Why capital suppliers (usually) hire workers: what we know and what we need to know

Gregory K. Dow\textsuperscript{a,}*; Louis Putterman\textsuperscript{b}

\textsuperscript{a} Department of Economics, Simon Fraser University, 8888 University Drive, Burnaby, BC, Canada V5A 1S6
\textsuperscript{b} Department of Economics, Brown University, Providence, RI 02912, USA

Received 21 May 1998; received in revised form 2 August 1999; accepted 31 August 1999

Abstract

A fundamental question for economics is why large firms in market economies usually assign control rights to capital suppliers rather than labor suppliers. A diverse collection of answers can be found in the literature. But unfortunately little theoretical consensus has emerged, and few attempts have been made to resolve this issue through systematic empirical investigation. This paper reviews a number of different hypotheses clustering around work incentives, the financing of firms, and collective choice. We identify the strengths and weaknesses of each approach, filling in theoretical gaps where necessary, and conclude with some suggestions for empirical research. © 2000 Elsevier Science B.V. All rights reserved.

JEL classification: D23; J54; L22; P13

Keywords: Firm ownership; Firm organization; Labor-managed firm; Control rights

1. Introduction

In market economies, the way in which production is organized is dictated neither by law nor by constitution. Each individual or household can be an independent producer, groups of workers can collectively organize production and purchase required inputs, or providers of inputs including capital can organize firms and hire workers. The actual outcomes are determined endogenously by the characteristics of technologies, by initial endowments, by access to information, and by preferences. A question of fundamental theoretical and policy importance is why when more than a few workers engage in a joint enterprise, they are typically hired by providers of financial resources (or by agents of the latter), rather than organizing the enterprise themselves and contracting with outsiders for the supply of...
capital. We can scarcely claim to understand what drives economic organization without an explanation for this phenomenon. Nor can we assess the potential costs and benefits of measures such as legislating participation by workers on company boards of directors (as in Germany), giving tax subsidies and other incentives to firms with profit-sharing and/or employee ownership plans (as in the US), or providing local government assistance to employee groups wishing to buy out ailing companies (as sometimes occurs in the US as well).

Although recent decades have seen a number of economists take up the challenge of explaining why capital providers typically hire workers, a student of economics who is interested in this question will be obliged to sort through an unwieldy array of alternative and sometimes conflicting answers. The purpose of this paper is to provide an overview of the major answers given, to fill in some gaps where necessary, and to synthesize sets of explanations that are closely related and complementary. These seem to us to be vital steps on the road to empirical research that might reveal which answers best account for observed patterns of firm organization.

We group the candidate explanations into three sets, focusing on work incentives; investment and financing problems; and issues of collective choice. In Section 2 we deal with conceptual preliminaries involving control rights, asset ownership, and residual claims. Sections 3–5 take up the work incentive, financing, and collective choice approaches in turn. Of course, the explanations are not entirely separable, and several may apply simultaneously or with different weights in different situations. We discuss this and other points by way of conclusion in Section 6.

Although some other works contain reviews related to this one, they focus on different questions or approaches. Hansmann (1996) covers similar territory but his work is grounded more in the law and economics literature, treats some of the arguments from economic theory more casually, and argues for the collective choice approach which is only one of those considered by us. Bonin et al. (1993) focus mainly on empirical propositions derived from the literature on labor-managed firms, a subset of the broader literature discussed here. We believe our paper is uniquely synthetic, and while it is informal in style, it is also carefully tied to relevant economic theory. We attempt to provide a coherent overview of what is known, organize extant ideas, and offer a convenient jumping-off point for further reading and future research. Of course, it is not possible to be fully comprehensive in the limited space available here, and we apologize in advance to authors whose work is not discussed explicitly.¹

2. Preliminaries

We ask “who controls firms?” rather than “who owns them?” or “who holds their residual claims?” Why make these distinctions and why focus on control? There is far too much packed into the ordinary concept of “firm ownership” for it to serve as a useful primitive in our analysis. As usually defined, “ownership of the firm” represents a bundle of rights including residual claims, control rights, and ownership of physical assets, along with the right to

¹ A more detailed treatment of our topic and a more complete set of references is provided in Dow and Putterman (1996). There we review briefly some further hypotheses, not mentioned here, which we believe warrant less attention.
transfer this bundle of rights to someone else. While this bundling at the conceptual level does reflect a genuine set of empirical regularities, it is also frequently true that important elements of the bundle are separated. For instance, in employee stock ownership plans (ESOPs) in the US, workers often hold large residual claims but without any meaningful control rights because they do not vote their own shares. For our purposes it is best to regard the bundling of these various rights as empirically contingent; that is, something that may or may not be observed in specific cases, and something that needs to be explained rather than assumed in advance.

We thus find it useful to focus on a key constituent element of the ‘ownership’ bundle, namely control rights, and to examine the conditions under which control might be bundled with other attributes such as residual claims or the ownership of physical assets. As we shall see below, one important class of theories emphasizes that in order to control firms, workers must in practice ‘own’ them by providing equity financing and holding associated residual claims. But other theories suggest that workers would not control their own firms even if equity financing were unnecessary (for instance, if workers could rent all required capital goods). These issues should not be swept under the rug by adopting the bundling of ownership attributes as a conceptual premise.

Moreover, firms — the organizations that carry out production and commercial functions in a market economy on scales exceeding that of an individual household — arise when several agents contribute to a joint production activity. As is widely recognized in the literature of the new institutional economics (Williamson, 1985; Hart, 1995), it is generally impossible to specify in advance all terms of the interactions among firm participants. If firms are production coalitions marked by incompleteness of contracts, it is crucial to determine who is entitled to fill in the details of these ongoing interactions. This motivates our definition of a conventional or capitalist firm as one where control rights are held by suppliers of capital or equity finance, and of labor-managed firms (LMFs) as ones where control rights are held by suppliers of labor.

Control could also be exercised by a combination of both types of agents, or by agents who are neither capital nor labor suppliers. Since our concern in this paper is with investor versus worker-controlled firms, we do not deal with cases where control rights are held by other input suppliers (e.g. agricultural cooperatives), by customers (consumer cooperatives), or by others (for instance, non-profit organizations).2 We treat mainly the polar cases of pure capital control or pure labor control, in the hope that hybrid cases such as partnerships which use both partner and hired labor, or firms where both worker and financier representatives sit on boards of directors, may be dealt with by interpolation between those poles. Due to our focus on control, firms with employee stock ownership plans or profit-sharing, but no worker decision-making rights, are not of central interest, although they do offer relevant evidence on some issues that concern us.

### 3. Monitoring and work incentives

In this section, we consider explanations of the assignment of control rights that focus on incentives for work effort and the monitoring of this effort. Of the three approaches

---

2 Hansmann (1996) includes a discussion of firms where control is held by neither capital nor labor suppliers.
considered here, two begin with the work incentive problem as such, while one comes to that problem from an initial concern with risk aversion. All assume that workers’ contributions to production are either costly to observe or else not contractible (that is, incapable of being made subject to legally enforceable contracts) for some other reason, such as the cost of writing detailed contracts or using the court system. In their most basic forms none of these theories provides an explanation of why capitalists (rather than customers, or providers of purely entrepreneurial inputs) hire workers. Instead, they merely suggest that control rights should not be held by workers themselves. One can nonetheless complete each of the three stories in such a way that the controlling party is identified as a capital supplier.

Both the monitoring story of Alchian and Demsetz (1972) and the budget-breaker story of Holmstrom (1982) focus on the problem of work incentives with joint or ‘team’ production. Output is assumed to result from the efforts of a number of individuals, and while output is easily observed, the contribution of any particular team member cannot be ascertained merely by observing output. If individual team members are compensated from their joint product, output is like a public good and a classical free rider problem obtains. The solution suggested by Alchian and Demsetz is that someone should undertake costly monitoring of inputs, and compensation should be tied to the resulting observations, providing appropriate motivation to each worker. But since monitoring, which allows workers to achieve higher utility levels, is itself a public good to the team members, a free rider problem also afflicts workers’ incentives to monitor. That problem is resolved, say Alchian and Demsetz, if one party is motivated to monitor each team member by being assigned the right to any residual remaining after payment of team members according to contributed effort. In effect, the returns to monitoring are converted from a public to a private good, breaking a potentially unending chain of “who monitors the monitor?” problems.

Holmstrom’s approach is radically different in that it obviates the need for monitoring. However, it shares with Alchian and Demsetz the idea that the team effort problem can be solved by introducing an asymmetry into (or adding an outside agent to) the team. Holmstrom shows that all team members can be motivated to provide optimal effort levels despite unobservability of individual effort, if the compensation of each is made to depend upon total team output in a sharply discontinuous manner. Specifically, each team member receives compensation sufficient to elicit her own optimal effort level if group output is at or above the efficient level. But to deter free riding, any shortfall in group output results in a large decline in compensation to ‘shirker’ and ‘non-shirker’ alike. To create this discontinuity, the team must commit to sharing less than its actual joint output. Holmstrom argues that such a commitment could be made by contracting with an outside party to withhold a portion of team product whenever output is too low, thus breaking the equality between output and the sum of payments to team members.

The third monitoring and incentive story is an adapted and pared-down version of one told almost three-quarters of a century ago by Knight (1964 [1927]). Suppose revenue is a stochastic function of the inputs of team members. This could result from uncertainties over mechanical break-downs and power failures, or from variability in market prices. Compensation out of joint revenue then exposes workers to risk. Risk averse workers would prefer to trade this uncertain earnings stream for a certain one with the same mean, but they can be guaranteed fixed payments only if they cede control over the production process to the insurer, since the payment guarantee eliminates incentives to provide value-enhancing
effort. The insurer must make payments contingent on input provision, and accordingly must monitor effort.

Knight’s approach anticipates modern principal-agent theory (Sappington, 1991) by highlighting the supply of insurance to risk averse agents (workers) by a risk-neutral principal (entrepreneur). Modern theory diverges from Knight by emphasizing the irreducible nature of worker control over effort, and the resulting inevitability of a trade-off between insurance and effort incentives. Unlike Knight, contemporary principal-agent theorists predict that workers will continue to bear some risk (and thus function as partial residual claimants) despite their assumed risk aversion. It is generally possible for the principal to reduce the size of the residual claims held by workers, and thus the costs associated with risk, by monitoring labor inputs. But monitoring is costly and imperfect as well, so workers remain only partially insured and retain some residual claim.

If worker incomes must depend on output or profit for incentive reasons, it is less clear that they will cede control rights to entrepreneurs. Indeed, profit sharing often leads to understandable demands for worker participation in decision-making. Holmstrom and Milgrom (1991, 1994) present models where an outside principal provides workers with partial insurance, and where for incentive reasons workers may or may not own physical assets, depending on exogenous parameters. Presumably workers are less likely to have decision-making powers when they have smaller residual claims and do not own assets, but this theoretical link has not yet been worked out explicitly.

Each of these three theories provides a reason why someone not belonging to the production team might appropriate its revenue stream and make payments to workers, but nothing so far requires that the person in question be a provider of capital or finance, or exercise other forms of control in the workplace such as hiring and firing. However, the authors of two of the theories do attempt to supply such links to real-world firm structure. Alchian and Demsetz argue that firms will own some of their own equipment to avoid the agency costs arising when assets are leased from outsiders (such as equipment abuse, and the costs of monitoring and deterring it). It is efficient that the central monitor of labor also be the equipment owner, they suggest, since his residual and equity claims provide appropriate incentives for the complementary monitoring of work effort and equipment use. Finally, they argue that directing workers’ activities and monitoring their effort provision are closely related, and that the right to determine team membership both enhances the value of the central monitor’s residual, and provides a useful disciplinary supplement to the adjustment of payments based on observed effort. Thus, in their theory residual earnings, determination of payments to workers, ownership of physical assets, control rights, and the right to hire and fire workers all fall to the same central agent.

The need for managerial control by the insuring party is implicit in Knight’s concern with the moral hazard of insuring labor suppliers. The idea that insurance supply should be bundled with the supply of capital could again be justified as a way of mitigating asset abuse, but Knight’s own discussion suggests that the roles of insurer and financier can be connected by extending moral hazard arguments to the sphere of equity supply, as will be done in the next section. Holmstrom, by contrast, does not explicitly connect ‘budget-breaking’ with capital supply. This linkage would be problematic if the inputs supplied by capitalists were to some degree non-contractible (for example, asset maintenance) as this would turn capitalists
into members of the production team, contrary to the requirement that the budget-breaker be a pure outsider.

The arguments of Alchian and Demsetz, Holmstrom, and Knight are each subject to conceptual criticisms. Alchian and Demsetz are not explicit about what reputational mechanisms make compensating the worker based on her observed effort more profitable to the residual claimant than falsely asserting that the worker shirked and withholding the requisite payment. Another problem is that responsiveness of worker effort to monitoring implies that the level of monitoring is known to the worker. But if that is so, it is unclear why monitoring services cannot simply be purchased by workers in specified amounts at an agreed-upon price. Third, even if residual claims are the most effective way to elicit monitoring, and spreading those claims across team members would reduce monitoring incentives, production technology may cause monitoring by a central party to be so much more costly than mutual monitoring by co-workers (which in many instances is a ‘free’ by-product of participation in teamwork) that mutual monitoring is the preferred solution, despite its motivational drawbacks. Finally, and relatedly, in a repeated game setting standard Folk Theorem results show that efficient effort levels can often be sustained under conditions of team production through threats of future retaliation against shirkers, without any need for an external monitor (Weitzman and Kruse, 1990).

Knight’s insurer-employer is subject to similar reputational problems as Alchian and Demsetz’s monitor. If the insurer offered a contractually fixed wage buttressed by supervision to assure compliance with a minimal effort standard, the problem would be to prevent unjustified dismissal, rather than to prevent the monitor from withholding wages as in the Alchian and Demsetz framework. As noted earlier, it is also unclear whether Knight’s theory can be reconciled with modern principal-agent ideas, which imply that workers will usually retain some residual claims for motivational reasons. Holmstrom’s budget-breaker story requires high levels of trust in the rationality of fellow workers, and could be susceptible to cheating by the budget-breaker and a colluding worker (Esvaran and Kotwal, 1984; Andolfatto and Nosal, 1997). In any case, the actions of the budget-breaker can be mimicked by separate divisions in a worker-run firm (MacLeod, 1987).

There also appears to be significant prima facie evidence against these stories. Mutual monitoring, reductions in supervisory expenses, and strong work incentives are widely

---

3 Without a worker response to monitoring, it is unclear how the monitor can profit from the activity. But such a response seems questionable if the worker cannot know how much monitoring is being done.

4 This argument about technology and mutual monitoring was earlier made by Putterman (1984). It is consistent with Alchian and Demsetz’s hypothesis that profit-sharing is more likely if workers are highly costly to monitor. But their argument involves the substitution of direct effort incentives (‘self-monitoring’) for monitoring by an outside party in such cases, rather than the use of mutual monitoring. Mutual monitoring may trump centralized monitoring even for tasks that are only moderately difficult to monitor.

5 As discussed in the efficiency wage literature, the firm could be contractually committed to pay a worker this period even if it finds cause to dismiss her (MacLeod, 1987). But importing this solution into the Knightian firm brings with it the rest of the efficiency wage problem — if workers are not to just move from firm to firm, collecting their wages and shirking, there must be an incentive to retain the job, a ‘job rent,’ which implies a non-clearing labor market (Dow, 2000).

6 Even if rationality is common knowledge, workers would be unwilling to make their pay too sharply contingent on the performance of untrusted co-workers, given the likelihood of multiple equilibria under a Holmstrom-style contract and the potentially severe losses that could result from coordination failures.
accepted as stylized characteristics of worker-owned firms. In particular, shirking does not appear to be a major problem (Bonin et al., 1993). Even when workers lack formal control rights, as under numerous profit-sharing plans, there is ample evidence that profit sharing tends to raise productivity in many firms (for the US, see Kruse (1993); for the UK, Wadhwani and Wall (1990); for Japan, Jones and Kato (1995)). As to budget-breaking, such payment schemes are rarely observed, perhaps reflecting limits on worker trust, violation of norms of fairness, or the impracticality of tying wages to output through legally binding contracts (who adjudicates output quality?). Evidence on the risk attitudes emphasized by Knight is difficult to obtain. In practice it may also be hard to distinguish between risk aversion effects that operate through work incentives and those that operate through ownership, discussed in the next section.

In our view, the hypothesis that control rights are held by capital suppliers due to work incentive problems is logically consistent but dubious. At the theoretical level it is necessary to tack on a series of supplemental arguments to explain why a central monitor or budget-breaker must also be a capital supplier, and apart from this there are conceptual problems with both the monitoring and budget-breaking stories. Empirically, there is a good deal of evidence that worker-managed firms are not at a disadvantage with respect to work incentives, and indeed that profit sharing frequently increases labor productivity within capitalist firms. The idea that workers’ interest in a relatively fixed wage leads to control and monitoring by capital suppliers seems less conceptually problematic, but also requires an auxiliary explanation of why the insurer must be a capital provider. Despite these reservations, we believe that coherent theories of the three types discussed in this section can be constructed, and we judge them worthy of serious empirical testing.

4. Investment, wealth, and diversification

A second group of theories focuses on the interactions between firm financing requirements and worker liquidity constraints or risk aversion. These theories come in many flavors, and most presentations have been incomplete in one respect or another. We believe the core of the argument has the following structure.

The main idea is that to control a firm, workers must supply financial resources as well as labor. This conclusion is reached in two steps. First, it is argued that efficiency requires most firms to own, rather than rent, at least some of the capital goods or assets they need. Second, it is argued that full outside financing of these owned assets would be incompatible with worker control. Therefore, if workers are to control firms, individual or collective worker financing of at least some assets is needed. But this is ruled out by limited worker wealth, limited access to credit markets, and/or the desire of risk averse workers to maintain diversified financial portfolios.

There are three major arguments as to why it is inefficient or impossible for firms to rent all of the capital goods they use. The first, already touched upon in the previous section, is that costs of asset abuse, and of the monitoring needed to detect and deter it, can be eliminated if user and owner are one and the same (Alchian and Demsetz, 1972). The second

---

7 A more detailed discussion is provided in our working paper (1996).
is that some assets are intangible and difficult to separate from the firm, such as reputation, customer goodwill, and technological capabilities (Jensen and Meckling, 1979). The third (often an outgrowth of the second) is that integrating ownership with use may be the best way to avoid 'hold-up' problems and associated bargaining costs in cases where assets are highly specific to the firm using them (Klein et al., 1978; Williamson, 1985; Hart, 1995). This last argument suggests that worker teams would not lease assets from outside owners when physical asset specificity is important, even if leasing is feasible in principle (Dow, 1993a). However, it might still be possible to have labor-managed firms where workers own such assets collectively, so auxiliary considerations (perhaps relating to financing) must be brought in to rule out this option. It should also be noted that problems arising from human capital specificity cannot be handled by means of integrated asset ownership unless it is workers who do the owning. Other things equal, firm-specific human capital thus argues in favor of labor-managed firms, because a capitalist firm creates unnecessary hold-up problems between owners and their employees (Grossman and Hart, 1986; Hart and Moore, 1990; Dow, 1993a).

Even if the firm must own some capital goods to avoid costs of monitoring asset usage or to deal with problems of asset specificity, in principle workers could control the firm without putting up any of their own wealth if they were able to procure full external financing. The argument that they cannot generally do this runs as follows. First, fully financing the firm using debt runs into familiar problems associated with high debt-to-equity ratios. Because the borrower but not the lender benefits from exceptionally good results on risky investments, while the lender can lose both his return and the principal of the loan in the event of exceptionally poor results, borrower and lender interests diverge, with borrowers being more prone to take risks than lenders. The severity of this conflict increases with the debt-equity ratio. This is the problem of moral hazard in the credit market, which can lead either to interest rates that rise with leverage or to outright credit rationing (Stiglitz and Weiss, 1981). Similar problems can arise from adverse selection, where lenders cannot distinguish ex ante between safe and risky borrowers. Again, such informational problems can lead to credit rationing where workers are unable to borrow at going (or even higher) interest rates despite having good projects, unless they can put up some equity or collateral to signal their true risk characteristics as borrowers.

Another option would be for the workers to hold all voting shares, while raising funds by selling non-voting equity shares to outside capital suppliers. But as in the case where moral hazard leads to interest rates on loans that rise with leverage, non-voting equity capital is likely to be made available only on unfavorable terms, if at all, unless workers can credibly signal that their interests converge with those of outside investors (to overcome moral hazard), and that their investment projects are of high quality (to overcome adverse selection). Because debt financing could be more fea-

---

8 Near-perfect convergence of interest between voting and non-voting shareholders may explain why non-voting shares can be issued by corporations, with non-voting shares possibly even fetching a premium if buyers believe they can free-ride on the governance efforts of voting shareholders (Lease et al., 1983). In contrast, Jensen and Meckling (1979) argue that equity suppliers could expect to be victimized by creative accounting and other contrivances were they to rely on controlling workers' pledges to share profits with them. Moral hazard-type considerations are emphasized in Jensen and Meckling (1976) and Hansmann (1988); adverse selection issues by Gui (1985), Schlicht and von Weizsacker (1977), and Stiglitz and Weiss (1981).
possible for easily redeployable assets, which can be pledged as collateral against loans, problems of worker financing may be more severe for highly firm-specific assets (Williamson, 1988).

A different way of stating these points is that because low debt-to-equity ratios mean that equity investors stand to lose along with debt-holders when firms invest in poorly performing projects, debt finance is more affordable for firms with substantial equity financing. And because financiers find investment in ventures over which they lack control rights to be excessively risky, that equity financing must come largely from those accorded control rights. If these arguments are right, then a worker-controlled firm will also have to be a worker-invested or worker-owned firm. But this means that worker-managers must come up with the means to finance some, and perhaps most, of their firms’ capital requirements, and that they must bear the risks associated with those investments. In the remainder of this section, we discuss three types of problems that may arise due to the need for workers to finance worker-controlled firms. These involve constrained liquidity, worker risk aversion and portfolio diversification, and the potential difficulties with mechanisms to implement collective ownership of capital goods.

If it is accepted that worker control requires worker financing, perhaps the most straightforward hypothesis regarding the rarity of worker-controlled firms is that workers usually lack sufficient wealth to finance their own firms. It is reasonable to suppose that in many cases (oil refineries, supertankers), firms use owned capital goods whose total value exceeds the aggregate wealth of their employees. Even if the ventures in question are promising enough that workers could in principle repay the personal debts required, banks may be unwilling to lend the necessary funds due to insufficient worker collateral or uncertainty about the durability of an individual worker’s connection to a given firm. It should be kept in mind that workers cannot generally pledge their own future labor income as collateral (Hart and Moore, 1994), creating a serious danger that loans to workers or to their firms may never be repaid.

However, pure liquidity constraints are unlikely to provide a full explanation of financing difficulties for a sizeable fraction of the workforce in the United States, where workers’ pensions and private insurance funds account for a significant fraction of stock market value. An obvious candidate to fill the resulting explanatory gap is worker risk aversion. If we look at investments made by worker pension funds, and where workers invest their own savings, we find that portfolio diversification plays a major role in most investment strategies. Through diversification the investor can minimize the impact of bad luck or bad judgment in the choice of any one investment. If controlling her firm would require a worker to forgo such diversification and invest most or all of her wealth in the firm’s assets, this may come at too high a cost for a risk averse worker. This could be true even if the labor market is competitive, so a worker can walk across the street and take another job at the same wage if need be. But in many jobs workers earn quasi-rents on firm-specific human capital, or enjoy rents for other reasons such as search costs, the use of efficiency wages, or simple bargaining power. In such cases a worker would not only be placing all of her financial eggs in one basket; she would also be investing her savings in an asset whose returns are highly correlated with the returns on her human capital. Such a worker only needs to be sufficiently risk averse, or to place a suffi-
ciently low value on holding control rights, to be dissuaded from joining a worker-owned firm. 9

A third set of reasons why the need for worker investment may prevent workers from seeking to control firms relates to difficulties in structuring collective ownership. Furubotn and Pejovich (1970) and others have shown that if the assets held collectively by a firm are not treated as the property of individual owners, there will be a disincentive for worker investment (including the retention of current earnings for collective capital formation) in worker-run firms. While workers in firms without individualized capital ownership can benefit from collective investment through returns to their labor during their period of employment, Furubotn and Pejovich argue that cessation of such returns at retirement reduces the effective rate of return on investments whose useful life exceeds a worker’s expected remaining tenure with the firm. In principle this could be corrected by guaranteeing the worker an appropriate return on investments made during her tenure and still paying off in her retirement, in the form of a retirement settlement or pension keyed to past investments in the firm. But this approach may fail due to a lack of commitment across generations of workers, or difficulties in imputing the economically relevant rates of return (problems surrounding depreciation and the calculation of appropriate risk premiums, for example, are notoriously hard to solve).

An appealing alternative might then be to allow workers to capitalize the value of membership in a worker-owned firm through a tradable membership right. Indeed, it has been shown (Sertel, 1982; Dow, 1986, 1993b, 1996; Fehr, 1993) that such tradable claims in theory resolve both the horizon problem discussed by Furubotn et al., and the problem of labor allocation across firms discussed in the literature on labor-managed firms (Ward, 1958; Domar, 1966; Vanek, 1970). In view of these theoretical results, we focus the rest of our discussion on the problems of operating a membership market. 10

Despite the theoretical appeal of markets for membership in worker-owned firms, they are rare—in fact the only well-documented case is that of the plywood cooperatives in the US northwest. This fact motivates another major finance-related explanation of the predominance of capital-controlled firms: if labor-managed firms perform poorly in the absence of well-functioning membership markets, then whatever explains the rarity of such markets (and the difficulties exhibited when they do arise) may help to explain the rarity of LMFs themselves. Candidate factors include two already discussed, liquidity constraints and risk aversion. It may be difficult to find buyers of membership rights if potentially interested workers lack adequate wealth or credit-worthiness, or if they are unwilling to accept the added risk from non-diversification.

9 Note, however, that firm-specific human capital was said earlier to favor worker control, from the standpoint of the ‘hold-up’ problem discussed by Williamson, by Klein et al., and by Hart. Thus, without some prior knowledge of the importance of worker risk aversion and of parameters like worker wealth, financing requirements, etc., it is difficult to say which type of control structure will in fact be favored by the presence of human capital specificity.

10 Another way to resolve horizon, but not allocative inefficiency, problems, is to record each workers’ investments in the firm and undertake to repay those investments with interest either upon demand, or upon the worker’s separation from the firm (including retirement). This is roughly the approach represented by the individual capital accounts of the well-known Mondragon cooperatives, where, however, a portion of capital investments are also held collectively. One problem with the accounts is that they make no attempt to track changes in the value of the firm’s investments, as would a well-functioning membership market.
A conceptually distinct explanation is that membership markets may suffer from low trading volumes relative to markets for corporate equity, simply because only those individuals who will actually work in a firm can purchase such memberships. Markets of this kind are likely to degenerate into infrequent bilateral bargaining between the firm and new applicants, without the continual monitoring and informational efficiencies of a stock market. Takeover mechanisms are unlikely to function smoothly because it would be necessary to organize a large competing team of workers in advance, since control can only be exercised by 51% of the firm’s labor suppliers. There is also an adverse selection problem: potential members may have difficulty appraising the prospects of the firm, and the incumbents who can provide useful information may have incentives to misrepresent those prospects. The problem of adverse selection goes deeper still, because a departing member might benefit financially by selling her position to an undesirable replacement, inflicting losses on those who stay behind. Such externalities interfere with anonymous trading of the sort found in equity markets. Any or all of these factors could render the membership market non-viable.

For the one known case of a membership market, Craig and Pencavel (1992) have estimated that selling prices of membership shares in plywood cooperatives typically fall short of discounted rent streams by upwards of 50%, suggesting discounting of shares due to some combination of risk aversion and adverse selection. Such systematic undervaluation may imply that the membership market does not fully correct the problem of deficient investment incentives in labor-managed firms originally identified by Furubotn and Pejovich.

Quickly summarizing the arguments of this section, the hypothetical possibility of worker control with pure rental of physical assets may be excluded by the requirement that enterprises own some assets to reduce the agency costs of potential asset abuse, and to avoid problems associated with asset specificity. Full debt financing may be ruled out by moral hazard or adverse selection in the credit market, and external equity financing through non-voting shares could prove unpromising for parallel reasons. The burden of worker finance will be highest in firms using large amounts of firm-specific assets per worker. This burden may deter worker ownership if workers lack the wealth to finance their firms, or if they are too risk averse to commit their wealth to the same venture as they are committing their labor. In addition to liquidity constraints and risk aversion, other problems may obstruct the LMF membership markets that can theoretically eliminate the difficulties surrounding collective ownership, investment incentives, and labor allocation across firms. Chief among these are thinness of membership markets, the impracticality of anonymous trading, and problems of adverse selection.\footnote{11}

\footnote{11 That so large and capital-intensive a company as United Airlines could come under the predominant ownership of a subclass of its employees may not be as much of an anomaly as it might otherwise appear in the light of these considerations. The company’s non-human assets are dominated by generic equipment and by rights to use routes and airport facilities that are readily tradable and difficult to degrade. Pilots are relatively highly paid, the suitableness of a candidate pilot-owner is relatively easy to assess, and the class of potential pilot-owners is reasonably large. Moreover, the very size of the company and the existence of equity traded among non-workers attracts intensive professional evaluation, so that a rough share value is widely agreed on and insiders have little relative advantage in assessing it. Gordon (1999) also argues, however, that employee control is sharply circumscribed by terms of the United buy-out that were drawn up to assure outside investors and creditors that the company would not deviate from conventional business goals.}
Unlike the arguments reviewed in the previous section, those discussed here have attracted little theoretical criticism. They are also consistent with the crude stylized facts that LMF membership markets are rare, that external equity financing of LMFs is almost non-existent, that worker cooperatives rely heavily on retained earnings for investment, that cooperatives are more common in industries with low to moderate capital-to-labor ratios, and that cooperatives appear to use less capital per worker than do conventional firms in the same industry. The task is to disentangle the contributions of risk aversion, liquidity constraints, and asymmetric information, and to investigate the scope of these factors as compared with other explanations for the dominance of conventional firms.

5. Collective choice and decision-making

The final group of explanations for the prevalence of capital-controlled firms is concerned with collective choice and decision-making. That worker-run firms may suffer from comparative disadvantages with respect to decision-making is an old and familiar idea. Perhaps the simplest version is the notion that people sort themselves into workers and entrepreneurs or managers on the basis of raw decision-making ability, and that a firm run by workers would be as competitively disadvantaged as an army run by common soldiers. Schemes that rotate managerial roles randomly amongst a workforce, or require day-to-day operations to be governed by time-consuming, ‘town meeting’ deliberations, are especially likely to suffer from such inefficiencies.

However, worker control imposes no requirement of ongoing participation or task rotation, the term ‘worker’ includes labor providers of every type (including managers), and a worker-run firm can be every bit as hierarchical as its capitalist counterpart, if that is the preference of the workers themselves (or what they find most prudent). The only required difference between a worker-controlled and capitalist firm is that in the former, managers or the board of directors are elected by the workforce, and in the latter, by the shareholders. With shareholders being diverse in their knowledge and decision-making abilities, and with the workforce of a firm including managerial and technical workers, differences in raw ability to comprehend one’s own interests or to choose good managers are unlikely to be significant.12

Most of the recent and compelling theories in which collective choice plays a central role have focused on heterogeneity of interests within the potential control groups of financiers and labor providers. In several works, Hansmann (1988, 1990a,b, 1996) argues that the prevalence of capital-controlled firms over labor-controlled firms is explained by problems of collective decision-making familiar from the public choice literature. The gist of the idea is that in the absence of unanimity (or, for single-issue voting, single-peaked preferences), majority rule may generate unstable outcomes or necessitate the imposition of various forms of agenda control. This is accompanied by an argument that capital suppliers are

---

12 To put it differently, it is unclear why a ‘common worker’ who is not comparatively advantaged in decision making can be expected to make good choices about agreeing to the authority of a manager or capitalist under a conventional employment relationship, but bad choices about delegating decision-making authority to more capable agents in an LMF.
likely to agree unanimously on their objectives (presumably, maximization of profit or present value), while workers have heterogeneous interests (concerning wages, job security, working conditions, time preference, risk attitudes, and so on) that leave them vulnerable to the full force of classic preference aggregation problems. In short, even if workers elected the board of directors, they would be unable to provide their representatives with any coherent guidance about the goals to be pursued by the firm.

In the strand of the literature associated with Hansmann, as well as Benham and Keefer (1991), the stress is on transaction costs of collective choice. Here the argument is not that workers are literally unable to resolve their conflicting interests, but rather that this is a costly process. For example, if decisions are reached through bargaining, there may be costs associated with incomplete information (Kennan and Wilson, 1993), or with the time consumed in multilateral bargaining among a large group of individual workers. On the other hand, if decisions are reached through voting, there is the usual problem of voting cycles: for any possible decision, there might be some other decision preferred by a majority of voters.

There are various ways to avoid the cycling problem but these measures have costs of their own. For instance, the firm could carefully screen incoming members to ensure uniformity of preferences; eschew an extensive division of labor, large skill differentials, or large wage differentials; or remain at a suboptimal scale in order to limit the size of the decision-making group. If such measures are too costly, the other option is to restrict the scope of enterprise democracy. This can be done by limiting opportunities for workers to put new proposals on the agenda; using representative rather than direct democracy; or disenfranchising segments of the workforce whose interests diverge from those of the controlling group (e.g. clerical workers or janitorial staff).

Zusman (1992) has attempted to specify the transaction costs of collective choice more concretely, suggesting that collective choice rules in firms are chosen to minimize the sum of bargaining costs and the members’ risk premia. Zusman handles the cycling problem by treating it as one of uncertainty: if the firm adopts a voting procedure where cycling is possible then every Pareto efficient choice is assigned equal subjective probability. Similarly, all possible bargaining coalitions are viewed as having equal probability of forming ex post, and if a constitutional rule creates a privileged position of authority, then all ex ante members of the firm are assumed to have the same subjective probability of becoming the boss.

Kremer (1997) has recently revived the longstanding idea that worker-run firms tend to be too egalitarian, using a model that involves voting over pay differentials. The claim is that a majority of workers in an LMF where workers differ in ability, e.g. with median ability below mean ability, can benefit by making differences in pay smaller than differences in productivity, thus transferring some returns from the most productive to the least productive workers. Because ability and effort are not easily distinguished, LMFs therefore impose distortionary taxes that tend to dull effort incentives. Although LMFs could gain by committing themselves in advance not to behave this way, Kremer argues that it is too difficult to entrench the necessary constitutional safeguards.

In the absence of mobility barriers, high-ability workers could respond by fleeing to jobs in firms owned by shareholders where they would be paid their marginal product. But if workers must invest some capital in the LMF in advance, and are uncertain about their
abilities ex ante, there will be some scope for redistributive activities. As Kremer recognizes, the argument would not go through if there were a perfect ex post market for membership rights. However, he suggests several reasons (notably adverse selection on worker ability) why such markets may fail in practice.

Flatness of pay has led to known problems in some cooperatives. ‘Distribution according to need’ rather than work has sometimes been seen as a problem in the Israeli kibbutzim, for example. While few worker-run firms are this egalitarian, the Mondragon cooperatives do have rules governing maximum pay differentials which they have had to adjust upwards over time (Bradley and Gelb, 1987). But as mentioned earlier, there is little evidence that worker-run firms are plagued by weak work incentives. Indeed, the evidence generally points in the opposite direction: labor productivity may well be higher on average in LMFs than in similar capitalist firms. This is as much a strike against Kremer’s story as it is against that of Alchian and Demsetz (1972). However, a greater emphasis on recruitment or retention of more capable workers, rather than work effort, could strengthen this approach empirically. It should also be kept in mind that investor-owned firms are not immune to concerns with pay equity (Baker et al., 1988), and that earnings compression may have beneficial effects on morale.

Some related work focuses on the potential instability of firms that are managed by workers but retain individually-tradable shares. Ognedal (1993) suggests that mixed patterns of share ownership where both workers and investors exercise some voting rights are likely to be unstable. The key idea is that majority voting can be used to redistribute profit away from a minority of shareholders. Skillman and Dow (2000) argue that because each firm’s capital needs can be met on a competitive market investors will tend to support profit maximization. But because labor markets are imperfectly competitive, workers typically have diverse preferences about firm policies. It follows that labor-managed firms are generally unstable: investors can offer to buy out such firms on terms that 51% of the worker owners will accept.

Hansmann (1996) and Benham and Keefer (1991) supply empirical evidence to support the idea that collective choice is relatively costly for worker-controlled firms. Various strategies for curbing these costs, such as staying small, screening new members, or using agenda control, are in fact observed in such firms. Hansmann points out that existing worker-controlled firms often have a limited division of labor among members, rather un-differentiated skills and task assignments, and participation rights that are restricted to a relatively homogeneous subset of the workforce. It is less clear that this approach can account for other important stylized facts (e.g. the observation that worker-owned firms seldom arise de novo in capital-intensive industries, or in industries where physical asset specificity is important, holding constant other factors such as firm size). We are unaware of systematic evidence that bears directly on the predicted instability in ownership structures, though there are anecdotal cases in both directions. United Airlines was recently bought out by two of its main employee groups (Gordon, 1999), but there are also examples where plywood cooperatives have sold out to conventional firms. Our general view of the collective-choice hypothesis is that it seems logically coherent and empirically plausible, but we are unsure how important these problems are in relation to other factors that could tip the balance toward control by capital suppliers (especially those sketched in Section 4 above).
6. Discussion and conclusion

The question of why most production in market economies occurs in firms where capital suppliers, rather than labor suppliers, wield control is among the most basic that could be addressed by the discipline of economics. However, many different answers have been suggested, there is at present no consensus, and very little headway has been made in settling the issue empirically. This is unfortunate, as the question is not only of intense intellectual interest, but has substantial policy import: for instance, in evaluating proposals for co-determination along European lines, or for tax and subsidy policies that bear on employee stock ownership plans, profit-sharing, and worker buy-outs.

In this paper, we have attempted to systematize the leading hypotheses, and bring into mutual confrontation ideas that have often been proposed and discussed in isolation from one another. The main ideas that we judged to have logical coherence and at least some measure of empirical plausibility are hypotheses about monitoring and incentives; worker wealth and credit constraints; portfolio diversification by workers; the operation of membership markets; and collective decision-making in firms.

More than one theoretical candidate is left standing at the end of our review. This should not be surprising, since the interactions among multiple causal factors are likely to matter. For instance, worker risk aversion may simultaneously lead to equity investments by outsiders for diversification reasons, and also relatively fixed wages with implications for work incentives. A single factor could also have ambiguous effects that are not easily disentangled through armchair reasoning. For example, firm-specific human capital may make worker control attractive as a mechanism for protecting quasi-rents, but unattractive from the standpoint of portfolio diversification.

For these reasons we think the best way forward to devise an appropriate strategy for analyzing the empirical record. One possible approach would be to exploit existing variation in the incidence of worker control across firms and industries, as a basis for inferring the contributions of the factors discussed in Sections 3–5. The idea would be to identify measures or proxies for such factors as monitoring difficulty, variability of labor returns, capital intensity, worker liquidity constraints, and workforce heterogeneity, and then to study the degree to which the prevalence of worker control is explained by these measures. For instance, if we had data on the proportion of employment accounted for by worker-controlled firms in each of N industries, and data on the aforementioned measures for each industry, we could estimate a multiple regression that would indicate which factors are significant for the prevalence of worker control, whether the signs of coefficients are consistent with the hypotheses, and the rough quantitative contribution of each significant factor to the extent of worker or financier control.

The types of data that would be useful for such an empirical exercise range from those that can be constructed fairly easily with existing series, to those that would require substantial additional survey research to acquire, to those, finally, that one cannot hope to obtain except on a case study basis. But since each of our main hypotheses is linked to at least some variables that are available without extensive new data collection, it seems that some progress can be made with a fairly modest incremental research effort. The central problem may be
to identify industries where worker-controlled firms are well represented but have close conventional firm counterparts.\textsuperscript{13}

Theoretical and empirical research in this area would both gain from a sustained dialogue. In particular, theorists need to inventory existing theories and facts instead of just churning out new hypotheses, which is perhaps all too easy. While a full consensus may prove elusive, one would hope that instructors of economics courses could at least be provided with a better-organized body of ideas about why our economy is a capitalist one. We see this paper as a first attempt at the required synthesis.

Acknowledgements

Both authors are grateful to the MacArthur Foundation for its support. Helpful comments on earlier versions of this paper were provided by members of the Foundation’s Costs of Inequality study group, by the audience at an ASSA session sponsored by the Association for Evolutionary Economics in 1996, and by participants in the conference on Employees and Corporate Governance at Columbia Law School in 1997. Dow would also like to acknowledge financial assistance from the Swedish Collegium for Advanced Study in the Social Sciences, and the Social Sciences and Humanities Research Council of Canada.

References


\textsuperscript{13} The series of studies by Craig and Pencavel, summarized by Pencavel (1996), is probably the best example of empirical work using a closely matched sample of worker-owned and conventional firms. It is restricted, however, to only one industry, the plywood industry in the US. In Ben-Ner et al. (2000), we attempt to implement some of the empirical agenda discussed in the last paragraph, but we are limited to a sample of basically conventional firms some of which exhibit modest degrees of worker participation in decision-making, profit-sharing, or employee ownership.


Pencavel, J., 1996. What has been learned about worker-owned firms from the plywood cooperatives of the Pacific North-west? Department of Economics, Stanford University, in press.


