gamma}}{1-\gamma} e^{-rt} A_{t,T}, \tag{21}$$

$$C(t, T) = \frac{be^q}{1-\gamma} \sum_{j=J(t)}^{J(T)} \frac{\Gamma(\delta_j, e^{(\max(t, t_j) - m)/b}) - \Gamma(\delta_j, e^{(\min(T, t_{j+1}) - m)/b})}{(y_{t_j} e^{-g_j t_j})^{\gamma - 1} q^{\delta_j}}, \tag{22}$$

where

$$\lambda(t) = y_t^{-\gamma} \exp((r - \beta)t + e^{-m/b}(1 - e^{t/b})), \tag{23}$$

$$Y_{t,T} = \sum_{j=J(t)}^{J(T)} y_{t_j} \frac{e^{(g_j - r)\min(T, t_{j+1})} - e^{(g_j - r)\max(t, t_j)}}{e^{g_j t_j - rt}(g_j - r)}, \tag{24}$$

$$A_{t,T} = e^{rt+h} b h^{-\alpha} [\Gamma(\alpha, he^{t/b}) - \Gamma(\alpha, he^{T/b})], \tag{25}$$

$$\Gamma(a, x) = \int_x^\infty t^{a-1} e^{-t} dt, \tag{26}$$

$$\alpha = b\left[\frac{r-\beta}{\gamma} - r\right], \quad h = \frac{e^{-m/b}}{\gamma}, \quad \delta_j = b[g_j(1-\gamma)-\beta], \quad q = e^{-m/b}. \tag{27}$$

A few observations about these formulas are instructive. In the budget constraint $W(t,T)$ in equation (20), $Y_{t,T}$ can be interpreted as the present value of human capital in the interval (t,T). The product $\lambda(T)^{-1/\gamma}\ A_{t,T}$ represents the present value of the optimal consumption stream. The solution is thus equivalent to the one that would be obtained for a model with an horizon of T and no borrowing constraint. The expression $\Gamma(a,x)$ in (26) is the incomplete Gamma function and it commonly arises in closed-form solutions when mortality follows a Gompertz distribution. It can usually be computed with built-in functions in computer software.[9] For example, Milevsky (2006) shows how to evaluate this function with Excel's functions GAMMADIST and GAMMALN.

The solution is then completed by determining the times $T_1, \ldots, T_M$ with the formulas listed above and the four-step algorithm presented in Appendix A. Several cases for the specification of these times are possible, the shape of the function $\lambda(t)$ determines which one applies. The solution in the appendix encompasses all possible scenarios, and the special case $M=2$ is described in the next section. Within any of the resulting non-binding phases (T_m, T_{m+1}), letting $T \equiv T_{m+1}$ the optimal consumption process is given by

$$c_t^* = \left(\frac{\lambda(T)e^{-rt}}{e^{-\beta t}p_{0,t}}\right)^{-1/\gamma} = y_T\left(e^{(r-\beta)(T-t)}p_{t,T}\right)^{-1/\gamma}. \tag{28}$$

Closed-form results are not limited to the power utility/Gompertz mortality case, Appendix B describes the solution for other cases.

4 Optimal retirement savings strategy: hump-shaped income profile

With the solution to Yaari's model established, it is now possible to address the question that is the object of this paper: What is the optimal retirement saving strategy in a life-cycle model with borrowing constraints? To introduce the retirement element in the general model, it is assumed that the individual has a hump-shaped income profile: earnings increase at the beginning of his career, they grow more slowly or decline at older ages, and they eventually drop at retirement. This pattern generally appears in aggregate data and permits the interpretation of all savings in the solution as retirement savings. Since most individuals do not have retirement savings when they start working, $w_0 = 0$.

When borrowing constraints are added to a life-cycle model, formulating the optimal saving strategy is a two-step process. First, the periods where the borrowing constraint is binding must be determined. Second, the saving strategy within the non-binding periods must be specified. With a hump-shaped income profile, the global

[9] If the incomplete Gamma function cannot be computed with built-in functions, it can be evaluated numerically with a series development (e.g., see Press *et al.*, 2007).

structure of the solution is fairly simple to describe.[10] There is a time τ where the individual starts to save. A corner solution $\tau=0$ is possible, i.e. it can be optimal to start saving at the beginning of the career. This case happens only if there exists a period $i>1$ with $W(0,\mathscr{T}_i(0))\geqslant 0$. After time τ, the individual initially builds up his savings and then draws them down after retirement until they are exhausted by a time T. In the two binding periods $[0,\tau]$ and $[T,\omega]$, the individual consumes his entire income. In the non-binding period (τ, T), the individual saves a fraction $s_t^*=(y_t-c_t^*)/y_t$ of his income for retirement. With that framework, a retirement saving strategy can be characterized by three key components: τ, T, and s_t^* in-between. The next subsections analyze each of these elements from a theoretical perspective and the section concludes with numerical illustrations.

4.1 Optimal onset τ

The optimal time of onset τ is jointly determined with T and the pair is obtained by solving the following system of two equations with two unknowns

$$W(\tau, T)=0 \quad \text{and} \quad \lambda(\tau)=\lambda(T). \tag{29}$$

The first condition $W(\tau, T)=0$ is simply the budget constraint. The second condition $\lambda(\tau)=\lambda(T)$ corresponds to the breakeven point for saving in equation (16). This can be interpreted as 'the individual initiates a savings phase when the marginal benefit (MB) of saving starts exceeding its marginal cost (MC)' by rewriting $\lambda(\tau)=\lambda(T)$ as

$$\underbrace{u'(y_\tau)}_{\text{MC of saving}} = \underbrace{u'(y_T)e^{(r-\beta)(T-\tau)}p_{\tau,T}}_{\text{MB of saving}}. \tag{30}$$

The result in equation (30) provides some insight regarding why some individuals may be reluctant to start saving for retirement: they either face high opportunity costs for saving or they discount future benefits too heavily. For instance, low incomes are an obstacle to saving because they imply high opportunity costs. Those anticipating high income growth are better off postponing savings to periods with higher income and relatively lower opportunity costs. Myopic individuals and those with high mortality prospects may postpone saving because they severely discount the future associated benefits. Favorable market conditions and preferential tax treatment should increase returns and make it relatively more attractive to start saving early. Poor economic conditions would have the opposite effect, especially when combined with a lowering of incomes. Equation (30) also shows that those with high pensions (i.e. y_T) have limited incentives to start saving, while those with little or no pensions would be more eager to save. Unfortunately, those who have both low incomes and low pensions are caught between a rock and a hard place: while saving now has a high opportunity cost, failing to do so can have negative consequences in the future.

[10] With a hump-shaped income profile, the problem is considerably simplified because the function $\lambda(t)$ has generally only two decreasing periods, i.e. $N=2$ in that case. With $N=2$ and $w_0=0$, the only values that M can take are $M=1$ and $M=2$. With respect to the notation in Sections 2 and 3, $\tau=0$ and $T=T_1$ when $M=1$ and $\tau=T_1$ and $T=T_2$ when $M=2$. If $N>2$, mild technical conditions can be imposed on the model's parameters to rule out the case $M>2$.

With a tractable solution, these intuitive results can be formalized by deriving comparative statics for τ. The derivatives are obtained by applying standard techniques to the equations in (29).[11] Based on these formulas, it can be shown that the comparative statics have the following signs

$$\frac{d\tau}{d\beta} > 0, \quad \frac{d\tau}{d\delta_t} > 0, \quad \frac{d\tau}{dr} < 0, \quad \frac{d\tau}{dy_0} < 0, \quad \frac{d\tau}{dy_T} > 0. \tag{31}$$

The results in (31) confirm the earlier predictions that workers start saving earlier for retirement if they are not myopic, they live longer, they enjoy greater returns on their investments, and they have higher incomes or lower pensions.

4.2 Optimal exhaustion T

In this model, a time T where savings are permanently depleted can always be found. Beyond some point, the odds of being alive become so slim that it is not worth saving for that contingency if one receives any form of pension. In a manner similar to that used for τ, the sign of several comparative statics can be established for T

$$\frac{dT}{d\beta} < 0, \quad \frac{dT}{d\delta_t} < 0, \quad \frac{dT}{dr} > 0, \quad \frac{dT}{dy_0} > 0, \quad \frac{dT}{dy_T} < 0. \tag{32}$$

Aside from small methodological differences, these comparative statics for T are similar to those obtained by Leung (2000).[12] Retirees will exhaust their savings later if they are not myopic, they expect to live longer, they have higher returns on their investments, they had higher incomes during their careers or they have lower pensions after retirement. Note that all the comparative statics here have opposite signs when compared to those obtained for τ. This is in part because, *ceteris paribus*, those who start saving earlier accumulate greater retirement savings and are able to spread these funds over a longer period of time after retirement. The results in (32) also reflect the fact that the speed at which individuals spend down their retirement savings depends on their preferences and situation. Those who are more myopic or have lower life expectancy will exhaust their savings more quickly. Those with higher pension income can also afford to run down their assets more rapidly because they will be in a relatively better situation when they have to live off their pension income after time T. By contrast, a higher rate of return would make retirement savings last longer and increase T for all.

4.3 Optimal level of savings

The last component of the retirement saving strategy is the saving rate s_t^* between times τ and T, which can be expressed explicitly as a percentage of income with the

[11] In the calculations, it is assumed that: the level of pensions y_T is held constant when computing $d\tau/dy_0$, changing pensions does not affect pre-retirement income when deriving $d\tau/dy_T$, and the result $d\tau/d\delta_t$ for the force of mortality applies to $t \in (\tau, T)$. If pensions change proportionally with income, then increasing y_0 does not affect τ. For $d\tau/dr < 0$, the result generally holds unless γ is very high. For $dT/dr > 0$, the result can be demonstrated for $N = 2$.

[12] Leung's (2000) results are based on a solution without an initial binding period. For mortality, he introduces a shift parameter ϕ that applies to all hazard rates and obtains $dT/d\phi < 0$. Similarly, for the income function he uses a single shift parameter ξ and derives $dT/d\xi \leqslant 0$.

following equation

$$s_t{}^* = \frac{y_t - c_t{}^*}{y_t} = 1 - \left(\frac{\lambda(t)}{\lambda(T)}\right)^{1/\gamma} = 1 - \frac{y_T}{y_t}\left(e^{(r-\beta)(T-t)}p_{t,T}\right)^{-1/\gamma}. \tag{33}$$

As explained in Section 2.3, for the case $w_0 = 0$ this solution is not a direct function of wealth, it is determined by the shape of the function $\lambda(t)$. This function $\lambda(t)$ has $u'(y_t)$ as a component and, thus, the income profile affects saving behavior. In that regard, equation (33) shows a new result: the saving rate s_t^* is linear and decreasing in the pension-to-income ratio y_T/y_t.[13] Thus, holding $(e^{(r-\beta)(T-t)}p_{t,T})^{-1/\gamma}$ constant, the saving rate should increase and decrease with income over the individual's working years. To examine the effect of mortality, consider the case $r = \beta$ where s_t^* is given by $1 - y_T/y_t \cdot p_{t,T}^{-1/\gamma}$. If $p_{t,T}^{-1/\gamma}$ was equal to one, then a less than 100% pension-to-income ratio would motivate saving automatically. However, when the individual considers that he might not be alive at time T when the savings are consumed, this has the same impact as increasing the pension-to-income ratio by a factor of $p_{t,T}^{-1/\gamma}$ For example, if $y_T/y_t = 70\%$, $p_{t,T} = 30\%$, and $\gamma = 3$, the effect of mortality is equivalent to using an effective pension-to-income ratio of $70\% \cdot (30\%^{-1/3}) = 105\%$, in which case the individual is not interested in saving. In that example, a low value for $p_{t,T}$ was chosen because, as will be illustrated later, T tends to happen much after retirement. Since $p_{t,T}$ increases with t, the dissuading effect of mortality is more pronounced earlier in the life-cycle. Last, the difference $r - \beta$ also affects s_t^*, a higher return increases savings, and a higher time discount lowers them. Small differences can have a significant impact when they are compounded over a long period of time $T - t$.

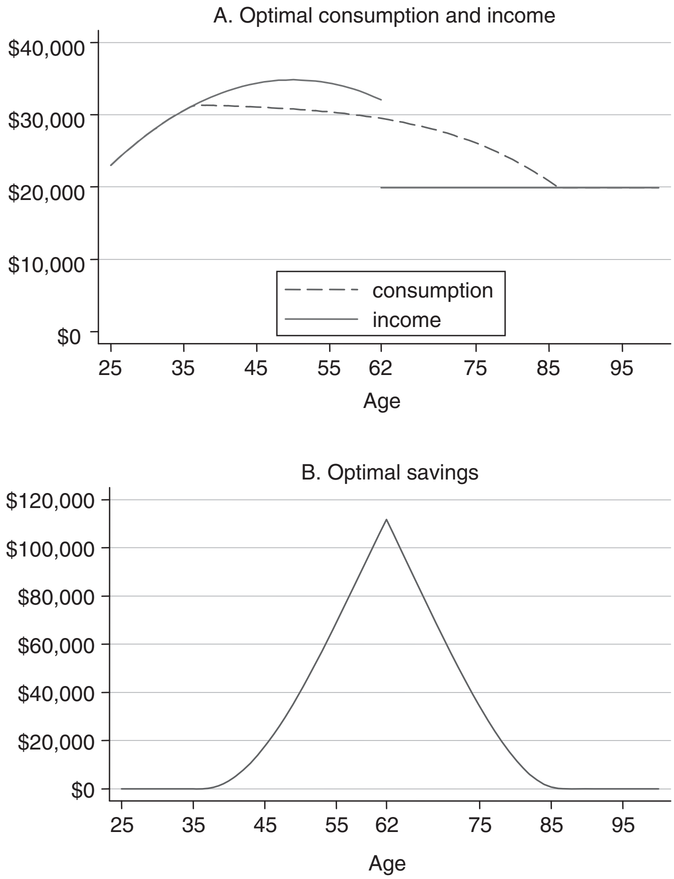
4.4 Numerical illustrations

The concepts discussed above can be illustrated more concretely with a numerical example based on the formulas developed in Section 3.2. The results are presented for a set of baseline assumptions before considering a wide range of sensitivity tests. For the baseline scenario, the problem starts at age 25 with initial wealth $w_0 = 0$. The real after-tax rate of return is set to $r = 3\%$, the time-discount factor to $\beta = 3\%$, and the coefficient of risk aversion to $\gamma = 3$. The survival probabilities are modeled according to a Gompertz distribution with parameters $m = 57$ and $b = 13$. These parameters are chosen to fit unisex mortality data from the National Center for Health Statistics (NCHS) and correspond to a life expectancy of 75 years.[14] The income process is based on data from the 2006 Current Population Survey grouped by age, with an adjustment for taxes.[15] The individual is assumed to retire at age 62, which is the most common retirement age in the United States. Pensions after retirement are assumed

[13] The linear relationship with the pension-to-income ratio can also be found with other utility functions such as members of the HARA family. The result for a general utility function is $s_t{}^* = 1 - u'^{-1}(u'(y_T)e^{(r-\beta)(T-t)}p_{t,T})/y_t$. Thus, pension income decreases savings in any case, although the relationship might not be linear.

[14] The mortality rates are based on 2005 data for men and women. These data are available online at http://www.cdc.gov/nchs/data/dvs/mortfinal2005_worktable_310.pdf.

[15] The median earnings (wages and salaries) from series PINC-08 of the Current Population Survey are \$30,174 for ages 25–34, \$36,758 for ages 35–44, \$39,126 for ages 45–54, and \$36,768 for ages 55–64. A quadratic equation is fitted to these values to interpolate income values for ages 25 to 62. Taxes are



Results computed with baseline assumptions given in Section 4.4.

Values at selected ages (in dollars)

Age	**36.4**	**45**	**55**	**62**	**75**	**86.3**
Income	31,343	34,373	34,356	19,908	19,908	19,908
Consumption	31,343	31,073	30,384	29,492	26,079	19,908
Retirement savings	0	17,517	69,014	111,780	34,208	0

Figure 1. Optimal strategy with baseline assumptions

to be fixed and represent 60 % of pre-retirement income, which is in line with the average income replacement rate observed at age 62 in the Health and Retirement Study.

The solution to the optimization problem for the baseline scenario appears in Figure 1: Panel A plots the optimal consumption process along with the income assumption and Panel B displays the optimal savings. The solution in Figure 1 is composed of three phases. First, the borrowing constraint is initially binding because income is relatively low and the individual does not start saving for retirement until age 36.4. The second phase in the solution is an accumulation/decumulation period

then deducted from these earnings according to the tax tables in the 2006 Internal Revenue Service 1040 form.

starting at age 36.4 and ending at age 86.3. At the beginning of this phase, the saving rate increases gradually as income grows and it eventually attains a rate close to 12%. At retirement, accumulated savings reach $111,780. After retirement, a spend-down phase is initiated until savings are exhausted by age 86.3. In the third and final phase of the solution after age 86.3, the borrowing constraint is binding and the individual consumes his pension income for the rest of his life. It should be noted that not everybody lives long enough to see this last phase where retirement savings are exhausted. In Figure 1, the probability of reaching age 86.3 is only 25%.

4.5 Robustness of results

This section examines the effect of alternative parameters and model assumptions on the solution. First, the impact of changing the parameters is illustrated graphically in Figure 2 and numerically in Table 1, which present the optimal ages τ and T as a function of the following factors: β, life expectancy, γ, r, y_0, and the income replacement rate at retirement. Figure 2 confirms that changing these parameters has the effect predicted by the comparative statics in (31) and (32).[16] The figure also displays the wide possible range of values that τ and T can take. The ages range from 25 to 62 years old for τ and from 62 to over 100 years old for T. For τ, the ages are generally less than 45 years old unless the individual is extremely myopic, he expects to die very early, or he is not risk averse at all. In most cases, it is not optimal to start to save for retirement right away at age 25. Since the only exception occurs when the rate of return exceeds 9%, this suggests that it takes strong artificial incentives to induce optimal savings at the beginning of one's career.

The same type of sensitivity exercise is performed in Figure 3 and in the last column of Table 1, which display the amount of retirement savings at age 62. The results show that retirement savings can range anywhere between almost zero and several hundred thousand dollars, depending on the parameters chosen. Except for the income panel, those who start saving earlier in Figure 2 also accumulate higher retirement savings in Figure 3. Note that this graph can be a useful tool to identify parameters that lead to extreme results in terms of retirement savings.

Next, the robustness of the results is tested by considering an alternative utility function. The dotted lines in Figure 2 give the results using an exponential utility function with coefficient $a = 0.0001$.[17] The results are quite similar, the onset of savings is typically within one year of the power utility results. The age of exhaustion of savings is about two years earlier versus power utility, with larger differences in Panel F for low replacement rates. This later result is explained by the steeper decline in utility at low levels of consumption for the power utility function. Overall though, the choice of the utility function is not critical for our conclusions as long as the level of risk aversion is similar to the one for the baseline case.

[16] The comparative statics for $d\tau/d\gamma$ and $dT/d\gamma$ were not included in these lists because these expressions cannot be signed for all parameters.

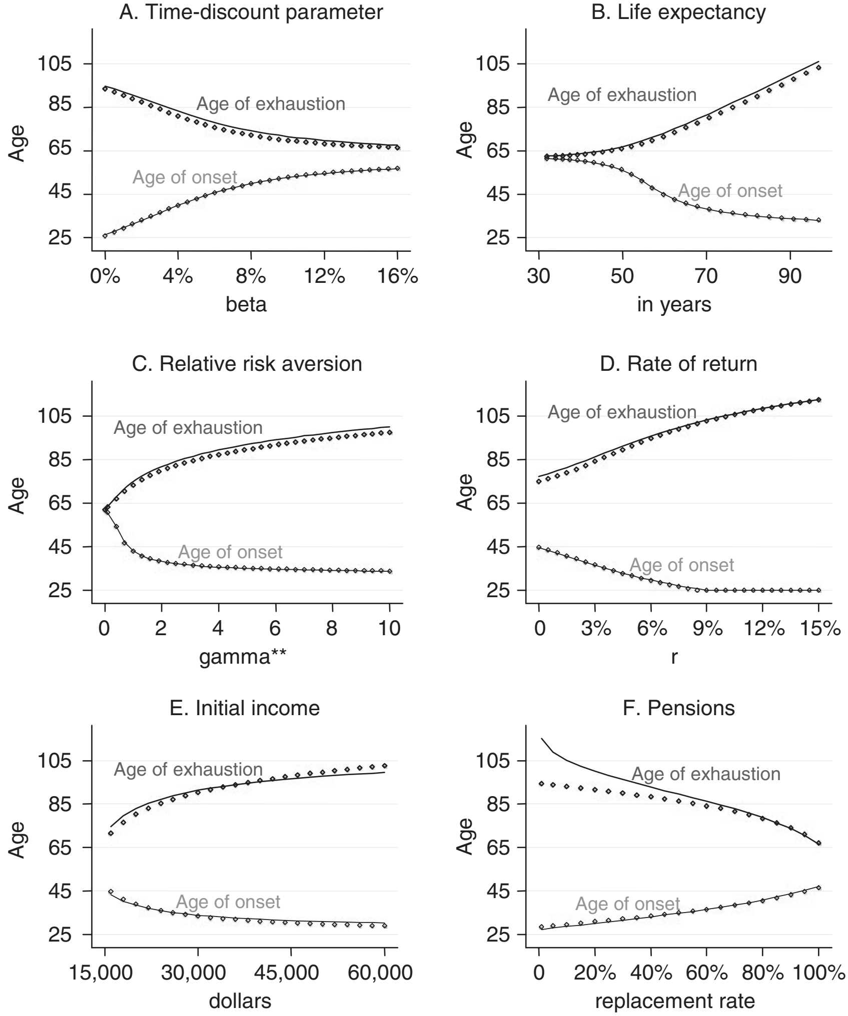
[17] The coefficient of relative risk aversion is $R^R = \gamma$ for the power utility function and $R^R = ac$ for the exponential function. To make results comparable, the parameter a is chosen to equate these coefficients, i.e. such that $\gamma = ac$. Given that consumption is roughly around $30,000 in the solution and that $\gamma = 3$, $a = 3/30{,}000 = 0.0001$.

Table 1. *Sensitivity analysis*

	Age of onset of retirement savings (τ)	Age of exhaustion of retirement savings (T)	Retirement savings at age 62 (in dollars)
A. Time-discount parameter			
$\beta = 0\%$	26.5	94.7	228,202
$\beta = 4\%$	39.8	83.4	88,211
$\beta = 8\%$	49.9	74.2	40,134
$\beta = 12\%$	54.6	69.8	23,957
B. Life expectancy			
50 years	55.9	67.0	16,665
70 years	38.2	81.7	89,791
90 years	33.7	99.8	162,246
C. Relative risk aversion			
$\gamma = 0.1$	60.6	63.5	4,540
$\gamma = 1$	42.9	74.9	55,200
$\gamma = 3$	36.4	86.3	111,780
$\gamma = 5$	35.0	92.1	135,743
$\gamma = 10$	33.7	100.2	163,495
D. Rate of return			
$r = 0\%$	44.4	77.3	58,828
$r = 4\%$	33.9	89.7	138,123
$r = 8\%$	26.3	101.1	269,254
$r = 12\%$	25.0	108.6	421,212
E. Initial income			
$y_0 = \$20,000$	38.4	82.8	66,482
$y_0 = \$40,000$	31.9	95.3	402,014
$y_0 = \$60,000$	30.3	99.6	765,640
F. Pensions			
Replacement rate: 5%	27.9	109.0	399,517
Replacement rate: 25%	30.8	98.1	272,277
Replacement rate: 50%	34.7	89.6	151,634
Replacement rate: 75%	39.6	80.8	59,640
Replacement rate: 100%	47.1	66.4	2,116

Note: Results computed with the baseline assumptions given in Section 4.4., except for parameter change in panel. Pensions are kept constant in Panel E and pre-retirement income is kept constant in Panel F.

By contrast, the income growth rate has more influence on the results. This variable was not included earlier in the comparative statics because it is not constant over the career. Intuitively, higher income growth rates should translate into a later onset of retirement savings. If one wants to finance a given amount of retirement savings, it is relatively cheaper (in terms of utility) to do so later if it is assumed that incomes grow over time. This effect should be more pronounced for those with steeper income growth because they face a greater price differential for saving early versus late. To illustrate this, Figure 4 shows the optimal consumption profiles for two income

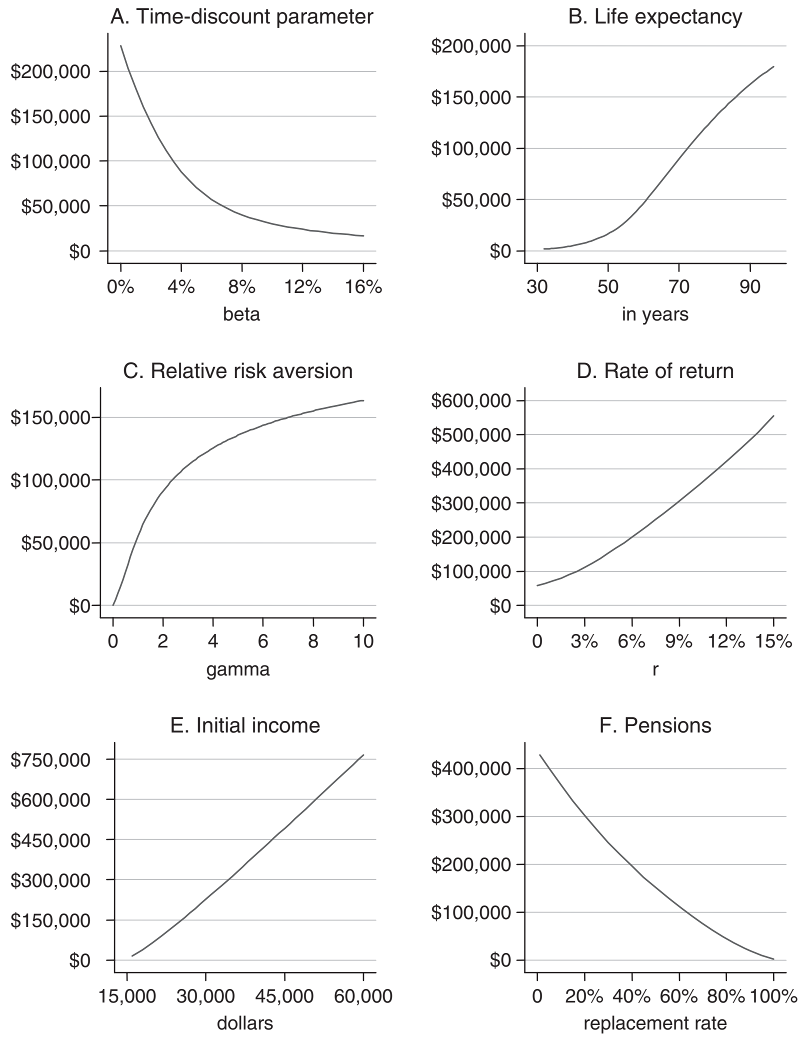


* Solid lines are for power utility and dotted lines are for exponential utility. Results computed with the baseline assumptions given in Section 4.4., except for parameter change in panel. Pensions are kept constant in Panel E and pre-retirement income is kept constant in Panel F. In Panel F, savings arise in the 100% replacement rate case because earnings decline before retirement.
** For exponential utility, horizontal axis is coefficient of absolute risk aversion multiplied by 30,000.

Figure 2. Sensitivity analysis: power vs. exponential utility*
(Optimal ages of onset and exhaustion of retirement savings)

patterns (one flat, one steep) that are obtained by multiplying the baseline growth rates by a factor. The income growth rates are multiplied by 0.2 (2) for the flat (steep) scenario. In Figure 4, pension income is set at 60 % of average career income for each case. The figure shows that savings start earlier at age 27 for the flat scenario and later at age 39 for the steep scenario. The results for the flat case should be viewed with caution because this type of income pattern is typically seen in groups with lower education and income (e.g. see Campbell and Viceira, 2002). Since those with very low incomes have high replacement rates from Social Security, the pension effect

https://doi.org/10.1017/S1474747221000284 Published online by Cambridge University Press

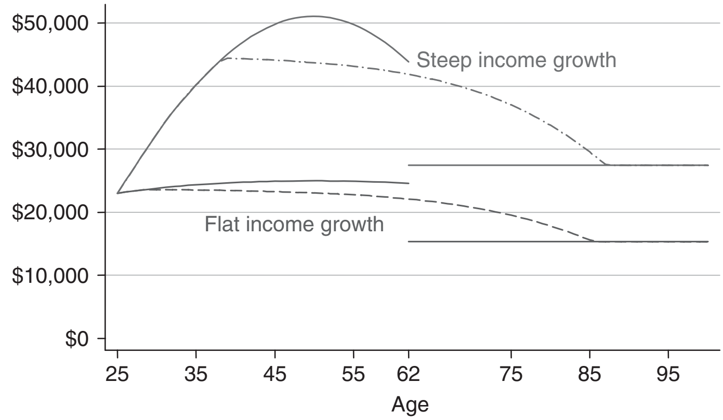


Results computed with the baseline assumptions given in Section 4.4., except for parameter change in panel. Pensions are kept constant in Panel E and pre-retirement income is kept constant in Panel F.

Figure 3. Sensitivity analysis with power utility
(Retirement savings at age 62)

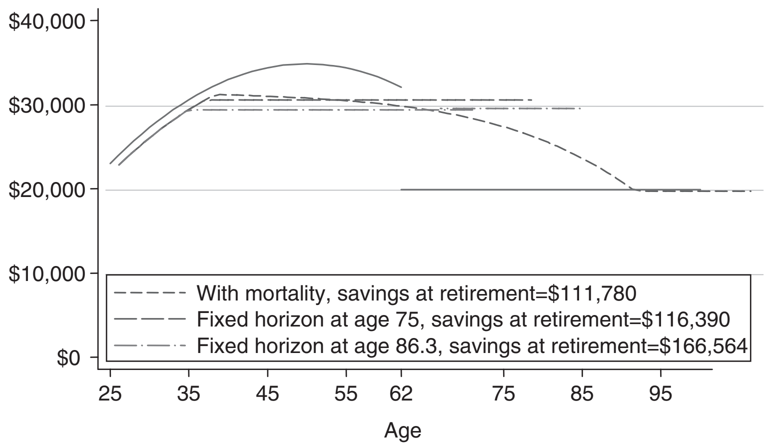
(as shown in Figure 2F) can offset part of the effect from the reduction in income growth, and savings may not start so early in practice.

The choice of the mortality distribution is another assumption that can be varied. To test whether the outcomes are affected by the use of the Gompertz distribution in place of actual mortality rates, the results were derived again using the formulas in Appendix B and a discrete mortality table based on the NCHS death rates truncated at age 100. The solutions are extremely close: $\tau = 36.5$ and $T = 86.3$ for the NCHS mortality table versus $\tau = 36.4$ and $T = 86.4$ for the Gompertz distribution. A more extreme robustness test is to consider the case without mortality by assuming a fixed horizon at age $\bar{T}$. The comparison between the cases with and without mortality



Solid lines represent income and dashed lines consumption. Results computed with the baseline assumptions given in Section 4.4, except that income growth rates are multiplied by 0.2 in the flat case and by 2 in the steep case. Pensions are 60% of average career income.

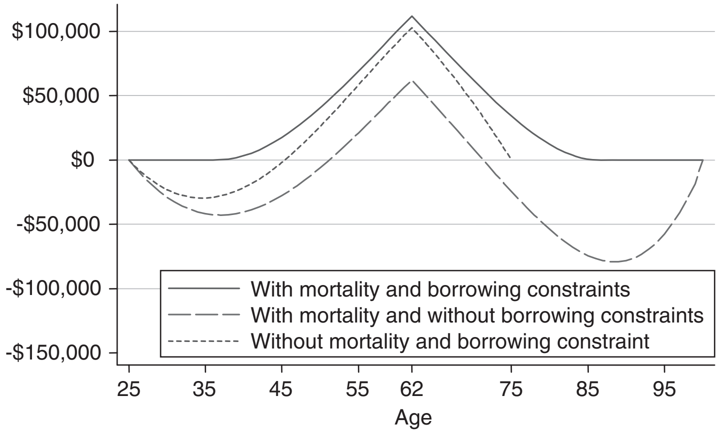
Figure 4. Optimal consumption and income growth



Solid lines represent income and dashed lines consumption. Results computed with the baseline assumptions given in Section 4.4, except for variations in mortality.

Figure 5. Consumption with and without mortality

depends greatly on the choice of $\bar{T}$. Figure 5 illustrates how the consumption profiles are affected in two cases, $\bar{T}=75$ (life expectancy) and $\bar{T}=86.3$ (the age at which assets are exhausted in the baseline case). Without the discount for mortality, post-retirement consumption carries a greater weight in the decision process, which should lead to earlier and higher retirement savings. Figure 6 supports this, τ decreases to 35.2 for $\bar{T}=75$ and to 33.3 for $\bar{T}=86.3$. Retirement savings are slightly higher for $\bar{T}=75$ at \$116,390 and much higher for $\bar{T}=86.3$ at \$166,564. Although retirement savings are relatively more attractive when mortality is removed, it is still suboptimal to start saving for retirement very early in the career, probably because of the anticipation of future income growth. It should also be noted that retirement savings



Results computed with the baseline assumptions given in Section 4.4, except that mortality is based on NCHS death rates.

Figure 6. Wealth with and without borrowing constraint

can be significantly overstated in models that use a fixed horizon much longer than life expectancy.

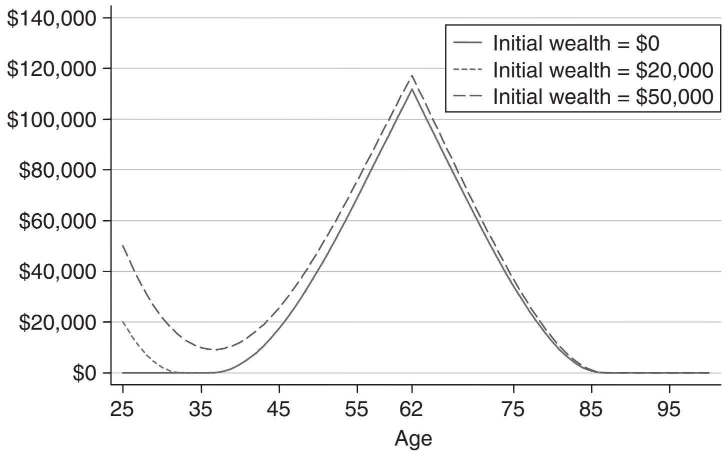
Relaxing the borrowing constraint has a meaningful impact on the results. As mentioned in the introduction, it is particularly important to add this constraint when lifetime is uncertain. Figure 6 illustrates this point by presenting the optimal wealth process when the borrowing constraint is removed for the cases with and without mortality. The case with mortality is based on the NCHS death rates. Figure 6 shows that the results are more distorted for the case with mortality. Young workers want to borrow in both cases, but retirees only want to borrow in the case with mortality. When young workers can borrow, they start saving for retirement later because they have to repay their debt first. Figure 6 shows that τ increases to 46 in the case without mortality and to 52 in the case with mortality. The results for cases with a limited amount of borrowing permitted should lie somewhere between those of Figures 1 and 6. Although not captured by the model, an additional concern is that retirement savings could be delayed further if the interest rate charged on the loan is greater than the risk-free rate.[18]

The case with $w_0 > 0$ can next be examined with the general solution developed in Section 2. As mentioned in that section, there is a difference between wealth that the individual builds up voluntarily and endowed wealth. In a case like the one in the baseline scenario where it is not optimal to start saving right away when $w_0 = 0$, any endowed wealth is automatically spent down over a period of time.[19] This may or may not affect retirement savings. If $w_0 < W(0, \tau)$, initial wealth is exhausted before time τ and the solution has $M = 3$: it starts in a non-binding phase and afterwards has the same three phases as for the case with $w_0 = 0$. If $w_0 > W(0, \tau)$, $M = 1$ and the individual does not draw down his entire savings before starting to save for retirement. In other

[18] The interaction between credit use and retirement savings would be an interesting topic to investigate in future research.

[19] In practice, the outcome may be different if retirement savings are illiquid or if a penalty tax applies.

https://doi.org/10.1017/S1474747221000284 Published online by Cambridge University Press



Results computed with the baseline assumptions given in Section 4.4, except for initial wealth.

Figure 7. Effect of positive initial wealth

words, a portion of the endowed wealth is used to increase retirement savings. The threshold $W(0, \tau = 36.4)$ is \$39,870 for the baseline case and Figure 7 illustrates two possible scenarios for the solution by contrasting the cases $w_0 = \$20{,}000$ and $w_0 = \$50{,}000$. The \$20,000 case is below the threshold and initial savings are drawn down by age 33. The solution is then identical to the one for the baseline scenario with savings starting again at age 36 and being exhausted by age 86. The case with $w_0 = \$50{,}000$ is above the threshold and there is only one long non-binding period which ends at a slightly later time $T = 86.6$.

There are several other model variations that would be interesting to consider (bequests, risky asset, income risk), but they imply losing the closed-form solution. Still, the expected results can be discussed briefly. As discussed in Section 2.3, adding income risk would add a layer of precautionary savings in the early part of the career. For risky pensions, the effect would depend on the specific modeling assumptions. To get a quick idea of the possible magnitude of the results, one can think of a worst-case scenario that indicates the earliest possible age of onset. For example, in the baseline scenario if pensions can decrease by at most 25%, the earliest retirement savings could start is age 33.9. A much more catastrophic scenario would be needed to induce savings at the beginning of the career. Next, adding a bequest function would likely affect the speed at which the individual exhaust his savings after retirement. On the other hand, this effect would be small if the bequest motive is mostly satisfied with sources outside liquid wealth such as life insurance or the survivor pension paid by Social Security.

Adding risky assets would mean higher expected returns, but also more volatility which would be penalized by risk averse investors. Merton's (1971) closed-form solution offers a simple strategy to estimate the combined effect of risk and return. In his solution for the optimal consumption process, the effect of adding a risky asset is to increase the rate of return from r to $r + 0.5(\mu - r)^2/\gamma\sigma^2$, where μ is the expected risky return and σ^2 the variance of the risky asset. For example, if the risk-free rate is 3%, the coefficient of risk aversion is $\gamma = 3$, the risky asset expected return is 8 %, and its standard deviation is $\sigma = 15\%$, $r = 4.9\%$ would be used instead of 3 %. Since the effect of adding a risky asset is similar to the effect of increasing the rate of return, the impact can be estimated with the results presented in Figure 2D. An earlier onset of retirement savings could arise, but only in cases where expected returns are relatively high when compared to risk penalties. In any case, it would take a significantly high risk premium for an immediate saving strategy to be optimal.

4.6 Onset of retirement savings in other life-cycle models

Our model's prediction for the age of onset of retirement savings can be compared to those of related life-cycle models by Davies (1981), Hubbard *et al.* (1995), Carroll (1997), Gourinchas and Parker (2002), and Cocco *et al.* (2005). These models are retained because they share a common core of assumptions with this model, they feature intertemporal consumption decisions, non-negative wealth, and declining income at the end of the career.[20] Some of these models are more restrictive than ours because they use a fixed horizon, impose parameter restrictions, or require that pensions be a multiple of earnings in the year before retirement. They all also differ in that they rely on numerical optimization techniques rather than an explicit solution. On the other hand, the last three models are more general in the sense that they include income risk and, in the case of Cocco *et al.* (2005), portfolio choice. Carroll (1997) and Gourinchas and Parker (2002) do not impose a borrowing constraint directly, but their assumption of a possible zero-income scenario guarantees that the individual never chooses to borrow. With this approach, it is not possible to obtain our results as a special case by removing income risk because wealth would become negative.

Table 2 provides a summary of the assumptions used in each model and in our baseline scenario, along with their predicted age for the onset of retirement savings.[21] The predictions of these models are consistent with the analysis in Section 4.5, they all find that young workers do not save for retirement, they wait until their thirties or forties. Our analysis in Figure 2 allows us to tie in all these results and offer an explanation for their differences by examining their choice of parameters. The relatively high income growth rate (5.4 %) in Davies (1981) can be responsible for the later onset between age 40 and 45. Carroll's (1997) later age range of 45–50 for operatives and managers is likely due to the fact that he uses a high discount rate relative to the rate of return (4 % versus 0 %). His assumption that unskilled laborers have no income growth after 40 can explain why he finds an earlier age around 35–40 for that group. Similarly, lower income growth for the case with high school education can justify the earlier age of 33 in Hubbard *et al.* (1995). Savings also start a

[20] This is not an exhaustive list, these papers should be considered as representative of other works, for example Campbell and Viceira (2002), Cagetti (2003), and Samwick (2006). Although Deaton (1991) is a common reference for borrowing constraints, it is not included here because its assumption of stationary income growth prevents it from predicting an onset of retirement savings.

[21] Most models have age-specific income growth rates. Since they all start with a growth phase and end with a declining (or lower growth) phase, the average income growth rates are presented for these two periods.

Table 2. *Optimal onset of retirement savings* (τ) *in related models*

	Assumptions						
Authors	β	Mortality	γ	r	Average income growth (by age range)	Pensions replacement rate (annual growth)	Optimal onset of retirement savings (τ)
Davies (1981)	1.5 %	Inverse logistic	1	3 %	20–56: 5.4 % 57–65:–4.8 %	100 % (−2.25 %)	38*
			4				44*
Hubbard *et al.* (1995)	3 %	Fixed 80	3	3 %	High school: 21–49: 1.8 %* 50–65: −1.8 %*	54 % (–1.9 %)*	33
					College: 21–49: 2.5 %* 50–65: −5.8 %*	100 % (–2.4 %)*	37*
Carroll (1997)	4 %	Fixed 80	2	0 %	Unskilled: 25–40: 3 % 41–65: 0 %	70 % (0 %)	35–40*
					Operatives: 25–50: 2.5 % 51–65: 1 %		45–50*
					Managers: 25–55: 3 % 56–65: −1 %		45–50*
Gourinchas and Parker (2002)	4 %	Fixed 87	0.5	3.4 %	26–50: 1.4 %* 51–65: −1.1 %*	See note	38*
Cocco *et al.* (2005)	4 %	NCHS	10	2 %	20–43: 2.4 %* 44–65: −0.5 %*	68 % (0 %)	35*
Baseline	3 %	Gompertz	3	3 %	25–50: 1.7 % 51–65: −0.6 %	60 % (0 %)	36

Note: An asterix indicates values estimated from graphical results. In Cocco *et al.* (2005), the risk premium is 4 %. In Gourinchas and Parker (2002), the problem is truncated at retirement and the retiree receives a multiple of the last permanent component of income. The value function used at retirement implies a multiple of about 2 % and a fixed horizon around age 87. Two set of ages for the onset of retirement savings are reported: 38 (based on a target wealth analysis) and 40–45 (based on the life-cycle/buffer-stock breakdown).

little sooner at age 35 in Cocco *et al*. (2005), possibly because of a high risk aversion assumption with $\gamma = 10$. At the other extreme, Gourinchas and Parker (2002) use a low level of risk aversion at $\gamma = 0.5$ and find an age of 38. As shown in Figure 2, the low risk aversion should normally push the onset of savings much later, but two offsetting forces in Gourinchas and Parker (2002) prevent this: the individual does not have much of a pension, and savings must be spread out over a relatively long fixed horizon going up to age 87.

It should be clarified that models featuring additional risks do not have a precise time (such as τ) that identifies when retirement savings start. This age is typically a rough estimate corresponding to the time where wealth starts building up more quickly in the solution, making it difficult to conduct analyzes such as the sensitivity illustrations in Figure 2. Some exceptions include Gourinchas and Parker (2002) who offer a strategy to detangle buffer-stock (precautionary) and life-cycle (retirement) savings. They define the life-cycle component as the solution to the model without income risk. While this is an interesting approach, the problem is that models such as Carroll (1997) and Gourinchas and Parker (2002) rely on income risk to prevent violations of the borrowing constraint. Without income risk, the results are those of a model without borrowing constraints and they display the same pattern as illustrated in Figure 6. As discussed earlier, removing the borrowing constraint delays the onset of retirement savings by several years. Indeed, Gourinchas and Parker (2002) report two different ages for the onset of retirement savings: 38 (based on a target wealth analysis) and 40–45 (based on the life-cycle/buffer-stock breakdown). Besides understating life-cycle savings, this approach also overstates precautionary savings because they are defined as the complement of life-cycle savings. This problem can be corrected by measuring retirement savings with a model with explicit borrowing constraints such as the one suggested in this paper.[22]

5 Conclusion

With a declining outlook for pensions, a larger share of retirement income will have to be derived from individual savings. In that context, the objective of this paper is to uncover the economic dynamics underlying the formation and exhaustion of individual retirement savings within the context of Yaari's (1965) model. This mechanism is often obscured in numerical solutions to increasingly complex life-cycle models. This paper develops a life-cycle model which captures enough realistic features to produce credible numerical illustrations, yet is sufficiently parsimonious to generate explicit solutions. Specifically, the theoretical contribution of this paper is threefold. First, it revisits Yaari's (1965) model and solves it for the general case. Second, it shows how to obtain closed-form equations for that solution for certain sets of realistic assumptions. Third, it derives several comparative statics for τ and T, the times of onset, and exhaustion of retirement savings. While the previous literature has often focused on wealth as a determinant of consumption and saving behavior, our

[22] In future work, an interesting extension to Gourinchas and Parker's (2002) approach would be to divide further the amount of precautionary savings into a component that is due to risky career income and one that is due to risky pension income.

https://doi.org/10.1017/S1474747221000284 Published online by Cambridge University Press

solution indicates that the evolution of this behavior over the life-cycle is explained by the shape of the function $\lambda(t)$. This function captures the combined effect of income profiles, risk aversion, time preferences, investment return, and mortality.

Analyzing the resulting solution reveals that the optimal retirement savings strategy can be divided into three stages over the life-cycle: a period $[0, \tau]$ without any savings, a period $[\tau, T]$ where savings are accumulated and then depleted, and, lastly, a period $[T, \omega]$ where savings are permanently exhausted. The same trade-off motivates the individual to either start saving or to save more: the marginal benefit of a dollar saved must be greater than its opportunity cost. This relationship highlights two of the main factors hindering retirement savings: (1) those with relatively lower incomes have higher opportunity costs and are likely to be better off paying their bills than saving for retirement, and (2) the benefits of saving may not loom large in the decision process because they are effectively discounted as if they were consumed at time T, which ranges from about 62 to 100 years old in our illustrations. Efforts to promote savings may be wasted if they fail to recognize these hurdles. These major obstacles also suggest that, for some, low observed retirement savings can be in line with optimal behavior rather than being a sign of deviation from normative behavior. When evaluating whether people save enough for retirement, it should be kept in mind that there is a difference between making *ex-ante* optimal saving decisions and having *ex-post* adequate retirement savings.

The model developed in this paper could be applied to quantify the impact on savings and welfare of various retirement-related programs. By contrast with simple retirement savings accumulation techniques, some advantages of the model proposed in this paper are that it: (1) recognizes the opportunity cost associated with saving, (2) takes into account the intricate relationship between optimal individual saving behavior, income, and pensions, and (3) measures and contrasts welfare in different scenarios.[23] While this model may not be as flexible as existing ones in terms of incorporating various sources of risk, it does not rely on numerical optimization and has an exact solution which is much faster to obtain. This is an attractive feature for problems requiring a large number of repetitions of the solution, for example when dealing with large datasets or solving for welfare equivalents. Of course, it would also be very valuable to explore ways to incorporate risks in the solution without resorting to numerical optimization.

This paper presents multiple opportunities for future research. For instance, it would be interesting to examine whether incentives such as favorable tax treatment and employer matching contributions make it optimal to save early. Another application is to assess the benefits for employers of offering defined contribution plans in light of the finding that saving early is not optimal. Lastly, a third area of research suggested in Section 4.5 is to analyze the interaction between credit use and the onset of retirement savings.

[23] An example of an application with welfare-equivalents was given in a previous version of the paper for the case of automatic enrollment. Sup-optimal saving behavior without automatic enrollment was represented by assuming that the individual delayed the onset of his savings by d years (after his optimal date τ) and adopted his optimal saving strategy afterwards. The results showed that automatic enrollment was associated with welfare losses for those with mild to moderate sub-optimal behavior and welfare gains for those with more serious issues ($d=20$).

References

Cagetti, M. (2003) Wealth accumulation over the life cycle and precautionary savings. *Journal of Business and Economic Statistics*, **21**(3): 339–353.

Campbell, J. Y. and Viceira, L. M. (2002) *Strategic Asset Allocation*. New York: Oxford University Press.

Carroll, C. (1997) Buffer-stock saving and the life cycle/permanent income hypothesis. *The Quarterly Journal of Economics*, **112**(1): 1–55.

Cocco, J. F., Gomes, F. J., and Maenhout, P. J. (2005) Consumption and portfolio choice over the life cycle. *The Review of Financial Studies*, **18**(2): 491–533.

Davies, J. B. (1981) Uncertain lifetime, consumption, and dissaving in retirement. *The Journal of Political Economy*, **89**(3): 561–577.

Deaton, A. (1991) Saving and liquidity constraints. *Econometrica*, **59**(5): 1221–1248.

Duflo, E., Orszag, P., Gale, W., Saez, E., and Liebman, J. (2007) Savings incentives for low- and moderate-income families in the United States: why is the saver's credit not more effective? *Journal of the European Economic Association*, **5**(2–3): 647–661.

Gourinchas, P.-O. and Parker, J. A. (2002) Consumption over the life-cycle. *Econometrica*, **70**(1): 47–89.

He, H. and Pages, H. F. (1993) Labor income, borrowing constraints, and equilibrium asset prices. *Economic Theory*, **3**: 663–696.

Hubbard, R. G., Skinner, J., and Zeldes, S. P. (1995) Precautionary saving and social insurance. *Journal of Political Economy*, **103**(2): 360–399.

Karatzas, I. and Shreve, S. E. (1998) *Methods of Mathematical Finance, Applications of Mathematics*, Volume 39, New York: Springer-Verlag.

Leung, S. F. (1994) Uncertain lifetime, the theory of the consumer, and the life-cycle hypothesis. *Econometrica*, **62**(5): 1233–1239.

Leung, S. F. (2000) Why do some households save so little? A rational explanation. *Review of Economic Dynamics*, **3**: 771–800.

Leung, S. F. (2001) The life-cycle model of saving with uncertain lifetime and borrowing constraint: characterization and sensitivity analysis. *Mathematical Social Sciences*, **42**: 179–201.

Leung, S. F. (2007) The existence, uniqueness, and optimality of the terminal wealth depletion time in life-cycle models of saving under uncertain lifetime and borrowing constraint. *Journal of Economic Theory*, **134**: 470–493.

Mariger, R. P. (1987) A life-cycle model with liquidity constraints: theory and empirical results. *Econometrica*, **55**(3): 533–557.

Merton, R. C. (1971) Optimum consumption and portfolio rules in a continuous-time model. *Journal of Economic Theory*, **3**: 373–413.

Milevsky, M. A. (2006) *The Calculus of Retirement Income: Financial Models for Pension Annuities and Life Insurance*. New York: Cambridge University Press.

Mitchell, O. S., Utkus, S. P. and Yang, T. (2007) Turning workers into savers? Incentives, liquidity, and choice in 401(k) plan design. *National Tax Journal*, **60**(3): 469–489.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (2007) *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. New York: Cambridge University Press.

Samwick, A. A. (2006) Saving for retirement: understanding the importance of heterogeneity. *Business Economics* (January): 21–27.

Thaler, R. H. and Benartzi, S. (2004) Save more tomorrow: using behavioral economics to increase employee saving. *Journal of Political Economy*, **112**(1): 164–187.

Yaari, M. E. (1965) Uncertain lifetime, life insurance, and the theory of the consumer. *The Review of Economic Studies*, **32**(2): 137–150.

Appendix A: Solution and proof to the problem of Section 2

Proposition 1. Let $\lambda(t)$ and $W(t,T)$ be the continuous functions defined respectively in equations (5) and (11). Assume that $\lim_{t\to\omega}\lambda(t)=0$ and $w_0\geqslant 0$. For any two periods i and k, define $\mathscr{T}_i(t)$ as an inverse function such that $t\in D_k$, $\mathscr{T}_i(t)\in D_i$, and $\lambda(t)=\lambda(\mathscr{T}_i(t))$. The solution to the optimization problem in (1)–(3) takes two possible forms, depending on the borrowing constraint being binding or not. The times $T_1,\ldots,T_M$ separate the binding and non-binding periods and are constructed with the following four-step process:

Step 1. Identify the periods D_i where $\lambda(t)$ is strictly decreasing
Let $D_i=[t_i^l,\ t_i^u)$, denote the periods $i=1,\ldots,N$ where the function $\lambda(t)$ is strictly decreasing, i.e. $\lambda'(t)<0$ for all $t\in(t_i^l,\ t_i^u)$ and $\lambda'(t)\geqslant 0$ for all other t. If $N=1$, $t_1^l=0$, and $w_0=0$, then $M=0$ and the borrowing constraint is binding for all t. Otherwise, proceed to Step 2.

Step 2. Determine whether the borrowing constraint is binding at $t=0$
In the first period $[0,\ T_1]$, the borrowing constraint is non-binding in Case 1 and binding in Case 2. Case 1 applies if $w_0>0$ or if $w_0=0$ and there exists a period $i\geqslant 1$ ($i>1$ if $t_1^l=0$) with a solution $t_i\in D_i$ to $W(0,t_i)=0$ with $\lambda(t_i)\geqslant\lambda(0)$. Proceed to Step 3 in Case 1 and to Step 4 in Case 2.

Step 3. Solve for T_1 in $W(0,\ T_1)=w_0$
For each period $i\geqslant 1$ ($i>1$ if $w_0=0$ and $t_1^l=0$) with $W(0,t_i^l)\leqslant w_0$ and $W(0,t_i^u)>w_0$, there exists a unique value $t_i\in[t_i^l,t_i^u)$ such that $W(0,t_i)=w_0$. There exists at least one period i such that a time t_i exists. If more than one solution t_i exists, select the period i^* with the highest associated value for $\lambda(t_i)$ and set $T_1=t_i{}^*$. If $i^*=N$, then $M=1$ and the solution is complete. Otherwise, proceed to Step 4.

Step 4. Solve for T_m in $W(T_m, \mathscr{T}_i(T_m))=0$ (with $T_{m-1}\in D_k$)
For each period $i\in(k,N]$, let $[\bar{t}_i^l,\ \bar{t}_i^u]$ denote the range of values of $t\in D_k$ for which $\mathscr{T}_i(t)$ is defined and $t\geqslant T_{m-1}$. If $[\bar{t}_i^l,\ \bar{t}_i^u]$ is non-empty, $W(\bar{t}_i^l,\mathscr{T}_i(\bar{t}_i^l))\leqslant 0$, and $W(\bar{t}_i^u,\mathscr{T}_i(\bar{t}_i^u))>0$, there exists a unique value $t_i\in[\bar{t}_i^l,\ \bar{t}_i^u)$ such that $W(t_i,\mathscr{T}_i(t_i))=0$. There exists at least one period $i\in(k,N]$ such that a time t_i exists. If more than one solution t_i exists, select the period i^* with the highest associated value for $\lambda(t_i)$ and set $T_m=t_{i^*}$ and $T_{m+1}=\mathscr{T}_{i^*}(t_{i^*})$. If $T_{m+1}\in D_N$, then $M=m+1$ and the solution is complete. Otherwise, repeat Step 4.

Accordingly, the optimal consumption and wealth processes within any non-binding period $(T_m,\ T_{m+1})$ are given respectively by equations (10) and (11). Within any binding period $[T_m,\ T_{m+1}]$, these processes become $c_t^*=y_t$ and $W_t^*=0$.

Proof. The first step in solving an optimization problem such as the one in (1)–(3) is to formulate its Lagrangian as

$$\mathscr{L}=\int_0^{\omega} f(t)u(c_t)dt+\mu\left[w_0-\int_0^{\omega} e^{-rt}(c_t-y_t)dt\right]+\int_0^{\omega}\eta(t)\left[w_0-\int_0^{t} e^{-rs}(c_s-y_s)ds\right]dt, \tag{34}$$

https://doi.org/10.1017/S1474747221000284 Published online by Cambridge University Press

where $\mu \geqslant 0$ and $\eta(t) \geqslant 0$ denote respectively the Lagrange multipliers for the budget constraint and the borrowing constraint. Following He and Pages (1993), the problem can be simplified by introducing the process $X(t) = \mu + \int_t^w \eta(s)ds$ and rewriting the Lagrangian more compactly as

$$\mathscr{L} = \int_0^\omega f(t)u(c_t)dt + X(0)w_0 + \int_0^\omega X(t)e^{-rt}(y_t - c_t)dt. \tag{35}$$

For c_t^* to be an optimal control, there must exist a process $X(t)$ such that the following five Kuhn–Tucker necessary conditions are satisfied

$$\frac{\partial \mathscr{L}}{\partial c} = 0 \qquad \Rightarrow \qquad c_t{}^* = u'^{-1}\left(\frac{X(t)e^{-rt}}{f(t)}\right), \tag{36}$$

$$c_t{}^* = y_t \quad \text{whenever} \quad W_t{}^* = 0, \tag{37}$$

$$X'(t)W_t{}^* = 0 \quad \text{and} \quad X'(t) \leqslant 0 \quad \text{for all } t, \tag{38}$$

$$W_0{}^* = w_0, \tag{39}$$

$$W_t{}^* \geqslant 0 \quad \text{for all } t, \tag{40}$$

where

$$W_t{}^* = \int_t^\omega e^{-r(s-t)}(c_s{}^* - y_s)ds. \tag{41}$$

Below, it is verified that these conditions are satisfied by a candidate solution consisting of the process

$$X(t) = \begin{cases} \lambda(t) & \text{for } t \in [T_m, T_{m+1}] \text{ in binding periods,} \\ \lambda(T_{m+1}) & \text{for } t \in [T_m, T_{m+1}] \text{ in non-binding periods} \end{cases} \tag{42}$$

and the times $T_1, \ldots, T_M$ defined in Proposition 1. The first condition in (36) is met because the optimal consumption process c_t^* in (10) and (14) is obtained by substituting $X(t)$ in (36). The second condition in (37) is satisfied because $c_t^* = u'^{-1}(\lambda(t)e^{-rt}/f(t)) = y_t$ whenever $W_t^* = 0$. The third condition in (38) holds because (1) $X'(t) = d\lambda(T_{m+1})/dt = 0$ whenever $W_t^* > 0$, and (2) by construction of the times $T_1, \ldots, T_M$, $X'(t) = \lambda'(t) \leqslant 0$ whenever $W_t^* = 0$. The fourth condition in (39) follows from the requirement that T_1 satisfies $W_0{}^* = W(0, T_1) = w_0$ in Case 1 and that $W_0{}^* = 0 = w_0$ in Case 2. The fifth condition in (40) ($W_t^* \geqslant 0$ for all t) necessitates a lengthier proof and dual methods are used here to present a more compact argument.[24] For that purpose, let $c_t^* = u'^{-1}(X(t)e^{-rt}/f(t))$ and define the dual utility function

$$J(X^*(t), t) = \min_{X \in \mathscr{D}} J(X(t), t) = \int_t^\omega f(s)u(c_s{}^*)ds - \int_t^\omega X(s)e^{-rs}(c_s{}^* - y_s)ds. \tag{43}$$

As demonstrated in He and Pages (1993), the optimization problem in (1)–(3) is equivalent to finding the process $X \in \mathscr{D}$ which minimizes $J \equiv J(X(t),t)$, where $\mathscr{D}$ represents the set of non-negative and decreasing processes. The connection between

[24] An alternative proof which is not based on dual methods is also available from the author upon demand.

the minimization of J and the borrowing constraint can be established by observing that $J_x(X^*(t),t)=e^{-rt}W_t^*$. If J is minimized, the borrowing constraint is satisfied because it must be the case that $J_x(X^*(t),t)\leqslant 0$ along the optimal path (otherwise, J could be decreased by picking a lower $X(t)$). To show that our solution indeed minimizes J, first observe that the optimal process $X(t)$ must take the form given in equation (42) to satisfy conditions (36), (37), (38), and (40). The argument can then be completed by showing that J would increase if, among the possible solutions that satisfy the budget constraint (i.e. the times t_i such that $W(0,t_i)=w_0$ and $W(t_i,\mathscr{T}_i(t_i))=0$), a t_i other than the one with the highest value of $\lambda(t_i)$ was chosen. For Case 1, J decreases with $X(0)=\lambda(t_i)$ because $J_x(X(0),0)=-w_0\leqslant 0$; setting $T_1=t_i$ with $\lambda(t_i)<\lambda(t_{i^*})$ would thus increase J. For Case 2 with $T_{m-1}\in D_k$, J decreases with $X(t)=\lambda(t_i)$ for all $t\in(t_i,t_k^u)$ because over that range $J_x(X(t),t)=-\int_t^{\mathscr{T}_i(t)} e^{-rs}(c_s{}^*-y_s)\, ds=\int_{t_i}^t e^{-rs}(c_s{}^*-y_s)ds<0$. Choosing any of the other admissible $t_i\in D_k$ with $\lambda(t_i)<\lambda(t_{i^*})$ for T_m would imply that $X(t)$ decreases for all $t\in(t_{i^*},t_i)\subset(t_{i^*},t_k^u)$ and that consequently J increases.

Finally, Proposition 1's proof can be completed by demonstrating the existence and uniqueness of the times t_i. Within each period D_i, this stems from the continuity of the functions $W(0,t)$ and $W(t,\mathscr{T}_i(t))$ and by noting that in Case 1 $dW(0,t)>0$ for all $t\in(t_i^l,t_i^u)$ and that in Case 2 $dW(t,\mathscr{T}_i(t))/dt>0$ for all $t\in(t_i,\bar{t}_i^u)$ if $W(t_i,\mathscr{T}_i(t_i))=0$. The result follows by applying the intermediate value theorem in conjunction with the boundaries defined in Proposition 1. Similar arguments can be used to prove the global existence of at least one time t_i in each problem. In that case, it must also be considered that in Case 1 $W(0,0)=0$ and $\lim_{t\to\omega} W(0,t)=\infty$, and that in Case 2 it can be shown that (1) $W(\bar{t}_i^l,\mathscr{T}_i(\bar{t}_i^l))\leqslant 0$ for all i with $[\bar{t}_i^l,\bar{t}_i^u]$ non-empty and (2) there must exist at least one period i with $W(\bar{t}_i^u,\mathscr{T}_i(\bar{t}_i^u))>0$.[25]

Remarks on Proposition 1. In the equations, $W(0,t_i)=w_0$, $W(t_i,\mathscr{T}_i(t_i))=0$, and $\lambda(t)=\lambda(\mathscr{T}_i(t))$, the times t_i and $\mathscr{T}_i(t)$ can be inverted numerically by applying the bisection method with the bounds defined in Proposition 1. The set $[\bar{t}_i^l,\bar{t}_i^u]$ is non-empty when $\lambda(t_i^l)\geqslant\lambda(t_k^u)$ and $\lambda(t_i^u)<\lambda(t_{m-1})$ and the bounds are given by $\bar{t}_i^l=\max[T_{m-1},\mathscr{T}_k(t_i^l)]$ and $\bar{t}_i^u=\min[t_k^u,\mathscr{T}_k(t_i^u)]$. If $\lambda'(t)=0$ for all t between t_i^u and t_{i+1}^l, some minor adjustments have to be made to Proposition 1: $[t_i^l,t_i^u)$ becomes $[t_i^l,t_i^u]$, $[\bar{t}_i^l,\bar{t}_i^u)$ becomes $[\bar{t}_i^l,\bar{t}_i^u]$, and t_{k+1}^l is substituted to T_{m-1}. Last, to be more precise, i^* is such that $\lambda(t_i)<\lambda(t_{i^*})$ for all $i<i^*$ and $\lambda(t_i)\leqslant\lambda(t_{i^*})$ for all $i>i^*$.

If $\lambda(t)$ does not converge to zero as t gets larger, then the solution may end in a non-binding phase. The condition $\lim_{t\to\omega}\lambda(t)=0$ should generally be satisfied if the force of mortality increases quickly enough, i.e. the chance of survival at extremely old ages is virtually zero. When mortality is not included in the model or when the force of mortality converges to a positive constant in an infinite horizon model, the condition may or may not be satisfied depending on the model's specific combination of parameters. Nevertheless, results can easily be extended to the case where $\lambda(t)$ does not converge to zero. Essentially, when solving for the time at which a non-binding

[25] The assumption that $\lambda(t)$ converges to zero implies that $\lim_{t\to\omega} W(0,t)=\infty$ and also that, if $k<N$, there must exist at least one period $i>k$ with $\lambda(t)>\lambda(\bar{t}_i^u)$ for all $t\in(\bar{t}_i^u,\mathscr{T}_i(\bar{t}_i^u))$, and thus $W(\bar{t}_i^u,\mathscr{T}_i(\bar{t}_i^u))>0$.

period ends, the corner solution at the last possible age ω (or ∞ with an infinite horizon) has to be considered. More precisely, the procedure for solving for T_1 in $W(0,T_1)=w_0$ has to be adjusted to include the case $W(0,\omega)=w_0$ if $\lambda(\omega)>\lambda(0)$. Similarly, for $W(T_m,T_{m+1})=0$, the case $W(T_m,\omega)=0$ has to be considered if $\lambda(\omega)>\lambda(T_m)$. The rest of the process is the same (including the test for the highest λ), but a constant λ is substituted to $\lambda(T_1)$ in equation (6) and $\lambda(T_m)$ is substituted to $\lambda(T_{m+1})$ in equation (9).

The results can also be extended to the case with discontinuities in $\lambda(t)$: in the limit, these jumps can be replicated by a very rapid change in $\lambda(t)$. The solution is mostly unaffected when the break in $\lambda(t)$ occurs when the function is increasing. In that case, functions evaluated at discontinuous points are simply replaced with their left-side limit. If the jump occurs at a time $\tilde{t}$ where the function $\lambda(t)$ is decreasing, the methodology used to solve for the times t_i has to be slightly modified. Specifically, three possible scenarios have to be considered: the solution can be before, at, or after time $\tilde{t}$. Standard techniques such as the intermediate value theorem can be used to identify which case applies. When the solution is at time $\tilde{t}$ in Case 1, a constant λ is substituted to $\lambda(T_{m+1})$ and the problem becomes: solve for λ in $W(0,\tilde{t})=w_0$. In Case 2, the problem changes to $W(\tilde{t},t_i)=0$ if $\tilde{t}\in D_k$ and to $W(t_i,\tilde{t})=0$ with $\lambda(T_{m+1})$ replaced by $\lambda(T_m)$ if $\tilde{t}\in D_{i>k}$.

Appendix B: Other utility and mortality functions

This appendix discusses some additional cases where closed-form equations can be derived. The results for c_t^* and W_t^* for the log utility case $u(c_t)=\ln(c_t)$ are obtained by substituting $\gamma=1$ in equations (27) and (28). The exponential utility function $u(c_t)=-\exp(-ac_t)$ with $a>0$ also yields explicit results for the optimal consumption and wealth processes, which become

$$c_t{}^*=y_T-\frac{(r-\beta)(T-t)+e^{-m/b}(e^{t/b}-e^{T/b})}{a}, \tag{44}$$

$$\begin{aligned} W_t{}^*=\;&y_T\frac{1-e^{-r(T-t)}}{r}-\frac{r-\beta}{ar}\left(T-t-\frac{1-e^{-r(T-t)}}{r}\right)\\ &-\frac{e^{-m/b}}{a}\left[\frac{e^{T/b-r(T-t)}-e^{t/b}}{-r+1/b}-e^{T/b}\frac{1-e^{-r(T-t)}}{r}\right]-Y_{t,T}. \end{aligned} \tag{45}$$

More generally, other members of the HARA family of utility functions, i.e. those that take the form

$$u(c_t)=\frac{1-\theta}{\theta}\left(\frac{\kappa c_t}{1-\theta}+\eta\right)^{\theta} \tag{46}$$

can produce explicit equations for c_t^* and W_t^*. However, if $\eta\neq 0$, the condition $\lim_{c\to 0}u'(c)=\infty$ is violated. Less technically, this means that zero consumption could be optimal and the equations given above would have to be adjusted slightly to reflect periods with zero consumption, if any.

In addition, the mortality distribution is not restricted to the Gompertz case. For example, the case of the Makeham distribution (which adds a constant K to the

Gompertz force of mortality) involves only a minor adjustment to our previous results, equation (28) would have to be multiplied by $\exp(-Kt/\gamma)$, and $-bK/\gamma$ would be added to α in equation (27). The case with a constant force of mortality μ at all ages is also straightforward, but should be used with caution as it makes it difficult to reflect very high mortality rates at older ages. More generally, the results can be adjusted to fit any discrete mortality table with survival probabilities $p_k, p_{k+1}, \ldots, p_l$. Assuming a constant force of mortality within each period, the function $A_{t,T}$ in (25) would be replaced by

$$A_{t,T} = e^{rt} \sum_{s=k}^{l} p_s^{1/\gamma} \left(\frac{p_{s+1}}{p_s} \right)^{-\max(s,t)/\gamma} \frac{e^{\sigma_t \min(s+1,T)} - e^{\sigma_t \max(s,t)}}{\sigma_s}, \quad \text{with } \sigma_s = \frac{r - \beta + \ln(p_{s+1}/p_s)}{\gamma} - r, \; t \in [k, k+1) \text{ and } T \in [l, l+1). \tag{47}$$