Unobserved Ability and Entrepreneurship

Deepak Hegde
Stern School of Business, New York University

Justin Tumlinson
Ifo Institute at the University of Munich

Preliminary Draft: Please Do Not Circulate or Cite

Abstract

Why do individuals become entrepreneurs? When do they succeed? We develop a model in which individuals use pedigree (*e.g.*, educational qualifications) as a signal to convince employers of their unobserved ability. However, this signal is imperfect, and individuals who correctly believe their ability is *greater* than their pedigree conveys to employers, choose entrepreneurship. Since ability, not pedigree, matters for productivity, entrepreneurs earn more than employees of the same pedigree. Our preliminary empirical analysis of two separate nationally representative longitudinal samples of individuals residing in the US and the UK supports the model's predictions that (A) Entrepreneurs have higher ability than employees of the same pedigree, (B) Employees have better pedigree than entrepreneurs of the same ability, and (C) Entrepreneurs earn more, on average, than employees of the same pedigree, and their earnings display higher variance. We discuss the implications of our findings for entrepreneurship, education, and public policy.

1 Introduction

Why do individuals become entrepreneurs? When are they successful? Social scientists, across different disciplines, have proposed different explanations. A leading psychological theory predicts that individuals who believe their performance depends on their own actions—those with an internal locus of control—are more likely to become entrepreneurs (McClelland 1964). Camerer and Lovallo (1999) provide experimental evidence that entrepreneurs are characterized by overconfidence and hubris. Sociologists have argued that those who feel "rejected" by the mainstream population, or "social misfits," are more likely to become entrepreneurs (Min 1984). Economists have observed that lower-paid wage workers, who have changed jobs more times than those with comparable attributes, enter self-employment more (Evans and Leighton 1989), and that entrepreneurs may have a distaste for following orders, for which they suffer a cut in their income (Hamilton 2000). Lazear (2004, 2005) suggests that entrepreneurs have more diversified skills than employees. Åstebro, Chen and Thompson (2011) find that a disproportionate number of entrepreneurs have either very high or low incomes compared to the general population and propose that entrepreneurs are drawn from the tails of the ability distribution. Kihlstrom and Laffont (1979) argue that the self-employed are less risk-averse. In summary, the current wisdom suggests that entrepreneurs are exceptional in either their preferences or abilities.

Here we present and test an alternative theory, which explains entrepreneurial choice across the spectrum of occupations—from restaurant owners to technology moguls. Rather than being fundamentally exceptional in abilities or tastes, we argue that an individual of any ability level has incentive to start his own venture if potential employers perceive his productive potential as lower than he does. We develop a formal model in which the labor market (for employment in existing firms) rewards individuals with jobs and wages based on their pedigree.

We define pedigree as the set of an individual's attributes observable to the labor market. Employment, work history and educational attainment, as measured by degrees from universities of varying prestige, are natural components of pedigree. Pedigree is simply the observable signal of unobservable ability. However, this signal is imperfect—if a worker believes his ability is greater than the labor market can infer from his observable characteristics, then he has incentive to choose entrepreneurship, where he will earn according to his ability rather than in traditional employment, where he will earn according to his pedigree.

Our model is driven by (1) asymmetric information about worker ability and (2) the fact that entrepreneurs earn according to their own ability rather than noisy perceptions of it. The former feature distinguishes our theory from a rich literature on job-matching pioneered by Roy (1951) and Jovanovic (1979), where occupational choice is a matter of matching one's multidimensional skill set to a job where those skills are most productive, like fitting

the pieces of a puzzle.[1, 2] In our model, unidimensional ability is general—better employees would make also better entrepreneurs and vice versa. One significant benefit of considering ability generally is that the model's predictions yield more readily to empirical testing, because general ability has been often measured in studies of human capital acquisitions and individual income dynamics (e.g., Angrist and Krueger 1991, Heckman and LaFontaine 2006). We capitalize on this advantage here.

We view–and define–an entrepreneur as an individual who is the residual claimant of his own *ability*. In contrast, wage workers receive (initial) compensation offers based on a limited set of information presented though a CV, a manner of discourse exhibited at an interview, letters of reference and so on, all meant to reflect underlying productivity, but do so imperfectly. While the firm may learn more over time, the muddling effects of teamwork, incentive and organizational structures may continue to obscure employees' true productivity for years. Even if a firm knew employees' productivity perfectly, it is unlikely that it would set wages to match, since it maximizes profits within the constraint that another firm, that can only compensate based on immediate observables, cannot profitably entice employees away. By focusing on a single facet of the entrepreneur, the model can describe occupational choice across the ability and pedigree spectrum: from the corner food vendor lacking a high school diploma to the founder of a revolutionary, biotech startup with a PhD from MIT—something holistic models of entrepreneurship cannot easily do.

Of course, we acknowledge that unobserved ability is not the only reason that drives self-employment. Individuals could become entrepreneurs for several reasons such as love for risky ventures, desire to be one's own boss and overconfidence. However, these explanations, unlike ours, do not yield clear implications for success in entrepreneurship. So long as pedigrees are used as signals, our theory proposes that high-ability individuals select into entrepreneurship *and* earn higher wages relative to comparably pedigreed wage workers. Hence, we also speak to the type of individuals who are most likely to succeed as entrepreneurs.

We formally derive that individuals who choose entrepreneurship have higher ability, higher income and higher income variance than other individuals with the same pedigree, and entrepreneurs have lower pedigrees than those of similar ability who are traditionally employed. We also show that these results hold whether pedigree arises exogenously or is acquired endogenously.

We test the propositions of our model using data drawn from the US *National Longitudinal Survey of Youth* (NLSY79), first administered to those born between 1957 and 1964, and

[1] Jovanovic argues that job turnover occurs as employee and employer simultaneously learn about the employee's job match. His model incorporates *imperfect* information, but there is no *asymmetry* in information between employer and employee.

[2] A recent body of work (e.g., Klepper and Thompson 2010) examines employee-entrepreneurship, and suggests that employees with superior ideas might prefer to exit wage employment and start ther own firms. The exits might partially be driven by disagreement between the employer and employee about the quality of the latter's ideas.

resident in the US in 1979, and the British *National Child Development Study* established in 1958. We find that in both samples, entrepreneurs scored higher on ability tests overall, and have, on average, higher scores conditioned on their educational qualifications. Despite their higher ability scores, entrepreneurs have systematically lower academic qualifications. We also find evidence that entrepreneurs earn more, on average, conditional on pedigree, and that their wages exhibits a higher dispersion. These results appear robust to controlling for a variety of factors including the family wealth and hackgrounds of individuals, their non-cognitive traits (such as risk-taking and locus of control) and other demographic features. Our main results also condition out industry-specific effects, but we also note that empirical evidence for our propositions is strongest in high-value generating industries such as Information Technology, Professional and Technical Services, Finance and Insurance, Wholesale and Retail Trade and Arts and Entertainment. We interpret these findings as suggestive evidence for our theoretical propositions based on asymmetric information.

We also explore an alternative explanation for our findings based on multi-dimensional ability and sorting. If entrepreneurship requires different abilities than employment, in particular, if the comparative advantage of specific skills acquired through formal education *relative* to general ability favor productivity in traditional employment, then individuals with more formal education relative to innate ability will sort into traditional employment, leaving the rest to exercise their natural gifts in entrepreneurship. Then, individuals who choose entrepreneurship will have higher general ability than other individuals with the same pedigree, and entrepreneurs have lower pedigrees than those of similar ability who are traditionally employed—predictions shared by our asymmetric information based explanation. However, we find that among those who are persistently wage employed during the span of our study, the returns to education diminish and returns to ability increases over their careers—patterns consistent with asymmetric information that resolves over employment tenure but difficult to reconcile with the sorting-based explanation. We also find that ability (conditional on educational qualifications) is a particularly strong predictor of the decision to be self-employed early in individuals' careers (when asymmetric information about ability is likely to be high). This, along with our third observation of higher entrepreneurial earnings, as well as higher variance in entrepreneurial earnings, are not easy to reconcile with the sorting-based explanation. Thus, we conclude that asymmetric information about ability is an important driver of entrepreneurship.

The rest of the paper is organized as follows. We develop our theoretical propositions in the next section. We introduce our sample and empirical findings, based on the US longitudinal surveys, in the third section. We deal with the matching-based alternative explanation in the fourth section. The fifth section describes our findings based on the UK longitudinal surveys. The sixth section concludes.

2 Theory

Our theory of entrepreneurial choice may be thought of as an application of Akerlof's (1970) "market for lemons" to labor. In the base model, workers' privately known, general abilities are statistically associated with exogenously assigned, public signals (*i.e.* pedigree).[3] Workers are both the more informed sellers and the traded good, while employers are the less informed buyers. Employers offer wages based on observable pedigrees, but to the extent that better informed workers believe their ability exceeds the labor market's estimate of it, they have an incentive to choose entrepreneurship, where they will be residual claimant of their talents.[4, 5] The result of this asymmetric information driven sorting is that for every pedigree, the most productive workers choose entrepreneurship. To use Akerlof's terminology, conditional on a given set of observables, *cherries* become entrepreneurs and employees are *lemons.* This is our first main theoretical result, and to the extent we can measure productivity that the labor market cannot see, it gives rise to the following empirical hypothesis: *conditional on pedigree, entrepreneurs are more able.*

Since wage offers can only depend on observables, employees bearing the same observables should be paid the same wage. Entrepreneurs with this pedigree believe they could earn more as entrepreneurs. Thus, we empirically predict *conditional on pedigree, entrepreneurs earn more.* Furthermore, since entrepreneurial reward stems directly from ability, and entrepreneurial productivity spans the space from just above the relevant wage offer to the upper limits of productivity, we hypothesize that *conditional on pedigree, entrepreneurial earnings exhibit higher variance.* Note that while we have no reason to reject the common notion that entrepreneurship is risky, venture uncertainty is not required to create extreme variance in entrepreneurial outcomes.

Above, we framed the question of occupational choice as, "Given my entrepreneurial

[3] According to theories of statistical discrimination, *any* observable attribute can become informative in equilibrium. Even observable attributes, which are *ex ante* orthogonal to ability, can indicate unobservable ability *ex post*, due to individuals' collective strategic response to *beliefs* about the information contained in the signal that self-reinforce and become *true in equilibrium* (Tumlinson 2011).

[4] While assuming that workers know their productive potential better than the labor market is reasonable and standard, this should not be a foregone conclusion. Workers and employers learn about worker ability over time but not necessarily at the same rate (Holström 1999). The muddling effects of teamwork and incentives, together with the limitations of monitoring, put employers at a learning disadvantage relative to workers themselves even after a long period of time. We do not model these dynamics, but they should not fundamentally change the model's results—it is agnostic about the timing of the occupational decision.

A related worry is that while workers have better information about themselves, their estimates of own productivity might be biased (Camerer and Lovallo, 1999). The model would suggest that the entrepreneurial ventures of workers, who are not undervalued but only think they are, will be short lived. Our empirical tests show this is the case.

[5] We acknowledge that by narrowly focusing on a single facet (residual claimancy), our model unintentionally describes selection into occupations (like commission based salesmen), which share this facet with, but are not generally considered entrepreneurship.

productivity, what minimum *wage* would I need to keep me from entrepreneurship?" But since different pedigrees induce different wages, pedigrees can be ordered according to their associated wages, and thus every worker could equivalently ask, "Given my entrepreneurial productivity, what minimum *pedigree* would I need to keep me from entrepreneurship?" Hence, at every ability level, there exists a threshold pedigree, such that those whose pedigree exceeds that threshold will receive an acceptable offer; those whose pedigree lies below the threshold will not—they open their own business. So, our theoretical result and empirical hypothesis is that *conditional on ability, entrepreneurs have lower pedigree.*

We extend this base model to comprehend endogenously acquired signals, like educational certification, by building on Spence's (1973) "job-market signaling" work. We show that mild structure on the pedigree acquisition process gives rise to precisely the statistical relationship between ability and signals, which we assume in the base model of exogenous pedigrees. Thus, all of the results of base model hold in one of endogenous pedigree acquisition.

The economics of asymmetric information are, by now, so familiar that our key theoretical results can be presented verbally. However, just as Akerlof's and Spence's models respectively require assumptions to prevent market unraveling and to guarantee the existence of separating equilibria, so does ours. Our formal model explicates these minimum assumptions in the entrepreneurship context. It also facilitates differentiation from an alternative model, driven by multidimensional skill matching rather than asymmetric information, which we present after providing evidence for the empirical implications of our theory. All proofs in the following formal analysis are relegated to the Theoretical Appendix for clarity.

2.1 Model

Suppose individuals have privately known ability $\theta \in \left[\underline{\theta}, \overline{\theta}\right]$ and publicly observable pedigree (or signal of ability) $P \in \left[\underline{P}, \overline{P}\right]$ distributed such that the posterior density of ability $F'\left(\theta \middle| P\right)$ has complete support over $\left[\underline{\theta}, \overline{\theta}\right]$ for all P. To guarantee that equilibrium wage offers increase in pedigree, we assume that the monotone-likelihood-ratio-property (MLRP) *strictly* holds—formally, for all $\theta_h > \theta_l$ and $P_l < P_h$[6]

$$\frac{F'\left(\theta_h \middle| P_h\right)}{F'\left(\theta_h \middle| P_l\right)} > \frac{F'\left(\theta_l \middle| P_h\right)}{F'\left(\theta_l \middle| P_l\right)}$$

An individual chooses to work either as an entrepreneur, where he produces θ (*i.e.* normalized to his ability) and keeps all of his produce, or accepts wage work at a firm. Individual productivity in a firm equals $\pi\left(\theta\right)$, finite and increasing in its argument. A firm makes symmetric take-it-or-leave-it (TIOLI) wage offers $w\left(P\right)$ to all individuals with pedigree P.

[6] The MLRP means that observing any two individuals with pedigrees P' and P'' such that $P' < P''$ and hypothesiszing any two ability levels θ' and θ'' such that $\theta' < \theta''$, it is more likely that the higher ability θ'' belongs to the individual with the higher pedigree P''.

We make two regularity assumptions to guarantee that the first-order approach is valid: (1) $F(\theta|P)$ is log-concave for all P, and (2) $\pi(\theta)$ exhibits weakly decreasing differences with respect to θ (*i.e.* for all θ, $(\pi(\theta)-\theta)' \leq 0$). The former assumption is satisfied by most probability distributions commonly used in theoretical economics.[7] The latter assumption means incrementally increasing an individual's innate ability will not *improve* his productivity as a wage worker more than his productivity as an entrepreneur. It could still be the case that all individuals are more productive as employees than as entrepreneurs. Note that although these two assumption guarantee that the firm's wage setting problem is well-behaved (*i.e.* the second order condition is satisfied for all critical points), they are stronger than necessary and are not used for any other purpose in the model.

We assume two additional properties to ensure that the equilibrium separates workers; *i.e.* both entrepreneurship and traditional employment coexist. In order for traditional employment to exist all, traditional employment must be more productive than entrepreneurship over, at least, some ability range.[8] Therefore to guarantee that employment exists (*i.e.* the firm can profitably make at least one offer that will be accepted by some individual), we assume that the least able individuals are more productive in the firm (*i.e.* $\pi(\underline{\theta}) > \underline{\theta}$). To ensure that the firm cannot profitably entice everyone to join the firm, we assume that extremely high ability is sufficiently rare—in particular, $\lim_{\theta \to \overline{\theta}} F'(\theta|P) = 0$, at least for some P. Again, as we will show in the analysis, these assumptions guarantee separation, but other reasonable assumptions would suffice instead. We begin the next section by proving these results.

2.2 Analysis

Individuals with pedigree P reject traditional employment if and only if their entrepreneurial outside option (ability) strictly exceeds the firm's offer of $w(P)$. Thus, for every pedigree P, the firm chooses wage w to solve

$$\max_{w} \int_{\underline{\theta}}^{w} (\pi(\theta) - w) F'(\theta|P)\, d\theta$$

[7] For example, Uniform, Normal, Exponential, Logistic, Extreme Value, Laplace, Power Function, Weibull, Gamma, Chi-Squared ($c \geq 2$), Chi ($c \geq 1$), Beta ($\nu \geq 1$, $\omega \geq 1$), Maxwell, Rayleigh, Pareto, and Lognormal distributions have log-concave cdfs (Bagnoli and Bergstrom 2005).

[8] When entrepreneurship pays based on ability directly, and traditional employment pays on pedigree, which is only imperfectly correlated to ability, a potential "lemons problem" in employment exists. To see this, suppose that employees of every ability were exactly as productive inside the firm as outside (*i.e.* $\pi(\theta) = \theta$). For any wage offer w, the range of individuals accepting the offer will have ability in the range $[\underline{\theta}, w]$ and average productivity strictly less than w. Since the firm must pay all the accepting workers w, such a high offer is clearly unprofitable. The firm may reduce the wage but then the most talented individuals who accepted before will now reject and the problem remains—in fact, no matter what the firm offers, the wage will always exceed the average productivity of those who accept.

yielding a first-order-condition (FOC) for every pedigree P

$$(\pi(w) - w) F'(w|P) = F(w|P) \tag{1}$$

such that the marginal benefit of attracting more able employees with pedigree P (LHS) equals the cost of raising the wages of all less able employees with pedigree P (RHS). Under our regularity assumptions, the firm's problem is well-behaved:

Lemma 1 *Any interior solution to the FOC is the unique global optimum.*

Proof. See Theoretical Appendix. ■

Now we describe the conditions for such an interior solution to exist. The employer can profitably offer $w(P)$ that some individuals will accept if and only if the marginal benefit of hiring individuals of some ability level (LHS of (1)) exceeds the marginal cost of hiring all less able workers (RHS), or equivalently

Remark 1 *Traditional employment exists if and only if there exists* $\hat{\theta} \in \left[\underline{\theta}, \overline{\theta}\right]$ *and* $\hat{P} \in \left[\underline{P}, \overline{P}\right]$ *such that* $\pi\left(\hat{\theta}\right) - \hat{\theta} \geq F\left(\hat{\theta}\middle|\hat{P}\right) / F'\left(\hat{\theta}\middle|\hat{P}\right)$.

Since $\pi(\underline{\theta}) > \underline{\theta}$ and $F'(\theta|P)$ has full support over $\theta \in \left[\underline{\theta}, \overline{\theta}\right]$, for all P, this condition is always satisfied for $\hat{\theta} = \underline{\theta}$; the LHS is positive and the RHS equals zero. To see why this condition suffices, notice that the firm generically faces adverse selection in hiring. If it offers a wage of $w(P)$ to a set of imperfectly pedigrees individuals, the average ability of individuals who accept the offer (if any) is generally less than $w(P)$, because all more able individuals reject the offer in favor of entrepreneurship. If though, traditional employment is, over at least some ability range, more productive than entrepreneurship, a lemons unraveling problem can be averted. The assumption $\pi(\underline{\theta}) > \underline{\theta}$ suffices, because it is always profitable for the firm to hire the least able. Of course, while $\pi(\underline{\theta}) > \underline{\theta}$ is sufficient, it is not necessary. The advantage of traditional employment over entrepreneurship might instead exist only at higher ability levels, but if this is the case then the advantage must be large enough that it can overcome losses from hiring those at the bottom of the ability distribution (with the same pedigree) for the same price.

Considering the opposite end of the ability spectrum, the employer cannot profitably attract all individuals with pedigree P if and only if the marginal benefit of hiring individuals of some ability level (LHS of (1)) is less than the marginal cost of hiring all less able workers (RHS), or equivalently

Remark 2 *Entrepreneurship exists if and only if there exists* $\hat{\theta} \in \left[\underline{\theta}, \overline{\theta}\right]$ *and* $\hat{P} \in \left[\underline{P}, \overline{P}\right]$ *such that* $\pi\left(\hat{\theta}\right) - \hat{\theta} < F\left(\hat{\theta}\middle|\hat{P}\right) / F'\left(\hat{\theta}\middle|\hat{P}\right)$.

Since $\pi\left(\overline{\theta}\right) - \overline{\theta}$ is finite and we have assumed that $\lim_{\theta \to \overline{\theta}} F'\left(\theta | P\right) = 0$, for all P, this condition is always satisfied for $\hat{\theta} = \overline{\theta}$. Note that our standard distributional assumptions imply that entrepreneurship will coexist with traditional employment even when traditional employment is much more productive than entrepreneurship over the *entire ability range.* Alternatively we might have assumed that at high ability levels the productive advantage of traditional employment is sufficiently small.

Therefore, we have proved that

Lemma 2 *An interior, separating equilibrium exists, in which positive measures of individuals choose entrepreneurship and traditional employment.*

Furthermore, since all P pedigreed individuals with ability strictly great than $w\left(P\right)$ choose entrepreneurship and all with weakly lower ability choose traditional employment, the following proposition is immediate:

Proposition 1 *Entrepreneurs are more able than employees of the same pedigree P.*

Also observe that all employees with a given pedigree earn the same wage, despite the fact that they have a range of abilities, whereas since entrepreneurs earn according to their ability, their incomes are not only higher but exhibit greater spread. That is,

Corollary 1 *The incomes of entrepreneurs have higher mean and variance than those of employees of the same pedigree.*

While the intuition behind Proposition 1 is clear, one may ask why firms, realizing that those who reject an offer are more talented than those who accept, do not subsequently offer higher wages to those who reject (*i.e.* entrepreneurs). If the firm offered entrepreneurs more, it would face another lemons problem and thus will not do so in equilibrium. To see this, suppose that the firm offered $w_E\left(P\right) > w\left(P\right)$ to all entrepreneurs with pedigree P. First, notice that very talented entrepreneurs (*i.e.* $\theta > w_E\left(P\right)$) would still choose entrepreneurship and could not be enticed without raising the wage so high that paying mundane entrepreneurs (*i.e.* those with θ such that $w\left(P\right) < \theta \leq w_E\left(P\right)$) the same wage would be unprofitable. This problem is obviously the same as when wages were based on P alone, and this is not surprising, because rejecting $w\left(P\right)$ has effectively become part of the individuals' pedigree, or observable characteristics. One could then offer ever greater wages to those who reject multiple offers: $w_{E^n}\left(P\right) > \ldots > w_{E^2}\left(P\right) > w_E\left(P\right) > w\left(P\right)$ *ad infinitum.* But this solution relies on incomplete reasoning. An individual who would accept $w\left(P\right)$ could "play hard to get" and receive the offer $w_{E^n}\left(P\right)$ by costlessly refusing all lower offers, and indeed if the firm offered such wages, it would be optimal to do so. Thus, in equilibrium, the "pedigree of entrepreneurship" contains no *ex ante* information for employers—paying for it would be irrational. Of course, were it costly to be an entrepreneur,

then the decision to become an entrepreneur may indeed contain information, and having a history of costly entrepreneurship would simply be part of one's pedigree, just like education or any other observable characteristic. Then the firm would have to offer a wage that exceeds the individuals' *future* entrepreneurial payoffs. If part of the productivity advantage that we assumed for traditional employment above is in fixed costs to start a business or the risk of entrepreneurial failure, then indeed hiring away an already successful entrepreneur could be much more expensive than hiring her before she began her venture. Our model is agnostic about how pedigree is obtained but simply requires that pedigree be related to underlying ability by the MLRP at the time a wage offer is made.

The simplicity of Proposition 1 stems from the fact that for every pedigree the unique minimum ability of entrepreneurs coincides with the unique maximum ability of employees. It is natural then to ask whether we might just as easily compare the pedigrees of entrepreneurs and employees of the same ability. Before making the analogous argument for pedigrees, we prove an intuitive property of the equilibrium wage function:

Lemma 3 *Wages strictly increase in pedigree (i.e.* $w'(P) > 0$*).*

Proof. See Theoretical Appendix. ■

Since MLRP implies that (unconditional on occupational choice) a higher pedigree always *probabilistically* indicates higher ability, and productivity increases in ability, it is intuitive that equilibrium wages increase in pedigree.[9] Because wages increase in pedigree, an individual who would accept the offer to P pedigreed individuals would accept all offers to higher pedigreed individuals. More formally, monotonicity implies $w(P)$ is invertible, and $w^{-1}(\theta)$ is the minimum pedigree, for which the firm makes an offer that an θ ability individual would accept. Of course, the firm might not make any offers which entice the most able individuals. Similarly, there may be very low ability individuals who would accept any offer the firms makes.

Therefore, the following proposition is immediate:

Proposition 2 *Employees have better pedigrees than entrepreneurs of the same ability* θ *(if there exist* θ *able individuals engaged in both occupations).*

Figure 1 illustrates the simple intuition behind Propositions 1 and 2.

Figure 1 here

The parsimonious nature of the model enables the sharp results of Propositions 1 and 2. Of course, we should not expect to find these predictions hold for every individual in an

[9] The proof of Lemma 3 also makes clear that our requirement that MLRP *strictly* holds is stronger than is generally necessary—in fact if MLRP holds weakly everywhere and strictly at any pedigree less than the wage, that suffices for the equilibrium wage to increase.

empirical setting; however, these propositions together with Corollary 1, taken as average effects, are our primary empirical hypotheses:

1. From Proposition 1: Entrepreneurs are, on average, more able than employees *of the same pedigree.*

2. From Proposition 2: Entrepreneurs have, on average, lower pedigrees than employees *of the same ability.*

3. From Corollary 1: *Conditional* on pedigree, entrepreneurs earn more, and their earnings exhibit greater variance.

2.3 Endogenous Pedigree Acquisition

In the model analyzed above, pedigrees arise *exogenously.* In the case of observable signals of ability, which a person has little control over, this may accurately describe the pedigree assignment process. For example, immutable physical attributes may signal abilities as a fashion model or basketball player. But an individual can control other signals of ability. Our leading example, formal educational attainment, is clearly a case where the individual *endogenously* influences the signal. Does this difference play a meaningful role in the theory?

To answer this, consider an additional stage, expanding on Spence's (1973) canonical job-market signaling model, that occurs *before* the occupational choice decision—one where each individual chooses how much effort to exert in the acquisition of (educational) pedigree. As above, individuals have privately known ability θ distributed according to log concave density F', such that $\lim_{\theta \to \overline{\theta}} F'(\theta) = 0$. In childhood (Stage 1) individuals invest effort in observable (educational) pedigree. Pedigree is the sum of chosen effort e and noise ε distributed according to log-concave density G' (*i.e.* $P(e) = e + \varepsilon$). Effort in education costs $C(e, \theta)$. Higher effort levels cost more ($C_1 > 0$), but any level of effort is less costly for more able individuals ($C_2 < 0$). We also assume that C is *sufficiently* convex for individuals' second-order-condition (SOC) to hold (see proof of Lemma 6 in the Theoretical Appendix) and the standard Spence-Mirrlees (or single crossing) property $C_{12} < 0$ holds. In adulthood (Stage 2) individuals choose either to enter entrepreneurship, where they earn their ability θ, or accept a wage $w(P)$ as we described in the main model.

First, note by Bayes' Rule

$$F'(\theta | P) = \frac{\Pr\{P | \theta\} F'(\theta)}{\Pr\{P\}} = \frac{G'(P - e(\theta)) F'(\theta)}{\int_{\underline{\theta}}^{\overline{\theta}} G'(P - e(t)) dF(t)} \tag{2}$$

Thus, $F'(\theta | P)$ is log-concave, because multiplication preserves log-concavity (the denominator is a constant, given P). Thus, all results from Lemma 1 through Corollary 1 hold.

In order for Lemma 3 and Proposition 2 of our main model to hold for *endogenously* determined pedigrees, the resulting posterior distribution of ability given pedigrees must satisfy the MLRP in equilibrium. In the Theoretical Appendix we show that it does:

Lemma 4 *Equilibrium effort exerted in pursuit of pedigree (and average pedigree) increases in ability, such that the posterior density of ability $F'(\theta|P)$ satisfies the MLRP.*

Proof. See Theoretical Appendix. ■

Since optimal effort (and pedigree) increases in ability, a continuously separating equilibrium exists. Therefore, Lemma 3 and Proposition 2 of our main model also hold:

Proposition 3 *When workers endogenously acquire pedigree before making an occupational choice, all results of Section 2.2 continue to hold.*

In summary, the employer does not care whether the signals are exogenously endowed or endogenously acquired *per se*. She only cares about their statistical relationship to unobservable present and future productivity. This is why the addition of a pre-stage where workers endogenously cultivated pedigrees does not change results from the base model, where signals are acquired exogenously.

It may seem counterintuitive that entrepreneurs would invest in pedigree at all, because in our simple model, pedigree does not help them. Indeed, if pedigree were a deterministic function of ability and effort, then individuals of some ability levels would all become education eschewing entrepreneurs, and individuals of other ability levels would engage in just enough education to signal their ability (and fitness for wage work). But the assignment of pedigree is stochastic. So, all individuals of a given ability in the pre-stage have the same motivation and the same *ex ante* chance of becoming an entrepreneur. They invest optimally based on the prior odds. Effort exerted pursuing pedigree is only a function of ability in equilibrium—all individuals of a given ability try just as hard to earn a degree. *Ex post,* though, those with good random draws of signal will accept wage offers, while those with poor draws will become entrepreneurs.

This extension covering endogenous pedigree acquisition is one of pure signalling—that is, effort exerted in acquiring education does not alter ability. But it could be that education fundamentally improves productivity (as in Becker 1964 and Mincer 1970). Does this parsimony motivated omission matter? In our model, the amount of effort exerted is, in equilibrium, an increasing function of ability. Thus, if the amount of human capital added by education is an increasing function of effort and innate ability, in equilibrium it becomes a transform of ability itself. So, although education changes ability, and individuals seek more of it for its productive value, pedigree will still have the required statistical properties with respect to the new ability level, and our model's propositions hold.[10]

[10] Formally, so long as the human capital enhanced by education is modelled as a noisy function of effort and ability, such that the posterior distribution of ability given pedigree remained log-concave and satisfied the MLRP, our results will continue to hold.

3 The US National Longitudinal Survey of Youth 1979: Sample and Analysis

3.1 The Sample

We empirically test our theoretical predictions using data from the US National Longitudinal Survey of Youth 1979 (NLSY79). The NLSY79 project follows the lives of a cohort of 12,686 men and women, who were 14-22 years old when first surveyed in 1979. The sample was designed to represent the population of youth residing in the United States on January 1, 1979. These individuals were interviewed annually from 1979 through 1994, and since 1994 on a biennial basis.[11]

Of the 12,686 surveyed individuals in 1979, 6,111 belonged to a sample of individuals designed to represent the non-institutionalized civilian segment of the US youth population in 1979. 5,295 individuals belonged to a supplemental sample of civilian Hispanic or Latino, black, and economically disadvantaged non-black/non-Hispanic respondents living in the US in 1979. 1,280 respondents represented the population serving in one of the four branches of the US military as of September 30, 1978. Following 1984, 1,079 members of the military sample were no longer eligible for interview and only 201 randomly selected respondents from the military sample remained in the survey.

Following 1990, none of the 1,643 members of the economically disadvantaged, non-black/non-Hispanic sample were eligible for interview, leaving 9,964 or 78.54 percent of the original sample of respondents.[12] A small percentage (less than 10 percent) of this set of 9,964 individuals available in 1991 drop out in every subsequent year, primarily due to death and shifting residence out of the US, such that we have information on the responses of 7,757 survivors (77.8 percent of 9,964 individuals eligible to be part of the survey during each round of surveying) for 2008, the last year in our study.[13] To ensure that our results are not driven by survivorship and non-response biases, or idiosyncratic aspects of any one round of the NLSY79, we analyze individuals' occupational choices and outcomes using information from each of the 1994, 1998, 2002 and 2008 rounds of the NLSY79. Thus, our record of individuals' occupational choices from 1994, fairly early in their careers (age 30-37), to 2008, up to and beyond their peak productive ages (age 43-52), is reasonably complete. For brevity's sake, we only report and discuss results obtained from the first (1994) and last (2008) years of our

[11] The interviews were conducted with paper-and-pencil interviewing (PAPI) till 1990 and were executed with computer-assisted personal interviewing (CAPI) in subsequent interviews.

[12] The dropping of a segment of the military subsample reflect a decision by NORC not to attempt to interview sample members who were determined to be extremely difficult to interview. The decision to stop interviewing a large subsample of the disadvantaged was due to funding cutbacks.

[13] As of 2010, 573 main respondents (5.8 percent of the respondents eligible for interview) had been reported as deceased. The response rate for those believed to be alive is 80.6 percent. https://www.nlsinfo.org/content/cohorts/nlsy79/intro-to-the-sample/retention-reasons-noninterview#rni

study.

We chose to analyze the NLSY79 by cross-sections corresponding to survey years, rather than longitudinally, because our theoretical model treats individual ability an educational qualifications (our pedigree measure) as fixed and determined prior to the timing of occupational choice—the decision between wage- and self-employment. In the data, measures of intrinsic ability are static. Since most individuals (about 80 percent in our sample) obtain their highest degrees by age 30, and employment/earnings data during the period of educational attainment are quite noisy, we choose the 1994 round of the survey (when respondents are 30-37 years old) to first examine employment choices. This point both captures the final educational pedigree and early career employment choices for most respondents, so as to most closely match the timing of our theoretical model.

We collect detailed information on the following aspects of the NLSY79 respondents:

- Labor market behavior (employment, industry of employment, and occupation),
- Education (high school, college, training, highest degree earned),
- Family background (including data on parental education and occupation collected from the respondents' parents in 1979),
- Armed Forces Qualification Test (AFQT) Scores (measures knowledge and skills including reading and mathematics, test administered in 1980),
- Non-cognitive and attitude test scores (from tests on risk-taking, locus of control, self-esteem, trusting of others, sociability, etc. in 1979),
- School and college identities of respondents from the NLSY79 Geocode survey, matched with college rankings (in 2002) obtained from the National Center for Education Statistics,
- Family life (marital status and family size), and
- Assets, income, number of hours worked and pay rate (all dollar figures are normalized to Y2010 dollars).

Table 1 shows key characteristics of the 1994, 1998, 2002 and 2008 samples. 78-85 percent of the respondents in the four NLSY79 rounds we analyze reported being employed full-time (that is, being engaged in work for 40 hours or more each week). We classify the full-time employed as wage-employed and self-employed (entrepreneurs). Individuals are considered self-employed if they reported (a) owning at least 50 percent of a business, (b) being the principal managing partner of a business, (c) filed a form SE for federal income taxes, or (d) being an independent contractor, independent consultant, or freelancer. This definition of self-employment is consistent with those used in surveys such as the Current Population Survey (CPS), the official source of data on employment and unemployment in the United States, as well as previous studies of entrepreneurship (e.g., Evans and Leighton 1989, Hamilton 2000).

Table 1 here

Table 1 shows that the self-employment rate (as a fraction of all full-time workers) varies between 8.2 percent in 1994 (when respondents are 30-37 years old) and 12.7 percent in 2008 (when respondents are 44-51 years old). According to the CPS, U.S.-wide self-employment rates were 10-11 percent during this period (Hippel 2010). Increasing self-employment rates in later NLSY79 rounds corroborates the finding from other US population surveys that the probability of self-employment increases with age. For example, according to CPS data, the self-employment rate in 2009 for those between 25-34 years was 6.5 percent, for those between 35-44 years was 10.9 percent, and for those between 45-54 years was 13.5 percent—numbers comparable with our sample self-employment rates (see Hippel 2010, Page 24).

Next, we report sample statistics for the key variables of our study.

3.2 Variables

Self-employment (Entrepreneurship)

The NLSY79 surveys and our analysis, treat all respondents as self-employed if they reported (a) owning at least 50 percent of a business, or (b) being the principal managing partner of a business, or (c) filing a form SE for federal income taxes, or (d) being an independent contractor, independent consultant, or freelancer. In robustness checks, we also distinguish between "incorporated self-employed"–self-employed individuals, who reported owning an incorporated business–and "unincorporated self-employed,"–those who did not.[14] Accurate information on business incorporation is available only for rounds 2002 and 2008 of the NLSY79: in each, entrepreneurs who own incorporated businesses account for approximately 12 percent of the self-employed individuals (less than 2 percent of the full samples).

Ability

We use each respondent's Armed Forces Qualifications Test score (AFQT) as the primary measure of ability. Each NLSY79 respondent was administered the Armed Services Vocational Aptitude Battery (ASVAB) test in 1980, after payment of a $50 honorarium. A composite score derived from select sections of the ASVAB battery was then used to derive the Armed Forces Qualifications Test score (AFQT) for each youth. The AFQT, developed by the US Department of Defense (DOD), is considered a general measure of intelligence and trainability, and is used as a primary criterion for enlistment into the Armed Forces. It is constructed by summing the respondent's performance on the arithmetic reasoning, word knowledge, paragraph comprehension, and numerical operations section of the ASVAB battery. To account for the variation in ages of the test-takers in 1980 (16-23 years old), we apply a correction to the scores as recommended by Altonji, *et al.* (2008). The scores are then expressed as percentiles. The scores measure individuals' inherent ability and thus

[14] In a recent working paper, Levine and Rubinstein (2013) suggest that self-employed individuals who own incorporated businesses are high-ability individuals who also earn significantly more than average wage employees and self-employed individuals who do not own incorporated businesses.

remains constant across all NLSY79 rounds.

Table 2 here

Pedigree (Educational qualifications)

An ideal measure of pedigree should capture all the observable attributes that potential employers observe, and treat, as related to an individual's ability and productivity. Here, we employ the educational qualifications of individuals, measured by the total number of years of education, as well as the highest degree attained, as proxies for pedigree.

Some educational qualifications may be indispensable for entrepreneurship in some sectors. For example, a J.D. may be essential to start a law partnership (and to practice law at a law firm), an M.D. to start one's own medical practice (and to practice medicine at a hospital), and so on. In these situations, attaining the educational degree might itself may not serve as an adequate signaling device for potential employers. Accordingly, we also gather detailed information on the rankings of the academic institutions associated with the respondents' degrees. Information on the colleges and universities attended by respondents is available from the NLSY79 Geocode survey. We then match these institutions (over 4,000 unique institutions in our sample) to their rankings obtained from the National Center for Education Statistics. The rankings information is based on the ACT scores of the applicants, SAT scores of the applicants, and selectivity (percentage of applicants that are admitted). We construct four tiers of institutional quality based on their rankings as an additional measure of pedigree.

Choosing 1994 as the start-year for our analysis ensures that a majority of our respondents have completed their highest degrees (since they are between 30 and 37 in 1994). A small number of respondents acquire educational qualifications after 1994. Accordingly, we measure the highest educational qualifications (and the ranking of the institution associated with the highest qualifications) achieved by the individual at a point in time on the individual's occupational choices. Hence, the pedigree of individuals can be different across the 1994, 1998, 2002 and 2008 rounds of the NLSY79 data. Table 2 reports the percentage of individuals associated with different levels of educational qualifications in the 1994 and 2008 NLSY79 rounds.

Earnings and wealth

We use the annual income reported by individuals in the 1994, 1998, 2002 and 2008 rounds of the NLSY79 to capture their earnings (in the previous calendar year). The NLSY79 surveys report income after checking the information against the individuals' information gathered from Employer Surveys and Current Population Surveys (CPS).[15]

[15] Personal income is gathered by adding the amounts in response to two questions (1): during [the previous year], how much did you receive from wages, salary, commissions, or tips from all jobs, before deductions for taxes or anything else? and (2) during [the previous year], how much did you receive after deducting expenses from your farm or business? (NOTE: Top 2% of values are replaced with the group average).

It is possible that self-employed individuals' incomes and wealth are higher, not because they are compensated more for their work, but because they work more. Hence, we also measure individual earnings through their hourly pay rates.

Finally, according to some scholars, entrepreneurs under-report their income by as much as 30 percent (see Sarada 2010). Hence, we also use the reported net worth of respondents in the different rounds as a measure of their overall wealth.[16] All dollar figures are in 2010 dollars. All the earnings and wealth variables are top-coded in the NLSY79 data (that is, top 2% of values are replaced with the corresponding variable's group average), suppressing the variation in the right-tails and reducing the likelihood that outliers are driving the mean values of the variables. Table 3 reports summary statistics for our earnings and wealth variables.

Table 3 here

Control Variables

We employ several control variables, including information on individuals' family background (parental education, occupation, and wealth), gender, race, risk-preferences, personality traits (mastery/control, self-esteem, trusting of others, and sociability), occupation (CPS occupation codes) and industry of employment (SIC 3 digit industries) in our multivariate regression analyses to account for correlates of individual ability and occupational choice. We describe the less familiar controls below.

- The Pearlin Mastery Scale is designed to measure self-concept and references the extent to which individuals perceive themselves in control of events and situations that significantly impact their lives. Total score could range from 7 to 28 points, such that greater scores represent greater mastery. The Pearlin test was administered in the 1992 round of the NLSY79.
- The Rosenberg Self-Esteem Scale is designed to measure the self-evaluation that an individual makes and customarily maintains. Total score could range from 0 to 30 points such that a higher score indicates higher self-esteem (test administered in 1987).
- "Trusting of others" was measured by a question asking respondents "Generally speaking, how often can you trust other people?" Answers ranged from 1 to 5, respectively for responses "always, most of the time, about half the time, once in a while, and never" (the question was asked in the 2008 round of NLSY79).
- Sociability was measured by a question: "Thinking of yourself as an adult, would you describe yourself as: extremely shy (1), shy (2), somewhat outgoing (3) or extremely outgoing (4)" (the question was asked in the 1985 round of NLSY79).
- Risk preferences are measured through responses to the question: "Are you generally a person who is fully prepared to take risks or do you try to avoid taking risks? Rate

[16] Net Worth is calculated by summing all asset values and subtracting all debts. Missing assets and debt values are imputed. Top 2% of all values are topcoded.

yourself from 0 to 10, where 0 means unwilling to take any risks and 10 means fully prepared to take risks." (the question was asked in the 2010 round of NLSY79).[17]

To control for the effects of individuals' initial wealth endowments and economic background on their occupational choices, we include variables measuring the total income of the individuals' families in 1979 as well the total net worth of their families in 1985, the earliest years in the survey for which this information is available (again, these numbers are normalized to 2010 dollars).

In addition to the above, we control for individual's sex, race, marital status, parental education levels, parents' occupation (whether they were self-employed), and industry of the individual's occupation. Table 4 reports sample statistics for these control variables (we do not tabulate parental education levels, parents' occupation and industry of individual's occupation due to space constraints here).

Table 4 here

3.3 Descriptive Evidence

Are the model assumptions valid?

Before presenting evidence related to our main propositions, we check whether patterns in our NLSY79 sample are consistent with the explicit and implicit assumptions made while deriving our theoretical propositions.

Recall that one of our key assumptions (MLRP) guarantees that equilibrium wage offers increase in pedigree (formally derived as Lemma 3). Table A1 of the Appendix shows the median wages of individuals by their pedigree measured by years of education, highest degree achieved, and institutional ranking of latest degree. The Table confirms that wages are systematically higher for higher pedigree levels.

A second implicit assumption, required for our measures of ability and productivity to be valid is that the measures should be positively correlated with each other. Figures A1, A2 and A3 in the Appendix plot our three measures of productivity—annual income, pay rate, and net worth—against the AFQT scores of the respondents. Despite the noise in the data, all three measures of productivity are strongly positively correlated with each other, with Pearson correlations between 0.30 and 0.36.

Third, for educational qualifications to be sustainable signals in equilibrium, they should be positively correlated with ability. Table A2 shows that higher values of educational

[17] We also measured risk through responses to the following question "Suppose you have been given an item that is either worth nothing or worth \$10, 000. Tomorrow you will learn what it is worth. There is a 50-50 chance it will be worth \$10,000 and a 50-50 chance it will be worth nothing. You can wait to find out how much the item is worth, or you can sell it before its value is determined. What is the lowest price that would lead you to sell the item now rather than waiting to see what it is worth? Higher prices represented a greater love for risk (the question was asked in the 2006 round of NLSY).

attainment are systematically associated with higher AFQT scores.

These patterns in the data assure us that some of the basic criteria required of our measures to meaningfully test the theoretical propositions are met. Next, we proceed to examine the data descriptively for evidence related to our main propositions.

Descriptive evidence

To ensure that our findings are not driven by sample-selection or non-response bias, we analyze four NLSY79 subsamples starting from 1994. We emphasize that the choice of 1994 as the first year to analyze allows sufficient time for a majority of the NLSY79 respondents to complete acquiring their educational qualifications (our measure of pedigree) and choose their occupations. We are interested in the occupational choices made by individuals when they have completed acquiring their formal educational qualifications, since it is at this moment that the mechanisms underlying our propositions are likely to be most influential.

As individuals acquire work experiences and other attributes, they add to the set of observables that can influence future employment and wages, thus diluting the signal conveyed by educational qualifications alone. Individuals may also be more likely to choose vocations later in their careers as a function of a broader set of variables (e.g., desire to be own boss, lifestyle preferences, etc.) which could inject noise to our measurement of occupational choices as a function of ability, educational qualifications, and expected wages. Earlier samples also minimize respondent attrition (see Table 1).

For the above reasons, we focus on interpreting and reporting the findings from the NLSY79 1994 round (we use the terms "round" and "sample" interchangeably). Where relevant, we also analyze and discuss findings from the NLSY79 2008 round, the latest year for which we have comprehensive NLSY79 data. As suggested above, we additionally verify our findings in the 1998 and 2002 rounds, but do not report them here for brevity's sake.

Before investigating our first proposition, that entrepreneurs have higher ability, conditional on pedigree, we compare the unconditional ability distributions of self-employed and wage-employed individuals. Figure 2 plots the density distributions of AFQT scores for self-employed and wage-employed individuals in the 1994 round.

Figure 2 here

Recall that 8.2 percent of the 6,921 full-time employed respondents are self-employed in the 1994 sample, and the rest are wage-employees. Figure 2 suggests that self-employed individuals are systematically more frequent on the right-side of the ability distribution (since AFQT scores are expressed as percentiles, a score of 50 bifurcates the distribution into the less and more able regions). In particular, that the wage-employee ability density crosses the self-employed ability density just once from above, indicates that the ability distribution of the self-employed *first order stochastically dominates* the ability distribution of wage employees. In other words, for *any* arbitrary level of innate ability x, the proportion of self-employed individuals with ability greater than x exceeds the proportion of wage employees above that

same threshold. This finding is not consistent with the view advanced by some previous scholars that entrepreneurs are outliers: that is are systematically more likely to be found in both tails of the ability distribution. Figure A4 of the Appendix plots and confirms that entrepreneurs are more likely than wage-employees to be drawn from the higher ability part of the distribution.

Table 5 here

Next, Table 5 reveals that self-employed individuals have higher AFQT scores, on average, for almost every level of educational attainment in the 1994 sample. T-tests for differences in means show that the differences are significant for the more common educational qualifications in the sample, perhaps because they afford a larger number of observations for the statistical test. These descriptives provide preliminary evidence for our first proposition that conditional on pedigree, self-employed individuals have higher ability. Table A3 of the Appendix shows that despite the likely switching of many wage-employees into self-employment as they grow older (recall that self-employment rate rises to 12.7 percent in the 2008 sample), entrepreneurs appear to have higher ability for a majority of the most common educational qualifications even in the 2008 sample.

Our second proposition states that conditional on ability, entrepreneurs have lower pedigree (educational qualifications).

In order to investigate the second proposition descriptively, we assigned each respondent to an ability decile based on his or her age-adjusted AFQT score. We then compare the average educational attainment of wage-employed and self-employed individuals within each ability-decile. We calculate the mean values of educational attainment after assigning each educational qualification an ordinal value. Thus, we measure "Education Level" on a scale on which the value 1 indicates the lowest educational level (completion of 1st grade) and the value 20 indicates the highest educational level (PhD or advanced Professional Degree such as MD, LLD or DDS). Similarly, we measure "Highest Degree Completed" using a scale on which the value 1 indicates the lowest degree (High-School diploma or equivalent) and the value 6 indicates the highest degree that can be attained (PhD or advanced Professional Degree such as MD, LLD or DDS). We measure College Selectivity Tier such that the value 1 indicates the respondent received her latest degree from a Top 50 institution, 2 indicates Top 50-200 institution and 3 indicates an institution ranked 200+ (we omit the observations for which we do not have information on the selectivity rank of the respondents' educational institutions).

Table 6 here

Table 6 shows that for most of the ability deciles, self-employed individuals have lower educational attainment, on average, relative to wage employees. The differences appear

more likely to be statistically significant when attainment is measured by educational level, rather than by highest degree achieved or selectivity tier perhaps because the latter variables are more crudely measured and reduce the variability among individuals in educational attainment (the selectivity tier variable is only available if the individuals completed four-year college, and in an institution for which admissions selectivity ranking is available from the National Center for Education Statistics). Overall, these numbers provide evidence in support of our second proposition that conditional on ability, self-employed individuals have lower educational qualifications.

Table A4 of the Appendix reports, and confirms, similar findings from the NLSY79 2008 sample.

Our third proposition states that holding pedigree constant, entrepreneurial earnings are higher on average, and have greater variance. We begin by comparing the unconditional earnings/assets distributions of entrepreneurs and wage employees.

Figure 3 plots the unconditional distributions of logged annual income (Figure 3a), logged pay rate (Figure 3b) and logged net worth (Figure 3c) for the wage-employed and self-employed in the 1994 sample. Figures 3a and 3b reveal the grater variance of entrepreneurial earnings. The unconditional average earnings of entrepreneurs also appear to be higher than that of wage-employees.

Figure 3 here

Figure 3c shows that the density of net assets for the self-employed, somewhat akin to their ability distribution, is right-shifted. The net worth of the average self-employed individual is nearly twice the net worth of the average wage-employee. The discrepancy in the annual income/pay rate distribution and the net worth distribution can be explained by the findings of previous literature which suggests that entrepreneurs systematically underreport annual income and that their annual income does not capture their true earnings which may be invested in business related assets or other types of holdings (Astebro and Chen 2014, Sarada 2010). Figure A5 confirms these patterns in the NLSY79 2008 round.

Next, we compare the earnings and wealth of the self-employed and wage-employed, conditional on educational qualifications. Figure 4 presents box plots comparing the distributions of annual income and net worth for wage employees and entrepreneurs based on the 1994 sample. Table 7 reports the corresponding descriptive statistics.

Figure 4 here

Table 7 here

The Table reveals that the average annual income of the self-employed was greater than that of the wage-employed for 12 out of the 15 educational levels. The average pay rate of the

self-employed was greater than that of the wage-employed for 11 out of the 15 educational levels. The average net worth of the self-employed was greater than that of the wage-employed for 13 out of the 15 educational levels. These differences are statistically significant for the more common education levels in the sample (e.g., for those that completed 12th grade). The variance (inferred from the standard deviation of the income/wealth variables in the Table) of earnings and wealth is also systematically higher for the self-employed for a majority of education levels. Table A5 shows that these patterns prevail even in the NLSY79 2008 round. We do not tabulate these descriptives for the other measures of pedigree (highest educational degree achieved and selectivity tier of institution) but confirm that even for different levels of these alternative measures, entrepreneurs have higher income and wealth.

Before testing our propositions more formally in multivariate regressions that control for potential correlates of occupational choice, education, and ability, we also descriptively examine differences in these control variables between entrepreneurs and wage employees.

Table 8 here

Table 8 shows the distribution of wage employees and entrepreneurs in the different industries and occupations in the 1994 sample. The self-employed concentrate in Construction, Personal Services and Business and Repair Services, while the fractions of wage-workers in these industries are notably lower. On the other hand, a much larger fraction of wage workers are employed in Manufacturing than the self-employed. Many self-employed are managers and proprietors, but a smaller proportion work in clerical positions than the traditionally employed. Table A6 reports the corresponding frequencies in the 2008 Sample. Although, the NLSY79 2008 sample uses a finer classification of industry and occupational classifications based on census 4-digit codes, unlike the NLSY79 1994 sample which uses a classification based on census 3-digit codes, these basic patterns appear to persist as the cohort ages.

Table 9 here

Table 9 shows that self-employed individuals, on average, have greater self-mastery (that is, a belief that they exert greater control over their life-events), higher self-esteem, though the differences are very slight. More notably, they love taking risks, and hail from wealthier backgrounds. They are less likely to be female, or belong to the minority communities. Entrepreneurs are more likely to be married than wage employees.

3.4 Multivariate Analysis

Here, we investigate our propositions more formally in multivariate regressions that control for the effects of demographic, attitudinal and background influences that might be correlated with both our outcomes of interest (occupational choice and earnings) and the main explanatory variables (educational qualifications and ability).

Table 10 here

Table 10 reports Probit estimates of the probability of being self-employed as a function of ability and educational attainment. We start by estimating the relationship between our measure of ability (AFQT scores expressed as percentiles) and the likelihood of being an entrepreneur (in the NLSY79 1994 sample, when the respondents are between 30-37 years of age). The estimates suggest a statistically significant positive effect of ability on entrepreneurship: a percentage point increase in the ability percentile is associated with a 0.05-0.08 percent (marginal effect calculated from the Probit estimates) higher likelihood of becoming an entrepreneur. Column 2 suggests that achieving a higher degree, on average, is associated with a 0.6-0.9 percent decrease in the likelihood of becoming an entrepreneur.

The estimated effects remain robust to controlling for attitudinal attributes, family background and wealth, and other demographic characteristics of the individuals, as well as controls for the occupations of the individuals (based on census-3 digit occupational codes). Many of the control variables have a statistically significant effect in directions that the previous literature on entrepreneurial determinants leads us to anticipate. For example, being male, being white, married, older, having a self-employed father, more educated mother, coming from a more wealthy family, and having a love for risk, all positively predict self-employment.

The estimated positive effect of ability, conditional on education, appears to be quite robust in later-year subsamples (although the estimated economic and statistical significance is slightly lower in the 1998 and 2008 subsamples) and survives alternative measurements of educational qualifications. Educational attainment also appears to be significantly negatively related to entrepreneurship in most estimations (Table A7 of the Appendix displays the estimates obtained through the alternative specifications).

These findings provide further evidence in support of our first two propositions. Self-employed individuals appear to have higher ability, conditional on educational qualifications (pedigree); they also have lower educational qualifications, holding ability constant, ceteris paribus.

Next, we proceed to check whether self-employed individuals earn more, on average, as anticipated by our third proposition.

Table 11 here

Table 11 reports OLS estimates of annual income, hourly pay rate and net worth as a function of being self-employed. The estimates suggest that on average, entrepreneurs earn more annually (by about $9,000), per hour (by about $3.3 per hour) and have greater net worth (by about $85,000) after controlling for educational qualifications and other variables.

One concern could be that these OLS regressions estimate the effect of the independent variables on mean earnings/income, and it is well known that entrepreneurial earnings are

more likely to have outliers. However, the NLSY79 reports earnings, pay rate and net worth values after top-coding them, such that the top 2 percent observations for each variable are all assigned the average of the top-2 percent. Thus, the maximum value for annual income was $385,282.6, hourly pay rate $294, and net worth $1.4 million in the 1994 sample. This top-coding, to an extent, mitigates the effect of influential outliers driving the OLS estimates.

Still, to get a better sense of the earnings/wealth gap between wage-employees and entrepreneurs, we estimated quantile regressions at different quintiles both on the raw variables and after log-transforming them (Table A8 of the Appendix reports the corresponding estimates). We find that systematically entrepreneurs have higher annual incomes above the 50th percentile for these variables, and higher net worth at every percentile (pay rate regressions, not reported here, provide estimates that are qualitatively comparable to the income regressions). These findings provide further evidence that the distribution of entrepreneurial income has a thicker tail, but also suggests that our OLS findings are not entirely driven by outliers. Further, entrepreneurs appear to enjoy a higher net worth at nearly all assets ranges. The patterns hold in the 1998, 2002, and 2008 samples.

Note that the regressions estimate the earnings and wealth of entrepreneurs after conditioning out the effect of the included observable factors, but do not necessarily correspond to the wages of entrepreneurs who chose self-employment due to asymmetric information about their ability. To investigate the relative wages and worth of such entrepreneurs, who are of particular interest to our study, we estimated 2-stage least squares regressions after explaining the choice of occupation (i.e., between self-employment and wage-employment) as a function of AFQT scores (not included in the second-stage) and other variables (included in the second-stage). The first-stage regression corresponding to this estimation is thus similar to the one presented in Column 6 of Table 10 (except that it is re-estimated as a linear probability model though OLS rather than Probit), and Table A9 of the Appendix reports the corresponding second-stage estimates. The estimates of entrepreneurial wages and wealth, relative to wage employees, thus obtained are substantially larger in magnitude than the corresponding OLS estimates (although less precise). This 2SLS estimation suggests that self-employed individuals with ability higher than their pedigree indicates earn more than their comparably pedigreed counterparts in wage-employment.

Overall, these findings are consistent with our third proposition that conditional on educational qualifications, self-employed individuals earn higher on average, and that their earnings show greater variance.

Robustness checks

First, we show that support for our three propositions appears stronger when we restrict the definition of the self-employed to those owning incorporated businesses (Table A10 of the Appendix reveals that the average incorporated self-employed have about 15 percent higher AFQT scores, but only 6 percent higher education relative to the wage-employed. They earn nearly 100 percent more and are worth nearly 200 percent more than the wage-employed).

Second, we have used the AFQT scores of respondents as a measure of their general ability–a factor which we argue affects individuals' productivity in both wage-employment and self-employment. One might argue instead that ability is multidimensional—some individuals have abilities that make them better suited to be entrepreneurs and others have distinct abilities that make them better at wage-employment—and our measure of ability captures the set of factors that are related to entrepreneurial productivity alone. However, not only have several previous studies used AFQT scores as a measure of general intelligence (Neal and Johnson 1996, Hansen et al 2004), but also the military uses the tests as a basis for recruitment. Still, to be sure that we are not uniquely capturing entrepreneurial abilities through the measure, we compared the scores received by self-employed and wage-employees on several other commonly employed measures of general intelligence including PSAT score, SAT score, and ACT score. Table 12 presents the corresponding results which show that self-employed individuals systematically scored higher on all these tests of intelligence, and arguably of scholastic aptitude. These results are striking, given our finding that entrepreneurs systematically have lower educational qualifications, especially when the ostensible purpose of SAT and ACT tests is sorting individuals into institutions of higher education suitable to their abilities.

Table 12 here

Third, we exclude those belonging to the government and non-profit sectors from among the wage-employed and confirm that the patterns hold (Appendix Tables to be added).

Fourth, we contruct two separate subsamples–one comprising of individuals working in high-value industries (defined as industries that employed individuals who enjoyed above median-net worths in 2008), and another comprising of individuals working in low-value industries (industries that employed individuals who enjoyed above median-net worths in 2008). High-value industries were: Agriculture, Forestry and Fishing, Mining, Utilities, Manufacturing, Wholesale Trade, Retail Trade, Information Technology, Finance and Insurance, Real Estate and Rental and Leasing, Professional and Technical Service, Educational Services, and Arts, Entertainment, and Recreation Low-value industries were: Construction, Transportation and Warehousing, Management and Administrative jobs, Health Care and Social Assistance, and Accomodations and Food Services. The rates of self-employment in 2008 were comparable in the two sets of industries. Running our regressions separately for the two sets of industries, we found evidence for our propositions strongest in the sub-sample of high-value industries (Appendix Tables to be added).

4 Alternative Explanation: Matching

Our primary theory based on asymmetric information can explain the above empirical findings even when ability is general—those innate attributes which make one a productive

entrepreneur (like IQ) are also productive within the firm. But, if one assumes that (i) ability is multidimensional, (ii) our measures of innate ability (AFQT) and pedigree (formal educational certificate) capture different kinds of (potentially correlated) abilities, and (iii) innate ability is relatively more productive in entrepreneurial work, while education matters relatively more for wage work, then a matching model (without asymmetric information) also delivers Propositions 1 and 2. In this section, we present and analyze such an alternative model. As before, our base model will be one in which pedigree is exogenous.

Suppose when entering the labor market, each individual is endowed with publicly observable innate ability θ and education P distributed stochastically (but not necessarily independently). Productivity in entrepreneurship and wage work are given by the respective constant returns to scale, Cobb-Douglas functions $S(\theta, P) = k_S \theta^{\alpha} P^{1-\alpha}$ and $\pi(\theta, P) = k_\pi \theta^{\beta} P^{1-\beta}$, where (non-equal) α and β lie in the unit interval. For simplicity we assume the labor market is sufficiently competitive that each worker is offered his productivity in wage work (recall that here all skills are publicly observable). Thus, an individual with ability θ and pedigree P chooses entrepreneurship if and only if $S(\theta, P) > \pi(\theta, P)$ or equivalently $(\theta/P)^{\alpha-\beta} > k_\pi/k_S$, which holds if and only if either (1) $\left(\theta > P\left(k_\pi/k_S\right)^{\frac{1}{\alpha-\beta}}\right) \wedge (\alpha > \beta)$ or (2) $\left(\theta < P\left(k_\pi/k_S\right)^{\frac{1}{\alpha-\beta}}\right) \wedge (\alpha < \beta)$.

Consider the first case: innate ability is relatively more important for entrepreneurial productivity ($\alpha > \beta$). Then, an individual chooses entrepreneurship if and only his innate ability θ exceeds the threshold level $P\left(k_\pi/k_S\right)^{\frac{1}{\alpha-\beta}}$, which is a constant given education P. Thus, Proposition 1 is again immediate. Similarly, conditional on innate ability θ, an individual chooses entrepreneurship if and only if his education is less than the threshold level $\theta\left(k_\pi/k_S\right)^{-\frac{1}{\alpha-\beta}}$, which proves Proposition 2.

Note, though, that while conditional on P, entrepreneurs enjoy higher incomes (i.e. the condition for choosing the occupation), we cannot straightforwardly make claims about variance. The difference in the matching model is that when managers can see both innate ability and education, there is no reason to bunch wage offers on education, but she will also base wages on varied innate abilities. In the asymmetric information model varied unobserved ability was only compensated in entrepreneurship, managers could not see it directly, and so ability directly introduced income variation only in entrepreneurship. Thus, while the first part of Corollary 1 holds in this case of the matching model, there is no reason to assume a matching explanation should lead to the second part of the corollary.

Now consider the second case: innate ability is relatively less important for entrepreneurial productivity ($\alpha < \beta$). Now an individual chooses entrepreneurship if and only if his innate ability is less than the threshold value $P\left(k_\pi/k_S\right)^{\frac{1}{\alpha-\beta}}$, a constant when education level P is observed. Clearly, Proposition 1 can never hold in this case. Likewise, conditional on innate ability θ, an individual chooses entrepreneurship if and only if his education exceeds the threshold level $\theta\left(k_\pi/k_S\right)^{-\frac{1}{\alpha-\beta}}$, which violates Proposition 2.

Thus, given our observed empirical support for Propositions 1 and 2, this matching model can only be true if $\alpha > \beta$, and we are left with no particular explanation for the second part of Corollary 1.

Nevertheless, since sorting might explain a majority of the empirical patterns which we have so far observed and argued to be evidence for asymmetric information about ability, we explore other empirical patterns in the data that could help us tell apart the two competing explanations. First, one aspect of asymmetric information is that it is likely to have the greatest influence on occupational choice when there is least observable information on individuals' productivity–that is, early on in their careers. As individuals work longer in any given job, employers receive additional signals, update their priors on the individuals' ability, and adjust wages accordingly. Hence, over a period of time, we expect asymmetries to be mitigated and to play less of a role in determining the choice between wage-employment and self-employment. In contrast, the sorting explanation suggests that higher ability individuals (relative to pedigree) are more likely to switch into entrepreneurship and higher pedigreed individuals (relative to ability) are more likely to switch into wage-employment, even if they were incorrectly assigned to vocations during the early-stages of their career (perhaps because individuals' learn about their types on the job, *à la* Jovanovic 1979). We investigate these subtle differences in the implications of the two models by examining the relationship between ability (conditional on pedigree) and occupational choice for the NLSY79 respondents in 1994 and then again for the same respondents in 2008. Table 13 presents the corresponding results.

Table 13 here

The first two columns reveal that the positive relationship between ability (conditional on pedigree) and the probability of self-employment, and the negative relationship between educational qualifications (conditional on ability) and the probability of self-employment, are strongest in the 1994 round; in fact, ability and education appear not to have a statistically significant relationship with the probability of being self-employed in 2008, when our respondents are between 44 and 52 years old. This finding is consistent with the resolution of asymmetric information towards the later years of individuals' careers.

Second, we investigate support for our propositions by separating the 2008 respondents' sample into a sub-sample of "stayers" (those who persistently stayed in either wage-employment or self-employment) and a sub-sample of "switchers" (those who switched from wage-employment to self-employment, or vice versa, between 1994 and 2008). Columns 3 and 4 reveal that evidence for P1 and P2 is strongest in the subsample of stayers. Indeed, if the sorting model (or sorting with imperfect information) has to explain our main empirical findings, then there is no reason to believe that the estimated effect of ability (and education) on the probability of being self-employed will be different for the stayers and switchers (since even if one allows for learning about multidimensional ability, individuals should be assigned

to their correctly matched vocation after a delay). Because the evidence for our propositions appears strongest for the set of individuals who were relatively certain of their future employment path early on in their career (the stayers), asymmetric information, rather than sorting, appears to be a more plausible explanation for the conditional correlations reported in the previous section.

Third, Table 14 investigates changes in the returns to education and ability across time as pertinent to the sorting and asymmetric information theories. The presence of asymmetric information about ability should result in a higher return to education and a lower return to unobserved ability during the early stages of the careers of wage employees. As the wage-employees' productivity is gradually revealed to their employers, the employers should compensate them based on their ability more and based on their educational qualifications less. Hence, declining returns to education with time and increasing returns to ability with time for the persistently wage-employed should support asymmetric information (and the importance of educational qualifications as a screening device in the early stages of employees' careers). The first two Columns of Table 14 show that the returns to education are remarkably stable over time for the entire sample of full-time employed respondents (i.e., across the 1994 and 2008 NLSY79 rounds). Yet, if one restricts the sample to those who are persistently wage-employed (these are the individuals for whom employers are more likely to have updated their beliefs about the employees' productivity), it appears that the returns to education has decreased, while the returns to ability has significantly increased from 1994 to 2008.

Table 14 here

Again, this finding is not consistent with sorting, even if one assumes imperfect information about ability. Baumol, Schilling and Wolff (2009) analyze the biographies of noted inventors and entrepreneurs and provide suggestive evidence that the returns to education for entrepreneurs are not necessarily lower—an independent finding which is difficult to reconcile with this alternative sorting-based explanation. Thus, overall, we conclude that the empirical patterns are consistent with signalling, rather than sorting.

5 The UK National Child Development Study: Sample and Analysis (NOTE: Incomplete)

In this section we describe the first data set we use to test our theory empirically, the first of the *British Cohort Studies* known as the *National Child Development Study (NCDS).* The data is maintained for research purposes by the *UK Data Service.* From their website:[18]

[18] http://www.esds.ac.uk/longitudinal/access/ncds/l33004.asp (Accessed 30 July 2013).

"The NCDS is a continuing longitudinal study that seeks to follow the lives of all those living in Great Britain who were born in one particular week in 1958. The aim of the study is to improve understanding of the factors affecting human development over the whole lifespan.

The NCDS has its origins in the Perinatal Mortality Survey (the original PMS study is held in the Data Catalogue under SN 2137). This study was sponsored by the National Birthday Trust Fund and designed to examine the social and obstetric factors associated with stillbirth and death in early infancy among the 17,000 children born in England, Scotland and Wales in that one week. Selected data from the PMS form NCDS sweep 0, held alongside NCDS sweeps 1-3, under SN 5565.

To date there have been eight attempts to trace all members of the birth cohort in order to monitor their physical, educational and social development. The first three sweeps were carried out by the National Children's Bureau in 1965, 1969 and 1974. These sweeps form NCDS1-3, held together with NCDS0 under SN 5565. In 1985 the NCDS moved to the Social Statistics Research Unit (SSRU) - now known as the Centre for Longitudinal Studies (CLS) - and the fifth sweep was carried out in 1991. The sixth sweep was conducted in 1999-2000, the seventh in 2004 and the eighth in 2008-2009.

The NCDS has gathered data from respondents on child development from birth to early adolescence, child care, medical care, health, physical statistics, school readiness, home environment, educational progress, parental involvement, cognitive and social growth, family relationships, economic activity, income, training and housing."

Recall our empirical hypotheses require data on (1) employment status, (2) ability and (3) pedigree. We descriptively examine patterns in the NCDS survey data and check whether they support our theoretical propositions. In particular, we focus on identifying patterns related to Propositions 1, and 2. The propositions require individual-level measures of ability, pedigree, and occupational choice (i.e., entrepreneur or employee).

5.1 Variables and Measurement (NCDS)

To gather data on measures of ability, pedigree and occupation, we start with the latest sweep of survey responses that are complete (rather than NCDS8, we use NCDS7 data since the latter has the responses of a larger number of individuals and more comprehensive coverage of individuals' employment choices). Accordingly, we use the NCDS7 survey responses, which were collected in the year 2004, and identified the occupations of 9,534 individuals when they were forty-six years of age.

1. Occupational choice: In NCDS7, 57.9% of the individuals were employed full time (30 or more hours per week) and 10.9% of the individuals were self-employed full time (entrepreneurs). These numbers are quite comparable to the current proportion of self-employed in the U.K. population of employed individuals (according to the UK office of National Statistics, 14% of the sub-population of employed individuals in 2012 were self-employed)[19] and the fraction of self-employed individuals in the U.S. (according to the Bureau of Labor Statistics, 1 in about 10 workers in the U.S. was self-employed in 2009)[20]. A majority of the rest of the 1958 cohort of individuals worked part-time (10.9% part-time employees and 2.1% part-time entrepreneurs).

For our tests, we focus on comparing the full-time employees with the full-time entrepreneurs. Hence, we only retain the 6,562 individuals who were occupied full-time in our sample: either as full-time employees (84.2%) or as full-time entrepreneurs (15.8%).

2. Pedigree: By pedigree, we mean all the observable attributes that individuals are endowed with and acquire, which they use to signal their ability. Of course, the most common pedigree individuals use to signal their ability is education. The effectiveness of education as a signal is because it is costly to acquire (thus cannot be acquired by low-ability individuals), and easily observed by employers. An ideal measure of education (or pedigree) should capture not only its extent (e.g., high-school, undergraduate degree, PhD), but also other aspects (e.g., selectiveness and reputation of the individual's educational institution, individual's field of study) that may be correlated with the ability of associated individuals. However, we do not currently have data on these other aspects and exclusively use information on individuals' successful completion of various levels of education in the British system as an indicator of pedigree.

Table 14 shows the highest educational attainment (which we henceforth use synonymously with pedigree) of the 6,562 individuals who were occupied full-time (when they were 46 years of age). Again, the distribution of educational attainment in our sample is comparable to the distribution in the U.K. population as inferred from the 2011-2012 U.K. Annual Population Survey.

Table 14 here

Clearly, relative to employees, entrepreneurs are over-represented at lower educational levels and under-represented at higher levels associated with higher pedigree. The distributions at the two ends of the educational attainment spectrum are illustrative: of those individuals who do not have any academic qualifications, 17.6% are entrepreneurs and the rest are employees; of those who have the highest academic degrees (post-graduate degrees)

[19] http://www.ons.gov.uk/ons/rel/lmac/self-employed-workers-in-the-uk/february-2013/rpt-self-employed-workers.html

[20] www.bls.gov/opub/mlr/2010/09/art2full.pdf

only 12.6% are entrepreneurs and the rest are employees (compared with the overall sample of occupied individuals which has 15.8% entrepreneurs).

Although these descriptives are unconditional, they provide suggestive support to Proposition 2 that entrepreneurs have lower pedigree than employees.

3. Ability: An ideal measure of individual ability should perfectly predict long-term productivity (and wages). Our theoretical premises assume that such perfect measures of ability are not available to employers. The empirical challenge then is to produce measures of ability that employers do not observe, but the econometrician does.

To start with, we assume that some attributes of individuals' accomplishment, such as grades received in tests that do not directly count towards admission to schools and also escape the scrutiny of employers, can convey information on individuals' ability beyond the educational attainment of individuals. We use two such measures of individuals' ability collected and coded in NCDS2 and NCDS4.

(i) Outstanding ability at age ten: The NCDS2 collected information on individuals' ability, when they were ten years old, based on their performance in standardized tests and interviews with their teachers. We use the teachers' responses to one question (about the ability of their students, the latter being the primary subjects of the survey) in particular:

"Compared with other children at this age, does he/she reveal outstanding ability in any area? E.g. writing stories, drawing, chess, modelling, music, science, sport, etc."

Of the 6,562 fully occupied individuals we identified (based on their employment status when they were 46 years of age through NCDS7 responses) teacher response to the above question was available for 5,602 individuals. Of these, the teachers identified 24.4% of the students as having outstanding ability (i.e. answered with a "Yes" to the above question) and 75.6% of the students as not having such outstanding ability. These responses were collected uniquely for the survey, and the teachers' assessment of the student ability were not made available to the students. The survey question was also meant to capture dimensions of students overall ability not necessarily captured by the typical tests of scholastic ability. Hence, it is not plausible that they were used by the individuals to signal their ability to potential employers in the future.

(ii) A-level performance score at age nineteen: This variable, ranging from 0-15 was put together after summing the grades obtained by the students in A-level examinations, which are taken by students for three or four subjects after studying the subjects for typically two years between ages 16-18. The scores were calculated on a 0-15 point scale (each A-grade in an exam counted for 5 points, while the lowest grade, "E" was worth just one point; an individual getting more than three A's still only received a maximum of 15 points). Students that did not take the A-level exams (typically only those students who want to pursue higher educations in universities take the exams) earned a zero on the variable. Of the 5,725 occupied individuals for whom we had information on A-level scores, the vast majority (80%) did not take the exams and were accordingly scored zero. The mean score,

conditional on taking the exam, was 6.4.

5.2 Descriptive Evidence (NCDS)

Table 14 showed that the entrepreneurs in the NCDS7 sample have lower pedigree, on average, and thus, although unconditioned descriptives, provided suggestive support to Proposition 2. Proposition 1.predicts that self-employed have higher ability on average, conditional on pedigree.

Table 15 here

Table 15 shows the unconditional means of our two ability scores for the sub-samples of employed individuals and entrepreneurs. We emphasize here that the two scores were obtained at respectively ages ten and seventeen, and the occupational status of the individuals at age forty-six. Entrepreneurs appear to score systematically higher on the two scores. Even if we restrict the sample to only those who took the A-level exams (about 81 % of our occupied individuals sample) and thus could not have earned a zero on the exam, entrepreneurs appear to score on average, higher. All these differences are statistically significant at $p<0.01$.

Figure 5 here

Figure 5 also shows the distribution of the A-level scores (conditional on taking the exam) for the entrepreneurs and the employees. It is clear that the distribution of the scores for the entrepreneurs is denser towards the higher range of scores. To the extent that these can be arguably seen as measures of individual ability, these unconditional patterns suggest Proposition 1 that entrepreneurs have higher ability, conditional on pedigree.

Table 16 here

Next, we descriptively examine whether, entrepreneurs are more able than employees, conditional on educational qualifications. Table 16 presents the averages of our two ability scores for the different pedigrees one might acquire. It is clear from the Table that for every class of educational attainment (except AS level and Diploma, both of which had a small number of observations for the entrepreneurs, respectively 1 and 38, thus making it hard to precisely estimate the means), entrepreneurs were more likely to have been rated as outstandingly able by their teachers when they were ten years of age. Likewise, conditional on having taken the A-level exam, which serves the primary purpose of gaining entrance into institutions of higher education, entrepreneurs appear to score higher. This supports our Proposition 1 that entrepreneurs tend to have higher ability than employees of the same pedigree.

Table 17 here

Finally we check whether, for a given ability level, employees have higher pedigree. Table 17 shows the percentage of employees and entrepreneurs that chose different educational levels, conditioned on belonging to the two ability levels (outstanding ability = 0 or outstanding ability = 1). Broadly, the table suggests that entrepreneurs are more likely to have lower educational attainment for both ability levels, hence providing support for Proposition 2.

Thus, our preliminary and descriptive analyses appear to broadly support the theoretical Propositions 1, 2 and Corollary 1.

6 (Preliminary) Conclusion

In this preliminary paper, we have formally generated three main hypotheses:

1. Entrepreneurs are, on average, more able than employees *of the same pedigree.*
2. Entrepreneurs have, on average, lower pedigrees than employees *of the same ability.*
3. *Conditional* on pedigree, entrepreneurs earn more, and their earnings show greater variance.

We find preliminary evidence for our main propositions from a longitudinal dataset of individuals born between 1957 and 1964 in the US, as well as a longitudinal dataset of individuals born in one particular week in 1958 in the UK. The evidence suggests that entrepreneurs scored systematically higher on measures of ability collected when they were in their youth, that the entrepreneurs were underrepresented at higher levels of educational attainment, and had higher ability scores relative to employees with comparable educational attainment. Conditional on educational attainment, entrepreneurs appear to have higher incomes and wealth levels. These findings appear to be robust after accounting for other variables, such as non-cognitive and attitudinal traits (locus of control, risk preferences, self-confidence), income and educational level of parents, and other demographic characteristics that might drive both occupational choice and earnings.

Taken together, our findings provide a novel explanation for why individuals become entrepreneurs: entrepreneurs reject the traditional labor market, because it undervalues them. Asymmetric information about ability leads to only "lemons" or unproductive workers being traded on the market, that is, hired as employees. The workers who reserve their superior talents for their own benefits become entrepreneurs. This implication, that entrepreneurs are, in fact, "cherries" stands in contrast to a large literature in social science, which casts entrepreneurs individuals as "lemons"—those who either cannot find, cannot handle or cannot stand real jobs.

Our unique findings have several implications for entrepreneurship, education, and public policy. For example, our findings suggest that some entrepreneurs would be more productive as employees, but imperfect information about their ability creates a market failure. The only reason the "lemons problem" does not unravel the labor market is that for a large segment of the talent pool, traditional employment is more productive than entrepreneurship. However, if educational institutions start providing more precise signals of ability, they could induce more precise ability-based sorting into higher education, and the information asymmetry, and with it adverse selection, can be reduced. Hence, reduced information asymmetry might also imply a lower rate of entrepreneurship, although those who choose entrepreneurship will then be those whose productivity is highest in the occupation.

Going forward, we plan to more carefully test the conditional correlations suggested by the propositions as also more carefully examine alternative theories that may explain the patterns we have identified in the data.

References

[1] Akerlof, G. A., 1970. "The market for 'lemons': Quality uncertainty and the market mechanism." *The Quarterly Journal of Economics*, 488-500.

[2] Altonji, Joseph G, Prashant Bharadwaj and Fabian Lange, "Changes in the Characteristics of American Youth: Implications for Adult Outcomes," *NBER Working Paper* No. W13883, March 2008.

[3] Angrist, J. D. and Krueger, Alan B., 1991. "Does Compulsory School Attendance Affect Schooling and Earnings?" *The Quarterly Journal of Economics*, 106 (4): 979-1014.

[4] Åstebro, T., and Chen, J.,2014. "The entrepreneurial earnings puzzle: Mismeasurement or real?" *Journal of Business Venturing*, 29(1), 88-105.

[5] Åstebro, T., Chen, J., and Thompson, P. 2011. "Stars and Misfits: Self-Employment and Labor Market Frictions," *Management Science*, 57 (11): 1999-2017.

[6] Bagnoli, M., and Bergstrom, T. 2005. "Log-concave probability and its applications." *Economic Theory* 26: 445-469.

[7] Baumol WJ, Schilling MA, Wolff EN. 2009. "The superstar inventors an entrepreneurs: how were they educated?" *Journal of Economics and Management Strategy* 18(3):711–728.

[8] Becker, Gary S. 1964. *Human Capital*. New York: National Bureau of Economic Research., 1964

[9] Camerer, C., and Lovallo, D. 1999. "Overconfidence and excess entry: An experimental approach." *American Economic Review*, 306-318.

[10] Evans, D. S. and Leighton, L. S. 1989. "Some Empirical Aspects of Entrepreneurship." *American Economic Review*, 79 (3), pp. 519–535.

[11] Evans, D. S., and Boyan, J. 1989 "Estimates of a model of entrepreneurial choice under liquidity constraints." *Journal of Political Economy*, 97(4):808-827.

[12] Hamilton, Barton 2000 "Does entrepreneurship pay? An empirical analysis of the returns to self-employment." *Journal of Political Economy*, 108(3):604-631.

[13] Hansen, Karsten T. & Heckman, James J. & Mullen, K.J.Kathleen J., 2004. The effect of schooling and ability on achievement test scores, *Journal of Econometrics*, Elsevier, vol. 121(1-2), pages 39-98.

[14] Heckman, J. J., and LaFontaine, P. A., 2006. "Bias-Corrected Estimates of GED Returns." *Journal of Labor Economics*, 24(3).

[15] Hipple, S. F., 2010. "Self-employment in the United States." *Monthly Labor Review*, 133(9), 17-32.

[16] Holmström, B., 1999. "Managerial incentive problems: A dynamic perspective." *The Review of Economic Studies*, 66(1), 169-182.

[17] Jovanovic, Boyan, 1979. Job Matching and the Theory of Turnover, *Journal of Political Economy*, University of Chicago Press, vol. 87(5), pages 972-90, October.

[18] Kihlstrom, Richard E., and Jean-Jacques Laffont 1979: A general equilibrium entrepreneurial theory of firm formation based on risk aversion." *Journal of Political Economy*, 87(4):719-748.

[19] Klepper S, and Thompson P. 2010. Disagreements and intraindustry spinoffs. *International Journal of Industrial Organization* 28(5): 526–538.

[20] Lazear, E. P. 2005. "Entrepreneurship." *Journal of Labor Economics*, 23(4):649-680.

[21] Lazear, E. P. 2004. "Balanced skills and entrepreneurship." *American Economic Review*, 208-211.

[22] Levine, R., & Rubinstein, Y., 2013. "Smart and Illicit: Who Becomes an Entrepreneur and Does it Pay?" (No. w19276). National Bureau of Economic Research.

[23] McClelland, D. C. 1964. *The Achieving Society*, Princeton, NJ, D. Van Norstrand.

[24] Min, P. G. 1984. From white-collar occupations to small business: Korean immigrants occupational adjustment. *Sociological Quarterly* 25(3):333–352.

[25] Mincer, Jacob. 1970. On-the-Job Training: Costs, Returns and Some Implications. *Journal of Political Economy* 70, no. 5, pt. 2 (October 1962): 50-79.

[26] Neal, Derek and William R. Johnson. 1996. The Role of Premarket Factors in Black-White Wage Differences. *Journal of Political Economy* 104(5): 869-895.

[27] Roy, A. D. 1951. "Some thoughts on the distribution of earnings." *Oxford economic papers*, 3(2), 135-146.

[28] Sarada, 2010. "The unobserved returns to entrepreneurship." Working paper. Department of Economics, UCSD.

[29] Spence, Michael. 1973. "Job Market Signaling." *The Quarterly Journal of Economics*, 87(3). 355-374.

[30] Tumlinson, Justin. 2011. "Adverse Selection in Team Formation under Discrimination." *Working Paper.*

7 Theoretical Appendix

7.1 Base Model

Lemma 1 *Any interior solution to the FOC is the unique global optimum.*

Proof. Taking the derivative of (1)with respect to w yields the following Second Order Condition (SOC):

$$\left(\pi\left(w\right)-w\right)'F'\left(w|\,P\right)+\left(\pi\left(w\right)-w\right)F''\left(w|\,P\right)-F'\left(w|\,P\right)<0 \tag{3}$$

Note that we can write the FOC from (1) as

$$\left(\pi\left(w\right)-w\right)\frac{F'\left(w|\,P\right)}{F\left(w|\,P\right)}=1 \tag{4}$$

Thus, dividing (3) by $F\left(w|\,P\right)$ and multiplying the final term by the LHS of (4), we see that the SOC holds at a critical point iff

$$\left(\pi\left(w\right)-w\right)'\frac{F'\left(w|\,P\right)}{F\left(w|\,P\right)}+\left(\pi\left(w\right)-w\right)\left(\frac{F''\left(w|\,P\right)}{F\left(w|\,P\right)}-\left(\frac{F'\left(w|\,P\right)}{F\left(w|\,P\right)}\right)^{2}\right)<0$$

which is equivalent to

$$\left(\pi\left(w\right)-w\right)'\frac{F'\left(w|\,P\right)}{F\left(w|\,P\right)}+\left(\pi\left(w\right)-w\right)\left(\ln F\left(w|\,P\right)\right)''<0$$

Note that the first term is negative due to decreasing differences in the production technologies (*i.e.* for all θ, $(\pi(\theta)-\theta)' \leq 0$), and the second term is also negative because (a) any w satisfying (4) is less than $\pi(w)$ and (b) $F(w|P)$ is log concave. Thus, the SOC holds at all solutions of the FOC. ■

Lemma 3 *Wages strictly increase in pedigree (i.e. $w'(P) > 0$).*

Proof. Rewrite the FOC (eqn. (1))

$$\left((\pi(w)-w) - \int_{\underline{\theta}}^{w} \frac{F'(\theta|P)}{F'(w|P)} d\theta\right) F'(w|P) = 0$$

Since we have shown in Lemma 1 that the SOC holds at all critical points, from the Implicit Function Theorem (IFT), it suffices to show that for all P,

$$\frac{\partial}{\partial P}\left(\left((\pi(w)-w) - \int_{\underline{\theta}}^{w} \frac{F'(\theta|P)}{F'(w|P)} d\theta\right) F'(w|P)\right) > 0$$

or equivalently

$$\begin{aligned}&\left(-\int_{-\infty}^{w} \frac{\partial}{\partial P}\left(\frac{F'(\theta|P)}{F'(w|P)}\right) d\theta\right) F'(w|P) \\ &+\left((\pi(w)-w) - \int_{\underline{\theta}}^{w} \frac{F'(\theta|P)}{F'(w|P)} d\theta\right) \frac{\partial}{\partial P}\left(F'(w|P)\right) > 0\end{aligned}$$

The first factor of the second term is zero wherever the FOC is satisfied, yielding the condition

$$\int_{\underline{\theta}}^{w} \frac{\partial}{\partial P}\left(\frac{F'(\theta|P)}{F'(w|P)}\right) d\theta < 0$$

where MLRP implies that the integrand is negative for all θ, since $w > \theta$. ■

7.2 Endogenous Pedigree

In this subsection we show that under the assumptions of the endogenous pedigree first stage, the resulting posterior distribution of ability conditional on pedigree $F'(\theta|P)$ satisfies the MLRP. When more able individuals exert more effort in acquiring pedigree, this is indeed the case:

Lemma 5 *If equilibrium effort increases in ability, then the posterior density of ability $F'(\theta|P)$ satisfies the MLRP.*

Proof. Differentiating the likelihood ratio yields (using equation (2))

$$\begin{aligned}\frac{\partial}{\partial P}\frac{F'(\theta|P)}{F'(w|P)} &= \frac{F'(\theta)}{F'(w)}\frac{\partial}{\partial P}\frac{\Pr\{P|\theta\}}{\Pr\{P|w\}} = \frac{F'(\theta)}{F'(w)}\frac{\partial}{\partial P}\frac{G'(P-e(\theta))}{G'(P-e(w))} \\ &= \frac{F'(\theta)}{F'(w)}\left(\frac{G''(P-e(\theta))G'(P-e(w)) - G'(P-e(\theta))G''(P-e(w))}{G'(P-e(w))^2}\right)\end{aligned}$$

Thus, $\frac{\partial}{\partial P}\frac{F'(\theta|P)}{F'(w|P)} < 0$ for all $\theta < w$ if

$$\frac{G''\left(P - e\left(\theta\right)\right)}{G'\left(P - e\left(\theta\right)\right)} < \frac{G''\left(P - e\left(w\right)\right)}{G'\left(P - e\left(w\right)\right)}$$

which holds, because the log-concavity of G' implies $\left(\log\left(G'\right)\right)'' = \left(G''/G'\right)' < 0$, and $e\left(\theta\right) < e\left(w\right)$ implies $P - e\left(\theta\right) > P - e\left(w\right)$. ■

Note that average pedigree moves with ability exactly as effort does, because $E\left[P\left(e\left(\theta\right)\right)|\,\theta\right] = e\left(\theta\right) + E\left[\varepsilon\right]$ and the expected noise term is independent of ability. Thus, Lemma **??** can be equivalently written

Corollary 2 *If in equilibrium, average pedigree increases in ability, then the posterior density of ability $F'\left(\theta|\,P\right)$ satisfies the MLRP.*

In other words, so long as a separating equilibrium exists in the first stage signalling game, such that higher ability individuals receive higher pedigrees on average, all of the results of the baseline model also hold when pedigree is endogenously determined. As previously stated, that pedigrees increase in ability can readily be observed in the data, but for completeness we establish that this is true under our assumed conditions, which are standard to signaling models:

Lemma 6 *Equilibrium effort exerted in pursuit of pedigree increases in ability.*

Proof. An individuals' Stage 1 problem can be written

$$\max_{e} \theta \Pr\left\{\theta > w\left(P\right)\right\} + E\left[w\left(P\right)|\,\theta \leq w\left(P\right)\right]\Pr\left\{\theta \leq w\left(P\right)\right\} - C\left(e,\theta\right)$$

or in integral form

$$\max_{e} \theta G\left(w^{-1}\left(\theta\right) - e\right) + \int_{w^{-1}(\theta)-e}^{+\infty} w\left(e+\varepsilon\right) dG\left(\varepsilon\right) - C\left(e,\theta\right)$$

Differentiating with respect to e yields the following FOC

$$-\theta G'\left(w^{-1}\left(\theta\right) - e\right) + w\left(e + w^{-1}\left(\theta\right) - e\right) G'\left(w^{-1}\left(\theta\right) - e\right) + \int_{w^{-1}(\theta)-e}^{+\infty} w'\left(e+\varepsilon\right) dG\left(\varepsilon\right) - C_1\left(e,\theta\right) = 0$$

which readily simplifies to the marginal benefit of all possible 'good' pedigree realizations resulting from an increase in effort must equal the marginal cost of that effort increase:

$$\int_{w^{-1}(\theta)-e}^{+\infty} w'\left(e+\varepsilon\right) dG\left(\varepsilon\right) = C_1\left(e,\theta\right)$$

Thus, a (continuously) separating equilibrium exists if the solution to this implicit equation for e strictly increases in θ. We have assumed that cost is sufficiently convex for the SOC to be satisfied. Thus, from the IFT, $\frac{de}{d\theta}$ has the same sign as

$$\begin{aligned} & \frac{\partial}{\partial \theta}\left(\int_{w^{-1}(\theta)-e}^{+\infty} w'\left(e+\varepsilon\right) dG\left(\varepsilon\right) - C_1\left(e,\theta\right)\right) \\ = & -w'\left(w^{-1}\left(\theta\right)\right) G'\left(w^{-1}\left(\theta\right)-e\right) \frac{\partial}{\partial \theta}\left(w^{-1}\left(\theta\right)\right) - C_{12}\left(e,\theta\right) \\ = & G'\left(w^{-1}\left(\theta\right)-e\right) - C_{12}\left(e,\theta\right) > 0 \end{aligned}$$

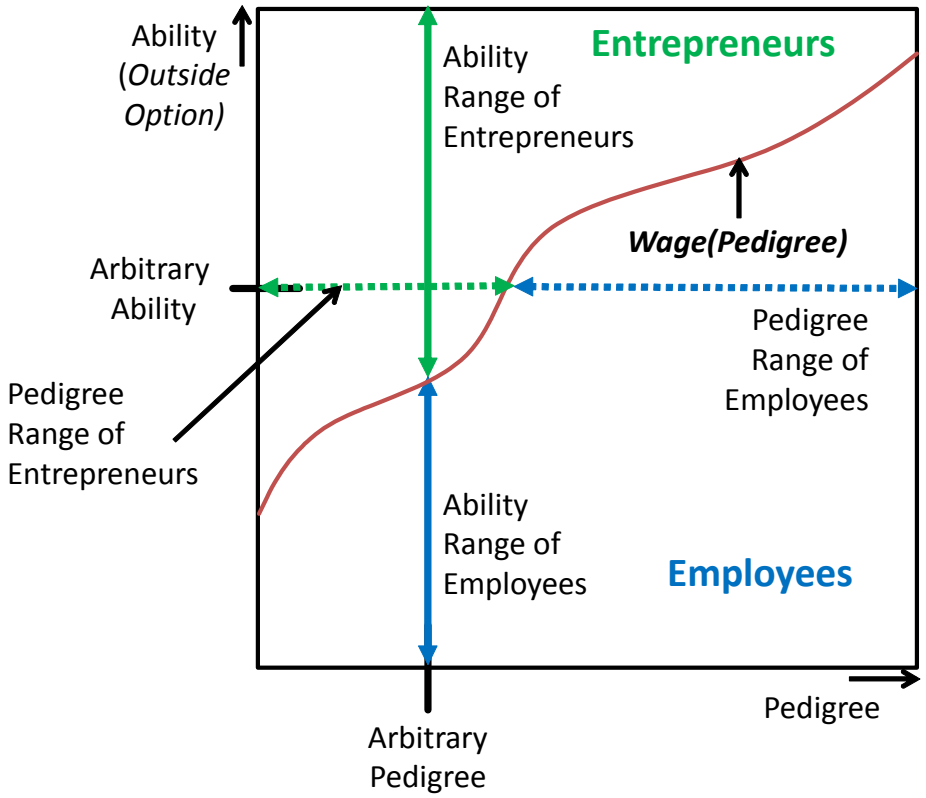
which is positive, because G' is a probability density and $C_{12} < 0$ by assumption (*i.e.* the Spence-Mirrlees condition). ■

Lemma 4 *Equilibrium effort exerted in pursuit of pedigree, and average pedigree, increases in ability, such that the posterior density of ability $F'\left(\theta \middle| P\right)$ satisfies the MLRP.*

Proof. The result follows immediately from Lemmas 6 and 5, together with the fact that average pedigree moves with ability exactly as effort does, because $E\left[P\left(e\left(\theta\right)\right) \middle| \theta\right] = e\left(\theta\right) + E\left[\varepsilon\right]$ and the expected noise term is independent of ability. ■

Figures and Tables

Figure 1: Ability, Pedigree, and Entrepreneurship



The figure represents the pedigree-ability sample space. The horizontal axis denotes increasing pedigree. The vertical axis denotes increasing ability (equivalently entrepreneurial payoff). Wage, a function of pedigree, bisects the space, such that all individuals of an arbitrary pedigree with entrepreneurial productivity exceeding the wage commanded by that pedigree choose entrepreneurship. Thus, conditional on pedigree, entrepreneurs have higher ability than employees. The inverse wage function expresses the minimum pedigree an individual with a given ability would require to choose employment. Thus, conditional on ability, entrepreneurs have lower pedigree than employees.

Figure 2: Distribution of ability scores, Self-employed v/s Wage-employed, NLSY 1994 sample

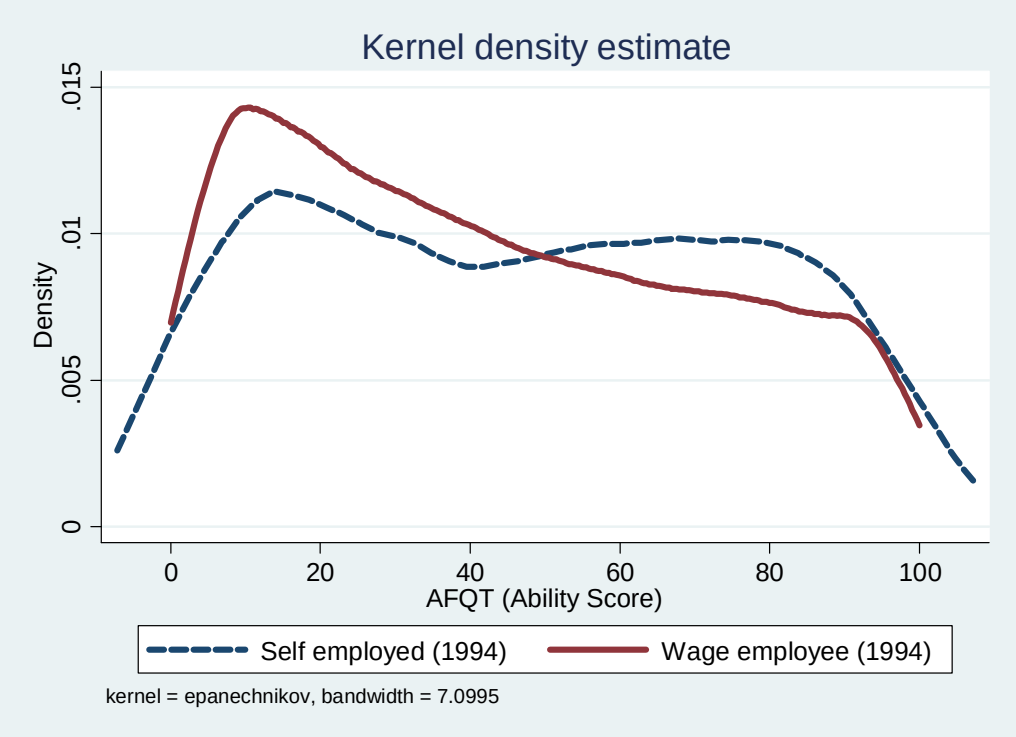


Figure shows the kernel density distribution of Armed Forces Qualification Test (AFQT) scores, expressed in percentiles, for self-employed individuals and wage employees. AFQT scores are based on the arithmetic reasoning, word knowledge, paragraph comprehension, and numerical operations sections of the Armed Services Vocational Aptitude Battery (ASVAB). NLSY respondents took the ASVAB battery in 1980. Employment status represents respondent employment status in 1993 (from NLSY 1994 sample).

Figure 3: Distribution of income and wealth, Self-employed v/s Wage-employed, NLSY 1994 Sample

Figure 3a: Distribution of Log Annual Income

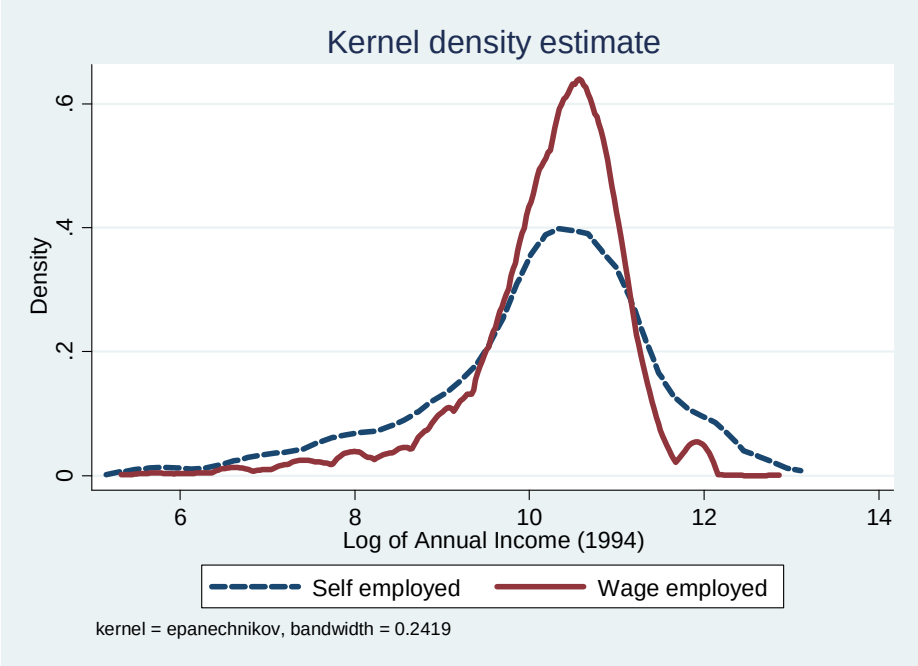


Figure 3b: Distribution of Log Pay Rate

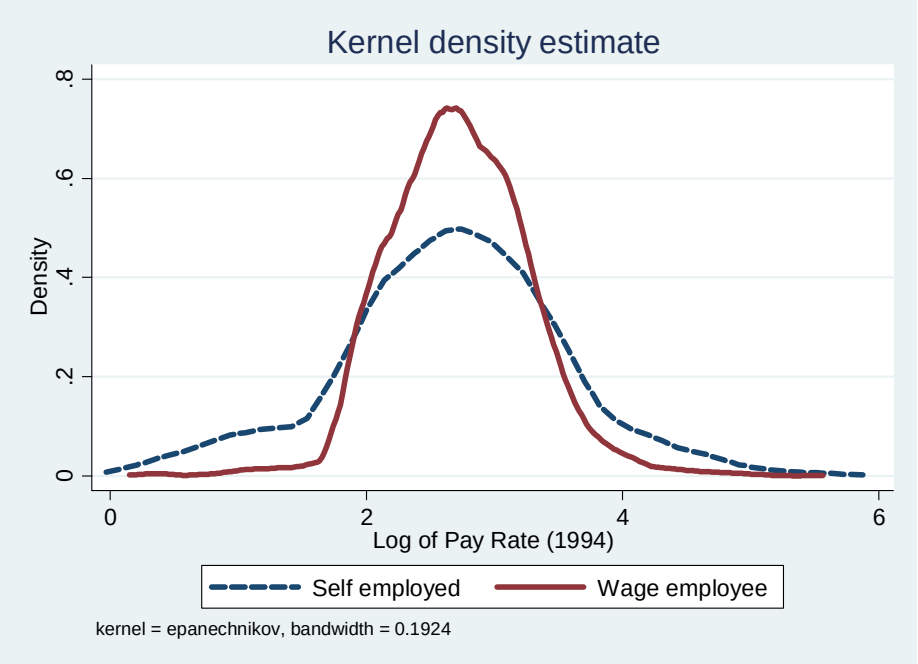
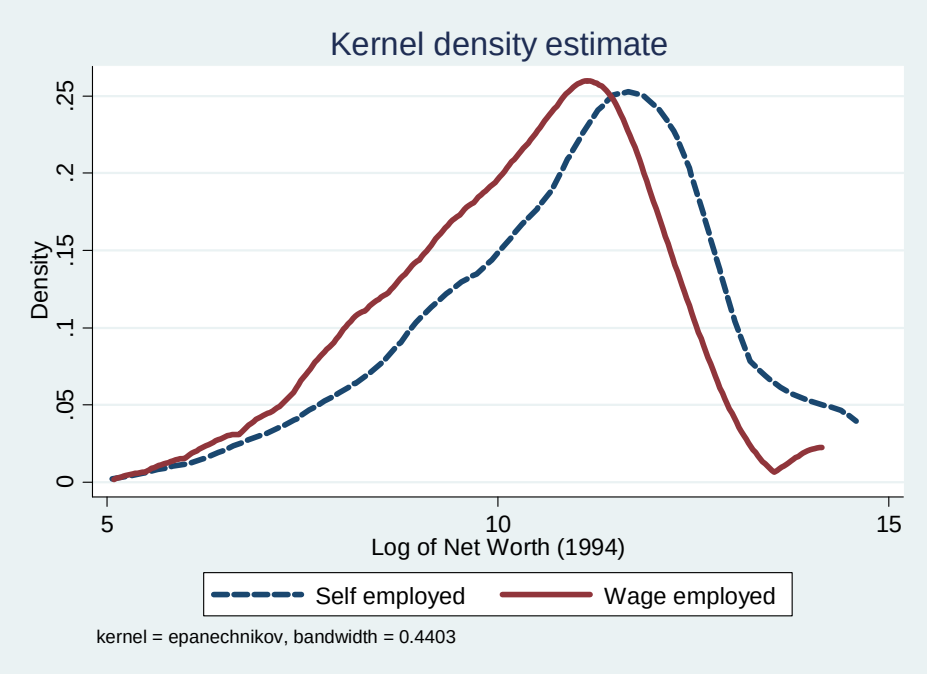


Figure 3c: Distribution of Net Worth



Figures show the kernel density distribution of logged annual income (Figure 3a), logged pay rate expressed in $/hr (Figure 3b), and net worth (Figure 3b) for self-employed individuals and wage employees in the NLSY 1994 sample. 8.2 percent of the 6,921 full-time employed individuals in the 1994 NLSY sample were self-employed, and the rest were wage-employees. Following Table reports the sample descriptives for the three variables.

	Net Annual Income ($)			Hourly pay rate ($/hr)			Net Worth ($)		
	Mean	Median	SD	Mean	Median	SD	Mean	Median	SD
Wage-employed	36,609	32,340	26,201	17.9	14.8	13.0	86,570	30,870	192,140
Self-employed	45,805	29,400	54,727	21.0	14.7	24.7	179,439	63,578	329,601

Figure 4: Distribution of income and wealth, Self-employed v/s Wage-employed, NLSY 1994

Figure 4a: Distribution of Annual Income

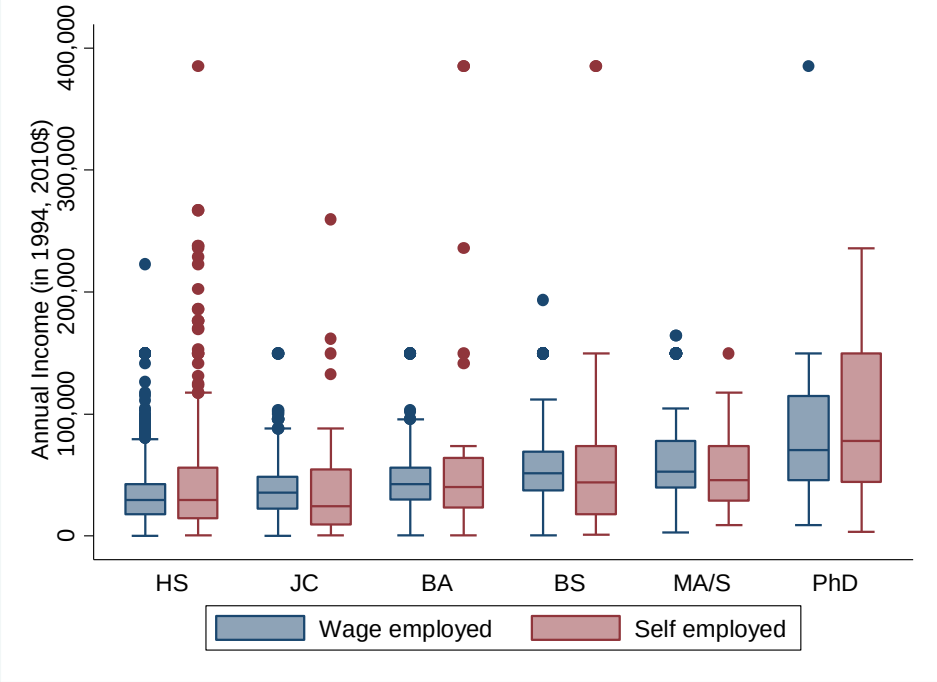


Figure 4b: Distribution of Net Worth

X-axis denotes highest degree achieved, such that HS indicates High School Diploma (or equivalent), JC = Junior/Associate College, BA=Bachelor of Arts, BS=Bachelor of Science, MA/S=Master's, and PhD includes doctoral and other advanced professional degrees (MD, LLD, DDS).

Figure 5: Distribution of ability scores, Self-employed v/s Wage-employed, U.K. NCDS Sample

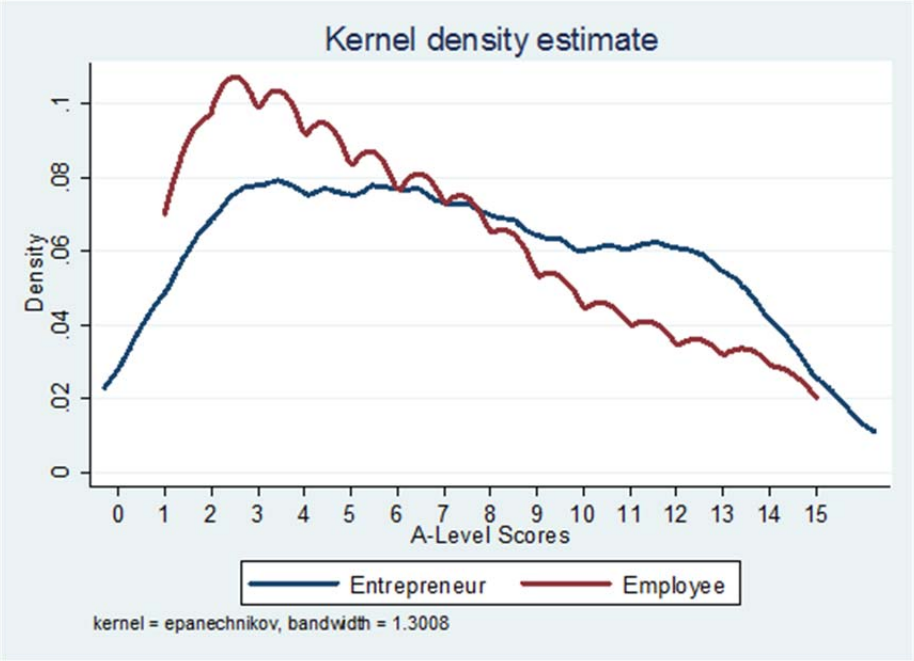


Table 1: The National Longitudinal Survey of Youth (NLSY) Samples and Self-employment Rates

NLSY Sample-year	Respondents	Respondents age	Fulltime Employed	NLSY Self-employment rate	CPS Self-employment rate
1979	12,686	14-22	5,108	3.5%	
1994	8,891	30-37	6,921	8.2%	11.0%
1998	8,399	34-41	7,154	8.8%	10.2%
2002	7,724	38-45	6,595	10.3%	10.1%
2008	7,757	44-52	6,355	12.7%	10.8%

Table shows the number of respondents in each of the different years of the NLSY survey. The first round of the NLSY surveyed 12,686 individuals in 1979, representative of the population of U.S. youths between 14 and 22. Subsequent rounds of the NLSY then tracks and surveys the same individuals either annually or on a biennial basis. Following 1984, 1,079 members of the military subsample of NLSY were no longer eligible for the survey. Following 1990, 1,643 members of the economically disadvantaged, nonblack/non-Hispanic sample dropped out of the sample, leaving 9,964 or 78.5 percent of the original sample of respondents. Each subsequent survey experienced a 5-10 percent drop out rate, primarily due to death or exit out of the U.S. of the respondents. NLSY Self-employment rate (Column 5) is calculated as a percent of full-time employed individuals (worked for 40 hours or more per week for more than size months of the previous calendar year) who reported being self-employed for their primary job. Current Population Survey (CPS) self-employment rates (Column 6) are from Hipple (2010) and show the percent of full-time employed individuals that are self-employed in the U.S. population.

Table 2: Educational attainment in the 1994 and 2008 NLSY Samples

PANEL A

Education Level	1994	2008
None	0.1%	0.1%
1ST Grade	0.0%	0.0%
3RD Grade	0.1%	0.1%
4TH Grade	0.1%	0.1%
5TH Grade	0.1%	0.1%
6TH Grade	0.3%	0.3%
7TH Grade	0.5%	0.4%
8TH Grade	1.6%	1.3%
9TH Grade	2.8%	2.2%
10TH Grade	3.3%	2.1%
11TH Grade	3.9%	2.6%
12TH Grade	43.8%	43.1%
1ST Year College	9.2%	9.0%
2ND Year College	9.7%	10.7%
3RD Year College	4.9%	5.3%
4TH Year College	11.9%	12.0%
5TH Year College	2.8%	3.0%
6TH Year College	2.7%	3.9%
7TH Year College	1.2%	1.6%
8TH Year College or more	1.1%	2.2%
Observations	7,010	6,446

PANEL B

Highest Degree	1994	2008
High school diploma (or equivalent)	68.7%	64.8%
Associate/Junior College (AA)	9.2%	11.2%
Bachelor of Arts Degree (BA)	6.6%	5.1%
Bachelor of Science (BS)	10.5%	11.4%
Master s Degree (MA, MBA, MS, MSW)	3.8%	6.2%
PhD, MD, LLD, DDS	1.3%	1.4%
Observations	7,010	6,446

PANEL C

College Rank	1994	2008
1 (Top 50)	1.3%	1.2%
2 (Top 50-200)	3.5%	3.5%
3 (>200)	83.3%	82.6%
4 (No info)	11.9%	12.7%
Observations	4,646	4,909

Panel A shows the distribution of respondents by their educational level, that is, number of years of school/college attended. Panel B shows the distribution of respondents by their highest degree. Panel C shows the distribution of respondents by the ranking of the institution of their latest degree. Rankings are constructed based on admissions selectivity data of the institutions available from the National Center for Education Statistics.

Table 3: Income and Earnings in the NLSY Samples

	Observations	Mean	Median	S.D.	Min	Max
1994 Sample						
Total income	6,966	$36,544	$30,870	$30,525	$9	$385,283
Total family income	7,004	$60,098	$48,804	$53,586	$0	$279,179
Total net family assets	7,546	$92,576	$27,195	$215,664	$0	$1,390,000
Hourly pay rate	7,152	$18	$15	$15	$0	$294
1998 Sample						
Total income	6,641	$42,365	$33,848	$37,368	$15	$403,467
Total family income	6,864	$64,038	$53,600	$56,205	$0	$327,420
Total net family assets	7,160	$148,875	$48,240	$345,294	$0	$2,230,000
Hourly pay rate	7,314	$20	$16	$21	$0	$870
2002 Sample						
Total income	6,228	$48,593	$37,510	$48,437	$0	$524,233
Total family income	6,535	$74,239	$58,080	$74,552	$0	$472,701
Total net family assets	7,004	$190,013	$63,500	$424,833	$0	$2,720,000
Hourly pay rate	6,634	$22	$17	$24	$0	$636
2008 Sample						
Total income	6,228	$49,079	$37,885	$48,922	$0	$529,475
Total family income	6,727	$74,120	$56,560	$76,262	$0	$459,284
Total net family assets	6,881	$293,371	$110,090	$566,961	$0	$3,480,000
Hourly pay rate	6,420	$22	$17	$22	$0	$806

Table shows sample statistics of various income and wealth variables for the 1994, 1998, 2002 and 2008 NLSY samples. All figures are in 2010 dollars. The NLSY top-codes the earnings and wealth variables by using the mean value of the top 2 percent for all respondents that belong to the top 2 percent for the corresponding variable.

Table 4: Control Variables

PANEL A: Continuous and Ordinal Variables

Control Variable	Observations	Mean	Median	SD	Min	Max
Self Esteem	7,827	23.52	23	4.12	3	30
Pearlin Mastery Score	7,984	22.16	22	3.2	7	28
Sociability	8,082	2.88	3	0.7	1	4
Trusting of others	7,076	3.02	3	1.01	1	5
Risk-averse or risk-taking	6,911	4.84	5	2.91	0	10
Family net worth in 1985	7,886	$23,833	$4,586	$78,490	-$412,000	$540,071
Family net income in 1979	6,694	$50,901	$42,000	$38,836	$60	$225,003

PANEL B: Binary Variables

Control Variable	Percent of Sample
Sex	
Male	50.5%
Female	49.5%
Non-missing observations	8,753
Race	
White	70.1%
Black	29.9%
Non-missing observations	8,753
Marital Status in 1990	
Never Married	33.1%
Married	52.0%
Separated	5.2%
Divorced	9.3%
Widowed	0.4%
Non-missing observations	8,286

Table shows sample statistics for our control variables. The variables are collated from various rounds of the NLSY surveys. Respondents' Occupation (CPS 3-digit), Respondents' Industry (SIC 3- digit), and Education and occupation of respondents' parents are also used as control variables in some estimations, but the sample statistics for these variables are not reported here due to space constraints.

Table 5: Ability (AFQT) Scores by Educational Qualification, Self-employed v/s Wage-employed, (NLSY 1994)

PANEL A

Education Level	Percent of 1994 Sample		Mean AFQT (Ability Score)			
	Wage employees	Self-employed	Wage employees	Self-employed	Difference	S.E. (ttest)
3RD GRADE	0.0%	0.0%	6			
4TH GRADE	0.1%	0.3%	2.3	1	-1.3	
5TH GRADE	0.0%	0.1%	1			
6TH GRADE	0.3%	0.4%	6.7	3.5	-3.2	4.5
7TH GRADE	0.3%	0.7%	12.6	16.5	3.9	7.7
8TH GRADE	1.5%	2.0%	9.7	11.1	1.4	2.7
9TH GRADE	2.4%	2.3%	13.9	14.3	0.4	3.7
10TH GRADE	2.7%	4.0%	17.4	24	6.6	3.2*
11TH GRADE	3.4%	3.6%	17.6	22	4.4	3.2
12TH GRADE	43.7%	41.1%	34.6	38.4	3.8	1.4**
1ST YEAR COL.	9.3%	9.7%	42.4	48	5.6	3.0+
2ND YEAR COL.	9.9%	10.3%	49.9	57.1	7.2	3.1*
3RD YEAR COL.	5.1%	4.5%	49.8	62.5	12.7	4.4**
4TH YEAR COL.	12.8%	12.6%	66.3	73.9	7.6	2.6**
5TH YEAR COL.	3.1%	2.1%	68.1	73.3	5.2	6.6
6TH YEAR COL.	3.0%	2.6%	75.3	73.4	-1.9	5.1
7TH YEAR COL.	1.3%	1.6%	79.8	86.2	6.4	6.1
8TH YEAR COL.	1.1%	2.1%	79.8	87.5	7.7	5.1
Observations	6,428	771				

PANEL B

Highest Degree	Percent of 1994 Sample		Mean AFQT (Ability Score)			
	Wage employees	Self-employed	Wage employees	Self-employed	Difference	S.E. (ttest)
High school diploma	67.2%	66.6%	38.3	44.1	5.8	1.3**
Associate/Junior College	9.5%	9.3%	49.7	59.6	9.9	3.6**
Bachelor of Arts	6.9%	7.5%	67	73	6	3.8+
Bachelor of Science	11.1%	10.8%	67.9	74.2	6.3	3*
Master s Degree	4.2%	3.1%	75.6	81.1	5.5	5.2
PhD, MD, LLD, DDS	1.2%	2.6%	86.4	90.9	4.5	4.3
Observations	5,330	611				

PANEL C

College Tier	Percent of 1994 Sample		Mean AFQT (Ability Score)			
	Wage employees	Self-employed	Wage employees	Self-employed	Difference	S.E. (ttest)
1 (Top 50)	1.5%	0.5%	76.4	90	13.6	19.9
2 (Top 50-200)	3.7%	3.9%	63.3	66.6	3.3	8.2
3 (>200)	83.1%	86.3%	54.9	62.6	7.7	1.5**
4 (No info)	11.8%	9.4%	47.7	52.5	4.8	4.7
Observations	3,574	416				

Table compares the mean values of AFQT test scores, our measure of ability, for self-employed and wage-employed individuals by the individuals' educational qualifications. PANEL A reports the scores by educational level, PANEL B by highest degree obtained, and PANEL C by the institutional rank of the last degree obtained. Employment status represents respondent employment in 1993 (collected by the NLSY 1994 survey). Data on highest educational attainment reflect each respondent's highest attainment as of 1994, when they were aged between 30 and 37. For t-tests of differences in means, ** represents $p<0.01$, * $p<0.05$, and + $p<0.1$

Table 6: Educational qualifications by ability decile, Self-employed v/s Wage-employed (NLSY 1994 Sample)

AFQT Decile	Education Level				Highest Degree				College Selectivity Tier			
	Wage employed	Self employed	Difference	S.E. (ttest)	Wage employed	Self employed	Difference	S.E. (ttest)	Wage employed	Self employed	Difference	S.E. (ttest)
1	10.84	10.18	-0.66	[0.36]+	1.01	1	-0.01	[0.03]	2.67	2	-0.67	[0.67]
2	11.66	11.46	-0.2	[0.19]	1.16	1.15	-0.01	[0.08]	2.56	3	0.44	[0.68]
3	12.47	11.7	-0.77	[0.20]**	1.29	1.12	-0.17	[0.12]	2.69	2.75	0.06	[0.28]
4	12.79	12.5	-0.29	[0.20]	1.32	1.29	-0.03	[0.11]	2.78	3	0.22	[0.18]
5	13.11	12.68	-0.43	[0.24]+	1.55	1.25	-0.3	[0.15]*	2.8	2.75	-0.05	[0.25]
6	13.43	12.9	-0.53	[0.26]*	1.71	1.61	-0.1	[0.17]	2.88	3	0.12	[0.15]
7	13.71	13.29	-0.42	[0.25]+	1.84	1.47	-0.37	[0.16]*	2.81	2.92	0.11	[0.13]
8	14.17	14.28	0.11	[0.27]	2.03	2.08	0.05	[0.18]	2.85	2.64	-0.21	[0.13]+
9	14.65	15.07	0.42	[0.29]	2.43	2.49	0.06	[0.21]	2.67	2.86	0.19	[0.15]
10	15.93	15.68	-0.25	[0.24]	3.33	3.23	-0.1	[0.19]	2.62	2.74	0.12	[0.13]

Table shows means, differences in means, and results of ttests for differences in mean educational attainment of wage-employed and self-employed individuals by decile of AFQT (ability) scores. "Education Level" is measured on an ordinal scale such that 1 = completion of 1st grade, 2 = completion of 2nd grade, ..12= completion of 12th grade (high-school), 16=completion of four-year college, and so on. "Highest Degree" is measured on a scale indicating 1=High School Diploma (or equivalent), 2=Associate/Junior College, 3=Bachelor of Arts, 4=Bachelor of Science, 5=Master's, and 6 = PhD or other advanced professional degrees (MD, LLD, DDS). College Selectivity Tier is measured such that 1 = Top 50, 2 = Top 50-200 and 3 >200 (We omit the observations for which we do not have information on the selectivity rank of the respondents' educational institutions). Data on highest educational attainment reflect respondents' highest attainment as of 1994, when they were aged between 30 and 37. For t-tests, ** $p<0.01$, * $p<0.05$, + $p<0.1$

Table 7: Income and Wealth by Educational Qualifications, Self-employed v/s Wage-employed (NLSY 1994 Sample)

PANEL A: Annual Income by Education Level

Education Level	Mean Annual Income ($)				SD Annual Income ($)	
	Wage employed	Self-employed	Difference	S.E. (ttest)	Wage employed	Self-employed
6TH GRADE	22,352	30,870	8,518	[10,025]	18,214	17,681
7TH GRADE	21,225	22,050	825	[9,308]	17,532	12,181
8TH GRADE	23,158	60,448	37,290	[7,228]**	15,011	54,984
9TH GRADE	23,222	20,627	-2,595	[4,310]	16,025	13,952
10TH GRADE	23,519	24,885	1,366	[3,918]	15,277	31,672
11TH GRADE	23,644	31,529	7,884	[3,998]*	15,887	34,048
12TH GRADE	30,422	41,965	11,543	[1,536]**	19,425	49,627
1ST YEAR COL.	34,668	46,478	11,811	[3,675]**	23,349	56,861
2ND YEAR COL.	37,245	37,434	189	[3,018]	22,547	32,214
3RD YEAR COL.	36,984	61,670	24,686	[6,182]**	22,115	78,111
4TH YEAR COL.	52,929	60,397	7,468	[4,133]+	31,004	72,048
5TH YEAR COL.	47,881	68,938	21,057	[9,673]*	27,434	93,770
6TH YEAR COL.	60,848	57,313	-3,535	[8,921]	35,218	44,720
7TH YEAR COL.	70,268	37,241	-33,027	[13,805]*	43,325	39,394
8TH YEAR COL.	71,165	91,104	19,939	[16,589]	54,526	64,335

Table 7, PANEL B: Pay Rate ($/hr) by Education Level

Education Level	Mean Pay Rate ($/hr)				SD Pay Rate ($/hr)	
	Wage employed	Self-employed	Difference	S.E. (ttest)	Wage employed	Self-employed
6TH GRADE	11.3	19.2	7.9	[4.7]	5.1	19.0
7TH GRADE	10.3	9.5	-0.8	[2.2]	4.0	6.8
8TH GRADE	13.5	32.4	18.9	[4.9]**	12.9	32.8
9TH GRADE	13.1	13.8	0.7	[2.6]	9.1	8.6
10TH GRADE	13.5	13.5	0.0	[2.5]	10.7	17.3
11TH GRADE	14.6	13.5	-1.2	[3.5]	17.1	9.9
12TH GRADE	15.0	18.5	3.5	[0.7]**	9.2	22.8
1ST YEAR COL.	16.6	19.0	2.4	[1.3]+	9.5	15.7
2ND YEAR COL.	18.5	19.0	0.4	[1.6]	11.5	15.6
3RD YEAR COL.	18.5	35.6	17.1	[3.7]**	10.3	56.0
4TH YEAR COL.	24.6	24.6	0.0	[1.6]	13.1	22.1
5TH YEAR COL.	23.9	26.5	2.6	[4.6]	15.5	23.8
6TH YEAR COL.	27.6	24.0	-3.6	[4.4]	17.1	18.5
7TH YEAR COL.	30.6	26.2	-4.4	[5.3]	16.5	15.2
8TH YEAR COL.	39.6	58.4	18.9	[13.8]	46.6	47.1

Table 7, PANEL C: Net Worth by Education Level

Education Level	Mean Net Worth ($)				SD Net Worth ($)	
	Wage employed	Self-employed	Difference	S.E. (ttest)	Wage employed	Self-employed
6TH GRADE	44,514	65,650	21,136	[22,822]	45,471	40,876
7TH GRADE	23,138	4,410	-18,728	[26,606]	25,638	
8TH GRADE	24,511	58,785	34,274	[11,828]**	33,955	57,798
9TH GRADE	29,501	47,745	18,244	[16,333]	51,409	95,819
10TH GRADE	45,615	37,627	-7,989	[28,508]	135,033	44,874
11TH GRADE	34,262	84,145	49,883	[14,305]**	50,918	119,931
12TH GRADE	64,674	172,845	108,171	[11,188]**	140,701	319,767
1ST YEAR COL.	86,969	148,892	61,923	[26,712]*	196,378	250,138
2ND YEAR COL.	99,888	205,921	106,032	[29,394]**	203,469	354,267
3RD YEAR COL.	128,360	340,921	212,561	[60,560]**	270,238	480,477
4TH YEAR COL.	146,480	262,434	115,953	[29,373]**	241,786	379,510
5TH YEAR COL.	118,325	206,359	88,034	[57,060]	198,016	342,623
6TH YEAR COL.	203,279	354,097	150,818	[91,087]+	326,518	524,400
7TH YEAR COL.	260,001	287,486	27,485	[152,291]	401,102	467,468
8TH YEAR COL.	214,565	499,477	284,912	[126,754]*	370,968	524,294

Table shows mean and standard deviation of annual income, hourly wage rate, and net family worth (assets - liabilities) for wage-employed and self-employed individuals in the NLSY 1994 sample. Income and wealth variables are in 2010 US dollars, and wages in dollars per hour. Income and wage data reflect respondent earnings associated with their employment in 1993 (data collected in 1994). Descriptives are based on the responses of 6,910 full-time employed individuals who responded to the 1994 round of the NLSY79 and had non-missing values for education level, employment and wages/wealth. For t-tests, ** $p<0.01$, * $p<0.05$, + $p<0.1$

Table 8: Industry and Occupation of Wage-Employed and Self-Employed (NLSY 1994 Sample)

PANEL A

Industry	Wage employed	Self-employed
Agriculture, Forestry and Fishing	2.0%	6.6%
Mining	0.5%	0.1%
Construction	6.2%	17.3%
Manufacturing	19.4%	5.8%
Transportation, Communication, Utilities	8.6%	5.2%
Wholesale Trade	2.9%	1.8%
Retail Trade	13.6%	11.4%
Finance, Insurance, Real Estate	6.6%	3.8%
Business and Repair Services	6.9%	15.5%
Personal Services	3.1%	16.1%
Entertainment and Recreation	1.0%	2.0%
Professional and Related Services	22.8%	13.0%
Public Administration	6.5%	1.4%
Observations	6,093	710

PANEL B

Occupation	Wage employed	Self-employed
Professional, Technical, Kindred	17.8%	14.6%
Managers, Proprietors	12.3%	17.1%
Sales Workers	3.9%	4.7%
Clerical and Kindred	19.6%	6.4%
Craftsmen, Foremen and Kindred	10.8%	16.3%
Armed Forces	0.1%	0.0%
Operatives and Kindred	13.7%	7.4%
Laborers, except farm	6.0%	9.6%
Farmers and farm managers	0.1%	2.1%
Farm Laborers and Foremen	0.7%	0.8%
Service Workers	14.8%	17.4%
Private Household Workers	0.2%	3.3%
Not reported	0.1%	0.3%
Observations	6,182	719

Table 9: Control Variables—Wage-employed v/s Self-employed (NLSY 1994 Sample)

PANEL A: Continuous and Ordinal Variables

Control Variable	Wage Employed			Self-Employed		
	Observations	Mean	SD	Observations	Mean	SD
Self Esteem	6,072	23.6	4.1	717	23.9	4.1
Pearlin Mastery Score	6,269	22.2	3.2	747	22.4	3.3
Sociability	6,266	2.9	0.7	749	3.0	0.7
Trusting of others	5,466	3.0	1.0	638	3.0	1.0
Risk-averse or risk-taking	5,323	4.8	2.9	619	5.4	2.9
Family net worth in 1985	6,122	$22,466	$75,070	719	$42,430	$108,504
Family net income in 1979	5,214	$51,333	$38,398	615	$60,199	$46,541

PANEL B: Binary Variables

Control Variable	% of Wage-employed	% of Self-employed
Sex		
Male	51.9%	59.9%
Female	48.1%	40.2%
Observations	6,471	782
Race		
Hispanic	19.3%	15.7%
Black	29.5%	18.8%
Non-Black Non-Hispanic	51.2%	65.5%
Observations	6,471	782
Marital Status in 1990		
Never Married	33.6%	27.7%
Married	52.0%	59.0%
Separated	5.1%	3.8%
Divorced	9.1%	9.2%
Widowed	0.3%	0.3%
Observations	6,203	736

Control variables are collected from various rounds of the NLSY survey. Employment status is as reported in the NLSY 1994 round.

Table 10: Probit estimates of the relationship between the Probability of Self-employment and Ability/Education

D.V. = Self-employed (0/1)	1	2	3	4	5	6
Ability (AFQT) Score	0.003**	0.005**	0.004**	0.004**	0.003*	0.004**
	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]
Highest Degree		-0.047**	-0.032**	-0.033**	-0.038**	-0.026+
		[0.010]	[0.010]	[0.011]	[0.012]	[0.013]
Male			0.171**	0.167**	0.165**	0.053
			[0.042]	[0.042]	[0.049]	[0.056]
White			0.241**	0.233**	0.235**	0.226**
			[0.056]	[0.056]	[0.068]	[0.070]
Log Age			1.437**	1.433**	1.671**	1.843**
			[0.457]	[0.458]	[0.542]	[0.564]
Risk Loving				0.012*	0.018**	0.018**
				[0.006]	[0.007]	[0.007]
Sociable				0.014	0.069*	0.068+
				[0.016]	[0.033]	[0.035]
Trusting				-0.016+	-0.016	-0.013
				[0.009]	[0.010]	[0.011]
Self- Esteem				-0.003	-0.007+	-0.007+
				[0.003]	[0.004]	[0.004]
Self-Mastery				0	-0.004	-0.004
				[0.004]	[0.005]	[0.005]
Father Self-employed?					0.230**	0.260**
					[0.077]	[0.080]
Mother Self-employed?					-0.213	-0.267
					[0.170]	[0.177]
Mother's education					0.018*	0.015+
					[0.008]	[0.008]
Father's education					0.004	0.004
					[0.005]	[0.006]
Family Income in 1979					0.049	0.043
					[0.032]	[0.033]
Family assets in 1985					0.001**	0.001**
					[0.000]	[0.000]
Other Dem. Controls	N	N	Y	Y	Y	Y
Occupational Controls	N	N	N	N	N	Y
Constant	-1.382	-1.442	-7.25	-7.206	-8.806	-8.873
Log-likelihood	-2309.24	-2235.89	-2203.96	-2200.46	-1630.49	-1518.27
Observations	6,887	6,718	6,717	6,717	5,195	5,165

Table displays Probit estimates of the relationship between the probability of being self-employed and explanatory variables. Educational attainment is measured by highest degree obtained on a scale indicating 1=High School Diploma (or equivalent), 2=Associate/Junior College, 3=Bachelor of Arts, 4=Bachelor of Science, 5=Master's, and 6 = PhD or other advanced professional degrees (MD, LLD, DDS). Employment status represents respondent employment in 1993 (data collected in 1994). Other Demographic Controls include dummies for marital status, family size, and country of birth not shown due to space constraints. Occupational controls include 13 dummy variables for SIC-3 based occupational classifications associated with individuals' employment in 1993. Robust standard errors are shown in brackets. We use **, *, and + to denote $p<0.01$, $p<0.05$ and $p<0.1$, respectively.

Table 11: OLS estimates of the relationship between Earnings/Net worth and Self-employment

Dependent Variable	1 Annual Income	2 Annual Income	3 Annual Income	4 Pay Rate	5 Net Worth
Self-employed (0/1)	9,196.2**	10,262.1**	9,171.2**	3.3**	85,290.7**
	[2,129.5]	[2,158.7]	[2,505.2]	[1.2]	[15,300.6]
Highest Degree		3,730.3**	2,065.9**	0.7**	8,622.8**
		[175.0]	[203.8]	[0.1]	[1,446.4]
Male			14,662.5**	3.3**	-1,626.10
			[913.7]	[0.4]	[7,537.4]
White			2,358.5**	1.8**	41,126.3**
			[772.0]	[0.4]	[5,788.8]
Log Age			49,018.9**	17.6**	309,249.3**
			[7,935.5]	[4.3]	[69,082.9]
Risk Loving			68.7	0.1	1,164.10
			[103.9]	[0.1]	[820.9]
Sociable			605.7	0	398
			[464.6]	[0.3]	[5,070.5]
Trusting			-413.0**	-0.3**	-1,038.70
			[151.4]	[0.1]	[1,129.0]
Self- Esteem			197.2**	0.1**	505.2
			[61.0]	[0.0]	[418.4]
Pearlin Mastery			192.1+	0	1,532.4**
			[113.6]	[0.1]	[520.7]
Father Self-employed?			4,269.1*	1.6+	24,630.9+
			[1,796.1]	[1.0]	[14,691.7]
Mother Self-employed?			-2,370.00	-1.9+	46,097.20
			[2,567.7]	[1.0]	[33,532.4]
Mother's education			141.6	0.1	3,419.2**
			[99.2]	[0.1]	[740.4]
Father's education			77.7	0	-382.2
			[67.9]	[0.0]	[660.0]
Family Income in 1979			1,803.4**	0.8**	5,239.50
			[510.6]	[0.3]	[4,417.5]
Family assets in 1985			0.0**	0.0+	0.5**
			[0.0]	[0.0]	[0.1]
Constant	36,609	32,766	-180,282	-59	-1,311,363
Other Demographic Controls	N	N	Y	Y	Y
Occupational Controls	N	N	Y	Y	Y
Industry Controls	N	N	Y	Y	Y
Observations	6,598	6,441	4,873	5,045	4,567
R-squared	0.008	0.095	0.298	0.17	0.172

Table displays OLS estimates of the relationship between various measures of earnings and wealth and being self-employed, after controlling for other factors. Earnings/Wealth and Employment status are from the NLSY 1994 sample. Robust standard errors are shown in brackets. We use **, *, and + to denote $p<0.01$, $p<0.05$ and $p<0.1$, respectively.

Table 12: Ability Test Scores, Wage-employed v/s Self-employed

Ability-Test Scores	NLSY 1994 Round			NLSY 2008 Round		
	Wage employed	Self-employed	Difference	Wage employed	Self-employed	Difference
AFQT	43.2	47.8	4.6	42.7	45.4	2.7
PSAT MATH	44.1	46.5	2.4	44.1	45.7	1.6
PSAT VERBAL	40	41.5	1.5	39.9	40.9	1
ACT MATH	16.6	18.9	2.3	16.9	18.3	1.4
ACT VERBAL	17.1	18	0.9	17.2	17.8	0.6
SAT MATH	443.5	444.9	1.4	438.7	452.5	13.8
SAT VERBAL	403.6	412.5	8.9	401.5	410.3	8.8

Table reports means for the various ability scores available in the NLSY surveys. Armed Forces Qualification Test (AFQT) scores, expressed in percentiles, are compiled from the arithmetic reasoning, word knowledge, paragraph comprehension, and numerical operations sections of the Armed Services Vocational Aptitude Battery (ASVAB). The respondents took the ASVAB battery in 1980. The Preliminary Scholastic Aptitude Test (PSAT) is a test that is intended to provide firsthand practice for the Scholastic Aptitude Test (PSAT). The Scholastic Aptitude Test (SAT) is intended to assess a student s readiness for college in the United States. The American College Testing scores (ACT) is based on a standardized test for high school achievement and college admissions in the United States produced by ACT, Inc.

Table 13: Probit estimates of the relationship between the Probability of Self-employment and Ability/Education in 1994 and 2008 (for "Stayers" and "Switchers")

D.V. = Self-employed (0/1)	1994 (Full Sample)	2008 (Full Sample)	2008 (Stayers)	2008 (Switchers)
Ability (AFQT) Score	0.004**	0.002	0.005*	0.001
	[0.001]	[0.001]	[0.002]	[0.001]
Highest Degree	-0.026+	-0.025	-0.094**	0.002
	[0.013]	[0.021]	[0.034]	[0.024]
Male	0.053	0.158**	0.276**	0.145*
	[0.056]	[0.057]	[0.102]	[0.066]
White	0.226**	0.138*	0.162	0.103
	[0.070]	[0.062]	[0.114]	[0.072]
Log Age	1.843**	-0.238	0.311	-0.791
	[0.564]	[0.520]	[0.921]	[0.623]
Risk Loving	0.018**	0.037**	0.059**	0.031**
	[0.007]	[0.007]	[0.016]	[0.009]
Sociable	0.068+	0.045	0.015	0.029
	[0.035]	[0.033]	[0.055]	[0.040]
Trusting	-0.013	0.01	-0.013	0.004
	[0.011]	[0.023]	[0.043]	[0.027]
Self- Esteem	-0.007+	-0.003	-0.010+	0
	[0.004]	[0.004]	[0.006]	[0.005]
Self-Mastery	-0.004	0.001	-0.006	0.002
	[0.005]	[0.005]	[0.009]	[0.006]
Father Self-employed?	0.260**	0.084	0.210+	0.095
	[0.080]	[0.083]	[0.124]	[0.097]
Mother Self-employed?	-0.267	-0.113	-0.213	-0.087
	[0.177]	[0.161]	[0.274]	[0.184]
Mother s education	0.015+	0.011	0.033*	0.004
	[0.008]	[0.007]	[0.014]	[0.008]
Father s education	0.004	0.005	0.008	0.003
	[0.006]	[0.005]	[0.010]	[0.006]
Family Income in 1979	0.043	-0.021	0.042	-0.046
	[0.033]	[0.032]	[0.055]	[0.037]
Family assets in 1985	0.001**	0.001**	0.001**	0.001+
	[0.000]	[0.000]	[0.000]	[0.000]
Other Demographic Controls	Y	Y	Y	Y
Occupational Controls	Y	Y	Y	Y
Constant	-8.873	-0.492	-3.418	2.065
Log-likelihood	-1518.27	-1831.59	-557.55	-1286.42
Observations	5,165	4,581	3,014	3,712

Table displays Probit estimates of the relationship between the probability of being self-employed and explanatory variables for all NLSY respondents in the 1994 round (Column 1), the 2008 round (Column 2), for those respondents in the 2008 round that reported being in the same occupation (i.e., self-employed or wage-employed) in that year as they did in 1994 (Column 3, labelled "Stayers"), and those that reported changing from self-employment to wage-employment, or vice versa, between 1994 and 2008. All variables are as described under Table 10 Notes. Robust standard errors are shown in brackets. We use **, *, and + to denote $p<0.01$, $p<0.05$ and $p<0.1$, respectively.

Table 14: Returns to education and ability in 1994 and 2008

DV= Log Income	1994 (Full Sample)	2008 (Full Sample)	1994 (Persistent Wage-employed)	2008 (Persistent Wage-employed)
Highest Degree	0.850**	0.860**	0.826**	0.726**
	[0.132]	[0.156]	[0.148]	[0.156]
Ability (AFQT) Score	0.001	0.002*	0.001	0.003**
	[0.001]	[0.001]	[0.001]	[0.001]
College Tier	-0.009	-0.075+	-0.075*	-0.078*
	[0.034]	[0.042]	[0.038]	[0.039]
Male	-0.398**	0.479**	-0.357**	0.347**
	[0.032]	[0.042]	[0.036]	[0.045]
White	0.027	-0.051	0.02	-0.086+
	[0.040]	[0.047]	[0.045]	[0.047]
Family Size in 1990	-0.056**	-0.049**	-0.059**	-0.047**
	[0.013]	[0.014]	[0.017]	[0.015]
Log Age	1.089**	1.086**	1.404**	1.384**
	[0.327]	[0.394]	[0.383]	[0.416]
Risk Loving	-0.002	0.004	-0.008	0.004
	[0.004]	[0.006]	[0.005]	[0.007]
Sociable	0.003	0.02	0.015	0.001
	[0.021]	[0.023]	[0.023]	[0.024]
Trusting	-0.007	-0.003	0.004	-0.018
	[0.007]	[0.016]	[0.017]	[0.017]
Father Self-employed?	0.106*	-0.026	0.039	-0.075
	[0.043]	[0.063]	[0.053]	[0.066]
Mother s education	-0.001	-0.005	0.002	-0.003
	[0.004]	[0.006]	[0.005]	[0.005]
Family assets in 1985	0	0	0	0
	[0.000]	[0.000]	[0.000]	[0.000]
Constant	4.243	4.261	3.184	2.105
Observations	2,831	2,413	1,767	1,701
Adjusted R-squared	0.281	0.288	0.316	0.301

Table displays OLS estimates of the relationship between (logged) income, and education/ability, calculated in 1994, when respondents were between 30-37 and again in 2008 when the respondents had turned 44-51. The first two columns report the estimates for the full sample of respondents and the last two columns report the estimates for the sub-sample of respondents who remain wage-employed between 1994 and 2008. Robust standard errors are shown in brackets. We use **, *, and + to denote $p<0.01$, $p<0.05$ and $p<0.1$, respectively.

Table 15: Educational Attainment, Wage-employed v/s Self-employed (U.K. NCDS7 Sample)

Highest educational attainment	Wage-employed	Self-employed	Total
No academic qualification	15.4%	17.4%	15.7%
CSEs2-5, other Scottish Quals	14.8%	15.7%	14.9%
GCSE A-C, good O levels Scot Standards	35.0%	36.2%	35.2%
AS Levels or 1 A Level	0.6%	0.1%	0.5%
2+ A Levels, Scot Higher/6th	8.4%	7.8%	8.3%
Diploma	4.8%	3.7%	4.6%
Degree, PGCE, Other Degree Level	17.1%	16.1%	17.0%
Higher Degree	4.0%	3.1%	3.9%
Total number of individuals	5523	1039	6562

Table 16: Ability Scores, Wage-employed v/s Self-employed (U.K. NCDS7 Sample)

	Outstanding Ability at 10 (0/1)	A-level score at 17	A-level score at 17 conditional on taking exam
Employee	0.23	1.16	6.21
Entrepreneur	0.30	1.54	7.24

Table 17: Ability Scores by Educational Attainment, Wage-employed v/s Self-employed (U.K. NCDS7 Sample)

Highest educational attainment	Outstanding Ability at 10 (0/1)			A-level score at 17 conditional on taking exam		
	Wage-employed	Self-employed	Difference	Wage-employed	Self-employed	Difference
No academic qualification	0.13	0.23	0.09	7.54	6.00	-1.54
CSEs2-5, other Scottish Quals	0.16	0.23	0.07			
GCSE A-C, O levels Scot Standards	0.22	0.25	0.03	2.88	3.50	0.63
AS Levels or 1 A Level	0.25	0.00	-0.25	3.00		
2+ A Levels, Scot Higher/6th	0.29	0.35	0.06	4.48	3.74	-0.74
Diploma	0.33	0.30	-0.03	3.73	5.75	2.02
Degree, PGCE, Other Degree Level	0.33	0.47	0.14	6.94	8.64	1.70
Higher Degree	0.38	0.50	0.12	7.80	8.90	1.10

Table 18: Educational Attainment by Ability, Wage-employed v/s Self-employed (U.K. NCDS7 Sample)

Highest educational attainment	Outstanding ability = 1		Outstanding ability = 0	
	Employee	Entrepreneur	Employee	Entrepreneur
No academic qualification	8.6%	13.3%	17.4%	19.0%
CSEs2-5, other Scottish Quals	10.0%	12.5%	16.2%	17.5%
GCSE A-C, good O levels Scot Standards	33.2%	31.3%	36.0%	38.8%
AS Levels or 1 A Level	0.6%	0.0%	0.6%	0.2%
2+ A Levels, Scot Higher/6th	10.7%	9.0%	7.9%	7.0%
Diploma	6.7%	3.5%	4.2%	3.4%
Degree, PGCE, Other Degree Level	24.1%	25.4%	14.7%	12.0%
Higher Degree	6.0%	5.1%	3.0%	2.1%
Total	100.0%	100.0%	100.0%	100.0%

Empirical Appendix

Figure A1: Net Income (from NLSY 1994 round) and Ability (AFQT) Score

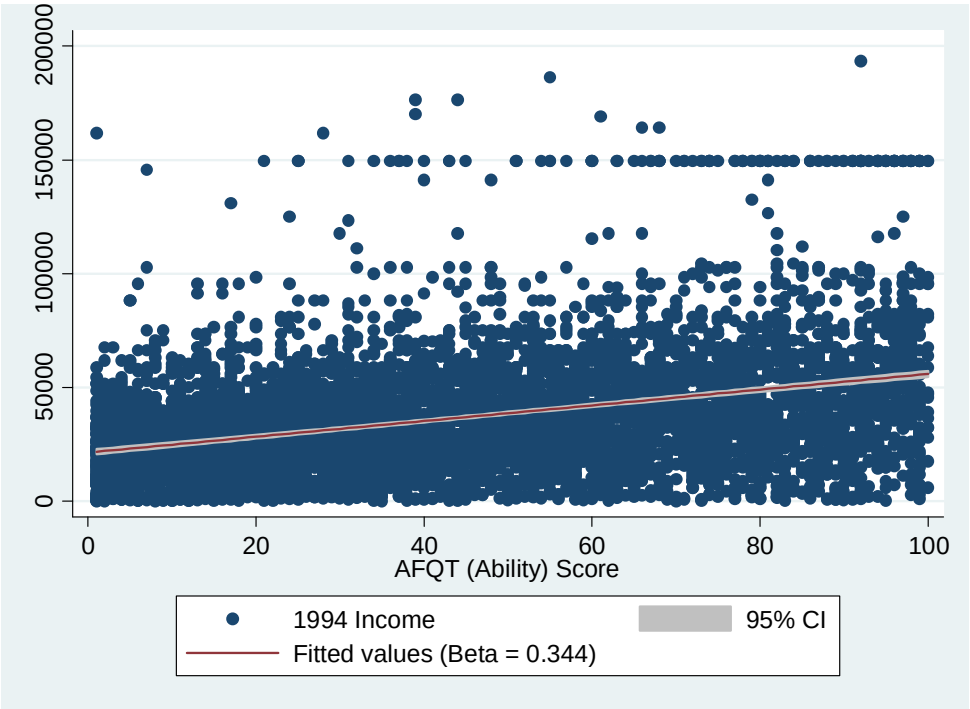


Figure A2: Pay Rate (from NLSY 1994 round) and Ability (AFQT) Score

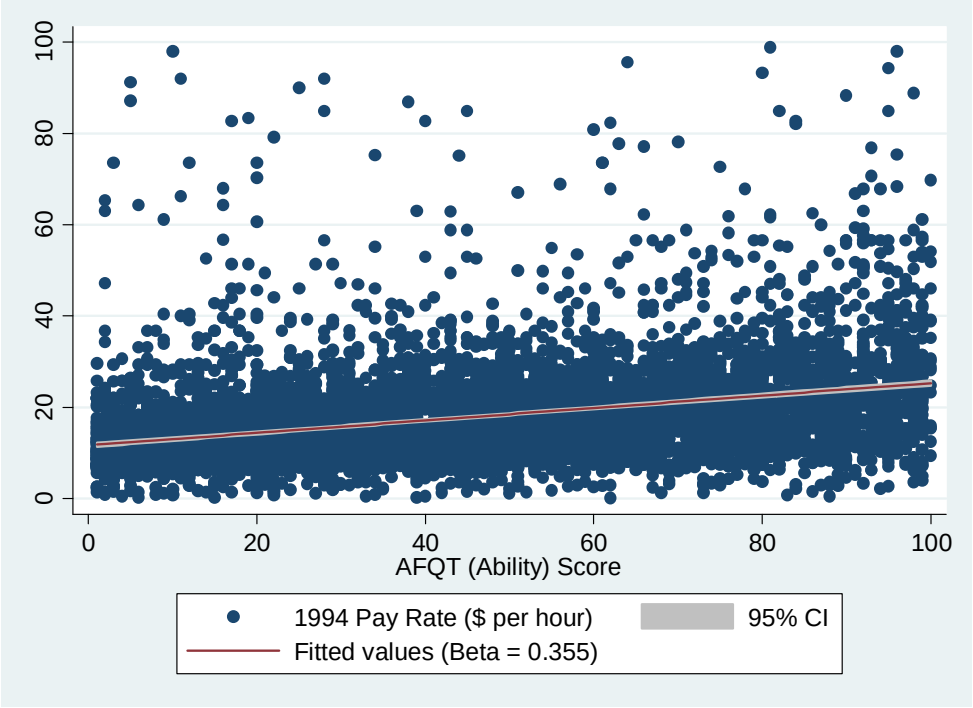


Figure A3: Net Worth (from NLSY 1994 round) and Ability (AFQT) Score

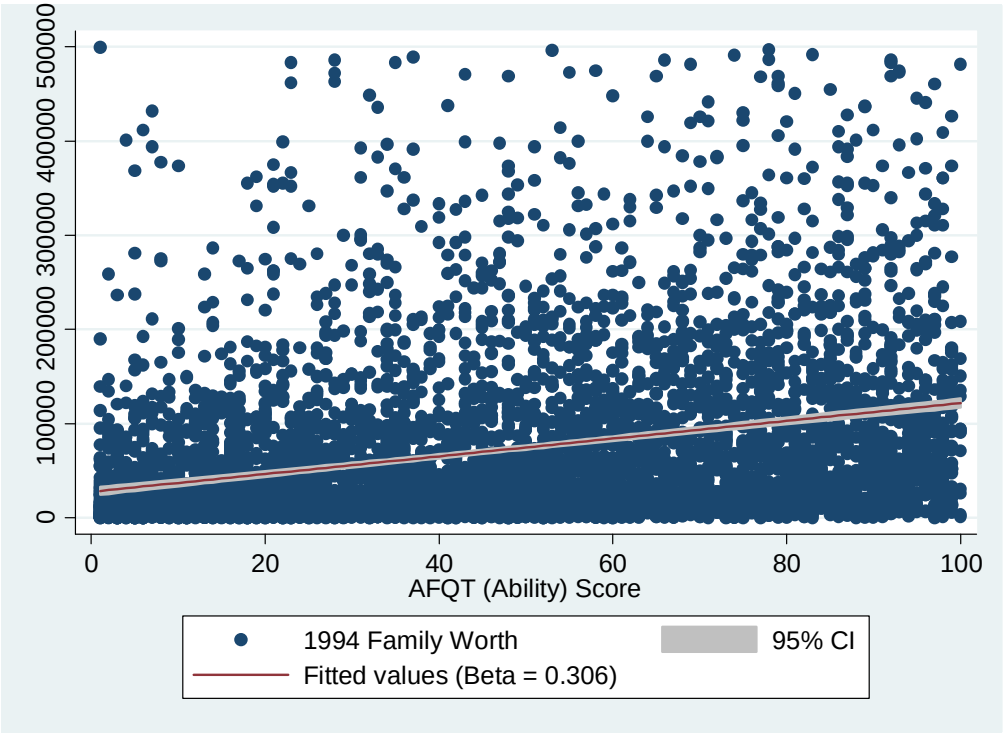


Figure A4: Distribution of ability scores, Self-employed v/s Wage-employed, NLSY 2008 round

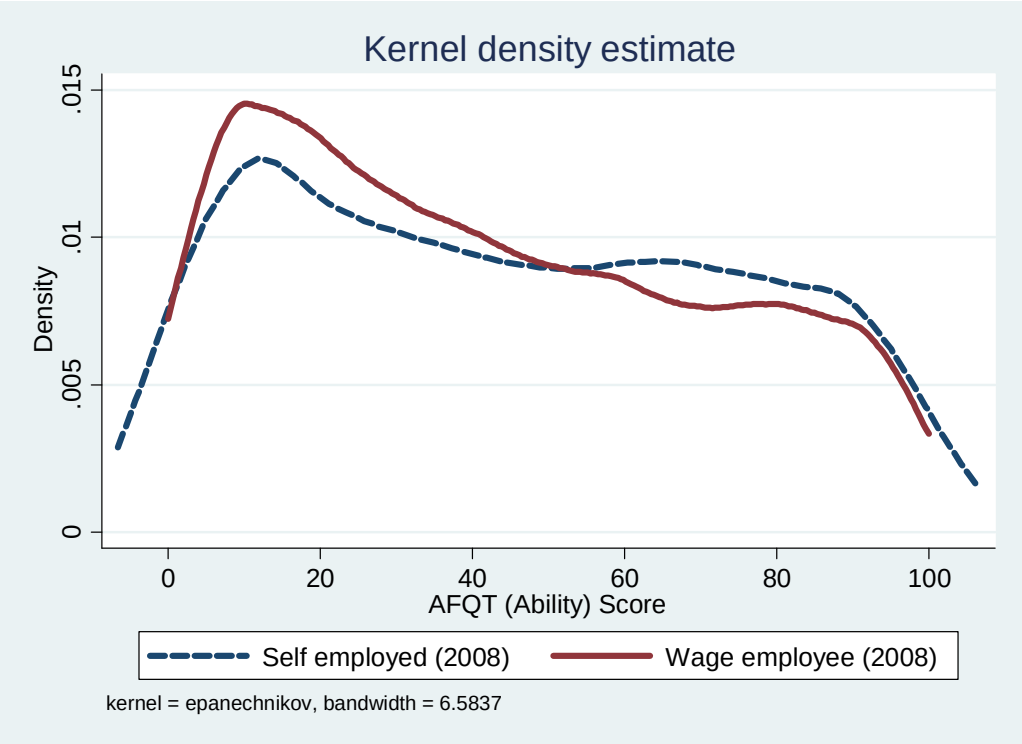


Figure A5: Distribution of income and wealth, Self-employed v/s Wage-employed, NLSY 2008 Round

Figure A5a: Distribution of Log Annual Income

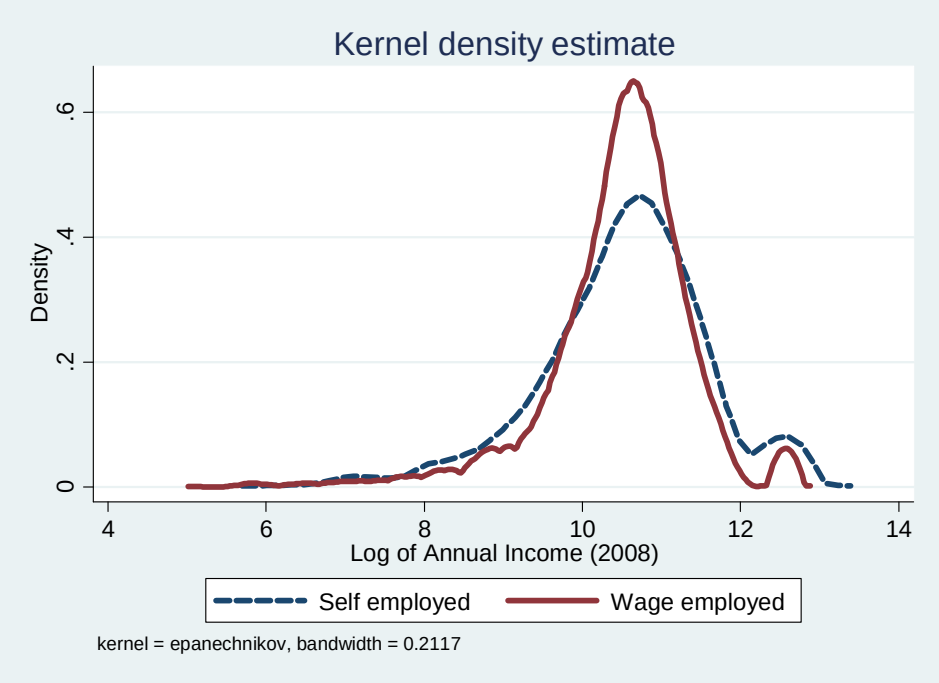


Figure A5b: Distribution of Log Pay Rate

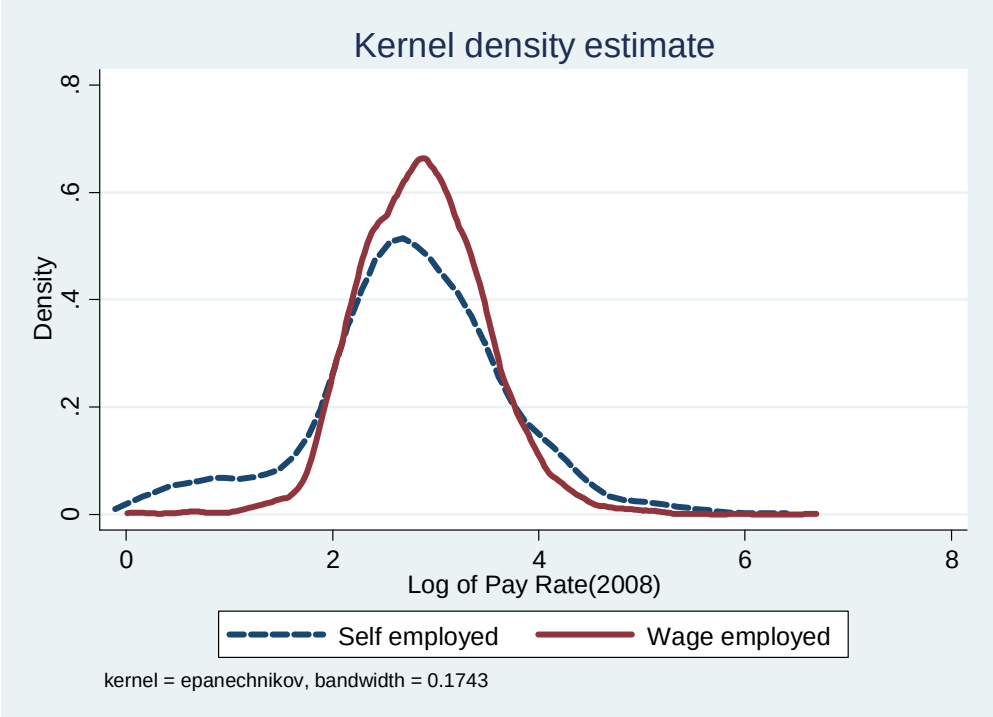
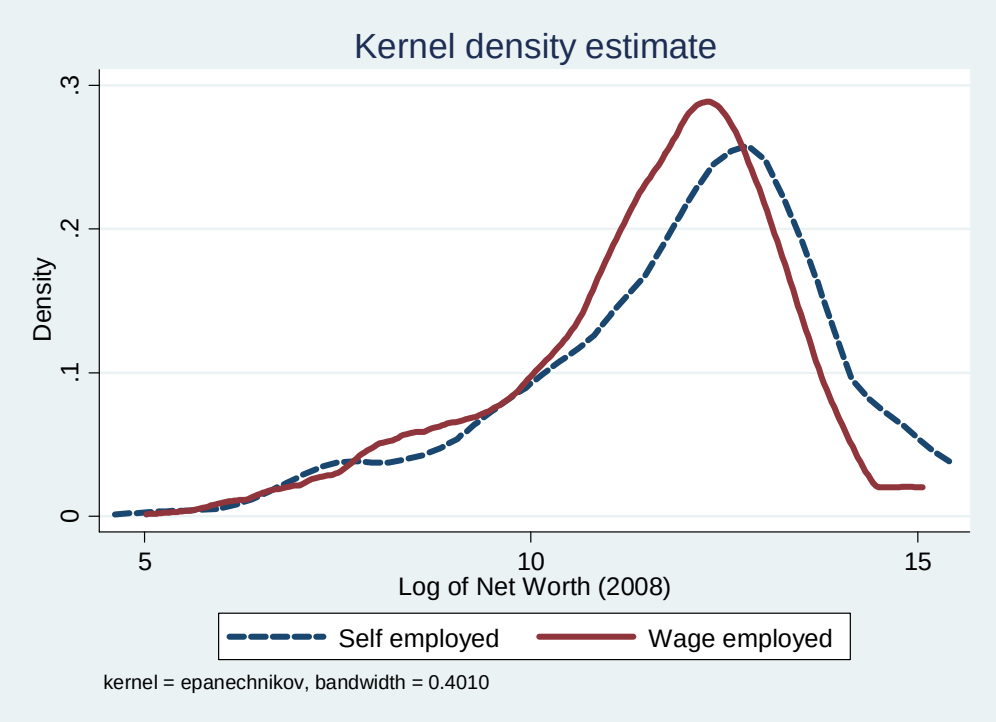


Figure A5c: Distribution of Net Worth



Figures show the kernel density distribution of logged annual income (A5a), logged pay rate expressed in $/hr (A5b), and net worth (A5c) for self-employed individuals and wage employees in the NLSY 2008 sample. 12.7 percent of the 6,355 full-time employed individuals in the 2008 NLSY sample were self-employed, and the rest were wage-employed. Following Table presents the sample descriptives for the three variables without log transformation.

	Net Annual Income ($)			Hourly pay rate ($/hr)			Net Worth ($)		
	Mean	Median	SD	Mean	Median	SD	Mean	Median	SD
Wage-employed	48,794	39,107	44,696	21.6	17.5	20.1	270,600	122,210	490,030
Self-employed	59,485	40,574	65,881	23.5	15.5	31.9	457,733	182,810	786,306

Table A1: Earnings and Wealth by Educational Qualification, (NLSY 1994 and NLSY 2008)

PANEL A

Educational Attainment	NLSY 1994 Round			NLSY 2008 Round		
	Annual income	Net worth	Payrate ($/hr)	Annual income	Net worth	Payrate ($/hr)
3RD GRADE	20,580	2,940	8.2	12,832	6,060	10.6
4TH GRADE	19,845	26,460	16.6	29,330	96,960	15.1
5TH GRADE	7,350	64,757	6.8	36,663	25,250	7.7
6TH GRADE	19,110	27,930	9.3	14,665	70,448	9.9
7TH GRADE	19,845	2,793	9.2	26,275	17,423	9.2
8TH GRADE	23,520	8,894	11.0	28,108	20,200	12.1
9TH GRADE	20,580	5,439	11.0	24,442	19,695	11.6
10TH GRADE	22,050	7,975	11.5	20,776	20,907	10.6
11TH GRADE	22,050	8,820	11.6	27,008	24,240	11.4
12TH GRADE	27,930	24,108	13.2	34,219	84,109	14.9
1ST YEAR COL.	31,321	33,884	14.8	36,663	106,050	16.6
2ND YEAR COL.	33,810	47,334	16.4	39,107	140,138	18.9
3RD YEAR COL.	33,590	40,976	16.5	40,329	148,975	18.2
4TH YEAR COL.	47,040	83,790	22.3	57,744	295,425	25.9
5TH YEAR COL.	44,100	69,090	21.2	53,467	242,721	25.8
6TH YEAR COL.	49,980	92,538	23.3	57,439	351,480	28.0
7TH YEAR COL.	58,800	113,337	27.6	56,217	408,646	28.1
8TH YEAR COL.	59,535	102,557	28.8	76,992	415,110	35.0
Observations	7,247			6,619		

PANEL B

Highest Degree	NLSY 1994 Round			NLSY 2008 Round		
	Annual income	Net worth	Payrate ($/hr)	Annual income	Net worth	Payrate ($/hr)
High school	29,400	27,195	13.8	34,219	79,790	14.6
Associate/Junior College	33,810	53,332	17.4	42,774	142,915	18.7
Bachelor of Arts	42,042	66,885	19.8	50,190	260,580	22.8
Bachelor of Science	51,450	88,200	22.8	61,105	294,920	27.3
Master s (MA, MBA, MS)	52,920	107,310	25.0	57,439	373,953	28.7
PhD, MD, LLD, DDS	74,970	162,435	35.5	101,434	550,880	42.4
Observations	5,961			6,446		

PANEL C

Selectivity Tier of Latest Degree	NLSY 1994 Round			NLSY 2008 Round		
	Annual income	Net worth	Payrate ($/hr)	Annual income	Net worth	Payrate ($/hr)
1 (Top 50)	62,475	97,755	29.5	84,325	626,480	35.0
2 (Top 50-200)	49,980	92,610	22.8	61,961	399,960	30.1
3 (>200)	36,750	48,510	17.1	43,996	181,952	20.2
4 (No info)	32,340	30,135	15.3	42,774	127,765	19.4
Observations	3,991			3,856		

Panel A of the Table shows the median earnings and assets of respondents (all figures in 2010 $) by their number of years in school/college. Panel B of the Table shows the median earnings and assets of respondents by their highest degree. Panel C of the Table shows the median earnings and assets of respondents (all figures in 2010 $) by the ranking of the institution of their latest degree, and their mean and median AFQT scores. Rankings are based on admissions selectivity data of the institutions available from the National Center for Education Statistics.

Table A2: Ability (AFQT) Scores by Educational Qualification, (NLSY 1994 and NLSY 2008)

PANEL A

Educational Attainment	NLSY 1994 Sample			NLSY 2008 Sample		
	Percent	Mean AFQT	Median AFQT	Percent	Mean AFQT	Median AFQT
3RD GRADE	0.1%	1.2	1	0.1%	1.9	1
4TH GRADE	0.1%	2.0	1	0.1%	2.8	3
5TH GRADE	0.0%	1.0	1	0.1%	0.7	1
6TH GRADE	0.4%	4.3	2.5	0.3%	3.7	1
7TH GRADE	0.4%	11.2	9	0.3%	9.2	6.5
8TH GRADE	1.6%	9.6	7	1.1%	8.5	5
9TH GRADE	2.4%	13.8	9	1.9%	12.1	8
10TH GRADE	2.9%	18.1	13	1.8%	15.2	10
11TH GRADE	3.5%	17.9	13	2.4%	16.8	12
12TH GRADE	43.4%	34.8	30	42.2%	32.4	27
1ST YEAR COL.	9.3%	43.0	40	9.1%	41.1	38
2ND YEAR COL.	9.9%	50.7	49	10.9%	47.0	43
3RD YEAR COL.	5.0%	51.1	54	5.6%	48.0	48
4TH YEAR COL.	12.7%	67.1	73	12.7%	64.8	69
5TH YEAR COL.	3.0%	68.4	73.5	3.2%	66.1	70
6TH YEAR COL.	2.9%	75.1	82	4.2%	70.7	78
7TH YEAR COL.	1.3%	80.6	86.5	1.8%	75.4	85
8TH YEAR COL.	1.2%	81.2	89.5	2.4%	75.8	82
Observations	7,247			6,619		

PANEL B

Highest Degree	NLSY 1994 Sample			NLSY 2008 Sample		
	Percent	Mean AFQT	Median AFQT	Percent	Mean AFQT	Median AFQT
High school	67.2%	38.0	34	64.8%	33.0	28
Associate/Junior College	9.4%	50.1	48.5	11.2%	47.3	45
Bachelor of Arts	6.9%	68.2	73	5.1%	65.2	70
Bachelor of Science	11.1%	69.6	74	11.4%	67.0	72
Master s (MA, MBA, MS)	4.0%	76.8	83.5	6.2%	72.1	78
PhD, MD, LLD, DDS	1.3%	87.2	92	1.4%	83.7	90
Observations	5,961			6,446		

PANEL C

Selectivity Tier of Latest Degree	NLSY 1994 Sample			NLSY 2008 Sample		
	Percent	Mean AFQT	Median AFQT	Percent	Mean AFQT	Median AFQT
1 (Top 50)	2.5%	78.8	89	2.4%	78.5	89
2 (Top 50-200)	6.7%	75.1	82	6.1%	73.6	81
3 (>200)	79.2%	54.3	55	79.1%	53.7	54
4 (No info)	11.6%	48.5	46	12.4%	47.7	45
Observations	3,991			3,856		

Panel A of the Table shows the percentage of respondents by their number of years in school/college and mean and median AFQT scores. Panel B of the Table shows the percentage of respondents by their highest degrees, and mean and median AFQT scores. Panel C of the Table shows the percentage of respondents by the ranking of the institution of their latest degree, and their mean and median AFQT scores. Rankings are constructed based on admissions selectivity data of the institutions available from the National Center for Education Statistics. AFQT Scores are expressed in percentiles, and respondents' highest educational attainment is as reported in the 1994 and 2008 rounds of the NLSY.

Table A3: Ability Scores by educational qualification for self-employed and wage-employed (NLSY 2008 Round)

PANEL A

Education Level	Percent of 2008 Sample		Mean AFQT (Ability Score)			
	Wage employees	Self-employed	Wage employees	Self-employed	Difference	S.E. (ttest)
3RD GRADE	0.1%	0.2%	3.7	2	-1.7	
4TH GRADE	0.1%	0.3%	1	5	4	
5TH GRADE	0.0%	0.1%	1	1	0	
6TH GRADE	0.2%	0.3%	5.6	6	0.4	
7TH GRADE	0.3%	0.2%	10.6	15	4.4	7.5
8TH GRADE	1.0%	1.2%	8.3	10.9	2.6	2.6
9TH GRADE	1.8%	2.1%	11.2	17.1	5.9	3+
10TH GRADE	1.8%	1.7%	16.1	12.6	-3.5	3.8
11TH GRADE	2.4%	2.2%	17.1	16.5	-0.6	3.3
12TH GRADE	42.1%	42.5%	32.3	34.1	1.8	1.2
1ST YEAR COL.	9.1%	9.3%	40.8	43.1	2.3	2.7
2ND YEAR COL.	11.2%	9.6%	46.3	50.8	4.5	2.6+
3RD YEAR COL.	5.8%	5.0%	47.5	50.3	2.8	3.5
4TH YEAR COL.	12.5%	14.1%	63.8	69	5.2	2.2*
5TH YEAR COL.	3.3%	2.7%	66	66.7	0.7	5
6TH YEAR COL.	4.3%	4.1%	70.9	69.8	-1.1	3.8
7TH YEAR COL.	1.7%	2.0%	74.4	79.7	5.3	5.7
8TH YEAR COL.	2.3%	2.6%	74	83.3	9.3	4.4*
Observations	5,455	1,122				

PANEL B

Highest Degree	Percent of 2008 Sample		Mean AFQT (Ability Score)			
	Wage employees	Self-employed	Wage employees	Self-employed	Difference	S.E. (ttest)
High school diploma	64.6%	64.4%	32.4	34.7	2.3	1.0*
Associate/Junior College	11.6%	9.9%	46.6	49.3	2.7	2.6
Bachelor of Arts	4.9%	6.5%	62.7	69.1	6.4	3.5+
Bachelor of Science	11.3%	12.2%	65	67	2	2.3
Master s Degree	6.5%	5.0%	71.4	71	-0.4	3.4
PhD, MD, LLD, DDS	1.2%	2.2%	82	85.4	3.4	4.4
Observations	5,318	1,086				

PANEL C

College Tier	Percent of 2008 Sample		Mean AFQT (Ability Score)			
	Wage employees	Self-employed	Wage employees	Self-employed	Difference	S.E. (ttest)
1 (Top 50)	1.2%	1.2%	77.4	70	-7.4	11.2
2 (Top 50-200)	3.1%	5.1%	60	69.1	9.1	5.6
3 (>200)	83.3%	81.3%	53.7	57.9	4.2	1.3**
4 (No info)	12.3%	12.4%	47.1	47.5	0.4	3.4
Observations	3,203	651				

Table compares the mean values of AFQT (ability) test scores for self-employed and wage-employed individuals, by the individuals' highest educational attainment. PANEL A lists attainment by educational level, PANEL B by highest degree obtained, and PANEL C by the institutional rank of the last degree obtained. Employment status represents respondent employment in 2007 (data collected in 2008). Data on highest educational attainment reflect each respondent's highest attainment as of 2008, when they were aged between 44 and 52. For t-tests, ** $p<0.01$, * $p<0.05$, + $p<0.1$

Table A4: Educational attainment by ability decile for self-employed and wage-employed (NLSY 2008 Round)

AFQT Decile	Education Level				Highest Degree				College Selectivity Tier			
	Wage employed	Self employed	Difference	S.E. (ttest)	Wage employed	Self employed	Difference	S.E. (ttest)	Wage employed	Self employed	Difference	S.E. (ttest)
1	11.06	11.22	0.16	[0.26]	1.03	1	-0.03	[0.03]	2.5			
2	12.06	11.78	-0.28	[0.16]+	1.19	1.18	-0.01	[0.06]	2.63	2.5	-0.13	[0.36]
3	12.71	12.55	-0.16	[0.18]	1.34	1.4	0.06	[0.09]	2.78	2.6	-0.18	[0.17]
4	13.12	12.86	-0.26	[0.19]	1.55	1.38	-0.17	[0.11]	2.81	2.86	0.05	[0.18]
5	13.47	13.57	0.1	[0.21]	1.68	1.88	0.2	[0.13]	2.83	2.8	-0.03	[0.12]
6	13.74	13.32	-0.42	[0.24]+	1.9	1.78	-0.12	[0.15]	2.85	2.77	-0.08	[0.11]
7	14.02	13.65	-0.37	[0.25]	2.02	1.68	-0.34	[0.16]*	2.8	2.91	0.11	[0.14]
8	14.62	14.48	-0.14	[0.25]	2.42	2.38	-0.04	[0.17]	2.85	2.61	-0.24	[0.11]*
9	15.01	14.93	-0.08	[0.28]	2.69	2.61	-0.08	[0.20]	2.75	2.71	-0.04	[0.13]
10	16.32	16.38	0.06	[0.21]	3.64	3.64	0	[0.16]	2.63	2.65	0.02	[0.10]

Table shows means, differences in means, and results of ttests for differences in the mean educational attainment of wage-employed and self-employed individuals by decile of AFQT (ability) scores. "Education Level" is measured on an ordinal scale such that 1 = completion of 1st grade, 2 = completion of 2nd grade, ..12= completion of 12th grade (high-school),..16=completion of four-year college, and so on. "Highest Degree" is measured on a scale indicating 1=High School Diploma (or equivalent), 2=Associate/Junior College, 3=Bachelor of Arts, 4=Bachelor of Science, 5=Master's, and 6 = PhD or other advanced professional degrees (MD, LLD, DDS). College Selectivity Tier is measured such that 1 = Top 50, 2 = Top 50-200 and 3 >200 (We omit the observations for which we do not have information on the selectivity rank of the respondents' educational institutions). Data on highest educational attainment reflect respondents' highest attainment as of 2008, when they were aged between 44 and 52. For t-tests, ** $p<0.01$, * $p<0.05$, + $p<0.1$

Table A5: Income and Wealth for self-employed and wage-employed by education level (NLSY 2008 Round)

PANEL A: Annual Income by education level

Education Level	Mean Annual Income ($)				SD Annual Income ($)	
	Wage employed	Self-employed	Difference	S.E. (ttest)	Wage employed	Self-employed
6TH GRADE	22,075	17,415	-4,660	[12,062]	22,583	11,524
7TH GRADE	24,398	5,499	-18,898	[15,725]	15,056	
8TH GRADE	30,362	41,251	10,889	[8,556]	21,906	36,322
9TH GRADE	27,985	38,940	10,955	[6,303]+	17,535	40,647
10TH GRADE	26,851	28,006	1,155	[6,570]	21,153	21,430
11TH GRADE	30,568	32,585	2,017	[5,111]	18,696	25,529
12TH GRADE	38,471	46,111	7,640	[1,836]**	29,753	44,790
1ST YEAR COL.	43,917	55,377	11,460	[4,900]*	32,139	65,759
2ND YEAR COL.	44,685	45,815	1,130	[3,997]	31,505	42,801
3RD YEAR COL.	44,854	67,996	23,141	[6,625]**	31,505	72,630
4TH YEAR COL.	72,458	80,081	7,623	[6,331]	59,712	82,211
5TH YEAR COL.	63,552	103,209	39,657	[11,827]**	49,156	85,310
6TH YEAR COL.	79,722	95,362	15,640	[13,608]	71,811	99,660
7TH YEAR COL.	90,542	94,906	4,363	[22,000]	82,814	93,074
8TH YEAR COL.	98,220	144,067	45,847	[18,402]*	79,325	104,081

Table A5, PANEL B: Pay Rate ($/hr) by education level

Education Level	Mean Pay Rate ($/hr)				SD Pay Rate ($/hr)	
	Wage employed	Self-employed	Difference	S.E. (ttest)	Wage employed	Self-employed
6TH GRADE	10.8	7.7	-3.1	[2.2]	3.3	6.0
7TH GRADE	9.1	11.1	2.0	[3.4]	4.2	10.2
8TH GRADE	13.2	15.7	2.6	[2.2]	7.0	6.2
9TH GRADE	12.8	30.6	17.7	[6.6]**	6.1	65.6
10TH GRADE	12.5	10.4	-2.1	[1.6]	6.1	6.3
11TH GRADE	13.0	10.4	-2.6	[1.5]+	6.1	8.8
12TH GRADE	16.8	19.2	2.3	[0.6]**	9.0	21.2
1ST YEAR COL.	19.2	18.9	-0.3	[1.4]	10.2	20.6
2ND YEAR COL.	21.5	22.7	1.2	[2.4]	12.2	51.5
3RD YEAR COL.	20.7	29.5	8.8	[2.8]**	11.5	39.5
4TH YEAR COL.	33.3	30.5	-2.8	[3.5]	38.4	30.4
5TH YEAR COL.	31.8	28.5	-3.3	[6.1]	32.4	19.9
6TH YEAR COL.	34.9	36.8	1.9	[4.9]	24.6	43.9
7TH YEAR COL.	37.2	30.8	-6.4	[7.1]	29.7	19.0
8TH YEAR COL.	41.7	56.6	14.9	[7.9]+	29.1	57.1

Table A5, PANEL C: Net Worth by education level

Education Level	Mean Net Worth ($)				SD Net Worth ($)	
	Wage employed	Self-employed	Difference	S.E. (ttest)	Wage employed	Self-employed
6TH GRADE	147,622	751,680	604,058	[471,256]	177,447	1,528,084
7TH GRADE	46,965	29,694	-17,271	[35,748]	58,240	23,637
8TH GRADE	80,671	173,472	92,800	[50,560]+	83,608	276,151
9TH GRADE	60,913	179,935	119,022	[35,976]**	78,064	272,113
10TH GRADE	87,419	78,266	-9,153	[42,957]	167,100	80,175
11TH GRADE	99,075	166,585	67,510	[47,596]	159,387	240,081
12TH GRADE	185,454	361,790	176,336	[22,217]**	311,972	694,591
1ST YEAR COL.	185,208	461,330	276,122	[44,716]**	269,771	705,184
2ND YEAR COL.	258,619	460,845	202,226	[55,080]**	423,364	739,569
3RD YEAR COL.	246,068	705,584	459,516	[79,649]**	345,541	1,036,627
4TH YEAR COL.	517,734	710,604	192,871	[69,827]**	713,525	954,227
5TH YEAR COL.	431,021	416,659	-14,363	[111,653]	580,406	361,997
6TH YEAR COL.	577,613	783,371	205,759	[144,096]	810,468	968,745
7TH YEAR COL.	792,855	745,403	-47,452	[249,922]	982,859	849,076
8TH YEAR COL.	700,451	1,107,318	406,867	[204,323]*	866,750	1,262,052

Table shows mean and standard deviation of annual income, hourly wage rate, and net worth (assets - liabilities) for wage employees and self-employed individuals in our NLS 2008 sample. Income and wealth variables are in 2010 US dollars, and wages in dollars per hour. Income and wage data reflect respondent earnings associated with their employment in 2007 (data collected in 2008). The Tables are based on the responses of 6,351 full-time employed individuals who responded to the 2008 round of the NLSY and had non-missing values for education level, employment and wage/wealth data. For t-tests, ** $p<0.01$, * $p<0.05$, + $p<0.1$

Table A6: Industry and Occupation of Self-Employed and Wage-Employed (NLSY 2008 Round)

PANEL A

Occupation	Wage employees	Self-employed
Executive, Admin & Managerial	10.7%	15.0%
Management Related Occupations	4.0%	4.7%
Mathematical & Computer Scientist	2.4%	0.8%
Engineers, Architects, Surveyors	1.5%	1.4%
Physical Scientists	0.2%	0.4%
Social Scientists & Related Worker	0.2%	0.4%
Life, Physical and Social Scientists	0.3%	0.2%
Counselors, Social & Religious Workers	1.9%	1.6%
Lawyers, Judges & Legal Support Workers	0.6%	0.9%
Teachers	4.7%	2.4%
Education, Training & Library Workers	1.6%	0.4%
Entertainers, Performers, Sports	0.6%	1.8%
Media & Communications Workers	0.4%	1.2%
Health Diagnosing & Treating Practice	2.5%	1.2%
Health Care Technical & Support Workers	4.7%	1.9%
Protective Service Occupations	2.8%	1.0%
Food Preparation & Serving Workers	4.5%	1.2%
Cleaning & Building Service Workers	3.7%	6.3%
Entertainment Attendants & Related	0.2%	0.6%
Personal Care & Service Workers	1.6%	7.5%
Sales & Related Workers	6.6%	11.2%
Office & Administrative Support	15.6%	6.4%
Farming, Fishing & Forestry Occupations	0.4%	1.0%
Construction Trade & Extraction	5.0%	11.2%
Installation, Maintenance & Repair	3.8%	4.0%
Production & Operating Workers	2.4%	0.6%
Food Preparation Occupations	0.3%	0.1%
Setters, Operators & Tenders	6.0%	2.7%
Transportation & Material Moving	7.0%	6.7%
Others	3.9%	5.7%
Non-missing observations	5,491	1,132

PANEL B

Industry	Wage employees	Self-employed
Agriculture, Forestry and Fishing	0.3%	2.2%
Mining	0.4%	0.8%
Utilities	1.5%	0.6%
Construction	5.9%	15.1%
Manufacturing	13.2%	5.7%
Wholesale Trade	3.0%	2.4%
Retail Trade	7.6%	6.8%
Transportation and Warehousing	5.9%	5.7%
Information	2.5%	1.9%
Finance and Insurance	4.5%	3.3%
Real Estate and Rental and Leasing	1.4%	3.4%
Professional and Technical Services	4.0%	6.5%
Management and Administration	4.3%	7.2%
Educational Services	23.7%	14.1%
Health Care and Social Assistance	1.1%	1.9%
Arts, Entertainment, and Recreation	4.5%	2.9%
Accommodation and Food Services	3.0%	11.0%
Other Services	7.5%	1.8%
Public Administration and Military	0.1%	0.0%
Others	5.7%	6.8%
Observations	5,272	1,085

Table A7: Probit estimates of the relationship between the Probability of Self-employment and Ability/Education

D.V. = Self-employed (0/1)	1	2	3	4	5
NLSY Sample	1994	1994	1998	2002	2008
Ability (AFQT) Score	0.005**	0.004*	0.002+	0.004**	0.002+
	[0.001]	[0.002]	[0.001]	[0.001]	[0.001]
Education Level	-0.042**				
	[0.015]				
Highest Degree		-0.017	-0.005	-0.033*	-0.011
		[0.023]	[0.014]	[0.013]	[0.022]
Selectivity Tier		0.029			
		[0.073]			
Demographic Controls	Y	Y	Y	Y	Y
Attitudinal Controls	Y	Y	Y	Y	Y
Family Background Controls	Y	Y	Y	Y	Y
Occupational Controls	Y	Y	Y	Y	Y
Constant	-7.65	-6.34	-7.78	-5.07	-1.89
Log-likelihood	-1331.44	-845.87	-1485.6	-1633.55	-1703.66
Observations	4,566	2,893	5,284	4,876	4,501

Table A8: Quantile regression estimates of the relationship between Earnings/Net worth and Self-employment

Dependent Variable	Log Annual Income				Log Net Worth			
Quantile	0.2	0.4	0.6	0.8	0.2	0.4	0.6	0.8
Self-employed (0/1)	-0.39**	-0.12**	0.02	0.18**	0.47**	0.43**	0.59**	0.61**
	[0.08]	[0.04]	[0.03]	[0.03]	[0.12]	[0.10]	[0.09]	[0.09]
Highest Degree	0.09**	0.06**	0.05**	0.05**	0.14**	0.13**	0.13**	0.07**
	[0.01]	[0.01]	[0.00]	[0.01]	[0.02]	[0.01]	[0.01]	[0.01]
Log Age	0.73	0.72**	1.16**	1.20**	3.85**	4.70**	3.89**	3.04**
	[0.48]	[0.25]	[0.21]	[0.22]	[0.74]	[0.62]	[0.57]	[0.56]
Risk Loving	0	0	0	0	0.02*	0.01	0.01+	0.01+
	[0.01]	[0.00]	[0.00]	[0.00]	[0.01]	[0.01]	[0.01]	[0.01]
Sociable	0.01	0.02	0.01	0.01	-0.03	0.01	0.02	0.02
	[0.03]	[0.02]	[0.01]	[0.01]	[0.05]	[0.04]	[0.04]	[0.04]
Trusting	-0.01	-0.01	-0.01*	-0.01*	-0.04**	-0.03*	-0.02*	-0.02+
	[0.01]	[0.00]	[0.00]	[0.00]	[0.01]	[0.01]	[0.01]	[0.01]
Self- Esteem	0.01**	0.01**	0.01**	0.00**	0.02**	0.01+	0.01+	0.01
	[0.00]	[0.00]	[0.00]	[0.00]	[0.01]	[0.00]	[0.00]	[0.00]
Pearlin Mastery	0.02**	0.01**	0.01**	0.01**	0.03**	0.02**	0.02**	0.01+
	[0.00]	[0.00]	[0.00]	[0.00]	[0.01]	[0.01]	[0.01]	[0.01]
Constant	5.14	5.99	5.08	5.53	-7.62	-11.52	-6.75	-2.64
Background Controls	Y	Y	Y	Y	Y	Y	Y	Y
Demographic Controls	Y	Y	Y	Y	Y	Y	Y	Y
Occupational Controls	Y	Y	Y	Y	Y	Y	Y	Y
Industry Controls	Y	Y	Y	Y	Y	Y	Y	Y
Observations	4,687	4,687	4,687	4,687	3,903	3,903	3,903	3,903
Pseudo R Square	0.190	0.190	0.194	0.199	0.209	0.219	0.197	0.169

Table displays quantile regression estimates of the relationship between measures of earnings and wealth, and being self-employed, after controlling for other factors. We divide the sample into five quantiles based on individuals' earnings/wealth. Earnings/Wealth and Employment status are from the NLSY 1994 sample. Robust standard errors are shown in brackets. We use **, *, and + to denote $p<0.01$, $p<0.05$ and $p<0.1$, respectively.

Table A9: 2SLS estimates of the relationship between Earnings/Net worth and Self-employment

Dependent Variable	1 Annual Income	2 Pay Rate	3 Net Worth	4 Log Annual Income	5 Log Pay Rate	6 Log Net Worth
Self-employed (0/1)	260,367.8*	139.5*	402,661.00	6.20*	6.49*	10.78+
	[109,207.7]	[61.9]	[346,529.9]	[2.85]	[2.94]	[6.33]
Highest Degree	1,908.6**	0.4	10,078.2**	0.07**	0.02	0.12**
	[594.0]	[0.3]	[1,832.8]	[0.02]	[0.02]	[0.03]
Male	13,010.1**	2.9*	441.5	0.41**	0.15*	0.02
	[2,644.2]	[1.3]	[8,714.2]	[0.07]	[0.06]	[0.12]
White	-4,641.80	-2.2	39,361.2**	-0.1	-0.11	0.33
	[3,876.4]	[2.1]	[10,918.6]	[0.10]	[0.10]	[0.23]
Log Age	-5,357.20	-9.4	262,692.6*	-0.36	-0.59	0.8
	[34,562.5]	[18.5]	[123,493.4]	[0.89]	[0.87]	[2.10]
Risk Loving	-552.1	-0.2	687.2	-0.01	-0.01	-0.01
	[419.7]	[0.2]	[1,457.8]	[0.01]	[0.01]	[0.02]
Sociable	-1,524.30	-0.9	-3,513.40	-0.04	-0.04	-0.06
	[1,677.8]	[0.9]	[6,081.7]	[0.04]	[0.04]	[0.08]
Trusting	20.5	0	-862.6	0	0	0
	[498.6]	[0.3]	[1,690.1]	[0.01]	[0.01]	[0.03]
Self- Esteem	368.3+	0.3*	793	0.01*	0.01*	0.02
	[216.8]	[0.1]	[564.9]	[0.01]	[0.01]	[0.01]
Pearlin Mastery	193.1	0	1,558.2*	0.01	0	0.01
	[245.8]	[0.1]	[669.8]	[0.01]	[0.01]	[0.01]
Father Self-employed?	-7,426.40	-4.7	11,062.60	-0.19	-0.25	-0.3
	[6,715.3]	[3.8]	[20,200.2]	[0.17]	[0.18]	[0.28]
Mother Self-employed?	3,348.20	0.4	49,632.10	0.03	0.05	-0.06
	[8,189.8]	[4.2]	[37,045.9]	[0.20]	[0.20]	[0.37]
Mother s education	-660.5	-0.3	2,573.5+	-0.02	-0.01	-0.01
	[448.5]	[0.2]	[1,495.9]	[0.01]	[0.01]	[0.03]
Father s education	-42.1	-0.1	-863.8	0	0	0
	[211.1]	[0.1]	[786.8]	[0.01]	[0.01]	[0.01]
Family Income in 1979	-728	-0.7	-8.9	0	-0.02	-0.03
	[1,955.0]	[1.1]	[6,836.5]	[0.05]	[0.05]	[0.10]
Family assets in 1985	0	0	0.5**	0	0	0
	[0.0]	[0.0]	[0.1]	[0.00]	[0.00]	[0.00]
Constant	109,936.10	88.8	-1,102,906.3	13.07	6.7	7.47
Other Demographic Controls	Y	Y	Y	Y	Y	Y
Occupational Controls	Y	Y	Y	Y	Y	Y
Industry Controls	Y	Y	Y	Y	Y	Y
Observations	4,687	4,543	4,083	4,687	4,543	3,903

Table displays 2-Stage Least Squares (2SLS) estimates of the relationship between various measures of earnings and wealth and being self-employed, after controlling for other factors. The regressions use AFQT (ability) scores to instrument for self-employment. Thus, the first-stage specification is identical to those reported in Table 10, except that it is estimated using OLS rather than Probit. Earnings/Wealth and Employment status are from the NLSY 1994 sample. Robust standard errors are shown in brackets. We use **, *, and + to denote $p<0.01$, $p<0.05$ and $p<0.1$, respectively.

Table A10: Mean Ability, Education and Earnings, Incorporated and Unincorporated Self-employed

2008 Status	Percentage	Mean AFQT Score	Mean Education	Mean Income	Family Worth
Wage Employed	82.9%	42.4	13.5	$48,794.0	$270,600.1
Unincorporated	15.6%	44.6	13.4	$57,324.6	$420,519.2
Incorporated	1.5%	49.1	14.3	$83,214.6	$855,534.6
Observations	6,623				