

Educational Signaling, Credit Constraints and Inequality Dynamics

Marcello D'Amato and Dilip Mookherjee¹

Preliminary Version, June 8 2008

Abstract

We develop a dynamic theory of educational signaling and income distribution in a two sector economy with missing credit markets. Education and ability matter in a modern/formal sector, not in the traditional/informal sector. Education signals unobserved ability to modern sector employers. Both quantity and quality of human capital evolve endogenously. The model generates a Kuznets inverted-U pattern in skill premia similar to those observed in historical US and UK experience: initially rising, later falling as skill accumulation progresses. In the first phase the social return to education exceeds the private return: under-investment owing to credit market imperfections dominate over-investment owing to signaling distortions. In the later stage the pattern is reversed.

¹University of Salerno and Boston University, respectively. Mookherjee's research has been supported by NSF Grant no. SES-0617874. We are grateful to Bob Margo for familiarizing us with the historical literature on skill premia in the US and in the UK.

1 Introduction

There are two major market imperfections in the context of education: *signaling* and *credit constraints* that form the focus of two distinct literatures. Signaling models in the Spence (1974) tradition are usually static, and abstract from credit constraints or income distribution. According to this approach, there tends to be over-investment in education, and market based rates of return to education overstate the social returns. Accordingly estimates of the rate of return to education based on market wages need to be adjusted downward for ‘ability-bias’. Empirical estimation of the magnitude of this ‘ability bias’ is the subject of a large literature in the economics of education (see, e.g., Card (2001)).

Credit constraint based models (e.g., Ray (1990), Galor-Zeira (1993), Ljungqvist (1993), Freeman (1996), Mookherjee and Ray (2003)) on the other hand tend to focus on dynamics of income growth and distribution, highlighting long-term persistence of income differences between countries and families. In these models there tends to be under-investment in education among poor families, and policy recommendations involve public educational subsidies for poor families.² These models have not generated a comparable empirical literature devoted to attempts to measure the extent of under-investment in human capital that owe to financial market imperfections. Yet at the level of theory they are widely accepted as relevant and potentially important. If it is the case that there is significant under-investment in education, that would imply a *negative* ability bias in market-based measures of the returns to investment. The natural question, particularly relevant for developing countries, therefore concerns the relative significance of signaling and financial market distortions, and the net effect on ‘ability bias’.

Despite the obvious practical importance of both signaling and credit constraints, very

²To be sure, such policies raise per capita income, income mobility (in the short run) and lower inequality, but are not necessarily Pareto-improving. Steady states can be constrained Pareto-efficient: e.g., with a continuum of occupations Mookherjee and Ray (2003) show that there is a unique steady state which is (unconstrained) Pareto efficient. However with two occupations (the context of this paper) they show there is a continuum of steady states, some of which are Pareto-inefficient and involve under-investment, while other steady states are Pareto-efficient.

few papers examine models in which both co-exist.³ The two phenomena are likely to be intertwined: signaling costs will depend on credit constraints and income distribution, since education will be more difficult to afford by poor families when they are credit constrained. And income distribution in the long run will in turn depend on signaling: education represents a channel of upward mobility for poor families which could help them escape under-investment poverty traps. In that case even if skilled wages exceed the marginal product of skilled workers owing to signaling distortions, increasing education may be welfare-enhancing. To assess the net impact of signaling and underinvestment distortions which operate in opposing directions, we first need a conceptual framework where signaling and credit constraints co-exist.

This paper constructs a dynamic theory of educational signaling and income distribution in the context of a two sector model of development. There is a traditional sector where neither education nor ability matter: productivity and wages are exogenously fixed at some low level. One can think of this either as a rural or an urban informal sector. Productivity tends to be higher in the modern sector, but workers need to be educated to work in this sector. Moreover, abilities of educated workers vary in the modern sector and are unobserved by employers; they are drawn randomly (i.i.d.) from a given distribution in any generation. Following Loury (1981), parents cannot take out loans on behalf of their children and must pay for their children's education by sacrificing current consumption. Parents are altruistic in a paternalistic fashion towards their children: they care about their own consumption and their children's future wealth.⁴ Parents observe the ability of their children and decide whether or not to provide them an education.

Employers in the modern sector operate with a constant returns to scale technology; they can check educational qualifications but not the ability of any applicant.⁵ More able agents

³Exceptions include Stiglitz (1975), Hendel et al (2005): these are described further in Section 7.

⁴This is the formulation of the bequest motive adopted in Becker and Tomes (1979), Mookherjee and Ray (2002) and Mookherjee and Napel (2007).

⁵In addition we abstract from traditional effects of supply of skilled people on the skill premium by assuming that the productivity of any given worker is independent of the number or type of other workers employed. Numerical simulations show that the principal results are robust in the context of a CES technology with imperfect substitutability between skilled and unskilled workers.

can acquire education at cheaper cost, hence education is a signal of ability. Education is productive insofar as it enables an agent to move from the traditional to the modern sector. Conditional on being educated, worker productivity increases with ability in the modern sector. In this world, lack of education may reflect either low ability or low parental wealth (or both), thus complicating the signal extraction problem of employers.

In contrast to standard macro models of human capital accumulation with credit constraints⁶, quality and quantity of the workforce in the modern sector both constitute state variables. Owing to signaling problems, wages equal average rather than marginal products in the modern sector, and the composition of the skilled workforce (e.g., distribution of parental wealth) matters in the determination of labor allocation, modern sector wages and per capita income. This complicates the dynamics of the model considerably, but also generates some novel implications.

Our first main result is that the model has a unique steady state, despite the lack of any technological source of diminishing returns. There are no possible macro-poverty traps or history-dependence.⁷ In these respects it resembles a neoclassical Solow growth model and contrasts with credit-constraint based theories of human capital accumulation. Income differences across countries are transitory and can be explained (partly) by differences in the proportion of workforce that becomes educated. Restricting initial conditions to appropriate subsets of the state space (i.e., the skill proportion below the steady state, and a skill premium that is not too low)⁸, there is income growth in the short run owing to human capital accumulation, and countries with a higher skill ratio tend to have higher per capita income. The process of development is then described by the non-steady-state dynamics.

Apart from rising per capita education, the development process also involves changes

⁶Ray (1990), Galor-Zeira (1993), Ljungqvist (1993), Freeman (1996), Mookherjee and Ray (2002, 2003), or Mookherjee and Napel (2007).

⁷This result is driven partly by the existence of random ability shocks in the model, which limits the extent of long run persistence considerably, as explained in Mookherjee and Napel (2007). An additional assumption responsible for this result is that the wage in the traditional sector is exogenously fixed, so there are no wealth effects associated with human capital accumulation among unskilled families.

⁸If the skill premium is too low, too few people are attracted to the modern sector and liquidity constraints bind more strongly, causing the skilled proportion in the economy to decline.

in the quality of the educated workforce, reflected by the wage gap between the two sectors (or the skill-premium). In general, the quantity-quality dynamic is quite rich: any combination of rising/falling quantity and rising/falling quality can result under suitable initial conditions. If the initial skill premium is not too high, it subsequently exhibits an inverse-U Kuznets curve of the sort observed in the historical experience of the US and UK (Williamson (1985), Margo (2000), Goldin and Katz (2007)): rising initially and falling subsequently as the supply of skills increases over time. The initial stage is characterized by a process of upward mobility wherein talented children from poor (i.e., unskilled) families acquire education and migrate to the modern sector, raising the average quality of the workforce in the latter. During this stage the under-investment associated with the capital market imperfection dominates the over-investment associated with signaling, and the ability bias is negative. At the later stage, enough talented people have already arrived in the modern sector. Further additions to this sector represent moves by people less talented on average than those already there. This causes average quality and the skill premium to fall, so ability bias is positive at later stages of development.

However, neither the Kuznets pattern nor convergence to steady state can be established theoretically. In general we can show that there exist a region of initial conditions for which the subsequent dynamic for the succeeding two generations exhibits the Kuznets pattern. We therefore investigate numerical solutions of the model with logarithmic utility, uniform ability distributions and alternative specifications of the production function. In all of these, both the Kuznets pattern and convergence to steady state continue to obtain.

The novel feature of this model is that it is possible for skill premia and skill proportion in the economy to increase at the same time (i.e., the first stage of the Kuznets process). This cannot be explained by conventional models where wages are determined by marginal products, and an unchanging technology. Indeed, the co-existence of rising skill premia with rising skill accumulation is commonly seen by empirical economists as *prima facie* evidence for existence of skill-biased technical change. Our model provides an alternative explanation for the same phenomenon, based on changes in the composition of the educated workforce.⁹

⁹Hendel, Shapiro and Willen (2005) and Regev (2007) provide similar explanations based on signaling models; these are reviewed in more detail in Section 7.

More generally, our theory focuses attention on the determination of dynamic skill supply patterns, which determines skill premium dynamics jointly with technical change, but has not received comparable attention.¹⁰

The dynamic patterns predicted by our model are shown to be robust with respect to two alternative formulations of expectations concerning future wages: *static* expectations where educational decisions are taken under the assumption that the current skill premium will prevail in future, and *rational* expectations where households are more sophisticated and have perfect foresight. Both expectational processes are associated with the same steady state, and differ only with respect to non-steady-state behavior. We show that weak assumptions concerning selection of static equilibria with rational expectations imply that the dynamics are qualitatively similar.

Section 2 provides a more detailed discussion of historical evidence concerning skill premium dynamics in the context of the US and UK. Section 3 introduces the model. Section 4 presents analysis of steady states, while Section 5 deals with non-steady-state dynamics. Section 6 provides numerical simulations that demonstrate the Kuznets pattern under varying specifications of parameter values and technologies. Section 7 discusses related literature. Finally, Section 8 concludes by discussing normative implications for ‘ability bias’ and educational policy that need to be explored in more detail in future research.

2 Historical Evidence on Skill Premia

Williamson (1985) provides a comprehensive treatment of earning inequality in the UK over the course of the 19th century. Irrespective of the inequality measure used, the evidence shows an increase in inequality from 1827 until 1851, and a subsequent fall between 1851 and 1901. For instance, the economy-wide Gini coefficient for male earnings rose from 0.293 in 1827 to 0.358 in 1851, falling thereafter to 0.328 in 1881 and 0.331 in 1901. Decomposing these inequality changes into the role of employment shifts across sectors, changing intra-occupational inequality and changing inter-occupational inequality, the dominant source

¹⁰This point is elaborated further in Section 2.

for these trends was accounted for by inter-occupational inequality. In particular the ‘pay ratio’ or disparity between skilled and unskilled wages displayed the Kuznets pattern and accounted for “three quarters of the rise in total earning inequality both in the economy as a whole and in non-agricultural employment” (Williamson (1985, p.43)). The pay ratio (using ‘variable’ weights, i.e., different census year observations) in the economy as a whole rose from 2.452 in 1815 to 3.486 in 1861, and fell thereafter to 2.483 in 1911 (Williamson (1985, Table 3.7)). Williamson (1985, Ch. 10) subsequently argued that the two key factors driving these patterns in skill premia were ‘unbalanced productivity advance’ and ‘skills per worker growth’, supplemented by changes in world market conditions.

The evolution of skill premia in 19th century experience of the United States has been the subject of some controversy. Williamson and Lindert (1980) assembled a variety of previously published evidence concerning wages of skilled artisans and unskilled workers to argue that skill premia followed an inverted U in the US case. They claimed a sharp rise in skill premia from roughly 1820 to 1860 corresponding to early industrialization, followed by a more modest rise and then plateau in the late 19th century, and then a decline in the 20th century. These findings were criticized by subsequent historians (e.g., Margo and Villaflor (1987)) who failed to find similar patterns using other sources of evidence concerning the ratio of wages of skilled artisans to unskilled workers. However Margo (2000) subsequently provided evidence that in the four decades prior to the Civil War, real wages of white-collar workers grew faster (32%) than those of unskilled workers (21%) or artisans (15%). Combining his own estimates with those of Goldin (1998), Margo argues the evidence shows that the relative wage of white-collar workers remained stationary between 1850s to the late 19th century. Since the beginning of the 20th century the work of Goldin and Katz (2007, Figure 6, p.148) indicates that the wage premiums earned by both college and high school graduates fell sharply between 1915 and 1950 (the log of both wage ratios fell from around 0.6 to below 0.35 during this period). Putting together these accounts, it appears that a Kuznets pattern characterized skill premia in the US between 1820 and 1950: rising between 1820–60, stationary until the turn of the century and falling thereafter until 1950.

The evolution of skill premia since 1950 in the US has been the subject of considerable research and discussion (e.g., see summaries in Goldin and Katz (2007) or Acemoglu (2002)).

Goldin and Katz argue that an important factor underlying the rise in skill premia since the late 1970s is a slowdown in the rate of increase in supply of skills, which failed to keep up with rates of skill biased technical change. For instance, Table 1 in Goldin-Katz (2007, p. 153) shows the annual rate of change in skill supply slowed from 3.83% during 1960–80 to 2.43% during 1980–2005, while the change in relative demand for skilled workers remained stationary (3.85% in the former period, and 3.76% in the latter). The slowdown in rates of skill accumulation reflect slower growth in educational attainment among natives, which slowed from 3.83% to 2.43% across these two periods. The causes of this are not explored further by Goldin and Katz, though they argue it is unlikely to result from reaching an ‘upper bound for educational attainment’, since returns to further educational investments continue to be substantial (Goldin-Katz (2007, p.157)).

Most accounts of skill premia dynamics focus on the ‘race between technology and education’ in a traditional supply-demand framework: rises in skill premia are explained by derived demand increases in the relative demand for skilled workers owing to skill-biased technical changes that outstrip increases in supply of skilled workers. Factors explaining technical change receive considerable discussion, and is treated either as exogenous or endogenous (e.g., Acemoglu (2002) argues that such technical change is endogenous and reacts to changes in the stock of skilled workers relative to unskilled workers). The factors underlying changes in supply of skills usually receives less discussion, except for changes in public schooling or educational subsidies: e.g., the decline in skill premia between 1915-1950 in the US is explained by Goldin and Katz by an increase in educational attainment owing to reforms in public schooling. Our model emphasizes other factors such as signaling and capital market imperfections which affect the supply of skills, which have hitherto received less attention.¹¹

¹¹See, however, Acemoglu (2002, pp. 65–68) who dismisses the possibility that changing composition of educated workers can explain the rise in ‘residual inequality’ in the US, on the basis of theoretical and empirical arguments. This is discussed further in Section 7.

3 Model

The traditional sector has a fixed wage v . The endogenous wage in the modern sector is denoted w . Agents' innate abilities are denoted by n ; in any generation t these are drawn randomly from a given distribution with c.d.f. F , which has full support on $[0, \bar{n}]$, and has a continuous density function f . Education is a 0 – 1 decision. Productivity in the modern sector equals $e.n$, where e denotes education and n the ability of a worker. Productivity in the traditional sector equals v for all agents. Hence working in the modern sector requires education, unlike the traditional sector. Production operates according to constant returns to scale, and both sectors produce a common consumption good. We assume that the average ability in the population En exceeds v — this will ensure that the modern sector wage will always exceed the traditional sector wage.

There is a continuum of families indexed $i \in [0, 1]$. Each family has a single agent in a given generation, whose payoff is $U(c_{it}) + V(y_{i,t+1})$, where c_{it} denotes consumption of this parent, $y_{i,t+1}$ denotes the income of its child, and both U, V are strictly increasing, strictly concave and twice differentiable functions.

The parent in household i at t observes the ability draw of its child $n_{i,t+1}$ and then decides whether to invest in the latter's education. Education costs $x(n)$ for a child of ability n , where x is strictly decreasing, differentiable, with $x(\bar{n}) = 0$. If w_{t+1} is the skilled wage expected to prevail at $t+1$, a parent with income $y_{it} \in \{v, w_{it}\}$ and a child with ability n will select an education decision $e = e_{i,t+1} \in \{0, 1\}$ to maximize

$$U(y_{it} - e.x(n)) + V(ew_{i,t+1} + (1 - e)v). \tag{1}$$

Here U represents the utility of the parent from its own consumption, and V the altruistic benefit it derives from the future earnings of its child. We assume both are continuously differentiable, strictly increasing and strictly concave functions. Implicit in this formulation is the assumption that education loan markets are missing.

Clearly if $w_{i,t+1} < v$ then no parent in generation at t will invest, whereas if $w_{i,t+1} > v$ some parents (with gifted children) will invest. In case of indifference we shall assume that investment will take place. For any given skilled wage $w^e \geq v$ expected in the next

generation, the investment decision of a parent with income y is described by an ability threshold $n^*(w^e, y)$ at which the parent is indifferent:

$$U(y) - U(y - x(n^*(w^e, y))) = V(w^e) - V(v). \quad (2)$$

Then children with ability at or above this threshold receive education, and others do not.

Let λ_t denote the fraction of population that is skilled at t , and w_t^e the skilled wage at $t + 1$ anticipated by parents of generation t . Then the evolution of the skill proportion is given as follows:

$$\lambda_{t+1} = \tilde{\lambda}(w_t^e; w_t, \lambda_t) \equiv \lambda_t[1 - F(n^*(w_t^e, w_t))] + (1 - \lambda_t)[1 - F(n^*(w_t^e, v))] \quad (3)$$

Bertrand competition among employers in the modern sector implies the skilled wage in the next generation is (with $m(n^*)$ denoting $E[n|n \geq n^*]$):

$$w_{t+1} = \tilde{q}(w_t^e; w_t, \lambda_t) \equiv \frac{[m(n^*(w_t^e, w_t))\lambda_t[1 - F(n^*(w_t^e, w_t))] + m(n^*(w_t^e, v))(1 - \lambda_t)[1 - F(n^*(w_t^e, v))]}{\lambda_{t+1}}, \quad (4)$$

provided $\lambda_{t+1} > 0$. In case $\lambda_{t+1} = 0$, we shall set $w_{t+1} = \bar{n}$.¹²

It remains to specify wage expectations. We shall consider two expectational processes: *static expectations (SE)* where $w_t^e = w_t$ and *rational expectations (RE)* where $w_t^e = w_{t+1}$. This generates the following definitions of competitive equilibrium dynamics.

Definition 1 *A dynamic competitive equilibrium sequence with static expectations (ESE) given initial conditions (w_0, λ_0) is a sequence $(w_t, \lambda_t), t = 1, 2, \dots$ such that $\lambda_{t+1} = \tilde{\lambda}(w_t; w_t, \lambda_t), w_{t+1} = \tilde{q}(w_t; w_t, \lambda_t)$ for all $t = 0, 1, 2, \dots$. A dynamic competitive equilibrium sequence with rational expectations (ERE) given initial conditions (w_0, λ_0) is a sequence $(w_t, \lambda_t), t = 1, 2, \dots$ such that $\lambda_{t+1} = \tilde{\lambda}(w_{t+1}; w_t, \lambda_t), w_{t+1} = \tilde{q}(w_{t+1}; w_t, \lambda_t)$ for all $t = 0, 1, 2, \dots$*

¹²In other words, if there are no agents that are educated at $t + 1$, the skilled wage is set equal to the highest ability in the population. This assumption prevents the possibility of the economy getting trapped in trivial steady states where $w < v$ and $\lambda = 0$. We do this to ensure that perceived average quality is continuous with respect to the expected wage at $w^e = v$.

Note that ESE is recursively determined: the wage and skill proportion at any date uniquely determine the wage and skill proportion at the next date. Not so for ERE, where the market-clearing wage in the modern sector at $t + 1$ is a fixed point of the function $\tilde{q}(\cdot; w_t, \lambda_t)$.

4 Steady State

It is obvious from the definitions above that both static and rational expectations processes are associated with the same steady states.

Definition 2 *A steady state (SS) is w^*, λ^* such that $\lambda^* = \tilde{\lambda}(w^*; w^*, \lambda^*)$, $w^* = \tilde{q}(w^*; w^*, \lambda^*)$.*

Hence in looking for steady states we may as well confine attention to stationary points of the static expectations dynamic.

Proposition 1 *There exists a unique SS.*

Proof. Define $n^R(w) \equiv n^*(w, w)$ and $n^P(w) \equiv n^*(w, v)$. These are both continuous functions mapping $[0, \bar{n}]$ to itself.

Next define

$$L(w, \lambda) = \lambda[1 - F(n^R(w))] + (1 - \lambda)[1 - F(n^P(w))] \quad (5)$$

$$Q(w, \lambda) = \frac{m(n^R(w))\lambda[1 - F(n^R(w))] + m(n^P(w))(1 - \lambda)[1 - F(n^P(w))]}{\lambda} \quad (6)$$

which map $[En, \bar{n}] \times [0, 1]$ to itself. (Note that $En > v$ ensures that $L(w, \lambda)$ is strictly positive for every $w \geq En$, so Q is well-defined.) Since F and n^P, n^R are continuous functions, (L, Q) is a continuous map, so must have a fixed point, which establishes steady state existence.

To establish steady state is unique, note that given any $w \geq En$, $L(w, \lambda)$ is a contraction map in λ alone, since:

$$L(w, \lambda) = 1 - F(n^P(w)) + \lambda[F(n^P(w)) - F(n^R(w))]$$

and $0 < F(n^P(w)) - F(n^R(w)) < 1$ as $w \geq En > v$. Hence given w the map L has a unique fixed point which we denote by $\lambda(w)$:

$$\lambda(w) = \frac{1 - F(n^P(w))}{1 - \{F(n^P(w)) - F(n^R(w))\}}. \quad (7)$$

Clearly every steady state must satisfy $\lambda = \lambda(w)$. It must also satisfy $w = q(w) \equiv Q(w, \lambda(w))$. Using the fact that $[1 - \lambda(w)][1 - F(n^P(w))] = \lambda(w) - \lambda(w)[1 - F(n^R(w))] = \lambda(w)F(n^R(w))$, we can express

$$q(w) = \int_{n^R(w)}^{\bar{n}} nf(n)dn + \frac{F(n^R(w))}{1 - F(n^P(w))} \int_{n^P(w)}^{\bar{n}} nf(n)dn. \quad (8)$$

This implies

$$q_w = -n^R f(n^R) n_w^R - n^P f(n^P) \frac{F(n^R)}{1 - F(n^P)} n_w^P + \left[\frac{f(n^R)}{1 - F(n^P)} n_w^R + \frac{F(n^R) f(n^P)}{[1 - F(n^P)]^2} n_w^P \right] \int_{n^P}^{\bar{n}} nf(n)dn$$

so

$$q_w = [m(n^P) - n^R] f(n^R) n_w^R + [m(n^P) - n^P] \frac{f(n^P) F(n^R)}{1 - F(n^P)} n_w^P$$

which is negative since $m(n^P) > n^P > n^R$ and $n_w^P, n_w^R < 0$. Hence q cannot have more than one fixed point. ■

One important reason for steady state uniqueness is the fixed nature of the wage v in the traditional sector. This owes to the constant returns assumption, as well as the irrelevance of ability in that sector. With diminishing returns to labor, increasing out-migration would drive up the traditional wage. Then (as in Mookherjee-Napel (2007)) there could be multiple steady states, as higher wages in the traditional sector relax liquidity constraints and allow more unskilled households to educate their children. On the other hand, if ability is positively related to productivity in the traditional sector as well, out-migration would tend to drive down the traditional wage. The consequences of this are complicated: a lower wage in the traditional sector lowers the capacity of parents in that sector to educate their children, but it encourages their motivation to do so as the wage gap between the two sectors grows. If the former effect dominates we would expect uniqueness again.

One part of the argument is that q the average quality of the workforce in the modern sector is decreasing in the wage w . This owes to the greater ‘pull’ of the modern sector where wages there rise, inducing a decline in the ability of the marginal type from within the traditional or modern sector that receive education. There is however a complicating compositional effect: those migrating into the sector from the traditional sector come from poorer families, compared with children of families already in the modern sector. Hence the former are more talented than those coming from within the modern sector. If the proportion of the former rises appreciably, average quality in the modern sector could rise following a rise in w . If the proportions are such as to maintain steady state (i.e., $\lambda = \lambda(w)$) the proof shows that this compositional effect is not powerful enough to allow multiple steady states. Out of steady state, however, it can cause quality and wage in the modern sector move in the same direction, as we shall see in the next section.

5 Non-Steady State Dynamics

5.1 Static Expectations

We start with the case of static expectations. This may be considered plausible from a behavioral standpoint. In any case it is simpler to work with, being recursively determined. We shall show later that the main results continue to apply with rational expectations.

Proposition 2 *Consider any competitive equilibrium sequence with static expectations. There exist functions $\lambda(w), \lambda_1(w)$ mapping $[v, \bar{n}]$ into $[0, 1]$ both of which pass through the steady state (λ^*, w^*) , with $\lambda(w)$ given by (7), and $\lambda_1(w) < (>)\lambda(w)$ according as $w > (<)w^*$, such that (as depicted in Figure 1):*

(a) $\lambda_{t+1} > (=, <)\lambda_t$ according as $\lambda_t < (=, >)\lambda(w_t)$, and

(b) $w_{t+1} > (=, <)w_t$ according as $\lambda_t < (=, >)\lambda_1(w_t)$.

Proof. (a) follows from the contraction property of L in λ for given w , since $\lambda_{t+1} = L(w_t, \lambda_t)$ and $\lambda(w_t)$ solves for λ in $\lambda = L(w_t, \lambda)$.

To prove (b), note that $Q(w, \lambda) = \alpha m(n^R(w)) + (1 - \alpha)m(n^P(w))$ where α denotes $\frac{\lambda[1-F(n^R)]}{\lambda[1-F(n^R)]+(1-\lambda)[1-F(n^P)]}$. It is easily verified that α is increasing in λ . Moreover, Q is decreasing in α since $m(n^P(w)) > m(n^R(w))$. So $Q(w, \lambda)$ is decreasing in λ , implying that $Q(w, \lambda) - w$ is decreasing in λ .

If there exists $\lambda_1 \in (0, 1)$ such that $Q(w, \lambda_1) - w = 0$, define this to be $\lambda_1(w)$. If $Q(w, \lambda) - w < 0$ for all w , set $\lambda_1(w) = 0$. If $Q(w, \lambda) - w > 0$ for all w , set $\lambda_1(w) = 1$. Note that if $w > w^*$ then $Q(w, \lambda(w)) = q(w) < w$, implying $\lambda_1(w) < \lambda(w)$. Conversely, if $w < w^*$ then $Q(w, \lambda(w)) = q(w) > w$, implying $\lambda_1(w) > \lambda(w)$. ■

Note that while $\lambda(w)$ is upward-sloping, it is difficult to sign the slope of $\lambda_1(w)$. From the Implicit Function Theorem $\lambda'_1 = \frac{1-Q_w}{Q_\lambda}$, so $1 > Q_w$ ensures it is downward sloping. And

$$Q_w = (1 - \alpha)m'(n^P(w))n_w^P + \alpha m'(n^R(w))n_w^R + \alpha_w[m(n^R) - m(n^P)] \quad (9)$$

The first two terms on the right-hand-side of (9) are negative, reflecting the lowering of quality of the marginal person receiving education from within the pool of unskilled and skilled families as the wage in the modern sector grows. The third term involves changing composition of the pool of the educated between these two groups. This compositional effect cannot be signed unambiguously, since $\alpha_w > 0$ if and only if

$$\frac{f(n^P)}{1 - F(n^P)}[-n_w^P] < \frac{f(n^R)}{1 - F(n^R)}[-n_w^R]. \quad (10)$$

In other words, it depends on the relative hazard rates of the ability distribution at the respective thresholds of the two groups, weighted by the slope of the threshold with respect to the wage. If (10) holds, then $Q_w < 0$, and $\lambda_1(w)$ is downward-sloping. But it is possible that (10) does not hold at some w , i.e., an increase in the modern sector wage elicits a much larger response from children in families located in the traditional sector, than those in the modern sector. In that case the compositional effect contributes to an improvement in the quality of the workforce in the modern sector. If it is strong enough to overwhelm the direct

effect of quality of each group separately, it is possible that increasing wages improve the quality of the skilled workforce.

However, “on average” the $\lambda_1(w)$ function must be downward sloping, in the following sense. Define \tilde{w} by the solution to $m(n^P(w)) = w$, and \hat{w} by the solution to $m(n^R(w)) = w$, if these solutions exist. Clearly $\tilde{w} > w^*$ since at w^* we have $m(n^P(w^*)) > q(w^*) = w^*$. If \bar{n} is large enough in the sense that $m(n^P(\bar{n})) < \bar{n}$, then \tilde{w} is well defined and lies in the interval (w^*, \bar{n}) . Then $\lambda_1(w) = 0$ for all $w > \tilde{w}$, since $Q(w, \lambda) < Q(w, 0) = m(n^P(w)) < w$ for all $w > \tilde{w}$ and all $\lambda > 0$. Conversely, note that \hat{w} is well-defined and lies in the interval (v, w^*) since $m(n^R(v)) \geq En > v$, and $m(n^R(w^*)) < q(w^*) = w^*$. Then for all w in the interval (v, \hat{w}) we must have $\lambda_1(w) = 1$. So the $\lambda_1(\cdot)$ function slopes down on average in the sense that it equals 0 above \tilde{w} , w^* at w^* , and 1 below \hat{w} . It will slope downwards at any point where the compositional effect is not strong enough in the sense that $Q_w < 1$.

One set of sufficient conditions for $\lambda_1(w)$ to be downward-sloping throughout the interior of the state space is provided below.

Remark 1 *Suppose the hazard rate of the ability distribution $\frac{f(n)}{1-F(n)}$ is non-increasing in n ¹³, and education cost $x(n)$ is linear or concave in n . Then $\lambda_1(w)$ is everywhere decreasing in the interior of the state space.*

Proof. Note that

$$-n_w^P = \frac{V'(w)}{U'(v - x(n^P))[-x'(n^P)]} \quad (11)$$

while

$$-n_w^R = \frac{V'(w) + U'(w - x(n^R)) - U'(w)}{U'(w - x(n^R))[-x'(n^R)]}. \quad (12)$$

By definition of n^R, n^P we have

$$U(w) - U(w - x(n^R)) = U(v) - U(v - n^P) = V(w) - V(v)$$

implying that $U(w - x(n^R)) > U(v - x(n^P))$. Therefore $U'(w - x(n^R)) < U'(v - x(n^P))$. Since $n^P > n^R$, the concavity or linearity of x implies $-x'(n^R) \leq -x'(n^P)$. Then (11, 12)

¹³An example is an exponential distribution, where $f(n) = ke^{-\mu n}$, whence the hazard rate is constant.

imply $-n_w^R > -n_w^P$. Combined with (10) and the non-increasing hazard rate, we obtain $\alpha_w > 0$. This implies $Q_w < 0$. In the interior of the state-space $\lambda_1(w)$ is the solution to $Q(w, \lambda) = w$, so $\lambda_1'(w) = \frac{1-Q_w}{Q_\lambda} < 0$. ■

In numerical computation of the equilibrium dynamics for log utility and uniform ability distributions (described in the next Section), the $\lambda_1(w)$ function turns out to be downward-sloping throughout. So for the purpose of the remaining discussion of this section we shall proceed on this assumption, whence the inverse of λ_1 function is well-defined.

Proposition 2 shows that the non-steady-state dynamics can be characterized by a partition of the state space (λ, w) into four regions, as depicted in Figure 1:

- I.** $\lambda_t < \lambda(w_t), \lambda_t < \lambda_1(w_t)$: here $w_{t+1} > w_t, \lambda_{t+1} > \lambda_t$. Both quality and quantity of the modern work force grows.
- II.** $\lambda_t < \lambda(w_t), \lambda_t > \lambda_1(w_t)$: here $w_{t+1} < w_t, \lambda_{t+1} > \lambda_t$. The quantity of the modern work force grows, but its quality declines.
- III.** $\lambda_t > \lambda(w_t), \lambda_t > \lambda_1(w_t)$: here $w_{t+1} < w_t, \lambda_{t+1} < \lambda_t$. Both quality and quantity of the modern work force shrink.
- IV.** $\lambda_t > \lambda(w_t), \lambda_t < \lambda_1(w_t)$: here $w_{t+1} > w_t, \lambda_{t+1} < \lambda_t$. Quality improves, but quantity declines.

Consider a country with low per capita income owing to a low proportion and quality of workforce in the modern sector. The quantity of skilled workforce is low in the sense that $\lambda_t < \lambda(w_t)$. Then we are in either region I or II. If the quality is also low in the sense that $w_t < \lambda_1^{-1}(\lambda_t)$, we are in region I. Both quality and quantity of the modern work-force will grow from t to $t + 1$. Both will contribute to a rise in per capita income:

$$y_{t+1} - y_t = [\lambda_{t+1} - \lambda_t][w_t - v] + \lambda_{t+1}[w_{t+1} - w_t] \quad (13)$$

and the social rate of return to education exceeds the market rate of return:

$$\frac{y_{t+1} - y_t}{\lambda_{t+1} - \lambda_t} = (w_t - v) + \lambda_{t+1} \frac{w_{t+1} - w_t}{\lambda_{t+1} - \lambda_t} \quad (14)$$

During this early phase of development, there are a sufficiently large proportion of new entrants into the modern sector from the traditional sector. These new entrants come from poorer backgrounds and are more able than those in the modern sector in the previous generation. Upward mobility goes hand-in-hand with a positive externality: the marginal entrants from the traditional sector are smarter on average than those already in the modern sector, causing the wage to rise, which benefits all others in the modern sector.¹⁴ Over this range, the under-investment effect owing to the capital market imperfection dominates the over-investment effect owing to signaling.

This dynamic will propel the economy into region II, as enough people migrate into the modern sector.¹⁵ Subsequently the proportion of the educated will continue to grow, while quality will fall. During this subsequent stage, the rise in the modern sector wage attracts new types with lower ability than those in the modern sector in the previous generation, which causes the wage to fall. In this case, the ‘ability bias’ is positive: the social return to education falls below the private return. Per capita income growth is likely to slow down both because the increase in the skill proportion is likely to slow down, and the quality of the skilled workforce starts to fall. Over this range the signaling externality overwhelms the capital market imperfection.

It is possible that the economy converges thereafter to a steady state, though we have not been able to prove any results concerning convergence. We explore this issue in the context of the numerical solutions below.

In general, however, the dynamics are quite complicated. Regions III and IV are those where there is ‘too much’ education in the economy relative to the wage, causing the proportion of educated to fall. Quality also declines in region III, so per capita income

¹⁴If increasing scarcity of labor in the traditional sector causes wages there to rise, then the migration benefits those remaining in the traditional sector as well.

¹⁵If $Q_w < 0$, as in the case described in Remark (1), and $w_t < w^*$, then the economy must move to Region II in the next generation. This is because the skill ratio will move towards λ^* but cannot overshoot it (since $\lambda_{t+1} < \lambda(w_t) < \lambda(w^*) = \lambda^*$). On the other hand, the fact that Q is decreasing, $\lambda_1^{-1}(\lambda_t)$ is the unique fixed point of $Q(\cdot; \lambda_t)$ and $w_t < \lambda_1^{-1}(\lambda_t)$ implies that $w_{t+1} > \lambda_1^{-1}(\lambda_t) > \lambda_1^{-1}(\lambda_{t+1})$, the last inequality following from the fact that $\lambda_{t+1} > \lambda_t$ and the assumption that λ_1 is a decreasing function.

definitely falls. In region IV, quantity declines and quality increases, so the effect on per capita income is ambiguous. The economy could converge to the steady state if the initial position is to the south-east of the steady state. If it is to the south-west, it could transit into Region I.

Note also that reverse transitions from region II to region I cannot be ruled out. It appears possible then that even with $\lambda_1(\cdot)$ downward sloping, the economy could flip-flop between these two regions. Hence the dynamics could be more complicated than the simple Kuznets pattern: periods of falling skill premia can be interspersed with periods of rising skill premia.

Nevertheless the dynamics of the skill premium provides a useful guide to the divergence between social and private returns to education. A rising premium indicates the social return lies above the private return, while a falling one indicates a positive ‘ability bias’.

5.2 Rational Expectations

We now consider the case of rational expectations. With forward-looking agents, the equilibrium sequence cannot be recursively computed. A related problem is that short-run competitive equilibrium of the modern sector labor market may not be unique. Recall the definition of the perfect foresight equilibrium skilled wage w_{t+1} , i.e., given the state (w_t, λ_t) , it is a fixed point of $\tilde{q}(\cdot; w_t, \lambda_t)$. Owing to the compositional effect (explained above) this function can be non-monotone: a rise in w_{t+1}^e could raise the average quality of the workforce in the modern sector over some ranges. So there may be multiple wage equilibria.

If we focus on a locally stable equilibrium (where \tilde{q} is downward-sloping), the wage will be locally decreasing in λ_t . An increase in λ_t (for given w_{t+1}^e) raises the proportion of children coming from wealthier backgrounds, which lowers the average quality of the workforce in the next generation. It is therefore natural to select equilibria so that this property is globally satisfied.

Similarly, an increase in w_t for given w_{t+1}^e raises the proportion of children with educated parents that choose to be educated, lowering average quality of the educated workforce at

$t + 1$.

If the highest fixed point or the lowest fixed point (corresponding to the most optimistic or most pessimistic expectations) is always selected, the perfect foresight equilibrium wage function $w_{t+1} \equiv Q^R(w_t, \lambda_t)$ will be decreasing and (almost everywhere) continuously differentiable in both w_t and λ_t .¹⁶

Proposition 3 *Suppose that with rational expectations, the equilibrium wage w_{t+1} is given by a function $Q^R(w_t, \lambda_t)$ which is decreasing and (almost everywhere) continuously differentiable in w_t and in λ_t . Then there exists a non-increasing function $w^R(\lambda)$ mapping $[0, 1]$ into $[v, \bar{n}]$, and an (a.e.) continuous function $\lambda^R(w)$ mapping $[v, \bar{n}]$ into $[0, 1]$ such that (as depicted in Figure 2):*

- (a) $\lambda_{t+1} > (=, <) \lambda_t$ according as $\lambda_t < (=, >) \lambda^R(w_t)$;
- (b) $w_{t+1} > (=, <) w_t$ according as $w_t < (=, >) w^R(\lambda_t)$;
- (c) both functions pass through the steady state λ^*, w^* ;
- (d) the function $\lambda^R(w)$ is nondecreasing at w if $L^R(w, \lambda)$ is nondecreasing in w .

Proof. The rational expectations dynamics are given by

$$\lambda_{t+1} = \tilde{\lambda}(w_{t+1}; w_t, \lambda_t) = \tilde{\lambda}(Q^R(w_t, \lambda_t); w_t, \lambda_t) \equiv L^R(w_t, \lambda_t)$$

and

$$w_{t+1} = Q^R(w_t, \lambda_t).$$

¹⁶Existence is ensured by the fact that for any w, λ , the function $\tilde{q}(\cdot; w, \lambda)$ maps $[v, \bar{n}]$ into itself continuously. Since utility functions and the distribution functions are C^1 functions, \tilde{q} is C^1 . Standard arguments imply that for a generic set of values of w, λ , the function $\tilde{q}(\cdot; w, \lambda)$ will have a finite number of equilibria that are locally stable and locally C^1 . The Implicit Function Theorem ensures each locally stable equilibrium is locally decreasing. Next, note that \tilde{q} approaches \bar{n} as w_{t+1}^e approaches v , the lowest fixed point must be locally stable. The highest fixed point must also be locally stable, since \tilde{q} is bounded away from \bar{n} as w_{t+1}^e approaches \bar{n} . Since an increase in w or λ causes the function \tilde{q} to shift downwards, the highest or lowest fixed point must be everywhere decreasing.

Fix any $\lambda \in [0, 1]$. Then $Q^R(\cdot; \lambda)$ is decreasing and (a.e.) C^1 . Define

$$w^R(\lambda) \equiv \sup\{w | Q^R(w, \lambda) \geq w\}$$

whence (b) follows. If Q^R is continuous in w at $w^R(\lambda)$ then $w^R(\lambda)$ must be the fixed point of $Q^R(\cdot; \lambda)$. In that case it is evident that $w^R(\cdot)$ is decreasing at λ . If Q^R jumps downward at $w^R(\lambda)$ then $Q^R(w, \lambda) < w$ in a left neighborhood of $w^R(\lambda)$ and $Q^R(w, \lambda) > w$ in a right neighborhood of $w^R(\lambda)$. If there exist $\hat{\lambda}$ and $\tilde{\lambda} > \hat{\lambda}$ such that $\tilde{w} \equiv w^R(\tilde{\lambda}) > \hat{w} \equiv w^R(\hat{\lambda})$ then there exist $\epsilon, \delta > 0$ such that $Q^R(\tilde{w} - \epsilon, \tilde{\lambda}) > \tilde{w} - \epsilon > \hat{w} + \delta > Q^R(\hat{w} + \delta, \hat{\lambda})$, contradicting the fact that Q^R is decreasing.

Next, note that $\lambda' > \lambda$ implies $n^R(Q^R(w, \lambda'), w) > n^R(Q^R(w, \lambda), w)$ and $n^P(Q^R(w, \lambda')) > n^P(Q^R(w, \lambda))$. Therefore

$$L^R(w, \lambda') - L^R(w, \lambda) < [\lambda' - \lambda][F(n^P(Q^R(w, \lambda))) - F(n^R(Q^R(w, \lambda)), w)] < \lambda' - \lambda. \quad (15)$$

This implies $L^R(w, \cdot)$ has at most one fixed point. Now define

$$\lambda^R(w) \equiv \sup\{\lambda | L^R(w, \lambda) \geq \lambda\}$$

Then for all $\lambda \leq \lambda^R(w)$ we have $L^R(w, \lambda) \geq \lambda$, while for λ in a right neighborhood of $\lambda^R(w)$ we have $L^R(w, \lambda) < \lambda$. Property (15) then implies that $L^R(w, \lambda) < \lambda$ for all $\lambda > \lambda^R(w)$. This establishes (a).

(c) follows (a) and (b). Finally, for (d), if $L^R(\cdot, \lambda)$ is increasing in w then $w' > w$ implies $L^R(w', \lambda) \geq L^R(w, \lambda) \geq \lambda$ for all $\lambda < \lambda^R(w)$, implying that $\lambda^R(w') \geq \lambda^R(w)$. ■

We thus obtain qualitatively similar dynamics with rational expectations. The difference from static expectations is that the threshold function $w^R(\lambda)$ dividing the space between states where wages are rising and where they are falling, is now a nonincreasing function in general. On the other hand, the threshold $\lambda^R(w)$ defining the condition for λ to increase, cannot be guaranteed in general to be upward sloping. The reason is that an increase in w_t raises the supply of skilled people from skilled households, lowering w_{t+1} . *This causes the supply of skilled people from unskilled households to decrease: n^P rises.* This is in contrast to the case of static expectations, where the supply from both types of households

increase with higher w , since everyone expects the current wage next period. With rational expectations it is therefore possible that increasing the skilled wage at t lowers the aggregate supply of skilled people at $t + 1$. Then the $\lambda^R(w)$ locus could be downward sloping.

6 Numerical Analysis

Numerical solutions for equilibrium dynamics can be computed with static expectations for specific utility functions and ability distributions. These permit us to check convergence to steady state, verify theoretical results concerning skill premia dynamics, and examine the impact of shifts in parameters and technology specification.

Figure 3(a) presents the equilibrium dynamic for the skilled wage corresponding to logarithmic utility (for both U and V), uniform ability distribution on $[0, 1]$, education cost $x(n) = 1 - n$, $v = 0.1$ and initial values $w(0) = 0.8, \lambda(0) = 0.01$. A Kuznets pattern is evident: both the wage and skill ratio rise initially. Then the skilled wage falls while the skill ratio rises, converging thereafter to a steady state (in the sense that all the trajectories plotted include up to around 15-20 observations where the differences between subsequent observations is zero up to five decimal points). Figures 3(b)–(d) successively vary the value of v to 0.2, 0.3 and 0.4. This lowers the initial skill premium, and causes the Kuznets pattern to be replaced by a dynamic whereby the skill premium declines monotonically while the skill proportion rises. The steady state skill ratio is not much affected, while the steady state skilled wage is lowered progressively. This owes to the greater scope for upward mobility allowed by a higher wage in the traditional sector, which motivates less able individuals to migrate to the modern sector.

Figure 4 shows the effects of lowering the initial level of the skilled wage $w(0)$ to 0.6, while keeping other parameters the same as in Figure 3. This lowers the motivation of parents to educate their children, raising the ability thresholds in both sectors, and causing a steeper initial rise in the skilled wage. The skilled wage in generation 1 is now higher than in Figure 3. This causes a steeper fall in the skilled wage from generation 1 to 2, as parents are now more motivated to educate their children, and those in the modern sector

are less credit-constrained. Hence the Kuznets pattern is more pronounced if the skilled wage is lower at the outset. The process converges eventually to the same steady state.

Figure 5 explores the effect of widening the support of the ability distribution to $[0, M]$, while preserving the parameter values of Figure 4. The education cost is now $x(n) = M - n$. Figure 5(a) reproduces Figure 4, i.e., sets $M = 1$, while parts (b-d) raise M to 1.5, 2 and 2.5 respectively. The Kuznets pattern continues to obtain, though convergence now occurs to a higher steady state skilled wage and a lower skill proportion. The long run wage is higher because the economy now has more productive individuals relative to education costs at the upper end of the ability distribution. The steady state skill ratio is lower because individuals at the lower end of the ability distribution (who have the same ability) now have higher education costs.

In all of the previous cases, the first phase of the Kuznets pattern where the skill premium and skill ratio rise at the same time lasts only for one period, while the second phase operates for all successive periods. Even if the economy starts in Region I (where the social return to education exceeds the private return) it seems to spend a negligible proportion of time in the long run in that region. This may owe to the lack of a realistic age structure in the model. We now explore the implications of more realistic demographic patterns.

Consider the following extension of the model. Any given cohort works for K periods. A date t cohort is educated at $t - 1$, starts working at t , and works until $t + K$. The parent of cohort t belongs to cohort $t - T$, so T is the age gap between parents and children. The proportion of cohort t that becomes educated depends on the wage of their parent and on the wage at $t - 1$ (the latter representing the wage they expect in their lifetime):

$$\lambda_t^c = \lambda_{t-T}^c [1 - F(n(w_{t-1}, w_{t-T}))] + (1 - \lambda_{t-T}^c) [1 - F(n(w_{t-1}, v))]$$

All cohorts are equal in size, so the workforce size is constant. The proportion of the entire economy's workforce that is skilled at t is then given by

$$\lambda_t = \sum_{k=0}^K \lambda_{t-k}^c$$

Assuming that employers cannot discriminate by age, the wage at t equals

$$w_t = \frac{1}{\sum_{k=0}^K \lambda_{t-k}^c} \sum_{k=0}^K \lambda_{t-k}^c [1 - F(n(w_{t-k-1}, w_{t-T-k}))] m(n(w_{t-k-1}, w_{t-T-k})) \\ + (1 - \lambda_{t-k}^c) [1 - F(n(w_{t-k-1}, v))] m(n(w_{t-k-1}, v))$$

The equilibrium sequence can now be recursively computed.

Figure 6 presents computations of the equilibrium dynamics where we set $K = 5$, and initial values of the skilled wage and skill ratio for periods 0–4 are 0.65 and 0.3 respectively, while λ_c is set at 0.041. The ability distribution is uniform on $[0, 1]$, and values of v are varied from 0.05 to 0.1, 0.2 and 0.25 in parts (a–d). In part (a) we see a fall in the skilled wage, representing a movement from Region II to Region I. The following four periods we see the first phase of the Kuznets pattern. In parts (b–d) the first phase lasts the first five periods. Thereafter the second phase operates.

Finally Figure 7 considers the effect of imperfect substitutability between skilled and unskilled labor, as well as allowing the quality of unskilled labor to vary. The production function now has constant elasticity of substitution between efficiency units of skilled and unskilled labor. Efficiency units of either kind of labor is obtained by weighting proportions of the labor force in each category by their average ability. The ability distribution is uniform as before on $[0, 1]$. Initial values of the skilled wage, unskilled wage and skill ratio are set at 0.3, 0.1 and 0.01 respectively. The skilled (resp. unskilled) wage is calculated by multiplying the average ability of skilled (resp. unskilled) workers by the marginal product of skilled (resp. unskilled) work. Dynamics for four values of elasticity of substitution are shown in parts (a–d). Raising the elasticity of substitution prolongs the duration of the first phase of the Kuznets pattern: in (d) the second phase does not appear at all. It also causes the steady state skill ratio and skilled wage to fall, a natural consequence of the increasing ability of firms to substitute skilled with unskilled labor.

7 Related Literature

The model of this paper is most closely related to models of human capital accumulation or occupational choice with credit market imperfections (Ray (1990, 2006), Banerjee and

Newman (1993) Galor and Zeira (1993), Ljungqvist (1993), Freeman (1996), Mookherjee and Ray (2003), Mookherjee and Napel (2007)). These papers focus on the implications of credit market imperfections, and abstract from signaling distortions in labor markets or in occupational choice. Our model can be viewed as a natural extension of this literature to incorporate signaling problems. In terms of results, one distinction is the lack of long run history dependence (in the sense of multiple steady states) in our model, whereas most of the previous literature emphasizes history dependence. Our focus is thus on non-steady state dynamics, which is more complicated owing to the need to keep track of both quantity and composition of the educated labor force. Few of the earlier models focus on non-steady-state dynamics (e.g., Ray (1990, 2006), Galor and Zeira (1993), Mookherjee and Napel (2007)), where conditions for convergence to steady state are investigated. We are unable to provide convergence conditions in general, but in our simulations the dynamics have always converged. A key distinction from the earlier literature is the nature of non-steady-state dynamics: in all preceding models wages equal marginal products, implying that skill premia decline in the process of development – rendering them incapable of generating co-movements of skill premia and ratios, or a Kuznets pattern.

The role of education screening for the analysis of income inequality and education policies has been explored by a number of papers, in particular Stiglitz (1975) and Lang (1994). Stiglitz (1975) was the first to study the implications of screening for inequality and the allocation of resources to education. His paper focuses on the determinants of over-investment effects in a static setting. Lang (1994) discusses the implications of the human capital *vis-a-vis* signaling debate for development policy, in the context of a static signaling model.¹⁷ Both these papers abstract from credit market imperfections and stay within a static setting.

¹⁷The main point argued by this paper is that it is incorrect to argue that a greater extent of imperfect information in the labor market among employers should increase ability bias. The argument is that both theories of human capital (where education raises productivity, and individuals of greater innate ability acquire more education) and imperfect information (where education need not raise productivity, and more able individuals acquire more education to signal their ability to employers) create ability bias, as part of the measured return to education is the effect of unmeasured innate ability. However, the former creates a greater ability bias than the latter.

The most closely related paper to ours is Hendel, Shapiro and Willen (2005), which studies the effect of combining credit constraints with educational signaling for skill premia. The main point of their paper is that expanding educational subsidies can increase skill premia, since they enable high ability individuals from poor backgrounds to acquire education, which lowers the wages of the uneducated. Hence skill premia and the skill ratio in the economy can move in the same direction, one of the results of our paper. However their model cannot allow any over-investment owing to a number of restrictive assumptions: there are two ability types in the population, and low ability types cannot obtain education regardless of initial wealth. Whether high ability types acquire education depends on their family wealth. We consider a more general model where over-investment can arise owing to signaling distortions, so as to examine the interaction between signaling and credit market imperfections, and implications of these for ability bias or Kuznets patterns of skill premia.

A different body of literature less directly related to our paper deals with explanations of recent evolution of skill premia in advanced countries. This literature (surveyed in Acemoglu (2002)) focuses mainly on skill biased technological change, technology- skill complementarities, and effects of international trade on the demand for skill in the labor market. Relatively little attention is devoted to the dynamics of skill supply, with the exception of Goldin and Katz (2007) who stress the role of slowing of skill supply in explaining the rise in skill premia over the past three decades in the US. We have already discussed this literature in Section 2.

Acemoglu (2002, pp.65-68) argues that changing composition of skilled workers cannot explain rising skill premia. His theoretical argument implicitly assumes a perfect capital market, whence there is a single threshold for unobserved ability for acquiring education. In such a context, a rising supply of skills is accompanied by lowering average ability and wage of both skilled and unskilled workers. Our model demonstrates that with capital market imperfections, there are two thresholds corresponding to whether the corresponding parent is skilled or unskilled. Upward mobility of children from unskilled backgrounds can then cause average ability of the skilled to rise, while that of the unskilled falls or remains the same.

In contrast, Regev (2007) provides a static signaling model where changing composition of skilled workers can explain rising skill premia at the same time that skill ratios are rising.¹⁸ Capital market imperfections or dynamic considerations play no role in this paper. However, it contains a discussion of recent empirical evidence concerning the importance of compositional effects in explaining skill premia in the US.

8 Normative Implications

The heterogeneity of ability combined with credit constraints implies that competitive equilibria are not efficient: children of a given ability but coming from families of varying wealths have different rates of return to their education. This is manifested by the differing ability thresholds for education in poor and rich families. The social return from education is greater for children from poorer backgrounds, as they are more able on average. With perfect credit markets these differences would vanish.

These imperfections presume that employers cannot identify the family background of their employees. Or even if they could they may not be allowed to discriminate on that basis. Otherwise they would offer higher wages to those from poorer backgrounds, controlling for educational status. This would tend to equalize rates of return between workers of varying family backgrounds.

Whether a social planner can engineer efficiency improvements depends on the state of the economy, and the extent to which they can target educational interventions on the basis of the background of agents. Detailed policy implications need to be investigated in future research. But we can offer a few conjectures.

Out of steady state the economy is typically prone to a divergence between social and private rates of return to education. If economies follow the Kuznets type of development

¹⁸In this model, education is equally costly for high and low ability workers. Employers are able to learn worker abilities to some extent, causing more able workers to perceive a higher return to education. But employers do not learn ability perfectly, so education still has signaling value. In this context a rise in college costs can cause (owing to strategic interactions between education decisions of different ability individuals) a rise in the proportion of individuals that acquire education, as well as a rise in the skilled wage.

path for the reasons described by our model, uniform (non-targeted) educational policies should subsidize education at early stages and tax it at later stages. As the economy converges to steady state the scope for such policies could vanish.

Educational policies should however be targeted on the basis of family background. On both efficiency and equity grounds policies would presumably aim to equalize educational opportunity: induce greater investments among the poor and discourage them among the rich. Pareto improvements could possibly be realized by providing public educational loans designed to appeal only to poor families with able children. These could be financed by government bonds that appeal only to rich families with less able children, who would prefer to transfer them financial rather than human capital bequests. Such interventions mirror absent credit market transactions in which rich parents of low ability children lend to poor parents of high ability children. Such credit markets may be absent owing to problems of enforcing loan repayment. If the government is better able to enforce repayments than are private lenders, such policies may be Pareto improving. Alternatively, if planners aim to maximize a social welfare function which positively values average per capita consumption and its distribution, provision of educational subsidies to the poor may be financed by taxes on incomes of the rich. The welfare effects of equalizing educational opportunity in the setting of this model need to be explored in future work.

REFERENCES

- Acemoglu D.** (2002), “Technical Change, Inequality and the Labor Market,” *Journal of Economic Literature*, XL (March 2002), 7–72.
- Acemoglu D.** (2003), “Patterns of Skill Premia,” *Review of Economic Studies*, 70(2), (April 2003), 199–230.
- Banerjee, A. and A. F. Newman** (1993), “Occupational Choice and the Process of Development ,” *Journal of Political Economy*, 101(2), (April 1993), 274–298.
- Becker G. and Tomes, N.** (1979), “An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility,” *Journal of Political Economy*, 87(6), 1153–89.
- Card D.** (2001), Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems. *Econometrica* 69 (September 2001).
- Freeman, S.** (1996), “Equilibrium Income Inequality among Identical Agents.” *Journal of Political Economy* October 1996, 104(5), pp. 1047–1064.
- Galor, O. and Zeira, J.** (1993), “Income Distribution and Macroeconomics.” *Review of Economic Studies* January 1993, 60(1), pp. 35–52.
- Goldin C. and L. Katz** (2007), “Long-Run Changes in the Wage Structure: Narrowing, Widening, Polarizing,” *Brookings Papers on Economic Activity*, 2:2007, 135–165.
- Hendel I., Shapiro J. and P. Willen** (2005), “Educational Opportunity and Income Inequality,” *Journal of Public Economics*, June 2005, 89(5/6), 841–870.
- Lang K.** (1994), “Does the Human Capital/Educational-Sorting Debate Matter for Development Policy?,” *American Economic Review*, March 1994, 84(1), 353–358.
- Ljungqvist, L.** (1993), “Economic Underdevelopment: The Case of Missing Market for Human Capital,” *Journal of Development Economics*, 40, 219–239.
- Loury, G.** (1981), “Intergenerational Transfers and the Distribution of Earnings.” *Econometrica* July 1981, 49(4), pp. 843–867.
- Margo R.** (2000), *Wages and Labor Markets in the United States, 1820–1860*, University of Chicago Press: Chicago and London.

- Margo R. and G. Villaflor** (1987), “The Growth of Wages in Antebellum America: New Evidence?” *Journal of Economic History* 47 (December 1987): 873-895.
- Mookherjee D. and S. Napel** (2007), “Intergenerational Mobility and Macroeconomic History Dependence,” *Journal of Economic Theory*, 137(1), Nov 2007, 49-78
- Mookherjee, D. and D. Ray** (2002), “Is Equality Stable?,” *American Economic Review* 92 (Papers and Proceedings), 253–259.
- Mookherjee, D. and D. Ray** (2003), “Persistent Inequality,” *Review of Economic Studies*, 70(2), April 2003, 369-394.
- Ray, D.** (1990), “Income Distribution and Macroeconomic Behavior,”. Mimeo, Boston University.
- Ray D.** (2006), ‘On the Dynamics of Inequality,’ *Economic Theory* 29, 291–306.
- Regev T.** (2007), “Imperfect Information, Self-Selection and the Market for Higher Education,” working paper 2007-18, Federal Reserve Bank of San Francisco.
- Spence M.**(1974) *Market Signaling: Information Transfer in Hiring and Related Screening Processes*, Cambridge MA: Harvard University Press.
- Stiglitz, J.E.** (1975), “The Theory of ”Screening”, Education, and the Distribution of Income,”. *American Economic Review*, 65(3), June 1975, 283-300.
- Williamson J.** (1985), *Did British Capitalism Breed Inequality?*, Allen and Unwin: Boston and London.
- Williamson J. and P. Lindert** (1980), *American Inequality: A Macroeconomic History*, Academic Press: New York and London.

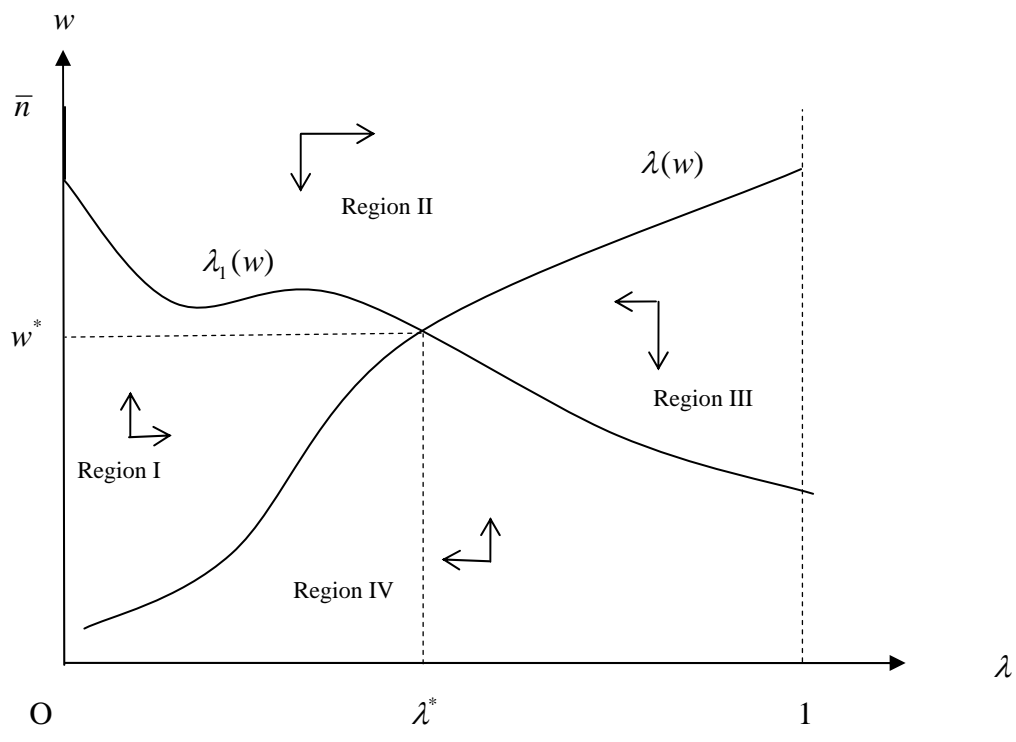


Figure 1: Static Expectations Dynamics

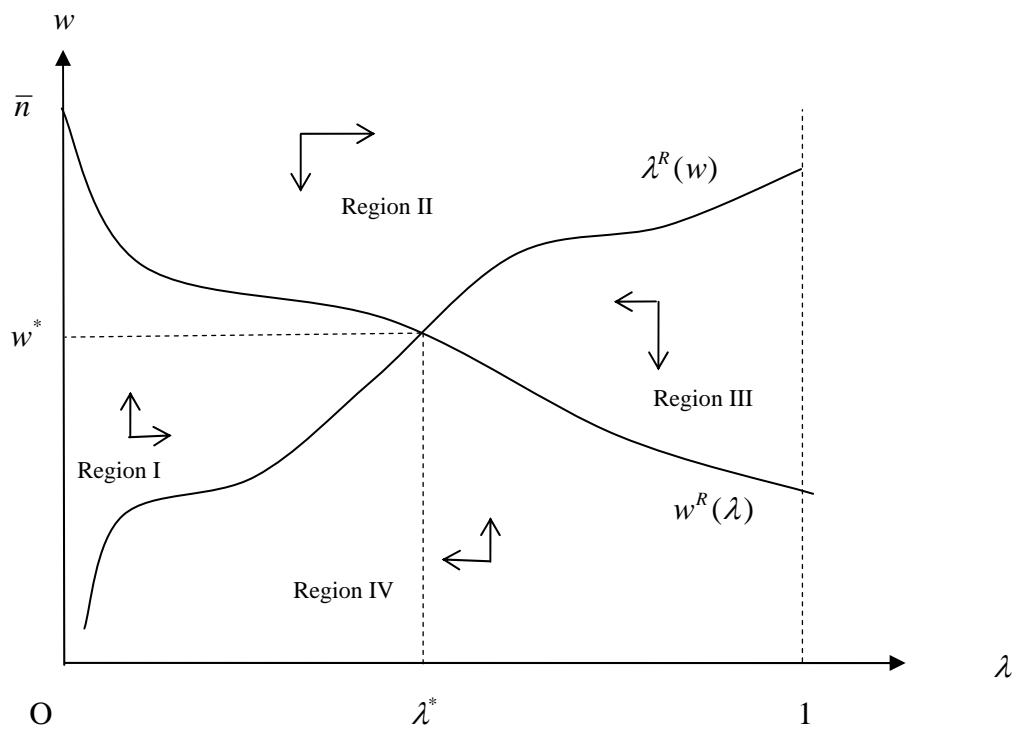
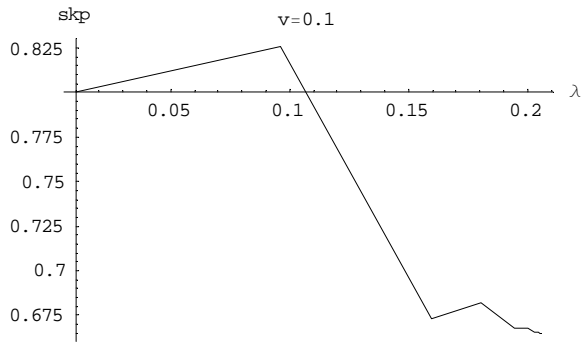
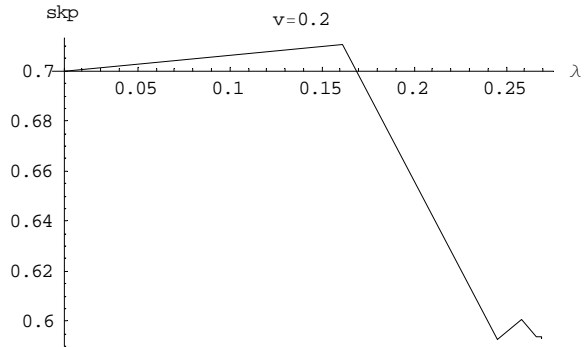


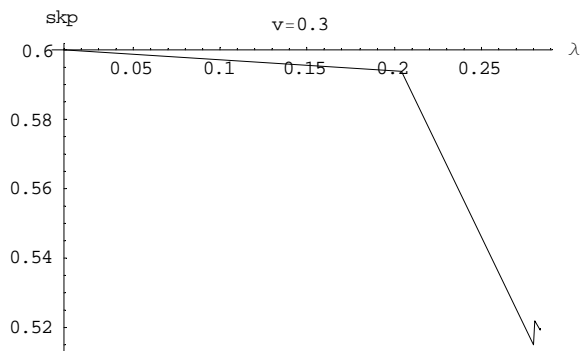
Figure 2: Rational Expectations Dynamics



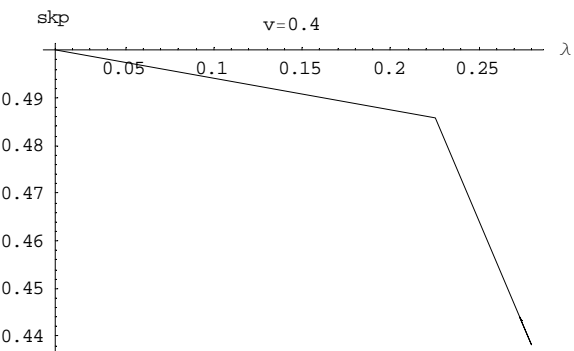
(a)



(b)



(c)



(d)

Figure 3.
 Uniform distribution on $[0,1]$, log utility, CRS; $w(0) = 0.9, \lambda(0) = 0.01$.
 (a) $v=0.1$, (b) $v=0.2$, (c) $v=0.3$, (d) $v=0.4$.

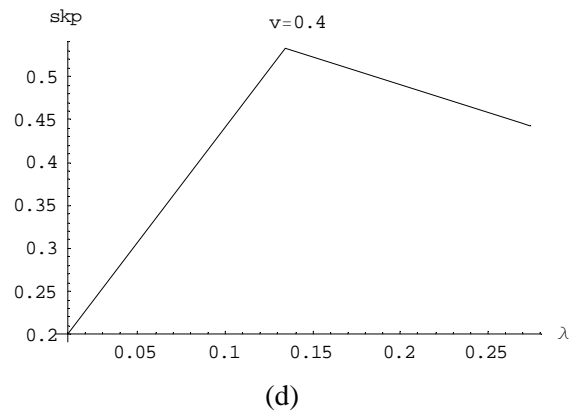
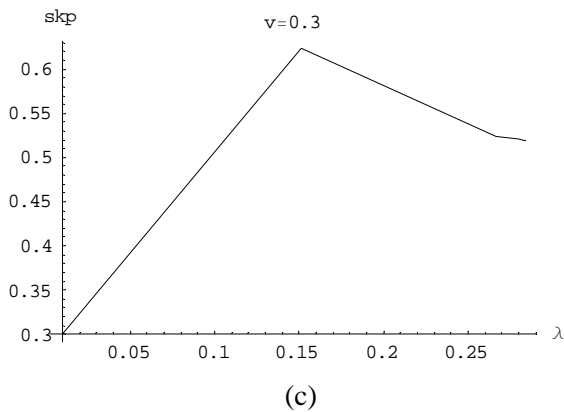
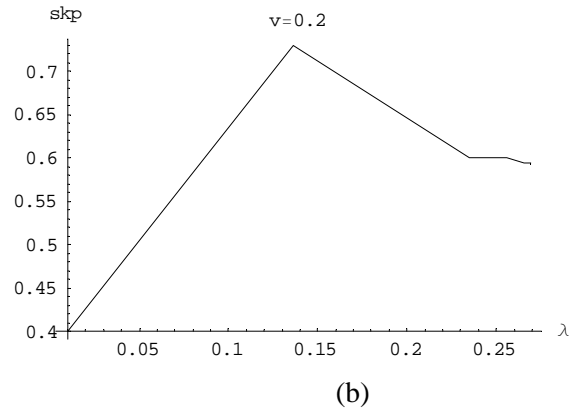
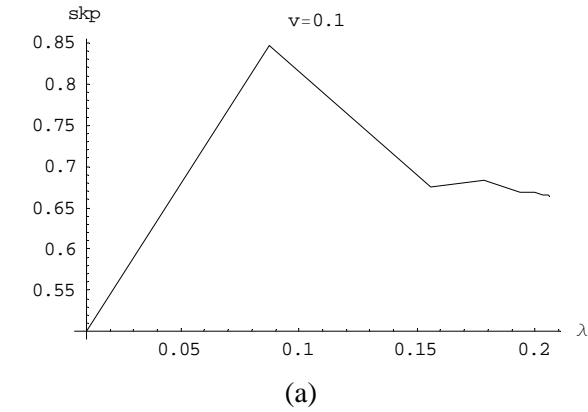


Figure 4.
 Uniform distribution on $[0,1]$, log utility, CRS; $w(0) = 0.6, \lambda(0) = 0.01$.
 (a) $v=0.1$, (b) $v=0.2$, (c) $v=0.3$, (d) $v=0.4$.

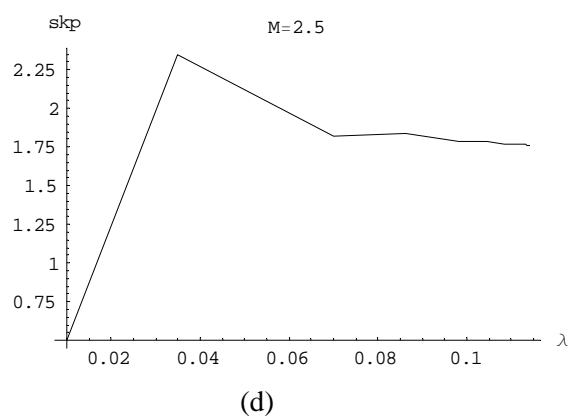
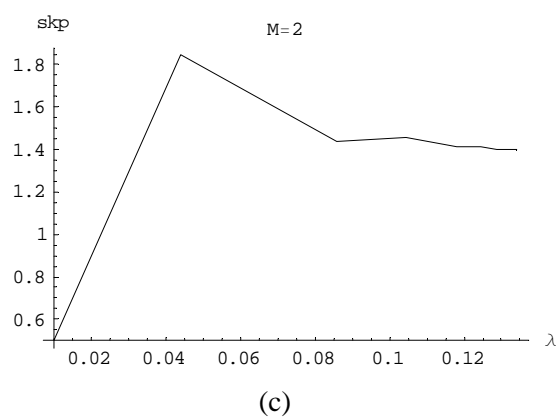
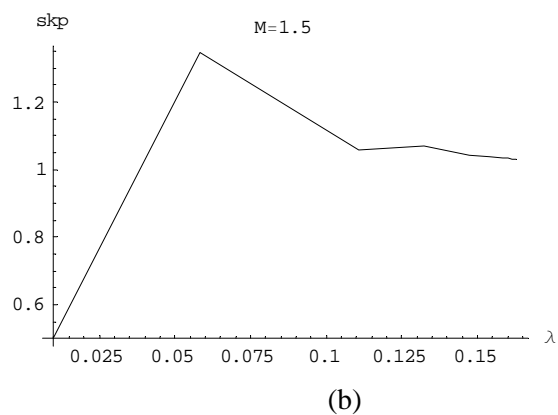
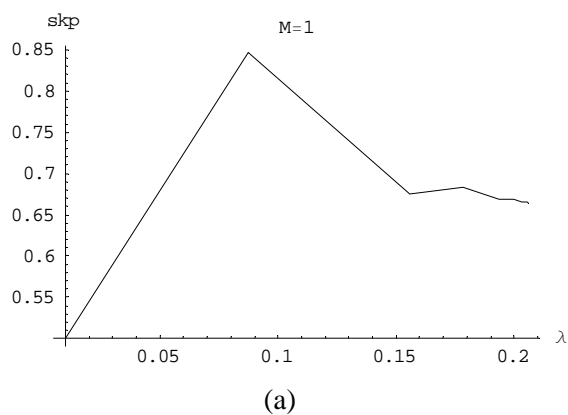


Figure 5.
 Uniform distribution on $[0, M]$, log utility, CRS; $w(0) = 0.6$, $\lambda(0) = 0.01$, $v = 0.1$.
 (a) $M=1$, (b) $M=0.2$, (c) $M=0.3$, (d) $M=0.4$.

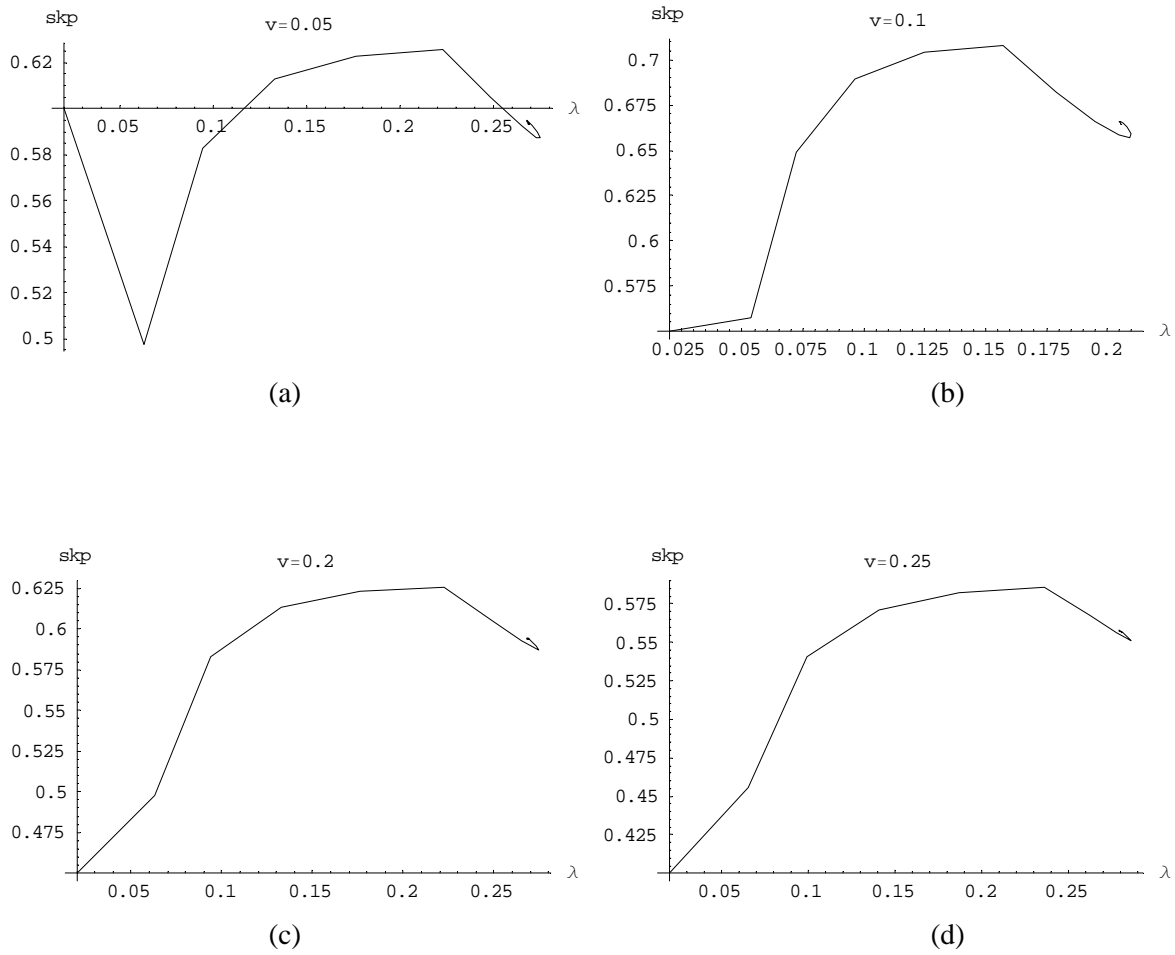


Figure 6.

Slowed down dynamics with 5 periods of working life per cohort, static expectations, Uniform distribution on $[0,1]$, log utility, CRS; $w(0) = \dots = w(4) = 0.65$, $\lambda_c(0) = \dots = \lambda_c(4) = 0.041$, $\lambda(0) = \dots = \lambda(4) = 0.1$.

(a) $v=0.05$, (b) $v=0.1$, (c) $v=0.2$, (d) $v=0.25$.

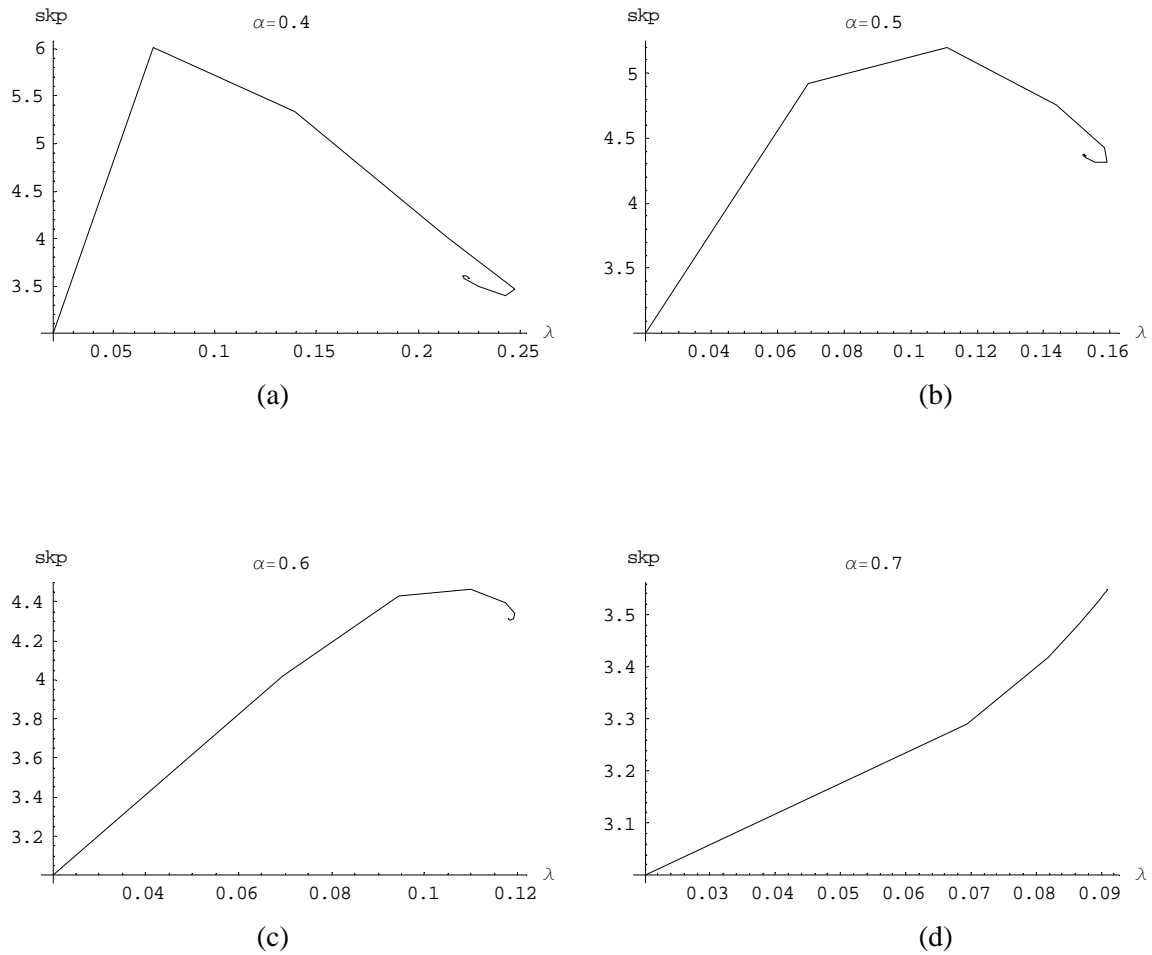


Figure 7.

CES production function $y = [(A_h H)^\alpha + (A_l L)^\alpha]^{1/\alpha}$. Static expectations, Uniform distribution on $[0,1]$, log utility, CRS; A_h : average ability in the skilled sector, $A_l=0.1$, $w(0) = 0.3$, $v(0)=0.1$, $\lambda(0) = 0.01$.
 (a) $\alpha=0.4$, (b) $\alpha=0.5$, (c) $\alpha=0.6$, (d) $\alpha=0.7$.