Heterogeneous time preference and the distribution of wealth

David Andolfatto and James Redekop
Economics Department
University of Waterloo
Waterloo, Ontario N2L-3G1

July, 1998

Abstract

This paper analyzes a dynamic model in which physical capital can be accumulated or depleted, and labour supply is endogenous. The distribution of income is then endogenously determined by both technological parameters of production, and the distribution of agents' discount parameters. Degenerate wealth distributions, in which only the most patient agents have any wealth, are avoided by having a fraction of the agents die each period, and bequeath their wealth to descendants with independently random discount parameters. On average, more patient agents will have higher wealths and incomes, but in the short run agents' stocks of wealth depend on their inherited wealth. If a patient individual lives long enough, she will retire and live on only investment income, while if an impatient individual lives long enough, he will deplete all his wealth and live on only labour earnings. The effects of a general increase in patience are an increase in the wage rate, a lowering of the return on capital, and general increases in wealth, income, and utility. Possibilities for engineering such an increase, by promoting "artificial patience", could include favourable taxation of investment income, forced savings such as payroll-tax financed pension plans, or public subsidies for education and health.

1 Introduction

Explanations of individuals' high or low wealths and incomes often focus on idiosyncratic traits or events: He has a tremendous talent for playing basketball, he is a career minor-leaguer, she had the look, voice and songs that millions of people instantly adored, she pays her dues 200 nights a year singing in bars, he lost his life savings in the Bre-X fiasco, she was abandoned by the father of her children, he won the lottery, she is the daughter of a billionaire, and so on. Less often, we hear explanations that involve economic decisions on the part of the individuals: She went to medical school, he got a Ph. D., she built her small business into a large one, he mismanaged his small business into bankruptcy, she saved for her retirement, he relied on his Social Security pension.

The economist has to explain income and wealth distributions in economic terms, and hope or assume that the idiosyncratic effects will be negligible in the aggregate. We believe that the non-idiosyncratic variables determining wealth include investment behaviour of all kinds, and not much else. We include inherited wealth as a non-idiosyncratic effect, since it is the legator's investment behaviour that determines the size of the bequest, and hence the wealth of the legatee and the distribution of wealth after the legator's death.

The economic literature on the *measurement* of wealth or income distributions is vast; but there have been very few attempts to build models in which these distributions are endogenously determined, by preferences and technologies. It is an old idea that an individual's wealth depends significantly on her willingness to invest, or implicitly on the discount the agent applies to future utilities; see Rae [5]. A formal model with this feature appears in Epstein and Hynes [3]. In the model, n infinitely-lived agents have heterogeneous discount parameters $\beta_1 < \beta_2 < \ldots < \beta_n < 1$, and opportunities to accumulate or deplete their capital stocks according to $k_{t+1} = (1 - \delta + r)k_t - c_t$. An individual with discount parameter β_i chooses to accumulate $(k_{t+1} > k_t)$ if and only if $\beta_i(1 - \delta + r) > 1$, so in a steady state we must have $k_i = 0$ for $i < n$, and $\beta_n(1 - \delta + r) = 1$. The wealth distribution is degenerate, with only the most patient agent owning any wealth.

The recursive preferences in Lucas and Stokey [4] avoid this degeneracy problem, with the assumption that agents valuing consumption streams according to

$$V(c_0, c_1, \ldots) = W(u(c_0), V(c_1, c_2, \ldots))$$

satisfy "increasing impatience", in which $W_{22} < 0$. As a special case of recursive preferences, we could have $W(u, V) = u + \beta V$, leading to the discounted utility preference

$$V(c_0, c_1, \ldots) = \sum_{j=0}^{\infty} \beta^j u(c_j);$$

so the partial W_2 ($= \beta$, a constant, in the discounted utility case) is a local or conditional measure of the agent's patience. If $W_{22} < 0$, then agents with identical preferences, but different initial capital stocks, will at first behave heterogeneously. The poorer agents will act as if they had higher discount parameters, and will accumulate; and the wealthier agents will behave like impatient agents, with $\beta_i(1 - \delta + r) < 1$, and deplete their capital stocks. In a steady state, the wealth distribution will be concentrated at a point.

Of the two models, we believe that the first has the less accurate prediction for aggregate wealth distributions, while the second has the less realistic implications for individual behaviour. Specifically, there is a significant fraction of the population having positive amounts of wealth, and not all of these people are especially patient accumulators; and there seems to be no tendency for people who have accumulated wealth to accumulate less as they grow wealthier.

In our model we assume that (given a constant interest rate) people can be classified from birth as "accumulators" or "decumulators", and that what prevents the accumulators from accumulating indefinitely is their finite lifespan (and not our finite limit on this amount, which we impose for purely technical reasons). A nondegenerate stationary wealth distribution arises from the eventual deaths of the accumulators, and their replacement by descendants who are on average not as patient.

We believe that we can explain any wealth distribution *or* labour income distribution, just by adjusting the exogenous distribution of individuals' patience parameters. Our computations have shown that we can also fit the outcomes of the model to reasonable-looking values for wages and rental rates. With more difficulty we can approximate *joint* distributions of wealth-income pairs, for our individuals fall into only four main categories–"retired" or not, and accumulating or decumulating–and the "retired" decumulators are bound to be quite rare in any equilibrium. When age is taken into account, the model has very specific predictions for individuals' behaviour, and therefore cannot be made to fit arbitrary joint distributions of wealth, labour income, and age.

In the remainder of the paper, we lay out the structure of the model in section 2, derive some aspects of individuals' behaviour in section 3, and describe the equilibria that occur in the model, especially its steady states, in section 4. We develop a parametric example of the economy in section 5, and describe the computational methods we use in section 6. The general result of the paper, that general increases in patience cause general increases in wealth, income, and utility, is *not* a theorem of the model; but it is well-substantiated by computation of equilibria as discussed in section 7. Section 8 concludes, with the idea of *artificial patience*, meaning the enactment of policies that make agents behave as if they were generally more patient, and so generate the increases described in section 7. Appendices I and II contain, respectively, proofs of the theoretical results and a summary of the computational results.

2 Assumptions of the model

2.1 Individuals

Time is discrete and denoted by $t = 0, 1, ..., \infty$. The economy consists of individuals with preferences defined over random streams of consumption $c_t \geq 0$ and leisure $l_t \geq 0$ that are represented by a discounted expected utility function

$$U_t = \mathrm{E}_t \left[\sum_{j=0}^{\infty} \beta^j u(c_{t+j}, \ell_{t+j}) \right],$$

or equivalently by the recursive specification

$$U_t = u(c_t, \ell_t) + \beta \, \mathrm{E}_t \left[U_{t+1} \right].$$

The real–valued function u is strictly concave and strictly increasing in each argument. The discount factor, or patience parameter β varies across individuals, but is fixed through time for a given individual. Individuals are subject to a constant probability of death $0 < d < 1$ in each period. Upon death, an individual is replaced by a descendant that inherits the family fortune, but not necessarily the family preferences. Specifically, we assume that each descendant is endowed with a new patience parameter that is drawn from a fixed distribution function $G(\beta)$.

Consumption is bought with labour income and the proceeds of capital rental, or interest on savings. Physical capital k_t is accumulated or depleted according to

$$k_{t+1} = (1 - \delta) k_t + y_t - c_t,$$

in which

$$y_t = r_t k_t + w_t n_t$$

is the agent's total income, from the proceeds of capital rentals and from labour income. The rental rate r_t and the wage rate w_t are expressed in units of the consumption good, and δ is the rate of depreciation. Because of some (for us) insurmountable technical difficulties, we assume that individuals cannot accumulate physical capital beyond some finite bound $\bar{K} > 0$. Labour supply and leisure are the only uses for time in this model, so we restrict these amounts by

$$n_t + \ell_t = 1.$$

2.2 Economy-wide variables

Throughout this paper we assume that there is no aggregate uncertainty. Although there is some randomness in the future, from the agents' point of view, we assume that in the aggregate, exactly the proportion d of the agents will die at the end of period t. Moreover, we assume that these fractions of deaths hold uniformly across the set of possible ordered pairs (k, β). In later analysis, it helps to assume that these quantities are bounded from below by zero, and from above by $\bar{K}$ and $\bar{B}$ respectively; in particular we need $\bar{B}(1-\delta) < 1$ in the section to follow.

The economy is completely characterized by the distribution of these parameters across individuals; so let Φ denote this distribution or measure, on the set of individuals' possible characteristics

$$C = [0, \bar{K}] \times [0, \bar{B}].$$

We can normalize Φ so that $\Phi(C) = 1$; then, for intervals $[k_1, k_2]$ and $[\beta_1, \beta_2]$, $\Phi\{[k_1, k_2] \times [\beta_1, \beta_2]\}$ will be the proportion of individuals with physical capital stocks between k_1 and k_2, and patience parameters between β_1 and β_2. Alternatively, this number is the probability that a randomly chosen individual will have a parameter pair (k, β) lying in this rectangle.

We assume that there is an economy-wide production function $F : \mathbf{R}_+^2 \to \mathbf{R}_+$, defined on aggregate physical capital and aggregate labour. For simplicity we assume that F is twice continuously differentiable and strictly increasing on $\mathbf{R}_+^2$, is concave, and exhibits constant returns to scale.

Given a distribution Φ_t of agents' characteristics at time t, the aggregate amount of physical capital is easy to define, by

$$K_t = \int_C k d\Phi_t(k, \beta).$$

The aggregate amount of labour supplied is more problematic, since from an agent's point of view, labour supply depends on the endogenous wage rate w_t and rental rate r_t, as well as on the agent's state variables k and β. However, supposing that individuals' labour supplies are a well-defined function, say n^*, of k and β, an aggregate measure N_t is well-defined, by

$$N_t = \int_C n^*(k, \beta) d\Phi_t(k, \beta).$$

We assume that the labour and capital inputs are used by a price-taking profit maximizing firm, or equivalently by a sector of competitive firms, each with the same linearly homogeneous production function F. In either case aggregate output is

$$Y_t = f(K_t, N_t),$$

the competitive rental rate and wage rate are

$$r_t = f_K(K_t, N_t), \quad w_t = f_N(K_t, N_t),$$

and Euler's theorem then gives

$$Y_t = r_t K_t + w_t N_t.$$

3 Individual agents' behaviour

From the point of view of an individual agent, present and future factor prices will form a deterministic sequence. Therefore we begin by studying the behaviour of an agent with patience parameter β, endowed with capital k, and given deterministic sequences w and r of factor prices. Given such a pair of sequences

$$w = (w_0, w_1, \ldots, w_t, \ldots) \quad r = (r_0, r_1, \ldots, r_t, \ldots),$$

we use the notation

$$_t w = (w_t, w_{t+1}, \ldots), \quad _t r = (r_t, r_{t+1}, \ldots).$$

We seek a value function of these variables and streams of ws and rs, say V, which must satisfy

$$V(k, \beta, w, r) = \max \ u(c, \ell) + \beta(1-d)V(k', \beta, {}_1w, {}_1r)$$

$$\text{subject to}$$

$$k' = (1-\delta)k + r_0 k + w_0 n - c \leq \bar{K}, \quad n + \ell \leq 1,$$

and nonnegativity. Hence we can rewrite the recursive equation as

$$T(V)(k, \beta, w, r) = \max_{n,c} \ [u(c, 1-n)$$

$$+\beta(1-d)V((1-\delta)k + r_0 k + w_0 n - c, \beta, {}_1w, {}_1r)],$$

where the optimization is now subject only to nonnegativity, $n + \ell \leq 1$, and $(1-\delta)k + r_0 k + w_0 n - c \leq \bar{K}$. Standard arguments show that the operator $T : X \to X$ is contractive, with contraction parameter $\beta(1-d) < 1$; here X is the set of all real-valued continuous functions on $C \times \mathbf{R}_+^\infty \times \mathbf{R}_+^\infty$, and we endow X with the uniform (sup-norm) topology. The set X is a complete metric space, so there is a unique fixed point V^* for T. The basic properties of V^* are summarized in

Proposition 1: V^* is strictly increasing and strictly concave in k (with β, w, and r held fixed).

Moreover, agents' behaviour given deterministic w and r can be found by maximizing V^*, as follows.

Proposition 2: The function of n and c

$$u(c, 1-n)$$

$$+\beta(1-d)V^*((1-\delta)k + r_0 k + w_0 n - c, \beta, {}_1w, {}_1r)$$

is strictly concave. Consequently, the maximizers n^* and c^* are unique. These maximizing values, and hence

$$k^* = (1-\delta)k + r_0 k + w_0 n^* - c^* \quad \text{and} \quad \ell^* = 1 - n^*,$$

can therefore be written as continuous functions of (w, r, k, β).

We summarize some important properties of individuals' behaviour in

Proposition 3: The function V^* is differentiable with respect to k. The function $k^*(w, r, k, \beta)$ is (weakly) increasing in k and in β.

The following proposition is familiar:

Proposition 4: Fix a pair of factor prices w and r, and assume that $w_t = w$ and $r_t = r$ for all t. Then

$$k^*(k, \beta, w, r) \geq k$$

if and only if

$$\beta \geq \frac{1}{(1-\delta+r)(1-d)} = \beta^*.$$

4 Equilibrium analysis

4.1 Existence of steady states

We now restrict attention to constant sequences of factor prices (w, r), which we write into the value function as scalars. That is, we write

$$V^*(k, \beta, w, r) = \max_{n,c} \; u(c, 1-n)$$

$$+\beta(1-d)V^*((1-\delta)k + rk + wn - c, \beta, w, r)$$

with the understanding that factor prices are constant through time at w and r. We also view the policy functions n^* and c^*, and hence k^*, as continuous functions of the four scalar variables k, β, w, and r.

Our immediate goal is to analyze the sequence of distributions on C, the space of agents' characteristics, that results from agents' behaviour and our assumptions on bequests, without worrying about whether or not the aggregate behaviour is consistent with the marginal product equations

$$r = f_K(K_t, N_t), \quad w = f_N(K_t, N_t).$$

For a fixed pair of factor prices, equilibrium or otherwise, and an initial distribution Φ_0 on C, we generate the sequence of distributions $\{\Phi_t\}$ that would follow from our agents' behaviour. That is, for $t \geq 0$, we define the distribution Φ_{t+1} on C by specifying it on products of intervals, say

$$I_k \times I_\beta = [k_1, k_2] \times [\beta_1, \beta_2],$$

and then extending it to the Borel subsets of C. On the intervals, we have

$$\begin{aligned} &\Phi_{t+1}\{I_k \times I_\beta\} \\ &= (1-d)\Phi_t\left\{(k,\beta) \middle| k^*(k,\beta,w,r) \in I_k, \beta \in I_\beta\right\} \\ &\quad + d\Phi_t\left\{(k,\beta) \middle| k^*(k,\beta,w,r) \in I_k\right\} \times G(I_\beta). \end{aligned} \tag{1}$$

The first term in the sum is the fraction of people who had patience parameters in $[\beta_1, \beta_2]$, did not die, and found it optimal to accumulate a capital stock in the range $[k_1, k_2]$. In the second term we count the people who died and left behind capital in the range $[k_1, k_2]$. Of these peoples' descendants, the fraction $G([\beta_1, \beta_2])$ will have βs in the required ranges.

In what follows, we assume that the initial marginal distribution of β is just the distribution G; therefore, the marginal distribution of β will continue to be G in all future time periods. According to Proposition 4, if factor prices are fixed, then in order to get nondegenerate equilibrium wealth distributions, we need to assume that the patience distribution G is not concentrated exclusively on the "accumulation range" $[\beta^*, 1)$, or on the "decumulation range" $[0, \beta^*]$. For example:

Assumption N1: The distribution G satisfies

$$G([0, \beta^*)) > 0 \quad \text{and} \quad G((\beta^*, \bar{B}]) > 0.$$

Theorem 1: Fix w and r, and assume Assumption N1. Then there is a unique distribution Φ_* on C such that for any initial distribution Φ_0 on C, the sequence Φ_t (1) converges weakly to Φ_*.

Up to now we have held factor prices fixed, and have not made any requirement that the resulting steady state factor supplies

$$K^*(w,r) = \int_C k\Phi_*(dk, d\beta|w,r) = \int_C k^*(k,\beta,w,r)\Phi_*(dk, d\beta|w,r)$$

and

$$N^*(w,r) = \int_C n^*(k,\beta,w,r)\Phi_*(dk,d\beta|w,r)$$

actually solve

$$F_K(K^*(w,r),N^*(w,r)) = r, \quad F_N(K^*(w,r),N^*(w,r)) = w.$$

However, these factor supplies are bounded, by

$$0 \leq K^*(w,r) \leq \int_C \bar{K}\, d\Phi_t(k,\beta) = \bar{K}, \quad 0 \leq N^*(w,r) \leq 1 = \bar{N},$$

and marginal products are assumed to be continuous functions of K and N, so we can restrict w and r by

$$0 \leq r \leq \max\ f_K(K,N) = \bar{R}, \quad 0 \leq w \leq \max\ f_N(K,N) = \bar{W},$$

where in both maximizations (K,N) is restricted to the compact set $[0,\bar{K}] \times [0,\bar{N}]$. Then we can specify a function

$$p : [0,\bar{W}] \times [0,\bar{R}] \to [0,\bar{W}] \times [0,\bar{R}]$$

by setting

$$p(w,r) = \left(f_K(K^*(w,r),N^*(w,r)), f_N(K^*(w,r),N^*(w,r))\right),$$

and hope to find equilibria as fixed points of p. In view of Brouwer's theorem, and the assumed continuity of f_K and f_N, it is enough to establish that the functions $K^*(w,r)$ and $N^*(w,r)$ are continuous, in order to prove:

Theorem 2: There exists at least one steady state $(\bar{w},\bar{r})$.

4.2 Uniqueness

We intend to study the effects of changing various policy parameters on the resulting steady state; however, it makes little sense to speak of such causes and effects unless the equilibria are always unique. Moreover, for computational reasons it is useful to know that a unique steady state always exists. However, we are able to rule out multiple equilbria only by invoking some behavioural assumptions, rather than imposing conditions directly on the data of the economy.

In a steady state we have

$$K^*(w,r) = \int_C k^*(k,\beta,w,r)\Phi_*(dk,d\beta|w,r)$$

$$= \int_C [(1-\delta+r)k + wn^*(k,\beta,w,r) - c^*(k,\beta,w,r)]\Phi_*(dk,d\beta|w,r)$$

$$= (1 - \delta + r)K^*(w,r) + wN^*(w,r) - C^*(w,r),$$

with

$$C^*(w,r) = \int_C c^*(k,\beta,w,r)\Phi_*(dk,d\beta|w,r).$$

It follows that

$$C^*(w,r) + \delta K^*(w,r) = rK^*(w,r) + wN^*(w,r) = Y^*(w,r),$$

say, which is the steady state amount of output that individuals will buy, if possible.

We would like to rule out the possibility of two different equilibria, say (w,r) and (w',r'); so suppose temporarily that these quantities exist. Because of the constant returns to scale assumption, any equilibrium pair of factor prices (w,r) must allow the competitive sector to produce at the normalized unit cost of output. Thus if

$$c^*(w,r,Y) = \min[wN + rK] \quad \text{subject to} \quad F(K,N) = Y,$$

then constant returns to scale in production implies that

$$c^*(w,r,Y) = Yc^*(w,r,1) = Yc(w,r);$$

and any pair of equilibria (w,r) and (w',r') must satisfy $c(w,r) = c(w',r') = 1$. Since c is increasing in both arguments, we must have (without loss of generality)

$$r' > r \quad \text{and} \quad w' < w.$$

Now consider the effects of increasing the rental rate from r to r' and then decreasing the wage from w to w'. In production, the capital-labour ratio of the factor demands will necessarily change according to

$$\frac{K'}{N'} < \frac{K}{N}.$$

But from the individuals' point of view, the resulting change in factor supplies is less predictable. If substitution effects dominated income effects, then we could conclude that $K^*(w',r') > K^*(w,r)$ and $N^*(w',r') < N^*(w,r)$; and then the purported pair of equilibria could not exist, and we would have a proof of uniqueness of steady states. But none of the required comparative statics derivatives is signable (see [1]); so we merely assume enough here to guarantee the result we want, and state the resulting uniqueness theorem:

Assumption N2: Steady state factor supplies satisfy:

$$\text{If } \frac{w'}{r'} < \frac{w}{r}, \text{ then } \frac{K^*(w',r')}{N^*(w',r')} \leq \frac{K^*(w,r)}{N^*(w,r)}.$$

Theorem 3: Under assumption N2, there is a unique steady state factor price pair $(\bar{w}, \bar{r})$. There is a unique distribution Φ_* on C, such that for any initial distribution Φ_0 on C, the sequence (1) converges weakly to Φ_*.

In numerical examples and simulations, we can verify assumption N2 only at finitely many pairs of factor prices; but it is reassuring that with natural specifications for utilities and production functions, no counterexamples to the assumption have yet been found. However, with some effort and ingenuity one could construct distinct pairs of equilibrium factor prices, or even a continuum of such pairs. All that is required is a utility function and a pair of factor price pairs (w, r) and (w', r') violating assumption N2; for then one can build a production function whose unit isocost passes through both factor price pairs, and whose partial derivatives at those points are

$$c_w(w, r) = N^*(w, r) \quad c_r(w, r) = K^*(w, r)$$
$$c_w(w', r') = N^*(w', r') \quad c_r(w', r') = K^*(w', r').$$

4.3 Dynamic equilibria

Our primary goal is to explain income distribution, rather than its behaviour over time, as a function of human preferences, technology, and policy parameters. Therefore our main notion of equilibrium, and the endogenous object whose sensitivities we will try to estimate, is the unique steady state distribution of characteristics and incomes. It seems only slightly hypocritical to argue that dynamics matter, in the sense that intertemporal preference parameters determine the final outcome of interest, but dynamics *don't* matter, in the sense that transitions to steady states are going to be ignored.

Our dynamic equilibrium notion takes as given an initial distribution Φ_0 on C, and then requires an entire deterministic sequence of (w, r) pairs, and a sequence of distributions Φ_1, Φ_2, ..., with the property that: If agents are distributed according to Φ_t, and optimize with respect to future ws and rs, then the distribution Φ_{t+1} will follow, and the ws, rs, and aggregate inputs will be consistent with each other.

Now suppose that in addition to a pair of sequences w and r, we are also given a sequence of distributions

$$\Phi = (\Phi_0, \Phi_1, \ldots, \Phi_t, \ldots).$$

Stack these objects into a sequence of triples

$$S = (S_0, S_1, \ldots, S_t, \ldots),$$

where

$$S_t = (w_t, r_t, \Phi_t),$$

and write ${}_t\Phi = (\Phi_t, \Phi_{t+1}, \ldots)$. We can now seek equilibria as fixed points, of a function to be specified shortly, in the set of all possible sequences S. The function must give time t aggregate capital stocks in terms of the distribution Φ_t, and aggregate labour supplies as functions of Φ_t and optimizing behaviour with respect to ws and rs; it also has to specify next-period distributions in terms of current distributions, optimizing behaviour, and random deaths. Specifically, we map a sequence S into another sequence

$$g(S) = S' = \{(w'_t, r'_t, \Phi'_t) | \, t \geq 0\}$$

by letting

$$K_t = \int_C k \, d\Phi_t(k, \beta), \quad N_t = \int_C n^*({}_t w, {}_t r, k, \beta) \, d\Phi_t(k, \beta)$$

and then

$$r'_t = f_K(K_t, N_t) \qquad w'_t = f_N(K_t, N_t).$$

We define $\Phi'_0 = \Phi_0$; but then we need to introduce some notation to define Φ'_{t+1} from Φ_t and agents' behaviour. Let

$$k^*(k, \beta, {}_t w, {}_t r) = (1 - \delta)k + r_t k + w_t n^*({}_t w, {}_t r, k, \beta) - c^*(k, \beta, {}_t w, {}_t r),$$

the capital stock accumulated by an agent with characteristics (k, β), who faces current and future prices ${}_t w$ and ${}_t r$. To define a distribution on C it is enough to specify it on rectangles, or sets of the form

$$I_k \times I_\beta = [k_1, k_2] \times [\beta_1, \beta_2].$$

The distribution we need is

$$\begin{aligned} &\Phi'_{t+1}\{[k_1, k_2] \times [\beta_1, \beta_2]\} \\ &= (1 - d)\Phi_t \{(k, \beta) | \, k^*(k, \beta, {}_t w, {}_t r) \in [k_1, k_2], \beta \in [\beta_1, \beta_2]\} \\ &\quad + d\Phi_t \{(k, \beta) | \, k^*(k, \beta, {}_t w, {}_t r) \in [k_1, k_2]\} \times G([\beta_1, \beta_2]). \end{aligned}$$

We can now state an equilibrium existence theorem. In what follows, the notation $\mathcal{D}(C)$ refers to the set of all probability measures on the Borel sets in the characteristics space C, and

$$\mathcal{S} = \{S | \, S_t = (w_t, r_t, \Phi_t)\},$$

the set of all sequences of the type discussed above. Since aggregate capital and efficiency units of labour are bounded by

$$0 \leq K_t \leq \int_C \bar{K} d\Phi_t(k, \beta) = \bar{K}, \quad 0 \leq N_t \leq 1 = \bar{N},$$

and marginal products are assumed to be continuous functions of K and N, we can restrict w_t and r_t to lie in $[0, \bar{W}] \times [0, \bar{R}]$, as in the last two subsections.

A topological vector space is a vector space X for which the operations $(x, y) \to x + y$ and $(\alpha, x) \to \alpha x$ are continuous mappings, say $a : X \times X \to X$ and $m : \mathbf{R} \times X \to X$. It is locally convex if for every element x of every open set O, there is a convex open set U with $x \in U \subset O$.

Theorem 4: Endow C with its usual Euclidean topology, $\mathcal{D}(C)$ with the closed convergence topology, and $\mathcal{S}$ with the product topology. Then $g : \mathcal{S} \to \mathcal{S}$ is continuous. The set $\mathcal{S}$ is a compact convex subset of a locally convex topological vector space, so the Fan-Glicksberg Theorem applies and g has a fixed point.

A fixed point for g is an entire sequence of triples (w_t, r_t, Φ_t) that is dynamically consistent with perfect foresight optimizing behaviour and market clearing. An example of such an equilibrium would be a constant sequence, beginning in one of the steady states discussed already, and staying there. But if the initial state S_0 is not a steady state, then we have no guarantee that the sequence of states converges to anything in particular. All we can offer is computational evidence for the following speculation:

Conjecture: Assume assumption N2, and let $\Phi_0 \in \mathcal{D}(C)$. Then there is a unique sequence S for which $g(S) = S$ and $S_0 = (w_0, r_0, \Phi_0)$, for some factor price pair (w_0, r_0). This dynamic equilibrium sequence satisfies $w_t \to \bar{w}$ and $r_t \to \bar{r}$, for the unique steady state factor price pair $(\bar{w}, \bar{r})$; and Φ_t converges weakly to Φ_*, the steady state characteristics distribution.

In various simulations we explore the possibilities of changing the equilibrium distribution Φ_* by manipulating some policy parameters that affect the capital and labour markets. In each case we assume that the economy is initially in a steady state, followed by a once-and-for all policy change, after which the economy approaches a new steady state along an equilibrium path of the type described in the Conjecture.

5 A Parametric Example

We now specify an atemporal utility function, a production function, and a class of distributions of discount factors that will allow numerical computation of equilibria and eventually, we hope, shed some light on the likely effects of some policy changes that are the subject of the last section of the paper. We parameterize the atemporal utility function first, via

$$u(c, \ell) = c^{\alpha} \ell^{\gamma},$$

with $\alpha > 0$, $\gamma > 0$, and $\alpha+\gamma < 1$; and before any analysis of production or patience distributions, we work out some of the details of optimal behaviour with this utility function, a fixed discount factor β, and fixed factor prices w and r.

We can maximize

$$\sum_{t=0}^{\infty} \beta^t u(c_t, \ell_t)$$

subject to

$$k_{t+1} = Rk_t + wn_t - c_t, \tag{2}$$

and

$$k_t \geq 0,\ 0 \leq n_t \leq T,\ \text{and}\ \ \ell_t + n_t = T \quad \text{for all}\ t, \tag{3}$$

by separating the temporal from the atemporal, as follows. First, we suppose that at some time the agent has some fixed amount Y to spend optimally on consumption and leisure. The indirect utility calculation is

$$\max\ c^\alpha \ell^\gamma \quad \text{subject to} \quad c + w\ell = Y, \quad c \geq 0,\ 0 \leq \ell \leq T,$$

leading to the choices

$$\ell = \frac{\gamma Y}{w(\alpha+\gamma)} \quad c = Y - w\ell = \frac{\alpha Y}{\alpha+\gamma},$$

provided that this choice of ℓ is feasible. If $\frac{\gamma Y}{w(\alpha+\gamma)} > T$, then we get

$$\ell = T \quad c = Y - wT.$$

The indirect utility function is thus

$$U(Y) = \begin{cases} Cw^{-\gamma} Y^{\alpha+\gamma} & \text{if}\ Y \leq \frac{wT(\alpha+\gamma)}{\gamma} \\ (Y - wT)^\alpha T^\gamma & \text{if}\ Y \geq \frac{wT(\alpha+\gamma)}{\gamma} \end{cases}, \tag{4}$$

with

$$C = \frac{\alpha^\alpha \gamma^\gamma}{(\alpha+\gamma)^{\alpha+\gamma}}.$$

In (4), the two functions of Y are both increasing and concave, and agree at the point $Y = \frac{wT(\alpha+\gamma)}{\gamma}$. Moreover, the first function lies above the second, for it is the value of the maximization

$$\max\ c^\alpha \ell^\gamma \quad \text{subject to} \quad c + w\ell = Y,$$

which is less constrained than

$$\max\ c^\alpha \ell^\gamma \quad \text{subject to} \quad c + w\ell = Y, \quad 0 \leq \ell \leq T.$$

Consequently, the two functions of Y have the same derivative at $Y = \frac{wT(\alpha+\gamma)}{\gamma}$, and U is therefore differentiable.

Our agent is implicitly endowed with the amount wT of labour income in each period, and has opportunities to accumulate or deplete a stock of capital. But within each time period, the agents optimization problem is already solved, once we know how much Y is in the above problem. Thus the original maximization can be rewritten as

$$\begin{aligned} \max \quad & \sum_{t=0}^{\infty} \beta^t U(Y_t) \\ \text{subject to} \quad & Y_t = wT - s_t \geq 0 \\ & k_{t+1} = Rk_t + s_t \\ & k_t \geq 0, \quad k_0 \text{ given}. \end{aligned} \tag{5}$$

Here s_t is the amount of savings (positive or negative) that is carried over into the next period's capital stock. The problem is now relatively standard, and has *interior* solutions characterized by

$$U'(Y_t) = \beta R U'(Y_{t+1}),$$

or

$$Y_{t+1} = \begin{cases} Y_t (R\beta)^{\frac{1}{1-\alpha-\gamma}} \\ wT + (Y_t - wT)(R\beta)^{\frac{1}{1-\alpha}} \end{cases}$$

$$= \begin{cases} Y_t G_1 & \text{if } Y_t, Y_{t+1} \leq \frac{wT(\alpha+\gamma)}{\gamma} = Y^* \\ wT + (Y_t - wT)G_2 & \text{if } Y_t, Y_{t+1} \geq Y^* \end{cases}$$

and a more complicated formula, if Y_t and Y_{t+1} are on opposite sides of Y^*. The condition

$$\beta R^\alpha = \left(\frac{G_2}{R}\right)^{1-\alpha} < 1,$$

or equivalently

$$G_2 < R,$$

is necessary for the maximization problem to have a solution.

We summarize the relevant properties of optimal behaviour, and hence computational aids, as follows:

Proposition 5C: If $R\beta = 1$, then optimal behaviour is constant, and consists of setting

$$s_t = s_0 = (1-R)k_0, \quad Y_t = Y_0 = wT + (R-1)k_0,$$

and getting

$$V(k_0) = \frac{1}{1-\beta} \begin{cases} Cw^{-\gamma}(wT + (1-R)k_0)^{\alpha+\gamma} & k_0 \leq \frac{wT\alpha}{\gamma(R-1)} \\ ((R-1)k_0)^\alpha T^\gamma & k_0 \geq \frac{wT\alpha}{\gamma(R-1)} \end{cases}.$$

If $R\beta > 1$, then it is optimal to accumulate. We distinguish between "accumulating retirees" and "accumulating workers" by the amount of capital they begin with, as follows:

Proposition 5AR: If

$$k_0 \geq k_0^* = \frac{wTR\alpha}{(R - G_2)\gamma},$$

then the individual sets

$$\ell_t = T, c_t = G_2^t(1 - \frac{G_2}{R})k_0, \quad k_t = G_2^t k_0,$$

and gets

$$V(k_0) = DT^\gamma k_0^\alpha,$$

with

$$D = (1 - \frac{G_2}{R})^{\alpha - 1}.$$

Proposition 5AW: Define

$$k_{-N}^* = \frac{wT}{R^{N-1}} \left[\left(\frac{G_1^N - R^N}{G_1 - R} \right) \frac{(\beta + \gamma)}{\gamma G_1^N} + \frac{\beta}{(R - G_2)\gamma} - \left(\frac{1 - R^N}{1 - R} \right) \right],$$

or

$$k_{-N}^* = \frac{wT}{(N+1)R^N} \left[\frac{R\beta}{(R - G_2)\gamma} + \frac{N(\beta + \gamma)}{\gamma} - R \left(\frac{1 - R^N}{1 - R} \right) \right]$$

in the case $R = G_1$. If $k_0 = k_{-N}^* \geq 0$, then it is optimal to set

$$Y_t = G_1^{-N+t} Y^*,$$

and

$$k_t = k_{-N+t}^* \text{ for } 0 \leq t \leq N.$$

In general we have

$$k_0^* > k_{-1}^* > \ldots > k_{-\bar{N}}^* > 0 > k_{-(\bar{N}+1)}^*$$

for some $\bar{N}$. If

$$k_{-N}^* < k_0 < k_{-N+1}^*$$

then in an optimal path

$$k_{-N+t}^* < k_t < k_{-N+t+1}^*$$

and

$$Y_t = G_1^t Y_0 \quad \text{for } 0 \leq t < N.$$

If $R\beta < 1$, then it is optimal to decumulate; and again there is a qualitative difference between those with initial capital above some critical point, and those with less. That is, those with enough initial capital will not work until they are poor enough, and then will work longer hours until they have depleted all their capital.

Proposition 5D: If $k_0 = 0$, then the individual sets

$$Y_t = wT,\ \ell_t = \frac{\gamma T}{\alpha+\gamma}, c_t = \frac{\alpha w T}{\alpha+\gamma},\ k_t = 0$$

for all t, and gets

$$V(0) = \frac{1}{1-\beta} C T^{\alpha+\gamma} w^{\alpha}.$$

Now define

$$k_N^* = \frac{wT(1-G_1)}{G_1}\left[\frac{1}{(1-c)(1-d)} + \frac{c^N}{(d-c)(1-d)} - \frac{d^N}{(d-c)(1-c)}\right],$$

where

$$c = \frac{1}{G_1}, \quad d = \frac{1}{R}.$$

Assume that N is small enough so that

$$G_1^N \geq \frac{\gamma}{\alpha+\gamma}.$$

If

$$k_0 = k_N^*,$$

then in an optimal path

$$k_t = k_{N-t}^*, \quad Y_t = wTG_1^{-N+t},$$

$$c_t = \frac{\alpha w T G_1^{-N+t}}{\alpha+\gamma},$$

and

$$\ell_t = \frac{\gamma T G_1^{-N+t}}{\alpha+\gamma},$$

for $0 \leq t \leq N$. For $t \geq N$, $k_t = 0$ is optimal, and the individual reverts to the behaviour already described. In general

$$0 = k_0^* < k_1^* < \ldots < k_{\bar{N}}^*$$

for some $\bar{N}$, defined to be the largest N for which $G_1^N \geq \frac{\gamma}{\alpha+\gamma}$. If $k_0 > k_{\bar{N}}^*$, then optimal behaviour consists of setting $\ell_t = T$ for $0 \leq t \leq T$ for some T, setting

$$k_{\bar{N}-j}^* \leq k_{T+j} \leq k_{\bar{N}-j+1}^*$$

for $1 \leq j \leq \bar{N}$, and setting $k_t = 0$ for $t > T + \bar{N}$.

6 Computation

The preceding results allow for exact calculation of the value function at certain points, namely the amounts k^*_{-N} and k^*_N. For this reason, the iterative algorithm

$$V^{K+1}(k,\beta) = \max_{n,c} u(c, T-n) + \beta(1-d)V^K(Rk + wn - c, \beta)$$

can be started with an especially good trial guess V^0, consisting of linear interpolations between the points where V is known exactly. In calculating equilibria we take as given a factor price pair (w, r) and find the corresponding value function and policy functions. Only then do we postulate a distribution G for the patience parameter. For the fixed factor prices we simply iterate the mapping (1), starting with an initial conditional distribution of capital that is increasing in β. When the iterates of the mapping seem to have converged, we then finally calculate a production function whose partials match the given factor prices at the input values $K^*(w, r)$ and $N^*(w, r)$. Thus calculating an *initial* equilibrium does not entail searching for a fixed point of the mapping p; we can set w and r to plausible values in advance, and fit a production function to those values last of all. By adjusting the units of measurement, we can even ensure that in an initial equilibrium, $N^*(w, r)$ also has a plausible value; but because $K^*(w, r)$ has the same units as output, the capital-output ratio $K^*/F(K^*, N^*)$ will be meaningful, and not independently variable in the same sense.

In this paper, the comparisons we would like to make all involve shifting the distribution G to a distribution G' that is generally more patient, in the sense that $G' \succeq G$. This time we have to repeat the exercise of calculating an initial equilibrium, all the way back to the calculation of the value functions, each time we postulate a new factor price pair and check for the closeness of $p(w, r)$ to (w, r). Various pieces of information enable a reasonably efficient search procedure.

When we postulate a new factor price pair, close to a previous one, we can begin our iterations toward the value function $V^*(k, \beta, w', r')$ by

$$V^0(k, \beta, w', r')$$

$$= V^0(k, \beta, w', r') + V^*(k, \beta, w, r) - V^0(k, \beta, w, r).$$

Here $V^0(k, \beta, w', r')$ is the initial value function that is known at the points $k^*_{\pm N}$, and interpolated between them; $V^0(k, \beta, w, r)$ is the same object with the earlier factor prices; and $V^*(k, \beta, w, r)$ is an adequate approximation to the fixed point of the value function mapping. Thus the previously calculated V^* can be used as a sort of control variate. In early experiments leading toward a new equilibrium w and r, we can afford to be rather cavalier about the exactitude of $V^*(k, \beta, w', r')$; for all we really need to know about the factor prices is whether w'/r' is too high or too low, given the new patience distribution.

Having calculated $V^*(k, \beta, w', r')$, at least accurately enough to believe that it will give the right information on w'/r', we can start our iterations toward a new Φ_* by beginning with the conditional distribution of capital implicit in the most recent Φ_*. Thus

$$\Phi_0(k_i, \beta_j, w', r') = G'(\beta_j)\Phi_*(k_i|\beta_j, w, r)$$

$$= G'(\beta_j)\Phi_*(k_i, \beta_j, w, r)/G(\beta_j),$$

in a discretized version of these distributions. In the second and later trials to determine the new equilibrium w and r, we use another control variate approximation; we predict that Φ_0 will be wrong in about the same way as it was in previous trials, and start with

$$\tilde{\Phi}_0(k_i, \beta_j, w^{n+1}, r^{n+1}) = G'(\beta_j)\Phi_*(k_i|\beta_j, w^n, r^n)$$

$$+\alpha[\Phi_*(k_i, \beta_j, w^n, r^n) - \Phi_0(k_i, \beta_j, w^n, r^n)]$$

with the fraction α chosen by trial and error. The results of these computations are summarized in the second Appendix.

7 Patience and the distribution of wealth

In our model, a more patient individual will save more and work more, for any given level of capital, than a less patient individual; that is, both $k^*(k, \beta, w, r)$ and $n^*(k, \beta, w, r)$ are weakly increasing in β. If the more patient person is an accumulator and lives long enough, then in general that person will retire after accumulating enough wealth, but continue to accumulate; and if the less patient person is a decumulator and lives long enough, then this person will eventually have no capital and consume out of only labour earnings. Thus the *long-run* value of n^* is higher for low values of β than it is for high values of β, even though in the short run the opposite effects occur. Finally, for any given initial capital stock and age, the more patient person is wealthier.

What is true for individuals is not always true for societies, but in our parametric model it does seem to follow that more patient societies are wealthier than less patient societies. Furthermore, the effects of a general increase in patience are progressive, in the sense that for a given low level of patience, a person is better off being born into the more patient society than into the less patient society. The effect is most pronounced for the poorest and least patient individuals; for a person with no wealth and no inclination to accumulate any ranks steady state societies according to their wage levels only.

The long-run effect of an increase in patience is especially favourable to the least patient individuals because of a strange sort of externality. When patient individuals accumulate more capital, they increase the marginal product of labour in two ways. First, there is the a direct effect on F_N, owing to the fact that $F_{NK} > 0$ for a concave constant returns to scale production function with two inputs (differentiate the Euler theorem equation with respect to N). And second, patient individuals who live long enough eventually retire, causing a reduction in labour supply and another increase in F_N, this time owing to the fact that $F_{NN} < 0$. Even though more patient individuals work more, for given levels of capital, simulations suggest that over an average lifetime they work less.

We summarize the computational results in

Proposition 6: Let G_0 and G_1 be two probability distributions of discount factors, with $G_1 \succeq G_0$, and let (w_1, r_1) and (w_0, r_0) be the corresponding equilibrium factor price pairs. Then $w_1 \geq w_0$ and $r_1 \leq r_0$. The equilibrium marginal distributions of capital, say M_1 and M_0, satisfy $M_1 \succeq M_0$. The equilibrium distributions of *atemporal* utility, say U_1 and U_0, satisfy $U_1 \succeq U_0$. For any fixed β, the distribution of $V(k, \beta, w_1, r_1)$ in the G_1 equilibrium dominates the distribution of $V(k, \beta, w_0, r_0)$ in the G_0 equilibrium.

Note that there is no claim here, even if it is true, that the individuals' value functions $V(k, \beta, w, r)$ have a dominating distribution in the more patient equilibrium. Such a claim would involve comparisons between utility levels of individuals with fundamentally different preferences. In general, the value function is increasing in β; but we cannot attach any positive or normative significance to this fact.

8 Conclusion

This last result suggests several avenues of further research, all based on the idea of *artificial patience*, in which societies enact policies designed to make individuals behave as if they were more patient than they really are.

For example, a social planner could force individuals to save more than they would choose, by taxing labour or investment income and investing the proceeds in the capital market. Such a tax could be made roughly equivalent to a government pension plan financed by payroll deductions; and the effect presumably would be to create a somewhat wealthier society, at least with fixed factor prices. It would not lead to a Pareto improvement, of course, for from the *impatient* individuals' point of view, the government is reducing current consumption and increasing future consumption, at rates that the individuals could have chosen if they wished. From the point of view of sufficiently *patient* individuals, these investments on their behalf would be perfect substitues for their own investments, and would just crowd them out with no effect on these individuals' welfare. Any positive welfare effect of the policy would have to rely on a general equlibrium adjustment in factor prices–in this case, in favour of labour at the expense of capital. However, once the policy had been in effect for a long time, the short-term sacrifices of the impatient dead would allow for a wealthier society in the long run, with or without fixed factor prices.

As another example, notice that the effect of lowering the death probability d is exactly the same as the effect of shifting the entire distribution of β upward, with the resulting effects detailed in the Proposition. We could interpret such a change as the result of improved health care, medical advances, more stringent safety regulations, a less belligerent and warlike foreign policy, clearance of landmines, or a reduction in violent crime rates. In any of these scenarios there is the potential for a government, hitherto nonexistent in the model, to encourage or enforce the reduction in d. The required policies might be costly, but could be financed in the long run by the resulting increase in wealth.

One can also speculate that differential taxation rates on labour and investment income could be engineered so as to create a more patient, more wealthy society, at least in the long run. Suppose we calculate an equilibrium in which all income is taxed at the rate τ, and then change the rates so that capital rentals and labour income are taxed at the rates $\tau_K < \tau < \tau_N$. The immediate effect is to change factor prices faced by individuals from w and r to $(1-\tau_N)w$ and $(1-\tau_K)r$ so that, among other things, the new value of β^* will be lower. There will be some individuals who used to be decumulators, who will become accumulators; and in general all individuals will act as if they had become more patient. In the long run there will be more capital and wealth than before, and the new equilibrium will be dominance comparable to the old, in the senses of the Proposition.

Further possibilities arise when we allow individuals to invest in human capital, with or without a market in physical capital. In the case of human capital, it seems reasonable to have individuals start their lives without any initial bequests from immediate ancestors. Also, the effect of human capital should be to increase an individual's wage, through an efficiency-unit-of-labour function; it could also improve an individual's ability to enjoy leisure. In either case, the more patient individuals will become wealthier in both physical and human capital, and general increases in patience, artificial or not, will lead to general increases in wealth of all forms. In such a model, favourable tax treatment of capital investment could take the form of public expenditure on education, or perhaps subsidies for on-the-job training; and "forced saving" could arise from a policy of mandatory school attendance. One of the two progressive externality effects would be missing in this scenario, for the accumulation of human capital would presumably have a negative effect on the marginal product of "unskilled labour", supplied by impatient individuals with minimal human capital.

We also think we can explain the simultaneous existence of unemployment and "skills shortages" in this human capital framework, even if all individuals are exactly the same except for their discount factors. (Of course, to have unemployment at all we will have to introduce some form of nonlabour income into our model). It could happen that despite extremely high wages in some occupations, the impatient individuals will prefer not to make the required investments in human capital. A long-run self-financing dominance improvement could arise from the right type of artificial patience.

9 Appendix I

Proof of Proposition 1: For easy reference, here is the transformation T again:

$$T(V)(k, \beta, w, r) = \max_{n,c} \; [u(c, 1-n)$$

$$+\beta(1-d)V((1-\delta)k + r_0 k + w_0 n - c, \beta, {}_1w, {}_1r)$$

First, look at what the transformation T does to an initial guess V for the value function. We show that T maps a weakly increasing, weakly concave function into another function with the same properties. Since the set X' of all Vs with these (weak) properties is a closed set in X, it is also a complete metric space; and since T is contractive on X', the fixed point for T must lie in X'. But then, if we can show that T maps weakly increasing, weakly concave functions into strictly increasing, strictly concave functions, then it will follow that the fixed point for T must have these strict properties as well.

It is fairly easy to show that T maps weakly increasing functions into strictly increasing functions. Therefore, we prove only that T maps strictly increasing, weakly concave Vs into strictly concave $T(V)$s. The proposition then follows.

So assume that V is strictly increasing and weakly concave, let k_1 and k_2 be given, and let n_j and c_j solve the respective maximization problems. (These solutions must exist because the right side is a continuous function of these variables, and the constraint set is compact because $0 \leq c \leq (1-\delta)k + r_0 k + w_0 n \leq (1-\delta+r_0)\bar{K} + w_0$). Given the convex combination $\alpha k_1 + (1-\alpha)k_2 = k_\alpha$ of capital, it is easy to show that the convex combinations

$$(n_\alpha, c_\alpha) = \alpha(n_1, c_1) + (1-\alpha)(n_2, c_2)$$

are feasible in the maximization problem. Also, if we let

$$k'_j = (1-\delta)k_j + r_0 k_j + w_0 n_j - c_j,$$

then the labour-consumption pair n_α, c_α gives rise to the future capital stock

$$k'_\alpha = \alpha k'_1 + (1-\alpha)k'_2.$$

Therefore

$$T(V)(k_\alpha, \beta, w, r)$$

$$\geq u(c_\alpha, 1-n_\alpha)$$

$$+\beta(1-d)V((1-\delta)k'_\alpha + r_0 k'_\alpha + w_0 n_\alpha - c_\alpha, \beta, {}_1w, {}_1r).$$

Since u and V are concave, this quantity is

$$\geq \alpha u(c_1, 1-n_1) + (1-\alpha)u(c_2, 1-n_2)$$

$$+\beta(1-d)\left[\alpha V((1-\delta)k_1 + r_0 k_1 + w_0 n_1 - c_1, \beta, {}_1w, {}_1r)\right.$$

$$\left.+ (1-\alpha)V((1-\delta)k_2 + r_0 k_2 + w_0 n_2 - c_2, \beta, {}_1w, {}_1r)\right]$$

$$= \alpha V(k_1, \beta, w, r) + (1-\alpha)V(k_2, \beta, w, r).$$

So far we have established only that the fixed point V^* for T is weakly concave; now we have to assume that $0 < \alpha < 1$ and $k_1 \neq k_2$, and consider various cases to show that strict concavity follows. First, if any of $n_1 \neq n_2$ or $c_1 \neq c_2$ is true, then

$$u(c_\alpha, 1-n_\alpha) > \alpha u(c_1, 1-n_1) + (1-\alpha)u(c_2, 1-n_2)$$

follows from strict concavity of u, the most recent weak inequality is strict, and we are finished. So suppose that even though $k_1 \neq k_2$, there are equal solutions n_j and c_j to the two maximization problems. If $k_1 > k_2$, then $n_1 = n_2$ and $c_1 = c_2$ cannot both be true; in particular some $c_1' > c_1$ will increase the maximand.

Proof of Proposition 2: Strict concavity of

$$f(n, c) = u(c, 1-n)$$

$$+\beta(1-d)V^*((1-\delta)k + r_0 k + w_0 n - c, \beta, {}_1w, {}_1r)$$

is easy to show. The problem of maximizing f always has a solution because the constraint set is compact. Berge's theorem (e.g., [4], p. 62) guarantees continuity of these maximizers as functions of k and the price sequences.

Proof of Proposition 3: Fix any $k_0 > 0$ and let $c_0 > 0$ and n_0 maximize $f(n, c)$. Then

$$V^*(k_0, \beta, w, r) = u(c_0, 1-n_0)$$

$$+\beta(1-d)V^*((1-\delta+r_0)k_0 + w_0 n_0 - c_0, \beta, {}_1w, {}_1r).$$

Moreover, for k near k_0 we have

$$V^*(k, \beta, w, r) \geq W(k)$$

$$= u(c_0 + k - k_0, 1-n_0)$$

$$+\beta(1-d)V^*((1-\delta+r_0)k_0 + w_0 n_0 - c_0, \beta, {}_1w, {}_1r),$$

for W is the value of simply consuming any extra capital at time 0, or decreasing time 0 consumption if $k < k_0$. Obviously W is a concave function of k, W is differentiable at k_0, and

$$V^*(k_0, \beta, w, r) = W(k_0).$$

Thus V^* and W fit the hypotheses of [2] (see [6], p. 85), and V^* is differentiable at k_0. It is possible that $c_0 = 0$, in which case a similar W can be constructed, based on adjusting n instead of c at time 0, and leaving all future behaviour unchanged.

Proof of Proposition 4: Differentiating with respect to k the identity

$$V^*(k,\beta)$$

$$= u(c^*, 1-n^*) + \beta(1-d)V^*(k^*,\beta),$$

and using the first order condition

$$u_1(c^*, 1-n^*) = \beta(1-d)(1-\delta+r)V_1^*(k^*,\beta),$$

shows that

$$V_1^*(k,\beta) = \beta(1-d)(1-\delta+r)V_1^*(k^*,\beta).$$

Since V_1^* is weakly decreasing in its first argument, the proposition follows.

Proof of Theorem 1: We are going to view the distributions Φ_t as the probability distributions of the characteristics of randomly sampled individuals at times t. To study the convergence properties of this sequence of distributions, we follow the notation and analysis of Stokey and Lucas ([6], henceforth SL). Construct the transition function $P : C \times \mathcal{B} \to \mathbf{R}$, where $\mathcal{B}$ is the collection of Borel subsets of C, and for $s = (k,\beta)$ and $A = I_k \times I_\beta \subset C$ we have

$$P(s,A) = dI(k^*(s) \in I_k)G(I_\beta)$$

$$+(1-d)I(k^*(s) \in I_k, \beta \in I_\beta).$$

Here the notation $I(\cdot)$ refers to the indicator variable that is 1 or 0 depending on whether or not the given statement is true. For fixed s, $P(s,\cdot)$ can be extended to all of $\mathcal{B}$, and will be a probability measure. The N-step transition functions $P^N(s,A)$ are defined by $P^1(s,A) = P(s,A)$ and the equivalent alternative definitions

$$P^{N+1}(s,A) = \int_{x\in C} P(x,A)P^N(s,dx) = \int_{x\in C} P^N(x,A)P(s,dx).$$

In this notation,

$$\Phi_t(A) = \int_{x\in C} P^t(x,A)\Phi_0(dx).$$

Associated with P are two transition functions T and T^*, defined by

$$T(f)(s) = \int_{x\in C} f(x)P(s,dx)$$

and

$$T^*(F)(A) = \int_{x\in C} P(x,A)F(dx).$$

Here f is supposed to be a bounded function on C, and F is a probability distribution on C; thus if $B(C)$ is the space of all bounded functions on C then $T : B(C) \to B(C)$ and $T^* : \mathcal{D} \to \mathcal{D}$ are the domains and ranges of these transition functions. The transition P has the *Feller property* if the transition T maps continuous functions into continuous functions.

Lemma A.1: The transition P has the Feller property.

Proof: Assume that $f : C \to \mathbf{R}$ is continuous and let

$$g(k, \beta) = \int_C f(k', \beta') P((k, \beta), d(k', \beta')) = T(f)(k, \beta),$$

the expected value of $f(k', \beta')$ when (k', β') is a random characteristic pair for the period $t+1$ individual, who is either an agent with characteristics (k, β) at t, or is the immediate descendant of such an agent. We can write

$$g(k, \beta) = (1-d) f(k^*(k, \beta), \beta) + d \int_0^{\bar{B}} f(k^*(k, \beta), \beta') dG(\beta'),$$

in which the first term is a continuous function of (k, β), in view of Proposition 2. For continuity of the second term, assume $(k_n, \beta_n) \to (k, \beta)$, note that $f(k^*(k_n, \beta_n), \beta')$ converges pointwise to $f(k^*(k, \beta), \beta')$, use compactness of C and hence boundedness and uniform continuity of f, get an upper bound on $|f(k^*(k_n, \beta_n), \beta')|$, and finally use dominated convergence.

The claim in the Theorem is equivalent to: For any pair of factor prices w and r, there is a unique distribution Φ_* on C such that for any initial distribution Φ_0 on C, $T^{*n}(\Phi_0) \to \Phi_*$ in the sense of weak convergence. To establish the claim, we make use of the order properties of C and T, as developed in SL (section 12.4).

The space of characteristics is an interval in $\mathbf{R}^2$, in the sense that

$$C = [a, b] = \left\{ x \in \mathbf{R}^2 \middle| \, a \le x \le b \right\},$$

where $x \le y$ means $x_j \le y_j$ for all coordinates, and $a = (0, 0)$, $b = (\bar{K}, \bar{B})$. In multidimensional settings such as ours, the first-order stochastic dominance relation, written $F_1 \succeq F_2$, is defined to occur if

$$\int_C f(x) F_1(dx) \ge \int_C f(x) F_2(dx)$$

whenever $f : \mathbf{R}^2 \to \mathbf{R}$ is weakly increasing. The mapping $T^* : \mathcal{D} \to \mathcal{D}$ is *monotonic* if $F_1 \succeq F_2$ implies $T^*(F_1) \succeq T^*(F_2)$.

Lemma A.2: Our mapping T^* is monotonic.

Proof: We have to assume that $f : C \to \mathbf{R}$ is weakly increasing and that $F_1 \succeq F_2$, and show that

$$\int_C f(s') T^*(F_1)(ds') \ge \int_C f(s') T^*(F_2)(ds').$$

To show this, we view the expectations on the two sides as the outcomes of the two-stage randomizations
1. Draw a characteristic pair $s = (k, \beta)$ according to F_j.
2a. With probability d, operate $k^*(k, \beta)$ to get k', draw β' according to G, and calculate $f(k', \beta')$, and
2b. with probability $1 - d$, operate $k^*(k, \beta)$ to get k', let $\beta' = \beta$, and calculate $f(k', \beta')$.
Thus we can write

$$\int_C f(s')T^*(F_j)(ds') =$$

$$\int_C \left[d \int_{[0,\bar{B}]} f(k^*(k,\beta),\beta')G(d\beta') + (1-d)f(k^*(k,\beta),\beta) \right] F_j(ds).$$

To get the required inequality, we have to show that the square-bracketed function of $s = (k, \beta)$ is weakly increasing. By Proposition 2, $k^*(k, \beta)$ is weakly increasing, so for fixed β', $f(k^*(k, \beta), \beta')$ is weakly increasing and so is the inner integral and the first term. In the second term, $k^*(k, \beta)$, β, hence $f(k^*(k, \beta), \beta)$, and hence $(1 - d)f(k^*(k, \beta), \beta)$ are all weakly increasing in $s = (k, \beta)$.

For monotonic transition functions, sufficient conditions for the required convergence are the Feller property and (SL, p. 381):
Assumption 12.1: There exist $c \in C = [a, b]$, $N \geq 1$, and $\epsilon > 0$ such that:

$$P^N(a, [c, b]) \geq \epsilon \quad \text{and} \quad P^N(b, [a, c]) \geq \epsilon.$$

Theorem 12.12 (SL): Suppose that $T^* : \mathcal{D}[a, b] \to \mathcal{D}[a, b]$ is weakly increasing and the associated transition function has the Feller property; and assume that Assumption 12.1 holds. Then for all $M \in \mathcal{D}[a, b]$, the sequence of distributions $\{T^{*n}(M)\}$ converges to the same limit, namely

$$M^* = \lim_{n\to\infty} T^{*n}(\delta_a) = \lim_{n\to\infty} T^{*n}(\delta_b).$$

To verify Assumption 12.1 in our context, the required lower bounds are easiest to establish if we consider the subset of individuals whose ancestors have all lived for just one period. There will be d^N such individuals at time $N \geq 1$. We denote by $Q^N(k, A)$ the conditional probability that, if a sequence of individuals all live for exactly one period, and the first individual in the sequence is born with physical capital k and a parameter β drawn according to G, then the Nth individual will have characteristics in the set A. To define Q^N formally, it is enough to define Q^0 on rectangles, by

$$Q^0(k, I_k \times I_\beta) = I(k \in I_k)G(I_\beta),$$

extend Q^0 to all of $\mathcal{B}$, then define

$$Q^N(k, I_k \times I_\beta) = \int_C I(k^*(x) \in I_k)Q^{N-1}(k, dx)G(I_\beta)$$

and finally extend $Q^N(k,\cdot)$ to all of $\mathcal{B}$. Obviously, $Q^N(k,\cdot)$ is the product of a marginal distribution, say $M^N(k)$ on $[0,\bar{K}]$, with the marginal distribution G on $[0,\bar{B}]$. Also, P^N and Q^N are related by

$$P^N(s,A) \geq d^N Q^{N-1}(k^*(s),A) = d^N \left[M^{N-1}(k^*(s)) \times G\right](A),$$

where $s=(k,\beta)$ denotes a typical element of $C=[0,\bar{K}]\times[0,\bar{B}]$. Form the map $T_K^*:\mathcal{D}[0,\bar{K}]\rightarrow\mathcal{D}[0,\bar{K}]$ by defining

$$T_K^*(M)([0,k]) = \int_C I(k^*(x,\beta)\leq k)M(dx)G(d\beta)$$

$$= \int_{[0,\bar{K}]} M(x|k^*(x,\beta)\leq k)G(d\beta).$$

Verifying Assumption 12.1 for P now consists largely of verifying enough properties of T_K^* to apply Theorem 12.12 to this marginal-conditional transition function.

Lemma A.3: T_K^* is weakly increasing; that is, if $M_1 \succeq M_2$ then $T_K^*(M_1)\succeq T_K^*(M_2)$.

Proof: The function $k^*(\cdot,\beta)$ is weakly increasing; so if $M_1\succeq M_2$, then the integrand in the second equation defining $T_K^*(M_1)([0,k])$ is everywhere less than or equal to the corresponding integrand in $T_K^*(M_2)([0,k])$. Consequently $T_K^*(M_1)\succeq T_K^*(M_2)$ for all k, as required.

Lemma A.4: The transition function associated with T_K^* has the Feller property.

Proof: The proof consists of writing out exactly what the required continuous function is. The transition function associated with T_K^* takes as input a value $k\in[0,\bar{K}]$ and produces a probability measure on the same set. The conditioning assumptions are that all the agents die, a new generation inherits the last generation's capital stocks, and both generations' members all have patience parameters drawn independently according to G. Thus if $k\in[0,\bar{K}]$, then the transition function, say P_K, is defined by

$$P_K(k,[0,k']) = \int_{[0,\bar{K}]} I(k^*(k,\beta)\leq k')G(d\beta);$$

and we have to show that if $f:[0,\bar{K}]\rightarrow[0,\bar{K}]$ is continuous, then so is $g:[0,\bar{K}]\rightarrow[0,\bar{K}]$, defined by

$$g(k) = \int_{[0,\bar{K}]} f(k^*(k,\beta),\beta)G(d\beta).$$

This function is continuous, by the same arguments as used in showing that P itself has the Feller property.

Lemma A.5: Assumption 12.1 holds for P_K.

Proof: Here we make use of Assumption N for the first time. We have to show that there exists $\bar{k} \in [0, \bar{K}]$, $\epsilon > 0$, and $n \geq 1$, for which

$$P_K^N(0, (\bar{k}, \bar{K}]) \geq \epsilon \quad \text{and} \quad P_K^N(\bar{K}, [0, \bar{k})) \geq \epsilon,$$

or equivalently

$$T_K^{*N}(\delta_0)(\bar{k}, \bar{K}] \geq \epsilon \quad \text{and} \quad T_K^{*N}(\delta_{\bar{K}})[0, \bar{k}) \geq \epsilon.$$

In fact we can let $\bar{k}$ be any capital stock in $(0, \bar{K})$, say $\bar{K}/2$. By Assumption N1, choose $0 < \beta_1 < \bar{\beta} < \beta_2$ for which

$$\alpha = \min\left[G[0, \beta_1], G[\beta_2, \bar{B}]\right] > 0.$$

With probability at least α^n, a sequence of descendants of an initial agent (the 0th) with capital stock 0 will have as its nth agent someone with capital stock at least equal to k_n, where

$$0 < k_1 = k^*(0, \beta_2) < k_2 = k^*(k_1, \beta_2) < \ \ldots \ < k_n = k^*(k_{n-1}, \beta_2);$$

or possibly the sequence reaches $\bar{K}$ at some point and stays there. In either case, this monotonic sequence of capital stocks has a limit k_∞, which can only be $\bar{K}$ because by continuity of k^* we have $k_\infty = k^*(k_\infty, \beta_2)$. So if $\bar{n}_1$ is the first n for which $k_n > \bar{k}$, then

$$P_K^n(0, (\bar{k}, \bar{K}]) \geq \alpha^n$$

for all $n \geq \bar{n}_1$. With probability at least α^n, a similar sequence, in which the 0th agent has capital stock $\bar{K}$, will have as its nth agent someone with capital stock at most equal to k^n, where

$$\bar{K} > k^1 = k^*(\bar{K}, \beta_1) > k^2 = k^*(k^1, \beta_1) > \ \ldots \ > k^n = k^*(k^{n-1}, \beta_1);$$

or possibly this sequence reaches 0 and stays there. In either case, this sequence has a limit k^∞, which must be 0; and if $\bar{n}_2$ is the first n for which $k^n < \bar{k}$, then

$$P_K^n(\bar{K}, [0, \bar{k}]) \geq \alpha^n$$

for all $n \geq \bar{n}_2$. Now we can let $N = \max[\bar{n}_1, \bar{n}_2]$ and $\epsilon = \alpha^N$.

Lemma A.6: The original transition function P satisfies Assumption 12.1.

Proof: We have to find $c \in C = [0, (\bar{K}, \bar{B})]$, $N \geq 1$, and $\epsilon > 0$ for which $P^N(0, [c, (\bar{K}, \bar{B})]) \geq \epsilon$ and $P^N((\bar{K}, \bar{B}), [0, c]) \geq \epsilon$. We claim that if $\bar{\beta}$ is any patience parameter between β_1 and β_2, and

$$\gamma = \min\left[G([0, \beta_1]), G((\beta_2, \bar{B}])\right] > 0,$$

then

$$c = (\bar{k}, \bar{\beta}),$$

and the same N as in the last lemma, will work. We have

$$P^N(s,[c,(\bar{K},\bar{B})]) \geq d^N Q^{N-1}(k^*(s),[c,(\bar{K},\bar{B}))$$
$$= d^N \left[M^{N-1}(k^*(s)) \times G\right]([c,(\bar{K},\bar{B})) \geq d^N \alpha^{N-1}\gamma,$$

and

$$P^N(s,[0,c]) \geq d^N Q^{N-1}(k^*(s),[0,c])$$
$$= d^N \left[M^{N-1}(k^*(s)) \times G\right]([0,c]) \geq d^N \alpha^{N-1}\gamma > 0,$$

so we can take ϵ to be this last quantity.

Proofs of Theorems 2-4 are mostly contained in the text and are omitted.

References

[1] Andolfatto, D. and J. Redekop, "A Parametric Intertemporal Model of Labour-Leisure Choice".

[2] Benveniste, L.M. and J.A. Scheinkman (1979) "On the Differentiability of the Value Function in Dynamic Models of Economics", *Econometrica* v. 47, pp. 727-732.

[3] Epstein, L.G., and A. Hynes (1980), "...", *Journal of Political Economy* v. ?, pp. ??.

[4] Lucas, R.E., and N.L. Stokey (1984), "Optimal Growth with Many Consumers," *Journal of Economic Theory* v. 32, pp. 139-171.

[5] Rae, John (1834). *Statement of Some New Principles on the Subject of Political Economy.* Boston: Hilliard, Gray. Reprinted by Augustus M. Kelly, New York (1964).

[6] Stokey, N.L. and R.E. Lucas (1989). Recursive Methods in Economic Dynamics. Cambridge: Harvard University Press.