Discussion Paper No. 951

THE CAPITAL MARKET, THE WEALTH DISTRIBUTION AND THE EMPLOYMENT RELATION

by

Andrew F. Newman

August, 1991

Department of Economics
Northwestern University
2003 Sheridan Road
Evanston, IL, U.S.A. 60208

ABSTRACT

This paper offers an account of the occupational choice among wage work, self-employment and entrepreneurship which contrasts with the "Knightian" one based on risk attitudes. As shown by example, the latter can lead to perverse results. We propose a model in which imperfect capital markets arising from costly output verification cause the cost of capital to decline with an agent's wealth. Employment contracts, which require costly labor monitoring are then viewed as a substitute for financial contracts. The prevalence of employment contracts (as distinguished from self-employment) then depends on (1) how effective is the labor monitoring technology as a substitute for output verification in the capital market; and (2) how unequal is the distribution of wealth.

I am grateful to A. Banerjee, P. Bolton, A. Mas-Colell, E. Maskin, S. Matthews, A. Sen, Y. Spiegel, K. Spier and A. Wolinsky for helpful discussion and comments.

THE CAPITAL MARKET, THE WEALTH DISTRIBUTION
AND THE EMPLOYMENT RELATION

1. Introduction

Given technology, preferences and information, which manner of organizing production will survive in a competitive environment? Is hierarchy with monitors and employees a necessary outcome of efficient contracting, or are there conditions under which other organizational forms such as partnerships might be viable? If agents can freely enter into the different occupations (entrepreneur, manager, employee, partner, etc.) that each type of organization entails, then won't the factors which determine occupational choice also affect the competitive selection of organizational forms?

One set of answers to these questions goes back to Frank Knight (1921): individuals sort themselves between the safety of employment and the hazards of entrepreneurship according to their attitudes toward risk. This approach has been pursued by Kihlstrom and Laffont (1979) and Kanbur (1979). In it the firm is viewed as an organization for providing insurance: workers receive a sure wage, while entrepreneurs monitor them and bear the risks of production.[1] People vary in their risk aversion, either with a utility parameter or with wealth. Then the more risk averse (and, with declining absolute risk aversion, the poor) work for the less risk averse (rich), and the wage adjusts to make a marginal agent indifferent between the two occupations. At the very least, these versions of the theory explain the strong empirical association between wealth and occupation.

However, a problem with this theory is its implicit assumption that the risks are exogenous. On the one hand, firms are supposed to provide insurance to workers despite the costly incentive effects this has; on the other, entrepreneurs themselves are excluded from insuring each other. When instead entrepreneurs are permitted to insure, the Knightian story may lead to implausible results. The very least that should be required of a reasonable theory of the employment relation is that it predict that the rich hire the

---

[1]We presume that a worker's effort is not publicly observable; otherwise it is not clear why a firm is the appropriate institution for the provision of income insurance.

poor and not the other way around. But as we show in Section 3 by way of an example, when entrepreneurs are able to partially insure,[2] then even with declining absolute risk aversion, an equilibrium of the economy may entail that the agents who become workers are rich, while the employers who hire them are poor! The reason is straightforward enough: incentive compatibility requires that the amount of risk an employer bears increases with wealth. In the example, risk increases "too quickly," swamping the effect of decreasing risk aversion. Thus, the rich, not the poor, find it optimal to accept the safe wage contract.

This paper then takes a different approach to the question of firm formation, viewing the employment contract essentially as a substitute for an imperfect capital market. The basic idea is very simple: if transacting in the capital market is costly, those who don't already have capital may get together with those who do by working for them instead of borrowing from them, even though this entails costs of its own.

In the model presented in Sections 4 and 5, agency costs prevail in both the capital and labor markets. It is costly for an employer to monitor labor input and for a lender to audit a firm's output. The prevalence of employment contracts relative to self-employment arises from a trade-off between these two types of cost. If auditing is cheap, it may be optimal to avoid the costs of monitoring by letting the producer be the residual claimant, that is, self-employed. But if auditing costs are high, it may be better to avoid using the capital market and bear the costs of input monitoring, paying the producer a performance wage, so that he becomes an employee.

This story is very much in the Coasian spirit, where we identify monitoring of input as occurring *inside* the firm, while auditing of output is associated with transacting *outside*. Taking the classical approach of identifying firm ownership with residual claimancy on output,[3] we are

---

[2]Moral hazard prevents full insurance in the present model. Note that with full insurance, the occupational choice would be indeterminate since workers and entrepreneurs would have to receive the same compensation (agents differ only in wealth).

[3]An alternative approach, appropriate for an incomplete contract setting, identifies ownership with residual rights of control and/or the right of exclusion from the use of assets (Grossman and Hart, 1986; Hart and Moore, 1990); it is in a similar sense that Simon (1951) uses the term "employment relation." While it is certainly desirable to extend the results of the present paper to that environment, we begin with the complete contracts/return

asserting that each of two types of firm — self-employment and employment-contract — may be optimal, depending on the trade-off of the two types of agency costs. What we should like to know is, how does a competitive economy make this trade-off? Under what conditions will the employment contract predominate?

To answer these questions we will investigate a simple general equilibrium model of labor and capital markets in which agents trading in competitive markets freely enter different "occupations" — wage work, self-employment or entrepreneurship.[4] Agents are risk-neutral in end-of-period income and identical except for their endowments of nonlabor wealth; each possesses a unit of indivisible labor or effort. The economy has both a safe asset and a risky investment project which yields a high expected return if an agent puts effort into it. There is also a monitoring technology which permits one entrepreneur to observe the effort levels of a fixed number μ of workers (for simplicity, it is not possible to monitor monitors) working on μ projects.[5] Under this monitoring, projects aggregate essentially with constant returns to scale: if r is the (random) output of a single project, the output of an aggregated one is μr.

In the capital market, the key proposition is that agency costs decline with wealth. To model this idea, we use a particularly simple version of the Townsend (1979) costly-state-verification model, similar to the variant proposed by Bernanke and Gertler (1989). In this setting, income is constrained to be nonnegative, and by risk-neutrality, agents' utility is bounded below. Since the outcome of an investment project is not costlessly observable to outside investors, (random) audits are required to insure that borrowers do not falsely claim failure. All else the same, poorer agents have

---

streams approach because it is simpler. It appears that the results in this paper can be extended to the incomplete contracts setting, but anything definitive must await further research.

[4]We follow Kihlstrom and Laffont (1979) in using the term "entrepreneur" to mean someone who spends her effort monitoring others' labor inputs and who is a residual claimant to their output; "owner-manager" would be an alternative designation.

[5]It should be remembered that what we are calling a monitoring "technology" is as much social and organizational as it is purely technological. A similar interpretation might apply to the auditing technology.

more to gain by misreporting (given that the project has succeeded, they must repay a larger amount to the creditors), so they face higher audit probabilities which in turn raises their effective cost of capital relative to that of wealthier agents. Thus the poor have less incentive to transact in the capital market than do the wealthy and will tend to select employment contracts instead.[6]

Agents choose the occupation which gives the highest expected utility, given their wealth, the prevailing wage and the expected project returns (which also depend on wealth, as outlined above). Wages are determined competitively, i.e. by equating the number of agents who find wage work to be their best option to μ times the number who wish to be entrepreneurs. An employment contract equilibrium (ECE) obtains when a positive fraction of the population is engaged in entrepreneurship or wage work. A self-employment equilibrium (SEE) occurs when instead everyone works for himself.[7]

What is perhaps surprising is that under this formulation, the prevalence of the employment relation does not depend only on the technical terms of the trade-off between auditing and monitoring (i.e. the cost per audit or per monitor). Rather, it also depends on something which has not received much attention in the literature: inequality in the distribution of wealth. Since we have a constant-returns production technology, potentially everyone could be self-employed without any efficiency loss. Nevertheless, aggregated production can arise because of the trade-off between auditing and monitoring costs. Because the frequency of auditing declines with an agent's wealth, the trade-off operates most strongly between agents who differ most in wealth. Consequently, there must be a certain degree of inequality in order for any employment contracts to appear in equilibrium. Otherwise, the lowest wage at which people are willing to become workers will be too high for anyone to want

---

[6]As this paper was being completed, my attention was drawn to Eswaran and Kotwal (1989), who study the effects of standard debt contracts on occupational decisions. Their focus is on partial equilibrium and therefore differs from that of the present paper. Roemer (1982) considers a general equilibrium framework but is not explicit about the role of distribution; moreover, by ruling out credit markets altogether, he obtains employment contracts by default, as it were.

[7]Under the constant-returns-to-scale aggregation of projects assumed here, one could just as well interpret the self-employment equilibrium as a partnership of arbitrary size.

to hire them. These considerations permit us to derive necessary and sufficient conditions on the support of the wealth distribution for ECE to exist in our model.

This dependence of the prevalence of firm type on the wealth distribution holds much more generally than in the simple model studied here.[8] All that is necessary for our results to hold is that the cost of capital decline as wealth increases. Such a relation between agency costs and wealth is likely to arise whenever there are lower bounds on the utility that an agent can attain, as for instance when limited liability constraints are in force. Thus, instead of assuming costly state verification, one might emphasize the role of moral hazard in project choice and adverse selection of borrower ability, as is done in Bernanke-Gertler (1990). Another approach depends on the observation that it is difficult to enforce a financial contract insofar as enforcement depends on the ability to seize assets. An agent who borrows significantly more than his initial wealth may have little incentive to repay, if the worst penalty for reneging is to lose that wealth. Lenders will then be reluctant to lend more than some multiple of an agents's wealth, with the consequence that poor agents can invest only in small or inefficient projects.[9]

There is a small empirical literature which suggests that the dependence on wealth of the returns on investment projects, particularly the cost of outside finance, helps to determine the separation of agents into entrepreneurs and workers. For instance, Evans and Jovanovic (1989) have estimated a (partial equilibrium) model of occupational choice based on liquidity constraints. Using a specification which assumes that agents can borrow up to a constant multiple of their own wealth, they find that in a sample of 1500 white males the probability of entering self-employment increases with net family assets. Moreover, because of liquidity constraints, wealthier men earn higher returns on their businesses because they are able to start up with more efficient capital levels.

Two other factors which might determine occupational choice and the

---

[8]Indeed, it would likely hold even in a Kihlstrom-Laffont type model, although the exact nature of the relation between the distribution and the existence of ECE is much harder to gauge.

[9]This approach is pursued in Kehoe-Levine (1990) and Banerjee-Newman (1991).

organization of the firm are entrepreneurial ability and increasing returns in production. Increasing returns is frequently invoked to justify some kind of aggregated production (e.g. Vassilakis, 1989) and in particular the need for input monitoring (Alchian and Demsetz, 1972). But an implication of the model in this paper is that increasing returns in production is neither necessary nor sufficient for aggregated production. On the one hand, we obtain aggregated production even without increasing returns. On the other, if we allow the level of monitoring to vary between 1 and μ, then the capital market imperfection, which makes the marginal cost of "external" capital exceed the marginal cost of internal capital, can yield a determinate size of the firm which need not bear any relation to its efficient scale (see also Banerjee, Newman and Qian, 1991).[10] Note as well that even if returns to scale do entail monitoring, the technology alone does not determine the *identity* of the monitor.

As for ability, this is surely important, and has been explored by several authors, including Lucas (1978) and Calvo and Wellisz (1979, 1980). Here we simply want to emphasize the role of capital market imperfections in the determination of occupational choice and firm structure; for this purpose including ability in the model would only complicate matters. Entrepreneurial ability also has the unfortunate property of being very difficult to specify a priori; thus it is all too easy to infer that the larger firm, or the one with the monitor, must be the one belonging to the more able person.

In the next section we lay out the basic model by describing the agents' preferences and the technologies and occupations available to them. Section 3 follows with a consideration of the "Knightian" story as it applies to this model, showing how the story is subject to certain perversities. Section 4 begins the presentation of the imperfect-capital-market alternative by discussing the market for outside finance and deriving a value function for each occupation. Section 5 then analyzes the equilibrium for the whole economy by comparing the value functions and develops the above arguments concerning the roles of monitoring, the capital market and distribution. Section 6 concludes with a discussion of robustness and extensions.

---

[10] Since the monitoring technology is indivisible, it actually displays increasing returns if we assume that entrepreneurs may hire fewer than μ workers. Yet as we show in the Appendix, many entrepreneurs will choose to hire fewer than μ workers even though hiring μ of them is more efficient.

2. The Model

(a) *Preferences and Demographics*

The economy lasts one period and contains a large number of agents (indexed by the unit interval with Lebesgue measure) with identical preferences defined over income and effort. Agents are identically endowed with $\bar{e}$ units of effort, but differ in their wealth endowment ω. An agent's preferences may be summarized by the von Neumann-Morgenstern expected utility

$$E\ \{u(y) - e\},$$

where $y \geq 0$ is the realized lifetime income which depends, among other things, on her occupation, and e is the effort level chosen. Effort enters linearly only as a normalization, as will be evident when we introduce the production technology.

The income utility will have two forms in this paper. In the next section, where risk aversion is supposed to determine the occupational choice, we will set $u(y) = \ln y$. Elsewhere, we will not be concerned with risk effects of wealth on occupational choice, and we will assume instead that agents are risk-neutral: $u(y) = y$.

(b) *Technology*

The economy has a single storable consumption good which may also be used as capital. Agents' economic activity surrounds four technologies. First, there is a perfectly divisible safe asset (such as storage) which earns a gross return $\hat{r} > 0$.

Second, there is a risky investment project which comes in discrete units with capital requirement k; once sunk, this capital cannot be recovered. If the agent puts her unit of effort into the project, it succeeds with probability q, yielding the gross return R, and fails, yielding 0, with probability 1-q. With anything less than full effort, however, the project outcome is zero with certainty. Project returns are independent across agents. To avoid trivialities, we assume that the project is productive in the sense that its expected profit exceeds the disutility of effort, or

$$k\bar{r} - \bar{e} > k\hat{r}, \tag{2.1}$$

where the expected return $\bar{r} = qR$. The agent undertaking the project is the only one who learns its outcome for free.

Effort is not directly observable. The third technology allows it to be perfectly monitored. If an agent expends her own unit of effort, she observes the effort level of $\mu > 1$ other agents working on projects as well as the outcomes of those projects; if she shirks, however, the monitored agents know this and therefore have an incentive to shirk as well. The monitoring technology permits aggregation of the investment project. By monitoring μ agents, the agent can run a project of size μk, yielding μkr, where r has exactly the same distribution as for a single project: production is subject to constant return to scale.[11] Notice that we have implicitly assumed that the returns on the individual projects under one monitor are perfectly correlated. If one thinks of the monitor as gathering the projects together under one roof, then whether the risk is interpreted as originating in demand uncertainty or physical noise in the production process, the correlation assumption seems to be most reasonable.[12] However, the aggregated project returns remain independent across monitors. We assume that it is not possible to monitor a monitor, so that only one level of hierarchy is technically feasible.

Finally, the fourth technology permits verification of the outcome of the project by an "outside" party (i.e., someone other than the agent undertaking the project). Specifically, it costs γ per unit of capital to "count" the output and learn whether the project succeeded. Thus, it costs $\mu\gamma k$ to learn the outcome of an aggregated project.

(c) *The Occupations*

Given these technologies, there are four possible occupations. An agent can invest all of its wealth in the safe asset (or lend to other agents — these will yield the same return) and expend no effort. This will be called

---

[11]Insofar as an agent may monitor fewer than μ other agents with her unit of effort, the monitoring technology displays increasing returns over a range. For the most part, we will ignore this possibility and assume that monitors always monitor the full complement of μ agents. In any case, this should be distinguished from increasing returns in the production technology proper. Moreover, if we allowed monitoring to be divisible, so that it too displays constant returns, we would not greatly affect our results.

[12]Early manufactories, for example, often consisted of little more than a large room in which several workers would attend to machines identical to those which they had formerly had in their own homes under the putting-out system. The superiority of the manufactory as a device for monitoring the workers' activities should be obvious.

the *subsistence* option U, and it yields income $y_U(\omega) = \omega\hat{r}$. In some circumstances, equilibrium entails that no one occupy U, but it always provides an individual rationality constraint for the other occupations.

The second occupation is employment for a safe, competitive wage v. An agent who chooses to become a *worker* (W) expends his effort under the supervision of a monitor. Under these circumstances an employment contract written contingent on effort is feasible, and accordingly we shall define an employment contract to be an agreement to deliver a unit of effort in exchange for the wage v, regardless of the project outcome or actions of the monitor, with zero (which is his output) paid if less than full effort is delivered by the worker. The agent who chooses W also invests his wealth in the safe asset and therefore has income

$$y_W(\omega,v) = \omega\hat{r} + v,$$

since he will always choose to set $e = \bar{e}$ as long as his monitor does the same (otherwise he might as well take the subsistence option).

The last two occupations are *self-employment* S, in which one works directly on a project without being monitored; and *entrepreneurship* E in which the agents expends his effort on monitoring. In each case the income depends on the enterprise's output; the difference between the occupation lies not only in how the effort is spent, but also in the return streams and liabilities accruing to the agent. In particular, while the self-employed need only advance the capital requirement k toward the project, entrepreneurs must provide $\mu(k + v/\hat{r})$. The reason for the presence of the v term is that the entrepreneur must be able to guarantee payment of wages even in case of low output, since by assumption wages are safe; consequently, the present value of the wage bill must be laid out as part of the capital input.

The income accruing to the self-employed or entrepreneur will be denoted by the quintuple (y_R,y_0,y_a,y_l,p), although it is to be understood that the values these variables assume depend on which occupation the agent chooses. The first two elements describe the income accruing in case of high and low output respectively. Since in general the self-employed or entrepreneur will share some of the output with creditors, it may be necessary for those agents to engage in costly auditing of the firm's output: y_a is the income for the firm's owner if he is audited and found to be telling the truth; y_l is the audit income if he is caught misreporting output. Finally, p is the probability that the audit occurs at all.[13] The precise expressions for these variables, which will represent the optimal financial contract, are derived separately below, first with risk aversion and costless output verification (Section 3), and second with risk neutrality and costly output verification (Section 4).

3. Perverse Behavior with Endogenous Risk Bearing

The "Knightian" explanation for the prevalence of wage contracts and occupational choice depends on self-selection based on attitudes toward risk. With declining absolute risk aversion we should expect the poorer agents to accept a safe competitive wage while the wealthy hire them and bear the risks of entrepreneurship. In this section we will verify that this is true *provided the risks are exogenous.* As soon as we endogenize the risks with a simple, albeit imperfect, insurance scheme, however, we shall find that just the opposite can happen: although the wealthy are less risk averse than the poor, an incentive compatible insurance scheme may require them to bear so much risk that they actually prefer safe wages to entrepreneurship. Consequently, the poor will hire the rich!

Consider the following version of the model to allow for risk aversion and insurance. We simplify the example by assuming, as in Kihlstrom-Laffont (1979), that there are only two occupations, namely worker and entrepreneur. The production and monitoring technologies are as described above, but there is no cost to state verification ($\gamma = 0$). Preferences have the form $E(\ln y - e)$, where for entrepreneurs $y = y_R$ with probability q and y_a with probability $1 - q$; for workers, $y_W = \omega\hat{r} + v$. (Since auditing is costless, p is optimally set equal to unity, y_1 to zero, and y_0 is irrelevant.)

(a) *No Insurance for Entrepreneurs*

In this case, an entrepreneur starting with wealth ω puts $\mu(k + v/\hat{r})$ into the project, invests the remainder at $\hat{r}$ and receives μkr at the end of the

---

[13]In general, one would allow auditing probabilities to depend on the state reported. With only two states, however, it is only necessary to audit in case of a single report, namely the low one. Note as well that we restrict our attention to deterministic payments, since with risk neutrality nothing is gained by further randomization.

period. Income is therefore

$$[\omega - \mu(k + v/\hat{r})]\hat{r} + \mu kr$$

so that

$$y_R = \omega\hat{r} + \mu k(R-\hat{r}) - \mu v$$
$$y_a = \omega\hat{r} - \mu k\hat{r} - \mu v$$

and the entrepreneur's indirect utility ("value") is

$$V_E(\omega,v) = q\ln(\omega\hat{r} + \mu k(R-\hat{r}) - \mu v) + (1-q)\ln(\omega\hat{r} - \mu k\hat{r} - \mu v) - \bar{e},$$

while a worker's value function is

$$V_W(\omega,v) = \ln(\omega\hat{r} + v) - \bar{e}.$$

Note that entrepreneurs must have wealth of at least $\mu(k + v/\hat{r})$ in order for their utility to be defined.

Each agent chooses the occupation which, given his wealth and the prevailing wage, affords him the higher utility. Competitive employment contract equilibrium (ECE) requires that the supply of workers equals the demand, i.e. that the number (measure) of workers equal μ times the number of entrepreneurs. Since the value functions are smooth, there will be, for each wage v, a wealth level S(v) at which the agent is indifferent between wage work and entrepreneurship. In particular, at an equilibrium v^* we have

$$V_W(S(v^*),v^*) = V_E(S(v^*),v^*).$$

For this economy, it is the agents with lower wealth who choose wage work:

> <u>Proposition 1</u> In the ECE without insurance, there is a wealth level ω^* such that all agents with $\omega < \omega^*$ choose wage work and all agents with $\omega > \omega^*$ are entrepreneurs.

To see this, simply differentiate the two value functions with respect to wealth. It follows that $\partial V_E(\omega,v)/\partial\omega > \partial V_W(\omega,v)/\partial\omega$ whenever $V_W(\omega,v) = V_E(\omega,v)$. Uniqueness of the switch point $\omega^* = S(v^*)$ follows from this argument as well.[14]

(b) *Risk-sharing among Entrepreneurs*

Proposition 1 essentially gives the result in Kihlstrom-Laffont, with wealth acting as the index of risk aversion, and it generates the occupational

---

[14]One obtains $y_W = y_R^q y_a^{1-q}$ in equilibrium; comparing the slopes of the values and substituting gives $y_R^{-q} y_a^{q-1} < qy_R^{-1} + (1-q)y_a^{-1}$ or $(y_a/y_R)^q < q(y_a/y_R) + 1-q$, which always holds if y_a/y_R is positive and not equal to one.

choice that accords with the stylized facts. The story changes considerably, however, as soon as entrepreneurs are allowed to share risks.[15]

The optimal contract (y_R, y_a) for an entrepreneur with wealth ω solves (again $p = 1$, $y_1 = 0$, y_0 is ignored)

$$\max \quad q\ln y_R + (1-q)\ln y_a - \bar{e} \tag{3.1}$$

$$\text{s.t.} \quad q[\omega\hat{r} + \mu kR - y_R] + (1-q)[\omega\hat{r} - \mu k\hat{r} - y_a] \geq C\hat{r}, \tag{3.2}$$

$$q\ln y_R + (1-q)\ln y_a - \bar{e} \geq \ln y_a \tag{3.3}$$

where we have introduced the notation C to denote the project's capital requirement.

The first constraint says that the insurers must make nonnegative profit (total income in each state is $(\omega-C)\hat{r}$ plus the outcome of the project, which is then shared between the insurers and the entrepreneur). It must bind at an optimum, since if it did not, free entry of insurers would guarantee that the entrepreneur could do better by raising y_R.

The second constraint is a no-shirking condition for the entrepreneur: the left-hand side is the utility when he works and the right hand side the utility when he shirks, since in that case the project fails for sure. (Here, as elsewhere, the agent will only wish to set $e = \bar{e}$ or $e = 0$.) At an optimum, this constraint must also bind, since if it did not the entrepreneur could feasibly improve his lot by raising y_a and lowering y_R to keep (3.2) satisfied (this would just be a mean-preserving risk reduction).

Making (3.2) and (3.3) into equalities and substituting yields

$$y_R = Ay_a \tag{3.4}$$

$$y_a = \frac{\omega\hat{r}+\mu[k(\bar{r}-\hat{r})-v]}{Aq-q+1}, \tag{3.5}$$

where $A \equiv \exp(\bar{e}/q) > 1$. Note that a solution can be found for any positive wealth provided $v \leq k(\bar{r}-\hat{r})$, which condition must be satisfied in equilibrium. Observe that $y_R - y_a$ is a measure of the amount of risk an entrepreneur bears under this (imperfect) insurance scheme. Since y_a is increasing in ω, so is the amount of risk borne. As an agent becomes wealthier, her utility function flattens, so that it becomes necessary to increase the difference between the high income from working and the low income from shirking in order to keep the difference in utility equal to $\bar{e}$. In fact, in this example, the amount of

---

[15]In the extreme case in which entrepreneurs could insure perfectly, of course, the returns to wage work and entrepreneurship must be identical for everyone and the occupational choice is indeterminate.

risk borne increases "too quickly" relative to the rate at which risk aversion declines.[16]

Once again, in order to have an ECE, it is necessary that at some v^* and $\omega^* = S(v^*)$ we have $V_W(\omega^*,v^*) = V_E(\omega^*,v^*)$ since both value functions are smooth. Substituting (3.4) and (3.5) into the entrepreneur's objective function allows us to rewrite this equation as

$$\ln(\omega^*\hat{r}+v^*) = q\ln A\left(\frac{\omega^*\hat{r}+\mu[k(\bar{r}-\hat{r})-v^*]}{Aq-q+1}\right) + (1-q)\ln\left(\frac{\omega^*\hat{r}+\mu[k(\bar{r}-\hat{r})-v^*]}{Aq-q+1}\right).$$

Simplifying this expression, we obtain

$$\omega^*\hat{r}+v^* = A^q\left(\frac{\omega^*\hat{r}+\mu[k(\bar{r}-\hat{r})-v^*]}{Aq-q+1}\right).$$

Observe now that for $A > 1$ and $0 < q < 1$, we have $0 < A^q < Aq - q + 1$. Thus,

$$\omega^*\hat{r} + v^* < \omega^*\hat{r} + \mu[k(\bar{r}-\hat{r})-v^*] \tag{3.6}$$

On the other hand, the slopes of the value functions are given by

$$\frac{\partial V_E(\omega^*,v^*)}{\partial\omega} = \frac{\hat{r}}{\omega^*\hat{r}+\mu[k(\bar{r}-\hat{r})-v^*]},$$

$$\frac{\partial V_W(\omega^*,v^*)}{\partial\omega} = \frac{\hat{r}}{\omega^*\hat{r}+v^*}.$$

Thus (3.6) implies $\partial V_E(\omega^*,v^*)/\partial\omega < \partial V_W(\omega^*,v^*)/\partial\omega$, and therefore we have

<u>Proposition 2</u> In the ECE with insurance there is ω^* such that all agents with $\omega < \omega^*$ choose entrepreneurship and all agents with $\omega > \omega^*$ are workers.

In other words, an ECE of this economy, if it exists, must have the rich working for the poor!

There are other aspects of the Knightian model which seem to depart from the stylized facts. First, risk attitudes across individuals do not seem particularly correlated with either wealth or occupation. Purchasers of lottery tickets seem to be drawn disproportionately from among the less wealthy; many of the most dangerous jobs or those with the greatest cyclical variability appear to be occupied by the poor rather than the wealthy; after

---

[16]The result that the amount of risk borne increases with wealth depends only on the concavity of $u(\cdot)$, not on its logarithmic form.

all, it is the poor who toil in the mines (as workers, not entrepreneurs!). Yet the pure Knightian model would predict that these individuals, who evidently are not terribly concerned with risk, should be entrepreneurs.

These observations raise a second difficulty, namely that it is not especially obvious that the occupation of wage worker is safer than that of entrepreneur. As mentioned, workers are frequently subject to on-the-job risks such as dangerous working conditions. And there are pecuniary risks as well: many workers receive some sort of bonus compensation; many more are subject to layoffs. There are as well good theoretical reasons for expecting employment contracts to be somewhat risky, at least if monitoring is not perfect. Either way, one of the basic assumptions of the Knightian story is undermined by riskiness in employment contracts.

In what follows, however, we will retain assumption that employment contract are safe; it is practically entailed in the supposition that the monitoring technology is perfect, since then a safe wage is an optimal contract. This characterization simplifies the analysis, and it should not affect the results. If an agent chooses employment over self-employment because he has costly access to capital, then the fact that one occupation is risky while the other is safe will not be particularly important. Moreover, if instead both occupations are considered to be risky, then the Knightian story is almost irrelevant while ours is largely unaffected.

The difficulty we encountered with the Knightian model is illustrated in Figure 1(a). Quite simply, the worker's value function cuts the entrepreneur's the wrong way — from below rather than above. In our example, a small increase in wealth benefits a worker because he can invest it at the safe return. But the same increase in wealth benefits an equally-endowed entrepreneur by less than this amount because the increased wealth means he must bear more risk in his firm.

In the alternative to the Knightian theory which we present here, we reverse this condition, making the value functions appear as in Figure 1(b). When capital market imperfections arise from enforcement problems — costly output verification being a special case — the costs involved will tend to decline the smaller is the amount borrowed. Thus small increases in wealth are worth *more* than the safe asset return because they result in a decrease in the cost of all of the entrepreneur's outside finance. The following section will be spent deriving a model in which the value functions have this property.

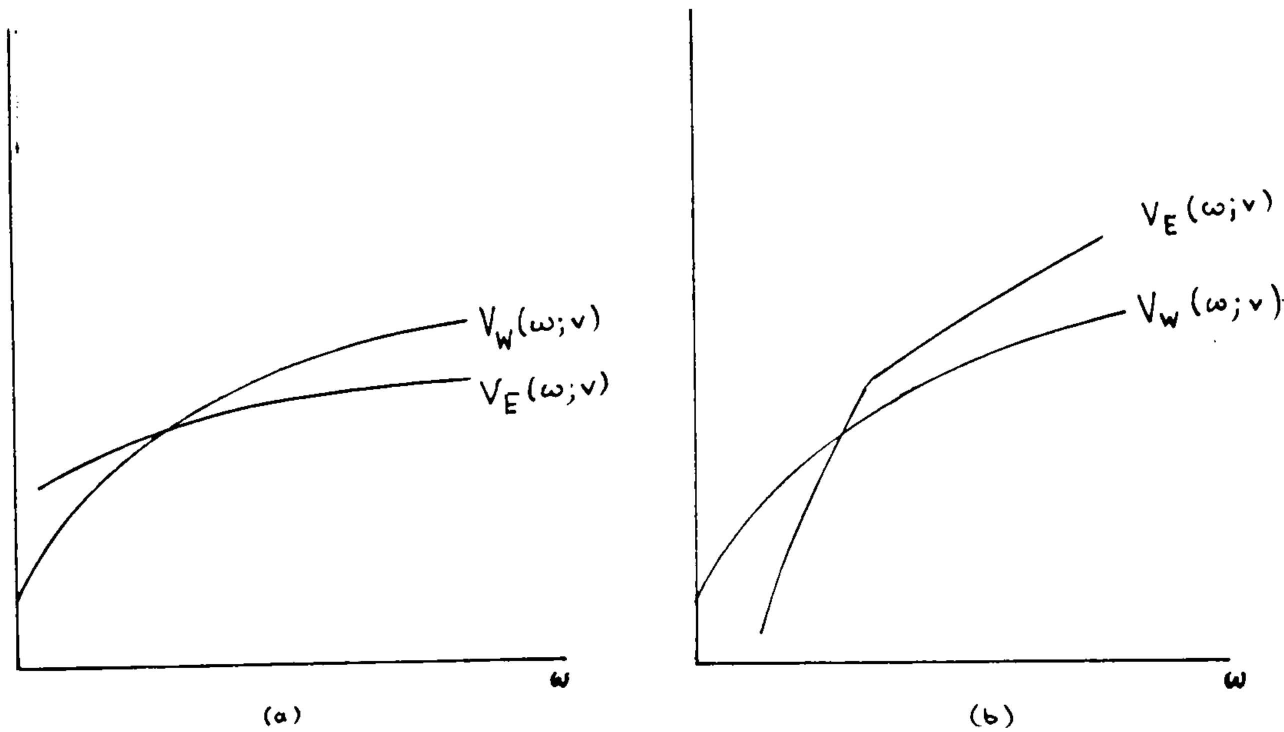


Figure 1. Relationship between entrepreneur's and worker's value functions.

4. The Financial Market with Costly Output Verification

We restore the assumption that output verification is costly to outside parties ($\gamma > 0$). However, since we are concerned with return rather than risk effects of wealth on occupational choice, we will simplify matters considerably by assuming that agents are risk neutral: $u(y) = y$. Since income is constrained to be nonnegative, this assumption implies that utility is bounded below. It is this property of the utility function, and not its curvature, that is crucial to our model.

The market for financial contracts is competitive with free entry. It is perhaps easiest to think of lenders as intermediaries acting on behalf of large numbers of individuals, so that lender's solvency is never an issue. As does Bernanke-Gertler (1989), we derive the optimal contract from the borrower's point of view — free entry assures that he can always shop around for such a contract — assuming for the moment that the borrower, be he self-employed or an entrepreneur, is committed to taking his project. In the next section, when we consider equilibrium, each agent will choose whether to undertake the project and will be in the occupation that, given his wealth, affords him the highest utility.

We continue to think of the financial contract as a sharing arrangement between creditor and borrower. However, since now only the borrower learns the project's outcome, he may have an incentive to misreport the output level; in order to induce truth-telling by the borrower, random audits are allowed. With probability p (determined as part of the contract)[17] the lender incurs the cost $K\gamma$ (where $K = k$ or μk depending on whether the borrower is self-employed or an entrepreneur) and verifies the return. It is perhaps best to think of the borrower as transferring wealth back to the lender: this way credible lies could occur only when output is high, since the inability to transfer wealth to the lender in case of low output would reveal that the project had failed. For this reason there will never be a need for auditing when the

---

[17] We are allowing random audits because they are generally more efficient than deterministic audits. Since increasing the efficiency of capital markets biases things against our theory, ruling out random audits would only strengthen our results concerning the dependence of capital costs on wealth (so would ruling out the implicit assumption that lenders can commit to the auditing probability p).

borrower reports that the project succeeded; p thus represents the audit probability in case failure is reported.

The optimal contract (y_R, y_0, y_a, y_1, p) for a borrower of wealth ω solves the following problem:

$$\max \quad qy_R + (1-q)(py_a + (1-p)y_0) - \bar{e} \tag{4.1}$$

$$\text{s.t.} \quad y_0 \geq 0,\ y_a \geq 0,\ y_1 \geq 0 \tag{4.2}$$

$$0 \leq p \leq 1 \tag{4.3}$$

$$y_R \geq (1-p)(y_0 + KR) + py_1 \tag{4.4}$$

$$q[\omega\hat{r} + KR - y_R] + (1-q)[\omega\hat{r} - p(y_a + K\gamma) - (1-p)y_0] \geq C\hat{r} \tag{4.5}$$

$$qy_R + (1-q)(py_a + (1-p)y_0) - \bar{e} \geq py_a + (1-p)y_0 \tag{4.6}$$

Except for (4.6) this problem is identical to Bernanke and Gertler's: the first two constraints require that income is always nonnegative and that p is a probability; (4.4) insures that it is in the borrower's interest to report the project outcome truthfully; and (4.5) is the individual rationality constraint for the creditor; it differs from (3.3) in that it has been modified to reflect the positive cost of auditing.

Inequality (4.6) is an additional incentive compatibility constraint necessitated by the costly observability of effort, the analog of the no-shirking constraint (3.2): the expected utility from working $(e = \bar{e})$ must be at least that from shirking $(e = 0)$. It turns out though, that this constraint will not bind under our assumptions: a solution to (4.1) subject only to (4.2) - (4.5) will automatically satisfy (4.6) for any agent who prefers self-employment or entrepreneurship to subsistence.

The solution to the optimal contract problem will allow us to derive the value functions for occupations S and E.

<u>Proposition 3</u> (a) For $\omega \leq C$, the optimal contract satisfies

$$y_a = y_1 = y_0 = 0,$$

$$y_R = (1-p)KR,$$

$$p = \min\left\{\frac{(C-\omega)\hat{r}}{qKR - (1-q)K\gamma},\ 1\right\};$$

(b) for $\omega > C$, an optimal contract satisfies

$$y_0 = (\omega - C)\hat{r},$$

$$y_R = (\omega - C)\hat{r} + KR,$$

$$p = 0.$$

The result (a) follows from the fact that constraints (4.4) and (4.5) bind at an optimum. If the first constraint did not bind, then it would be possible to lower the probability p, thereby saving on (average) auditing costs: lowering p would allow y_R to be increased, to the benefit of the borrower, without violating (4.5). By the same token, if (4.5) does not bind, the borrower can increase y_R, thereby increasing his utility without violating the other constraints. The nonnegativity constraints on y_a, y_0, and y_l all bind because this allows minimization of p and therefore of expected auditing costs. It is easy to check that the solution provided in part (b) satisfies the constraints: agents with wealth greater than C can always repay so they have no incentive to misrepresent project outcomes and need bear no auditing costs; these agents are indifferent between "borrowing" from themselves and from outside investors.

The important point to note from Proposition 3 is that for borrowers who require outside finance the probability of audit — and therefore the expected auditing costs and associated deadweight loss — is a declining function of their initial wealth. The reason is that for lower wealth levels, agents (borrowers) are nearer the lower bound on their utility and cannot be made to suffer large losses when caught misreporting; instead, the probability of loss (audit) has to be increased. But this implies that poorer agents bear higher auditing costs. What differentiates the rich from the poor with respect to the returns they receive is not simply that the rich have more money, but rather that they are far from the lower bound on utility.[18]

Substituting the formulas in Proposition 3 into the borrower's objective

---

[18] To see that it is bounded utility, not nonnegativity constraints per se, that result in agency costs which decline with wealth, observe that if we set $u(y) = \ln y$ and re-solve the financial contract problem (ignoring the no-shirking constraint), a contract yielding arbitrarily close to the first-best utility level, with correspondingly low probability of audit, can be found à la Mirrlees (1974) by setting y_l sufficiently close to zero.

gives the following value functions:

$$V_S(\omega) = \begin{cases} k\bar{r} + \alpha(\omega-k)\hat{r} - \bar{e}, & 0 \le \omega \le k \\ k\bar{r} + (\omega-k)\hat{r} - \bar{e}, & k \le \omega \end{cases}$$

$$V_E(\omega,v) = \begin{cases} \mu k\bar{r} + \alpha[(\omega-\mu k)\hat{r} - \mu v] - \bar{e}, & 0 \le \omega \le \mu(k+v/\hat{r}) \\ \mu k\bar{r} + (\omega-\mu k)\hat{r} - \mu v - \bar{e}, & \mu(k+v/\hat{r}) \le \omega \end{cases}$$

where $\alpha \equiv \dfrac{\bar{r}}{\bar{r}-(1-q)\gamma} > 1$ (if $\bar{r} \le (1-q)\gamma$ auditing costs are so high it is never efficient to borrow: subsistence would be preferable). Notice that these value functions automatically satisfy the no-shirking constraint (4.6) as long as the agent prefers S or E to subsistence: the right-hand side of (4.6) is zero, while subsistence always yields nonnegative utility. The value functions for all four occupations are illustrated for a particular wage v in Figure 2.

The value functions for W and U have constant slope equal to $\hat{r}$. But for S and E, the slope is higher for wealth levels below the capital requirement. This reflects the fact that the marginal value of wealth in this range exceeds $\hat{r}$; an extra unit of wealth not only yields the return from investing it in the safe asset, but also lowers the expected costs of audit on all outside finance. Due to agency costs in the capital market, therefore, we have increasing returns to wealth. The aggregated production entailed by the employment contract arises because of the increasing returns to wealth, and need not have anything to do with increasing returns to scale in production.

5. Equilibrium

Having derived the value functions for each of the occupations, we can now proceed to an analysis of equilibrium. An employment contract equilibrium will exist when a positive measure of agents choose to become workers at the going wage; this naturally entails that there is also a sufficient number of agents choosing entrepreneurship to hire them.

Formally, let $\omega(a)$ be the wealth of agent $a \in [0, 1]$, where $\omega(\cdot)$ is Lebesgue measurable. Let $\bar{A}_W(v)$ be the set of agents who prefer working to any other occupation, given that the wage is v:

$$\bar{A}_W(v) = \{a: V_W(\omega(a),v) = \max\{V_E(\omega(a),v), V_W(\omega(a),v), V_S(\omega(a)), V_U(\omega(a))\}\};$$

let $\underline{A}_W(v)$ be the set of wealth levels where this preference is strict. Define $\bar{A}_E(v)$ and $\underline{A}_E(v)$ similarly for entrepreneurs. The supply correspondence of

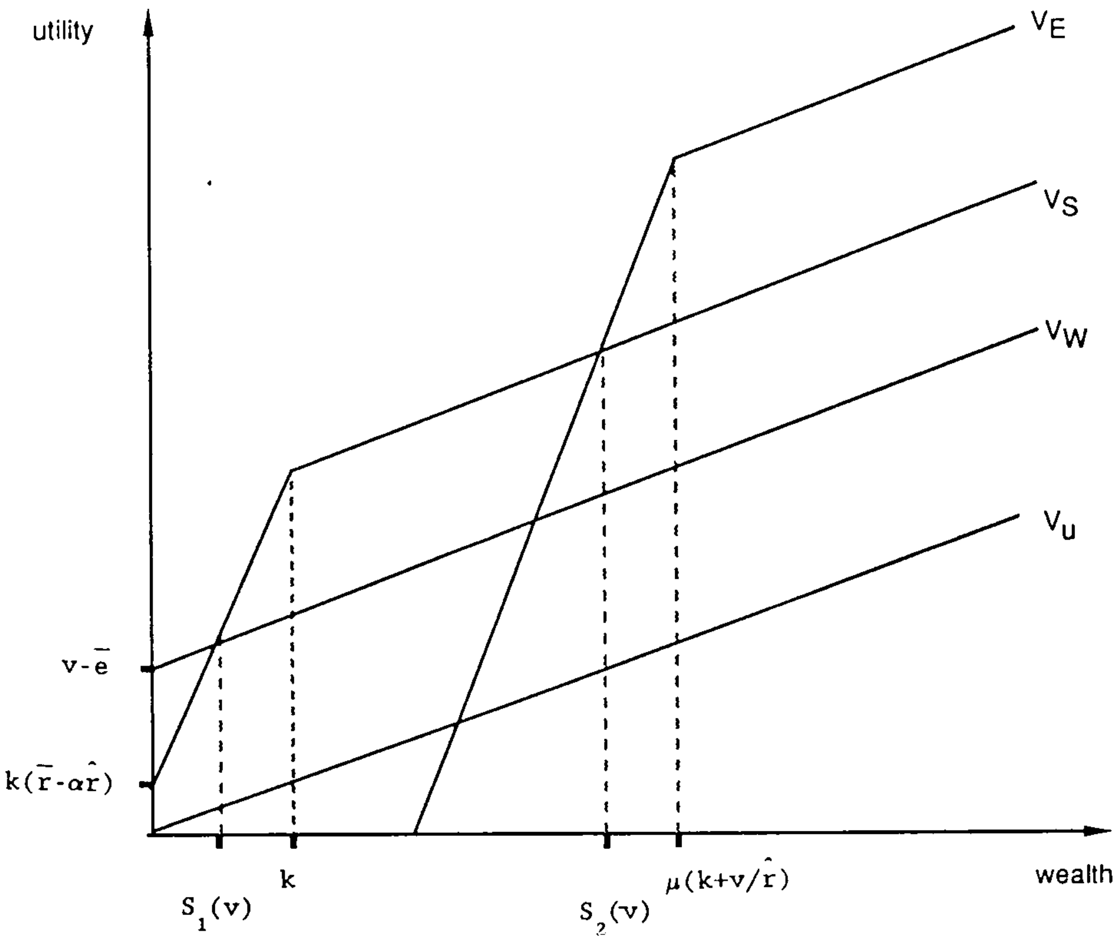


Figure 2. Value functions for the four occupations.

workers will then be an interval (possibly degenerate) with endpoints given by the Lebesgue measure $\Lambda(\cdot)$ of $\underline{A}_W(v)$ and $\bar{A}_W(v)$; the demand for workers is defined analogously using $\bar{A}_E(v)$ and $\underline{A}_E(v)$. Our economy will have an employment contract equilibrium if there is a wage $v > 0$ and disjoint sets $W(v)$ and $E(v)$ with $\underline{A}_W(v) \subseteq W(v) \subseteq \bar{A}_W(v)$ and $\underline{A}_E(v) \subseteq E(v) \subseteq \bar{A}_E(v)$ such that

$$\Lambda(W(v)) = \mu\Lambda(E(v)) > 0 \tag{5.1}$$

The proviso that the measures be positive must be added because it is perfectly possible that the equation is satisfied trivially with both sides equal to zero. This definition may appear cumbersome, but as will become clear following Proposition 4 below, it can be reduced to a rather simpler form.

An equilibrium with employment contracts need not always exist — this is one of the central points of this paper — and we shall spend this section developing necessary and sufficient conditions for its existence. As a rule, if an ECE does not exist, at least some agents will be self-employed (the rest may choose subsistence) and we shall call this alternative a self-employment equilibrium (SEE).

(a) *Paradox Lost*

First, however, we resolve the paradox of the Knightian theory developed in Section 3, and then simplify the definitions of ECE and of the alternative equilibria that can occur in this economy. Look again at Figure 2, and note that as the wage v increased, the V_W curve shifts up and the V_E curve shifts down and to the left, while V_S and V_U are unaffected. Since the "steep" sections of the V_S and V_E curves are parallel, there can only be one intersection between V_S and V_E, which we call $S_2(v)$ to denote its dependence on the wage.[19] Similarly, V_W and V_S can intersect at only one point which we denote $S_1(v)$. Thus if v is the going wage, all agents with wealth less than $S_1(v)$ will choose wage work, all agents with wealth greater than $S_2(v)$ will become entrepreneurs, and the rest will be self-employed. (For future reference, note that by equating the value functions, we obtain

[19]Strictly speaking, there is the possibility that V_S and V_E coincide above $\mu(k+v/\hat{r})$. It is straightforward to show that in an ECE, V_E can never lie with the steep portion coincident with or above the steep portion of V_S (the proof follows the lines of argument of that of Proposition 5).

$$S_1(v) = \frac{v-k(\bar{r}-\alpha\hat{r})}{(\alpha-1)\hat{r}} \tag{5.2}$$

and

$$S_2(v) = \frac{\alpha\mu v+k(\bar{r}-\hat{r})-\mu k(\bar{r}-\alpha\hat{r})}{(\alpha-1)\hat{r}} ; \tag{5.3}$$

note that both of these are increasing in v.)

The description must be modified slightly in case the wage is either so low that V_W and V_U coincide or so high that V_S and V_E coincide above $\mu(k+v/\hat{r})$. The first situation can arise only if $V_S(0) = k(\bar{r}-\alpha\hat{r}) - \bar{e} \leq 0$. Then v is at the minimum possible level $\underline{v}$ compatible with ECE, and some agents may choose subsistence rather than working. The other situation arises only if v is at its maximum level $\bar{v}$; some of the agents whose wealth exceeds $\mu(k+\bar{v}/r)$ choose self-employment, while the rest are entrepreneurs.[20]

In any event, we do have the following description of an ECE.

<u>Proposition</u> <u>4</u> In the ECE with costly output verification and risk neutrality, only the poorest agents in the economy become workers and only the richest become entrepreneurs.

This model would seem to accord better with the stylized facts than does our Knightian model of Section 3.

In light of this proposition, we can now provide a much simpler criterion for characterizing an employment contract equilibrium, at least if we assume that $H(\cdot)$, the cumulative wealth distribution function, is continuous. Then all we need is a $v \in (\underline{v}, \bar{v})$ such that

$$H(S_1(v)) = \mu[1 - H(S_2(v))] \tag{5.4}$$

$$H(S_1(v)) > 0 \tag{5.5}$$

The first condition is simply the market clearing equation we had before, while the second states that a positive fraction of the population are engaged in wage work. $H(S_1(v))$ therefore provides a natural measure of the prevalence

---

[20]By equating V_S and V_E at these high wealth levels, it is easy to check that $\bar{v} = \frac{\mu}{\mu-1} k(\bar{r}-\hat{r})$. For wages higher than this, V_E lies below V_S, and ECE is precluded. It is also easy to verify that $\underline{v} = \max\{\bar{e}, k(\bar{r}-\alpha\hat{r})\}$.

of employment contracts in the economy.[21]

(b) *Conditions for the Existence of an Employment Contract Equilibrium*

Once we have the sketch of a possible ECE configuration of value functions portrayed in Figure 2, our results on the roles of capital market imperfections, monitoring technology and wealth distribution in producing the prevalence of employment relation are easy to generate.

Necessary Conditions on the Costs of Auditing and Monitoring

As we suggested in the Introduction, employment contracts can be thought of as a substitute for costly financial contracts. If they are to be viable, they must not be too costly: some agents must prefer them to self-employment. Thus, the financial market needs some degree of inefficiency before it will be replaced by employment contracts. In the present model, this imposes a simple necessary condition on the relative costs of the auditing and monitoring technologies.

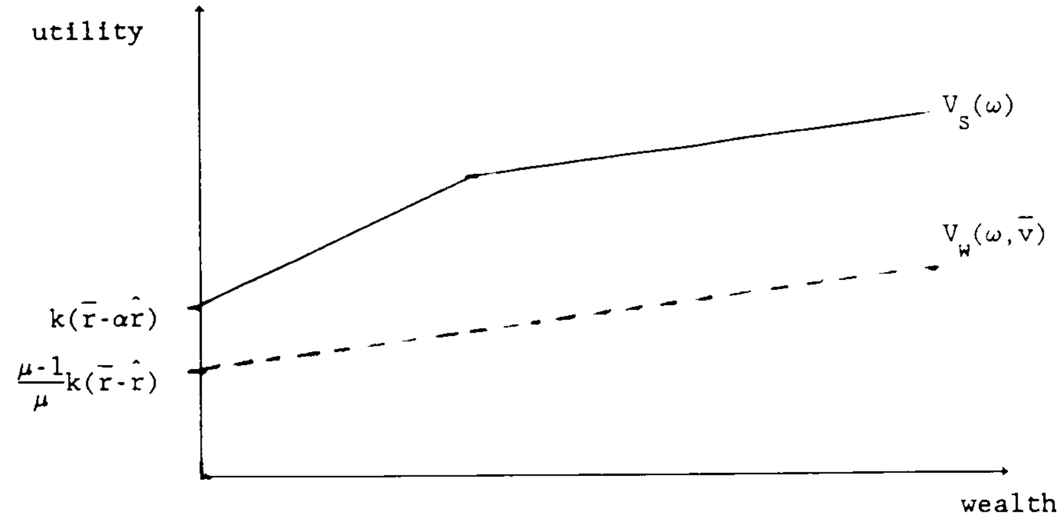
The severity of financial market inefficiency can be measured by the parameter γ: as it decreases, the deadweight loss from borrowing at a given wealth level falls. In the limit as γ approaches 0, the capital market is perfect. We can equally well use α as our measure of inefficiency, where $\gamma = 0$ implies $\alpha = 1$ and α increases with γ.

Figure 3 compares $V_s(\omega)$ and $\bar{v} = \frac{\mu-1}{\mu}k(\bar{r}-\hat{r})$. Since a necessary condition for ECE is that some agents prefer W to S, it is clear that such an equilibrium will be impossible if V_s lies everywhere above $V_W(\omega,\bar{v})$. This observation provides us with the following necessary relation between the quality of monitoring and cost of auditing.

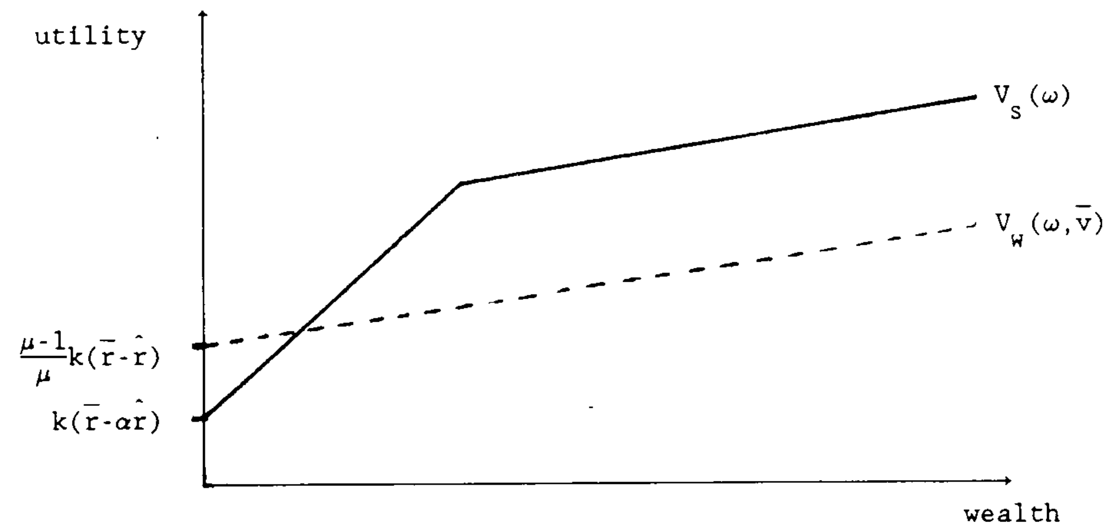
<u>Proposition 5</u> A necessary condition for ECE is that $V_S(0) \le V_W(0,\bar{v})$, i.e. that $\alpha-1 \ge \frac{\bar{r}-\hat{r}}{\mu\hat{r}}$, with strict inequality if the wealth distribution has no atom at 0. If this condition fails, then all agents will be

---

[21] We also have ECE if $H(S_1(\underline{v})) \ge \mu[1 - H(S_2(\underline{v}))] > 0$ (excess supply of workers) or $0 < H(S_1(\bar{v})) \le \mu[1 - H(S_2(\bar{v}))]$ (excess demand). Note that the appropriate measure of the prevalence of employment contracts in case of excess supply is $1-H(S_2(\underline{v}))$.



(a)



(b)

Figure 3. Relationship between quality of monitoring and cost of auditing (a) ECE precluded; (b) ECE possible.

self-employed.

The inequality illustrates the trade-off between the (purely technical components) of the two agency costs; as we should expect, the more costly the auditing, the less efficient the monitoring technology needs to be in order to make ECE possible. If effort monitoring is not a sufficiently good substitute for output auditing, then it will not be used, and our economy will be inhabited solely by agents who work for themselves. On the other hand, if monitoring is very efficient (μ is large), then even the smallest of capital market imperfections will make an ECE possible.

Recall we are being rather loose in our interpretation of the auditing and monitoring "technologies": they may have an important social and organizational dimension. In particular, we might expect that in communities with strong religious, cultural or family ties, output auditing — or credit arrangements generally — will be less costly. For instance certain East Asian immigrant groups to the United States are known to enjoy relatively low cost credit through rotating savings and credit associations. Our results suggest that these groups will also have a high rate of self-employment, as is indicated by the evidence in Light (1972) and Light and Bonacich (1988).

Note that as α approaches 1, the capital market becomes perfect; directly from Proposition 5 we therefore have a

Corollary 1 If the capital market is perfect, then an ECE is impossible; equilibrium entails that all agents are self-employed.

Note that with a perfect capital market, the equilibrium is first-best optimal, as self-employment is more efficient than the aggregated production under employment contracts.[22]

Distribution

The previous proposition provides a necessary condition for the existence of an ECE, but it is hardly sufficient. In this model, entrepreneurs have a

---

[22]With self-employment, $\mu+1$ agents produce an expected gross output of $(\mu+1)k\bar{r}$, while under entrepreneurship, those same agents produce only $\mu k\bar{r}$, since one is acting as monitor. The possible first-best properties of an SEE depend, of course, on risk neutrality.

comparative advantage at accessing capital because of their high wealth, so it pays for them to trade with (hire) workers provided the advantage is large enough. This will only be brought about by sufficient dispersion in the distribution of initial wealth.

For simplicity as well as empirical sensibility, let us restrict our attention to continuous distributions supported on intervals of the form $[\underline{\omega},\bar{\omega}]$. Two things are clear immediately. First, in order to have an ECE there must be a positive measure of the population with less than the highest wealth at which anyone would be a worker, $S_1(\bar{v})$, and a positive measure above $S_2(\underline{v})$, the lowest wealth at which anyone would be an entrepreneur; thus, $\underline{\omega} < S_1(\bar{v})$ and $S_2(\underline{v}) < \bar{\omega}$ (note, using (5.2), (5.3) and the definitions of $\bar{v}$ and $\underline{v}$ that $S_1(\bar{v}) < S_2(\underline{v})$). Second, the lowest wage at which someone is willing to work must be lower than the highest wage at which someone would be willing to hire him. Now since it is the poorest person in the economy who is willing to work at the lowest wage, and it is the richest person who will hire at the highest wage, this condition can be expressed by $S_1^{-1}(\underline{\omega}) < S_2^{-1}(\bar{\omega})$; the inequality is strict because with a continuous distribution, there must be an interval of wealth levels at which agents are willing to work ($S_1^{-1}(\cdot)$ and $S_2^{-1}(\cdot)$ are obtained by inverting expressions (5.2) and (5.3)). While these conditions are obviously necessary for ECE (it is straightforward to check that neither implies the other), it is also true that are sufficient:

<u>Proposition</u> <u>6</u> Let $H(\cdot)$ be continuous with support $[\underline{\omega},\bar{\omega}]$. Then an ECE exists if and only if

(1) $\underline{\omega} < S_1(\bar{v})$ and $S_2(\underline{v}) < \bar{\omega}$;

(2) $S_1^{-1}(\underline{\omega}) < S_2^{-1}(\bar{\omega})$.

The proof of sufficiency depends on examining the excess demand correspondence and verifying that conditions (1) and (2) guarantee the existence of an equilibrium wage v^* which clears the labor market with a positive measure of agents choosing to be workers. Details are provided in the Appendix.

One can rewrite condition (2) as

$$\bar{\omega} > \alpha\mu\underline{\omega} + k\mu(\bar{r}-\alpha\hat{r})/\hat{r} + k(\bar{r}-\hat{r})/((\alpha-1)\hat{r}),$$

thereby permitting us to examine the effects of parameter changes on the range of the wealth distribution necessary for an ECE. Specifically, if the capital market worsens (α increases), then given $\underline{\omega}$, $\bar{\omega}$ may decrease and we will still have an ECE ($\underline{\omega}$ must be less than k, since it must fall on the steep part of $V_S(\omega)$). With a less efficient capital market, smaller wealth differentials are needed to give the wealthy the comparative advantage in accessing capital required for employing the poor. As for changes in μ, there are two effects: greater efficiency will tend to raise the wage, so that given $\bar{\omega}$, $\underline{\omega}$ can increase; on the other hand, the capital requirement also increases with μ, so this tends to require that $\bar{\omega}$ rise for any given value of $\underline{\omega}$. It turns out that this second effect dominates, at least provided $\bar{r} - \alpha\hat{r}$ is positive, so the larger firm sizes associated with a less costly monitoring technology require a greater dispersion of wealth.[23]

The relation between conventional measures of inequality and of the prevalence, as distinct from existence, of employment contracts is more complicated, and although the relevant computations are straightforward, strong results concerning the monotonicity of the prevalence of employment contracts (as measured by $H(S_1(v^*))$ as a function of inequality do not appear to be available without considerable restrictions on the class of wealth distributions.[24] But we can provide bounds on the level of inequality necessary for a certain fraction, say ε, of the agents to be workers. For example, if we measure inequality by the variance and denote by $\hat{\omega}$ the mean wealth, then if the measure of workers is at least ε, the variance must exceed $\varepsilon(S_1(\bar{v})-\hat{\omega})^2 + (\varepsilon/\mu)(S_2(\underline{v})- \hat{\omega})^2$, which is increasing in ε. Otherwise put, the smaller the variance, the smaller can be the number of workers and entrepreneurs in the economy.

Proposition 6 also allows us to characterize a highly egalitarian economy.

Corollary 2 Suppose all agents have the same wealth $\hat{\omega} > 0$.

(a) If $k(\bar{r}-\alpha\hat{r}) - \bar{e} \geq 0$, then in equilibrium, all agents are self-employed.

---

[23]This is something of an artifact of the indivisibility of monitoring. If entrepreneurs can choose the size of the firm to be less than μ, the first effect will be more important: greater equality will be be compatible with ECE.

[24]Within the class of exponential distributions, for instance, it can be shown that a (small) mean-preserving increase in inequality always increases the number of workers and entrepreneurs at the expense of the self-employed. On the other hand, it is also possible to construct examples in which a mean-preserving spread of the initial distribution results in a decrease of the number of workers and entrepreneurs.

(b) Let $k(\bar{r}-\alpha\hat{r}) - \bar{e} < 0$.

(i) If $\hat{\omega} > \dfrac{\bar{e}-k(\bar{r}-\alpha\hat{r})}{(\alpha-1)\hat{r}}$, then all agents are self-employed in equilibrium;

(ii) if $\hat{\omega} < \dfrac{\bar{e}-k(\bar{r}-\alpha\hat{r})}{(\alpha-1)\hat{r}}$, then all agents choose subsistence.

In other words, with perfect equality, no matter what the wealth level shared by the agents, an employment contract equilibrium cannot exist. Whether all agents are self-employed or choose subsistence depends on the severity of the capital market imperfection. In part (a), even workers with zero wealth prefer self-employment to subsistence, since α is not terribly large. Part (b) states that if the capital market is so inefficient that $V_S(0) < V_U(0)$ and all agents are sufficiently poor (have wealth less than that at which V_S and V_E intersect), then they all choose subsistence; otherwise they all become self-employed. See Figure 4.

The results relating the wealth distribution to the existence of ECE finds an historical counterpart in Western Europe in the eighteenth century. Agriculture was dominated by the employment relation in Britain, where the enclosure movement and lack of a developed capital market resulted in concentrated ownership of land. In France, where land had been more equally distributed following the Revolution, peasants tended to remain self-employed. The analysis has been extended to try to account for France's supposed lag in industrial development (presumably the employment relation is good for growth and industrialization — see, for instance, North, [1981]), but that requires quite a separate argument.

6. Discussion

Insofar as the theory of the firm is meant to explain the predominance of particular organizational forms, the model presented here suggests that such a theory may not be complete without some consideration of its general equilibrium aspects. Specifically, we have shown that when two organizational forms compete in the market, the outcome cannot be predicted only by checking which form solves an optimal contract problem. Rather, it may be necessary to know economy-wide data before it will be possible to determine whether either form gets "priced out of the market." Thus, while it is a commonplace that institutions affect the wealth distribution, our model shows that the reverse

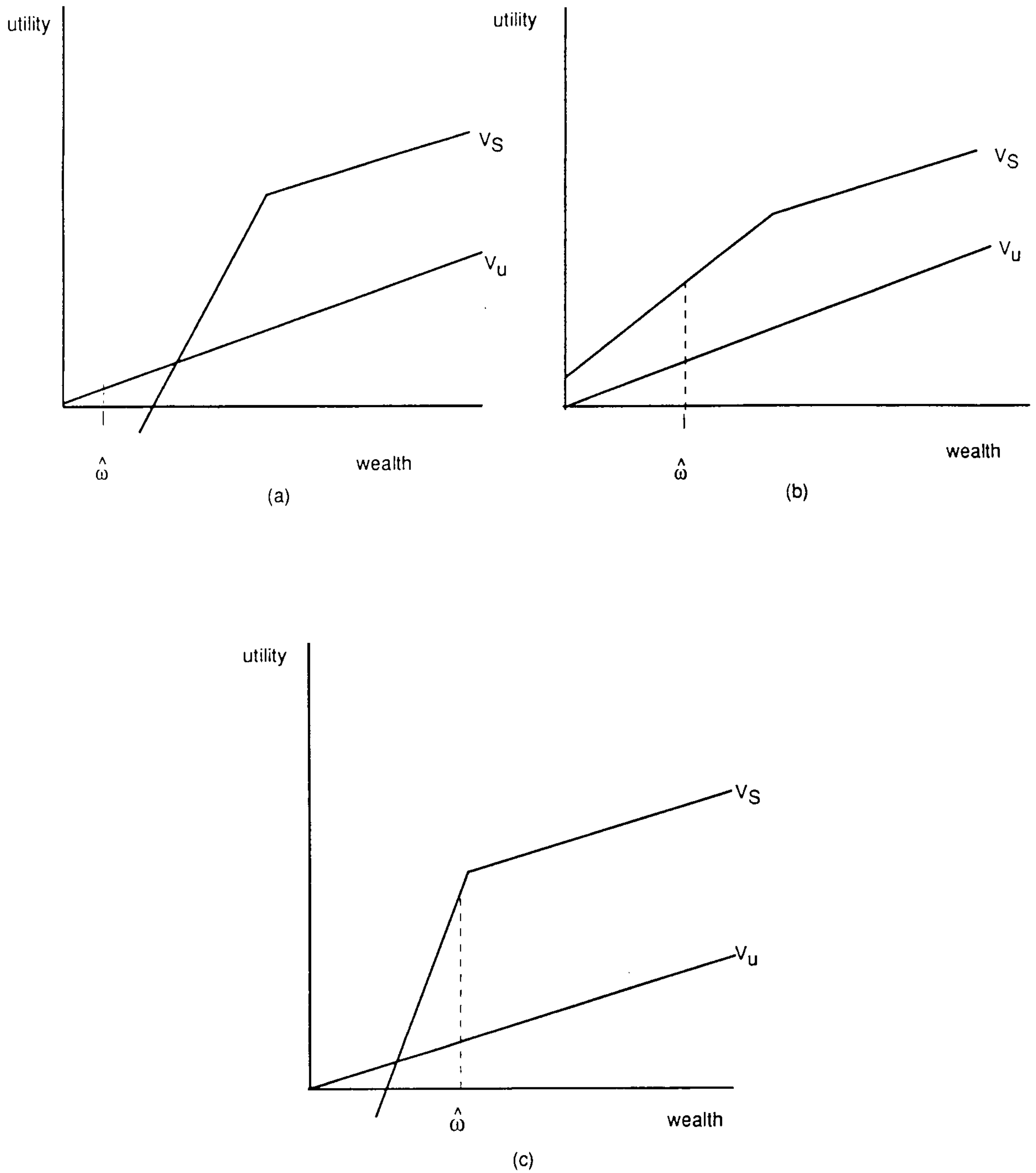


Figure 4 Equilibria with perfect equality. (a) Subsistence; (b) and (c) Self-employment

may be true as well: the distribution can determine the institutions!

Our discussion has been confined to the short run. A natural question arises on whether the dependence of the institutional structure on distribution remains true in the long run or whether, once the endogeneity of the distribution is taken into account, one or another type of contractual equilibrium is inevitable. In Banerjee and Newman (1991), a dynamic variant of the present model is used to approach this question by interpreting $u(\cdot)$ as the indirect utility for a bequest-granting agent and studying the dynamics of of the wealth distribution by following the stochastic process governing bequests. It turns out that depending on the initial distribution of wealth, the economy converges either to a stationary distribution which supports an SEE or to a different one which supports an ECE, at least if the technological parameters assume certain values. With other parameter values, the picture is not so rosy: again depending on the initial distribution, the economy evolves either to an ECE or to total collapse.

Another extension of the model involves endogenizing the safe asset return $\hat{r}$ by allowing it to be determined by supply and demand in the financial contract market. Supply is simply the sum of wealth invested at the safe return by all agents in the economy, while demand is the aggregate of the differences between the capital requirements for the various projects and the initial wealths of the borrowers. Recall that entrepreneurs must borrow enough to cover wages and that the wage fund is then put into escrow and paid at the end of the period. Thus one can write the market clearing condition as equating the total (or mean) wealth in the economy ω^* with the total (physical) capital requirements of all projects. The endogeneity of $\hat{r}$ does not change the correspondence of wealth and occupational choice depicted above, but the switch points do now depend explicitly on $\hat{r}$ as well as v. In an ECE, our equation will be given by

$$\omega^* = k[H(S_2(v,\hat{r})) - H(S_1(v,\hat{r}))] + \mu k[1 - H(S_2(v,\hat{r}))], \qquad (6.1)$$

while in an SEE we must have

$$\omega^* = k[1-H(S_1(\bar{e},\hat{r})]. \qquad (6.2)$$

(The notation in (6.2) reflects the fact that the switch from subsistence to self-employment is exactly the same as the switch from being a worker when the wage is $\bar{e}$). The market clearing condition may hold as an inequality if $\hat{r} = 0$.

Coupled with a labor market equation like (5.4), we have a complete general equilibrium system, and we would again like to know under what conditions an ECE will prevail. In particular, does the existence of an ECE still depend on the distribution, or does the endogeneity of the safe return imply that one or the other sort of organizational equilibrium will always prevail? A complete study of these equations is beyond the scope of this paper, but I do have a complete characterization for all positive values of ω^* outside of $[\frac{\mu-1}{\mu}k, k)$. The arguments are straightforward but detailed, so I do not present them here; rather I summarize the results.

First, if $\omega^* \geq k$, then the only equilibrium is an SEE with $\hat{r} = 0$. This should not surprise us terribly, because given the linear technology, capital in this case is not scarce, so there is nothing preventing everyone from carrying out his own project.

Second, it can be shown that (6.2) has a solution for any ω^*; call it $\hat{r}_0$. Suppose we are in an SEE, but that there are agents who would like to be entrepreneurs if they could pay a wage of $\bar{e}$ when the interest was $\hat{r}_0$. In other words, suppose $S_2(\bar{e},\hat{r}_0) < \bar{\omega}$. Then these wealthy agents would indeed become entrepreneurs (some of the subsisters will become workers at this wage), and the SEE unravels. On the other hand, if $S_2(\bar{e},\hat{r}_0) \geq \bar{\omega}$, then no one wishes to be an entrepreneur at any feasible wage, given the interest rate, and the SEE is "stable." The question is, does the instability of the SEE imply existence of an ECE and does the stability of an SEE imply nonexistence of an ECE, so that the type of organizational equilibrium may be properly attributed to the distribution?

It turns out that the existence of an ECE, at least for the case in which $\omega^* < \frac{\mu-1}{\mu}k$, depends precisely on whether $S_2(\bar{e},\hat{r}_0) < \bar{\omega}$; if the condition fails to hold, then we have an SEE. (This condition is analogous to the second part of condition (1) of Proposition 6; the other parts will be satisfied automatically when $\hat{r}$ is endogenous.) The argument depends on showing that for mean wealth in this range, the only possible ECE is one with a wage of $\underline{v}$, and further that with endogenous $\hat{r}$, $\underline{v}$ must equal $\bar{e}$. Then it remains to show that the capital market equation (6.1) has a solution when the wage is $\bar{e}$. Since the SEE cannot persist if $S_2(\bar{e},\hat{r}_0) < \bar{\omega}$, the ECE is the only equilibrium. At the same time, if $S_2(\bar{e},\hat{r}_0) \geq \bar{\omega}$, then the only interest rate which could satisfy (6.1) would exceed $\hat{r}_0$; but $S_2(\bar{e},\cdot)$ is increasing in $\hat{r}$, so there is no demand for labor, and the only equilibrium is therefore the SEE.

It may be objected that since $\hat{r}_0$ depends on the distribution, that we cannot say that the type of organizational equilibrium properly depends on the distribution in a nontrivial way. But observe that we may rewrite (6.2) as

$$\hat{r}_0 = \frac{\bar{e} - k\bar{r}}{(\alpha-1)H^{-1}(1-\omega^*/k)-\alpha k}.$$

Then our condition on the upper bound of the support of the distribution is clearly an independent one, since there are many distribution switch mean ω^* which have the same value of $H^{-1}(1-\omega^*/k)$ but which have different supports. Once again, in a rough way, a certain amount of inequality is necessary to bring about an ECE.

One might like to know to what extent Proposition 4 depends on risk neutrality. Since our results depend only on the comparing the value functions for the different occupations, we can appeal to the Maximum Theorem to conclude that there is no qualitative change resulting from the introduction of a small amount of risk aversion. For more severe risk aversion, the question is still open, but it would seem that a strict reversal of the proposition, like that we obtained in costless state verification case of Section 3, is unlikely provided that the cost of an audit is not too small.[25]

Finally, as has been suggested above in a footnote, it would be desirable to extend the analysis to the case of incomplete contracts which assign residual control rights and/or rights of exclusion (Grossman-Hart, 1986; Hart-Moore, 1990). The question which presents itself is whether there is an association between wealth and ownership similar to the one we have developed here. In practice, agents with low wealth are rarely observed to acquire such rights. Indeed, if one thinks in terms of the right of exclusion, then it would appear that those who initially have the assets (the wealthy) automatically receive this right. But the issue is complicated by the fact that wealthy individuals frequently give up control rights when, for instance, they lend money or hire managers.

---

[25]The optimal contract with risk aversion has been studied by Mookherjee and Png (1989), but they do not obtain results on the agent's payoff as a function of the principal's required return, which is the information we require. It can be shown for t· case of logarithmic utility (e.g. $u(y) = \ln(y+1)$) and the further restriction that $p = 0$ or 1 that results such as Proposition 2 cannot occur if the auditing cost γ is large enough.

APPENDIX

A.1 *Proof of Proposition 6:*

<u>Proposition</u> <u>6</u> Let $H(\cdot)$ be continuous with support $[\underline{\omega},\bar{\omega}]$. Then an ECE exists if and only if

(1) $\underline{\omega} < S_1(\bar{v})$ and $S_2(\underline{v}) < \bar{\omega}$;

(2) $S_1^{-1}(\underline{\omega}) < S_2^{-1}(\bar{\omega})$.

To see that these conditions are sufficient as well as necessary, consider the excess demand correspondence (we are ignoring v that are sufficiently greater than $\bar{v}$ that everyone wants to be a worker or sufficiently smaller than $\underline{v}$ that everyone wants to be an entrepreneur, since these values of v do not affect the argument):

$$\begin{array}{ll} -H(S_1(v)) & v > \bar{v} \\ [0,\mu(1-H(S_2(\bar{v})))] - H(S_1(\bar{v})) & v = \bar{v} \\ \mu(1-H(S_2(v))) - H(S_1(v)) & \underline{v} < v < \bar{v} \\ \mu(1-H(S_2(v))) - [0,H(S_1(v))] & v = \underline{v} \\ \mu(1-H(S_2(v))) & v < \underline{v} \end{array}$$

which is easily checked to be upper hemicontinuous, convex-valued and decreasing. Condition (1) tells us that $H(S_1(\bar{v}) > 0$ and $H(S_2(\underline{v})) < 1$, so the excess demand contains positive values for $v = \underline{v}$ and negative values at $v = \bar{v}$. Thus, there is a (unique) $v^* \in [\underline{v},\bar{v}]$ such that the market for workers clears. To verify that a positive measure of agents are workers at this wage, note that this is automatically true if either $H(S_2(\bar{v})) < 1$ or $H(S_1(\underline{v})) > 0$ since then we have $\mu[1-H(S_2(v^*))] \geq \mu[1-H(S_2(\bar{v}))] > 0$ or $H(S_1(v^*)) \geq H(S_1(\underline{v})) > 0$.

In the other case, observe that for wages exceeding $S_1^{-1}(\underline{\omega})$, the supply of workers is positive, while for wages less than $S_2^{-1}(\bar{\omega})$, the demand for workers is positive. Provided the equilibrium wage v^* wage lies in $(S_1^{-1}(\underline{\omega}),S_2^{-1}(\bar{\omega}))$, as condition (2) allows, we have an ECE. Computing the excess demand at these endpoints, we find first that it is positive at $S_1^{-1}(\underline{\omega})$. To see this, note that the excess demand at $S_1^{-1}(\underline{\omega})$ is $\mu[1-H(S_2(S_1^{-1}(\underline{\omega})))]-H(S_1(S_1^{-1}(\underline{\omega}))) = \mu[1-H(S_2(S_1^{-1}(\underline{\omega})))]$. Since $S_2(S_1^{-1}(\underline{\omega})) > S_1(S_1^{-1}(\underline{\omega})) = \underline{\omega}$, $H(S_2(S_1^{-1}(\underline{\omega}))) > 0$ and excess demand is positive (the only way it could not be is if $S_2(S_1^{-1}(\underline{\omega})) \geq \bar{\omega}$, which contradicts condition (2)). A similar argument shows that the excess demand is negative at $S_2^{-1}(\bar{\omega})$. Consequently there is an equilibrium wage somewhere between, as required.

A.2 *Variable Firm Size*

Suppose that entrepreneurs were permitted to hire fewer than μ employees. Assume first that it requires full effort to monitor one's employees, no matter how few one may have. It is a simple matter to verify that Proposition 3 remains true in this case, so that the value function for an entrepreneur contemplating hiring $\lambda < \mu$ employees would look like that in Figure 5, with a kink at $C(\lambda) = \lambda(k+v/\hat{r})$. Now if $C(\lambda)$ is less than her wealth ω, an increase in λ results in a gain of value of $k(\bar{r}-\hat{r})-v$ which exceeds zero for any wage in the allowed range. On the other hand, if $C(\lambda)$ exceeds ω, so that the entrepreneur must borrow to finance additional expansion of the firm, she gains $k(\bar{r}-\alpha\hat{r})-\alpha v$, which is negative for any v in the allowed range. Note that an agent would be indifferent between operating a firm with $\underline{\lambda} = k(\bar{r}-\hat{r})/[k(\bar{r}-\hat{r})-v]$ and being self employed. Thus agents with wealth between $\underline{\lambda}(k+v/\hat{r})$ and $\mu(k+v/\hat{r})$ will choose to be "small scale" entrepreneurs, setting firm size λ equal to $\omega/(k+v/\hat{r})$.

Such entrepreneurs are inefficient relative to larger entrepreneurs since they expend more resources per worker on monitoring. Given the wealth distribution, therefore, per capita output must fall (provided there were no initially no subsisters) when variable firm size is introduced. On the other hand, it is fairly clear that all else the same, the equilibrium wage must be higher when variable firm sizes are allowed (Proposition 4 is unaffected by the introduction of variable firm size, and indeed the supply function is unchanged, since this is just given by $H(S_1(v))$; but the demand must be higher at the old wage so equilibrium entails that the new wage is higher) so that workers as well as small entrepreneurs are better off than before; only the large entrepreneurs are made worse off. Thus capital market imperfections can determine the size of the firm, so that the distribution of wealth will have implications for the distribution of firm size as well as for the type of firms which inhabit the economy.[26] Notice that in reference to what we said about increasing returns to scale in the Introduction, we have just outlined a case in which there is increasing returns, but firms are typically smaller than the minimum efficient scale.

As for the existence and characterization of ECE in this case, we have to make some modifications to our definitions, but the general flavor of our

---

[26] See Banerjee, Newman and Qian (1991) for an elaboration of this point.

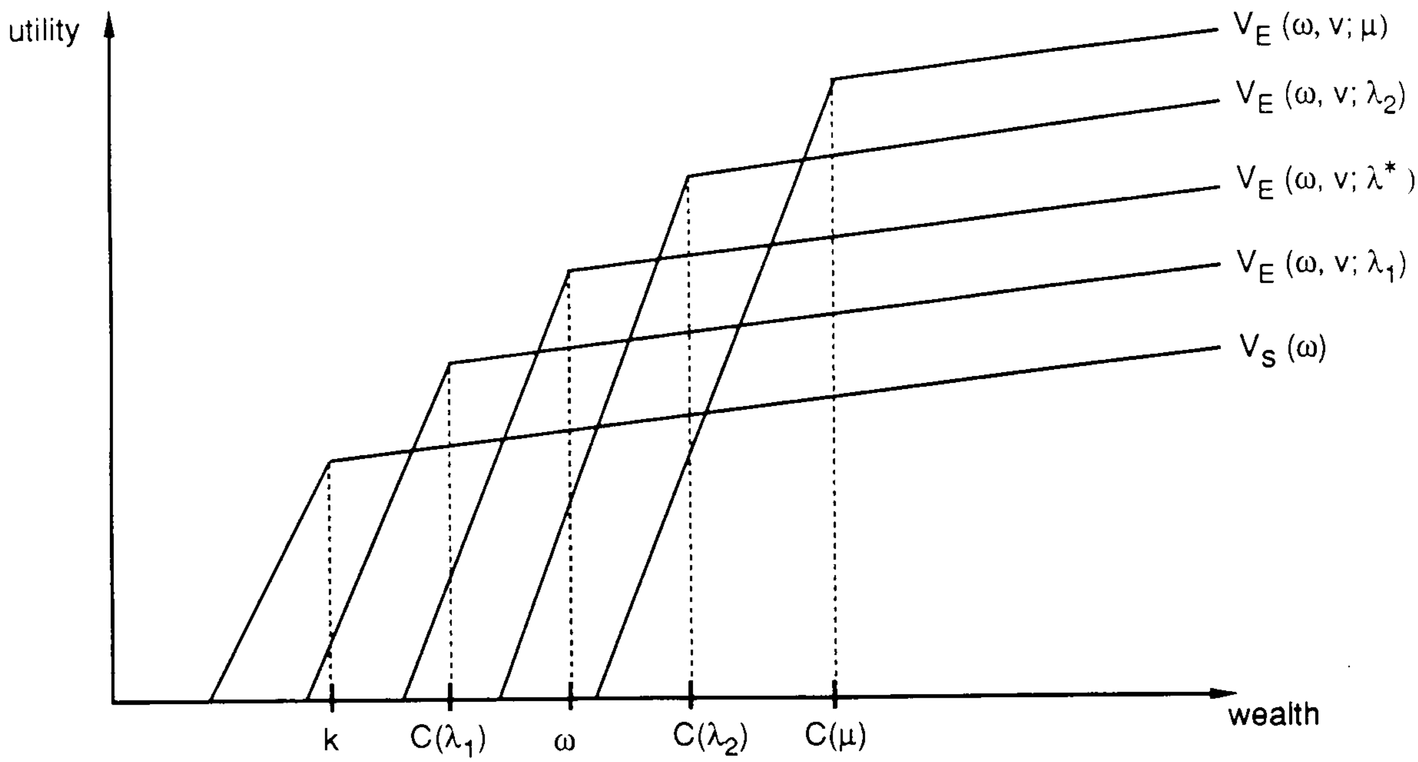


Figure 5. Choice of firm size when employment level is variable. Optimal size is given by $C(\lambda^*) = \omega$.

results is unchanged.

If we suppose that individual projects and effort are divisible, we get similar results. All agents will set up firms of size equal to their wealth and spend the remainder of their effort working for a wage. In this case we have "pure" constant returns to scale in production, but nevertheless get aggregated production. Of course, the occupations are a little less "pure," since everyone with positive wealth will spend at least a little time working for himself. But the poorer one is, the more time one spends as a worker. Moreover, no one with wealth less than k will be an entrepreneur of any sort.

Finally, observe that we could relax the assumption that monitors cannot be monitored. Even with no "information dissipation" (each supervisor can monitor the same number of workers below him, regardless of his level in the hierarchy), the capital market imperfection will still give the firm a determinate size and the hierarchy a determinate number of levels.

REFERENCES

Alchian, A. and H. Demsetz (1972), "Production, Information Costs, and Economic Organization," *American Economic Review*, vol. 62, 777-795.

Banerjee A.V. and A.F. Newman (1991), "Occupational Choice and the Process of Development," mimeo, Northwestern University.

Banerjee A.V., A.F. Newman and Y. Qian (1991) "The Capital Market and the Distribution of Firm Size," mimeo, Northwestern and Stanford Universities.

Bernanke B. and M. Gertler (1989), "Agency Costs, Net Worth, and Business Fluctuations," *American Economic Review*, vol. 79, no. 1, 14-31.

——— (1990), "Financial Fragility and Economic Performance," *Quarterly Journal of Economics*, vol. CV, no. 1, 87-114.

Calvo, G.A. and S. Wellisz (1979), "Hierarchy, Ability and Income Distribution," *Journal of Political Economy*, vol. 87, no. 5, 991-1010.

——— (1980), "Technology, Entrepreneurs, and Firm Size," *Quarterly Journal of Economics*, vol. XCV, no. 4, 663-677.

Eswaran, M. and A. Kotwal (1989), "Why Are Capitalists the Bosses?," *Economic Journal*, vol. 99, no. 94, 162-176.

Evans, D.S. and B. Jovanovic (1989), "An Estimated Model of Entrepreneurial Choice Under Liquidity Constraints," *Journal of Political Economy*, vol. 97, no. 4, 808-827.

Grossman, S.J. and O.D. Hart (1986), "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration," *Journal of Political Economy*, vol. 94, no. 4, 691-719.

Hart, O.D. and J. Moore (1990), "Property Rights and the Nature of the Firm," *Journal of Political Economy*, vol. 98, no. 6, 1119-1158.

Kanbur, S.M. (1979), "Of Risk Taking and the Personal Distribution of Income," *Journal of Political Economy,* vol. 87, no. 4, 769-797.

Kehoe, T.J. and D.K. Levine (1989), "Debt Constrained Asset Markets," mimeo Federal Reserve Bank of Minneapolis.

Kihlstrom, R.E. and J.-J. Laffont (1979), "A General Equilibrium Entrepreneurial Theory of Firm Formation Based on Risk Aversion," *Journal of Political Economy,* vol. 87, no. 4, 719-748.

Knight, F. (1921) *Risk, Uncertainty and Profit*, Boston: Houghton-Mifflin.

Light, I.H. (1972), *Ethnic Enterprise in America: Business and Welfare among Chinese, Japanese and Blacks*, Berkeley: University of California Press.

Light, I.H. and E. Bonacich (1988), *Immigrant Entrepreneurs*, Berkeley: University of California Press.

Lucas, R.E. (1978), "On the Size Distribution of Business Firms," *Bell Journal of Economics*, vol. 9, no. 2, 508-523.

Mirrlees, J. (1974), "Notes on Welfare Economics, Information, and Uncertainty," in M. Balch, D. McFadden, and S.-Y. Wu, eds., *Essays on Economic Behavior under Uncertainty*, Amsterdam: North-Holland.

Mookherjee, D. and I. Png (1989), "Optimal Auditing, Insurance, and Redistribution," *Quarterly Journal of Economics*, vol. CIV, no. 2, 399-416.

North, D.C. (1981), *Structure and Change in Economic History*, New York: Norton.

Roemer, J. (1982), *A General Theory of Exploitation and Class,* Cambridge, Massachusetts: Harvard University Press.

Simon, H. (1951), "A Formal Theory of the Employment Relation," *Econometrica*, Vol. 19, 293-305.

Townsend, R.M. (1979), "Optimal Contracts and Competitive Markets with Costly State Verification," *Journal of Economic Theory*, vol. 21, 265-293.

Vassilakis, S. (1989), "Increasing Returns and Strategic Behavior: the Worker-Firm Ratio," *RAND Journal of Economics*, vol. 20, no. 4, 622-636.