JOURNAL OF ECONOMIC THEORY **70**, 312–341 (1996)
ARTICLE NO. 0092

Wealth Effects, Distribution, and the Theory of Organization*

Patrick Legros

Department of Economics, Cornell University, Ithaca, New York 14853–7601

and

Andrew F. Newman

Department of Economics, Columbia University, New York, New York 10027

Received May 28, 1994; revised August 21, 1995

We construct a general equilibrium model of firm formation in which organization is endogenous. Firms are coalitions of agents providing effort and investment capital. Effort is unobservable unless a fixed monitoring cost is paid, and borrowing is subject to a costly state verification problem. Because incentives vary with an agent's wealth, different types of agents become attractive firm members under different circumstances. When borrowing is not costly, firms essentially consist of one type of agent and are organized efficiently. But when the costly state verification problem is sufficiently severe, firm organization will depend on the distribution of wealth: with enough inequality, it will tend to be dictated by incentives of rich agents to earn high returns to wealth, even if the chosen organizational form is not a technically efficient way to provide incentives. *Journal of Economic Literature* Classification Numbers: D2, D31, J41, L2.

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 60 years ago. In market economies, firms form by voluntary association: utility-maximizing individuals are free to move among them, create and dissolve them, and choose the way they are organized,

* We are grateful for comments from the referees, A. Banerjee, P. Bolton, D. Fudenberg, R. Guesnerie, O. Hart, M. Jackson, J.-J. Laffont, A. Manelli, S. Matthews, M. Meyer, J. Moore, R. Sah, J. Tirole, S. Weber, and numerous workshop participants. W. Pesendorfer, B. Smith, L. Stole, A. Wolinsky, and M. Holtz Wooders deserve special mention for their advice and patience.

0022-0531/96 $18.00

subject, of course, to the constraints of technology, information, and the wishes of other individuals. Any explanation for organization—whether based on transaction costs, agency and incentive problems, or difficulties of coordination and information transmission—must therefore take account of the competitive forces which voluntary association generates.

The theory of incentives, with the principal–agent model as its chief analytical tool, has dominated recent economic analysis of organization (see, e.g., Hart [11], Hart and Holmström [12], Holmström and Tirole [15], and Radner [30] for surveys). This framework accommodates voluntary association through two exogenous pieces of information: the opportunity cost (individually rational utility level) of the agent, and the assignment (or "match") of the principal to the agent. As a rule, these data are essential for determining the nature of the contract selected: the compensation scheme, the production and monitoring technology, and the efficiency of the organization. But the data of the principal–agent model are not economic fundamentals; a complete theory of organization would make the matches and individually rational utilities endogenous to tastes, technology, and endowments, and would therefore need to take account of the general equilibrium effects which inhere in this problem. The purpose of this paper is to construct a simple version of such a theory.

We study a model of firm formation in which the firms, their membership, and their organization are endogenous, and do so for an environment in which agents' payoffs feature a significant nontransferability: wealth effects.[1] Production requires a fixed capital investment and the efforts of a firm's risk-neutral members. Two agency problems have to be solved: one internal due to a moral hazard problem in effort provision, and one external due to an asymmetry of information about the output of the firm between its members and outside lenders. Each problem can be solved at no extra cost if there is enough wealth in the firm: the moral hazard problem can be solved if agents can post a large enough incentive bond, and the costly state verification problem is absent if the firm can be self-financed. Otherwise a costly monitoring (of effort) or auditing (of output) technology needs to be used.

This leads to two possible internal organizational forms: the *M*-firm uses the monitoring (of effort) technology and the *I*-firm does not. These forms represent the extremes of a trade-off between monitoring costs and team

[1] By wealth effects, we mean simply changes in marginal incentives that accompany changes in wealth; in the present paper they arise from nonnegativity constraints on income. As is well known (see, for instance, Milgrom and Roberts [27]), there is one special case where little is gained by endogenizing the individually rational utilities, namely when individuals' payoffs are fully transferable. There, a partial equilibrium approach suffices: one finds the contractual and technological choice and type matches that maximize the total surplus, and individual rationality can then be satisfied with lump-sum transfers.

free-rider problems. There is an additional, equally important, trade-off between the size of surplus generated and the flexibility to distribute it. The M-firm tends to have more flexibility, although it may yield a smaller surplus (for incentive compatibility, shares in the I-firm must exceed the smallest feasible share in the M-firm, but the M-firm uses extra resources in order to monitor). Sometimes the two trade-offs come into conflict, and the resolution may be technically inefficient in the sense that monitoring of effort is used while the firm could afford not to use it. Hence, an organizational form may serve as an instrument for wealthy agents to earn high returns to wealth, even if it is not an efficient means of providing incentives.

Whether this kind of outcome can arise depends jointly on the efficiency of the financial market and the distribution of wealth. When the external verification problem is absent (if the cost of auditing output is small) the map from wealth levels to equilibrium utility levels is invariant to the distribution. Moreover, in this case, every equilibrium allocation is equivalent to one in which firms share the surplus equally among their members and consist of agents with equal wealth. Of course, distribution may affect the aggregate surplus created by the economy. Nevertheless, this effect of the distribution is purely compositional.

With a severe enough financial market imperfection (when the cost of auditing is large), distribution matters in an important way, both for the allocation of surpluses in firms and for the types of firms that emerge. We show that M-firms which are technically inefficient may arise under certain conditions. In an example, when wealth is distributed unequally, all firms are technically inefficient M-firms.[2] When wealth is distributed more equally, all firms are (technically efficient) I-firms. The equilibrium map from wealth to surplus and firm type is no longer invariant: distribution matters.

Our model's structure is similar to those of club theory: a firm is simply a coalition of agents assembled to engage in production. The ways in which they generate and distribute surplus must adhere not only to technological constraints, as in the standard club-theoretic framework, but also must satisfy incentive compatibility constraints. We set up this model in the next section. In Section 3 we define the feasible allocations for coalitions and establish the existence and constrained Pareto optimality of an equilibrium. In that section we note that wealth effects manifest themselves in two distinct ways pertinent to the trade-offs between firm types. Section 4 begins with a numerical example which illustrates most of the main points of our approach and then presents some more general characterization results. We conclude in Section 5.

[2] It is not the use of the monitoring technology per se that is technically inefficient. For instance, if the economy consists of poor agents, M-firms could form and be technically efficient because the agents could not afford to post the incentive bond.

2. THE ECONOMY

2.1. *Preferences and Demographics*

The economy lasts one period and has a single storable consumption good which may also be used 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. All agents have one unit of indivisible effort and one unit of indivisible time, 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, $\Omega = \{\omega_0, ..., \omega_L\}$ where ω_i is increasing in i. Let h^i denote the measure of agents with wealth ω_i, $H(\omega)$ the measure of agents with wealth less than ω. Agents have identical risk-neutral preferences which may be summarized by the von Neumann–Morgenstern expected utility $E\{y - e\}$, where $y \geqslant 0$ is the realized lifetime income and $e \in \{0, 1\}$ is the effort level chosen. Observe that the source of wealth effects in this model is the lower bound on income, which simultaneously imposes a lower bound on an agent's utility. Thus, despite the fact that everyone is risk-neutral, payoffs are not necessarily fully transferable.

2.2. *Technology*

Agents' economic activity surrounds four technologies. First, there is a perfectly divisible safe asset which earns an exogenous gross return $r > 0$. By arbitrage, this return is also earned by lenders of capital (one could also think of our economy as small and open, with r the world gross interest rate). Second, a fixed cost of K_I must be incurred before production; 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 $1 - \pi_n$. We make the following assumptions about the function π:

Assumption 1. $\pi_n < 1$ *is nondecreasing in* n *with* $\pi_0 = 0$.

Assumption 2. $\exists \tilde{n} : \pi_n - \pi_{n-1}$ *is increasing for* $n \leqslant \tilde{n}$ *and nonincreasing for* $n \geqslant \tilde{n}$.

These assumptions give the expected output function $\pi_n R$ the standard sigmoid shape; since there is a region of increasing returns, firms will typically consist of two or more agents.

Productive effort is not directly observable without the use of the third technology. If in addition to K_I, the firm makes a fixed capital investment of $K_M - K_I$, then not only is it possible to produce, but also to verify its members' effort; think of $K_M - K_I$ as the cost of a factory building inside of which it is easy to monitor the project's participants. In order to ensure that projects can be viably operated with the monitoring technology (and *a fortiori* without it), we make

Assumption 3. $\exists n : \pi_n R - K_M r - n > 0$.

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). It costs γ to audit the output and learn whether the project succeeded; this information becomes public knowledge.

2.3. *Information*

Information is held symmetrically except as noted above. Specifically:

Assumption 4. *The output is common knowledge to the members of a firm, but is unobserved by nonmembers.*

Assumption 5. *The parties to a contract with an agent can observe all other contracts he may have signed.*

Assumption 6. *Wealth can be costlessly verified.*

The first assumption gives rise to the costly state verification problem. The other two allow us to ignore adverse selection; only moral hazard (in its hidden-action and hidden-information forms) plays a role in this paper.

2.4. *Occupations and Organizations*

Given the technological and information assumptions (including those on the indivisibility of time and effort), there are three things an agent can do in this economy. First, he may invest all of his wealth in the safe asset (or lend to other agents—these activities yield the same return) and expend no effort or time. This option is called *subsistence*, and it yields income $\omega(a)r$ to agent a. Often, no one chooses subsistence in equilibrium, but it always provides a lower bound on utility for the other occupations.

If he does not choose subsistence, an agent becomes a member of a firm, spending his unit of time on a project, and contributing part or all of his wealth to it (any remainder can be invested in the safe asset); a firm is simply the set of agents who spend their time on a given project. Since his time is indivisible, an agent can belong to at most one firm. We call a firm with a capital investment of K_k, $k = I, M$, a k-firm. Typically a firm's members will be *workers*, expending their effort on the project. However, because a member has an information advantage over an outsider, there may be situations in which some members are brought in merely for their capital and are not expected to work; we call the agents who select this third occupation *silent partners*.

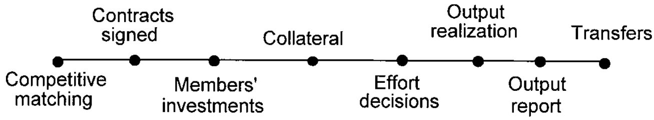

FIGURE 1

3. FEASIBILITY AND EQUILIBRIUM

The timing structure of our model is illustrated in Figure 1. It is quite standard for models of contracting and organization except for the presence of a competitive matching stage in which the firms form.

When an agent joins a firm, he contributes all or part of his wealth to its investment fund, which is then sunk directly into the firm's project, or—equivalently in this risk-neutral, complete-contract world—used as collateral by the firm when it goes to the financial market.

The "competitive" equilibrium concept we use here is the core, which has proven convenient in the study of organization in information-constrained economies (see, for instance, Boyd and Prescott [7] and Boyd *et al.* [8]), as well as in an earlier literature concerned with endogenous firm formation (e.g., Ichiishi [16]) and in the classical theory of clubs [31]. 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 members a payoff higher than what they are getting in their current coalitions. We interpret the coalitions as firms or enterprises. Lenders are not part of explicit coalitions, and could be thought of either as being outside the economy, or as costlessly operating intermediaries who accept the safe asset investments of the economy's agents as deposits.[3]

3.1. *Contracts and Feasibility*

Before defining our equilibrium concept formally, we must specify what is meant by a feasible contract for a firm. Let F be a (finite) set of agents. A *contract* for F is $c=(K_k, p, x, \bar{e})$, where K_k, $k=I, M$, is the capital

[3] As we discuss below, it turns out that following our assumptions on the technology and wealth space, whatever allocation an arbitrary coalition can achieve, can also be generated by a collection of finite coalitions. Thus, properly speaking, we will be applying a variant of the core concept known as the *f*-core (or "finite core," which is just like the standard core in continuum economies except that only finite coalitions are permitted to block), due to Kaneko and Wooders [17] and [18] (see also Hammond *et al.* [10] and Wooders [35]).

investment; p is the probability of audit, i.e., the probability that a lender employs the output verification technology following a report of failure by the borrowing firm; $\bar{e}: F \rightarrow \{0, 1\}$ defines the effort levels that the agents are expected to exert, $x: F \rightarrow \mathbb{R}_+$ defines the income of the agents when the project succeeds and when each agent exerts the effort $\bar{e}(a)$.[4] We now define the set of feasible contracts.

3.1.1. *Finance*

Our model of the financial market closely resembles the standard one in the costly state verification literature ([4, 6, 28, 33]). The supply side of the financial market is competitive with free entry (we treat it here as devoid of agency problems, although a natural extension would allow for active financial intermediation as a separate occupation, as in Diamond [9] or Boyd and Prescott [7]). The firm F puts up its members' wealth $\omega_F \equiv \sum_F \omega(a)$ as collateral, receives a loan K_k, and then carries out its production activities.[5] Once the outcome of the project becomes known to the firm's members, they report it to the lender. If they report success, each agent in the firm obtains his share $x(a)$.[6]

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. The state contingent transfers to the agents must therefore satisfy a truth-telling constraint; in addition they must insure the lender with a nonnegative expected profit and be consistent with the nonnegativity of income:

[4] We prove in our working paper [23] that there is no need to consider more complex contracts.

[5] Since it makes no difference whether the firm provides collateral $C < \omega_F$ and borrows $K_k - (\omega_F - C)$, we assume the collateral is ω_F and do not specify it in the contract. It is routine to show that agents have a weakly dominant strategy to contribute all of their initial wealth in their firm: as long as $\omega_F < K_k$, it is better to reduce the size of the loan or increase the size of the collateral than to invest in the safe asset.

[6] The reader will note that we have ruled out the possibility that the firm's members contract individually with the lender and send separate messages to him. Under the standard assumptions of contract theory, such arrangements can be designed to elicit information about the firm's success at no cost to the principal (in particular without auditing), since this information is common among the firm's members. For simplicity, we follow Tirole [32] in supposing that agents can perfectly enforce side-contracts. Hence, we suppose that when communicating with the lender, agents in a firm behave like a unique agent; in particular, there is no loss of generality in supposing that a unique message is transmitted from the agents to the lender. This assumption is extreme; in a companion paper [26], we develop a model in which the messages that are sent between the agents and the lender/principal are "falsifiable" and in which collusive agreements can be sustained non-cooperatively, i.e., in which whistle-blowing contracts are not feasible. One could also think of γ as including costs of extracting verifiable messages from the firm's members.

$$\sum_F x(a) \geqslant (1-p)R \tag{1}$$

$$\pi_n \sum_F x(a) + (1-\pi_n)\, p\gamma \leqslant \pi_n R + (\omega_F - K_k) r \tag{2}$$

$$\forall a \in F : x(a) \geqslant 0 \tag{3}$$

We show in our working paper [23] that equilibrium financial contracts maximize the firm's expected income $\pi_n \sum_F x(a)$, subject to these constraints.[7] As is well-known (see e.g. [4]), the solution to this program is derived by showing that (1) and (2) bind, and then solving the resulting equations simultaneously for $\sum_F x(a)$ and p. This procedure yields

$$\pi_n \sum_F x(a) = \begin{cases} \pi_n R + (\omega_F - K_k) r, & \text{if} \quad \omega_F \geqslant K_k \\ \pi_n R + (\omega_F - K_k)\, \alpha_n r, & \text{if} \quad \omega_F < K_k \end{cases} \tag{4}$$

$$p = \frac{(K_k - \omega_F) r}{\pi_n R - (1-\pi_n)\gamma} \tag{5}$$

where $\alpha_n \equiv (\pi_n R/(\pi_n R - (1-\pi_n)\gamma)) > 1$ whenever $(\pi_n R/(1-\pi_n)) > \gamma > 0$. (If $\pi_n R - (1-\pi_n)\gamma \leqslant 0$, then external finance is infeasible, while if $\gamma = 0$, external and internal finance are equally costly.) Here we see the first wealth effect: as measured by the size of α_n, firms which need external finance face a higher marginal cost of capital and a higher marginal return to initial wealth than do firms which can finance internally. (Equivalently, a firm with lower wealth faces a higher audit probability p; however, it is more convenient to use α_n than p, and we shall no longer have occasion to refer to the latter.) It is because of this wedge between internal and external financing costs that there is an incentive to bring in members as silent partners: because they are informed about output, they can effectively lend to the firm at less than the rate $\alpha_n r$ offered by the financial market. Of course, silent partners have an opportunity cost of their time (they could be workers in other firms), so they need not appear in equilibrium.

3.1.2. *Share*

The share contract x distributes the proceeds of the project among the firm's members. In the M-firm, which uses the monitoring technology, the

[7] Note that this is a *result* of competition, rather than a behavioral assumption, as in the principal-agent framework. There, it is well known [6, 4] that nothing is lost by supposing that firms get a zero income in failure states. This fact as well as the maximization of firm's expected income are more delicate to establish in our model because it is not always straightforward—especially in the presence of silent partners—to distribute the gains from increasing a firm's income among all its members without distorting any of their incentives to provide effort.

share contract is written contingent on the effort of members as well as on the firm's output. The I-firm contract can only be written contingent on the output of the enterprise.

I-firms. 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, and so has no incentive to work at the level that maximizes the expected surplus (Alchian and Demsetz [1]; Holmström [14]). Under some conditions, though, I-firms can achieve the maximum surplus (Williams and Radner [34], Legros and Matsushima [21]) and under others can approximate as closely as desired the maximum surplus (Legros [20], Legros and Matthews [22]). In the present model, in fact, an I-firm may be able to achieve the maximum surplus (an M-firm never does, since K_M units of capital are required to produce $\pi_n R$, while the I-firm uses only K_I).[8]

If an I-firm has n agents working, $\sum_F \bar{e}(a) = n$, and the share contract x must induce a game in effort levels in which $\bar{e}$ is a Nash equilibrium. Hence, writing $\Delta\pi_n \equiv \pi_n - \pi_{n-1}$, the following incentive compatibility constraints are satisfied:

$$\begin{aligned} &\bar{e}(a) = 1 \Rightarrow \pi_n x(a) - 1 \geqslant \pi_{n-1} x(a) \qquad \text{or} \quad \Delta\pi_n x(a) \geqslant 1 \\ &\bar{e}(a) = 0 \Rightarrow \Delta\pi_{n+1} x(a) \leqslant 1. \end{aligned} \tag{6}$$

Using the fact that $x(a) \geqslant 0$ and adding up over all agents in the firm, (6) implies that $\sum_F x(a) \geqslant n/\Delta\pi_n$. Combining this with (4) and rearranging yields

$$\begin{aligned} &\omega_F \geqslant K_I + \frac{\pi_n}{\alpha_n r}\left(\frac{n}{\Delta\pi_n} - R\right), \qquad \omega_F < K_I \\ &\omega_F \geqslant K_I + \frac{\pi_n}{r}\left(\frac{n}{\Delta\pi_n} - R\right), \qquad \omega_F \geqslant K_I, \end{aligned} \tag{7}$$

indicating that the firm requires a certain minimum wealth just to satisfy the incentive constraints. In effect the firm's members need to post incentive bonds; this is the second way that wealth effects manifest themselves, and is unique to the I-firm. For later use, let us denote the right hand side of Eq. (7) when $\omega_F < K_I$ by $\underline{\omega}(n, \gamma)$ and write $\underline{\omega}(n) = \omega(n, 0)$.

M-Firms. In this case, effort levels are observable, and shares can be contingent on output and effort levels. By the principle of maximum

[8] As we have said, the observed organizational form will result from a trade-off among the costs of each, so whether we imagine I-firms to be fully efficient or only nearly so doesn't matter for the general point. In either case, efficiency doesn't determine the outcome.

punishment, the cheapest way to implement a profile of effort levels $\bar{e}$ is by giving a share of zero to the agents who are detected shirking.[9] It follows that the incentive compatibility conditions are simply

$$\begin{aligned} \bar{e}(a) = 1 &\Rightarrow \pi_n x(a) \geqslant 1 \\ \bar{e}(a) = 0 &\Rightarrow \pi_n x(a) \geqslant 0. \end{aligned} \tag{8}$$

At this point we can point out the principle difference in the restrictions on sharing rules between *I*-firms and *M*-firms. As we have said, an *I*-firm generates a greater surplus for a fixed n provided it can satisfy (6). But note that for any working member, (6) implies that the expected income is equal to $\pi_n x(a) \geqslant \pi_n / \Delta\pi_n$, which is strictly larger than the effort disutility of unity if $n \geqslant 2$. Note that the minimum wealth/income constraint is crucial here: without it, (6) could always be satisfied without making expected income larger than unity, simply by making the payoff when there is failure low enough. On the other hand, in an *M*-firm, the expected compensation may be as low as unity and still satisfy (8): indeed, if it is individually rational to join an *M*-firm, it is also incentive compatible.

The difference between the minimum feasible compensations in each type of firm creates the possibility that an agent may prefer to join the inefficient *M*-firm rather than the more efficient *I*-firm: the smaller wage bill in the *M*-firm may leave him more surplus, even after allowing for the expenditure on monitoring. Of course, whether this possibility can be realized in equilibrium, that is whether the *M*-firm compensation is indeed bid low enough, is precisely the general equilibrium question we need to answer.

For a finite set F, let $C(F)$ be the set of feasible contracts, i.e., contracts $(K_k, p, x, \bar{e})$ that satisfy conditions (1), (2), (3), (6), and (8). Observe that the principle departure from the usual notion of feasibility for principal–agent contracts is the absence of any participation constraint: as we have said, the individually rational utility levels are endogenous variables here. For $a \in F$ and $c = (K_k, p, x, \bar{e})$ where $\sum_a \bar{e}(a) = n$, we define the *surplus* of agent a who is party to contract c by

$$u(a|c) = \pi_n x(a) - \omega(a) r - \bar{e}(a),$$

that is, as the utility gain yielded by the contract over subsistence.

[9] More precisely, let e be the observed vector of effort levels and let $x(a \mid e)$ be the income of agent a when the project succeeds and when the vector of observed effort levels is e. If $e = \bar{e}$ or if $e = 0$, let $x(a \mid e) = x(a)$. If $e \neq \bar{e}$ and if $e \neq 0$, let $G(e) = \{a \in F : e(a) \neq \bar{e}(a)\}$ be the set of agents who deviate from their contracted effort level and let $x(a \mid e) = 0$ if $a \in G(e)$ and $x(a|e) = x(a) + \sum_{G(e)} (x(b)/\#G(e))$ if $a \in F \backslash G(e)$. (It is necessary to have $\sum_F x(a|e) = \sum_F x(a)$ for all e, including $e = 0$, in order to satisfy the truth-telling constraint.)

3.2. *Equilibrium*

Defining equilibrium in this economy follows the standard club-theoretic approach: having specified what is feasible for any (finite) group of agents, we need only specify what groups will form. Since 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.[10] Call a partition of A into finite sets *measure-consistent* if for all positive integers m and for all i, $j = 1, ..., m$, the set of all ith members of size-m elements of the partition has the same measure as the set of all jth members of those elements (see Kaneko and Wooders [8]).[11]

We now provide our definition of equilibrium.

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

(i) for almost every $F \in \mathscr{P}^*$, $c^*(F) \in C(F)$,

(ii) For all finite $T \subseteq A$ and $c \in C(T)$, there exists $a \in T$ such that $u(a|c) \leqslant u(a \mid c^*(F(a)))$ where $F(a)$ is the element of $\mathscr{P}^*$ containing a.

The second condition (ii) is 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 and Wooders in ruling out the possibility of blocking by infinite coalitions (feasibility is not even defined for them). In the present context, we feel this restriction is justified. Recall that a continuum economy is simply an approximation to a large finite economy. 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

[10] For instance, let N be a fixed number and suppose that $\Omega = \{0, \omega\}$, where ω is large enough to cover the capital requirement of a project. Suppose that all agents in $[0, (1/(2N+1))]$ have wealth ω (are rich) and that the other agents have wealth 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 that same organization, i.e., one rich and N poor.

[11] We also place a further restriction on the equilibrium partition, namely that it satisfies "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.

blocks.[12] This upper bound is independent of the size of the finite economy and thus applies as well to the continuum economy taken as the limit of some sequence of finite economies. The same kind of argument shows that membership in all equilibrium firms will be uniformly bounded by $\hat{n}$.

The proof of existence is a straightforward application of the results in the literature on the f-core and is in the Appendix. Moreover, since from what we have just said, the grand coalition cannot achieve any allocation that is not achieved by a collection of finite coalitions, we also obtain Pareto optimality of the equilibrium allocation (optimality is, of course, of the constrained sort, by (1), (2), (3), (6), (8), and Assumptions 4–6).

PROPOSITION 2. *An equilibrium exists and is constrained Pareto optimal.*

We emphasize the Pareto optimality of the equilibrium, since the allocation typically will not satisfy other common efficiency notions such as surplus maximization or technical efficiency. As alluded to above, the lower bound constraint on agents' income implies that payoffs are not fully transferable, despite risk neutrality. A firm is *technically inefficient* if it is possible to produce the same output by using less input; it is *surplus inefficient* if there is a feasible way for it to generate more surplus. In our context, the output of a firm is the probability of success, and a technically inefficient firm could generate more surplus by *reorganization*: M-firms can be technically inefficient if the firm can be organized as an I-firm with the same number of agents working. Surplus inefficiency is the more familiar concept of inefficiency encountered in principal–agent models. It occurs, of course, whenever technical inefficiency does, but may also arise if, for instance, a firm operates with a number of working members other than the one which maximizes the surplus.

Since failures of technical and surplus efficiency are typically more evident than failures of Pareto optimality, they are often the focus of policy or popular discussion about possibilities for reorganization in corporations,

[12] The formal proof is somewhat involved and is omitted, but the intuitive argument is simple enough. The production technology implies that there is an upper bound $\bar{\bar{N}}$ on the number of working agents in the firm (by increasing the number of working agents beyond that number, the total surplus in the firm decreases and the incentive problem becomes more severe). The only problem is finding a bound on the number of agents who are not expected to work. It turns out that there is a finite bound $\bar{\bar{\omega}}$ on the total amount of wealth that a firm ever needs; 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 non-working agents who add wealth over $\bar{\bar{\omega}}$ or who add zero wealth do not contribute to the total surplus. Let ω^0 be the minimum positive wealth (which exists by finiteness of Ω). Thus $\hat{n}$ can be chosen equal to be any integer exceeding $\bar{\bar{N}} + \bar{\bar{\omega}}/\omega^0$.

health care, or the former Soviet Union. Nontransferabilities such as the wealth effects studied here give rise to conflicts between surplus or technical efficiency and Pareto optimality and help to explain why apparently desirable reorganizations often face considerable resistance.

4. SOME CHARACTERIZATIONS

We are now in a position to study how organization is determined in equilibrium. We will proceed first by studying a simple numerical example, and then will give some more general results.

4.1. *An Example*

Consider an economy in which $\pi_0 = 0$, $\pi_1 = 0.1$, $\pi_2 = 0.6$, and $\pi_n = 0.8$, for $n \geqslant 3$. Let $R = 15$, $r = 1$, $K_I = 1$, and $K_M = 2$. Finally, suppose there are just three wealth levels, namely 0, 2, and 4 with corresponding fractions of the population h^0, h^1, and h^2. Any single-person firm generates negative surplus, and average surplus is maximized at $n = 2$ for both *I*-firms and *M*-firms. The first-best allocation therefore consists entirely of two-agent *I*-firms.

The first best is achieved when $\gamma = 0$. To see this, note that in this case $\underline{\omega}(2)$ is negative, so that even agents with wealth zero (call them "poor") can satisfy (6) if they try to form I_2-firms (refer to an n-person k-firm as k_n-firm). Any allocation which did not give almost every agent at least one half of the I_2-firm surplus would be blocked by pairs of agents forming their own *I*-firms. Thus, all agents will choose to enter into this kind of contract, regardless of the distribution of wealth. Finally, the matching of types (i.e., wealth levels) within firms is indeterminate in this case, since everyone gets the same surplus no matter who is in the firm they join; in particular the surplus and firm type allocation is always the same as the one which arises when every firm contains agents of a single wealth level.

Things are rather different when γ is sufficiently large, however (say $\gamma > 45$). In this case, both types of firms, at any size, are either infeasible or generate negative surplus when composed entirely of poor agents. Thus, every firm will have to have at least one agent with positive wealth. Each firm will then have enough capital to satisfy the aggregate *I*-firm incentive compatibility constraint. Since every firm would then be fully financed internally, there would be no role for silent partners.

We shall be interested in how the organization of firms varies with the distribution. Agents with wealth 2 and 4 (the "rich") behave identically from the point of view of firm membership, so the parameter of interest for examining the effect of distribution on organization is h^0.[13]

We begin by asking under what conditions a rich agent will prefer to join each of the possible firm types and sizes. Observe first that he will always (weakly) prefer to have poor agents in his firm rather than other rich agents. Indeed, poor agents can never get more surplus than rich agents in equilibrium, since the rich agents could then break off and form their own firms. Denote the equilibrium M-firm compensation going to a poor agent by w (this is independent of the size of the firm), and the corresponding compensation in an I_n-firm by v_n. The rich agent will choose the organization and size of firm which give him the highest residual incomes, that is he will compare $\max_n\{\pi_n R - K_I r - (n-1)v_n\}$ and $\max_n\{\pi_n R - K_M r - (n-1)w\}$. It is easy to show that only M_3-firms and I_2-firms can arise in equilibrium.[14] In an M_3-firm, the total expected income is $\pi_3 R - K_M r = 10$ and in an I_2-firm it is $\pi_2 R - K_I r = 8$.

Suppose first that the poor are relatively scarce ($h^0 < 1/2$). Then even if each rich agent matches with only one poor agent, there will be leftover rich agents to keep bidding up their (I_2-firm) compensation until it reaches 4. It follows that all firms are I_2-firms in which everyone receives a compensation of 4; some are composed of one rich agent and one poor one; the remainder consist of two rich ones. Each of these firms is efficient in both the technical sense and the surplus-maximizing sense, and indeed the economy performs at its first-best level.

On the other hand, if there are many poor ($h^0 > 2/3$), then even if every rich agent matches with two poor, there are still unmatched poor. Thus w must be bid down to its minimum value of unity. This yields an income of 8 to the rich agent in an M_3-firm, which he prefers to the maximum of 6.8 that he can obtain in an I_2-firm. Some of the poor will remain in subsistence, although they are indifferent between that status and working for a wage of 1.

Note that each of these M_3-firms is technically inefficient since it would be feasible to reorganize them as I_3-firms, thereby producing the same expected output with less input: each firm already has enough wealth to

[13] 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.

[14] A rich agent can always collect 4 by splitting the proceeds of an I_2-firm formed with another rich agent. Since $\pi_3/\Delta\pi_3 = 4$, an I_3-firm can never yield him more than 3 (the maximum residual income to the rich agent is $11 - 2(\pi_3/\Delta\pi_3)$, and so will not appear in equilibrium. We can also rule out the M_2-firm as an equilibrium organization. To see this, note that the M_2-firm would be preferable to the M_3-firm only if $w \geqslant 3$. But w cannot exceed 3 because that yields the rich agent less than 4. If $w = 3$, $w > (\pi_2/\Delta\pi_2) = 1.2$, and a rich agent would prefer to get 5 by reorganizing as an I_2-firm.

satisfy (7) and therefore could find a contract paying each member at least $\pi_3/\Delta\pi_3 = 4$. The reason this does not happen in equilibrium is that the poor agents' outside opportunities are limited by the imperfect capital market, so that rich agents are better off paying them the low M_3-firm compensation brought about by the relative abundance of poor.

Of course these firms are also too large compared to their first-best size: net output would be larger if every two M_3-firms were replaced by three M_2-firms (or I_2-firms). This outcome would entail that more capital flow from the rich to the poor via the financial market; but the imperfection prevents this from happening.

For $1/2 < h^0 < 2/3$, equilibrium requires that rich agents be indifferent between the two types of firm, as must poor agents: $w = v_2$ and $8 - v_2 = 10 - 2w$ together imply that $w = 2$. The fraction μ of M_3-firms is determined by the requirement that the demand for the poor equals the supply:

$$2\mu[1 - h^0] + \mu[1 - h^0] = h^0.$$

The graphs of μ along with the incomes accruing to rich and poor agents are shown in Figs. 2a and 2b. Note that the income distribution becomes more equal when the wealth distribution does. Figure 2c shows what firm types prevail in different parts of the distribution space, and in particular that economies with different organizational structures may have the same mean wealth.

Besides illustrating how the internal organization of the firm depends on the external economy (that is, on the efficiency of the financial market and the distribution of initial endowments), this example also underscores how the *function* of the firm may also vary with these factors. Specifically, it is fair to say that the contractual form selected when $h^0 < 1/2$ is indeed the one which optimally provides incentives. But in the case $h^0 > 2/3$, a rent-seeking function of the organization predominates: incentives could be provided more cheaply by organizing as I_3-firms, but rich agents would earn less surplus that way. In cases like these, policies such as the taxation of high incomes or the imposition of minimum wages may cause firms to reorganize, possibly with a gain in social surplus. Note however that the equilibrium is (constrained) Pareto optimal. This illustrates the tension between surplus or technical efficiency and Pareto optimality brought about by wealth effects.[15]

[15] In this instance, a 100% tax on total income greater than 6 would cause all of the M_3-firms to be reorganized as I_3-firms, which would increase the total surplus by $1 - h^0$. But the gain in surplus cannot be turned into a Pareto improvement, because if the poor agents were to compensate the wealthy, their final incomes would either violate the nonnegativity constraint in the failure state, or would be too small in the success state to satisfy (6). Imposing a minimum wage of 2 would turn all enterprises into I_2-firms, although this would not raise total social surplus: capital could still not flow to enough people to increase the number of firms sufficiently to compensate for the decreased output of existing firms.

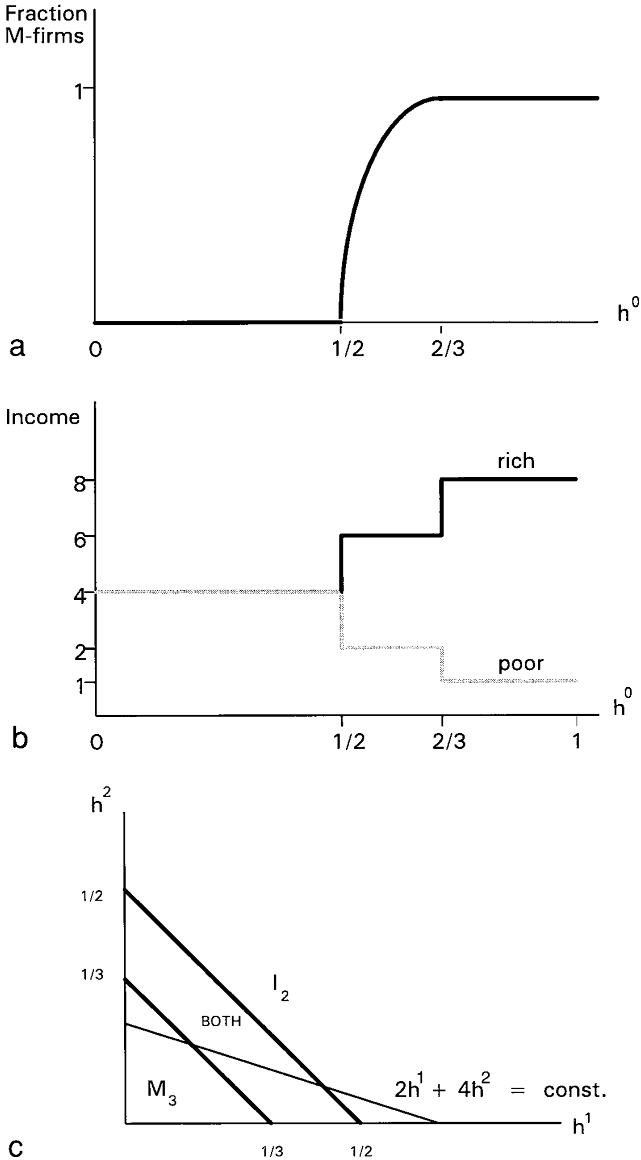
Fraction
M-firms
1
h0
a
0
1/2
2/3
Income
rich
8
6
4
2
1
poor
h0
b
0
1/2
2/3
1
h2
1/2
1/3
I2
BOTH
M3
2h1 + 4h2 = const.
h1
c
1/3
1/2

FIGURE 2

4.2. *The Roles of the Financial Market and the Wealth Distribution*

We proceed now to a more general consideration of the properties of equilibrium organizations. We divide our discussion into two cases corresponding to different assumptions about the efficiency of the financial market.

4.2.1. *The case* $\gamma = 0$

Our economies are parametrized by (Ω, H) and γ and r. For a fixed γ, we refer to a (Ω, H)-economy and when $\Omega = \{\omega\}$ is a singleton, we call the resulting economy an ω-economy.

When $\gamma = 0$, the characterization of equilibria of ω-economies is quite simple. First, note that all agents provide effort in equilibrium firms since it is not cheaper to borrow from within the firm than from the capital market. By equal treatment,[16] it follows that the agents will form firms that maximize the average surplus.

Let $V_k(n) \equiv \pi_n R - K_k r - n$ be the total surplus in a k-firm with n working members. Let $n(\omega)$ denote the size of an I-firm that maximizes the average surplus subject to the constraint that the agents have enough wealth to satisfy incentive compatibility:

$$n(\omega) = \begin{cases} \arg\max \left\{ \dfrac{V_I(n)}{n} \,\middle|\, n\omega \geqslant \underline{\omega}(n) \right\}, & \text{if} \quad \arg\max \neq \varnothing \\ 0, & \text{otherwise.} \end{cases}$$

Let N_M be the size of an M-firm that maximizes average surplus $(V_M(n))/n$.[17]

Then in equilibrium, almost every agent receives a surplus of

$$u^e(\omega) = \max \left\{ \frac{V_I(n(\omega))}{n(\omega)}, \frac{V_M(N_M)}{N_M}, 0 \right\} \tag{9}$$

with almost every equilibrium firm of the type for which $u^e(\omega)$ is defined.

The map defined by (9) represents the minimum surplus that an agent with initial wealth ω can obtain in any (Ω, H)-economy where $\omega \in \Omega$. Indeed, if there is a positive measure of agents of wealth ω who do not obtain $u^e(\omega)$ in the equilibrium of the (Ω, H)-economy, they could behave as in the ω-economy and obtain that surplus. One might think that when the economy consists of heterogenous types, the agents with larger wealth

[16] See Lemma 1 in our working paper [23] for a proof that agents of the same wealth *who are in the same type of firm* have the same equilibrium utility.

[17] It is straightforward to show that $n(\omega)$ and N_M are uniquely determined.

could obtain a higher surplus than $u^e(\omega)$ by forming firms with agents with lower wealth. This intuition is incorrect when $\gamma = 0$, and it is a remarkable fact that any equilibrium of a (Ω, H)-economy replicates the equilibria of the individual ω-economies.

PROPOSITION. *If $\gamma = 0$, then for any (Ω, H), any equilibrium $(\mathscr{P}^*, c^*)$, and almost every agent a, $u(a \mid c^*) = u^e(\omega(a))$. Moreover, almost every agent a belongs to the same type of firm as in the $\omega(a)$-economy equilibrium.*

Proof. Suppose that in an equilibrium of a (Ω, H)-economy, there is a positive measure of M-firms with n working members. Then, $V_M(n) \geqslant 0$ and moreover, each agent in the M-firm obtains at least the surplus given by (9). Therefore for each agent a in the firm, forming an M-firm in the $\omega(a)$-economy is optimal. Hence, by definition of N_M, equilibrium M-firms must be of size N_M, and each agent gets the average surplus $(V_M(N_M))/N_M$.

Consider now the case in which there is a positive measure of I_n-firms. We claim that for almost every I_n-firm and each a in the firm, $n\omega(a) \geqslant \underline{\omega}(n)$, i.e., that the I_n-firm leads to a positive surplus in the $\omega(a)$-economy. Suppose instead that for some agent a in the I_n-firm, $n\omega(a) < \underline{\omega}(n)$. By definition of $\underline{\omega}(n)$, this is equivalent to

$$\frac{\pi_n}{\Delta\pi_n} - \omega(a)r - 1 > \frac{V_I(n)}{n}. \tag{10}$$

Note that the left hand side is the minimum surplus a can obtain in an I_n-firm while remaining incentive compatible. Let b be an agent in a's firm for whom $n\omega(b) > \underline{\omega}(n)$ (such an agent must exist, since by hypothesis, the firm's wealth is at least $\underline{\omega}(n)$). Then (10) implies that there exists such an agent b whose surplus is strictly less than the average surplus $(V_I(n))/n$. But now we have a contradiction, since n agents with wealth $\omega(b)$ could obtain a greater surplus by forming an I_n-firm on their own.

We conclude that each agent a in an I_n-firm can afford this type of firm in the $\omega(a)$-economy, i.e., $n\omega(a) \geqslant \underline{\omega}(n)$. It follows that in almost all I_n-firms, the surplus is equally shared. But then, (9) implies that in an I_n-firm, $n = n(\omega(n))$. ∎

The intuition for the result is the following. If agents with large wealth form an I-firm with agents who could not afford that type of firm when they are restricted to match together, the richer agents have to subsidize the incentive compensation of the poorer agents: the poorer agents cannot post the incentive bond since otherwise they would be able to create the firm on their own. Hence, in any equilibrium of any (Ω, H)-economy, there is (weak) *segregation*:[18] agents who are matched together are those who have the same $u^e(\omega)$; often, this entails that agents matched together have the same wealth. In general, richer agents form I-firms and poorer agents form M-firms or are idle.

Clearly, in an ω-economy the total equilibrium surplus is maximized (which implies that each firm maximizes the total surplus of its members and is therefore technically efficient). In (Ω, H)-economies, equilibrium firms are still maximizing the surplus of their members but the equilibrium might fail to maximize the total surplus because of inefficiencies in matching.

For instance, suppose that there are two atoms, one at 0 and the other at ω. Suppose that M-firms are not feasible, that ω is large enough for the equilibrium firms of the ω-economy to be I_n-firms and that agents use subsistence in the 0-economy. From Proposition 3, in the (Ω, H) economy, agents with wealth ω form I_n firms and agents with wealth 0 are in subsistence. If $(n-1)\omega > \underline{\omega}(n)$ it would be feasible for $(n-1)$ agents with wealth ω and an agent of wealth 0 to create a I_n-firm. In equilibrium there are h^ω/n I_n-firms while it would be feasible to have $h^\omega/(n-1)$ firms of type I_n. Hence total surplus is not maximized in equilibrium. As we have argued above, the richer agents are unwilling to match with poor agents because poor agents can be made incentive compatible only if their compensation comes out of the wealth of the rich agents (something that the rich agents do not need to do when they match together since each of them is able to post a bond large enough for his own incentive compatibility).

When $\gamma = 0$, the only inefficiencies are in matching. Equilibrium firms maximize the total surplus of their members over the set of feasible contracts. What is striking is that this property is independent of the distribution: changing the distribution will not affect the equilibrium surplus of the agents nor the types of firms to which they belong (it may, of course, affect the total measure of firms of different types in the economy, but this effect is purely compositional). Hence when $\gamma = 0$, distribution does not matter: in order to understand the behavior of (Ω, H)-economies, it is enough to analyze the behavior of ω-economies.

4.2.2. *The case* $\gamma > 0$

As we have seen in the example of Section 4.1, when γ is positive, distribution does matter. The equilibrium map from individual wealth to

[18] This result holds more broadly. For instance, weak segregation holds in a model in which the monitoring technology is imperfect, i.e., enables the verification of effort levels with probability q, where q takes values in $[0, 1]$ and where $K(q, n)$ is the capital cost necessary to use monitoring technology q with n working members. In [24] we provide the necessary and sufficient condition on the characteristic function to obtain weak segregation in more general matching problems.

surplus and firm type is not invariant with respect to the distribution. Moreover, equilibrium firms are technically inefficient for some distributions. The object of this section is to explain the emergence of heterogenous matching and of technically inefficient firms.

To understand why heterogeneous matches can occur when γ is large, first consider I-firms. Note that opposite to the case $\gamma=0$, an I_n-firm can include agents who could not afford it if they had to be matched with agents of their own type. Indeed, if a cannot afford I_n in the $\omega(a)$-economy, then $n\omega(a) < \underline{\omega}(n, \gamma)$, which is equivalent to $\pi_n/\Delta\pi_n - \omega(a)r - 1 > V/n$, where $V = \pi_n R - \alpha_n K_I r + (\alpha_n - 1)\, n\omega(a) r - 1$ is less than $V_I(n)$. Therefore, it is possible for a rich agent to form an I_n-firm with poor agents while getting more than the average surplus $(V_I(n))/n$. Note that it must still be true that $n\omega(a) \geqslant \underline{\omega}(n)$, i.e., the poor agents would be able to form an I_n-firm if $\gamma=0$. This wedge between what agents can earn in matches with sufficiently wealthy agents and what they can gain in matches with agents of lower wealth provides the motive for heterogenous matching and rent seeking.

For M-firms, when γ is large enough, poor agents who form an M-firm obtain a positive surplus only if there is enough wealth inside the firm, i.e., only if there is a large number of silent partners. Therefore, if the cost of borrowing is large, poor agents who form an M-firm obtain a low average surplus. It is therefore possible for richer agents to attract poor agents in M-firms, pay them a low compensation, and obtain for themselves a higher residual surplus than if they formed an I-firm and had to pay an incentive compensation.[19]

Because only M-firms can be technically inefficient in this model, and because M-firms experience only the financial market wealth effect, in order to derive a result in which equilibrium firms are technically inefficient, two basic conditions must be met. First, there must exist a nonnegligible set of agents who cannot generate a large surplus by forming an M-firm because they have little wealth and have to pay a large interest rate to the lender. Second, there must not be too much competition among the remaining agents to attract the first set. Here we develop conditions under which these two effects appear.

To simplify the analysis and to focus on the role of the costly state verification problem, we consider economies in which the wealth effect due to the incentive problem would not be too severe if γ were small. We also assume that γ is large enough to prevent firms with zero wealth from attaining a positive surplus. Finally, we assume that there are positive masses of agents with wealth zero and with wealth greater than K_M.

[19] When $\gamma=0$, if rich agents can generate a positive surplus by creating an M-firm, *all* agents can also generate a positive surplus by creating an M-firm. Therefore, there is no possibility for a rich agent to claim more than the average surplus in any M-firm.

Assumption 7. *For all n such that* $\pi_n R - K_I r - n > 0$, $\underline{\omega}(n) < 0$.

Assumption 8. $\forall n$, $\pi_n R - \alpha_n K_I r - n < 0$.

Assumption 9. $\omega_0 = 0$, $\omega_L > K_M$, $h^0 > 0$, $h^L > 0$.[20]

We first provide a necessary condition for equilibrium firms to be technically efficient. Consider an equilibrium $(\mathscr{P}, c^*)$ and an agent a with wealth ω_L. From Assumptions 7 and 9, if a belongs to an M-firm, this firm is technically inefficient. Let

$$u^0 = \operatorname*{essinf}_{a' \in A} u(a' \mid c^*).$$

The surplus of a in an I-firm with n working members is

$$u(a \mid c^*) = \pi_n R - K_I r - \sum_{b \neq a} u(b \mid c^*) - n.$$

We show in our working paper [23] that in equilibrium the surplus $u(b \mid c^*)$ is nondecreasing in the wealth of the agent. Therefore, from Assumption 7, a weakly prefers to match with agents with minimal wealth $\omega_0 = 0$. Agents with zero wealth who work in an I_n-firm have a surplus at least equal to $(\pi_n / \Delta\pi_n) - 1$. To participate, they must obtain a surplus of at least u^0. Hence, an upper bound for the surplus that a can obtain by forming an I_n-firm is

$$u_I = \pi_{n_I} R - K_I r - (n_I - 1)\frac{\pi_{n_I}}{\Delta\pi_{n_I}} - 1,$$

where,

$$n_I = \arg\max_n \pi_n R - (n-1)\max\left\{u^0 + 1, \frac{\pi_n}{\Delta\pi_n}\right\}.$$

By definition of u^0, and the fact that $u(b \mid c^*)$ is increasing in the wealth of b, there exists a positive measure of agents with zero wealth whose equilibrium surplus is u^0. Agent a can form an M_n-firm with these agents and give them a compensation equal to $u^0 + 1$. It follows that the maximum surplus of a in an M-firm is

$$u_M = \pi_{n_M} R - K_M r - (n_M - 1)(u^0 + 1) - 1,$$

[20] Recall that ω_0 is the smallest wealth level in Ω, and ω_L the largest.

where

$$n_M = \arg\max_n \pi_n R - (n-1)(u^0+1).$$

If there are no inefficient firms, we must have $u_I \geqslant u_M$. To be precise, it is necessary that

$$u^0 \geqslant \frac{1}{n_M - 1}\left[(\pi_{n_M} - \pi_{n_I})R - (K_M - K_I)r + (n_I - 1)\frac{\pi_{n_I}}{\Delta\pi_{n_I}}\right] - 1. \qquad (11)$$

Because both n_I and n_M depend on u^0, the right hand side of (11) is also a function of u^0. This inequality is always satisfied if the right hand side is negative (for instance, if K_M is much larger than K_I). Hence, we assume.

Assumption 10. *The set U^0 of u^0 satisfying* (11) *is bounded away from zero.*

Denote by u^* the infimum of U^0. Assumption 10 then says that $u^* > 0$. Typically, the condition is satisfied if $K_M - K_I$ is small enough. We show in [23] that $n_M \in [N_M, \bar{N}]$, where N_M maximizes the per-capita surplus in an M-firm that does not borrow and has no silent partner and $\bar{N}$ maximizes the total surplus in a firm that does not borrow. That is, $N_M = \arg\max((\pi_n R - K_m r)/n)$ and $\bar{N} = \arg\max(\pi_n R - n)$. Since $\Delta\pi_n$ is decreasing in this region by Assumption 2, it follows that $n_I \leqslant n_M$, where the equality obtains when $u^0 = (\pi_{n_I}/\Delta\pi_I) - 1$.

If Assumption 10 is satisfied, then efficiency of all firms entails that the equilibrium allocation of surplus guarantees everyone a "large" payoff (i.e., one that exceeds u^*). If instead competition deals some agents less than u^*, some firms must be inefficient. We now provide a condition on the distribution of wealth for which (11) cannot hold. We establish the existence of a kind of "poverty line," below which agents are too poor to earn significant surplus when matched with others who are below the line; the condition on the distribution then ensures that the poor so defined are plentiful enough to ensure that they receive a low surplus in equilibrium.

PROPOSITION 4. *There exist $\omega^* \in \mathbb{R}$ satisfying $\omega_0 < \omega^* < \omega_L$ and a finite $n^* \geqslant 1$ such that if*

$$H(\omega^*) > (n^* - 1)(1 - H(\omega^*)), \qquad (12)$$

a positive measure of coalitions are technically inefficient M-firms.

Note that it is not necessary that $\omega^* \in \Omega$.

Proof. Let $S^* = \pi_{\bar{N}} R - K_I r - \bar{N}$ be the maximum surplus that any firm can generate. Then the maximum number of agents in a firm is bounded above by $n^* < \infty$, where $n^* = S^*/u^*$. It is straightforward to verify that $n^* \geqslant 1$.

Consider now the ω_i-economies; their equilibria are characterized by maximizing the average surplus of a firm. As in Section 4.2.1, denote this surplus by $u^e(\omega_i)$; note that by Assumption 8, $u^e(\omega_0) = 0$ and $u^e(\omega_L) = \max((\pi_n R - K_I r)/n) - 1 > u^*$. Clearly, $u^e(\omega_i)$ is nondecreasing in ω_i.

Let ω^* be such that $u^e(\omega_i) < u^*$ for $\omega_i < \omega^*$ and $u^e(\omega_i) \geqslant u^*$ for $\omega_i \geqslant \omega^*$. Thus, $\omega^* \in (0, \omega_L)$. Now, under the assumption on H in Proposition 4, if in equilibrium u^0 satisfies (11), there exists a positive measure of agents in $[\omega_0, \omega^*)$ who are matched with agents in the same interval. By definition of $u^e(\omega_i)$, it is not possible for agents in $[\omega_0, \omega^*)$ to create firms and give each agent a surplus of at least u^*. This contradicts the assumption that in equilibrium for each agent a, $u(a \mid c^*) \geqslant u^0 \geqslant u^*$. ∎

Condition (12) could be considerably weakened. First, it is not necessary that ω_0 be zero or that agents with that wealth earn zero surplus on their own, only that surplus be "low enough." If ω_1 is small, for example, then $u^e(\omega_1)$ is close to zero: in order to obtain a positive surplus, a firm consisting of agents of wealth close to zero must have a large number of silent partners, which decreases the average surplus. Second, the upper bound n^* on firm size could be made considerably tighter. The principal obstacle to obtaining a tight bound is controlling for the number of silent partners that may join a firm. For the example in Section 4.1 it can be can be checked that $n^* = 8/0.6$, ω^* is any number in (0, 2), and that condition (12) is equivalent to $h^0 > 37/40$.[21] On the other hand, since in that case there are no silent partners, it is clear that no firm can have more than three members, so condition (12) could be amended to $h^0 > 2/3$; this is still not as tight as possible, since we showed that it is only necessary that $h^0 > 1/2$ in order that inefficient M-firms exist.

The restrictions on the distribution for which technically inefficient firms arise in equilibrium are not generally reducible to conditions on standard measures of income inequality, which depend too much on what is happening in the middle of the distribution. Rather, the right sort of conditions to look for are related to the relative size of the upper and lower tails of the distribution. One way to capture this is to consider the number of people below appropriately chosen "poverty lines" and "affluence lines," defined not in the usual sense by consumption baskets, but instead in terms of the capital requirements for production. We develop this idea further in [25].

[21] One can check that n_M is 3 for $0 \leqslant u^0 \leqslant 2$, 2 for $2 < u^0 \leqslant 6.5$; n_I is 2 for $0 \leqslant u^0 \leqslant 6.5$ The right-hand side of (11) is then equal to 0.6 for $0 \leqslant u^0 \leqslant 2$ (hence $u^* =$ 0.6) and -0.8 for $2 < u^0 \leqslant 6.5$.

5. DISCUSSION

Our approach to modeling firm formation enables us to endogenize participation constraints and provides a natural setting for studying the impact of interacting market imperfections on the organization of the firm. It also suggests that the nature and role of organizations is best understood when attention is paid to general equilibrium effects. We devote the remainder of this section to a discussion of robustness and extensions.

Financial market. While we use a costly state verification model, what matters most is that the cost of borrowing is decreasing in total wealth. Our results could therefore be obtained using other models of imperfect capital markets, such as those in Bernanke and Gertler [3], Kehoe and Levine [19], and Hart and Moore [13].

Furthermore, we suppose that the interest rate r is exogenous. Closing the economy and letting r depend on the supply and demand would require some modification of our definition of equilibrium (for instance to accommodate a capital market clearing condition). However, the substance of Proposition 3 does not depend on r, except that the relevant comparison economies are ω-economies in which the prevailing interest rate is the one that clears the market in the original economy. We note that the equilibrium values of r are bounded above by $(\pi_{\bar{N}} R - \bar{N})/K_I$ (or else no one would have any reason to demand capital) and below by unity (since the consumption good is storable). One can therefore always find parameter values such that Assumptions 7–10 and condition (12) are satisfied, so Proposition 3 will still be valid. Our argument for the Pareto optimality of equilibrium would have to be modified, however, because financial market clearing would now involve an infinite coalition of agents.

Size of firms. When γ is zero, size—as measured by the number of working members—maximizes the average surplus in the firm. But when γ is positive, there is a tendency for firms to employ above or below this level. For instance, if M-firms form and do not have to borrow, then from the point of view of maximizing economy-wide surplus, it would be best for them to operate at a scale which maximizes per capita surplus. But in the presence of financial market imperfections, oversized firms need not break up into smaller ones because some agents may not have cheap enough access to capital.

In particular, when γ is positive, a rich agent who belongs to an M-firm wants to maximize his residual surplus; thus there is a tendency for the firm to be larger than the average-surplus-maximizing size N_M (refer to the discussion following Assumption 10 in Section 4.2.2 for definitions). As long as the firm does not borrow, its size is bounded above by the surplus-maximizing size $\bar{N}$. However, if it does need to borrow, it may have more than $\bar{N}$ workers, because this increases the probability of success, which in turn lowers borrowing costs.

As for I-firms, there is the possibility that they underemploy relative to the average-surplus-maximizing scale $N_I = \arg\max((\pi_n R - K_I r)/n)$. The reason is that the total incentive compensation (proportional to $n\pi_n/\Delta\pi_n$) may be increasing faster below N_I than is total surplus; a greater residual may therefore be available to a rich agent who belongs to an undersized I-firm. Further discussion of these issues is available in [23]. See also [3].

Lotteries. As is well known, when there are indivisibilities or nonconvexities, individual agents (or small groups of them) may have incentives to engage in lotteries over initial wealth. Clearly, our model can accommodate lotteries. Since they would occur *before* the matching stage, allowing them would simply cause wealth to be redistributed before the match occurs. In that case all that we say in the paper applies without qualification to the *ex post* distributions or to initial distributions which are "lottery-proof."[22] A natural question is whether the qualitative message of our paper is still valid when lotteries are allowed. In particular, is it still true that firms can be technically inefficient when the costly state verification problem is sufficiently severe?

A sufficient condition for the result to go through is the existence of a lottery proof distribution satisfying the conditions of Proposition 4. We now argue that all of the distributions in the example in Section 4.1 are lottery proof. To see this, observe that agents with zero wealth cannot engage in any nontrivial fair lotteries, since negative payoffs are infeasible. It is not hard to establish that the marginal value of wealth is always nonnegative (surplus is nondecreasing in wealth—see our working paper [23]). Moreover, for agents with wealth at least 2, the marginal value of wealth is exactly zero, since for all distributions in the example, they receive the maximum surplus in equilibrium. The only way an agent in that range would have an incentive to participate in a fair lottery is if there were some wealth level ω in the interval $[0, 2]$ for which $U_H(\omega) > U_H(2)$; but this contradicts the fact that $U_H(\cdot)$ is nondecreasing.

Risk aversion. An important source of wealth effects that we have not considered here is risk aversion. As a referee has pointed out, Proposition 3 need not hold strictly if agents are risk averse. With risk aversion, it may be somewhat easier to provide incentives to poor agents than to rich

[22] Let $U_H(\omega)$ be the equilibrium surplus achieved by an agent with wealth ω when the distribution is H (this is defined for all $\omega \in \mathbb{R}$, and in particular is not restricted to the support of H). Nonconcavities in this function generate incentives to participate in lotteries. Thus H is lottery proof if $U_H(\cdot)$ is equal to its concavification at all wealth levels in the support of H.

ones; thus even if there is no costly state verification problem, rich agents might prefer to form *I*-firms with poor agents who could not form *I*-firms by themselves. What is unclear is whether this effect would be strong enough to swamp the bad incentives the poor have as a result of their proximity to the lower bound on utility, which we view as a useful characterization of poverty [2]. Risk aversion introduces many other complications that go beyond the scope of this paper (see e.g. [29]). In any case, risk neutrality enables us to illustrate in a dramatic way the role of the financial market for the organization of production.

APPENDIX: EXISTENCE OF AN EQUILIBRIUM

Here we sketch the proof of existence of an equilibrium for our economy. Let $\mathscr{F}$ 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 and Wooders [18] and therefore the f-core of $(A, \mathscr{F}, \hat{U})$ is nonempty. Existence of an equilibrium for our economy will follow.

Let $F \in \mathscr{F}$, $C(F)$ be the corresponding set of feasible contracts and

$$U(F) \equiv \{u \in \mathbb{R}^{\#F} \mid \exists c \in C(F) : u = (u(a \mid c))_{a \in F}\}$$

be the corresponding set of surplus vectors. From the definition of $C(F)$, it is clear that $U(F)$ is closed in $\mathbb{R}^{\#F}$. Let $\hat{U}(F) = U(F) - \mathbb{R}^{\#F}_{+}$ be the comprehensive extension of $U(F)$. $\hat{U}(F)$ is 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, $i = 1, ..., L$. $\{A^i\}$ 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 $F \in \mathscr{F}$.

(ii) $\hat{U}(F) \times \hat{U}(G) \subseteq \hat{U}(F \cup G)$, for all F, $G \in \mathscr{F}$, $F \cap G = \varnothing$.

(iii) $\inf_{a \in A} \sup \hat{U}(\{a\}) > -\infty$.

(iv) $\forall F \in \mathscr{F}$, $u \in \hat{U}(F)$, $u' \in \mathbb{R}^{\#F}$, $u' \leqslant u \Rightarrow u' \in \hat{U}(F)$.

(v) $\forall F \in \mathscr{F}$, $\hat{U}(F) \backslash \bigcup_{a \in F} [\mathrm{int}[\hat{U}(\{a\}) \times \mathbb{R}^{\#F-1}]\}$ is non-empty 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 $\hat{U}(F) \backslash \bigcup_{a \in F} \{\mathrm{int}[\hat{U}(\{a\}) \times \mathbb{R}^{\#F-1}]\}$ is non-empty. 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 $\pi_{\bar{N}} R - K_I r - \bar{N} + \omega_{F \backslash \{a\}} r$. Consequently, each $u \in \hat{U}(F)$ is bounded above by the vector $\bar{u}^F = (\pi_{\bar{N}} R - K_I r - \bar{N} + \omega_F \backslash \{a\}\, r)_{a \in F}$. It follows that $\hat{U}(F) \backslash \bigcup_{a \in F} \{\mathrm{int}[\hat{U}(\{a\}) \times \mathbb{R}^{\#F-1}]\}$ is bounded.

$(A, \mathscr{F}, \hat{U})$ is a game without side payments. Let Π be the set of measure-consistent partitions of A and for $\mathscr{P} \in \Pi$, let $F(a) \in \mathscr{P}$ be the firm 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 $F \in \mathscr{F}$, v^F is the restriction of v to F. Define the following sets:

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

$$H \equiv \bigcup_{\mathscr{P}} H(\mathscr{P})$$

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

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

Theorem 1 of Kaneko and Wooders [8] establishes the nonemptiness of the f-core of $(A, \mathscr{F}, \hat{U})$ when a simple condition, called *per capita boundedness*, is satisfied. We shall need the following definitions.

DEFINITION 5. $(A, \mathscr{F}, \hat{U})$ has the *r-property* with respect to $\{A^i\}_{i=1}^L$ if for any $F \in \mathscr{F}$, for any $i = 1, \ldots, L$, and for any $a, b \in A^i$,

(i) if $a \notin F$, $b \notin F$, then $\hat{u} \in \hat{U}(F \cup \{a\}) \Leftrightarrow \exists \hat{u}' \in \hat{U}(F \cup \{b\})$ s.t. $\hat{u}(d) = \hat{u}'(d)$, for all $d \in F$ and $\hat{u}(a) = \hat{u}'(b)$;

(ii) if $a, b \in F$, $\hat{u} \in \hat{U}(F)$ then $\hat{u}' \in \hat{U}(F)$ where $\hat{u}(d) = \hat{u}'(d)$, for all $d \in F \backslash \{a, b\}$ and $\hat{u}(a) = \hat{u}'(b)$ 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 $(A, \mathscr{F}, \hat{U})$ has the r-property.

DEFINITION 6. For any $F \in \mathscr{F}$, 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 $i = 1, \ldots, L$.

DEFINITION 7. $\hat{U}$ is *per capita bounded* with respect to $\{A^i\}_{i=1}^L$ if there is a $\delta \in (0, 1)$ and a $Q \in \mathbb{R}$ such that if $F \in \mathscr{F}$ is a coalition satisfying $(1+\delta)\, h^i \geqslant (\#(F \cap A^i)/\#F) \geqslant (1-\delta) h^i$ and if $\hat{u} \in \hat{U}(F)$ has the equal treatment property, then $\hat{u}(a) < Q$ 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 $u \leqslant \bar{u}^F$. 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 $\delta = \frac{1}{2}$. For $F \in \mathscr{P}$, let $n^i = \#(A^i \cap F)$. We consider finite coalitions such that $\frac{3}{2}h^i \geqslant (n^i / \sum_{i=1}^{l} n^j) \geqslant h^i/2$, 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 $\hat{u} \in \hat{U}(F)$ 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. The maximum total payoff of the set of agents in $A^i \cap F$ is $\pi_{\bar{N}} R - K_I r - \bar{N} + \sum_{j \neq i} n^j \omega_j r$. From our assumption on the proportions of types in F, one can show that $n^j/n^i \leqslant 3(h^j/h^i)$. Therefore, if the payoff vector satisfies equal treatment, the maximum payoff to $a \in A^i \cap F$ is

$$Q^i = \frac{\pi_{\bar{N}} R - K_I r - \bar{N}}{n^i} + \sum_{j \neq i} \frac{3h^j}{h^i} \omega_j r.$$

Choosing $Q = \max\{Q^i\}$ then ensures that $\hat{U}$ satisfies per capita boundedness.

Theorem 1 of Kaneko and Wooders [18] now tells us that there exists $\hat{v} \in \mathscr{C}_f(\hat{U})$. Thus there is $\mathscr{P} \in \Pi$ such that for almost all $a \in A$, $\hat{v}^{F(a)} \in \hat{U}(F(a))$. By definition of the comprehensive extension of $U(F(a))$, for almost all a, there exists $v^{F(a)} \in U(F(a))$ such that $v^{F(a)} \geqslant \hat{v}^{F(a)}$. Clearly, since $\hat{v}$ cannot be improved upon, neither can v. Consequently, $v \in \mathscr{C}_f(U)$. Using the feasibility condition, for almost all a, there exists a contract $c^{F(a)}$ such that $v^{F(a)}(b) = u(b \mid c^{F(a)})$ for all $b \in F(a)$. This establishes the existence of an equilibrium for our economy.

REFERENCES

1. A. Alchian and H. Demsetz, Production, information costs and economic organization, *Amer. Econ. Rev.* **62** (1972), 777–795.
2. A. V. Banerjee and A. F. Newman, Poverty, incentives, and development, *Amer. Econ. Rev.* **84**, No. 2 (1994), 211–215.
3. A. V. Banerjee, A. F. Newman, and Y. Qian, Capital markets, inequality, and the size distribution of business firms, mimeo, Stanford University, 1990.
4. B. Bernanke and M. Gertler, Agency costs, net worth, and business fluctuations, *Amer. Econ. Rev.* **79** (1989), 14–31.
5. B. Bernanke and M. Gertler, Financial fragility and economic performance, *Quart. J. Econ.* **105** (1990), 87–114.
6. K. Border and J. Sobel, Samurai accountant: a theory of auditing and plunder, *Rev. Econ. Stud.* **54** (1987), 525–540.

7. J. H. Boyd and E. C. Prescott, Financial intermediary-coalitions, *J. Econ. Theory* **38** (1986), 211–232.
8. J. H. Boyd, E. C. Prescott and B. Smith, Organizations in economic analysis, *Can. J. Econ.* **21** (1988), 477–491.
9. D. Diamond, Financial intermediation and delegated monitoring, *Rev. Econ. Stud.* **51** (1984), 393–414.
10. P. Hammond, M. Kaneko, and M. H. Wooders, Continuum economies with finite coalitions: Core, equilibria, and widespread externalities, *J. Econ. Theory* **49** (1989), 113–134.
11. O. Hart, An economist's perspective on the theory of the firm, *Columbia Law Rev.* **89** (1989), 1757–1774.
12. O. Hart and B. Holmström, The theory of contracts, *in* "Advances in Economic Theory," (T. F. Bewley, Ed.), Cambridge Univ. Press, Cambridge, 1987.
13. O. Hart and J. Moore, A theory of debt based on the inalienability of human capital, mimeo, MIT and LSE, 1991.
14. B. Holmström, Moral hazard in teams, *Bell J. Econ.* **13** (1982), 324–340.
15. B. Holmström and J. Tirole, The theory of the firm, *in* "Handbook of Industrial Organization" (R. Schmalensee and R. Willig, Eds.), North-Holland, New York, 1989.
16. T. Ichiishi, Coalition structure in a labor-managed market economy, *Econometrica* **45** (1977), 341–360.
17. M. Kaneko 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.
18. M. Kaneko and M. H. Wooders, The core of a game with a continuum of players and finite coalitions: The model and some results, *Math. Social Sci.* **12** (1986), 105–137.
19. T. J. Kehoe and D. K. Levine, Debt constrained asset markets, *Rev. Econ. Stud.*, forthcoming.
20. P. Legros, "Efficiency and Stability in Partnerships," Ph.D. dissertation, California Institute of Technology, 1989.
21. P. Legros and H. Matsushima, Efficiency in partnerships, *J. Econ. Theory* **55** (1991), 296–322.
22. P. Legros and S. Matthews, Efficient and nearly efficient partnerships, *Rev. Econ. Stud.* **68** (1993), 599–611.
23. P. Legros and A. F. Newman, Wealth effects, distribution and the theory of organization, mimeo, Cornell and Columbia Universities, August 1992, revised April 1994.
24. P. Legros and A. F. Newman, Complementarities, minimum outside option and segregation in production, mimeo, November 1994, revised June 1995.
25. P. Legros and A. F. Newman, Matching, restricted markets and rent-seeking, preliminary report, November 1994, revised June 1995.
26. P. Legros and A. F. Newman, Falsifiability and collusion in organizations, preliminary report, December 1994.
27. P. Milgrom and J. Roberts, "Economics, Organization and Management," Prentice–Hall, Englewood Cliffs, NJ, 1992.
28. A. F. Newman, The capital market, inequality, and the employment relation, mimeo, Columbia University, 1992.
29. A. F. Newman, Risk-bearing and "Knightian Entrepreneurship," mimeo, Columbia University, 1995.
30. R. Radner, Hierarchy: The economics of managing, *J. Econ. Lit.* **30** (1992), 1382–1415.
31. S. Scotchmer, Public goods and the invisible hand, mimeo, 1993; *in* "Modern Public Finance" (J. Quigley and E. Smolensky, Eds.), forthcoming.
32. J. Tirole, Collusion and the theory of organization, *in* "Advances in Economic Theory" (J. J. Laffont, Ed.), Vol. 2, Sixth World Congress, Cambridge Univ. Press, Cambridge, 1992.

33. R. M. Townsend, Optimal contracts and competitive markets with costly state verification, *J. Econ. Theory* **21** (1979), 265–293.
34. S. Williams and R. Radner, "Efficiency in Partnership when the Joint Output is Uncertain," CMSEMS DP 760, Northwestern, 1988; *in* "The Economics of Informational Decentralization" (J. Ledyard, Ed.), forthcoming, Kluwer Academic, Dordrecht/Norwell, MA.
35. M. H. Wooders, Large games and economies with effective small groups, Bonn Discussion Paper B-215, 1992.