WHEN MORE SELECTION IS WORSE

JERKER DENRELL, CHENGWEI LIU AND GAËL LE MENS

Abstract. We demonstrate a paradox of selection: the average level of skill of survivors of selection may initially increase but eventually decrease. This result occurs in a simple model in which fitness is not frequency dependent, does not depend on the presence of other genes or practices, and there are no delayed effects. The fitness of an agent in any given period equals a skill component plus a noise term. We show that the average skill of survivors eventually decreases when the noise terms in consecutive periods are dependent and are drawn from a distribution with a hazard rate which is eventually decreasing.

1. Introduction

Does selection increase the proportion of the ‘fittest’ animals or firms? In general, this cannot be guaranteed (Wright, 1931; Levins, 1968; Holland, 1975; Nelson & Winter, 1982). When fitness is frequency dependent, a less fit but more common gene or practice can become dominant (Maynard Smith, 1982; Arthur, 1989; Carroll & Harrison, 1994). The fitness of any particular gene or practice may depend on the presence of other genes or practice, creating a ‘rugged’ fitness landscape with multiple local optima in which selection is unlikely to identify the global optima (Wright, 1931; Levinthal, 1997). Selection in most natural settings is myopic and tends to eliminate practices with positive long-term but negative short-term effects (Levinthal & March, 1981). Selection responds to past environmental states and may systematically fail to track a changing environment (Levins, 1968; Nelson & Winter, 1982). Finally, selection only operates upon the genes or practices that are present in the population and due to the huge possibilities of possible genes and practices the ‘fittest’ one may never have entered the population (Alchian, 1950; Holland, 1975).

Here we show that there is another, more basic, reason for why selection does not necessarily increase ‘fitness’. In our model we assume away all of the above complications. Fitness is not frequency dependent, does not depend on the presence of other genes or practices, and there are no delayed effects. Key to our argument is that the fitness consequences of any gene or practice is stochastic. A particular gene or practice may increase fitness on average but the effect varies stochastically. We model this in the simplest possible way: the fitness of an agent is equal to a deterministic component (‘skill’) plus a noise term: \( p_{i,t} = u_i + \varepsilon_{i,t} \). We are interested in whether selection increases average skill (the average value of \( u_i \) after selection). That is, suppose a population is subject to repeated selection. For example, in every period the 10% agents with the lowest fitness are eliminated. Does average skill \( \langle u_t \rangle \) increase over time?

We show that average skill does not necessarily increase over time even in this simple model. Rather, average skill may initially increase but then decrease. Whether average skill increases depends on the shape of the noise distribution and whether noise terms in consecutive periods are dependent or not. Selection increases average skill whenever the noise term is drawn from a distribution with a

\[ Date: \] January 12, 2016.
increasing hazard rate. When the hazard rate is not increasing, and the noise terms are correlated, the average skill of survivors may eventually decrease, implying that 'more' selection is 'worse'.

2. Set-up

We consider selection operating on a cohort of $n$ agents (an agent can be an individual, a firm etc) where $n$ is large. Selection works as follows. In each period the agents with the $w$ percent lowest 'performances' are removed from the population. There is no replacement; our focus is on the survivors of repeated selection. The performance of agent $i$ in period $t$, $p_{i,t}$, equals her skill, $u_i$ plus a noise or 'error' term, $\varepsilon_i$: $p_{i,t} = u_i + \varepsilon_{i,t}$. There are only two levels of skill: high ($u_i = 1$) and low ($u_i = 0$). Each agent is equally likely to have high or low skill. Let $I_{i,t}$ be an indicator variable equal to one if agent $i$ has survived during all periods 1, 2, ..., $t$ and equal to zero otherwise. We are interested in how $E[u_i|I_{i,t} = 1]$ varies with $t$. Does $E[u_i|I_{i,t} = 1]$ increase with $t$ implying that selection increases the proportion of agents with high skill?

3. Independent Error

Suppose the error terms are independent across agents and periods. That is, $p_{i,t} = u_i + \varepsilon_{i,t}$ and $\varepsilon_{i,t}$ are iid draws for all $t$ and $i$. Specifically, $\varepsilon_{i,t}$ are independently drawn, for each agent and period, from a continuous density $f(x)$ with support $(a, b)$.

The assumption of independent error terms implies that the performance distribution conditional upon skill, $f(p_{i,t}|u_i = k)$, is identical over time:

$$f(p_{i,t+1}|u_i = k) = f(p_{i,t}|u_i = k).$$

The reason is that the error terms are redrawn in every period. It follows that high skilled agents (whose performances equal $p_{i,t} = 1 + \varepsilon_{i,t}$) are, in every period, more likely to have have a higher performance than low skilled agents are (whose performances equal $p_{i,t} = \varepsilon_{i,t}$). As a result, the proportion of high skilled agents increases over time. To illustrate this, Figure 1 shows how the average skill of survivors changes over time when the noise term is drawn from A) a normal distribution and B) a t-distribution (with one degree of freedom). The student's t distribution is bell-shaped and symmetric around zero, like the normal distribution, but has 'heavier' tails. In both cases average skill increases over time. Theorem 1 demonstrates that this result holds for any (continuous) distribution of the error term:

**Theorem 1.** Suppose $p_{i,t} = u_i + \varepsilon_{i,t}$, where $\varepsilon_{i,t}$ are iid draws for all $t$ and $i$. Whenever $w \in (0, 1)$ expected skill increases over time: $E[u_i|I_{i,t+1} = 1] > E[u_i|I_{i,t} = 1]$ for all $t \geq 1$.

**Proof.** Let $\pi_t$ be the proportion of high skilled agents at the end of period $t$, i.e., $\pi_t = E[u_i|I_{i,t} = 1]$. Let $\pi_{t+1}$ be the minimum performance of the agents that survive during period $t + 1$. By Bayes rule,

$$\pi_{t+1} = \frac{\pi_t[1 - P(1 + \varepsilon_{i,t+1} < p_{t+1}^*)]}{\pi_t[1 - P(1 + \varepsilon_{i,t+1} < p_{t+1}^*)] + (1 - \pi_t)[1 - P(\varepsilon_{i,t+1} < p_{t+1}^*)]}.$$  

It follows that $\pi_{t+1} > \pi_t$ whenever $P(\varepsilon_{i,t+1} < p_{t+1}^*)$, the proportion of low skill agents that fail in period $t$, is larger than $P(1 + \varepsilon_{i,t+1} < p_{t+1}^*)$, the proportion of high skill agents that fail in period $t$. Moreover, $P(\varepsilon_{i,t+1} < p_{t+1}^*)$ is larger than $P(1 + \varepsilon_{i,t+1} < p_{t+1}^*)$ whenever $p_{t+1}^* < b + 1$. Finally, $p_{t+1}^*$ is smaller than $b + 1$ whenever $w \in (0, 1)$ because $p_{t+1}^*$ is the minimum performance of the surviving agents which must be below $b + 1$. $\square$
Figures 1. How average skill changes over time when the noise terms are independent and drawn from A) a normal distribution and B) a t-distribution. Each graph is based on 2 million simulations.

4. Dependent Error

Suppose now that the error terms are dependent across periods. Specifically, the error term in the first period $\varepsilon_{i,1}$ is drawn from a continuous density $f(x)$ with support $(a, b)$. The error terms in subsequent periods, however, are identical to the error term drawn in period one. That is, $\varepsilon_{i,2} = \varepsilon_{i,1}$, $\varepsilon_{i,3} = \varepsilon_{i,1}$, $\varepsilon_{i,4} = \varepsilon_{i,1}$, etc. It follows that the performance of agent $i$ does not vary over time but remains the same: $p_{i,t} = u_i + \varepsilon_{i,1}$ for all periods $t$. This set-up represents a situation of extreme dependency across periods. Such dependency can occur when initial 'luck' has long-term effects.

An important implication of this specification is that the performance distribution conditional upon skill for the survivors changes systematically over time. The reason is that selection in prior periods eliminates agents with low values of $\varepsilon_{i,1}$. Whether selection increases average skill or not then depends on the assumptions made about the shape of the noise term distribution, i.e., it depends on the shape of $f(x)$.

To illustrate this, Figure 2 plots how the average skill level of the survivors change over time when $\varepsilon_{i,1}$ is drawn from A) a normal distribution (with mean zero and variance one) and B) from a student’s t distribution (with mean zero and one degree of freedom) and when $w = 0.1$, i.e., the agents with the ten percent lowest performances are removed in each period.\footnote{Figure 2 is based on numerical computations, rather than simulations. Appendix A shows how the computations were done.} Average skill of survivors increases over time when $\varepsilon_{i,1}$ is drawn from a normal distribution (Figure 2A) but increases initially but eventually decreases when $\varepsilon_{i,1}$ is drawn from a t-distribution (Figure 2B). Hence, when the noise term is drawn from a student’s t distribution ‘more selection’, i.e., selection during more periods is ‘worse’, i.e., reduces the average level of skill of the survivors.

Why does selection increase average skill when the noise term is drawn from a normal distribution while selection eventually decreases average skill when the noise...
When more selection is worse

Figure 2. How average skill changes over time when the noise terms are dependent and drawn from A) a normal distribution and B) a t-distribution.

term is drawn from the student’s t-distribution? The reason is that the hazard rates of these two distributions are qualitatively different. The hazard rate, i.e., \( f(x)/(1 - F(x)) \), of a normal distribution is an increasing function of \( x \) while the hazard rate of the t-distribution initially increases with \( x \) but eventually decreases with \( x \). Theorem 2 shows that whether selection increases average skill or not depends on the hazard rate of the noise term distribution:

**Theorem 2.** Suppose \( p_{i,t} = u_i + \varepsilon_{i,t} \) and \( \varepsilon_{i,t} \) is drawn from a continuous density \( f(x) \) with support \((a,b)\). In this case

i) Expected skill increases as a result of selection during the first period: \( E[u_i | I_{i,1} = 1] > 0.5 \).

ii) The change in expected skill after the first period \((t > 1)\) depends on the shape of the hazard rate of the noise distribution as follows: a) When the noise distribution has a strictly increasing hazard rate, \( E[u_i | I_{i,t} = 1] \) increases with \( t \). b) When the noise distribution has a strictly decreasing hazard rate, \( E[u_i | I_{i,t} = 1] \) eventually decreases with \( t \). c) Let \( r_1(x) \) be the hazard rate for high skill agents and \( r_0(x) \) the hazard rate for low skilled agents. If there exists a \( c^* \) such that \( r_1(c) < r_0(c) \) for all \( c < c^* \) and \( r_1(c) > r_0(c) \) for all \( c > c^* \), \( E[u_i | I_{i,t} = 1] \) eventually decreases with \( t \).

**Proof.** See Appendix B.

Why does the effect of selection depend on the hazard rate of the noise term distribution? To understand why, consider first how the threshold for survival changes over time as a result of selection. Remember that when the error terms are dependent the performance of any surviving agent \( i \) remains the same over time. Because selection eliminates agents with low performance the 'threshold for survival' increases over time. Formally, let \( p_i^* \) be the minimum performance of the
agents that survive during period \( t \). \( p^*_t \) increases over time when agents with low performance are eliminated.

To understand the role of the hazard rate, suppose \( w \), the fraction of agents eliminated, is very small so only a small fraction of agents are eliminated in each period. Which agents are eliminated? In period \( t + 1 \) it will be the agents with a performance just above the threshold for survival in period \( t \), \( p^*_t \). What proportion of the high skilled agents have a performance close to \( p^*_t \)? The proportion of high skilled agents that remains after selection is \( 1 - F_h(p^*_t) \), where \( F_h(x) \) is the cumulative distribution function for high skilled agents. Among these agents \( f_h(p^*_t) \) have a performance close to \( p^*_t \). Thus, among the high skilled agents \( f_h(p^*_t)/(1 - F_h(p^*_t)) \) are eliminated in period \( t + 1 \). This expression is equal to the hazard rate, at \( x = p^*_t \), for the performance distribution of the high skilled agents. A similar argument shows that among the remaining low skilled agents \( f_l(p^*_t)/(1 - F_l(p^*_t)) \) are eliminated in period \( t + 1 \). It follows whether more high or low skilled agents are eliminated depends on whether the hazard rate of the performance distribution for the high skilled agents is higher or lower than the hazard rate of the performance distribution for the low skilled agents.

As Figure 3 shows, when the error term is normally distributed the hazard rate for the high skilled agents is always lower than the hazard rate for the low skilled agent. When the error term is drawn from a \( t \)-distribution, however, the hazard rate for the high skilled agents eventually becomes higher than the hazard rate for the

---

**Figure 3.** The density of performance for low and high skilled agents when the noise distribution is drawn from A) a normal distribution and B) a \( t \)-distribution. The bottom panels show the hazard rate for the high and low skilled agents.
low skilled agents. The implication of such crossing hazard rate functions is that the probability of failure will eventually become higher for high skilled agents than for low skilled agents. To demonstrate this, Figure 4 plots how the probability of failure conditional upon skill and survival, i.e., \( P(p_{i,t} < p^*_t | \forall j < t : p_{i,j} > p^*_j, u_i = k) \), varies over time for high and low skilled agents. When the noise term is drawn from a t-distribution (Figure 4B) the failure probability is eventually higher for high skilled agents. In contrast, when the noise term is drawn from a normal distribution, high skilled agents are, in every period, less likely to fail (Figure 4A).

Theorem 2 implies that the effect of selection, with dependent error terms, depends on whether the hazard rate of the noise term is increasing or decreasing. Table 1 lists several distributions and whether their hazard rates are increasing or decreasing (based on Bagnoli & Bergstrom (2005) and Glaser (1980)). Table 1 shows that selection always increases average skill whenever the noise term is drawn from a uniform distribution, triangular ('tent'-shaped) distribution, normal distribution, the logistic distribution, the extreme value distribution, and the Weibull distribution (density \( k e^{k-1} e^{-\epsilon^k} \)) with parameter \( k > 1 \). Selection eventually decreases average skill, however, for several distributions with 'fatter' tails than the normal distribution, including the the t-distribution, the Cauchy distribution, the log-normal distribution, the inverse gaussian and the Weibull distribution (density \( k e^{k-1} e^{-\epsilon^k} \)) with parameter \( k < 1 \). There exists fat-tailed distributions, however, for which selection never decreases average skill, such as the Laplace distribution which has fatter tails than the normal distribution but which nevertheless does not have a decreasing hazard rate.

Table 1 also implies that average skill will suddenly plateau if the noise terms are drawn from an exponential distribution, with constant hazard rate. To illustrate this, Figure 5 plots how the average skills of survivors change when the noise terms

Figure 4. How the failure proportion varies with the period for high and low skilled agents when the noise term is drawn from A) a normal distribution and B) a t-distribution. Minor oscillations in the graphs are due to numerical imprecision in the computations.
Table 1. Shape of Hazard Rates

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Hazard Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uniform</td>
<td>Increasing</td>
</tr>
<tr>
<td>Triangular</td>
<td>Increasing</td>
</tr>
<tr>
<td>Normal</td>
<td>Increasing</td>
</tr>
<tr>
<td>Logistic</td>
<td>Increasing</td>
</tr>
<tr>
<td>Poisson</td>
<td>Increasing</td>
</tr>
<tr>
<td>Extreme Value</td>
<td>Increasing</td>
</tr>
<tr>
<td>Exponential</td>
<td>Constant</td>
</tr>
<tr>
<td>Laplace</td>
<td>Initially increasing, eventually constant</td>
</tr>
<tr>
<td>Cauchy</td>
<td>Initially increasing, eventually decreasing</td>
</tr>
<tr>
<td>Log-normal</td>
<td>Initially increasing, eventually decreasing</td>
</tr>
<tr>
<td>Inverse Gaussian</td>
<td>Initially increasing, eventually decreasing</td>
</tr>
<tr>
<td>Weibull</td>
<td>Increasing when $k &gt; 1$, decreasing when $k &lt; 1$</td>
</tr>
<tr>
<td>U-shaped ($1.5b(x/b)^2$)</td>
<td>Initially decreasing, eventually increasing</td>
</tr>
<tr>
<td>Beta with parameter $\alpha &lt; 0.8$</td>
<td>Initially decreasing, eventually increasing</td>
</tr>
</tbody>
</table>

Figure 5. How average skill changes over time when the noise term is drawn from an exponential distribution.

are drawn from an exponential distribution with parameter $\lambda = 1$. Initially average skill increases but eventually the increase abruptly stops. It is as if selection stops. A similar plateau occurs if the noise term is drawn from a Laplace distribution, which eventually has a constant hazard rate.

Perhaps even more surprising, average skill may initially decrease and eventually increase. This happens when the hazard rate of the noise distribution is initially decreasing and eventually increasing. Consider, for example, an u-shaped noise distribution with density $1.5b(x/b)^2$ and support $x \in (-b, b)$. Figure 6 shows how average skill evolves over time as a result of selection when $b = 1$ and $w = 0.1$. Average skill increases in the first period, then decreases, reaches a minimum after about six periods after which it increases and reaches 1 (the maximum) after twelve periods.
5. ALTERNATIVE ERROR STRUCTURES

The case of independent and dependent error terms represent two extremes. In reality, errors may contain some dependency. For what types of dependency will selection always increase the proportion of high skilled agents and for what types of dependency may selection eventually decrease the average skill of survivors? It seems difficult to resolve these questions formally. Nevertheless, simulations show that selection may eventually decrease the proportion of high skilled agents for error structures other than the case of complete dependency. Here we consider two simple error structures.

First, suppose errors add up, as in a random walk: \( p_{i,1} = u_i + \varepsilon_{i,1}, \) \( p_{i,2} = u_i + \varepsilon_{i,1} + \varepsilon_{i,2}, \) \( p_{i,3} = u_i + \varepsilon_{i,1} + \varepsilon_{i,2} + \varepsilon_{i,3} \) etc. Generally, \( p_{i,t} = u_i + \sum_{j=1}^{t} \varepsilon_{i,j} \).

Simulations suggest that Theorem 2 extends to this error structure. That is, when the noise distribution has a strictly increasing hazard rate, \( E[u_i | I_{i,t} = 1] \) increases with \( t \) but when the hazard rate of the error term distribution eventually declines, \( E[u_i | I_{i,t} = 1] \) eventually decreases with \( t \). As an illustration, Figure 7 plots how the average skill of survivors changes over time when the noise terms are added up and A) the noise terms are drawn from a normal distribution and B) drawn from a t-distribution with one degree of freedom. As shown, average skill eventually declines when the noise terms are drawn from a t-distribution.

Second, suppose performance follows a modified random walk, as follows: \( p_{i,1} = u_i + \varepsilon_{i,1} \) and \( p_{i,t} = qu_{i}+(1-q)p_{i,t-1}+\varepsilon_{i,j} \) for \( t > 1 \). In this modified random walk the expected value in period \( t \) is \( qu_{i}+(1-q)p_{i,t} \) as compared to \( p_{i,t} \) for the random walk. Thus \( q \in (0, 1) \) represents how much skill, as opposed to past performance, impacts expected future performance. A low value of \( q \) represent strong dependence: average performance depends strongly on past performance. A value of \( q = 0 \) implies that only past performance impact expected future performance (as in section 5). Simulations show that if \( q \) is sufficiently low, results are, unsurprisingly, similar to when performance follows a random walk. For example, Figure 8 shows how average skill varies with the period when \( q = 0.05 \). Average skill eventually decreases when the noise terms are drawn from a t-distribution.
Figure 7. How average skill changes over time when the noise terms are added up and A) the noise term is drawn from a normal distribution and B) drawn from a t-distribution with one degree of freedom. Each graph is based on 2 million simulations.

Figure 8. How average skill changes over time when the performance follows a modified random walk and A) the noise term is drawn from a normal distribution and B) drawn from a t-distribution with one degree of freedom. Each graph is based on 2 million simulations.

6. Discussion: Do the fitter survive?

Does theorem 2 imply that less 'fit' agents are more likely to survive? This conclusion is incorrect for two reasons. First, our results show that the proportion of high skill agents may eventually decrease but the proportion remains above 50%, the initial proportion. Second, and more subtly, depending on how 'fitness' is
defining it could be argued that 'fitter' agents are in fact more likely to survive in each period in our models. To explain this, consider the model in section four in which the error term is drawn only once, in the first period. Survival in period $t+1$ depends on performance in period $t+1$, which equals $p_{i,t} = u_i + \epsilon_i,1$. Theorem 2 shows that the expected value of skill for the survivors may go down. That is, $E[u_i|I_{i,t+1} = 1]$ may be lower than $E[u_i|I_{i,t} = 1]$. This theorem does not imply, however, that expected performance in period $t+1$ will go down. Rather, expected performance increases: $E[p_{i,t}|I_{i,t+1} = 1]$ is higher than $E[u_i|I_{i,t} = 1]$ (because the better performers survive). While survivors have a lower value of $u_i$ in period $t+1$ than in period $t$, they have a higher value of $\epsilon_i,1$. Because survival in the next period depends on performance, and performance depends on $\epsilon_i,1$ in addition to skill, it would be wrong to conclude that the population eventually becomes less 'fit' if fitness is defined as the expected performance of an agent in the next period. If the error term is not redrawn, survivors do not become less fit, even if their average skill declined. Survivors of period $t + 1$ do become less 'fit', however, than the survivors of period $t$ if the error is redrawn, i.e., which can occur if the environment shifts. While it is obvious in evolutionary theory that fitness may decline if the environment change, the change in the environment needed here is a subtle one: a change in the correlation structure of the error terms rather than a change in the direction of selection.

7. Implications

Theorem 2 implies that more selection, in the sense of selection during more periods, can lead to a population of survivors with a lower expected level of skill. Lemma 1 (see the appendix) also shows that 'tougher' selection, in the sense of having a higher threshold for survival, can lead to a population of survivors with a lower expected level of skill. These results have interesting implications for understanding a wide variety of selection systems and for theories that examine the implications of more or less 'tough' selection.

7.1. Career Systems. An up or out system represents a selection system in which the lowest performers are eliminated. The results in this paper show that the survivors of such a system, i.e., the individuals promoted to the highest levels, can be worse than the individuals who do not reach the highest levels. This happens if a) performances across periods are dependent and b) the noise or 'luck' term is drawn from a heavy-tailed distribution such as the log-normal or the t-distribution.

7.2. Competition and Density Delay. The threshold for survival may increase if there is more competition. If only a fixed number of agents can survive, the threshold for survival increases when there are more actors around (Barnett et al., 2003). It has been suggested that actors that survive such a period may be better, because they reached a higher bar. The results in this paper imply that actors that survive during such a period may in fact be worse. This offers a different interpretation of the findings that density at founding has a persistent negative effect on survival rates (Carroll & Hannan, 1989).

References


APPENDIX A: NUMERICAL COMPUTATIONS

To compute $E[u_i|I_{i,t} = 1]$ when the noise distribution is continuous we proceed as follows. Let $c$ be the minimum performance of the agents that survive period $t$. The value of $c$ has to satisfy the equation

\[ 0.5S_1(c) + 0.5S_0(c) = (1 - w)^t \]

Here $S_1(c) = 1 - F_1(c)$ is the proportion of high skill agents ($u_i = 1$) with a performance above $c$, $S_0(c) = 1 - F_0(c)$ is the proportion of low skill agents ($u_i = 0$) with a performance above $c$, and $(1 - w)^t$ is the proportion of agents that survive period $t$.

Let $c_t$ be the value of $c$ that satisfies the above equation for period $t$ (assuming there is a unique value). The proportion of high skill agents after $t$ periods, i.e., $E[u_i|I_{i,t} = 1]$, can then be expressed as

\[ E[u_i|I_{i,t} = 1] = \frac{S_1(c_t)}{S_1(c_t) + S_0(c_t)}. \]

Hence, to compute $E[u_i|I_{i,t} = 1]$ we only need to find $c_t$ which can be done, for any noise distribution, by grid search (i.e. by trying a range of values for $c_t$).

Note that this method of computing $E[u_i|I_{i,t} = 1]$ assumes that the population is large ($n$ is large) so that the surviving agent with the lowest performance in period $t$ has a performance that equals $c_t$. Note also that if the error distribution were discrete there may be no value of $c$ that solves the above equation.

APPENDIX B: PROOF OF THEOREM 2

Here we characterize the noise distributions for which $E[u_i|I_{i,t} = 1]$ increases with $t$ and for which $E[u_i|I_{i,t} = 1]$ eventually decreases with $t$. We begin by analyzing how $E[u_i|p_i > c]$ varies with $c$.

Let $f(x)$ be the density and $F(x)$ the cumulative density function of performance for actors with low skill ($u_i = 1$). Let $(a, b)$ be the support of $f(x)$. Because $P(u_i + \varepsilon_i > x) = P(\varepsilon_i > x - u_i)$ the cumulative density function of performance for actors with high skill ($u_i = 1$) is $F(x-1)$. Its density is $f(x-1)$ with support $(a + 1, b + 1)$. Let $S_0(x)$ be the survival function for actors with low skill, i.e., $S_0(x) = 1 - F(x)$, and let $S_1(x)$ be the survival function for actors with high skill, i.e., $S_1(x) = 1 - F(x-1)$. Finally, let $r_1(c)$ be the hazard rate for high skill agents, i.e., $r_1(x) = f(x-1)/S(x-1)$. Similarly, $r_0(x) = f(x)/S(x)$ is the hazard rate for low skilled agents.

**Lemma 1.**

\( i \) Whenever $c \in (a, a + 1)$ $E[u_i|p_i > c]$ is an increasing function of $c$.

\( iia \) When $c \in (a + 1, b)$ and $r_1(c) > r_0(c)$ for all $c$ $E[u_i|p_i > c]$ is an increasing function of $c$.

\( iib \) When $c \in (a + 1, b)$ and $r_1(c) > r_0(c)$ for all $c$ $E[u_i|p_i > c]$ is a decreasing function of $c$.

\( iic \) When $c \in (a + 1, b)$ and $r_1(c) = r_0(c)$ for all $c$ $E[u_i|p_i > c]$ is a constant.

\( iid \) When $c \in (a + 1, b)$ and $r_1(c) < r_0(c)$ for all $c$ $E[u_i|p_i > c]$ is an increasing function of $c$ up until $c = c^*$ and a decreasing function thereafter.

\( iii \) Whenever $c \in (b, b + 1)$, $E[u_i|p_i > c] = 1$.

**Proof.** \( i \) Suppose $c \in (a, a + 1)$. In this interval case all high skill agents have a performance above $c$. Bayes rule then implies that $E[u_i|p_i > c]$ equals $1/[1 + S(c)]$ which is increasing in $c$. 

ii) Suppose \( c \in (a + 1, b) \). By Bayes rule we have

\[
E[u_i|p_i > c] = \frac{S_1(c)}{S_1(c) + S_0(c)} = \frac{1}{1 + \frac{S_0(c)}{S_1(c)}},
\]

Hence \( E[u_i|p_i > c] \) is an increasing function of \( c \) whenever \( S_0(c)/S_1(c) \) is a decreasing function of \( c \). Moreover,

\[
\frac{dS_0(c)/S_1(c)}{dc} = -\frac{f_0(c)S_1(c) + f_1(c)S_0(c)}{S_1(c)^2},
\]

which is negative when \( f_1(c)S_0(c) < f_0(c)S_1(c) \), or, \( f_1(c)/S_1(c) < f_0(c)/S_0(c) \), i.e., when \( r_1(c) < r_0(c) \). Similarly, \( E[u_i|p_i > c] \) is decreasing function of \( c \) when \( r_1(c) > r_0(c) \) and \( E[u_i|p_i > c] \) is a constant when \( r_1(c) = r_0(c) \).

iii) Only high skilled agents have a performance in the interval \( (b, b + 1) \). Hence, \( E[u_i|p_i > c] = 1 \).

Lemma 1 shows that, in the region of common support where \( c \in (a + 1, b) \), the shape of the hazard rate determines whether \( E[u_i|p_i > c] \) increases or decreases with \( c \). If the hazard rate of the noise term distribution is increasing for all values, implying that \( r_1(c) < r_0(c) \), then \( E[u_i|p_i > c] \) is increasing in \( c \). If the hazard rate is decreasing, implying that \( r_1(c) > r_0(c) \), \( E[u_i|p_i > c] \) is decreasing in \( c \). If the hazard rate initially increases but eventually declines with \( c \) then \( E[u_i|p_i > c] \) may initially increase and then declines with \( c \). To determine whether \( E[u_i|p_i > c] \) increases with \( c \) is it thus enough to know whether the hazard rate is increasing or decreasing.

Whether the hazard rate is strictly increasing or strictly decreasing can be determined from the shape of the density function as follows:

**Lemma 2.** Let \( g(x) \) be a differentiable density function such that \( g(x) \to 0 \) as \( G(x) \to 1 \). If \( \ln(g(x)) \) is a concave (convex) function of \( x \), then \( r(x) = g(x)/(1 - G(x)) \) is an increasing (decreasing) function of \( x \).

**Proof.** See Thomas (1971).

Using this criterion it can be shown that the hazard rate of the normal distribution has an increasing hazard rate (see, Luce (1986), p. 16-17). It follows that \( E[u_i|p_i > c] \) is an increasing function of \( c \) when the noise term is drawn from a normal distribution.

We next demonstrate that the threshold for survival increases over time.

**Lemma 3.** Suppose the \( w \) percent agents with the lowest level of performance are removed in every period \( t = 1, 2, \ldots \). Let \( p^*_t \) be the minimum performance of the agents that survive period \( t \). When the population is large \( (n \to \infty) \) and the noise distribution is continuous, \( p^*_t \) is an increasing function of \( t \).

**Proof.** In period \( t \) only agents with performances among the highest \( (1 - w)^t \) percent survive. Hence \( p^*_t \) has to satisfy the equation \( 0.5S_1(p^*_t) + 0.5S_0(p^*_t) = (1 - w)^t \). When the noise distribution is a continuous function the value of \( p^*_t \) that solves this equation is a strictly increasing function of \( t \). The threshold for survival hence keeps increasing as long as there are any agents left and when \( n \) goes to infinity there are always agents left.

By combining Lemmas 1 and 3 we can characterize the noise distributions for which \( E[u_i|I_{1,t} = 1] \) always increases with \( t \) and the noise distributions for which \( E[u_i|I_{1,t} = 1] \) eventually decreases with \( t \).
i) Consider the first period. The proportion of high skilled agents who survive the first period, $P(1 + \varepsilon_i > p_i^*)$ is larger than the proportion of low skilled agents who survive the first period, $P(\varepsilon_i > p_i^*)$. Hence, the proportion of high skilled agents is higher than 0.5 after the first period.

ii) Consider periods $t > 1$. a) Suppose the noise distribution has a strictly increasing hazard rate. Note that $E[u_i|I_{i,t} = 1]$ equals $E[u_i|p_i > p_i^*]$ and by Lemma 3 $p_i^*$ is an increasing function of $t$. It follows from Lemma 1 that when the noise distribution has a strictly increasing hazard rate $E[u_i|p_i > p_i^*]$ is an increasing function of $p_i^*$. b) Suppose the noise distribution has a strictly decreasing hazard rate. There exists a period $t$ such that $p_t^* > a + 1$. It follows from Lemma 1 that after this period, $E[u_i|I_{i,t} = 1]$ decreases with $t$. c) Suppose $r_1(c) < r_0(c)$ for all $c < c^*$ and $r_1(c) > r_0(c)$ for all $c > c^*$. Let $t^m$ be the first period in which $p_{t^m}^* > c^*$. Let $t^a$ be the first period in which $p_{t^a}^* > a + 1$. It follows from Lemma 1 that whenever $t > \max(t^m, t^a)$, $E[u_i|I_{i,t} = 1]$ is a decreasing function of $t$. Hence, $E[u_i|I_{i,t} = 1]$ eventually decreases with $t$. 

E-mail address: jdenrell@gmail.com