

Discussion Paper No. 1024

**WEALTH EFFECTS, DISTRIBUTION,
AND THE THEORY OF ORGANIZATION**

by

Patrick Legros*

and

Andrew F. Newman**

First Version: October, 1992

Revised February, 1993

*Department of Economics
Uris Hall
Cornell University
Ithaca, NY 14853

**Department of Economics
Andersen Hall
Northwestern University
Evanston, IL 60208

We are grateful for comments from Abhijit Banerjee, Drew Fudenberg, Roger Guesnerie, Matt Jackson, Alejandro Manelli, Meg Meyer, Raaj Sah, Shlomo Weber, and workshop participants at Paris, Stanford, Northwestern, Chicago, Cornell, Rochester-Cornell-Buffalo, Harvard/M.I.T., and Minneapolis. Wolfgang Pesendorfer, Lars Stole, Asher Wolinsky, and Myrna Holtz Wooders deserve special mention for their advice and patience.

ABSTRACT

We construct a general equilibrium model of firm formation in which organization is endogenous. Incentive-based wealth effects arise from lower bounds on wealth and utility, and these affect the way in which different organizational forms can divide the proceeds of production. Individuals may choose between organizing their firms as hierarchies or as partnerships; these decisions are mediated by agency costs in both labor and financial markets. The type of organization which emerges depends on the distribution of wealth and need not be surplus maximizing: the same output could be produced with less labor if some firms were forced to reorganize from their equilibrium form. This result suggests that instead of serving to provide incentives efficiently, organizations may sometimes act to transfer surplus from some agents to others, at potential social cost.

1. Introduction

What is the role of organization in a market economy? What determines the form that an organization assumes? These are among the central questions that the theory of the firm has sought to answer at least since Coase raised them some sixty years ago. The salient feature of firms in market economies is that they form by voluntary association — individuals are free to move among them, create and dissolve them, and choose the way they are organized. Moreover, an agent's willingness to join a firm depends on the gains he expects to secure there and on what he foregoes in outside opportunities. These benefits and costs in turn depend on supply of and demand for other agents, both those with similar attributes and opportunities and those with complementary ones. Thus any explanation for organization — whether based on transaction costs, agency and incentive problems, or difficulties of coordination and information transmission — must take account of the competitive forces which mobility generates. In short, the theory of organization ultimately needs a general equilibrium formulation, one which builds predictions about the forms and roles of organizations from the fundamental economic data of preferences, technology, resources and information structures.

The purpose of this paper is to construct a simple version of such a theory, providing a general equilibrium model of firm formation in which both the firms and their organization are endogenous. Our approach is essentially club-theoretic: firms are associations of individuals for the purpose of production that are governed by incentive compatible sharing rules and financial contracts.¹ The model permits one to study the effect on organizational choice of the interaction of several market imperfections, an

¹As our title suggests, we deal here with organization, rather than ownership; indeed there is nothing to preclude interpreting an equilibrium of our model to be descriptive of one giant firm with zillions of subdivisions. Our framework can be extended to allow for ownership à la Grossman-Hart-Moore, and we shall do so in future work.

aspect of the theory that is generally neglected by the more conventional, partial equilibrium approach. The results of this analysis suggest some new answers to questions concerning the efficiency of organizational form and a new understanding of the possible roles played by organization in a market economy.

One thing an explicit competitive model should allow is a closer examination of some of the assumptions underlying the predominant approach in organization theory. Usually, one seeks explanations of the role and structure of organization based on some notion of "efficiency." The basic presumption is that decentralized competition should generate organizations which respond efficiently to market failures. If a particular organizational form were not efficient, its members would dissolve their enterprise and earn more income for themselves by reorganizing in a different way (or an entrepreneur could offer higher incomes to anyone joining his better-organized firm).

At least if one uses a strong enough notion of efficiency (weaker ones such as the Pareto criterion are useful for eliminating possibilities, but generally cannot make very specific predictions), this argument is quite powerful, for it greatly simplifies the task of studying organization: one need only maximize a suitable objective function against the constraints of information, transaction, or coordination one is interested in. Indeed, using as an objective total surplus generated by the firm, the modern theory has taught us much about the workings of organizations.²

Of course, an agent doesn't care about social surplus or even that of his firm; he only cares about what he can get for himself. Given that it is individual motivations that are the driving force of competition, we need to consider environments in which the decentralized equilibrium does lead to the existence of surplus maximizing firms if this partial equilibrium approach is to be valid.

Environments in which wealth effects are not significant have the desired property.³ If by reorganizing an enterprise there is a way to increase the

²The literature on the internal organization of the firm is vast, and we cannot summarize it here. Surveys can be found in Hart-Holmström (1987), Holmström-Tirole (1989), Hart (1989), and Radner (1992).

³By wealth effects, we mean changes in an agent's marginal incentives as his wealth varies. Although potentially important, we will not be concerned with

joint surplus of a group of individuals, these gains can always be distributed among them so that they will all agree to the change. But when wealth effects are important, there is in general no reason to believe that the surplus maximization shortcut will yield the same answers as a model of competitive process of organization formation: as the phrase goes, distribution and efficiency can no longer be separated.

While the general point is well known (indeed, well enough to be reiterated throughout a textbook: see Milgrom and Roberts, 1992), in the practice of the theory of organization, it has mostly been ignored. But in many instances one would think that wealth effects matter, whether for the questions of economic history and development, the problems of urban poverty and entrepreneurship, or the issues of East European privatization and reform. And without a conceptual structure for systematically analyzing the complications introduced by wealth effects, it is hard to know what one is missing.

The framework we provide in this paper is designed to help to fill this gap. It focuses on wealth effects and the consequent role of distribution in the determination of the structure and function of organization. Our model is abstract and meant more as a conceptual guide than as an accurate account of observed organizations.

We consider an environment in which production occurs with increasing returns over a range, so there is a priori reason for enterprises to form. Organizational choice arises from the presence of two sorts of agency costs, one arising from the costly verification of output, the other from the costly verification of effort. Output can only be learned by participating as a member of a firm or through the use of a costly auditing technology; effort of another agent can only be learned if he is scrutinized with a costly monitoring technology. Given the assumptions we make about the technologies of production and information gathering, the organizational choice is quite simple: either an enterprise is set up as a hierarchy with one manager supervising several workers, or it is set up as a partnership in which the agents are induced to work through an appropriate incentive scheme.

preference-based wealth effects such as the normality of leisure or declines in risk aversion.

The incentive cost of a hierarchy is obvious in this setting: one agent is spending her effort monitoring instead of producing. The partnership, on the other hand, can overcome its free-rider problem efficiently, provided an appropriate incentive scheme can be implemented (agents are risk-neutral, so this can be done without welfare loss due to lack of insurance).

It turns out that the condition whether such a scheme is feasible depends on the total wealth of the members of the enterprise (Legros-Matsushima 1991; Legros-Matthews, forthcoming). If this wealth exceeds a certain level, then it is incentive compatible for all members to exert first-best effort; below that level, first-best effort cannot be achieved. This is the first of two sources of wealth effects in the model.

The second source stems from the costly verification of output, which leads to a capital market imperfection. The wealthier is a borrower, the lower the agency cost of borrowing. In fact, this sort of wealth effect can be generated from other sources besides the cost of observing output: it is capital market imperfections more generally that are at issue.

The wealth effects generate two reasons for the viability of hierarchies. First, poor agents who would be unable to form an efficient partnership might still form a hierarchy, if borrowing is not too costly. But a second reason for hierarchies emerges if borrowing is more costly. In this case poor agents can instead cheaply get together with wealthy ones inside firms. One possibility is that a rich agent forms a partnership with the poor; but he may strictly prefer to avoid paying the poor the high surplus needed to make them incentive compatible in a partnership and to opt for a hierarchy instead.

Now while it is clear enough that an individual will always try to capture a big slice for himself, even if it comes from a smaller pie (this could be modeled straightforwardly in a partial equilibrium principal-agent framework), what is not clear is when anyone can do this: the virtue of competition is that it disciplines people against such opportunism. It is only within a general equilibrium framework, one which explicitly models the effect of competition on the choice of organizational form, that we can evaluate the likelihood and scope of such possibilities.

As it turns out, the attraction of the hierarchy increase as the wage falls, and the ability to pay low wages depends on the supply and demand for workers, in particular the outside opportunities of would-be workers. These in turn are determined by the distribution of wealth: when it is sufficiently skewed to the low end, hierarchies emerge. Moreover, these hierarchies can be

inefficient; for some wealth distributions, an appropriate tax scheme (or outright ban) would turn every hierarchy into a partnership which produces the same output using less labor. But the resulting increase in surplus — and in poorer agents' incomes — could be achieved only at the expense of the wealthiest agents, whose incomes would fall.

Thus when individuals choose which firms to join and what contractual forms to use, we find that efficiency cannot explain organizational form: in a competitive equilibrium, efficient forms may not be used, and only inefficient ones may prevail. Rather than thinking of observed organizational structure as a consequence of efficiency, we might do better instead to see it as a reflection of the distribution of wealth. This point of view also allows us to view organizations not merely as part of the "social technology"; under some circumstances, they play the role of pure transfer devices, even if this entails social costs.

These points are illustrated by an example which we present in Section 6; some readers may want to skip to that section immediately after perusing the notation in Section 2, which introduces the model. Section 3 then develops the notion of feasibility for a firm, which essentially involves characterizing financial and share contracts. As a side result of the analysis, we note that since monitoring itself is not contractible, any share contract for a hierarchy must satisfy certain "upward" incentive compatibility conditions. Specifically, the manager must have an incentive to monitor, which means that if he does not, expected output must be low: the shares going to workers must not provide them with incentives to work if they are not monitored. If they did, there would be no need for the manager, and the hierarchy would not exist. Thus the very existence of a hierarchy in equilibrium places restrictions on the compensation schemes available to its members: roughly speaking, the manager must bear a certain amount of risk, while the workers must not bear too much.

Next, we make precise our notion of equilibrium. The outcome of competition must be a partition of the set of agents (equal to the unit interval) into firms and occupations; it is in this sense that our model has a club-theoretic flavor, and a variant of the core, known as the f -core (Kaneko-Wooders, 1984, 1986; Hammond-Kaneko-Wooders, 1989), turns out to be the most convenient solution concept. The f -core resembles the usual core for continuum economies, except that only finite coalitions are permitted to block. In Section 4, we define this concept for our model, discuss its

appropriateness, and establish the existence of an equilibrium.

We present in Section 5 the main results on characterizations of equilibrium. Proposition 2 there states that equilibrium financial contracts maximize the firm's expected income. This is not obvious in view of what we have said about the tenuousness of surplus maximization as a characterization of equilibrium, and the proof requires somewhat delicate construction of transfers to the firm's members which avoid violation of their incentive constraints.

Proposition 3 then shows that surplus maximization is indeed valid in case the capital market operates efficiently, specifically when the cost of auditing is zero. As much of organization theory proceeds by considering a single market imperfection at a time, this result suggests why failures of surplus maximization are apt to be missed. The next two propositions consider economies in which the capital market is far from perfect and provide sufficient conditions on the distribution of wealth for firms to consist exclusively of partnerships (Proposition 4) and for the presence in competitive equilibrium of inefficient hierarchies (Proposition 5); roughly speaking, the former occurs when the distribution is skewed to the right, the latter when it is skewed left. The numerical example appears in Section 6, and Section 7 concludes.

2. Model

(a) *Preferences and Demographics*

The economy lasts one period and has one physical good which may be used for consumption or as capital. There is a large number of agents (indexed by the unit interval with Lebesgue measure; denote this set A) with identical preferences defined over income and effort. Agents are identically endowed with one unit of indivisible effort and one unit of indivisible time,⁴ but differ in their wealth endowment: $\omega(a)$ is the wealth of agent a ; $\omega: A \rightarrow \Omega \subset \mathbb{R}_+$ is Lebesgue measurable. We assume that there are finitely many wealth levels, that is, $\Omega = \{\omega^1, \omega^2, \dots, \omega^J\}$, where ω^i is increasing in i . Let h^i denote the (strictly positive) measure of agents with wealth ω^i , $H(\omega)$ the

⁴The assumption on the indivisibility of time is made to ensure that agents may only be members of a single enterprise.

measure of agents with wealth less than ω .⁵ Agents have identical risk-neutral preferences which may be summarized by the von Neumann-Morgenstern expected utility $E\{\eta - e\}$, where $\eta \geq 0$ is the realized lifetime income and $e \in \{0,1\}$ is the effort level chosen.

Since we are emphasizing the role of wealth effects — in particular *incentive* wealth effects — we should point out here that they enter because of the lower bound on utility that inheres in risk neutral preferences defined over the nonnegative numbers. The main source of these wealth effects, though, should be understood to be the lower bound on utility, not the lower bound on wealth. (If agents were risk averse, it would be possible to render these wealth effects negligible just by assuming that utility is unbounded below as wealth approaches zero; risk aversion, though, would typically introduce wealth effects of its own.) When it comes to obtaining efficient contracts, the real difference between the rich and the poor is not that the rich have more money, but that they have more utility to lose.

(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 an exogenous gross return $r > 0$. By arbitrage, this return is also earned by lenders of capital — one could equivalently think of our economy as small and open, with r the world gross interest rate.

Second, there is a risky investment project which comes in discrete units with capital requirement K ; once sunk, this capital cannot be recovered. This project succeeds, yielding R , with a probability π_n (n is the number of agents expending effort on the project), and fails, yielding 0 , with probability $1 - \pi_n$. Returns are independent across projects. We make the following assumptions about the function π :

- (T1) π is nondecreasing with $\pi_0 = 0$;
- (T2) There is an $N \geq 2$ such that $\pi_n = \pi_N < 1$ for $n \geq N$
- (T3) N is the unique maximizer of $\pi_n R - Kr - n$
- (T4) $\pi_N R - N - 1 > Kr$.

⁵Most of our results extend straightforwardly to distributions with compact supports as long as $\inf \{\omega \in \Omega | \omega > 0\} > 0$.

The first assumption needs no comment. The second is a simple form of diminishing returns. (T3) states that the unique first-best efficient scale of operation is N agents. The fourth assumption implies that projects are viable when operated as hierarchies (and therefore as partnerships of size N) in the sense that they constitute a better use of capital than subsistence.

The other two technologies are for information gathering. Productive effort is not directly observable. The third technology allows it to be perfectly monitored. If an agent belonging to a firm expends her own unit of effort with this technology, she observes the effort levels of all other agents associated with that firm's project, as well as the project's outcome; if she shirks, however, the monitored agents know this and therefore might have an incentive to shirk as well⁶ (this rules out stochastic monitoring and the accompanying complications in compensation schemes). Assume for simplicity that it is not possible to verify (as distinct from observe) the effort of a manager, so that only one level of hierarchy is feasible.

Finally, the fourth technology permits verification of the outcome of the project by an "outside" party (i.e., someone other than the agents undertaking the project). Specifically, it costs γ to "count" the output and learn whether the project succeeded; this information becomes public knowledge.

(c) *Information*

Since we have already mentioned information gathering, there must be some asymmetries of information. We assume:

- (I1) productive effort is not freely observable; the monitoring effort of the manager is observed by the other agents in the firm before they decide on their own effort levels.
- (I2) all agents who are members of a firm learn the outcome of its project costlessly; no one else does;
- (I3) the parties to a contract with an agent can observe all other contracts he may be party to.
- (I4) wealth can be costlessly verified.

Because an agent's time is indivisible, he will be involved with at most one

⁶Subject to the caveat that they haven't signed a contract which does not induce them to shirk. More on this later.

firm (this is the maximum number of projects whose output he can observe costlessly).

(d) *Occupations and Organizations*

Given the technology and information structure, there are four possible pure occupations. An agent can invest all of his wealth in the safe asset (or lend to other agents — these will yield the same return) and expend no effort. This will be called the subsistence option, and it yields income $\omega(a)r$ to agent a . Often, no one chooses subsistence in equilibrium, but it always provides an individual rationality constraint for the other occupations.

The members of a firm F belong to one of three disjoint sets, $W(F)$, $S(F)$, and $M(F)$. $W(F)$ is the set of *workers* in F , agents who are expected to provide a unit of effort; if n agents provide a unit of effort then the probability of obtaining R is π_n . $S(F)$ is the set of *silent partners*, members who contribute only their wealth to the partnership but are not expected to provide a unit of effort.⁷ Finally, $M(F)$ is the set of *managers*, members who are expected to expend a unit of effort to monitor the other agents in F .

If the set $M(F)$ is empty, then the only way to ensure that the agents in $W(F)$ exert themselves is by designing a share contract, contingent on the return of the project only, that induces each worker to exert her unit of effort. We call such firms *partnerships*. If $M(F)$ is not empty, then the share contracts can be written contingent on the return of the project and also on the effort levels of the agents in $F \setminus M(F)$. We call such firms *hierarchies*.

For simplicity and expositional clarity we assume that only pure occupations are possible for the agents: no one may spend part of his time or effort in one occupation and part in another, and no lotteries are allowed. The extension to the case of mixing of occupations is straightforward.

As we shall see, each of the two forms of organization in general has costs, the size of which will depend in part on the relative wealths of the

⁷Silent partners can play two roles. First, because members of a firm can observe the output while outsiders cannot, it is cheaper to borrow from silent partners than from the financial market. Hence, silent partners reduce the cost of borrowing. Even if the other agents have a total wealth greater than K , silent partners might be useful to implement the desired effort levels from the agents in W . This second role is akin to the budget-breaking solution of Holmström (1982).

members. What we want to do is investigate how a competitive economy makes the tradeoff between these two types of firms, and in particular how the predominance of one organizational form may depend on the distribution of wealth. The answers we get will depend crucially on the workings of the capital market. Thus, we shall have to examine two sorts of interactions, viz., those among members of the enterprise and those between the enterprise and outside investors.

We imagine that firms or enterprises act as intermediaries between agents and lenders. That is, the firm signs a share contract with its members which specifies how the proceeds of that firm's production (net of financial obligations to lenders) are distributed among the members of the enterprise. Lenders in turn sign financial contracts with firms, not with individual agents.

This separation assumption is an important simplification. The lenders' side of the financial market is treated here in the standard competitive fashion, free from agency problems. One could extend the model to allow for financial intermediation as a separate occupation, as in Diamond (1984) or Boyd and Prescott (1986).

More important, we limit the contracts between lenders and firms to consist only of income transfers to the firm as a whole, rather than to individual members. We feel this is a realistic assumption which also yields a tractable model.⁸ We should emphasize, however, that our arguments depend

⁸It might be objected that a lender should sign a contract with each agent in the firm, effectively giving him control of the agent's share. As will become apparent, the main obstacle to achieving fully efficient financial contracts stems from the asymmetry of information about the success of the project: the members of the firm have incentives to report failure even when there was success so that they can avoid repaying their obligations. If lenders could commit to "whistleblower" arrangements, they might be able to extract this commonly held information. But since such arrangements typically pay more to the lender when there is a whistleblower than when all agents report truthfully, the lender will always prefer to claim that there was a single whistleblower. If the content of messages from the firm's members to the lender are not verifiable, such contracts will not be feasible (and if the burden of proof is placed on the lender, the court will always have to decide in favor of the members, which then makes the whistleblower contract ineffective — a sole truth teller anticipates that he will not be able to claim his reward because all the other members will claim that they too told the truth). Thus with falsifiable messages, the only relevant messages that can be sent are all fails or all successes: the members have to act as though there were a single agent ("the firm") acting on their behalf.

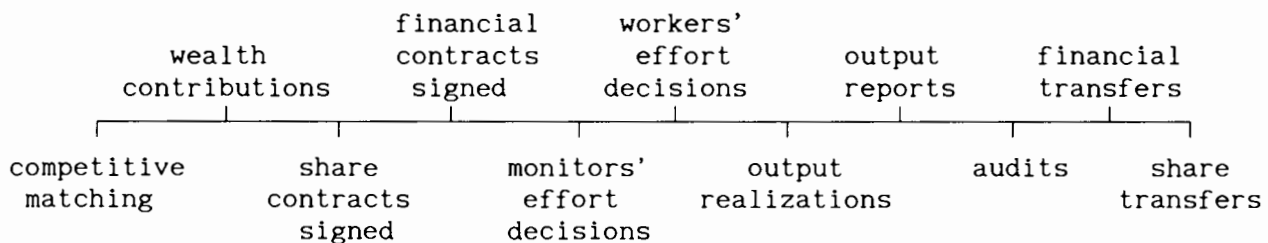
Commitment by some reputation mechanism appears to be equally difficult,

less on specific models of imperfect capital markets than on imperfect capital markets generally. Other models which would lead to similar results include Bernanke-Gertler (1990), Kehoe-Levine (forthcoming), and Hart-Moore (1991).

Figure 1 illustrates schematically the relationships among the loan market, firms, and their members, with the most general formulation in (a) and our formulation in (b).

(e) *The Economy*

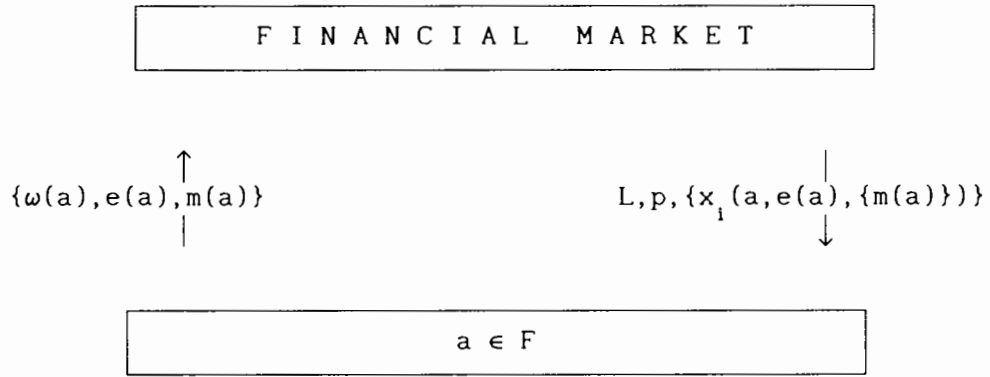
The timing structure of our model is illustrated below. It is quite



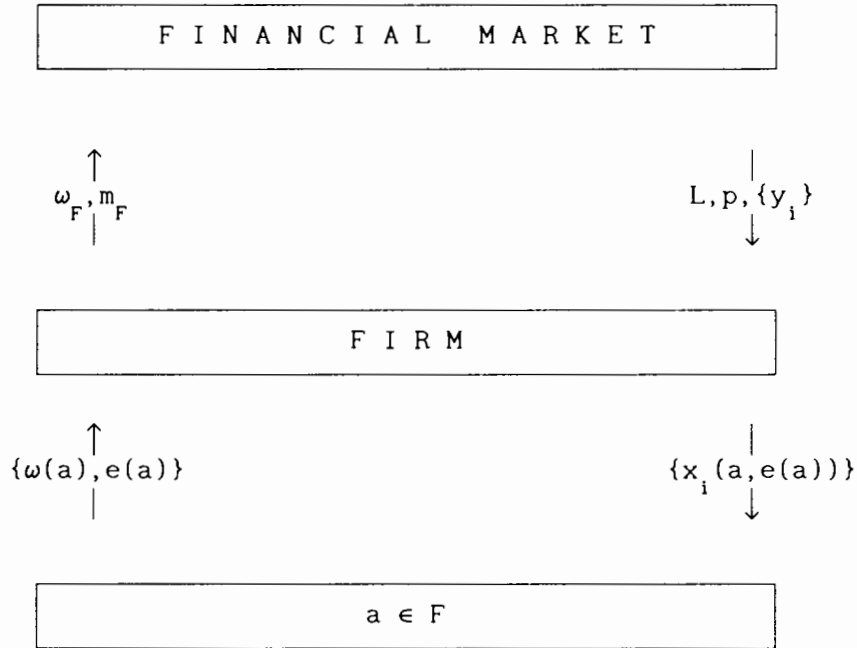
standard for models of contracting and organization except for the presence of the competitive matching stage in which the firms form. The set of firms in this economy is, of course, not given. Any group of agents may get together, create a firm, and sign a share contract (knowing what kind of financial contract they'll sign). Since no one will be part of more than one firm, what we are looking for as an outcome of our competitive matching process is a *partition* of the set of agents. The parallel with the Tiebout model of community formation should be apparent.

The competitive equilibrium concept we use is a variant of the so-called *f*-core, due to Kaneko-Wooders (1984, 1986; see also Kaneko-Wooders-Hammond, 1989 and Wooders, 1992) and related to the "social coalitional equilibrium" of Ichiishi (1977) as well as the classical theory of clubs. Here the set of agents partitions itself into finite coalitions, each of which achieves something feasible for itself (to be defined below, but basically this comes down to producing output and distributing it among the members and the lenders according to incentive compatible sharing rules and financial contracts) and such that no other (finite) coalition could form which would give each of its

since an outside observer of the lender's behavior can never know what really happened with respect to the set of messages sent.



(a) General formulation. Note individual shares depend on all agents' messages. L = loan amount, p = audit probability, $m(a)$ = a 's report, x_i = share in state i .



(b) Simplified formulation. $\omega_F = \sum \omega(a)$, m_F = firm's report, y_i = firm's income in state i .

Figure 1

members a payoff higher than what they are getting in their current coalitions. We interpret the coalitions as firms or enterprises; the lenders are not part of explicit coalitions, and in fact could be thought of as being outside the economy.⁹ As we argue below, any equilibrium of the economy is a constrained Pareto optimum (at least for a full measure of agents).

Before defining this concept formally, we must specify what is meant by a feasible contract for a firm. That is the task of the next section.

3. Contracts and Feasibility

As we remarked above, the firm F will be thought of as acting on behalf of its members in the financial market; the financial contract will consist only of income transfers to the firm y , an audit probability p , a loan amount L , and collateral C . The income in each state is then distributed among the members via the share contract x . Finally, it is useful to denote the organization of the firm by the sets W , M , and S , which identify the members in each occupation.

Formally, then, a contract for a firm F is a tuple $c = (y, p, L, C, x, W, S, M)$, where $y = (y_s, y_f, y_t, y_l)$ and $x = (x_s(a), x_f(a), x_t(a), x_l(a))_{a \in F}$ (the meaning of the subscripts will become clear below). Of course, feasibility entails that these parts of the contract cannot be defined arbitrarily and independently of each other. We examine the restrictions on each of them in turn.

(a) *Finance*

⁹Our approach seems more natural than an alternative, Walrasian solution concept which has also been used to discuss firm formation (see e.g. Drèze, 1976, 1989), in which beside the usual commodities, memberships in all possible firms are offered for sale (in principle, assuming only finite firms are feasible, the size of the set of commodities is of the order of that of the set of all finite subsets of the agent space). One can then assume price-taking behavior for each of these commodities; in equilibrium, most of them will not be "produced." This does not strike us as an especially natural application of the Walrasian equilibrium concept; in particular, the price-taking assumption doesn't seem very plausible since the market for each good is generally thin. Because of the strong complementarities exhibited by these "artificial" commodities (if there is to be a firm consisting of agents 1 through 12, then if the good "a position for person 1 in the firm consisting of 1 through 12" is offered for sale, so must the good "a position for person 9 in the firm consisting of 1 through 12") there may be many inefficient equilibria as well (see Hart 1980 or Makowski 1980). Since our equilibria turn out to be Walrasian anyway, adherents to the Walrasian approach should not be unhappy.

The supply side of the financial market is competitive with free entry. The firm F puts up its members' wealth $\omega_F \equiv \sum_{a \in F} \omega(a)$ as collateral, receives a loan L, and then carries out its production activities. Note that members are assumed to invest all their wealth in the firm; we could let them choose how much to invest, but under our assumptions, this would make no difference. Nor does it make any difference whether the firm provides collateral $C < \omega_F$ and borrows $L - (\omega_F - C)$; we will always simply assume the collateral is ω_F and not specify it in the contract.

Once the outcome of the project becomes known to the firm's members, they report it to the lender. If they report success, the firm receives an income of y_s . Since the firm typically has an incentive to report failure even if it succeeds (reporting success when there is failure is assumed infeasible), the lender will need to conduct random audits to insure truthful reporting.¹⁰ Let p be the probability of audit, y_f the firm's income if failure is reported and no audit is conducted, and y_t the income if the audit is conducted and the firm reports truthfully. Finally, y_l is the income if the firm is caught lying.¹¹

Of course, not any financial contract is admissible. Recall that the share contract and organization are known to the lender, so that the equilibrium success probability $\pi = \pi_{\#T}$ is also known (#T is the cardinality of the set T). In addition, the financial contract must satisfy the following constraints:

$$y_i \geq 0, \quad i = s, f, t, l \quad (\text{NN})$$

$$p \in [0, 1] \quad (\text{P})$$

$$\pi y_s + (1 - \pi)[p y_t + (1 - p) y_f] + (1 - \pi) p y_l \leq \pi R - L r + \omega_F r \quad (\text{NNP})$$

$$y_s \geq p y_l + (1 - p)(R + y_f) \quad (\text{TT})$$

The first constraint follows from the fact that each member of the enterprise must receive a nonnegative income and from the budget-balancing conditions of

¹⁰Of course, we are assuming that a certain amount of collusion among the firm's members is possible.

¹¹The implicit assumption that the lender commits to an auditing policy could be justified by a reputation argument; in this context, as distinct from that of whistleblower arrangements, the ability to establish a reputation seems plausible, since there is some chance that outsiders could observe whether the lender was adhering to its policy. Our results would be little affected by relaxing this assumption, however, since without commitment, the capital market would be even less efficient than we suppose.

the share contracts (see below). The second says that p is a probability. The third is a nonnegative profit condition on lenders' returns. The fourth is a truth-telling constraint which makes it in the firm's interest to report the true outcome of the project (if successful) rather than lie and keep the output. These feasibility conditions also arise in the optimal contract problems studied by Townsend (1979), Border-Sobel (1987) Bernanke-Gertler (1989) and have been used in a similar context to ours in Newman (1991).

The financial contracts actually used in equilibrium will have more structure than we have imposed thus far; in particular, it turns out that they maximize the firm's expected income. By solving the optimal contract problem, we can examine one of the two wealth effects more precisely. We defer further discussion of this issue to Section 5 on the characterization of equilibrium.

(b) *Share*

A share contract just distributes the proceeds of the project among the firm's members in such a way as to satisfy a budget balancing condition in every state of the world.¹² Henceforth, all contracts under consideration are assumed to satisfy the feasibility constraints (NN), (P), (NNP), and (TT) for financial contracts.

In the hierarchy, which uses the monitoring technology, the share contract is written contingent on the effort of members (except that of the manager), as well as on the firm's output. The partnership contract can only be written contingent on the output of the enterprise.

The firm F is partitioned into W , S , and M with $M = \emptyset$ when F is a partnership. Since the probability of success is bounded above by π_N and since there is no need to duplicate monitoring effort, in any equilibrium, the sets W and M must be finite, with $\#W \leq N$ and $\#M \leq 1$.

At this point, we cannot exclude the possibility that the set S is not finite. Below, we characterize the set of share contracts that are consistent with feasibility and incentive compatibility when firms have finite (indeed,

¹²Since the output is not supposed to be costlessly observable to outside parties, in particular the court, there may be a question of how any of these contracts can be enforced. However, since all members are symmetrically informed regarding output, we may imagine that if anyone attempts to take more than the contracted share, that other members may sue; the court can then discover the true output and require forfeiture of the defendant's share to the other members.

countable) membership. In the next section, we point out that since it is possible to show there is uniform finite bound on the size of firms and blocking coalitions, excluding all infinite firms from consideration is justified.

Denote by $x_s(a)$ the share received by agent a when there is success; $x_t(a)$ when his firm truthfully reports failure and is audited; $x_f(a)$ when failure is reported and there is no audit; and $x_l(a)$ when his firm is caught misreporting. A share contract is admissible (relative to a feasible financial contract (y,p,L) and organization (W,S,M)) if for each a in F , $x_i(a) \geq 0$ for $i = s, f, t, l$ and $\sum x_i(a) = y_1$. If c is a contract satisfying the financial contract constraints mentioned above, then an agent a earns *surplus* (i.e. utility over what he could achieve on his own just by choosing subsistence) equal to

$$u(a|c) \equiv \pi x_s(a) + (1-\pi)[p x_t(a) + (1-p)x_f(a)] - e(a) - \omega(a)r.$$

Of course, the only share contracts which are of interest are those which actually implement the organization (W,S,M) and generate the returns expected in the financial contract. We examine these feasibility restrictions for the partnership and the hierarchy separately.

(i) Partnership

Here, of course, the *prima facie* problem is that of the free-rider: once the contract is signed, each member typically gains less than he expends, so has no incentive to work at an efficient level (Alchian and Demsetz 1972; Holmström 1982). Now in fact, under some conditions, partnerships can achieve full efficiency (Williams-Radner 1988; Legros-Matsushima 1991) and under others, near-efficiency (Legros-Matthews, forthcoming). As we have said, the observed organizational form will result from a tradeoff among the costs of each, so whether we imagine partnerships to be fully efficient or only nearly efficient doesn't matter. In either case, efficiency doesn't determine the outcome. In the present model, in fact, a partnership may be able to achieve first-best efficiency (a hierarchy never does, since $N+1$ units of effort are required to produce $\pi_N R$, while the partnership uses only N).

If a partnership is to have n agents working, that is, $\#W = n$, then each worker faces the incentive compatibility constraint:

$$\begin{aligned} & \pi_n x_s(a) + (1-\pi_n)[p x_t(a) + (1-p)x_f(a)] - 1 \\ & \geq \pi_{n-1} x_s(a) + (1-\pi_{n-1})[p x_t(a) + (1-p)x_f(a)] \end{aligned}$$

or

$$\Delta\pi_n(x_s(a) - [px_t(a) + (1-p)x_f(a)]) \geq 1, \quad (\text{ICW})$$

where $\Delta\pi_n \equiv \pi_n - \pi_{n-1}$.

Consequently, there is a minimum surplus that a worker with wealth $\omega(a)$ must earn if she is to be incentive compatible in a firm with n working partners:

$$\underline{u}(\omega(a)) \equiv \frac{\pi_n}{\Delta\pi_n} - 1 - \omega(a)r. \quad (\text{MPS})$$

Finally, unless we can prevent silent partners from participating in the production process, a share contract that is to implement an effort level n must be designed in such a way that these agents do not have an incentive to exert effort, i.e.,

$$\begin{aligned} & \pi_n x_s(a) + (1-\pi_n)[px_t(a) + (1-p)x_f(a)] \\ & \geq \pi_{n+1} x_s(a) + (1-\pi_n)[px_t(a) + (1-p)x_f(a)] - 1, \text{ for all } a \in S \end{aligned}$$

or

$$\Delta\pi_{n+1}(x_s(a) - [px_t(a) + (1-p)x_f(a)]) \leq 1, \text{ for all } a \in S. \quad (\text{ICS})$$

As it turns out, this constraint will never bind in equilibrium: all silent partners strictly prefer not to work (see Lemma 2 in the Appendix.)

(ii) Hierarchy

A firm is now written $F=W \cup S \cup M$, where $\#M=1$. Because effort levels are observable, if the agent in M monitors, it is easy to punish agents in $W \cup S$ who deviate, that is, those in W who shirk and those in S who work. Indeed, the worst punishment that can be imposed on these agents is to give them a share of zero. Consequently, we might as well suppose that whenever one agent is detected deviating, the share of each agent in $W \cup S$ is zero. If there is monitoring (recall that the agents can observe when they are monitored) it is enough to consider the incentives for an agent in W to shirk (agents in S clearly cannot gain by exerting effort). Hence, the incentive compatibility conditions are

$$\pi_n x_s(a) + (1-\pi_n)[px_t(a) + (1-p)x_f(a)] \geq 1, \text{ for all } a \in W. \quad (\text{HICW})$$

Notice that this is a rather weak requirement, since it is implied by the individual rationality of the worker joining the hierarchy in the first place.

For managers, the situation is somewhat different, since monitoring is not contractible. The agents in $W \cup S$ know when they are being monitored, and therefore typically have incentives to shirk when the manager does. But if their shares happen to satisfy incentive compatibility conditions like (ICW) and (ICS), their response to shirking by the manager will be to work anyway. Thus, if the manager is not to shirk — that is, if we are to have a hierarchy

at all — it is necessary that enough of the incentive compatibility conditions (ICW) and (ICS) are violated by the hierarchy's share contract. Indeed, let π' be the probability of success when the manager b does not monitor. Then we need

$$\pi_n x_s(b) + (1 - \pi_n)[px_t(b) + (1-p)x_f(b)] - 1 \geq \pi' x_s(b) + (1 - \pi')[px_t(b) + (1-p)x_f(b)]$$

or

$$(\pi_n - \pi')(x_s(b) - [px_t(b) + (1-p)x_f(b)]) \geq 1. \quad (\text{ICM})$$

Note that (ICM) is the same as (ICW) only if $\pi' = \pi_{n-1}$. On the other hand, if all the agents in WUS receive noncontingent compensation (e.g. wages), then the (ICW) are violated, (ICS) satisfied, π' therefore is zero, and (ICM) is implied by the manager's individual rationality of joining the hierarchy. In any case, the manager always bears some risk.

Clearly, π' cannot be too large (e.g. $\pi' = \pi_n$), or b has no incentive to monitor. When the manager shirks, the agents in WUS behave as if they were in a partnership which earns incomes $\hat{y}_i = y_i - x_i(b)$. Consider partitions $\{W_m, S_m\}$ of WUS where

$$W_m \subseteq \{a \in WUS \mid \Delta\pi_m(x_s(a) - [px_t(a) + (1-p)x_f(a)]) \geq 1\}$$

and

$$S_m \subseteq \{a \in WUS \mid \Delta\pi_{m+1}(x_s(a) - [px_t(a) + (1-p)x_f(a)]) \leq 1\}.$$

The agents in WUS can implement π_m when they are not monitored if they can partition themselves into two such sets W_m and S_m satisfying $m = \#W_m$. The manager's constraint (ICM) will be hardest to satisfy if π' is chosen to be the largest of these implementable probabilities (call it $\pi_{m(F)}$, and we will impose this choice as a necessary condition for the feasibility of hierarchies.

To summarize, the feasible contracts are described in the following

Definition 1 For any countable coalition F, the set C(F) of feasible contracts is the set of tuples $c = (y, p, L, x, W, S, M)$ satisfying

(i) $x_i(a) \geq 0$ for $i = s, f, t, l$ and all $a \in F$;

(ii) the financial contract constraints ($\pi = \pi_n$ and $n = \#W$)

$$p \in [0, 1] \quad (\text{P})$$

$$\pi y_s + (1 - \pi)[py_t + (1-p)y_f] + (1 - \pi)p\gamma \leq \pi R - Lr + \omega_F r \quad (\text{NNP})$$

$$y_s \geq py_1 + (1-p)(R + y_f), \quad (\text{TT})$$

$$\text{where } \omega_F = \sum_{a \in F} \omega(a);$$

(iii) the budget balancing conditions

$$\sum_{a \in F} x_i(a) = y_i, \quad i = s, f, t, l; \quad (\text{BB})$$

(iv) the partnership constraints

$M = \emptyset$ and

$$\Delta\pi_n(x_s(a) - [px_t(a) + (1-p)x_f(a)]) \geq 1, \text{ all } a \in W \quad (\text{ICW})$$

$$\Delta\pi_{n+1}(x_s(a) - [px_t(a) + (1-p)x_f(a)]) \leq 1, \text{ all } a \in S; \quad (\text{ICS})$$

or

(iv') the hierarchy constraints

$M = \{b\}$ and

$$\pi_n x_s(a) + (1 - \pi_n)[px_t(a) + (1-p)x_f(a)] \geq 1, \text{ all } a \in W \quad (\text{HICW})$$

$$(\pi_n - \pi_{m(F)})(x_s(b) - [px_t(b) + (1-p)x_f(b)]) \geq 1, \quad (\text{ICM})$$

$$m(F) \equiv \max \{m\}$$

s.t. $\exists \{W_m, S_m\}$ a partition of $W \cup S$ satisfying

$$W_m \subseteq \{a \in W \cup S \mid \Delta\pi_m(x_s(a) - [px_t(a) + (1-p)x_f(a)]) \geq 1\}$$

$$S_m \subseteq \{a \in W \cup S \mid \Delta\pi_{m+1}(x_s(a) - [px_t(a) + (1-p)x_f(a)]) \leq 1\}.$$

Of course, the nonnegativity constraints (NN) on the y_i are implied by (i) and (BB).

4. Equilibrium of the Economy

Having defined feasibility for countable coalitions, we are almost ready for a formal definition of a competitive equilibrium. There is a small technical issue that needs to be taken care of first. Recall that we seek a partition of the set of agents. Because we have a continuum of agents, it is possible to find one-to-one maps from arbitrarily small sets of positive measure onto sets of large measure: we therefore must restrict ourselves to partitions which satisfy a measure-consistency criterion.¹³ Call a partition of A into finite sets *measure-consistent* if for all positive integers k and $i, j = 1, \dots, k$, the set of all i^{th} members of size- k elements of the partition has the same measure as the set of all j^{th} members of those elements (for a formal definition, see Kaneko-Wooders, 1986).

We now provide our definition of equilibrium.

Definition 2 An equilibrium (\mathcal{P}^*, c^*) is a measure-consistent partition \mathcal{P}^* of A into finite sets and a function c^* on \mathcal{P}^* such that

¹³For instance, if all agents in $[0, 1/(2N+1)]$ have K units of wealth (are rich) and the remainder have 0 (are poor), a reasonable candidate for equilibrium is that half of the poor agents join firms each of which has one rich and N poor agents, while the remaining poor are idle. Without measure consistency, however, it would be possible to have every poor agent join a firm with the same organization, i.e. one rich and N poor.

- (i) For almost every $F \in \mathcal{P}^*$, $c^*(F) \in C(F)$ is a feasible contract for F ;
- (ii) For all finite $T \subseteq A$ and feasible $c \in C(T)$, $u(a|c) \leq u(a|c^*(F))$ for some $a \in T$.

(We say that a property is true for almost every firm if it is true for all firms whose members are in a set of full measure.)¹⁴

The second condition is, of course, the core stability condition — deviating coalitions must make all of their members strictly better off in order to upset a putative allocation. Notice that this definition follows Kaneko-Wooders in ruling out the possibility of blocking by infinite coalitions (feasibility is not even defined for uncountable coalitions). In the present context, we feel this restriction is justified. Recall that a continuum economy is simply an approximation to a large finite economy. Consider a sequence of replications of a finite economy. Uncountable coalitions in the continuum economy (in particular those of positive measure) then correspond to positive fractions of the populations of the replica economies. On the other hand, it can be shown that there is a uniform finite upper bound \hat{n} on the size of "effective" blocking coalitions — if in a finite economy any coalition of size larger than \hat{n} blocks a putative allocation, there is a subcoalition no larger than \hat{n} which also blocks.¹⁵ This upper bound

¹⁴We shall actually be placing a further restriction on the equilibrium partition, namely that it satisfy "minimality." By this we mean that no element of the partition can be broken into subsets, each of which feasibly achieves the same surplus for all its members as in the original coalition. For instance, two separate projects are not considered to belong to the same firm. For our purposes, there is no loss of generality in adding this restriction.

¹⁵The formal proof is somewhat involved, and is omitted. But the intuitive argument is simple enough. The production and monitoring technologies dictate that there is never reason to have more than one manager or more than N workers. The only problem is finding a bound on the number of silent partners. It turns out that there is a finite bound $\bar{\omega}$ on the total amount of wealth that a firm ever needs (see expression (MPW+) in Section 5b; set $\bar{\omega}$ equal to the maximum $\underline{\omega}_{n+}$ over n such that $\Delta\pi_n$ is positive); above this level it gains nothing for its members that they could not achieve by investing in the safe asset or borrowing from the financial market. Thus any silent partners who add wealth over $\bar{\omega}$ or who add zero wealth are redundant in the same way that the $N+1^{\text{st}}$ worker or second manager are. Call the minimum positive wealth ω_0 (this exists because of the finite number of wealth levels). Thus n can be chosen equal to $N+1+\bar{\omega}/\omega_0$.

applies equally well to all finite economies (thus it is not even necessary to consider countably infinite coalitions). The same kind of argument shows that membership in all equilibrium firms will be uniformly bounded by \hat{n} (again, restricting attention to minimal partitions).

The first order of business is to establish that the equilibrium exists. The proof is a straightforward application of the results in the literature on the f -core. We simply state the result here and provide a sketch of the proof in the Appendix.

Proposition 1 An equilibrium exists.

We note that any equilibrium is constrained Pareto optimal because the grand coalition can only achieve what a disjoint union of finite coalitions can achieve; thus something that would be blocked by the grand coalition could actually be blocked by a finite coalition.

We now turn to an analysis of other properties of the equilibria of our model.

5. Some Characterizations

(a) Financial Contracts are Optimal

As we have said, it is not generally true that a firm's contracts will be surplus maximizing. As a corollary to this observation, it is not obvious that the financial contract will be income-maximizing for the firm: it may be possible that a financial contract which fails to maximize the firm's expected income could be adopted nevertheless, simply because its members are unable to distribute any extra surplus without violating some of the incentive compatibility constraints (ICW, ICS, ICM and HICW). This problem is potentially most severe in the hierarchy, where increases in the workers' surplus might increase the manager's incentive to shirk, thereby reducing the incomes of all members. Thus although we are essentially compelled by our assumptions on the verifiability of messages to look at financial contracts of the form (y,p,L) , there is no guarantee that the separation between share and financial contract is complete in the sense that the financial contract maximizes the firm's expected income.

Our first characterization result shows that, at least in the present

model, the separation is complete, that is, in a competitive equilibrium, any financial contract is income-maximizing with respect to the success probability implied by the share contract. Of course, the nonnegativity, nonnegative-profit and truth-telling constraints must be satisfied.

Proposition 2 Let (\mathcal{P}^*, c^*) be an equilibrium. For each $F \in \mathcal{P}^*$ and equilibrium contract $c^*(F) = (y, p, L, x, W, S, M) \in C(F)$, the financial contract (y, p, L) solves

$$\begin{aligned} & \max_{(\hat{y}, \hat{p})} \pi \hat{y}_s + (1-\pi)(p\hat{y}_t + (1-p)\hat{y}_f) \\ & \text{subject to (NN), (P), (NNP) and (TT).} \end{aligned}$$

The proof, which is in the Appendix, shows that any increase in the firm's income that satisfies the financial contract constraints can be translated into increases in each member's utility by finding transfers that do not violate anyone's incentive constraints.

In view of this result, we can provide a sharper characterization of the financial part of a firm's equilibrium contract. For the case $\omega_F < L$ straightforward calculation (see e.g. Bernanke-Gertler, 1989) shows that the (unique) solution to this program is:

$$y_f = y_t = y_1 = 0; \tag{F1}$$

$$y_s = (1-p)R; \tag{F2}$$

$$p = \min \left\{ \frac{(L-\omega_F)r}{\pi_n R - (1-\pi_n)\gamma}, 1 \right\} \tag{F3}$$

If auditing is costless, then one can always set $p = 1$ and (TT) will never bind. In this case one obtains a first-best financial contract (that is, one in which there are no agency costs from using external finance).

The first-best can also be achieved if $\omega_F \geq L$: F doesn't need external finance, and in any case can afford to satisfy (TT) with $p = 0$ by setting $y_s = R + (\omega_F - L)r/\pi$ and $y_f = y_t = y_1 = 0$. (Note that in either case when the first-best is achievable, there is typically a continuum of optimal contracts; the one we have specified, and will assume throughout, proves most convenient.)¹⁶

¹⁶ Provided of course, that y_s so defined is nonnegative. In view of the corollary to this proposition and assumption (T4), this is always true if $\pi = \pi_N$; things can be modified straightforwardly in case $R + (\omega_F - L)r/\pi < 0$, so we

Since the y_1 and p typically depend on n and ω_F , we will sometimes write $p(n, \omega_F)$ and $y_1(n, \omega_F)$ to reflect this dependence.

Given the imperfections in the capital market, it would be surprising if any firm borrowed more than K . Indeed, any borrowing in excess of K can only earn the market return r , but will cost at least this amount. Thus we have

Corollary 1 For all competitive equilibrium financial contracts (y, p, L) , there is no loss of generality in assuming $L = K$.

We prove this result in the Appendix.

(b) The Wealth Effects

As we pointed out above, there are two wealth effects in this model; both are commonly rooted in the lower bound on individual's utility. The first stems from the inefficiency in the capital market. Notice from (F3) that p is declining in the wealth of the enterprise: the less net borrowing there is by the firm, the smaller the incentive to misreport, so the auditing cost need not be borne as often.

As an index of the inefficiency of the financial market, define for $n \geq 1$, $\alpha_n \equiv \frac{\pi_n R}{\pi_n R - (1 - \pi_n)\gamma}$. Notice that this is equal to unity if and only if $\gamma = 0$ and increases with γ .¹⁷ And, since a greater number of working members implies higher expected output for the firm, α_n is decreasing in n (for $n < N$). The expected monetary surplus (i.e. net of effort costs) of the firm will then be

$$\begin{aligned} \pi_n R - \alpha_n K r + (\alpha_n - 1)\omega_F r, & \quad \text{if } \omega_F < K; & \text{(SNB)} \\ \pi_n R - K r, & \quad \text{if } \omega_F \geq K. & \text{(SSF)} \end{aligned}$$

(SNB) indicates the surplus for a net borrower, while (SSF) describes the level of surplus achievable when the firm can self-finance and avoid agency costs in the financial market. The (net of subsistence) value of extra wealth

ignore this complication in what follows.

¹⁷Of course, if $\pi_n R$ is too small relative to γ , then the set of feasible financial contracts is empty, and the financial market shuts down for any firm with only n working members and $\omega_F < K$. A natural parameter restriction, then, is that $\pi_N R > (1 - \pi_N)\gamma$.

is positive only when there is net borrowing: in that case, extra wealth lowers the cost of external finance. Note how the nonnegativity constraints on individuals (which by budget balancing lead to the firm's nonnegativity constraint), along with the truth-telling constraint, together generate the inefficiency. If only y_1 (more generally, agents' utility) could be set sufficiently low, auditing could be conducted and its costs incurred with negligible frequency.

Those firms and individuals which have low wealth face higher marginal capital costs than their wealthier counterparts. The wealthy are therefore in a position to earn rents that they could not extract without the capital market imperfection. As we shall see, the ability to do so depends on supply and demand conditions in the labor market, which in turn depend on the distribution of wealth.

The second wealth effect arises from the incentive compatibility constraints in the partnership. A necessary condition for incentive compatibility is obtained by summing the n conditions (ICW) and observing that by budget balancing, the maximum value of $\sum_{a \in F} (x_s(a) - [px_t(a) + (1-p)x_f(a)])$ is equal to $y_s(n, \omega_F)$. Consequently,

$$\Delta \pi_n y_s(n, \omega_F) \geq n \quad (\text{ICP})$$

From the results in the previous subsection characterizing optimal financial contracts, we have

$$y_s(m, \omega_F) = R + (\omega_F - K)r/\pi_n, \quad \text{if } \omega_F \geq K.$$

$$y_s(m, \omega_F) = R + (\omega_F - K)\alpha_n r/\pi_n, \quad \text{if } \omega_F < K.$$

Substituting each equality into the incentive condition (ICW), we obtain

$$\Delta \pi_n (R + (\omega_F - K)r/\pi_n) \geq n, \quad \text{if } \omega_F \geq K, \quad (\text{ICP+})$$

$$\Delta \pi_n (R + (\omega_F - K)\alpha_n r/\pi_n) \geq n, \quad \text{if } \omega_F < K \quad (\text{ICP-})$$

There is a minimum wealth at which each inequality can be satisfied:

$$\underline{\omega}^{n+} \equiv K - [R - \frac{n}{\Delta \pi_n}] \frac{\pi_n}{r}, \quad \text{if } \omega_F \geq K, \quad (\text{MPW+})$$

$$\underline{\omega}^{n-} \equiv K - [R - \frac{n}{\Delta \pi_n}] \frac{\pi_n}{\alpha_n r}, \quad \text{if } \omega_F < K. \quad (\text{MPW-})$$

Observe that if $R < n/\Delta \pi_n$, (ICP-) cannot be satisfied for any $\omega_F < K$, and that (ICP+) can be satisfied only for ω_F sufficiently greater than K . If $n/\Delta \pi_n \leq R$, then (ICP+) is always satisfied, and (ICP-) can be satisfied for some wealth level less than K .

We therefore find a second wealth effect, one which arises only in the partnership: if the agents in the firm do not have at least the level defined in (MPW+) or (MPW-) among themselves, they cannot satisfy incentive

compatibility. Note that when $\omega_F < K$ and $n/\Delta\pi_n \leq R$, this minimum wealth level is increasing in γ (i.e. in the cost of capital market transactions), but that even if auditing is costless ($\gamma = 0$), it is typically positive: partnerships consisting of very poor agents will generally not be feasible, particularly if capital markets are imperfect.

What we might expect of this model, then, is that partnerships will usually not consist only of poor agents. In order to be incentive compatible, partnerships will have to contain a sufficient number of the economy's wealthy. The interesting thing is that we cannot be sure that these rich agents *will* form partnerships, even if they can: if there are enough poor agents around, a rich person might prefer to form a hierarchy with them (their compensation will be bid down) than to form a partnership, in which their surplus will typically have to be higher in order to satisfy incentive compatibility ((MPS) provides the lower bound for a worker in a partnership, while, depending on supply and demand, 0 may be the lower bound for a worker in a hierarchy).

(c) *Efficiency*

We define the total surplus $V(F)$ generated by a firm F in which the total wealth is ω_F by

$$V(F) \equiv \pi_n y_s(n, \omega_F) + (1 - \pi_n)[p y_t(n, \omega_F) + (1 - p) y_f(n, \omega_F)] - \omega_F r - [n + \#M(F)]$$
if there is x such that $(y(n, \omega_F), p, K, x, W, S, M) \in C(F)$ and $n = \#W$; and $V(F) = 0$ otherwise. Obviously, in equilibrium, $V(F) \geq 0$ for all F .¹⁸

We now provide a precise efficiency criterion for evaluating organizational forms.

Definition 3 The firm F is *surplus inefficient* (or simply *inefficient*) if there is a subset $G \subseteq F$ such that $V(F) < V(G)$.¹⁹

¹⁸We do not impose the basic individual rationality condition $u(a|c) \geq 0$ when defining feasible surplus maps for F . That individual rationality condition must be satisfied in equilibrium since a would otherwise prefer to use her subsistence option. Hence, in any equilibrium firm with contract c , $u^F(a|c) \geq 0$ and $V(F) \geq 0$. Observe that in general there may be feasible contracts for F which generate $V(F) < 0$.

¹⁹Milgrom and Roberts (1992) refers to this criterion as "technical efficiency."

For instance, if a partnership and a hierarchy each produce the same expected output (and therefore have the same number of active workers), the hierarchy is inefficient because it uses one extra unit of effort in monitoring. This criterion is weak in the sense that it is clearly necessary but hardly sufficient for the maximization of total social surplus. The existence of a surplus inefficient firm is not ruled out by our definition of equilibrium because even though $V(G)$ might exceed $V(F)$, it need not be the case that each member of G receives more in G than she does in F . In fact, it is precisely this nontransferability of utility, that is, the nonseparation of distribution and efficiency, that is at issue here and leads to inapplicability of the surplus efficiency criterion for predicting the organizational outcome of equilibrium.

As a benchmark, let us examine what happens if capital markets are "perfect." In this case, the argument in Alchian-Demsetz (1972) that each type of organization arises when it is the better means for providing incentives to labor seems to be correct: hierarchy occurs in equilibrium only if partnership cannot be feasibly maintained. We have

Proposition 3: Suppose that auditing is costless ($\gamma = 0$). Then in equilibrium, almost every firm is surplus efficient.

Proof Appendix.

With perfect capital markets, any agent can borrow enough to belong to a hierarchy which produces the first-best output $Y_N \equiv \pi_N R - K r$; thus every agent, no matter how poor, is guaranteed at least $Y_N / (N+1)$. The leading candidate for an inefficient firm is a hierarchy with wealth at least K , since this could generate a higher surplus if reorganized as a partnership. But no agent could gain more than $Y_N / (N+1)$ in such a firm, and would do better to form a partnership where he can obtain Y_N / N .

So with well-functioning capital markets, efficiency in our sense does obtain. It is easy to see why surplus maximization should be expected whenever one uses the partial equilibrium approach, particularly in the presence of only one market imperfection: even with the partnership wealth effect, surplus

for a given set of agents in a firm is maximized by competitive equilibrium.²⁰ It appears that the interaction of several imperfect markets is necessary to bring about failures of surplus efficiency.

(d) *Distribution and Organization*

Consider now what happens when auditing is more costly. Now the determination of organizational form is quite different. Efficiency does not dictate what forms prevail, nor are those that survive necessarily efficient. In these circumstances, organizations can become ways to transfer surplus from some agents to others, possibly at a social cost. Equilibrium organizations might better be understood as reflections of the distribution of initial resources than as surplus maximizing solutions to incentive problems.

We begin illustrating these points with two results that link the distribution of wealth to the organization and efficiency of the equilibrium firms, and elaborate further with the example in the next section. To simplify the analysis, we will use the following additional assumptions (we suppose that $N \geq 3$):

- (A1) $\Delta\pi_N > \Delta\pi_n$ for $n \neq N$.
- (A2) $\pi_N/(N+1) > \pi_n/n$ for $n < N$
- (A3) $N/(N-1) < \pi_N/\Delta\pi_N$
- (A4) $R - Kr/\pi_N > N/\Delta\pi_N$
- (A5) $\pi_N R - \alpha_N Kr < N$

None of these conditions is particularly weak, but as we are interested mainly in providing examples of how organization within the firm may depend on distribution in the whole economy, there is nothing especially distasteful about adopting strong assumptions. As we show in Lemma 3 below, (A1) and (A2) imply that any equilibrium hierarchy consists of N workers. (A3) puts an upper bound on $\Delta\pi_N$ and implies that the wage bill in an N -worker hierarchy can be lower than that in an N -worker partnership. Assumption (A4) implies that a firm with wealth at least K can feasibly be organized as a partnership with N workers, thereby achieving first-best efficiency. (A5), along with (T3) and

²⁰In general, total social surplus may be increased by redistributing initial wealth if, for instance, a group of agents which had formed an efficient hierarchy is induced to form a partnership instead.

the fact that α_n is decreasing in n , implies that any firm must have a positive initial wealth to generate a positive surplus; this is essentially a restriction on the efficiency of the capital market.

We show below that despite (A4), and despite the fact that partnerships may be more efficient than hierarchies, inefficient hierarchies must arise in equilibrium as long as the distribution of wealth is skewed enough to the left.

The intuition for these results rests on supply-and-demand reasoning: when the distribution is skewed to the left, poor agents compete against each other to bid down wages, making hierarchies more attractive to the wealthy than partnerships. To facilitate the analysis, the following result is useful:

Lemma 3 If (A1) and (A2) hold, then $\#W = N$ for almost all equilibrium hierarchies.

The proof, given in the Appendix, shows that under (A1) and (A2), hierarchies with $\#W$ less than N always gain by adding some workers. If a small firm was viable, then (A1) and (A2) insure that its output would increase by enough to cover the effort cost when the extra workers are added. Since the workers are paid for performance, the income transferred to them does not depend on the initial size of the hierarchy, so it will be in the interest of all members to expand.

Given this result, we can now illustrate the dependence of organization on distribution by providing a condition which guarantees that in equilibrium, only partnerships arise. Indeed, under this condition, the allocation is first-best.

Proposition 4 Suppose (A1), (A2) and (A4) and that $H(K/N) < (N-1)(1-H(K))$. Then almost all equilibrium firms are partnerships of size N with a total wealth of at least K and are surplus efficient. Moreover, almost every agent has the same surplus.

Proof Appendix.

The condition in Proposition 4 says roughly that not too large a fraction of the population is "poor" in the sense of owning less than $1/N$ of the capital

requirement for a project. The essential idea is that under the specified condition, there is excess demand for agents with wealth less than K/N . This implies that the agents with wealth above K will compete to form firms with those agents. This competition ultimately leads to an equalization of the surpluses of the agents in the economy.

On the other hand, if the distribution is skewed to the left, i.e., if there are relatively more poor than rich agents in the economy, and if γ is large enough, then we can show that inefficient forms of organization must arise in equilibrium. We need some additional notation. Define

$$\begin{aligned} n^* &\in \operatorname{argmax} \{ \pi_n R - Kr - (n-1)\pi_n / \Delta \pi_n \} \\ \alpha^* &= \max \{ \alpha_n \mid \pi_n \geq \gamma / (R + \gamma) \} \\ \omega^* &= \min \{ K / (N+1), [(\pi_N - \pi_n^*)R + (n^* - 1)\pi_n^* / \Delta \pi_n^* - N] / (N\alpha^* r) \} \end{aligned}$$

The interpretation is that n^* is the partnership size that an agent with wealth K would join if her partners all received the minimum incentive compatible compensation. The maximum marginal cost of external finance compatible with the existence of financial contract is given by α^* . The wealth level ω^* turns out to be the lowest at which an agent could ever be admitted as a silent partner. In the proof of the following Proposition, we show, using (A2) and (A3), that ω^* is strictly positive.

Proposition 5 Suppose (A1)-(A5), that $H(\omega^*) > (N+1)[1-H(\omega^*)]$ and that $H(K) < 1$. Then in equilibrium, a positive measure of agents belong to inefficient hierarchies.

Proof Appendix.

Under the assumptions of Proposition 5, we can show that the maximum demand for "poor" agents (those with wealth less than ω^*) by each "wealthy" agent (wealth greater than ω^*) is bounded above by $N+1$. Therefore, in any equilibrium partition, there is a positive measure of poor agents who are not matched with wealthy ones. Condition (A5) implies that the cost of borrowing is too high for firms consisting only of poor agents to generate a nonnegative surplus. It follows that in equilibrium, there will be a positive measure of poor who have zero surplus. But then agents with wealth more than K will want to form hierarchies because they can offer an expected compensations of $1+\epsilon$ to N poor agents and be better off than by forming a partnership and having to

offer n^*-1 expected incomes of at least $\pi_n^*/\Delta\pi_n^*$.

6. An Example

We now consider a simple example in which we can completely characterize the organizations in the economy as a function of wealth distribution. Suppose that $N = 3$, $\gamma > 10$, $\pi_0 = 0$, $\pi_1 = 1/15$, $\pi_2 = 1/5$, $\pi_3 = 1/2$, $R = 14$, $r = 1$, $K = 1$. Consider distributions with support $\Omega = \{0,1,2\}$ and masses h^0 , h^1 , h^2 . Notice that the minimum compensation that a monitored worker would accept is 1. It can be checked that $\alpha_3 > 3.5$, $\omega^{3-} = 1-2/\alpha_3$ is positive, $\omega^* = 1/126\alpha_3 > 0$, partnerships of size 1 and 2 violate (ICP-), and that hierarchies with zero wealth generate a negative surplus (indeed, all of (A1)-(A5) are satisfied). Consequently, any agent with wealth 0 must form a firm with at least one agent who has positive wealth. If he forms a partnership with someone, condition (MPS) implies that he receives an expected compensation of at least $5/3$.

Now we can ask whether an agent with wealth 1 or 2 would prefer to enter a hierarchy or a partnership. First, there is no reason for an agent with positive wealth to be a silent partner; indeed, either in a partnership or a hierarchy, an agent with wealth 1 might as well replace a partner or the manager and appropriate the positive surplus that these agents must obtain in order to satisfy incentive compatibility. Notice that in either case, net borrowing is not necessary, so we can assume $y_s - y_f = R = 14$. Suppose workers get v . Then the rich person prefers the hierarchy if $3v < 10/3$, i.e. $v < 10/9$.

All distributions are parameterized by h^1 and h^2 , so we can represent them as in Figure 2. Three regions are distinguished there: (I) "unequal" distributions with $h^0 = 1-h^1-h^2 > 3(h^1+h^2)$; (II) "equal" distributions with $h^0 < 2(h^1+h^2)$; (III) "middling" distributions in between. Observe that all distributions lying along a segment of the form $h^1+2h^2 = \mu$ have the same mean (namely μ) and in particular that there are many such segments intersecting all three regions. (We chose an example with three rather than two wealth levels in order to allow variations in distributions with common means — organization then truly depends on the way wealth is distributed rather than merely on the aggregate amount of it in the economy).

Consider region I. If every rich agent (wealth 1 or 2) sets up a hierarchy, the demand for workers is no more than $3/4$. The supply is at least this amount since none of the poor could demand labor or form partnerships on

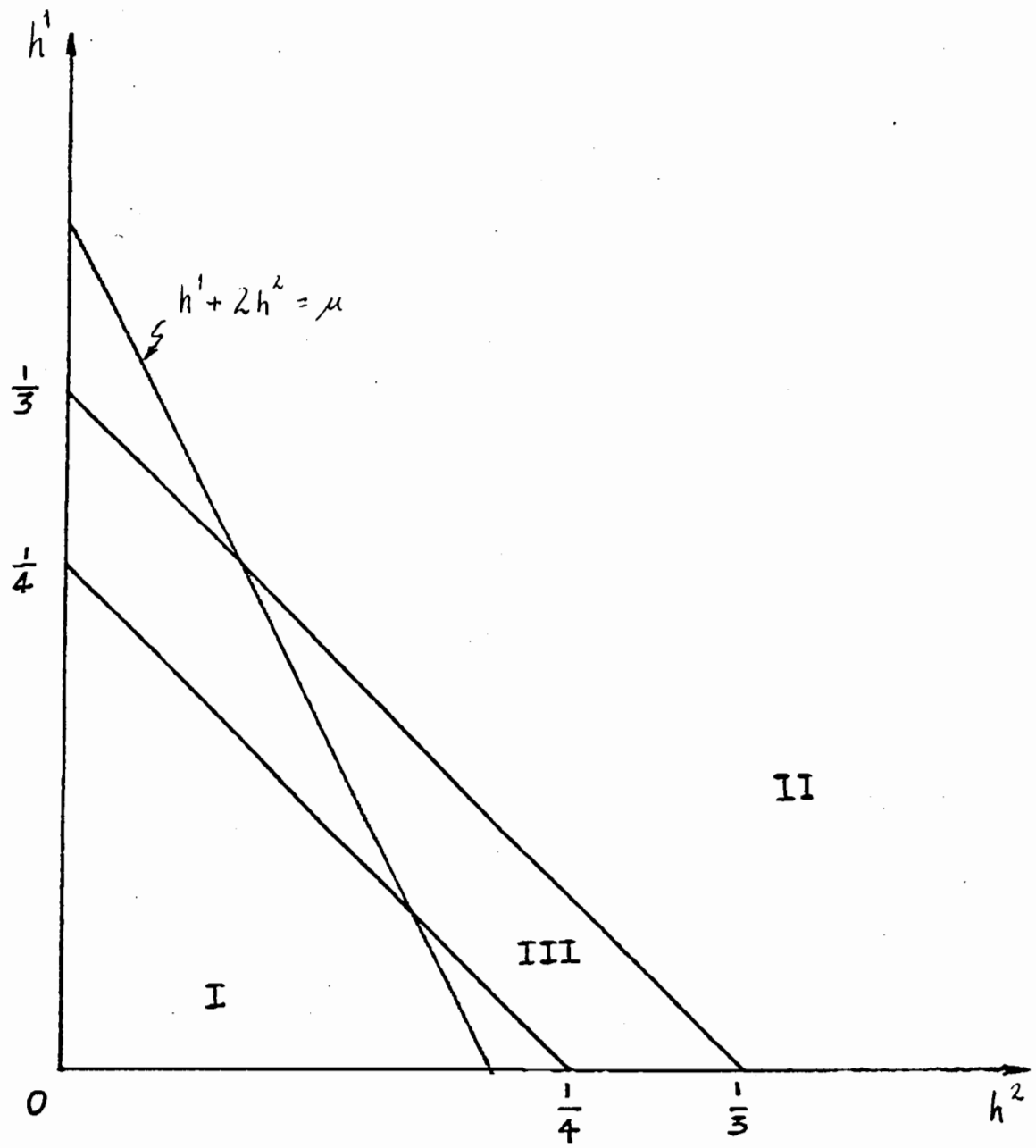


Figure 2

their own. So the wage would be bid down to its minimum of 1. This is an equilibrium, since at this wage, every rich agent does indeed strictly prefer monitoring three poor agents to forming a partnership with two of them.

Can there be any other equilibria for distributions in this region? The answer is no. Suppose there were a positive measure of partnerships. Then there must be some poor around who are idle (this is generically true even if there are only hierarchies). A rich partner could offer them slightly more than 1 to be monitored, thereby making both them and himself better off.

Thus, when the distribution of wealth is unequal enough, only hierarchies form, even though the surplus generated by each firm could be increased by taxing the use of the monitoring technology, taxing high incomes, or otherwise encouraging the use of partnerships. Observe that this increased surplus can be achieved *without* redistribution of initial wealth.

On the other hand, in region II, we know from Proposition 4 that only partnerships form. Indeed if any hierarchies formed, there would be excess demand for workers and the wage would be bid up to at least the minimum partnership compensation of $5/3$. But at this wage all rich agents strictly prefer to be partners. In this case, with the more equal distributions, the form of organization which results maximizes social surplus.

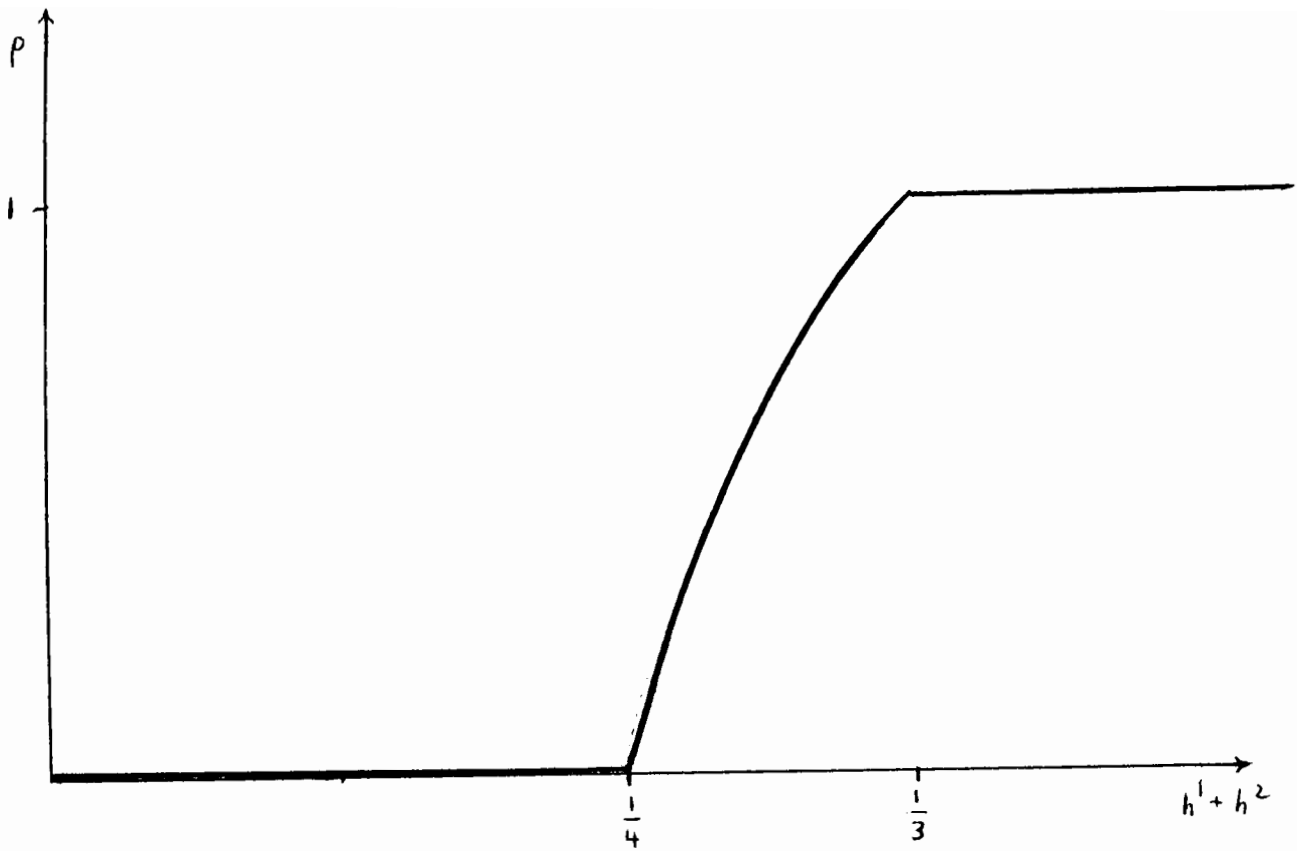
Finally in region III, a mixture of both types of firms obtains (This shows that the condition of Proposition 5 is sufficient but not necessary.) We can be much more specific. Let ρ denote the equilibrium fraction of partnerships in the economy. If ρ is positive, there can be no poor agents who choose subsistence: if there were, a rich agent who belonged to a partnership could form a hierarchy by offering them a wage of slightly over 1. On the other hand, we cannot have $\rho = 0$ since $h^0 < 3(h^1+h^2)$, i.e. there will be rich agents unmatched who would prefer to lure some workers away by offering them partnership positions. Thus the fraction of partnerships must satisfy

$$\rho 2(h^1+h^2) + (1-\rho)3(h^1+h^2) = 1-h^1-h^2,$$

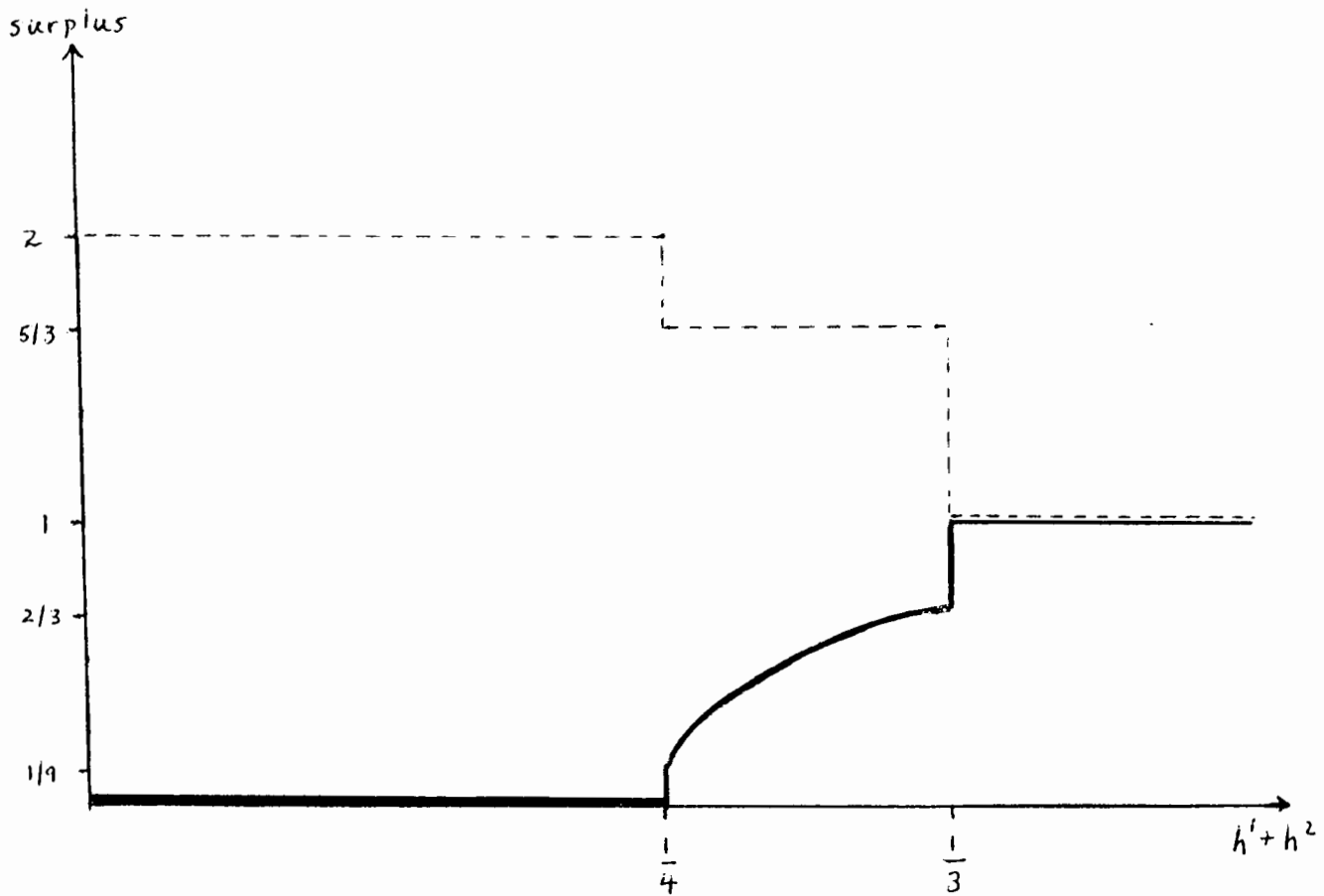
or

$$\rho = 4 - \frac{1}{h^1+h^2},$$

which is (weakly) monotonic in h^1+h^2 ; see Figure 3(a). Moreover, the equilibrium organizations are surplus efficient only when $h^1+h^2 \geq 1/3$. For all other values of h^1+h^2 the presence of hierarchies precludes surplus maximization, even though it would be feasible to have partnerships alone



(a)



(b)

----- 1's and 2's
 ———— 0's

Figure 3

simply by banning the use of the monitoring technology.²¹

Consider now the relationship between the distribution of surplus and the distribution of initial wealth, as illustrated in Figure 3(b). The surplus obtained by each type of agent have been calculated for the different wealth distributions. The figure reflects an indeterminacy in the equilibrium compensations at $1/4$ and $1/3$. In the interval $[1/4, 1/3]$ the curve representing the surplus of agents with wealth 0 is an average over all agents of this type, since ρ of them end up in partnerships and receive $2/3$ while the remainder join hierarchies and receive $1/9$. Clearly, as the wealth distribution becomes more equal, so does the surplus distribution. This contrasts to the first-best economy in which everyone would receive the same surplus of 1 (though not, of course, the same income) regardless of the distribution: with perfect capital markets, only partnerships would form.

This figure also illustrates the fact that the organization of the firm may serve essentially as a transfer device: if the hierarchies in region I were prevented from forming, some of the poor agents would receive a surplus of $2/3$ (if they became partners) while the rest would continue to receive zero (if idle). The wealthy agents' surplus would fall from 2 to $5/3$. Note that the monitoring technology and the capital market imperfection act in tandem to permit the wealthy to transfer surplus to themselves under competitive conditions: in the absence of either, only partnerships form, and poor agents get higher surplus at the expense of the wealthy.

6. Discussion

By avoiding consideration of the general equilibrium aspects of the problem, a theory of organization based solely on the efficiency criterion can generate conclusions that are incomplete and potentially misleading. There is nothing perverse about the existence of evidently inefficient organizational forms, for that is perfectly consistent with competitive equilibrium.²² As we have shown, efficiency, in the sense of maximum social surplus, cannot explain observed organizations. Indeed, it may be better to view the efficiency of

²¹This is not a recommended policy.

²²It may be fair to say, however, that if an inefficient organizational form persists, *someone* is benefiting by more than would be possible for that person under another organizational form.

organizations as an empirical question rather than a theoretical principle.

There are numerous ways in which the model could be extended. We mention just a few. A point we have not emphasized so far concerns occupational determinacy. In general, our model does *not* predict, for instance, that rich agents become managers, although it is true that (ICM) will tend not to bind as strongly for them as for the poor. Often, there is nothing preventing some reasonably wealthy agents from hiring a poor person to monitor them. Of course, the stylized facts seem to suggest a correlation between wealth and occupation. The implications of the hierarchy incentive constraints need to be explored further, not only to see if they can lead to a theory of occupational choice, but also to develop predictions about the distribution of income within the firm.

Nor have we discussed ownership and/or control in the sense of Grossman-Hart (1986) and Hart-Moore (1990). Ultimately, this is the direction we want to take. But our results should suggest that in that more complicated setting the same principle should apply: the ownership and boundaries of the firm will not be determined only by efficiency, but also by distribution.

A third consideration concerns endogenizing the interest rate by allowing it to depend on the supply of and demand for capital within the economy. Something like this is attempted in Newman (1991), where it is shown that the dependence of organizational form on distribution still holds. We expect that in the present model, both the form and efficiency of organizations would continue to display this dependence.

We have not dealt with risk aversion. This is partly for ease of analysis. Certainly insurance issues will affect the types of compensation schemes that are feasible and may have a bearing on the distribution of income within the firm. But several authors (e.g., Williamson 1985) have expressed skepticism about the relevance of risk attitudes in the determination of organizational form. Moreover, as shown in Newman (1991), their effect is quite likely ambiguous.

We could also explore the determination of firm size more thoroughly. Although there is a single efficient size for the firm, there is no necessity in general for equilibrium firm size to be equal to this scale without assumptions like (A1) and (A2). A different, but related approach to this question is pursued in Banerjee-Newman-Qian (1990).

Finally, we have only considered a very simple set of production and information gathering technologies. Ordinarily, not much of interest would be

added by looking at more general technologies (say those in which the capital stock is adjustable), but an interesting feature of the partnership is that in contrast to the hierarchy, its ability to achieve efficiency seems to be sensitive to the technology (Legros 1989; Legros-Matsushima 1991; Legros-Matthews, forthcoming). Also, since more complex technology will tend to add additional truth-telling and incentive-compatibility constraints, it is possible that analogs of Proposition 2 will no longer be true. Again, since our goal here is the weak one of pointing out that the general equilibrium approach to the theory of organization is necessary; we are a long way from its full development, and the consideration of a broader class of technologies is certainly an ingredient of a general theory.

APPENDIX

Throughout this Appendix, Y_n denotes $\pi_n R - Kr$, the net expected output of a project with n workers.

(a) Existence

Here we sketch the proof of existence of an equilibrium for our economy. Let \mathcal{F} be the set of finite subsets of A . We will construct a characteristic function \hat{U} which associates to each finite coalition F a subset of $\mathbb{R}^{\#F}$. \hat{U} will satisfy the properties of Theorem 1 of Kaneko-Wooders (1986) and therefore the f -core of $(A, \mathcal{F}, \hat{U})$ is nonempty. Existence of an equilibrium for our economy will follow.

Consider a partition of $F \in \mathcal{F}$ into three sets W, S and M , where $\#M \in \{0, 1\}$. For such a partition, let

$$C(F|W, S, M) \equiv \{c \in C(F) \mid c = (y, p, L, x, W, S, M)\}$$

be the set of feasible contracts for which the organization of F is given by (W, S, M) . Let

$$U(C(F|W, S, M)) \equiv \{u \in \mathbb{R}^{\#F} \mid \exists c \in C(F|W, S, M) \text{ s.t. } u = (u(a|c))_{a \in F}\}$$

be the corresponding set of surplus vectors. From Definition 1 it is clear that $U(C(F|W, S, M))$ is closed in $\mathbb{R}^{\#F}$.

Let $U(F) \equiv \bigcup U(C(F|W, S, M))$ where the union is taken over all partitions (W, S, M) of F in which $\#M \in \{0, 1\}$. Since F is finite, $U(F)$ is the finite union of closed sets, hence is closed. Let $\hat{U}(F) \equiv U(F) - \mathbb{R}_+^{\#F}$ be the comprehensive extension of $U(F)$. $\hat{U}(F)$ is also closed in $\mathbb{R}^{\#F}$.

Let $A^i \equiv \{a \in A \mid \omega(a) = \omega^i\}$ be the set of agents with wealth ω^i , $i = 1, \dots, J$. $\{A^i\}$ is a finite partition of A .

We first establish that \hat{U} is a characteristic function, that is, it satisfies the following:

- (i) $\hat{U}(F)$ is a nonempty, closed subset of $\mathbb{R}^{\#F}$, for all $F \in \mathcal{F}$
- (ii) $\hat{U}(F) \times \hat{U}(G) \subseteq \hat{U}(F \cup G)$, for all $F, G \in \mathcal{F}$, $F \cap G = \emptyset$
- (iii) $\inf_{a \in A} \sup \hat{U}(\{a\}) > -\infty$, for all $a \in A$,
- (iv) $\forall F \in \mathcal{F}$, $u \in \hat{U}(F)$, $u' \in \mathbb{R}^{\#F}$, $u' \leq u \Rightarrow u' \in \hat{U}(F)$
- (v) $\forall F \in \mathcal{F}$, $\hat{U}(F) \setminus \bigcup_{a \in F} [\text{int}(\hat{U}(\{a\}) \times \mathbb{R}^{\#F-1})]$ is nonempty and bounded.

Conditions (i), (ii) and (iv) follow from the construction of \hat{U} . We note that for any a in A , $\hat{U}(\{a\}) = (-\infty, 0]$. This proves (iii) and establishes that $\hat{U}(F) \setminus \bigcup_{a \in F} [\text{int}(\hat{U}(\{a\}) \times \mathbb{R}^{\#F-1})]$ is nonempty. By incentive compatibility, the

minimum surplus of an agent in F is $-\omega(a)r$ (i.e., when a invests her wealth in the firm and is compensated only for her effort). Therefore, the maximum that an agent can obtain is $Y_N - N + \omega_{F \setminus \{a\}} r$. Consequently, each $\hat{u} \in \hat{U}(F)$ is bounded above by the vector $\bar{u}^F = (\omega_{F \setminus \{a\}} r)_{a \in F}$. It follows that $\hat{U}(F) \setminus \cup_{a \in F} [\text{int}(\hat{U}(\{a\}) \times \mathbb{R}^{\#F-1})]$ is bounded.

$(A, \mathcal{F}, \hat{U})$ is a game without side payments. Let Π be the set of measure-consistent partitions of A and for $\mathcal{P} \in \Pi$ we denote by $P(a) \in \mathcal{P}$ the coalition to which a belongs. Let $L(A, \mathbb{R})$ be the set of measurable functions from A to \mathbb{R} ; for $v \in L(A, \mathbb{R})$ and $F \in \mathcal{F}$, v^F is the restriction of v to F . Define the following sets:

$$H(\mathcal{P}) \equiv \{v \in L(A, \mathbb{R}) \mid v^{P(a)} \in \hat{U}(P(a)), \text{ a.e. } a \in A\};$$

$$H \equiv \bigcup_{\mathcal{P} \in \Pi} H(\mathcal{P});$$

$$H^* \equiv \{v \in L(A, \mathbb{R}) \mid \exists \{v^k\} \subset H, v^k \rightarrow v\}, \text{ where the convergence is in measure.}$$

If $v \in H^*$, then $F \in \mathcal{F}$ can improve upon v if for some $\hat{u} \in \hat{U}(F)$, $\hat{u}(a) > v(a)$ for each $a \in F$. $C_f(\hat{U})$, the f -core of $(A, \mathcal{F}, \hat{U})$, consists of those elements of H^* that cannot be improved upon by any finite coalition.

Theorem 1 of Kaneko-Wooders (1986) establishes the nonemptiness of the f -core of $(A, \mathcal{F}, \hat{U})$ when a simple condition, called *per capita boundedness*, is satisfied. We shall need the following definitions:

Definition A1 $(A, \mathcal{F}, \hat{U})$ has the r -property with respect to $\{A^i\}_{i=1}^J$ if for any $F \in \mathcal{F}$, for any $i=1, \dots, J$, and for any $a, b \in A^i$,

(1) if $a \notin F, b \in F$, then $\hat{u} \in \hat{U}(F \cup \{a\}) \Leftrightarrow \exists \hat{u}' \in \hat{U}(F \cup \{b\})$ s.t.

$$\hat{u}(d) = \hat{u}'(d), \text{ all } d \in F \text{ and}$$

$$\hat{u}(a) = \hat{u}'(b);$$

(2) if $a, b \in F, \hat{u} \in \hat{U}(F)$ then $\hat{u}' \in \hat{U}(F)$ where

$$\hat{u}(d) = \hat{u}'(d) \text{ for all } d \in F \setminus \{a, b\}$$

$$\hat{u}(a) = \hat{u}'(b) \text{ and } \hat{u}(b) = \hat{u}'(a)$$

This condition says that any two agents of the same type are substitutes. It is clear that our $(A, \mathcal{F}, \hat{U})$ has the r -property.

Definition A2 For any $F \in \mathcal{F}$, a payoff vector \hat{u} has the *equal treatment property* if $\hat{u}(a) = \hat{u}(b)$ for all $a, b \in F \cap A^i$, all $i=1, \dots, J$.

Definition A3 \hat{U} is *per-capita bounded* with respect to $\{A^i\}_{i=1}^J$ if there is a $\delta \in (0, 1)$ and a $Q \in \mathbb{R}$ such that if $F \in \mathcal{F}$ is a coalition satisfying

$$(1+\delta)h^i \geq \#(F \cap A^i) / \#F \geq (1-\delta)h^i$$

and if $\hat{u} \in \hat{U}(F)$ has the equal treatment property, then $u(a) < Q$ for all $a \in F$.

We need to establish that \hat{U} is per-capita bounded with respect to $\{A^i\}$. We argued above that if $\hat{u} \in \hat{U}(F)$, then $u \leq \bar{u}^F$, where $\bar{u}^F(a) = \frac{Y_N - N + \omega_{F \setminus \{a\}} r}{n^i}$. Clearly, as more agents are added to a given F , $\bar{u}^F(a)$ increases. Per-capita boundedness requires that equal treatment payoff vectors feasible for F are uniformly bounded above, but only for those finite coalitions in which the relative proportions of types are close to the corresponding proportions in the whole economy. .

Let $\delta = 1/2$. For $F \in \mathcal{F}$, let $n^i \equiv \#(A^i \cap F)$. We consider finite coalitions such that $3h^i/2 \geq n^i / (\sum_{j=1}^J n^j) \geq h^i/2$, i.e., the proportion of types in F is δ -close to the proportion of types in A . Note that for all such F , $n^i > 0$ for each i .

Let $u \in \hat{U}(F)$ have the equal treatment property. In order to find the maximum that an agent in $A^i \cap F$ can obtain, it is enough to find the maximum total surplus for agents in $A^i \cap F$ and divide by n^i . From Definition 1, incentive compatibility implies that the maximum total payoff of the set of agents in $A^i \cap F$ is $Y_N - N + \sum_{j \neq i} n^j \omega^j r$. Therefore, if the payoff vector satisfies equal treatment, the maximum payoff to $a \in A^i \cap F$ is $(Y_N - N) / n^i + \sum_{j \neq i} n^j \omega^j r / n^i$. From our assumption on the proportions of types in F , one can show that $n^j / n^i \leq 3h^j / h^i$. Therefore, $a \in A^i \cap F$ has a payoff bounded above by $(Y_N - N) / n^i + \sum_{j \neq i} 3h^j \omega^j r / h^i$, which is no greater than $Q^i \equiv \frac{Y_N - N + \sum_{j \neq i} 3h^j \omega^j r}{h^i}$. Clearly, $Q^i < \infty$. Choosing $Q = \max \{Q^i\}$ then ensures that \hat{U} satisfies per-capita boundedness.

Theorem 1 of Kaneko-Wooders (1986) now tells us that there exists $\hat{v} \in \mathcal{C}_f(\hat{U})$. Thus there is $\mathcal{P} \in \Pi$ such that for almost all $a \in A$, $\hat{v}^{P(a)} \in \hat{U}(P(a))$. By definition of the comprehensive extension of $U(P(a))$, for almost all a , there exists $v^{P(a)} \in U(P(a))$ such that $v^{P(a)} \geq \hat{v}^{P(a)}$. Clearly, since \hat{v} cannot be improved upon, neither can v . Consequently, $v \in \mathcal{C}_f(\hat{U})$. Using the feasibility condition, for almost all a , there exists a contract $c^{P(a)}$ such that $v^{P(a)}(b) = u(b | c^{P(a)})$ for all $b \in P(a)$. This establishes the existence of an equilibrium for our economy.

(b) Proof of Proposition 2

Here the strategy is to show that a failure to maximize the firm's income could be blocked by the firm itself: it is possible to make all members

better off by increasing the firm's income. The proof is less than straightforward because the presence of the truth-telling constraints limits across-state transfers and the incentive constraints (especially those of silent partners and managers) limit the ability to distribute gains to all members. We use

Lemma 1 Let $c = (y, p, L, x, W, S, M) \in C(F)$ and $\delta > 0$. Then if

$$(\hat{y}, p) \equiv \begin{cases} (y_s + \delta, y_t + \delta/p, y_f, y_1, p), & \text{if } p > 0 \\ (y_s + \delta, y_t, y_f + \delta, y_1, 0), & \text{if } p = 0 \end{cases}$$

also satisfies NNP and TT, there is an alternative contract $\hat{c} = (\hat{y}, p, L, \hat{x}, W, S, M) \in C(F)$ such that for all $a \in F$, $u(a|\hat{c}) > u(a|c)$.

Proof For each $a \in F$, $\hat{x}_i(a) = x_i(a) + (\hat{y}_i - y_i)/\#F$, $i = s, f, t, l$. Then \hat{x} satisfies budget balancing. Moreover, for each a , $\hat{x}_s(a) - [p\hat{x}_t(a) + (1-p)\hat{x}_f(a)] = x_s(a) - [px_t(a) + (1-p)x_f(a)]$, so $\hat{x}(a)$ satisfies the incentive constraints for the share contract, and $u(a|\hat{c}) = u(a|c) + \delta/\#F$. ■

This result provides a way for an increase in a firm's surplus to be distributed among its members to make each of them better off without violating the incentive compatibility of their share contract. We shall repeatedly make use of this construction to characterize equilibrium, since by definition, such improvements cannot be made from an equilibrium allocation.

We now provide the separation result.

Proposition 2 Let (c^*, \mathcal{P}^*) be an equilibrium. For each $F \in \mathcal{P}^*$ and equilibrium contract $c^*(F) = (y, p, L, x, W, S, M) \in C(F)$, the financial contract (y, p, L) solves

$$\max_{(\hat{y}, p)} \pi \hat{y}_s + (1-\pi)(p\hat{y}_t + (1-p)\hat{y}_f) \quad (\text{FC})$$

$$\text{s.t. } \hat{y}_i \geq 0, \quad i = s, t, f, l \quad (\text{NN})$$

$$0 \leq \hat{p} \leq 1 \quad (\text{P})$$

$$\pi \hat{y}_s + (1-\pi)(p\hat{y}_t + (1-p)\hat{y}_f) + (1-\pi)p\gamma \leq \omega_F r - Lr + R \quad (\text{NP})$$

$$\hat{y}_s \geq (1-\hat{p})(\hat{y}_f + R) + \hat{p}\hat{y}_1 \quad (\text{TT})$$

Proof First note that without loss of generality we need only consider contracts in which $y_1 = 0$. If y_1 exceeds zero, it may be lowered without affecting the firm's payoff, and since this weakens (TT), does not diminish the set of feasible share contracts for the firm.

Second, observe that (NP) always binds. If not, let $\delta = \omega_F r - Lr + R - [\pi \hat{y}_s + (1-\pi)(\hat{p}y_t + (1-p)\hat{y}_f) + (1-\pi)p\gamma]$ and apply Lemma 1 to increase the utility of each $a \in F$, contradicting the assumption that (y, p) is part of an equilibrium. Call (ZP) the binding form of (NP).

Now there are two cases to consider, depending on whether the firm's wealth is greater or less than the size of its loan.

Case 1: $\omega_F < L$. (ZP) and (TT) together imply that $p > 0$ whenever $\omega_F < L$. This implies that we may without loss of generality consider contracts for which $y_f = 0$, for y_t may always be increased to keep (ZP) and the firm's payoff unchanged (since whether to audit is determined after messages are sent, this procedure does not affect any incentives).

We now claim that (TT) binds. If not, lower p to some $p' > 0$ and raise y_t to y'_t so that $py_t = p'y'_t$. This generates a surplus of $\delta = (1-\pi)(p-p')$ without violating any constraints. Now let $\hat{y}_s = y'_s + \delta$ and $\hat{y}_t = y'_t + \delta/p'$ and apply Lemma 1. Denote by (TTB) the binding form of (TT).

Denote by $\bar{C}(F) \subseteq C(F)$ the set of contracts which satisfy $y_1 = y_f = 0$, (ZP) and (TTB).

Now suppose in contradiction to the proposition that $\pi y_s + (1-\pi)py_t < \pi \hat{y}_s + (1-\pi)\hat{p}y_t$ for some (\hat{y}, \hat{p}) satisfying (N), (P), (ZP) and (TTB). We claim there is $\hat{c} \in \bar{C}(F)$ such that $u(a|\hat{c}) > u(a|c)$ for all $a \in F$. Note first that $\hat{p} < p$ by (ZP) and that by (TTB) $\hat{y}_s - y_s = (p-\hat{p})R > 0$. Substituting and using (ZP) once again yields

$$(1-\pi)(\hat{p}y_t - py_t) = [(1-\pi)\gamma - \pi R](p - \hat{p}).$$

The construction is slightly different, depending on the sign of the term in brackets.

Case 1a. $(1-\pi)\gamma - \pi R > 0$. Define $y'_s = \hat{y}'_s$, y'_t such that $\hat{p}y'_t = py_t$, and choose $\beta(a) > 0$ such that $\sum_{a \in W'} \beta(a) = y'_s - y_s$.

If F is a partnership ($M = \emptyset$), define $x'_s(a) = x_s(a) + \beta(a)$, $x'_t(a) = \frac{p}{\hat{p}}x_t(a)$ for $a \in W$, and $x'_s(a) = x_s(a)$, $x'_t(a) = x_t(a)$ for $a \in S$. These shares make workers more strongly incentive compatible since

$$x'_s(a) - p'x'_t(a) = x_s(a) + \beta(a) - px_t(a) > x_s(a) - px_t(a) \geq 1$$

and leaves silent partners unaffected. But since $\hat{p}y'_t < \hat{p}y_t$, they make a positive profit of $\delta = (1-\pi)\hat{p}(\hat{y}_t - y'_t)$ which can be distributed to all partners via Lemma 1.

If F is a hierarchy, the same reasoning applies except that the $\beta(a)$'s are distributed only to members of M and W' (recall W' is the set of agents in $W \cup S$ who, under the original contract, decide to work when they are not

monitored). This procedure strengthens the incentives of the manager to monitor.

Case 1b. $(1-\pi)\gamma - \pi R \leq 0$. In this case $py_t \geq \hat{p}\hat{y}_t$ while $\hat{y}_s > y_s$, so $\hat{y}_s - \hat{p}\hat{y}_t > y_s - py_t$. Define $y'_s = \hat{y}_s - \epsilon$, $y'_t = \hat{y}_t$, $p' = \hat{p}$, where ϵ satisfies $(\hat{y}_s - y_s) - (p\hat{y}_t - py_t) > \epsilon > 0$. Now choose $\beta(a) > 0$ such that $\sum_{a \in W \cup M} \beta(a) = y'_s - y_s$, and proceed as before, redefining shares which distribute the $\beta(a)$'s to $W \cup M$; since there is now a positive profit of $\pi\epsilon$, this may be distributed to all members of F by Lemma 1.

Case 2. $\omega_F \geq L$. If $p > 0$, then proceed as in Case 1. So instead assume $p = 0$. But the result is now trivial, since the value of the objective is fixed at $(\omega_F - L)r + \pi R$ by (ZP). ■

Following Proposition 2, there is no loss of generality in considering financial contracts such that $y_i = 0$ when $i \neq s$, i.e., the firm receives a positive payment only when there is success. Hence, we suppose that the firm receives $y = y_s$ with probability π_n , where y depends on the collateral and on the probability of success,

$$\begin{aligned} y(n, \omega_F) &= R + \alpha(n, \omega_F)(\omega_F - L)r/\pi_n \\ \alpha(n, \omega_F) &= 1 \quad \text{if } \omega_F \geq L \\ &= \alpha_n \quad \text{if } \omega_F < L. \end{aligned}$$

By budget balancing, it follows that the share contracts can specify a positive payment only when there is success, i.e., $x_1^F(a) = 0$ for all a and all $i \neq s$. Therefore, we write $x^F(a)$ to denote the compensation of an agent in (F, c^F) and we ignore from now on the reference to the state of the world.

In order to prove the Corollary, we need

Lemma 2: Let (F, c^F) be an equilibrium firm with $\#W(F) = n$. If $M(F) = \emptyset$, then for each $a \in S(F)$, $\Delta\pi_{n+1}x(a) < 1$. If $M(F) \neq \emptyset$, then for each $a \in S_{m(F)}$, $\Delta\pi_{m(F)+1}x(a) < 1$.

Proof Case 1. F is a hierarchy. For a contract x , let $m(F)$ be the maximum number of agents of $W \cup S$ who work in a Nash equilibrium when the monitor does not monitor, as specified in Definition 1. Because no confusion will arise, we write $m = m(F)$ below. Recall that we associate to m the partition (W_m, S_m) of $W \cup S$ such that $\Delta\pi_m x(a) \geq 1$ for $a \in W_m$ and $\Delta\pi_{m+1} x(a) \leq 1$ for $a \in S_m$. We remark that there is no loss of generality in supposing that $\min \{x(a) | a \in W_m\} \geq \max$

$\{x(a) | a \in S_m\}$. Indeed, if $x(a) < x(a')$, for $a \in W_m$, $a' \in S_m$, then $\Delta\pi_m x(a') > \Delta\pi_m x(a) \geq 1$ and $\Delta\pi_{m+1} x(a) < \Delta\pi_{m+1} x(a') \leq 1$, so $a' \in W_m$ and $a \in S_m$, contradicting $W_m \cap S_m = \emptyset$.

If $S_m = \emptyset$, there is nothing to prove. So suppose that $S_m \neq \emptyset$. Let $S_m \equiv \{a^1, \dots, a^s\}$ where $x(a^1) \geq x(a^2) \geq \dots \geq x(a^s)$. Suppose the lemma's conclusion is false, i.e. that $\Delta\pi_{m+1} x(a^1) = 1$.

Case 1a. $\Delta\pi_{m+1} \geq \Delta\pi_m$. Consider the set $W^1 = W_m \cup \{a^1\}$. Then for each $a \in W^1$, $\Delta\pi_{m+1} x(a) \geq 1$. By definition of m , $\Delta\pi_{m+2} x(a^2) > 1$. Since $a^2 \in S_m$, $\Delta\pi_{m+2} > \Delta\pi_{m+1}$. Define $W^2 = W^1 \cup \{a^2\}$. Then for each $a \in W^2$, $\Delta\pi_{m+2} x(a) \geq 1$. Continuing in this fashion, we can define recursively sets $W^i = W^{i-1} \cup \{a^i\}$ with the property that for each $a \in W^i$, $i \leq s-1$, $\Delta\pi_{m+i} x(a) \geq 1$ and $\Delta\pi_{m+i+1} x(a^{i+1}) > 1$. We obtain a contradiction since $W^s = W \cup S$ implies a contradiction of the manager's incentive constraint.

Case 1b. $\Delta\pi_{m+1} < \Delta\pi_m$. Since $\min \{x(a) | a \in W_m\} \geq x(a^1)$, $\Delta\pi_{m+1} x(a) \geq \Delta\pi_{m+1} x(a^1) \geq 1$. The proof continues as in the previous case.

Case 2. F is a partnership. Let $S(F) = \{a^1, \dots, a^s\}$ be the set of silent partners, where $x(a^1) \geq \dots \geq x(a^s)$. Suppose that there exists $a \in W$ such that $x(a) < x(a^1)$. Let $\hat{x}(a^1) = x(a^1) + 1/\pi_n$ and $\hat{x}(a) = x(a) - 1/\pi_n$. Then $\Delta\pi_n \hat{x}(a^1) \geq 1$, $\Delta\pi_{n+1} \hat{x}(a) < 1$, and $\pi_n \hat{x}(a^1) - 1 = \pi_n x(a^1)$ and $\pi_n \hat{x}(a) = \pi_n x(a) - 1$. Hence, we can shift the occupations of a and a^1 . Consequently, if (F, c^F) is a partnership, there exists a contract $\hat{c} \in \bar{C}(F)$ such that $u(a, c^F) = u(a, \hat{c})$ for each $a \in F$ and such that $\min \{\hat{x}(a) | a \in \hat{W}\} \geq \max \{\hat{x}(a) | a \in \hat{S}\}$. We suppose below that c^F already has this property.

Note first that if $n = N$, the lemma's conclusion is true since $\Delta\pi_{N+1} = 0$. So suppose $n < N$. Then if the conclusion of the lemma is not true, replication of the reasoning in Case 1 lets us conclude that the set $W^s = W \cup S$ is such that for each $a \in W^s$, $\Delta\pi_{n+s} x(a) \geq 1$. Because the probability of success is larger, $(\pi_{n+s} > \pi_n)$, there is a strict increase in the expected transfer, say $\delta > 0$. Let $x'(a) = x(a) + \delta/\#F$. Then, each agent in W is made strictly better off (such an agent works before and after reorganization). For an agent $a^1 \in S$, we note that $\Delta\pi_{n+s} x(a^1) \geq 1$ by construction. Therefore, $\pi_{n+s} (x(a^1) + \delta/\#F) - 1 > \pi_{n+s-1} x(a^1) \geq \pi_n x(a^1)$ and a^1 is made better off by working in the reorganized partnership than by not working in the initial partnership. This contradicts our assumption that (F, c^F) is an equilibrium firm. ■

Corollary 1 If $\pi > 0$, there is no loss of generality in supposing that $L = K$.

Proof Let (F, c^F) be an equilibrium firm. Suppose that the firm borrows $L > K$. From (F2) and (F3), the transfer y is a decreasing function of L . If $\omega_F > K$, then y is constant for $L \in [K, \omega_F]$ and decreases with L if $L > \omega_F$.

Suppose that F is a partnership. If $\omega_F > L > K$, then we might as well suppose that $L = K$ since the transfer to the firm is the same. If $L > \max\{\omega_F, K\}$, then by decreasing L to $\max\{\omega_F, K\}$, the transfer to the firm increases — supposing that the probability of success is the same. Let y be the initial transfer to the firm and y' be the transfer when the firm borrows only $\max\{\omega_F, K\}$. Then $\delta = y' - y > 0$. If $S(F) = \emptyset$, let $x'(a) = x(a) + \delta/\#F$. Then $\Delta\pi_n x'(a) > 1$ since a was initially incentive compatible. Clearly, each agent is made better off. If $S(F) \neq \emptyset$, then by using Lemma 2, there exists, for each $a \in S(F)$, $\varepsilon(a) > 0$ such that $x(a) + \varepsilon(a) < 1/\Delta\pi_{n+1}$. Therefore, we can find for each $a \in F$, $\delta(a) > 0$ such that $\sum_F \delta(a) = \delta$ and such that $x(a) + \delta(a) < 1/\Delta\pi_{n+1}$ for each $a \in S(F)$. Obviously, for each $a \in W(F)$, $x(a) + \delta(a) > 1/\Delta\pi_n$. Therefore, by borrowing $\max\{\omega_F, K\}$ instead of L , the firm can receive a larger transfer, generate the same probability of success, and each agent can be made better off. This is a contradiction.

If F is a hierarchy, we can replicate the above reasoning by replacing $S(F)$ by the set $S_{m(F)}$. ■

(c) *Proof of Proposition 3*

Consider an equilibrium $\{(F, c^F); F \in P^*, c^F \in \bar{C}(F)\}$ in which P^* minimal partition. The only constraint for the design of contracts on the financial market is the zero profit condition. There is no truth telling constraint since the probability of audit can be set equal to 1. Therefore, as long as $\pi y_s + (1-\pi)y_t = \pi R + (\omega_F - K)r$, and $y_s, y_t \geq 0$, the financial contract $(y_s, y_t, 0, 0, 1, K)$ is admissible. In other words, there is no problem of transferring income across states (while we know that when $\gamma > 0$ the possibilities of transferring income from state s to state t are limited by the truth telling constraint).

We note that there are no silent partners: it is not cheaper to borrow from a silent partner than from the market. Therefore silent partners should obtain a zero surplus and there is no loss of generality in supposing that equilibrium firms have at most $N+1$ members.

While there is no wealth effect at the financial market level, there is

still a wealth effect for incentive compatibility inside the firm since the zero profit condition sets an upper bound on the maximal total marginal compensation $y(n, \omega_F)$.

Proposition 3 Suppose that $\gamma = 0$. Then almost every equilibrium firm is surplus efficient.

Proof Suppose that (F, c^F) is not surplus efficient. Then there exists (G, c^G) , $c^G \in C(G)$ such that $V(F, c^F) < V(G, c^G)$. Let $n \equiv \#W(F)$ and $m \equiv \#W(G)$.

$$\sum_F [\pi_n x_s^F(a) + (1 - \pi_n) x_t^F(a)] = Y_n + \omega_F r. \quad (\text{A.1})$$

$$\pi_n x_s^F(a) + (1 - \pi_n) x_t^F(a) \geq 1 + \omega(a)r, \text{ for all } a \in F. \quad (\text{A.2})$$

$$Y_m > Y_n - (\#F - \#G). \quad (\text{A.3})$$

(A.1) is the zero profit condition. (A.2) is the equilibrium condition $u(a, c^F) \geq 0$ for each $a \in F$ (where we use our previous observation that there is no silent partner, i.e., that each member of a firm must work). (A.3) is implied by the assumption that F is surplus inefficient.

Our proof is constructive and we need to consider the cases in which G is a partnership or a hierarchy.

Case 1. $M(G) = \emptyset$ (G is a partnership). Since there exists $c \in C(G)$ such that $V(G, c^G) > V(F, c^F)$ (we use the zero profit constraint)

$$Y_m + \omega_G r \geq m / \Delta \pi_m. \quad (\text{A.4})$$

We now construct a contract that has the property that for each $a \in G$, $u(a, c^G) > u(a, c^F)$. Consider the set of contracts satisfying

$$x_s^G(a) = x_t^G(a) + 1 / \Delta \pi_m,$$

$$x_t^G(a) \geq 0,$$

$$\sum_{a \in G} x_t^G(a) = Y_m + \omega_m r - m / \Delta \pi_m.$$

Note that this set of contracts is not empty since (A.4) holds.

Because all agents work in F and in G , comparing the surpluses is equivalent to comparing the expected incomes. In G , the sum of the expected incomes is equal to $Y_m + \omega_G r$.

$$Y_m + \omega_G r > Y_n + \omega_G r - (\#F - \#G) \quad \text{by (A.3)}$$

$$= Y_n + \omega_F r - (\#(F \setminus G) + \omega_{F \setminus G} r)$$

$$\geq Y_n + \omega_F r - \sum_{a \in F \setminus G} [\pi_n x_s^F(a) + (1 - \pi_n) x_t^F(a)] \quad \text{by (A.2)}$$

$$= \sum_{a \in G} [\pi_n x_s^F(a) + (1 - \pi_n) x_t^F(a)] \quad \text{by (A.1).}$$

Therefore, there exists x^G such that for each $a \in G$,

$$\pi_m x_s^G(a) + (1 - \pi_m) x_t^G(a) > \pi_n x_s^F(a) + (1 - \pi_n) x_t^F(a).$$

But this implies that there exists a contract $c^G \in C(G)$ such that for each $a \in G$, $u(a|c^G) > u(a|c^F)$. This contradicts the fact that (F, c^F) is an equilibrium firm.

Case 2. $M(G) = \{b\}$ (G is a hierarchy). We can replicate the reasoning of Case 1 by considering the following share contract (b denotes the monitor).

$$x_s^G(b) = x_t^G(b) + 1/\pi_m$$

$$x_s^G(a) = x_t^G(a) \text{ for all } a \in G \setminus \{b\},$$

$$x_t^G(a) \geq 0, \text{ for all } a \in G,$$

$$\sum_{a \in G} x_s^G(a) = Y_m + \omega_G r - 1.$$

We leave the details to the reader. ■

(d) Proofs of Propositions 4 and 5

We shall require some additional notation. Let $u^0 \equiv \text{essinf } u(a, c^{F(a)})$ be the essential infimum of the surpluses of the agents in an equilibrium $\{(F, c^F)\}$. By definition of u^0 , the set $B = \{a \in A | u(a|c^{F(a)}) < u^0\}$ has measure zero.

For any $\varepsilon > 0$, the set $A_\varepsilon \equiv \{a \in A; u(a, c^{F(a)}) \leq u^0 + \varepsilon\}$ is the set of agents who have an equilibrium surplus ε -close to u^0 . By definition of u^0 , A_ε has positive measure for all $\varepsilon > 0$. Note that $B \subseteq A_\varepsilon$ for any $\varepsilon > 0$.

To simplify, we suppose that A_0 has positive measure. The proofs in Lemma 3 and Propositions 4 and 5 can be generalized in a straightforward way when A_0 has measure zero: for any firm $F \subseteq A \setminus B$, there exists $\varepsilon(F) > 0$ such that for each $a \in F$, $u(a, c^F) > u^0 + \varepsilon(F)$ and there exists a positive measure of agents with surplus in the interval $(u^0, u^0 + \varepsilon(F))$. We can replace the set A_0 by the set $A_{\varepsilon(F)}$.

Lemma 3 If (A1) and (A2) hold then $\#W(F) = N$ for almost all equilibrium hierarchies.

Proof The idea of the proof is the following. Let $F \subset A \setminus B$. If $\#W(F) = n < N$,

then we form a new set G by adding $N-n$ agents from A_0 to F (this is possible because $F \subset A \setminus B$). The additional agents become workers. We then construct a contract for the larger set G of agents such that each of these agents is made strictly better off than in the equilibrium situation.

We suppose that each additional agent in A_0 invests all her wealth in the safe asset. Therefore, the total wealth in G is the same as the total wealth in F . Clearly, if we can prove that each agent in G can be made strictly better off under this assumption, each agent can also be made strictly better off if agents in $G \setminus F$ invest their wealth in the firm. Hence, if $a \in G \setminus F$, a 's expected surplus in (G, c^G) is $\pi_N x(a) - 1$. In order to make a strictly better off, we shall need $\pi_N x(a) > u^0 + 1$.

Let (F, c^F) be an equilibrium hierarchy. Suppose that $\#W(F) = n < N$. Consider $G \supset F$ where $G \setminus F \subset A_0$ and $\#G \setminus F = N-n$. Let $b \in F$, $M(F) = \{b\}$.

Case 1. Suppose that $u^0 + 1 < \pi_N / \Delta \pi_N$. We organize G as follows: $W(G) = W(F) \cup G \setminus F$, $M(G) = M(F) = \{b\}$, $S(G) = S(F)$. Consider the following share contract:

$$\begin{aligned} x^G(a) &= \pi_n x^F(a) / \pi_N, \text{ for all } a \in W(F) \cup S(F), \\ x^G(a) &= (u^0 + 1) / \pi_N, \text{ for all } a \in G \setminus F \\ x^G(b) &= y(N, \omega_F) - (N-n)(u^0 + 1) / \pi_N - \pi_n y(n, \omega_F) / \pi_N + \pi_n x^F(b) / \pi_N \end{aligned}$$

We need to show that the resulting contract c^G is in $C(G)$. Budget balancing is immediate since by definition $\sum_{a \in F} x^F(a) = y(n, \omega_F)$. Nonnegativity of $x^G(a)$ for $a \neq b$ is also immediate. For the monitor b , it is enough to prove that

$$\pi_N y(N, \omega_F) - (N-n)(u^0 + 1) - \pi_n y(n, \omega_F) \geq 0.$$

Simple algebra shows that the above expression is equal to

$$E = (\pi_N - \pi_n)R + [\alpha(N, \omega_F) - \pi_n \alpha(n, \omega_F)](\omega_F - K)r - (N-n)(u^0 + 1)$$

From (A2),

$$(\pi_N - \pi_n) / (N-n) > \pi_n / n \text{ for any } n < N. \quad (\text{A.5})$$

Because each agent in F has a surplus greater than u^0 , each worker in $W(F)$ must have an expected income greater than $u^0 + 1$. Therefore, if $u^0 + 1 \geq \pi_n R / n$, it follows that each agent in F has a surplus greater than $(\pi_n R - n) / n$. This contradicts the fact that $V(F, c^F) \leq \pi_n R - Kr - (n+1)$. Therefore, $u^0 + 1 < (\pi_n R - n) / n$.

If $\omega_F \geq K$, this proves that $x^G(b) > 0$ since $\alpha(N, \omega_F) = \alpha(n, \omega_F) = 1$, and therefore that $y(N, \omega_F) - (N-n)(u^0 + 1) / \pi_N - \pi_n y(n, \omega_F) / \pi_N > 0$.

If $\omega_F < K$, note that $(\pi_N - \pi_n)R + (\alpha_N - \alpha_n)(\omega_F - K)r > (\pi_N - \pi_n)R \geq (\pi_N - \pi_n)y(n, \omega_F)$, since $\alpha_N < \alpha_n$. Suppose that $E < 0$; then $u^0 + 1 > (\pi_N - \pi_n)y(n, \omega_F) / (N-n) > \pi_n y(n, \omega_F) / n$ (by (A.5)). We obtain a contradiction,

since the n workers in $W(F)$ could not have shared more than $y(n, \omega_F)$ without some silent partners getting a negative surplus. Therefore, $x^G(b) > \pi_n x^F(b) / \pi_N$.

We now check incentive compatibility. Let $m(F)$, $m(G)$ be the maximum number of agents who work in any Nash equilibrium when the monitor does not monitor in c^F and c^G respectively. We need to establish that $(\pi_N - \pi_{m(G)}) x^G(b) \geq 1$.

We remark that $m(F) < n$ since b is incentive compatible in (F, c^F) . If $m(G) > m(F)$, all agents in $G \setminus F$ must satisfy $\Delta \pi_{m(G)}(u^0+1) / \pi_N \geq 1$; otherwise, there exists a set $W_{m(G)}$ of $m(G)$ agents from $W(F) \cup S(F)$ for which $\Delta \pi_{m(G)} x^G(a) \geq 1$ and such that for all $a \in W \setminus W_{m(G)}$, $\Delta \pi_{m(G)+1} x(a) \leq 1$. This is impossible by definition of $m(F)$.

Since $u^0+1 < \pi_N / \Delta \pi_N$, (A1) implies that $u^0+1 < \pi_N / \Delta \pi_{m(G)}$. Therefore, no agent in $G \setminus F$ wants to work when b does not monitor and from our observation in the previous paragraph, $m(G) \leq m(F)$. But then, incentive compatibility for b in (G, c^G) follows since

$$\begin{aligned} (\pi_N - \pi_{m(G)}) x^G(b) &> (\pi_N - \pi_{m(F)}) \pi_n x^F(b) / \pi_N \\ &\geq (\pi_n - \pi_{m(F)}) x^F(b) \\ &\geq 1 \end{aligned}$$

The first inequality is by definition of x^G , the fact that $m(G) \leq m(F)$ and our previous observation that $x^G(b) > \pi_n x^F(b) / \pi_N$. The second inequality follows from the fact that $\pi_n < \pi_N$. The last inequality follows from incentive compatibility of b in (F, c) .

Finally note that for all $a \in F$, $u(a, c^G) = u(a, c^F)$ and that for b , $u(b, c^G) > u(b, c^F)$. Moreover, b 's incentive compatibility in (G, c^G) is not binding, $(\pi_N - \pi_{m(G)}) x^G(b) > 1$. Therefore, it is possible to decrease $x^G(b)$ by $\delta > 0$, δ small enough that b still satisfies incentive compatibility, and to increase each $x^G(a)$, $a \neq b$, by $\delta / (\#G - 1)$. For δ small enough, there are still only $m(G)$ agents in $G \setminus \{b\}$ who want to work when b does not monitor. Therefore, each agent in G is strictly better off than in equilibrium. A contradiction.

Case 2. Suppose that $u^0+1 > \pi_N / \Delta \pi_N$. Then for any $a \in W(F)$, $\pi_n x^F(a) > \pi_N / \Delta \pi_N$ by (A1). Consider $G \supset F$, where $G \setminus F \subset A_0$. We show that G can be re-organized as a partnership which generates higher surplus for all its members. Let $W(G) = W(F) \cup G \setminus F$, $S(G) = S(F) \cup \{b\}$ and $M(G) = \emptyset$. Consider the same sharing contract as in Case 1. Since for $a \in W(G)$, $x^G(a) > 1 / \Delta \pi_N$, the workers are

incentive compatible in the partnership. Note that b not only has a greater expected income but also saves the unit of effort that she was expending as a monitor in (F, c^F) . Therefore, it is possible to decrease $x^G(b)$ by $\delta > 0$ and increase each $x^G(a)$ by $\delta/(\#G-1)$. This will decrease b's surplus by $\pi_N \delta$, increase each other agent's surplus by a positive amount and will weaken incentive compatibility. Therefore it is possible to make every agent strictly better off. A contradiction. ■

Assumption (A3) implies that a firm with wealth $\omega_F \geq K$ could sustain a partnership. The proof of Proposition 4 proceeds as follows. First, we show that each agent with wealth greater than K/N must have a surplus of at least $Y_N/N - 1$. Then we argue that the assumption on the distribution implies that agents with wealth greater than K will compete for forming firms with agents who have less than K/N and that this competition leads to an equalization of the surpluses in the economy.

We use the following two facts.

Fact 1 $u(a|c^{F(a)}) \geq Y_N/N - 1$ for almost every a such that $\omega(a) \geq K/N$.

Fact 2 Let $\omega(b) \geq K$. Let $\varepsilon^* \leq Y_N/N - \pi_N/\Delta\pi_N$. Let for each $a \neq b$, $0 \leq \varepsilon(a) \leq \varepsilon^*$. Then there exists a firm (F, c^F) containing b such that for each $a \neq b$, $u(a|c^F) = Y_N/N - 1 - \varepsilon(a)$ and $u(b|c^F) = Y_N/N - 1 + \sum_{a \in F \setminus b} \varepsilon(a)$.

Proof of Fact 1 Let F be a partnership consisting of N agents of wealth greater than K/n . Then the partnerships has a total wealth greater than K and it is possible to choose a financial contract such that $y \geq R \geq N/\Delta\pi_N$ (by A2) and share contracts such that $x(a) = y/N > 1/\Delta\pi_N$. Therefore, it is possible to share the total surplus $Y_N - N$ equally among the members of the partnership. Suppose now that there is a positive measure of agents with wealth greater than K/N who have a surplus strictly less than $Y_N/N - 1$. Then, from the previous observation, N of these agents would form a partnership and be better off. ■

Proof of Fact 2 Observe that we can choose $\varepsilon^* > 0$ by (A3). Let $b \in F$, $\#F = N$. F is organized as a partnership: $F = W(F)$. Consider the share contract in which for $a \neq b$, $x(a) = [\omega(a) r + Y_N/N - \varepsilon(a)]/\pi_N$. Then, $x(a) > 1/\Delta\pi_N$ since $\varepsilon(a) \leq \varepsilon^*$. Moreover, $u(a|c^F) = Y_N/N - \varepsilon(a)$. The result is proved if we show that

b is incentive compatible. By budget balancing, (recall that $y \geq R$)

$$\begin{aligned} x(b) &= R + (\omega(b)-K)r/\pi_N - (N-1)Y_N/(N\pi_N) + \sum_{a \in F \setminus b} \varepsilon(a)/\pi_N \\ &= Y_N / N\pi_N + [\omega(b)r + \sum_{a \in F \setminus b} \varepsilon(a)]/\pi_N \\ &> Y_N / (N\pi_N) \\ &> 1/\Delta\pi_N \qquad \qquad \qquad \text{by (A3)} \end{aligned}$$

Therefore b is incentive compatible. ■

Proposition 4 Suppose (A1), (A2), (A3) and that $H(K/N) < (N-1)(1 - H(K))$. Then almost all equilibrium firms are partnerships of size N with a total wealth of at least K and are surplus efficient. Moreover, almost every agent has the same equilibrium surplus.

Proof Suppose that there is a set $A^K \subset A \setminus B$ of agents with wealth greater than K who form hierarchies and that $\lambda(A^K) > 0$. From Lemma 3, for each $a \in A^K$, $\#W(F(a)) = N$. Moreover, there is no loss of generality in supposing that for each such a, $S(F(a)) = \emptyset$. Therefore, $\#F(a) = N + 1$ for each $a \in A^K$. The maximum expected utility of such an agent is attained by employing N agents in A_0 , offering them a surplus of u^0 each, and obtaining a surplus of $Y_N - N(u^0+1) - 1$.

It is clear that if $a \in A^K$ forms a hierarchy, she will form it with N agents in A_0 . If this is not the case, a's surplus is less than $Y_N - N(u^0+1) - 1$ and a can attract N agents in A_0 by offering them a surplus slightly greater than u^0 and make them and herself better off.

Because $\lambda(A_0) \leq H(K/N) < (N-1)(1-H(K))$, there exists a positive measure of agents with wealth greater than K who form partnerships. Call this set D. The expected utility of $a \in D$ is bounded above by the quantity

$$\max_n [Y_n - (n-1) \max \{u^0+1, \pi_n/\Delta\pi_n\}].$$

Let \bar{n} maximize that function. The maximum value of the function is bounded above by $Y_{\bar{n}} - (\bar{n}-1)(u^0+1)$. In equilibrium, a must be indifferent between forming a partnership and a hierarchy, so $Y_N - N(u^0+1) \leq Y_{\bar{n}} - (\bar{n}-1)(u^0+1)$. Therefore,

$$u^0+1 \geq (\pi_N - \pi_{\bar{n}})R / (N - \bar{n} + 1).$$

Since agents in A^K who form partnerships obtain a positive surplus,

$$u^0+1 < \pi_{\bar{n}}R / \bar{n}.$$

Combining these two inequalities, we obtain $\pi_{\bar{n}}(N - \bar{n} + 1) > \bar{n}(\pi_N - \pi_{\bar{n}})$, which is equivalent to $(N+1)\pi_{\bar{n}} > \bar{n}\pi_N$. This contradicts (A2) unless $\bar{n} = N$.

Therefore, each agent in D is matched with at least (N-1) other agents.

Because $H(K/N) < (N-1)(1-H(K))$, there are agents in D who are matched with agents whose wealth exceeds K/N . Applying Fact 2, an agent $a \in D$ who is matched with an agent of wealth greater than K/N can attract an agent in A_0 by paying a surplus of slightly over u^0 , thereby making herself and the other agents better off.

Because some agents in D must be matched with agents of wealth greater than K/N , $u^{0+1} = Y_N^0/N$. But then agents in A^K who form a hierarchy have a surplus less than $Y_N^0/N - 1$, which contradicts Fact 1.

Consequently, each agent of wealth greater than K forms a partnership of size N . Some must be matched with agents of wealth greater than K/N . From Fact 2, this can be an equilibrium situation only if $u^{0+1} = Y_N^0/N$. The conclusion of the Proposition follows. ■

Before proving Proposition 5, we show that $\omega^* > 0$. To see this, observe that the result is immediate if $n^* = N$, since (A3) is equivalent to $(N-1)\pi_N/\Delta\pi_N > N$. If instead $n^* < N$ (it is obvious that n^* never exceeds N), then by definition of n^* and (A4), we have

$$Y_n^* - (n^*-1)\pi_n^*/\Delta\pi_n^* > Y_N^0 - (N-1)\pi_N/\Delta\pi_N > \pi_N/\Delta\pi_N > 1.$$

It follows that $\pi_n^*R/n^* > 1$. Since $n^* < N$, (A2) implies that $(\pi_N - \pi_n^*)R/(N-n^*+1) > \pi_n^*R/n^*$, and therefore $Y_N^0 - N > Y_n^* - (n^*-1) > Y_n^* - (n^*-1)\pi_n^*/\Delta\pi_n^*$. Consequently, $\omega^* > 0$.

Proposition 5: Suppose (A1)-(A5), that $H(\omega^*) > N(1-H(\omega^*))$ and that $H(K) < 1$. Then there is positive measure of agents who belong to inefficient equilibrium hierarchies.

Proof Suppose that there is at most a zero measure of inefficient hierarchies in equilibrium. This implies that almost all agents with wealth greater than K join partnerships (any hierarchy in which one member has K is inefficient because by (A4) it could be reorganized as a partnership producing the same output). Let A^* be the set of agents with wealth less than or equal to ω^* and $u^* \equiv \text{essinf} \{u(a|c^{F(a)}) | a \in A^*\}$. By our assumption on H , u^* is well defined and $u^* \geq u^0$.

Claim 1 $u^0 \geq \alpha^* \omega^* r$.

Proof An agent a with wealth greater than K who forms a partnership will choose a size n that maximizes the expression

$$Y_n - (n-1)\max\{u^0+1, \pi_n/\Delta\pi_n\} \quad (A.6)$$

over the set of n such that $\pi_n R + (\omega(a)-K)r \geq n\pi_n/\Delta\pi_n$. For any values of u^0 and of $\omega(a)$, the function in (A.6) is no greater than $Y_n^* - (n-1)\pi_n^*/\Delta\pi_n^*$. Therefore, a necessary condition for a to form a partnership rather than a hierarchy with N agents from A_0 is that $Y_n^* - (n-1)\pi_n^*/\Delta\pi_n^* \geq Y_N - N(u^0+1)$. This implies $u^0 \geq \alpha^* \omega^* r$.

Claim 2: Let (F, c^F) be an equilibrium firm. Suppose that $S(F) \cap A^* \neq \emptyset$. Let $\delta(a) \equiv \pi_n [y(n, \omega_F) - y(n, \omega_F - \omega(a))]$. Then $\delta(a) \leq \alpha^* \omega^* r$ for each $a \in S(F) \cap A^*$.

Proof By definition of $\delta(a)$,

$$\delta(a) = \alpha(n, \omega_F)(\omega_F - K)r - \alpha(n, \omega_F - \omega(a))(\omega_F - \omega(a) - K)r.$$

The maximum value of $\delta(a)$ is obtained when $\omega_F < K$ and is equal to $\alpha_n \omega(a)r$. This is less than $\alpha^* \omega^* r$ since $a \in A^*$ and since $\alpha^* \geq \alpha_n$.

From Claim 1, $u^* > \alpha^* \omega^* r > \alpha_n \omega(a)r$. Thus, if there is no positive measure of inefficient hierarchies, $u^* > 0$.

From Claim 2, if agent $a^* \in S(F) \cap A^*$ is excluded from F , the set of agents $F \setminus \{a^*\}$ can find a financial contract in which $y(n, \omega_F - \omega(a^*)) > y(n, \omega_F) - u(a^* | c^F)$. Therefore, there is more marginal compensation to distribute in $F \setminus \{a^*\}$ when a^* is excluded than in the initial situation. Using Lemma 1, this surplus can be distributed among the agents in $F \setminus \{a^*\}$ to make them all better off without violating the initial incentive compatibility conditions. This contradicts the assumption that F is an equilibrium firm.

It follows that no equilibrium firm will have silent partners with wealth less than ω^* ; any agent $a \in A^*$ will belong to a firm only if she is a worker or a manager. (Note that while we suppose that there is no inefficient hierarchy, we do not eliminate the possibility of efficient hierarchies.) Therefore, each agent in $A \setminus A^*$ will be matched with at most $N+1$ agents in A^* . From our assumption on H , it follows that there is a positive measure of agents in A^* who are not matched with agents outside A^* . Let A^{**} be this set of agents. Then, for each $a \in A^{**}$, $F(a) \subset A^*$ and $\lambda(A^{**}) > 0$. From our previous observations, for each $a \in A^{**}$, $\#F(a) \leq N+1$ and $S(F(a)) = \emptyset$.

Suppose that in equilibrium there is a firm F with $\#W(F) = n$ consisting of agents in A^{**} . For this firm, $\omega_F \leq (N+1)\omega^* < K$ by definition of ω^* . Therefore, the firm must borrow and the total surplus available in F is

$$V(F) = \pi_n R + \alpha_n (\omega_F - K)r - \omega_F r - \#F,$$

and $V(F) \geq \#F u^*$. Since $u^* \geq \alpha^* \omega^* r$ and since $\omega_F \leq \#F \omega^*$, it follows that $V(F) \geq \alpha_n \omega_F r$. But then, $\pi_n R - \alpha_n K r - \#F \geq 0$ which contradicts (A4) since that expression is less than $\pi_N R - \alpha_N K r - N$.

Therefore, agents in A^{**} will be unemployed (i.e., $F(a) = \{a\}$ for $a \in A^{**}$), and they receive zero surplus. This contradicts the fact that the absence of a positive measure of inefficient hierarchies implies $u^* > 0$. ■

REFERENCES

- Alchian, A. and H. Demsetz, "Production, Information Costs and Economic Organization," *American Economic Review*, December 1972.
- Bernanke, B. and M. Gertler, "Agency Costs, Net Worth, and Business Fluctuations," *American Economic Review*, March, 1989.
- , "Financial Fragility and Economic Performance", *Quarterly Journal of Economics*, February, 1990.
- Banerjee, A.V., A.F. Newman and Y. Qian, "Capital Markets, Inequality, and the Size Distribution of Business Firms," mimeo, Stanford University, 1990.
- Border, K. and J. Sobel (1987), "Samurai Accountant: a Theory of Auditing and Plunder," *Review of Economic Studies*, 54, 525-540.
- Boyd, J.H. and E.C. Prescott, "Financial Intermediary-Coalitions," *Journal of Economic Theory*, April, 1986.
- Coase, R., "The Nature of the Firm," *Economica*, 4, 1937.
- Diamond, D. (1984), "Financial Intermediation and Delegated Monitoring," *Review of Economic Studies*, 51, 393-414.
- Drèze, J., "Some Theory of Labor Management and Participation," *Econometrica*, November, 1976.
- , *Labour Management, Contracts, and Capital Markets: a General Equilibrium Approach*, Oxford, Basil Blackwell, 1989.
- Grossman, S. and Hart, O. "The Costs and Benefits of Ownership: a Theory of Vertical and Lateral Integration," *Journal of Political Economy*, August, 1986.
- Hammond, P., M. Kaneko and M.H. Wooders, "Continuum Economies with Finite Coalitions: Core, Equilibria, and Widespread Externalities," *Journal of Economic Theory*, 49, October, 1989, 113-134.
- Hart, O. "Perfect Competition and Optimal Product Differentiation," *Journal of Economic Theory*, 1980.
- , "An Economist's Perspective on the Theory of the Firm," *Columbia Law Review*, November, 1989.
- Hart, O. and B. Holmström, "The Theory of Contracts," in *Advances in Economic Theory*, ed. T.F. Bewley, Cambridge: Cambridge University Press, 1987.
- Hart, O. and J. Moore, "Property Rights and the Nature of the Firm," *Journal of Political Economy*, December, 1990.
- , "A Theory of Debt Based on the Inalienability of Human Capital," mimeo, MIT and LSE, 1991.
- Holmström, B., "Moral Hazard in Teams" *Bell Journal of Economics*, 13, 1982.

- Holmström, B. and J. Tirole, "The Theory of the Firm," in *Handbook of Industrial Organization*, ed. R. Schmalensee and R. Willig, New York: North-Holland, 1989.
- Ichiishi, T., "Coalition Structure in a Labor-Managed Market Economy," *Econometrica*, March, 1977.
- Kaneko, M. and M.H. Wooders, "The Core of a Game with a Continuum of Players and Finite Coalitions: Nonemptiness with Bounded Sizes of Coalitions," IMA Preprint Series No. 126, University of Minnesota, 1984.
- , "The Core of a Game with a Continuum of Players and Finite Coalitions: the Model and some Results," *Mathematical Social Sciences*, 12, 1986.
- Kehoe, T.J. and D.K. Levine, "Debt Constrained Asset Markets," forthcoming, *Review of Economic Studies*.
- Legros, P. *Efficiency and Stability in Partnerships*, Unpublished Ph.D. dissertation, California Institute of Technology, 1989.
- Legros, P. and H. Matsushima, "Efficiency in Partnerships," *Journal of Economic Theory*, December, 1991.
- Legros, P. and S. Matthews, "Efficient and Nearly Efficient Partnerships," forthcoming, *Review of Economic Studies*
- Makowski, L., "Perfect Competition, the Profit Criterion, and the Organization of Economic Activity," *Journal of Economic Theory*, August, 1980.
- Milgrom, P. and J. Roberts, *Economics, Organization and Management*, Englewood Cliffs, New Jersey: Prentice Hall, 1992.
- Newman, A.F., "The Capital Market, Inequality, and the Employment Relation," mimeo, Northwestern University, 1991.
- Radner, R. (1992), "Hierarchy: The Economics of Managing," *Journal of Economic Literature*, 30, 1382-1415.
- Townsend, R.M. (1979), "Optimal Contracts and Competitive Markets with Costly State Verification," *Journal of Economic Theory*, 21, 265-293.
- Williams, S. and R. Radner, "Efficiency in Partnership when the Joint Output is Uncertain," CMSEMS DP 760, Northwestern, 1988. Forthcoming in *The Economics of Informational Decentralization*, ed. J. Ledyard, Kluwer.
- Williamson, O., *The Economic Institutions of Capitalism*, New York: The Free Press, 1985.
- Wooders, M.H., "Large Games and Economies with Effective Small Groups," Bonn Discussion Paper B-215, 1992.