Risk-Bearing and Entrepreneurship[1]

Andrew F. Newman[2]
Boston University and CEPR

March 2007

[1] I thank P. Bolton, B. Holmström, I. Jewitt, P. Legros, E. Ligon, E. Maskin, D. Mookherjee, J. Nachbar, an anonymous referee, and numerous seminar participants for helpful comments.

[2] Department of Economics, Boston University, 270 Bay State Road, Boston, MA 02215, afnewman@bu.edu

Abstract

In the "Knightian" theory of entrepreneurship, entrepreneurs provide insurance to workers by paying fixed wages and bear all the risk of production. This paper endogenizes entrepreneurial risk by allowing for optimal insurance contracts as well as the occupational self-selection. Moral hazard prevents full insurance; increases in an agent's wealth then entail increases in risk borne. Thus, even under decreasing risk aversion, there are robust instances in which workers are wealthier than entrepreneurs. This empirically implausible result suggests that risk-based explanations for entrepreneurship are inadequate.

Keywords: moral hazard, occupational choice, principal-agent model

JEL classification numbers: D2, D8, L2, O16

1 Introduction

Entrepreneurship has long fascinated economists. The way we understand it affects our thinking about the processes generating growth and development, policies for influencing productivity and mitigating unemployment, even mechanisms underlying business cycles. One influential and intuitively appealing theory of entrepreneurship can be traced back to Cantillon [5] and Knight [14] and was formalized more recently by Kanbur [11] and Kihlstrom and Laffont [12]. In this theory, entrepreneurs — through the institution of the fixed wage contract — are viewed essentially as providers of insurance. Individuals choose between the safety of wages and the hazards of entrepreneurship according to their attitudes toward risk. More risk averse people (and with decreasing risk aversion, the poor) receive sure wages and work for the less risk averse (rich), who are the residual claimants. An attractive feature of this theory, if we accept decreasing risk aversion, is that it easily explains one of the oldest stylized facts in economics, namely the tendency for entrepreneurs to be richer than workers.

Granting the basic presumption that being an entrepreneur is riskier than being a worker, the theory is nevertheless incomplete, for it does not account for why the risks are exogenous, or more precisely that the choice of occupation is the *only* institutional arrangement available for risk sharing. After all, there are alternatives, most prominent among them the market for insurance contracts. The first question that needs answering, then, is why entrepreneurs should be bearing risk at all. One response is that the relevant risks are aggregate, and therefore cannot be insured away. Another is that while the risks may be idiosyncratic, some information problem prevents full diversification.

Either way, one is led to ask whether the Knightian theory still make plausible predictions if we take proper account of reasons for the inability of the market to provide perfect insurance. This paper address this question by examining how moral hazard on the part of entrepreneurs modifies the basic story (we shall briefly say something about alternative reasons for imperfect insurance at the end). We embed the risk sharing choices of entrepreneurs into a standard principal-agent framework and show that for a broad class of utility functions, competitive equilibrium will entail that entrepreneurs are poorer than workers, rather than the other way around. Thus a plausible modification of the basic Knightian model leads to an implausible prediction. The fragility of this theory's empirical predictions suggests that we probably should look elsewhere for explanations of the roles and sources of entrepreneurship.

The intuition behind the result is quite simple and depends on the existence of an apparently little-noticed property of the principal-agent model: when utility is separable in income and effort, the wealthier an agent is, the more risk he needs to bear in order to remain incentive compatible at a given effort level. The reason for this is that at low wealth levels, the income utility is very steep, so it only takes a small spread in the income "lotteries" generated by different effort levels to maintain a utility differential that will offset the cost of effort. But as the income utility flattens with greater wealth, the spread in incomes must be increased in order to maintain the differential. Thus, wealthy agents need to bear more risk. Finally, for the class of utility functions we identify (namely those in which the marginal income cost of providing utility is a convex transform of the utility function), this "increasing risk effect" swamps the decreasing risk aversion that goes with higher wealth, and we are led to the result.

The same intuition underlies a related result that we obtain for the standard formulation of the principal-agent model: agents with higher (expected) wages are monitored more than those with lower wages. The evidence on this is perhaps less clear than that on the relative wealth of workers and entrepreneurs. Nevertheless, both results point to a larger issue, which is to sort out the relative *empirical* importance of the two fundamental trade-offs in the theory of moral hazard, namely that between risk sharing and incentives (our focus) and that between surplus extraction and incentives (the focus, for instance, of the efficiency-wage and credit-constraint literatures).[1] While the risk-sharing trade-off is the one commonly emphasized, the present results suggest that the surplus extraction trade-off is at least as pertinent in the real world.

2 Model

There is a single good which is produced from labor according to the stochastic production function $f(L, \theta)$, where L is the amount of labor hired (or worker effort) and θ is a random variable indexing the state of the world, representing the risks that inhere in the production and sale of goods. Raising θ raises both the total and marginal products of labor, that is, $f(L, \theta)$ and $f_1(L, \theta)$ are increasing in θ (for instance, θ is a multiplicative noise). These risks are independent and identically distributed across firms. The production function satisfies the standard properties: $f(0, \theta) = 0$, $f_{11} < 0 < f_1$, $\lim_{L \to 0} f_1(L, \theta) = \infty$, $\lim_{L \to \infty} f_1(L, \theta) = 0$. Each firm also requires the effort e of an entrepreneur. We could suppose that the entrepreneur expends this effort in coordinating production, marketing output, finding suppliers, etc. It enters the production technology only through the distribution of θ.

There is a continuum of agents, indexed by the unit interval. In order for the self selection underlying the Knightian story to have any relevance, agents must differ in some way. One way to do this is to focus on preferences, allowing, for instance, that all the variation be indexed by some utility parameter; this is the tack taken by Kihlstrom and Laffont [12]. Here we shall

[1] The latter trade-off arises when limited liability or other lower bounds on the agent's payoff limit the punishment for outcomes indicative of undesired behavior; as a substitute, rewards for outcomes indicative of desired behavior must be paid, and this is costly to the principal. Often, models focusing on surplus extraction assume a risk-neutral agent.

follow a special case of this approach, which is to assume that agents have identical preferences and vary instead in some other characteristic, namely initial wealth. Since standard assumptions, such as decreasing absolute risk aversion, then place restrictions on the relationship between risk attitudes and marginal incentives, this approach has the advantage of generating additional testable implications almost for free.

Thus, let all agents have identical preferences represented by the von Neumann-Morgenstern utility $u(y) - e$, where $y \geq 0$ is realized income and e is the effort expended. The income utility $u(\cdot)$ has the usual properties: $u' > 0 > u''$ with decreasing absolute risk aversion (i.e., $u'''u' > (u'')^2$). In order to ensure that only risk sharing is at issue here, it is convenient to assume that $u(\cdot)$ is unbounded below.[2] Agents vary in the amount of initial assets or wealth a, the distribution of which is exogenous. An agent may choose one of two occupations in which to expend his effort: either he can become a worker, earning a sure wage w (so that his income is $a + w$), or he can be an entrepreneur, hiring L units of labor and earning the residual profit from production, which we denote $y(\theta)$. Note that in general the amount of labor that an entrepreneur wishes to hire could depend not only on the wage, but also on his characteristics, which in this case are limited to his wealth a.

Make the following assumptions on $L, \theta,$and e: entrepreneurs may hire any nonnegative amount of labor, θ is drawn from a finite set, and e is a binary

[2] This is a standard condition (see e.g. Grossman and Hart [8]) that ensures that no nonnegativity constraints on income bind at an optimum, which in turn purifies the model of any surplus extraction effect (in the usual formulation of the principal-agent model where the principal offers the contract, it guarantees that the agent's participation constraint binds).

It also effectively lets us ignore financing issues (i.e., we assume entrepreneurs always have enough income ex-post to pay their workers; with utility unbounded below, this will never be a problem), so that we can focus purely on the role of risk aversion in the occupational choice.

variable taking values in $\{0, 1\}$ (the last assumption allows us to abstract from issues having to do with how effort levels might also change with parameters of the model; the independence of the effort cost function from occupation facilitates focus on risk bearing as the determinant of occupational choice). Denote by $q(\theta)$ the probability of θ when $e = 1$ and $p(\theta)$ when $e = 0$; $q(\cdot)$ stochastically dominates $p(\cdot)$ and the two distributions have common support. We assume throughout that it is optimal for both workers and entrepreneurs to set their effort levels equal to 1.

Each agent takes the wage w as given and chooses the occupation which gives him the higher utility, taking into account any incentive compatibility constraints that may apply. Competitive equilibrium in this economy consists of a partition of the set of agents into a set of workers and a set of entrepreneurs, and a wage w such that the total demand for workers, given the wage, is equal to the total supply. Our main concern here is with characterizing equilibrium, especially this partition (see Kihlstrom and Laffont [12] for a more formal definition of equilibrium and a treatment of existence).[3]

3 No Insurance for Entrepreneurs

This is the case considered in the literature. It is convenient to represent the indirect utility of an agent as a function of his wealth and occupation; let $V_E(a)$ be the utility achievable by an agent with wealth a if he becomes

[3] For completeness, we offer a slightly less general definition which is valid if the distribution of initial wealth is represented by an atomless measure $\eta(a)$. Suppose an entrepreneur with wealth a demands $L(w; a)$. An equilibrium is then a partition of the wealth interval into sets A_E and A_W such that

$$\int_{A_W} d\eta(a) = \int_{A_E} L(w; a) d\eta(a).$$

an entrepreneur, and $V_W(a)$ be the utility he achieves as a worker. Since by assumption, being a worker is safe, he obtains the income $w+a$, and $V_W(a) = u(w+a)-1$. Denote by $y(\theta)$ the income of an entrepreneur in state θ. In the absence of any insurance for entrepreneurs, we have $y(\theta) = f(L,\theta) - wL + a$ and $V_E(a) = \max_L E|_{e=1}\{u(f(L,\theta)-wL+a)\}-1$ (we assume that it is always optimal for the entrepreneur to set $e = 1$), where the expectation is taken with respect to the distribution of θ, conditional on effort being set at 1.

When contemplating becoming an entrepreneur, an agent will choose L to maximize his expected utility, taking as given the wage w. Since a larger firm will generate a larger spread among the possible output realizations, we might expect that entrepreneurs will attempt to self-insure by varying the sizes of their firms. Indeed, it is easy to show that the wealthier — and therefore the less risk averse — an entrepreneur, the larger will be his firm: the argument, provided in the Appendix, is a standard one from the theory of portfolio choice, and depends on the assumptions of decreasing risk aversion and on the properties of the production function stated above. Since the expected profit $E|_{e=1}\{f(L,\theta) - wL\}$ is increasing when L is below its first-best level, it follows that entrepreneurs' profits are increasing in wealth. This model therefore captures the old idea that profits are a return to risk-bearing. It also provides an account for the observation that firms in the same industry vary in size.

In equilibrium there is a wealth level $\bar{a}$ at which an agent is indifferent between the two occupations (since $V_W(a)$ and $V_E(a)$ are continuous functions of a, if there were no such $\bar{a}$, everyone would prefer one of the occupations, which cannot be an equilibrium). We now argue that this point is unique, and that all agents who become workers are below $\bar{a}$, while all entrepreneurs are above it. To see this, note that at $\bar{a}$, $u(\bar{a}+w) = E|_{e=1}\{u(f(L(w,\bar{a}),\theta) - wL(w,\bar{a}) + \bar{a})\}$, that is, the worker's income w is the certainty equivalent of the entrepreneur's income $f(L(w,\bar{a}),\theta) - wL(w,\bar{a})$ for this marginal agent. Consider an agent with wealth a' slightly greater than $\bar{a}$, and suppose she contemplated being an entrepreneur *using the same amount of labor* as the agent with $\bar{a}$: since the agent with a' is less risk averse than the agent with $\bar{a}$, she strictly prefers the lottery (i.e. the entrepreneur's income) to the worker's safe income; since as an entrepreneur agent a' would typically choose some level of L different from that chosen by $\bar{a}$, she prefers entrepreneurship all the more strongly. This argument implies that $V_E(a)$ cuts $V_W(a)$ from below wherever they are equal (in other words, $V_E'(\bar{a}) > V_W'(\bar{a})$) , and so $\bar{a}$ is unique.

To summarize, we have the following results, which are virtually the same as those in Kihlstrom and Laffont [12]:

Proposition 1 *When entrepreneurs cannot insure, in competitive equilibrium*

(a) Firm size — as measured by the amount of labor demanded — and expected profit increase with the entrepreneur's initial wealth;

(b) There exists a wealth level $\bar{a}$ such that all workers have wealth at most $\bar{a}$ and all entrepreneurs have wealth at least $\bar{a}$.

4 The Second-Best Economy with Moral Hazard

Proposition 1 makes the empirically plausible predictions that wealthy agents will tend to become entrepreneurs and poor ones workers, and accounts for variations in firm size according to the risk preferences of their owners. But observe that even though entrepreneurs are less risk averse than workers, they are still risk averse, and would be better off if they could share risks, say through a stock market. The model of the previous section has arbitrarily excluded them from engaging in these contracts. In this section, we consider one possibility for an improved risk sharing arrangement, one which nonetheless leaves some risks for entrepreneurs to bear.

There are, of course, several reasons why entrepreneurs might bear some (idiosyncratic) risk. Predominant among them is moral hazard — the entrepreneur's activities are not verifiable to the outside world. We shall now consider the optimal risk-sharing arrangement when the entrepreneur's effort is not observable (workers' effort and the level of L is assumed to be verifiable, however).

Assuming free entry into the insurance market and that each entrepreneur's contract satisfies incentive compatibility, the optimal risk sharing scheme can be written as a function $y(\theta)$ which satisfies

$$\max_{y(\theta),L} \sum q(\theta)u(y(\theta)) - 1 \tag{1}$$

$$\text{s.t. } \sum q(\theta)y(\theta) \leq \sum q(\theta)f(L,\theta) - wL + a \tag{2}$$

$$\sum q(\theta)u(y(\theta)) - 1 \geq \sum p(\theta)u(y(\theta)). \tag{3}$$

(We continue to assume that $e = 1$ is optimal; if it were not, then full insurance would always be possible, making entrepreneurship as safe as working.) Let $\hat{V}_E(a)$ denote the value of this problem and $(\hat{y}(\theta), \hat{L})$ its solution. At an optimum, both constraints bind: that (3) does is a standard result in the principal-agent literature (e.g. Bolton and Dewatripont [4]). As for (2), if it did not bind, it would be possible to raise the $y(\theta)$ in such a way as to increase utility by the same amount in each state. This in turn raises the payoff without affecting (3), a contradiction. Let us now examine properties of the solution in somewhat greater detail.

The first observation to make is that, regardless of initial wealth (or risk attitude), all entrepreneurs will choose the same size of firm $\hat{L}(w)$, namely that which maximizes $\sum q(\theta)f(L,\theta) - wL$ (following the argument in the previous paragraph, increasing the expected profit, thereby relaxing (2), allows the entrepreneur to raise his payoff).[4] Immediately, then, we have a depar-

[4] The demand function $\hat{L}(w)$ is the same as that in the first-best economy; however,

ture from the closed-market formulation considered in the previous section. Expected profits are also independent of the entrepreneur's wealth.

We shall also use the following

Lemma 2 *Let $\hat{V}_E(a)$ be the value function for problem (1). Then*

$$\frac{1}{\hat{V}_E'(a)} = \sum q(\theta) \frac{1}{u'(\hat{y}(\theta))} \tag{4}$$

Proof. Let α and β denote the multipliers on the constraints (2) and (3) respectively (as we pointed out above, they both bind at an optimum). The first-order condition with respect to $y(\theta)$ can be written

$$\frac{q(\theta)}{\alpha} = \frac{q(\theta)}{u'(\hat{y}(\theta))} - \frac{\beta}{\alpha}(q(\theta) - p(\theta));$$

adding over θ gives

$$\frac{1}{\alpha} = \sum q(\theta) \frac{1}{u'(\hat{y}(\theta))}.$$

Noting from the envelope theorem that $\hat{V}_E'(a) = \alpha$ completes the argument. ■

This result holds even if the there are many effort levels, provided utility is additively separable in income and effort: the same proof applies with only minor modification. It simply states that at an optimal contract, the marginal cost of providing the expected utility $V_E(a)$ to the entrepreneur (the left-hand side) is equal to the expected marginal cost of providing him with utility $u(\hat{y}(\theta))$ in each state (right-hand side; recall that the cost in terms of output of raising utility a small amount from $u(y(\theta))$ is $\frac{1}{u'(y(\theta))}$). A related result appears in Rogerson [23].

---

in the second-best equilibrium, the wage will be smaller, and the firms larger and fewer, than in the first-best.

4.1 Occupational Self-Selection in the Second-Best Economy

We have already noted one departure of the second-best scheme from the closed-market scheme, namely the absence of a scale effect. We now show that there is likely to be a second departure, namely in the way agents self-select into occupations.

Make one additional restriction on the income utility:

Condition 3 *There is a strictly convex function $h : \mathbb{R} \to \mathbb{R}_+$ such that*

$$\frac{1}{u'(x)} = h(u(x)).$$

The role of this assumption will be discussed below. Suffice for now to say that it is satisfied by a broad class of utility functions; all utilities in the constant-relative-risk aversion (CRRA) class $\frac{x^{1-\sigma}}{1-\sigma}$ with $\sigma > 1/2$ satisfy it (indeed, with CRRA, the condition is automatically entailed in the requirement that $u(\cdot)$ is unbounded below, for then $\sigma \geq 1$).

Once again the strategy is to compare the slopes of the value functions for the two occupations at a point of indifference between them. Just as in the case with no insurance, equilibrium entails that such a point exists; call it $\hat{a}$. Since for an agent with $\hat{a}$, being a worker and expending a unit of effort generates the same utility as being a (risk-bearing) entrepreneur and expending a unit of effort, we have

$$u(w + \hat{a}) = \sum q(\theta)u(\hat{y}(\theta)).$$

In other words, for the marginal agent, the worker's income is the certainty equivalent of the entrepreneur's. The well-known theorem of Arrow and Pratt tells us that anyone who is less risk averse than a consumer with utility function $u(\cdot)$ will strictly prefer the entrepreneur's income lottery. In other words, for any strictly convex increasing $h(\cdot)$,

$$h(u(w+\hat{a})) < \sum q(\theta) h(u(\hat{y}(\theta))).$$

But under Condition 3, $1/u'(\cdot) = h(u(\cdot))$ for some such $h(\cdot)$; thus,

$$\frac{1}{V_W'(\hat{a})} = \frac{1}{u'(w+\hat{a})} < \sum q(\theta) \frac{1}{u'(\hat{y}(\theta))} = \frac{1}{V_E'(\hat{a})}. \tag{5}$$

(The first equation follows trivially from the definition of V_W, and the second is from Lemma 2.) Thus we have

$$V_W'(\hat{a}) > V_E'(\hat{a}),$$

and we conclude that in the equilibrium of the second-best economy entrepreneurs are poorer than workers.

The principal properties of equilibrium in the second-best economy are summarized in

Proposition 4 *In a second-best equilibrium:*

(a) Both firm size and expected profit are independent of the entrepreneur's initial wealth;

(b) There exists a wealth level $\hat{a}$ such that all entrepreneurs have wealth at most $\hat{a}$ and all workers have wealth at least $\hat{a}$.

What this proposition shows is that the Knightian theory, when pushed to a reasonable standard of theoretical consistency, predicts that quite commonly we should find that the poor hiring the rich. The empirical implausibility of this result calls into question theories that view the primary determinant of who becomes an entrepreneur to be risk attitudes and the primary function of the entrepreneur as an insurer of workers. In particular, it seems that the evidence suggests that the risk-sharing-versus-incentives-trade-off is not the primary one operating in the occupational choice between entrepreneur and worker.

4.2 Analysis of the Main Result

Why the difference in predictions about the occupational choice? Some intuition for the result is contained in the following proposition, which states that as an entrepreneur's wealth increases, he needs to bear more risk in order to remain incentive compatible.[5]

Proposition 5 *Let $v(a)$ be the variance of the least-risk contract satisfying the binding forms of (2) and (3), i.e.*

$$v(a) = \min_{y(\theta)} \sum q(\theta)[y(\theta) - \sum q(\theta)y(\theta)]^2$$

$$s.t. \ \sum q(\theta)y(\theta) = \sum q(\theta)f(\hat{L}(w), \theta) - w\hat{L}(w) + a$$

$$\sum q(\theta)u(y(\theta)) - 1 = \sum p(\theta)u(y(\theta)).$$

Then $v'(a) > 0$.

Proof. Let γ be the multiplier on the first constraint and δ be the multiplier on the second. Then the envelope theorem tells us that $v'(a) = -\gamma$. We need only show that $\gamma < 0$. The first-order condition for this problem can be written

$$2\frac{q(\theta)(y(\theta) - X)}{u'(y(\theta))} = -\frac{\gamma q(\theta)}{u'(y(\theta))} + \delta(q(\theta) - p(\theta)),$$

where $X \equiv \sum q(\theta)f(\hat{L}(w), \theta) - w\hat{L}(w) + a$. Adding over θ yields

$$2\sum \frac{q(\theta)(y(\theta) - X)}{u'(y(\theta))} = -\gamma \sum \frac{q(\theta)}{u'(y(\theta))}.$$

[5] Subject, of course, to the well-worn caveat that increases in variance do not necessarily entail increases in risk in the sense of second-order stochastic dominance.

Thus γ is negative if and only if $\sum \frac{q(\theta)(y(\theta)-X)}{u'(y(\theta))} > 0$. Now, for all θ such that $y(\theta) > X$, we have, by concavity of $u(\cdot)$, $\frac{q(\theta)(y(\theta)-X)}{u'(y(\theta))} > \frac{q(\theta)(y(\theta)-X)}{u'(X)}$; similarly, for all θ such that $y(\theta) < X$, we also have $\frac{q(\theta)(y(\theta)-X)}{u'(y(\theta))} > \frac{q(\theta)(y(\theta)-X)}{u'(X)}$. Therefore, $\sum \frac{q(\theta)(y(\theta)-X)}{u'(y(\theta))} > \sum \frac{q(\theta)(y(\theta)-X)}{u'(X)} = 0$. ■

Observe that the only properties (besides differentiability) of the utility used in the proof are that $u(\cdot)$ is increasing and concave — in particular, no third-derivative conditions are required.[6] The intuition for the result is very simple: at low wealth levels the income utility is very steep, so a small spread among the $y(\theta)$, when weighted by the different distributions $q(\cdot)$ and $p(\cdot)$, will be enough to maintain a utility differential that will offset the cost of effort. As the income utility flattens, the spread in incomes must be increased in order to maintain the differential. Thus, wealthy agents need to bear more risk.

It is this increasing risk effect, which is not present in the closed market case, where the required risk for a given size of firm (as opposed to the chosen risk which changes because the firm size does) is constant, that leads to the result in Proposition 4. If the risk increases fast enough, compared to the rate at which risk aversion decreases, then the first effect dominates.

Now we can understand the role of Condition 3. It is essentially a requirement that risk aversion not decline too quickly (relative to the rate at which marginal utility does) and therefore guarantees that this trade-off goes at the right speed: denoting absolute risk aversion by ρ and its derivative by ρ', an equivalent way to write the condition is $\frac{\rho'}{\rho} > 2\frac{u''}{u'}$.[7]

[6]The result is also valid for multiple effort levels, at least if the "first-order approach" is valid (almost exactly the same proof applies), in which case it is properly interpreted to say that the amount of variance needed to implement a given effort level is increasing in the agent's wealth.

[7]In fact, this condition has appeared in the literature before, albeit in a different context. It is among the suite of conditions provided by Jewitt [10] to justify use of the first-order approach.

Meanwhile, recall that the nonnegative profit condition permits the entrepreneur to choose the first-best scale for her firm. In a first-best world, of course, the entrepreneur would have the entire half-space from which to choose; here she is confined to a (full-dimensional) subset of that set. Moral hazard doesn't really cause markets to close so much as it leaves them "half-open."

4.3 Monitoring and Task Assignment

Lemma 2 has other implications. Consider for instance the standard Principal-Agent setting in which the agent is now interpreted as an employee and the principal is a firm. We are interested in how the firm will choose the level of monitoring it applies to the worker.

To model this, assume that there is a family of monitoring technologies indexed by m. Continue to assume that a worker can choose from among two effort levels with probability distributions q and p. We assume that higher m implies a higher level of monitoring, i.e., that the higher is m, the more efficient is the information system $\{q(\theta, m), p(\theta, m)\}$.[8] For a fixed level of m and a worker whose outside option yields utility $\underline{u}$, the firm minimizes the cost of a worker $C(\underline{u}, m)$, i.e.

$$C(\underline{u}, m) \equiv \min_{y(\theta)} \sum q(\theta, m) y(\theta)$$

$$\text{s.t. } \sum q(\theta, m) u(y((\theta)) - 1 \geq \underline{u}$$

$$\sum q(\theta, m) u(y(\theta)) - 1 = \sum p(\theta, m) u(y(\theta)).$$

First observe that Lemma 2 applies to this problem and can be written

[8] Several rankings of information systems have been studied in the incentive literature (see, e.g. Holmström [9] or Kim [13]). For present purposes, it does not matter which we use.

$\frac{\partial C(\underline{u},m)}{\partial \underline{u}} = \sum \frac{q(\theta,m)}{u'(y(\theta))}$. More efficient information implies that $C(\underline{u}, m)$ is decreasing in m. Moreover, it is clear that $C(\underline{u}, m)$ is increasing in $\underline{u}$: workers with higher outside options receive higher average wages.

The firm has a series of tasks that it is assigning to workers with different outside options. The tasks differ in the observability of the effort expended on them; workers are equally capable at the tasks (more generally, the outside option $\underline{u}$ is uncorrelated with the worker's cost or the expected revenue generated by the worker at each effort level), and the question is which workers should be assigned to which task. Thus we assume that the tasks are ordered by m as above. (Alternatively, the firm is choosing the level of costly monitoring to be expended on a worker: more monitoring means a higher level of m.) It is a well-known consequence of the theory of assortative matching that if the firm is trying to minimize its total cost, the worker with the higher $\underline{u}$ will be assigned to the task with higher m if $C(\underline{u}, m)$ is submodular in $(\underline{u}, m)$ (the firm will choose to monitor workers more intensively if they have a higher level of $\underline{u}$.)

Now suppose that $u(y) = \ln y$. Then $1/u' = y$ and it follows that $\sum \frac{q(\theta,m)}{u'(y(\theta))} = \sum q(\theta, m)y(\theta) = C(\underline{u}, m) = \frac{\partial C(\underline{u},m)}{\partial \underline{u}}$. Thus if $m > \hat{m}$, we have $\frac{\partial C(\underline{u},m)}{\partial \underline{u}} < \frac{\partial C(\underline{u},\hat{m})}{\partial \underline{u}}$ which yields

Proposition 6 *If workers have logarithmic utility for income, then workers with higher average wages are assigned to more easily observed tasks and are monitored more intensively.*

Although we have shown this to be true in the logarithmic case (i.e. logarithmic utility is sufficient for submodularity of $C(\underline{u}, m)$), it certainly holds for a broader class of utility functions; whether it holds for a class as broad as the one satisfying Condition 3 without undue restrictions on the informativeness orderings of the monitoring technologies is a subject for future research.[9]

[9] In fact, limiting comparison to a perfectly monitored (therefore safe) task and an

While we are somewhat skeptical of the empirical plausibility of either implication of this proposition, there appears to be little systematic evidence on this question (an exception is MacLeod and Parent [18], which does suggest that it is the better paid who are monitored less). The main point is that the result does offer a possible avenue for research on the empirical significance of the risk-sharing/incentives trade-off.

5 Discussion

We have examined one prominent interpretation of the Knightian idea that entrepreneurship is a form of risk sharing and profits a return to risk-bearing and have shown that, when properly specified, it can easily lead to implausible predictions. The analysis suggests that alternative approaches are more likely to be useful for understanding entrepreneurship.[10]

imperfectly monitored (risky) one, it is a simple matter to show, along the lines of the proof of Proposition 4, that under Condition 3, the safe task is assigned to workers with $\underline{u}$ above some threshold, while the risky one is assigned to those below the threshold. In particular, if $\underline{u}$ is the utility that an agent can get by working for a wage, with variations in $\underline{u}$ due to variations in wealth, then one can interpret this model as one of occupational choice between working and entrepreneurship when the market for insurance contracts is monopolized: the occupational outcome preferred by the monopolist would replicate the one in Proposition 4.

[10] A couple come to mind. One is based on credit market imperfections (which in turn depend on a property of the agent's payoff quite different from risk aversion, namely whether it is bounded below), which can restore the predictions of Proposition 1, namely that the wealthy bear more risk, tend to be the entrepreneurs, and have larger firms (Banerjee and Newman [2], [3]).

Another looks at "ability" (Lucas [17], Lazear [16]), which is potentially fruitful to the extent that it can be measured and its effects disentangled from skills that might be acquired in the market (the credit market imperfection approach would apply to the

This is not to say that the risk-sharing-versus-incentives trade-off is irrelevant in other contexts. For instance, the evidence in Ackerberg and Botticini [1] that wealthy peasants in Renaissance Tuscany were likely to grow relatively safe crops such as cereals, while poor peasants tended toward risky ones such as vines, does square with the model.

More generally, the model underscores the importance of matching or selection effects in the study of contracts and organization. Notice that a straightforward interpretation of Proposition 5 would say roughly that as agents grow wealthier, the amount of risk they bear increases. But Proposition 4, which takes explicit account of the character of agents' outside opportunities (in this case the necessity that at least some of them take the wage contract, which in turn implies the existence of the "marginal" wealth levels $\bar{a}$ and $\hat{a}$), predicts almost the opposite: while increasing risk occurs over a range of low wealth levels, the overall effect is of an "inverted U," with the rich agents bearing no risk at all.

We turn now to a discussion of robustness of the main result.

5.1 Multiple Effort Levels

In the standard separable-utility specification of the principal-agent model, leisure (i.e. reduction of effort) is a normal good. If one assumes that effort is a continuous variable and that its cost is strictly convex, then as the agent's wealth increases, he will lower his effort, thereby reducing its marginal cost; this in turn weakens the increasing-risk effect which was essential to the poor-hire-the-rich result in Proposition 4. One then wonders whether the result is simply a consequence of the two-effort level formulation we have used.

There are two approaches to this question. First, we could try to characterize equilibria of a multiple-effort-level version of the model explicitly. In general, this appears to be a difficult exercise. We do have an example in

latter).

which the probabilities and effort cost are linear in $e \in [0,1]$ where entrepreneurs behave as in the two effort case, choosing $e = 1$, while workers choose effort in the interior of the interval. The existence of a marginal agent who is indifferent between the two occupations now implies that his income as a worker is strictly lower than his expected income as an entrepreneur, and under Condition 3, the chain of inequalities (5) remains valid (recall that (4) in Lemma 2 doesn't depend on the number of effort levels). Thus, in this case at least, the configuration of occupational choices is as stated in Proposition 4. Moreover, for strictly convex effort costs which are "close" to linear, the equilibrium configuration remains the same. See the Appendix.

An alternative response also exploits the fact that Lemma 2 is valid for any number of effort levels. Suppose that $u(y) = \ln y$ (and so satisfies Condition 3); effort cost is $c(e)$, where $c' \geq 0$, $c'' > 0$, $c(0) = c'(0) = 0$; a worker who exerts $e_W \geq 0$ earns $e_W w$; and that if an entrepreneur expends e_E, the probability of state θ is $q(e_E, \theta)$, where the distributions $q(\cdot, \cdot)$ have common support and increases in e_E result in first-order stochastic increases in the distribution of θ. Denote the expected profits of the marginal agent when he is an entrepreneur by $E_{e_E}\pi$. Again , since for $u(y) = \ln y$, $1/u'(y) = y$, Lemma 2 implies that $V_E'(\hat{a}) > V_W'(\hat{a})$ if and only if $e_W w + \hat{a} > E_{e_E}\pi + \hat{a}$. Now, effort is nonincreasing in wealth for both workers and entrepreneurs (this is easy to check from the first-order condition for workers, but is somewhat more complicated for entrepreneurs — see the Appendix). This implies that the wealthier a worker is, the less he earns; similarly, the wealthier is an entrepreneur, the lower are his profits on average (given the specified family of distributions, effort and labor are complements, so lower effort corresponds to a smaller and less profitable firm). Suppose then we have the realistic case in which the rich hire the poor. The argument we just made gives us the following

Proposition 7 *Suppose income utility is logarithmic and there is a wealth level a^* such that all entrepreneurs have wealth at least a^* and all workers* *have wealth at most a^*. Then every worker has higher (expected) earnings than every entrepreneur.*

This result suggests that allowing for multiple effort levels in the Knightian model need not reduce the likelihood of empirically implausible predictions.[11]

5.2 Other Interpretations of the Knightian Theory

Since embedding the Knightian theory into a standard moral hazard framework reveals the fragility of its predictions, it is natural to ask what happens in the presence of other causes of imperfect insurance.

We have made the conventional supposition, natural for the market economies to which the Knightian theory is supposed to pertain, that a firm's output (or profit) is verifiable. A logical alternative is that it is unobservable to all but the entrepreneur running the firm, so that any kind of insurance for her is simply infeasible. Indeed, the two cases considered in Sections 3 and 4 could be nested in a single model that also incorporates all the intermediate ones by supposing that only some garbling η of the output state θ is observable to the insurer; a matrix $P(\eta|\theta)$ represents the distribution of η given θ, and the contract is now an η-contingent transfer from the entrepreneur to the insurer. The two cases correspond to different assumptions about the form of this exogenous garbling matrix.[12]

[11] If there are multiple crossings of the value functions, then things are not so simple. But once again, we are forced to accept the predictions that either the poorest members of the economy are entrepreneurs who hire people wealthier than they, or the agents with the lowest initial wealth earn the highest incomes.

[12] At one extreme, when $P(\eta|\theta) = P(\eta|\theta')$ for all η, θ, θ', we have completely unobservable output; it is not hard to check that the optimal contract in this instance corresponds to our no-insurance case. At the other, where $P(\eta|\theta)$ is the identity matrix, we have our second-best case. By continuity, occupational patterns will resemble those of our two

Since difficulties with verifying profits stem largely from incentives by entrepreneurs to hide them, it is more appropriate to replace the exogenous information structure by one that takes account of these incentives and contractual responses.[13] One literature (e.g. Townsend [25], Mookherjee and Png [20]) that takes this approach assumes that (costly) ex-post audits are available; whether these provide an effective obstacle to efficient contracting depends in part on the presence of lower bounds on utility. Under the preference assumptions of this paper, a contract can allow for arbitrarily rare audits and punish an entrepreneur severely if he is caught lying about output, possibly reducing welfare costs to negligible levels relative to the second best. If utility is bounded below, on the other hand, auditing will generally be costly in equilibrium, but then as we have seen, the pure risk-sharing-versus-incentives trade-off is unlikely to be the only operative one.

Allowing for unobservable output may leave the door open for the empirical relevance of the Knightian theory, but it also has other empirical implications that warrant further exploration. Taking the verifiability of profit to be more of a problem in developing countries than in developed ones, our results suggest that the poor should bear little risk in the former, and rather more risk in the latter. Though comparison across countries is difficult, there is evidence (e.g. Morduch [19]) of the opposite: the poor in developing countries are the ones particularly vulnerable to financial risk.

Another interpretation of the risk-bearing story is that entrepreneurs bear the *aggregate* risks, rather than idiosyncratic ones. Then if one assumes that there is no possibility for entrepreneurs to insure against these risks, while workers are perfectly insured, one can easily show that the poor (and highly risk averse) agents become workers while the wealthy (and less risk averse) become entrepreneurs.

But once again, this fails to correspond to optimal risk sharing. For

cases for open sets of garbling matrices near the extremes.

[13] Dixit [6], [7], among others, has argued for this methodological imperative.

instance, with CRRA utility, it is not hard to show that if there is only aggregate risk, everyone — worker and entrepreneur, rich and poor — will bear a risk that is proportional to his initial endowment. In this case, the problem is not so much that the rich end up bearing smaller risks than the poor (they don't), it is that — from the risk-bearing point of view, at least — there is no meaningful distinction between the two occupations.

Adverse selection might also account for incomplete insurance. Perhaps entrepreneurs know more about the likelihood of success of their ideas than do outsiders. Specifically, suppose for simplicity that projects are of fixed size, and have two outcomes. Projects come in two qualities, measured by the probability of success; the quality of a project is private information. In this case, one can use a Rothschild and Stiglitz [24] screening model to show that in equilibrium, *regardless of wealth level* (or risk attitude) agents with high quality projects become entrepreneurs, while agents with low quality projects are screened off and become workers. This may be a socially efficient and perhaps even plausible outcome, but it has nothing to do with risk attitudes (the same result could come from a model in which everyone is risk neutral).

Finally, one could depart from the rational expectations framework and suppose that people have different beliefs about their success which they don't completely revise in the light of market-transmitted information. In this case, though, it is easy to construct "Edgeworth Box" examples where markets operate perfectly but in which a highly optimistic but risk averse trader bears all the risk while a more guarded but risk neutral one bears none, opposite the predictions of the Knightian model.[14]

[14]See Wu and Knott [26] for a recent model along similar lines. They also survey evidence that suggests that entrepreneurs' risk preferences are not distinguishable from those of workers.

5.3 Implications for the Theory of Moral Hazard

Our results have broader implications for the theory of moral hazard, at least as it expressed in the principal-agent model. In addition to the textbook view that "the" fundamental trade-off is the one between the provision of incentives and the sharing of risk (e.g. Mas-Colell, Whinston, and Green [21, p. 482] or Bolton and Dewatripont [4, p. 23]; Laffont and Martimort [15] seems to be the exception that proves the rule), principal-agent theory has revealed a trade-off between incentives and surplus extraction: if there is a lower bound on the utility that the agent can receive (contrary to what we have assumed in order to isolate the risk sharing aspects), it may be necessary to pay him a rent in order to maintain incentives. The distinction between these trade-offs is crucial for understanding the role of organizations, and for evaluating their economic performance. Which of them is more relevant to understanding our world is of course an empirical question, and it is therefore important to derive empirical implications of these trade-offs.[15] Though conclusive evidence is currently lacking, our results suggest that for real-world moral hazard, risk sharing may take a back seat to surplus extraction.

6 Appendix

Proof of Proposition 1(a): Differentiate the first order condition for L

$$E|_{e=1}\{u'(f(L,\theta) - wL + a)(f_1(L,\theta) - w)\} = 0$$

and observe that $\frac{dL}{da}$ has the same sign as $E|_{e=1}\{u''(f(L,\theta)-wl+a)(f_1(L,\theta)-w)\}$. Now $f_1 - w$ is monotonic and must be negative for low values of θ and positive for high values of θ if the first-order condition is to be satisfied. Let $\hat{\theta}$ be a value for which $f_1 - w < 0$ for $\theta < \hat{\theta}$ and $f_1 - w > 0$ for $\theta > \hat{\theta}$. Then

[15] A similar point about the empirical relevance of the risk-sharing tradeoff is made by Prendergast [22].

by decreasing risk aversion,

$$-\frac{u''(y(\theta))}{u'(y(\theta))} > -\frac{u''(y(\hat{\theta}))}{u'(y(\hat{\theta}))},\ \theta < \hat{\theta};$$

upon multiplying both sides by $-q(\theta)u'(y(\theta))(f_1(L,\theta) - w)$ and adding up to $\hat{\theta}$, one obtains

$$\sum_{\theta<\hat{\theta}} q(\theta)u''(f_1 - w) > \frac{u''(y(\hat{\theta}))}{u'(y(\hat{\theta}))} \sum_{\theta<\hat{\theta}} q(\theta)u'(f_1 - w).$$

Similarly, one can show that

$$\sum_{\theta>\hat{\theta}} q(\theta)u''(f_1 - w) > \frac{u''(y(\hat{\theta}))}{u'(y(\hat{\theta}))} \sum_{\theta>\hat{\theta}} q(\theta)u'(f_1 - w).$$

Taken together, these imply

$$Eu''(f_1 - w) > \frac{u''(y(\hat{\theta}))}{u'(y(\hat{\theta}))} Eu'(f_1 - w) = 0,$$

where the equality follows from the first-order condition.■

We now elaborate on the arguments in Section 5.1.

A Linear Example. Suppose that effort is chosen from the unit interval and that its cost is linear. Denote the income of a worker with wealth a by $y_W(a)$ and the effort level he chooses by $e_W(a)$. For the marginal agent with wealth $\hat{a}$,

$$u(y_W(\hat{a})) - e_W(\hat{a}) = Eu(y(\theta)) - e_E(\hat{a}),$$

where $y(\theta)$ is the entrepreneur's income function and e_E his effort. Thus if $e_E(\hat{a}) \geq e_W(\hat{a})$, we deduce that $u(y_W(\hat{a})) \leq Eu(y(\theta))$ and then use (5) to conclude that the entrepreneurs are poorer than the workers.

Suppose that when the entrepreneur chooses e, the probability of state θ is $eq(\theta) + (1-e)p(\theta)$, where $q(\cdot)$ stochastically dominates $p(\cdot)$ and the two distributions have common support (we shall suppress notating the dependence on θ in what follows). The entrepreneur's problem is

$$\max_{y(\theta),e} e \sum (q-p)u(y) + \sum pu(y) - e$$

$$\text{s.t. } e\sum(q-p)y + \sum py = \max_L e\sum(q-p)f(L) + \sum pf(L) - wL + a \quad (6)$$

$$e\sum(q-p)u(y) + \sum pu(y) - e \geq e'\sum(q-p)u(y) + \sum pu(y) - e', \forall e' \in [0,1]. \quad (7)$$

We now show that any nontrivial solution to this problem entails $e_E = 1$ for all a, and therefore $e_E \geq e_W$. Observe that (7) can be written

$$e[\sum(q-p)u(y) - 1] \geq e'[\sum(q-p)u(y) - 1];$$

thus $\sum(q-p)u(y) \geq 1$ (else $e = 0$). If $\sum(q-p)u(y) > 1$, then $e = 1$ is the only one satisfying (7), and we are done. If instead $e \in (0,1)$, then $\sum(q-p)u(y) = 1$. Raising e until it is equal to 1 therefore leaves (7) unchanged. If it relaxes (6), then utility in each state can be raised without affecting (7), contradicting the assumption that $e < 1$ is optimal. Increasing e relaxes (6) provided $\sum(q-p)(f(L(e)) - y) > 0$ (here $L(e)$ denotes the profit maximizing choice of L given e); we claim this condition must hold. Suppose not. From (6), $\sum py \leq \sum pf(L(e)) - wL(e) + a$. Then

$$\sum pu(y) \underset{(a)}{<} u(\sum py) \underset{(b)}{\leq} u(\sum pf(L(e)) - wL(e) + a) \underset{(c)}{\leq} u(\sum pf(L(0)) - wL(0) + a);$$

where (a) is by risk aversion and (c) is from the fact that $L(0)$ maximizes $\sum pf(L) - wL$. Since $\sum(q-p)u(y) = 1$, the entrepreneur's payoff is $\sum pu(y)$;on the other hand, receiving $\sum pf(L(0)) - wL(0) + a$ in every state is feasible, and we are thus led to a contradiction.

This result shows that at every equilibrium wage of the economy with linear utility, the occupational configuration is as in Proposition 4. Now consider perturbations of this economy of the following sort. Replace the cost function with one that is of the form $c_\lambda(e) = (1-\lambda)e + \lambda C(e)$,where $C(e)$ is some smooth, strictly convex, strictly increasing function. Since value functions are continuous in the parameters, the relative slopes of $V_E(a)$ and $V_W(a)$, thought of as functions of λ and w, will not change for values of λ in a neighborhood of zero. The direct effect of changing λ slightly will be small, and the indirect effect, through the change in w, will also be small: for some interval $(0, \bar{\lambda})$, every equilibrium wage of an economy in which the cost function is $c_\lambda(e)$, $\lambda \in (0, \bar{\lambda})$, must be close to one in the original economy. Thus for this family of strictly convex costs, at least, the main result in Proposition 4 is unchanged.

Entrepreneurial Effort Decreases with Wealth. The entrepreneur solves

$$\max_e G(e, a),$$

where

$$G(e, a) \equiv \max_{\{y(\theta)\}} \sum q(e, \theta) u(y(\theta)) - c(e)$$

$$\text{s.t.} \quad \sum q(e, \theta) y(\theta) = \max_L \sum q(e, \theta) f(L, \theta) - wL + a \equiv \pi(e) + a$$

$$\sum q(e, \theta) u(y(\theta)) - c(e) \geq \sum q(e', \theta) u(y(\theta)) - c(e') \ \forall e'.$$

It is enough to show that $\partial^2 G / \partial e \partial a < 0$ in order to guarantee that e_E is nonincreasing in a. From the envelope theorem and Lemma 2, $\partial G / \partial a = 1 / \sum \frac{q(e,\theta)}{u'(y(\theta))}$. Thus $\partial^2 G / \partial e \partial a < 0$ if and only if $\sum \frac{q(e,\theta)}{u'(y(\theta))}$ is increasing in e. From the envelope theorem, the fact that $f(L, \cdot)$ is increasing, and the definition of stochastic dominance, $\pi(\cdot)$ is increasing in e. Now, with $u(y) = \ln y$, $\sum \frac{q(e,\theta)}{u'(y(\theta))} = \sum q(e, \theta) y(\theta) = \pi(e) + a$ which is increasing in e, as desired.

References

[1] D. Ackerberg, M. Botticini, Endogenous matching and the empirical determinants of contract form, J. Polit. Economy 110,3 (2002), 564-591.

[2] A.V. Banerjee, A.F. Newman, Occupational choice and the process of development, J. Polit. Economy 101,2 (1993), 274-298.

[3] ______, Poverty, incentives, and development, Amer. Econ. Rev. 84,2 (1994), 211-215.

[4] P. Bolton, M. Dewatripont, Contract Theory, MIT Press, Cambridge, MA, 2005.

[5] R. Cantillon, Essay on the Nature of Commerce, tr. H. Higgs, Macmillan, London, 1755.

[6] A.K. Dixit Trade and insurance with imperfectly observed outcomes, Quart. J. Econ. 104,1 (1989),195-203.

[7] _______Trade and insurance with adverse selection, Rev. Econ. Stud. 56,2 (1989b), 235-247.

[8] S. Grossman, O.D. Hart, An analysis of the principal-agent problem, Econometrica 51,1 (1983) 7-45.

[9] B. Holmström, Moral hazard and observability, Bell J. Econ. 10 (1979), 74-91.

[10] I.D. Jewitt, Justifying the first-order approach to principal-agent problems, Econometrica 56,5 (1988), 1177-90.

[11] R. Kanbur, Of risk-taking and the personal distribution of income, J. Polit. Economy 87,4 (1979), 769-797.

[12] R. Kihlstrom, J.-J. Laffont, A general equilibrium entrepreneurial theory of firm formation based on risk aversion, J. Polit. Economy 87,4 (1979), 719-748.

[13] S.K. Kim, Efficiency of an information system in an agency model, Econometrica 63,1 (1995), 89-102.

[14] F. Knight, Risk, Uncertainty, and Profit, Houghton-Mifflin, Boston, 1921.

[15] J.-J. Laffont, D. Martimort, The Theory of Incentives: The Principal-Agent Model, Princeton Univ. Press, Princeton, NJ, 2001.

[16] E. Lazear, Entrepreneurship,J. Lab. Econ. 23,4 (2005), 649-680.

[17] R.E. Lucas, Jr., On the size distribution of business firms, Bell J. of Econ. 9,2 (1978), 508-523.

[18] B. MacLeod, D. Parent, Job characteristics and the form of compensation, Res. Lab. Econ. 18 (1999), 177-242.

[19] J. Morduch, Poverty and vulnerability, Amer. Econ. Rev. 84,2 (1994), 221-225.

[20] D. Mookherjee, I. Png, Optimal auditing, insurance, and redistribution, Quart. J. Econ. 104,2 (1989), 399-415.

[21] A. Mas-Colell, M.D. Whinston, J.R. Green, Microeconomic Theory, Oxford University Press, New York, 1995.

[22] C. Prendergast, The tenuous tradeoff between risk sharing and incentives, J. Polit. Economy 110,5 (2002), 1071-1102.

[23] W. Rogerson, Repeated moral hazard, Econometrica 53,1 (1985), 69-76.

[24] M. Rothschild, J. Stiglitz, Equilibrium in competitive insurance markets: an essay on the economics of imperfect information, Quart. J. Econ. 90,4 (1976), 629-650.

[25] R. Townsend, Optimal contracts and competitive markets with costly state verification, J. Econ. Theory 22 (1979), 265-293.

[26] X.B. Wu and A.M. Knott, Entrepreneurial risk and market entry, Small Business Administration Office of Advocacy Working Paper no. 249, 2005.