The Firm as an Incentive System

By BENGT HOLMSTROM AND PAUL MILGROM*

We explore the twin hypotheses (i) that high-performance incentives, worker ownership of assets, and worker freedom from direct controls are complementary instruments for motivating workers, and (ii) that such instruments can be expected to covary positively in cross-sectional data. We also relate our conclusions to empirical evidence, particularly that on the organization, compensation, and management of sales forces. (*JEL* L22, D23, M31)

One of the most intensively studied topics in the modern theory of the firm is the make-or-buy decision. This is the firm's decision either to acquire some intermediate input by having an employee make it under the employer's direction, using the employer's tools, and usually being paid a fixed wage, or instead to hire an independent contractor who chooses his or her own tools and methods and is paid proportionally to the quantity supplied. Most analyses of the make-or-buy decision have focused on just one of the differences that tend to distinguish employment from independent contracting. For example, Ronald Coase (1937) and Herbert Simon (1951) emphasize the discretion that the employer has to direct the employee's activities; Benjamin Klein et al. (1978), Oliver Williamson (1985), and Sanford Grossman and Oliver Hart (1986) focus on the firm's ownership of assets; and the principal–agent literature (e.g., Armen Alchian and Harold Demsetz, 1972; Holmstrom, 1982) stresses monitoring and compensation issues. What is not adequately explained by any of these analyses is how these choices are intertwined. Why does inside procurement tend to involve production by a worker who is supervised by the firm *and* uses the firm's tools *and* is paid a fixed wage? Why does outside procurement tend to involve purchases from a worker who chooses his or her own methods and hours *and* owns the tools used *and* is paid only for quantities supplied?

In this paper, we explore the hypothesis that these tendencies emerge because they describe two alternative *systems* for managing incentives for the wide array of tasks for which a single worker may be responsible. Firms use a variety of incentive *instruments* in such systems. Perhaps the most direct incentive is to pay the agent based on measured performance in a given task or set of tasks. But monitoring is imperfect and costly, enabling only a narrow set of activities to be rewarded effectively this way. Asset ownership is often a broader, more powerful incentive instrument. When an agent owns a set of productive assets, she maintains those assets more effectively. She also reaps the many implicit returns that accrue through such ownership, notably those stemming from an enhanced bargaining position. A third major incentive instrument is the design of the job: the tasks included in the job description, the activities that are expressly excluded (such as working for other firms), and the specification of work rules, working hours, and similar policies that restrict the freedom of the worker. The three theories mentioned in the opening paragraph, each focus on just one of these instruments. Our

*School of Organization and Management, Yale University, Box 1A, New Haven, CT 06520, and Department of Economics, Stanford University, Stanford, CA 94305-6072, respectively. We thank the National Science Foundation for financial support and James Brickley, Oliver Hart, James Heckman, Preston McAfee, Margaret Meyer, and Ariel Pakes for extensive discussions and comments on this work.

main interest is in understanding whether a coordinated use of the instruments explains their typical covariation.

Our analysis is built around the observation that, when several activities compete for a worker's attention, there is a tendency for the *levels of incentives* provided for the different activities of a worker to be *complementary* in the incentive problem. The intuitive idea is that increasing the incentive for just one task could cause a worker to devote too much effort to that one task while neglecting other aspects of the job, and that increasing incentives for all of the agent's activities avoids that cost. Asset ownership, contingent rewards, and job restrictions, all influence different dimensions of the worker's task portfolio. If, as suggested, there is a desire to keep the various incentives in balance, then one would expect that in an optimal system, the three instruments would have to be similarly balanced. Weak incentives for maintaining asset values should go with weak incentives for narrowly measured performance and significant restrictions on worker freedom (excluding an activity is the same as setting its incentive to zero). These, of course, are the attributes that typically characterize employment.

While this intuition may sound plausible, it is incomplete. All the instruments are endogenous variables in the problem of structuring incentives. So one has to ask what explains the choice between different incentive systems: why are some workers employees and other workers independent contractors? To address this, we have to introduce exogenous parameters that move the system solution around. Variations in the cost of measuring performance, in asset specificity, and in uncertainty about the future are all good candidates. The question is: will changes in these exogenous parameters move the incentive intensities and instruments in the same direction as presumed in the intuition about balance?

The answer depends on the parameter we look at. A major contribution of the analysis is to show how to assess which exogenous parameters lead to comovements in the incentive instruments and which do not. It is important to observe that, even when the instruments are complementary (i.e., using one more intensively increases the marginal benefit of using the others more intensively), this still does not assure that the instruments will move in the same direction in response to a shift in an exogenous parameter, such as measurement costs. What is required in addition is that an increase in measurement cost will move all the marginal incentive benefits in the same direction: the returns from ownership, from performance rewards, and from worker freedom, should all either go up or down with such a change.

The final step in the analysis is to identify conditions under which comparative-static comovements translate into statistical covariations in pooled data. For this we employ the concepts of associated and affiliated random variables, which are ordinal concepts of statistical covariation. Using the assumption that the parameters are associated or affiliated allows us to make statistical predictions even when some relevant parameters are censored in the data.

We will analyze these issues using the multitask principal–agent model introduced in our 1991 paper. This model lends itself naturally to the study of work incentives in an environment characterized by competing worker activities. It also offers a convenient way for introducing the various incentive instruments that we have discussed. We begin by presenting the model in a very general form. The purpose is not to derive our results in this general form; for that the model has too little structure. Rather, the intention is to present it as a theoretical laboratory that can be easily adapted to particular empirical contexts. In any given specification of the model, comovements can be readily analyzed. Exogenous parameters that are troublesome, in the sense that they will not give rise to comovements, can be identified, informing researchers about how to implement the model empirically. Too often, it seems, theorists do not go far enough in advising empiricists about the need for controlling certain exogenous sources of variation. Our model is one in which these issues can be given definite answers if one is careful about tailoring the model to the empirical context.

As an illustration of our general approach, we will analyze a stylized model of industrial selling described in the empirical studies of Eric Anderson and David Schmittlein (1984) and Anderson (1985). They were interested in testing transactions-cost explanations of why industrial selling is sometimes performed by in-house salespeople and sometimes by independent representatives (often the same firm uses both methods: direct selling in some regions and independent representation in other regions). As they describe it, this is a case in which sales agents have many competing demands on their time, and presumably because of the different incentives in place, this time gets allocated differently depending on whether the agent is an employee or an independent representative.

In our model of industrial selling, we can derive precise conditions under which the incentive instruments are complements, study which exogenous variables lead to comovements, and finally provide conditions on the distribution of the exogenous variables that are sufficient for the comovements to result in statistical covariations among the instruments and parameters—that is, in positive correlations between each pair of them. Our main finding is that, when the cost of measuring sales performance is high (e.g., because it involves team selling) or when hard-to-measure nonselling activities are important, it is more likely that the agent's optimal incentives will conform with the attributes of employment: modest commissions, firm ownership of customers, and no right for the agent to sell the products of other manufacturers. On the other hand, when performance is easy to measure or when nonselling activities are unimportant, incentives will conform with the attributes of independent representation: strong output-based incentives, customer ownership, and freedom to sell the products of other manufacturers. We also find that changes in the marginal value of direct selling and in nurturing customer relationships may lead to negative comovements and therefore may have to be controlled for in empirical work.

Our results fit well with the main findings in the Anderson-Schmittlein studies. Their data revealed that the most important variables determining the choice between in-house and independent sales agents were the "difficulty of evaluating performance" and the "importance of nonselling activities." Also, they report that independent sales representatives were compensated entirely by commissions and were allowed to sell other manufacturers' products, while employee agents were paid by salary (with small additional commissions) and were not allowed to sell other manufacturers' products, all in line with our theoretical predictions. Notably, variables meant to reflect Williamson's (1985) version of transaction-cost theory, emphasizing asset specificity, uncertainty, and the interactions between the two, all proved much less significant.

The paper proceeds as follows. Section I describes the general framework. Section II presents our methodological approach. It uses the theory of supermodular functions to derive comparative-statics results and the theory of associated and affiliated random variables to translate these into statistical predictions about positive covariations and conditional expectations. Our statistical analysis is of independent interest. It can be applied to other models based on complementarities—for example, to the modern manufacturing model of Milgrom and John Roberts (1990). Section III is devoted to the sales-agency application and includes a more detailed discussion of the Anderson-Schmittlein findings as well as additional evidence bearing on our theory. Section IV concludes the paper.

I. A General Model

Our analysis is based on the dynamic principal–agent model that we introduced in Holmstrom and Milgrom (1987). There we showed that in suitably stationary environments in which the agent can continuously monitor his own performance, the optimal incentive contract coincides with the optimum of a certain particularly tractable reduced-form static model. In the reduced form, the principal is constrained to pay the agent a linear function of the vector of

time-aggregated observations, and the agent is constrained to choose a single vector of efforts, which corresponds to the instant-by-instant allocation of efforts in the original dynamic model. Our analysis below is conducted using this reduced form.

The agent can allocate effort among several activities $n = 1, \ldots, N$, with the full vector of efforts denoted $\mathbf{t} = (t_1, \ldots, t_N)$. The agent's choice of $\mathbf{t}$ cannot be directly observed, but it can be monitored indirectly via a collection of measures $\mathbf{X} = (X_1, \ldots, X_I)$:

$$(1) \qquad X_i = F_i(t) + \varepsilon_i \qquad i = 1, \ldots, I.$$

The terms ε_i are normally distributed measurement errors with zero means. Each function F_i could represent the individual's measured contribution to profit, revenue, or cost, or some other performance indicator, or it could represent estimates of inputs such as the total time or energy that the agent devotes to the job.

As a modeling strategy, we introduce X_i's to represent any information the principal and agent can contract upon, even if that item is a return that appears elsewhere in the model. For example, if the parties can contract directly on the agent's costs or on some component of them, then even though that cost will be represented elsewhere in the model, we assume that there is some X_j to represent it. With this convention, the optimal incentive scheme takes the linear form:

$$(2) \qquad s(\mathbf{X}) = \sum_i \alpha_i X_i + \beta$$

where the coefficients $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_I)$ are called *commission rates* and β is *salary* (although this could be negative in our model). We restrict the α_i's to be nonnegative to avoid creating an incentive for the agent to conceal performance.

The agent in our model incurs a private cost $C(\mathbf{t})$ if he chooses $\mathbf{t}$. This cost is transferable only to the extent that information about it is recorded by one of the variables X_i. Effort is the standard example of such a private cost, but there are others. The agent's private return C can include the value of any increments to the agent's human capital (accounted for as a negative cost) as well as the opportunity cost of time spent on the principal's job. The model allows that some of the t_n's may not affect the private cost C at all (manipulating accounts could be one example).

There is a corresponding private benefit $B(\mathbf{t})$ that accrues to the principal. As in the agent's case, $B(\mathbf{t})$ could be nontransferrable because it relates to the principal's human capital or it could represent returns to assets that for unmodeled reasons are best assigned to the principal in all circumstances (see below). If any of the agent's activities impose private costs on the principal, these of course are recorded as negative benefits. In principle, it is possible that $B(\mathbf{t}) = -C(\mathbf{t})$, in which case the principal and the agent would share intrinsic preferences over the choice of $\mathbf{t}$ and there would be no incentive problem to solve.

In addition to the nontransferrable returns B and C, there are two other categories of returns included in the model. One is a *transferrable* return $Y(\mathbf{t})$ associated with asset ownership. This return can be allocated between the principal and the agent by choice of an ownership structure denoted by $\boldsymbol{\lambda}$. The most general formulation would specify an arbitrary division:[1]

$$(3) \qquad Y(\mathbf{t}) = Y_A(\mathbf{t}; \boldsymbol{\lambda}) + Y_P(\mathbf{t}; \boldsymbol{\lambda}).$$

For instance, if we think of $\boldsymbol{\lambda}$ as a configuration of ownership rights in incomplete-contract models of organization (Grossman and Hart, 1986; Hart and John Moore, 1990), then $Y_A(\mathbf{t}; \boldsymbol{\lambda})$ is the share of returns that the agent is able to obtain in the *ex post* bargain with the principal under configuration $\boldsymbol{\lambda}$.[2] Incomplete-contract mod-

[1] We do not consider the possibility that $\boldsymbol{\lambda}$ alters performance measures X_i. For a model in which ownership changes the informativeness of stock prices, see Holmstrom and Jean Tirole (1993).

[2] Using an inequality in (1) would allow one to model cases in which some of the returns are wasted in the process of negotiating the division of Y.

els examine the mapping from $\boldsymbol{\lambda}$ to $Y_A(\mathbf{t};\boldsymbol{\lambda})$ and seek to derive the reduced form (3) from what are allegedly more primitive assumptions about the process of bargaining. These models frequently treat ownership as the only means by which incentives can be affected. Our approach is to take the reduced form $Y_A(\mathbf{t};\boldsymbol{\lambda})$ as given and proceed to consider additional instruments that can be used to affect incentives. In doing so, we will not be as general as (3), but instead will assume that $Y(\mathbf{t})$ is the sum of a collection of returns $Y_j(\mathbf{t})$, $j=1,\ldots,J$, each of which can be allocated either fully to the principal or fully to the agent. We let $\lambda_j=1$ $(=0)$ indicate that the return Y_j accrues to the agent (principal). An ownership configuration then is a vector $\boldsymbol{\lambda}=(\lambda_1,\ldots,\lambda_J)$ of 0's and 1's. Moreover, we allow that asset returns may be random and assume that the random return Y_j takes the form

$$(4)\quad Y_j = G_j(\mathbf{t}) + \varepsilon_{I+j} \qquad j=1,\ldots,J.$$

Here again, G_j could be profits, costs, revenues, and so on. The stochastic portion of the return, ε_{I+j} is assumed to be normally distributed with zero mean.

Finally, there is a hybrid category of returns, those that, while not transferable, are nevertheless *excludable*:

$$(5)\quad Z_k = H_k(\mathbf{t}) + \varepsilon_{I+J+k} \qquad k=1,\ldots,K.$$

The error terms in (5) have the same interpretation and properties as those in (4). Excludable returns accrue to the agent if the contract permits the agent to collect them.[3] For each Z_k, there is a dichotomous contractual variable δ_k, such that if $\delta_k=1$ the agent can enjoy Z_k, while if $\delta_k=0$, Z_k is zero (or dissipated). The difference between Y_j and Z_k is that the principal can never enjoy Z_k. The implicit assumption is that it is costless to monitor whether the benefits Z_k are being enjoyed and to exclude them by contract. For example, Z_k may represent a sales agent's commissions from selling the products of other principals, or his private savings from choosing particular tools, or work methods, or hours of work.

Because the general model distinguishes returns from information and incorporates the use of various kinds of incentive instruments, it is quite a flexible tool for describing different agency contexts. The demands of stationarity in the underlying dynamic model impose the most significant restrictions on what the model can represent. In this version, one important implication is that all the stochastic terms $\boldsymbol{\varepsilon}=(\varepsilon_1,\ldots,\varepsilon_{I+J+K})$ are jointly normally distributed and enter additively into the observations. Let $\boldsymbol{\Sigma}$ be the variance-covariance matrix of these random terms. Later, we will allow the possibility that the intensity of monitoring may be chosen contractually, which affects the covariance matrix $\boldsymbol{\Sigma}$, at a cost $K(\boldsymbol{\Sigma})$.

To summarize, the principal and the agent have the following organizational design variables to choose from:

(i) commission rates: $\boldsymbol{\alpha}=(\alpha_1,\ldots,\alpha_I)$, $\alpha_i\geq 0$,
(ii) allocation of transferrable returns: $\boldsymbol{\lambda}=(\lambda_1,\ldots,\lambda_J)$, $\lambda_j\in\{0,1\}$,
(iii) exclusion of private returns: $\boldsymbol{\delta}=(\delta_1,\ldots,\delta_K)$, $\delta_k\in\{0,1\}$,
(iv) monitoring intensity: $\boldsymbol{\Sigma}$.

The implied financial payoffs from the organizational design $(\boldsymbol{\alpha},\boldsymbol{\lambda},\boldsymbol{\delta},\boldsymbol{\Sigma})$ are:

$$(6)\quad \text{Principal: } P = B(\mathbf{t}) - \sum_i \alpha_i X_i + \sum_j (1-\lambda_j) Y_j - \beta - K(\boldsymbol{\Sigma})$$

$$(7)\quad \text{Agent: } A = \sum_i \alpha_i X_i + \sum_j \lambda_j Y_j + \sum_k \delta_k Z_k + \beta - C(\mathbf{t}).$$

[3]There is never a reason to exclude any positive returns to the principal, since the principal supplies no productive inputs.

Observe that for any choices of $\boldsymbol{\alpha}$, $\boldsymbol{\Sigma}$, and $\mathbf{t}$, both P and A will be normally distributed.

We assume that the principal is risk-neutral and that the agent has constant absolute risk aversion:

$$U(A) = -\exp(-rA) \tag{8}$$

where r is the coefficient of risk aversion. Since A is normally distributed, we can write the agent's utility measure (8) in the certainty-equivalent form:

$$\begin{aligned}\text{ACE}(\mathbf{t},\boldsymbol{\alpha},\boldsymbol{\lambda},\boldsymbol{\delta},\boldsymbol{\Sigma}) &= \sum_i \alpha_i F_i(\mathbf{t}) + \sum_j \lambda_j G_j(\mathbf{t}) \\ &\quad + \sum_k \delta_k H_k(\mathbf{t}) - C(\mathbf{t}) \\ &\quad - \tfrac{1}{2} r V(\boldsymbol{\alpha},\boldsymbol{\lambda},\boldsymbol{\delta},\boldsymbol{\Sigma})\end{aligned} \tag{9}$$

where $V(\boldsymbol{\alpha},\boldsymbol{\lambda},\boldsymbol{\delta},\boldsymbol{\Sigma})$ is mnemonic for the variance of income implied by $(\boldsymbol{\alpha},\boldsymbol{\lambda},\boldsymbol{\delta},\boldsymbol{\Sigma})$ and ACE is the *agent's certainty equivalent*. We will not write out the expression for V; it is a quadratic function of the commission rates $\boldsymbol{\alpha}$ and a linear function of $\boldsymbol{\Sigma}$.

Since a dollar of certainty equivalent can always be transferred between the agent and the principal by transferring a physical dollar without affecting the agent's performance incentives, an efficient $(\mathbf{t},\boldsymbol{\alpha},\boldsymbol{\lambda},\boldsymbol{\delta},\boldsymbol{\Sigma})$ must maximize the total certainty equivalent:

$$\begin{aligned}\max[\text{TCE}(\mathbf{t},\boldsymbol{\alpha},\boldsymbol{\lambda},\boldsymbol{\delta},\boldsymbol{\Sigma})] &\equiv B(\mathbf{t}) + \sum_j G_j(\mathbf{t}) + \sum_k \delta_k H_k(\mathbf{t}) \\ &\quad - C(\mathbf{t}) - K(\boldsymbol{\Sigma}) - \tfrac{1}{2} r V(\boldsymbol{\alpha},\boldsymbol{\lambda},\boldsymbol{\delta},\boldsymbol{\Sigma})\end{aligned} \tag{10}$$

subject to the agent's optimal choice,

$$\mathbf{t} = \underset{\mathbf{t}'}{\operatorname{argmax}} \left\{ \sum_i \alpha_i F_i(\mathbf{t}') + \sum_j \lambda_j G_j(\mathbf{t}') + \sum_k \delta_k H_k(\mathbf{t}') - C(\mathbf{t}') \right\}.$$

Here $\text{TCE}(\mathbf{t},\boldsymbol{\alpha},\boldsymbol{\lambda},\boldsymbol{\delta},\boldsymbol{\Sigma})$ is the *total certainty equivalent* or *joint surplus*. Note that we have dropped the risk term from the agent's maximization problem, since it is unaffected by the agent's actions. If F_i, G_j, and H_k are all concave functions and C is strictly convex, the agent's choice will be characterized by a set of first-order conditions that uniquely define the agent's response function, denoted $\mathbf{t}(\boldsymbol{\alpha},\boldsymbol{\lambda},\boldsymbol{\delta})$. If we substitute this response function into the objective TCE, we have reduced (10) to an unconstrained optimization over the organizational choice variables $(\boldsymbol{\alpha},\boldsymbol{\lambda},\boldsymbol{\delta},\boldsymbol{\Sigma})$:

$$\begin{aligned}\max_{\boldsymbol{\alpha},\boldsymbol{\lambda},\boldsymbol{\delta},\boldsymbol{\Sigma}} [T(\boldsymbol{\alpha},\boldsymbol{\lambda},\boldsymbol{\delta},\boldsymbol{\Sigma})] &\equiv \text{TCE}(\mathbf{t}(\boldsymbol{\alpha},\boldsymbol{\lambda},\boldsymbol{\delta}),\boldsymbol{\alpha},\boldsymbol{\lambda},\boldsymbol{\delta},\boldsymbol{\Sigma}).\end{aligned} \tag{11}$$

Expression (11) is the general form of the maximization problem that we will be studying.

II. The Analytical Approach

Let $\mathbf{x}$ denote the optimal choice of $(\boldsymbol{\alpha},\boldsymbol{\lambda},\boldsymbol{\delta},\boldsymbol{\Sigma})$ in program (11). Our general objective is to determine how the design instruments in $\mathbf{x}$ covary across a set of agency relationships. This requires that we introduce explicit parameters of TCE that characterize the heterogeneity of the population we are interested in studying. Let $\boldsymbol{\pi}$ represent a generic parameter vector and $\mathbf{x}(\boldsymbol{\pi})$ the corresponding optimal design $\mathbf{x}$. Expressed in this form, we are interested in the properties of the distribution of $\mathbf{x}(\boldsymbol{\pi})$ induced by the underlying variation in the population of $\boldsymbol{\pi}$.

The first step is to characterize how $\mathbf{x}$ varies with $\boldsymbol{\pi}$. For this step we will apply the theory of supermodular functions: *supermodular* is the term for a function in which every pair of arguments are Edgeworth complements. When the supermodularity condition is satisfied, there is an expedient and general approach for dealing with the complexities that arise from interactions between the endogenous variables $\mathbf{x}$.

The second step is to establish conditions under which the distribution of $\boldsymbol{\pi}$ induces

positive covariations in $\mathbf{x}$. For this step we apply the theories of associated and affiliated random variables. All of these theories have been discussed in detail elsewhere,[4] so we can be relatively brief and not overly formal.

A function $f\colon \mathbb{R}^k \to \mathbb{R}$ is *supermodular* if for any two arguments $\mathbf{z}$ and $\mathbf{z}'$:

$$f(\mathbf{z}\vee\mathbf{z}')+f(\mathbf{z}\wedge\mathbf{z}')\geq f(\mathbf{z})+f(\mathbf{z}')$$

where $\mathbf{z}\vee\mathbf{z}'$ denotes the component-wise maximum and $\mathbf{z}\wedge\mathbf{z}'$ the component-wise minimum of $\mathbf{z}$ and $\mathbf{z}'$. The definition implies that if all the variables of a supermodular function are increased simultaneously, the function value increases by more than if we were to sum up the value changes from increasing the variables one at a time. If f is smooth, supermodularity is equivalent to the condition: $\partial^2 f/\partial z_i \partial z_j \geq 0$ for all $i \neq j$. In this form it is transparent that supermodularity reflects complementarities between the variables: when one goes up, the marginal return from increasing the other variables also goes up. Note that nothing is assumed about $\partial^2 f/\partial z_i^2$; thus supermodularity neither implies nor is implied by concavity. It is immediate that if f and g are two supermodular functions, then $f+g$ is supermodular.

Let $f(\mathbf{x},\boldsymbol{\pi})$ be supermodular, where $\mathbf{x}$ is interpreted as an n-vector of endogenous variables and $\boldsymbol{\pi}$ as an m-vector of parameters. It is intuitive that if we maximize $f(\mathbf{x},\boldsymbol{\pi})$ over $\mathbf{x}$ and consider how the optimal value $\mathbf{x}(\boldsymbol{\pi})$ varies with changes in $\boldsymbol{\pi}$, then an increase in any component of $\boldsymbol{\pi}$ will trigger a sequence of upward adjustments in the components of $\mathbf{x}$, all of which reinforce each other. The net result should be that $\mathbf{x}(\boldsymbol{\pi})$ increases with $\boldsymbol{\pi}$. The formal statement of this conclusion, given below, takes into account the possibility of multiple optima and includes the additional observation that, with multiple optima, the component-wise supremum and infimum of the set of maximizers are also maximizers. What makes this result particularly useful for us is that it holds even if the choice of $\mathbf{x}$ is restricted to certain subsets $\mathbf{X}$ of $\mathbb{R}^n$, called sublattices. $\mathbf{X}$ is a *sublattice* if $\mathbf{x}\vee\mathbf{x}'$ and $\mathbf{x}\wedge\mathbf{x}'$ belong to $\mathbf{X}$ whenever $\mathbf{x}$ and $\mathbf{x}'$ belong to $\mathbf{X}$. We will later use the fact that all product sets in $\mathbb{R}^n$ are sublattices.

THEOREM 1: *Let $f(\mathbf{x},\boldsymbol{\pi})$ be a continuous supermodular function and $\mathbf{X}$ a compact sublattice of $\mathbb{R}^n$. Let $\mathbf{X}(\boldsymbol{\pi})=\operatorname{argmax}\{f(\mathbf{x},\boldsymbol{\pi})\mid \mathbf{x}\in\mathbf{X}\}$ be the set of maximizers with least upper bound $\mathbf{x}^*(\boldsymbol{\pi})=\sup\mathbf{X}(\boldsymbol{\pi})$ and greatest lower bound $\mathbf{x}_*(\boldsymbol{\pi})=\inf\mathbf{X}(\boldsymbol{\pi})$. Then $\mathbf{x}^*(\boldsymbol{\pi})$ and $\mathbf{x}_*(\boldsymbol{\pi})$ are both elements of $\mathbf{X}(\boldsymbol{\pi})$ and both $\mathbf{x}^*(\cdot)$ and $\mathbf{x}_*(\cdot)$ are nondecreasing functions (from $\mathbb{R}^m$ to $\mathbb{R}^n$).*

In case of multiple optima, we will focus on the largest one, corresponding to $\mathbf{x}^*(\boldsymbol{\pi})$ in Theorem 1, and henceforth denoted by $\mathbf{x}(\boldsymbol{\pi})$. All the results to follow could be restated with a provision for multiple optima as in Theorem 1.

Suppose $\boldsymbol{\pi}$ is a one-dimensional parameter and $f(\mathbf{x},\boldsymbol{\pi})$ a supermodular function. Then, since $\mathbf{x}(\boldsymbol{\pi})$ is monotone in $\boldsymbol{\pi}$, all the components of $\mathbf{x}(\boldsymbol{\pi})$ will covary positively in a population, irrespective of how $\boldsymbol{\pi}$ is distributed. But what if $\boldsymbol{\pi}$ is a multidimensional parameter? The answer depends on the distribution of $\boldsymbol{\pi}$. For example, if two components of $\boldsymbol{\pi}$ are negatively correlated, say $\operatorname{Cov}(\pi_1,\pi_2)<0$, $x_1=\pi_1$ and $x_2=\pi_2$, then obviously $\operatorname{Cov}(x_1(\boldsymbol{\pi}),x_2(\boldsymbol{\pi}))<0$: two nondecreasing functions of the same vector of parameters can be negatively correlated. Indeed, even if $\operatorname{Cov}(\boldsymbol{\pi})\geq 0$, there is no guarantee that $\operatorname{Cov}(\mathbf{x}(\boldsymbol{\pi}))\geq 0$, because covariation is not preserved under arbitrary monotone transformations of the random variables.[5]

[4]The basic results on supermodular optimization were developed by Donald Topkis (1978). An elementary treatment can be found in Milgrom and Roberts (1990) and additional results are reported by Milgrom and Chris Shannon (1994). The theory of affiliated random variables is due to Milgrom and Robert Weber (1982), who also summarize some of the main results about associated random variables.

[5]The notation $\operatorname{Cov}(\boldsymbol{\pi})\geq 0$ means that all the elements of the covariance matrix are nonnegative.

A more useful concept of covariation for our analysis is the one known as *association*. A vector $\boldsymbol{\pi}$ of random variables is associated if for all real-valued nondecreasing functions f and g, $\mathrm{Cov}(f(\boldsymbol{\pi}), g(\boldsymbol{\pi})) \geq 0$. An equivalent and more directly usable statement of this condition is that

$$(12) \quad E[f(\boldsymbol{\pi})g(\boldsymbol{\pi})] \geq E[f(\boldsymbol{\pi})]E[g(\boldsymbol{\pi})].$$

We list below some useful properties of associated random variables.

THEOREM 2: *Let* $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_m)$ *be a vector of associated random variables. Then:*

(i) any subvector of $\boldsymbol{\pi}$ *is associated; in particular, any single random variable is associated;*

(ii) $\mathrm{Cov}(\pi) \geq 0$*;*

(iii) if γ *is a random variable, independent of* $\boldsymbol{\pi}$*, then* $(\boldsymbol{\pi}, \gamma)$ *is associated; in particular, any vector of independent random variables is associated;*

(iv) if $x_i: \mathbb{R}^m \to \mathbb{R}$, $i = 1, \ldots, n$, *is a collection of nondecreasing functions, then* $(\mathbf{x}(\boldsymbol{\pi}), \boldsymbol{\pi})$ *is associated; in particular,* $\mathbf{x}(\boldsymbol{\pi})$ *is associated.*

Part (i) follows by applying (12), taking f and g to be independent of the excluded variables. Part (ii) is (12) with f and g as projections. Part (iii) follows from applying the law of iterated expectations; we will see a similar argument in Theorem 3. Part (iv) is true because compositions of nondecreasing functions are nondecreasing.

Applied to our problem, parts (ii) and (iv) of Theorem 2, together with Theorem 1, imply that the design instruments $\mathbf{x}$ will be associated whenever the parameters $\boldsymbol{\pi}$ are associated and the objective function TCE is supermodular in $(\mathbf{x}, \boldsymbol{\pi})$. Moreover, if a subset $\boldsymbol{\pi}'$ of the parameters can be observed, then by parts (iv) and (i), $(\boldsymbol{\pi}', \mathbf{x}(\boldsymbol{\pi}))$ is associated.

Note that we are not making any parametric assumptions about the distribution of $\boldsymbol{\pi}$. This suggests that we apply nonparametric inequality tests to check whether the data reject positive covariation (see R. Barlow et al., 1972). Since the hypothesis of association is preserved by order-preserving transformations of the variables, all the relevant information is contained in the rank-order data. One way to test association is to test the hypothesis that the smallest of the sampled rank-order covariances is greater than or equal to zero. Depending on the most plausible competing economic hypotheses, other tests might be preferred.

The restrictions on endogenous variables in Theorem 2 are about as powerful as one could hope for. However, they require strong conditions, too. Since TCE will be supermodular only for certain subsets of parameters, $\mathbf{x}(\boldsymbol{\pi})$ will typically not be monotone in the full vector $\boldsymbol{\pi}$, and this will limit the applicability of the results. We will momentarily discuss how to deal with parameters that upset monotonicity. But first, we should point out that certain kinds of nonmonotonicities cause no problems. To give a trivial example, let $\mathbf{x}: \mathbb{R}^n \to \mathbb{R}$ be arbitrary and suppose $f: \mathbb{R} \to \mathbb{R}$ is nondecreasing. Then, $f(\mathbf{x}(\boldsymbol{\pi}))$ and $\mathbf{x}(\boldsymbol{\pi})$ are associated irrespective of how the vector $\boldsymbol{\pi}$ is distributed and whether $\mathbf{x}$ is monotone in $\boldsymbol{\pi}$. This is a special case of the following more general result.

THEOREM 3: *Suppose* $\boldsymbol{\gamma}$ *is a vector of random variables, which is independent of a vector of associated random variables* $\boldsymbol{\pi}$*. Let* $\mathbf{x}(\boldsymbol{\gamma}, \boldsymbol{\pi})$ *be real-valued and monotone nondecreasing in* $\boldsymbol{\pi}$ *for all* $\boldsymbol{\gamma}$ *and let* $\mathbf{y}(\mathbf{x}, \boldsymbol{\pi})$ *be a nondecreasing vector-valued function of* $\mathbf{x}$ *and* $\boldsymbol{\pi}$*. Then* $(\boldsymbol{\pi}, \mathbf{x}, \mathbf{y})$ *is a vector of associated random variables.*

PROOF:

We need to show that for arbitrary nondecreasing functions f and g,

$$(13) \quad E[f(\boldsymbol{\pi}, \mathbf{x}, \mathbf{y})g(\boldsymbol{\pi}, \mathbf{x}, \mathbf{y})] \geq E[f(\boldsymbol{\pi}, \mathbf{x}, \mathbf{y})]E[g(\boldsymbol{\pi}, \mathbf{x}, \mathbf{y})].$$

By the law of iterated expectations,

$$(14) \quad E[fg] = E[E[fg|\boldsymbol{\pi}]] \geq E[E[f|\boldsymbol{\pi}]E[g|\boldsymbol{\pi}]]$$

where the inequality follows by applying (12) to the conditional expectation inside the brackets. This we can do since f and g are nondecreasing functions of the single random variable $\mathbf{x}$ when $\boldsymbol{\pi}$ is held fixed.

Next, we claim that $E[f|\boldsymbol{\pi}]$ and $E[g|\boldsymbol{\pi}]$ are nondecreasing functions of $\boldsymbol{\pi}$. Written out,

$$(15)\quad E[f|\boldsymbol{\pi}] = \int f(\boldsymbol{\pi}, \mathbf{x}(\boldsymbol{\gamma}, \boldsymbol{\pi}), \mathbf{y}(\mathbf{x}(\boldsymbol{\gamma}, \boldsymbol{\pi}), \boldsymbol{\pi}))\, dF(\boldsymbol{\gamma})$$

where we have written the distribution function as $F(\boldsymbol{\gamma})$ instead of $F(\boldsymbol{\gamma}|\boldsymbol{\pi})$ because $\boldsymbol{\gamma}$ and $\boldsymbol{\pi}$ are independent. Since compositions of nondecreasing functions are nondecreasing, f is nondecreasing in $\boldsymbol{\pi}$ for each $\boldsymbol{\gamma}$. Hence $E[f|\boldsymbol{\pi}]$ is nondecreasing in $\boldsymbol{\pi}$. The argument for $E[g|\boldsymbol{\pi}]$ is identical.

Because $\boldsymbol{\pi}$ is associated, we can apply (12) again to conclude

$$(16)\quad E[E[f|\boldsymbol{\pi}]E[g|\boldsymbol{\pi}]] \geq E[E[f|\boldsymbol{\pi}]E[E[g|\boldsymbol{\pi}]] = E[f]E[g].$$

Combining (14) and (16) yields (13).

Had we assumed that $\mathbf{x}$ is monotone in $\boldsymbol{\gamma}$ and that $\boldsymbol{\gamma}$ is a vector of associated random variables, then the conclusion of Theorem 3 would have followed directly from Theorem 2. The thrust of Theorem 3 is that parameters may enter in a nonmonotone way as long as they are independent of the other parameters and enter through a single endogenous variable. This may seem like a minor relaxation, but it will prove useful. One application is the following. Suppose that we are studying the covariation between a single exogenous parameter π and a single endogenous variable $x(\pi, \boldsymbol{\gamma})$, where $\boldsymbol{\gamma}$ represents all the other parameters of the model. If x is increasing in π and π is independent of $\boldsymbol{\gamma}$, then no matter how x behaves as a function of $\boldsymbol{\gamma}$, x and π will be associated. This shows how a traditional one-parameter comparative-statics result, namely that x is increasing in π, can be used to predict a positive covariation between the parameter and the endogenous variable in a population characterized by multiple sources of heterogeneity—even when dependencies are nonlinear.

It is worth pointing out a second result that can be proved essentially the same way as Theorem 3. Let $\mathbf{x}$ be a vector of associated random variables and suppose that the components of $\mathbf{x}$ are observed with error. Let $\tilde{\mathbf{x}}$ denote the observed value of $\mathbf{x}$. If the error terms are independent of each other and of $\mathbf{x}$, then $\tilde{\mathbf{x}}$ is a vector of associated random variables. This shows that the conclusions of Theorem 2 are robust to independent measurement errors in both endogenous and exogenous variables.[6]

Now, let us get back to the general case. Suppose $\mathbf{x}(\boldsymbol{\pi})$ is nondecreasing for some, but not all the parameters of the vector $\boldsymbol{\pi}$. Let the subvector of parameters for which $\mathbf{x}$ is nondecreasing be $\boldsymbol{\pi}_1$ and the vector of troublesome parameters be $\boldsymbol{\pi}_2$; that is, $\boldsymbol{\pi} = (\boldsymbol{\pi}_1, \boldsymbol{\pi}_2)$. The standard empirical approach is to keep the parameters in $\boldsymbol{\pi}_2$ fixed by controlling for them in a regression (assuming that they can be measured). Indeed, when testing traditional one-parameter comparative-statics results, $\boldsymbol{\pi}_2$ should constitute all but the one parameter π_1 that is being varied. In reality, of course, this is impossible so the empiricist proceeds to control for only those variables that are thought to be potentially troublesome. In our approach, in which parameters are viewed as random variables, this works, but it requires a stronger assumption on the parameter distribution than association. The relevant condition is known as *affiliation*. A vector of random variables $\boldsymbol{\pi}$ is affiliated if conditional on each sublattice $\mathbf{X}$ the variables are associated; that is,

[6] Instead of assuming that the measurement errors are independent of each other, we could have assumed that they are associated but additive; that is, $\tilde{x}_i = x_i + \gamma_i$, or more generally, that $\tilde{x}_i$ is increasing in γ_i. Systematic biases in individual observations of $\mathbf{x}$ could give rise to association; for instance, some firms may overstate all their data while other firms may understate them, or there could be an uncontrolled scale effect that shifts all the data for a firm in the same direction.

$E[f(\boldsymbol{\pi})g(\boldsymbol{\pi})|\mathbf{X}] \geq E[f(\boldsymbol{\pi})|\mathbf{X}]E[g(\boldsymbol{\pi})|\mathbf{X}]$ for all nondecreasing functions f and g and all sublattices $\mathbf{X}$. As a special case, independent random variables are affiliated. Since $\{\boldsymbol{\pi}|\boldsymbol{\pi}_2 = \bar{\boldsymbol{\pi}}_2\}$ is a sublattice, the parameters in $\boldsymbol{\pi}_1$ are associated conditional on fixing $\boldsymbol{\pi}_2$. Now we can apply our earlier results to this sublattice. Conditional on $\boldsymbol{\pi}_2 = \bar{\boldsymbol{\pi}}_2$, the endogenous variables $\mathbf{x}(\boldsymbol{\pi})$ and the parameter vector $\boldsymbol{\pi}_1$ will be associated.

When one encounters parameters that upset supermodularity, one typically has to be content with predictions conditional on holding those parameters fixed. However, one need not revert to the traditional mode of testing comparative statics, in which all but one parameter is held fixed. Predictions of association between endogenous and exogenous variables, when subsets of parameters are allowed to vary, constitute an important part of the theory and should not be overlooked. Indeed, regressions that control for parameters that theoretically should not affect covariation can be misguided. An empiricist who finds that the correlation between two variables is of the wrong sign and proceeds to "correct" the sign by adding controls that should not be consequential is throwing out critical evidence. In our application, the theory has specific implications about which controls should be necessary and which should not.

So far we have discussed conditioning on parameters and found that if they are affiliated there is no problem. We will also be interested in statements conditional on endogenous variables. For instance, we may want to know whether firms that employ independent contractors are likely to allow these agents more freedom than firms that use in-house employees. Conditioning on endogenous variables in this way is more problematic, because even if the underlying parameters are affiliated, there is no guarantee that the endogenous variables will be affiliated. However, the following theorem shows that when we condition on a dichotomous endogenous variable things work out well.

A set $\mathbf{A}$ in $\mathbb{R}^k$ is called *increasing* if $\mathbf{z} \in \mathbf{A}$ and $\mathbf{z}' \geq \mathbf{z}$ implies that $\mathbf{z}' \in \mathbf{A}$. The complement of a set $\mathbf{A}$ is denoted $\bar{\mathbf{A}}$.

THEOREM 4: *A random vector* $\mathbf{z}$ *is associated if and only if for every nondecreasing function f, and increasing set* $\mathbf{A}$,

$$(17)\quad E[f(\mathbf{z})|\mathbf{z} \in \mathbf{A}] \geq E[f(\mathbf{z})] \geq E[f(\mathbf{z})|\mathbf{z} \in \bar{\mathbf{A}}].$$

The case of most relevance for us is when $\mathbf{A}$ is defined by conditioning on an ownership assignment λ_i or a permission δ_k. The set of parameter values for which $\lambda_i = 1$ (or $\delta_k = 1$) will be shown to be an increasing set. Similarly, $\{\alpha_i|\alpha_i \geq \bar{\alpha}_i\}$ is an increasing set, and hence, if the vector of commission rates $\boldsymbol{\alpha}$ is associated and f is nondecreasing, $E[f(\boldsymbol{\alpha})|\alpha_i \geq \bar{\alpha}_i] \geq E[f(\boldsymbol{\alpha})]$. Such inequalities are directly testable.

III. Employment vs. Independent Contracting

A. *The Model*

We now turn to a simple application of the general paradigm. It is a stylized model of the situation described in the empirical studies of Anderson and Schmittlein (1984) and Anderson (1985). Anderson (1985) provides a detailed discussion of the nature of the sales job in the electronic-components industry (the subject of the studies), describing the wide range of activities that sales agents have to handle and pointing out the distinctions between independent representatives and in-house sales agents in dealing with these tasks and in their contractual relationship with the firm. Based on her account, we have grouped the agent's responsibilities into four major tasks in our model: (i) direct selling, (ii) investing in future sales to customers, such as prospecting for new customers and building customer goodwill, (iii) nonselling activities, such as helping other agents in the work force, and (iv) selling the products of other manufacturers. We denote the efforts devoted to these activities by t_1 through t_4 and the corresponding returns per unit from these activities by p_1 through p_4, respectively, each strictly positive.

In general, one might expect substantial crossover effects among these activities. For example, a good job coordinating the response to a customer complaint would save costs for the manufacturer but typically also enhance future sales. We sidestep this possibility in a purely formalistic way by selecting the axes in the agent's choice space so that what is called the nonselling activity is that combination of increased coordination activity and reduced other activities that reduces the cost of providing customer service without affecting the expected level of future sales. With nonselling activities so defined, there is no reason to suppose that the agent's marginal cost of conducting nonselling activity is everywhere positive nor that the levels of the various activities are all nonnegative. Accordingly, we impose no such constraints in our analysis.

The only indicator on which direct compensation can be based is the volume of current sales ($X_1 = t_1 + \varepsilon_1$), and this is compensated with a commission αX_1. In addition, by making a commitment not to transfer the sales agent to a new region or by using an independent agent, who cannot be transferred, the firm may effectively transfer ownership of the return stream Y (customer ownership) associated with the effort t_2. The agent then receives λY, where the variable λ is 1. For an employee agent, $\lambda = 0$. The final instrument for controlling agent incentives is the decision to let the agent accept commissions for selling the products of other manufacturers during customer visits ($\delta = 1$). An agent who can accept such commissions receives the additional return Z.[7,8] A contract specifies a salary β, a commission rate α, the transfer (customer-ownership) variable λ, and the 0–1 permission variable δ. Because the salary β is just a transfer and our model is one with quasi-linear preferences, β does not enter into our analysis of efficient contracts.

Let subscripted σ^2's denote the variances of the various error terms ε, and suppose that the errors are independently distributed. Noting that $\delta = \delta^2$, the risk premium borne by the agent under the contract is

$$(18)\quad R(\boldsymbol{\alpha}, \boldsymbol{\lambda}, \boldsymbol{\delta}, \boldsymbol{\sigma}^2) = \tfrac{1}{2} r\left[\alpha^2\sigma_X^2 + \lambda^2\sigma_Y^2 + \delta p_4^2\sigma_Z^2\right].$$

Notice that R does not depend on any of the agent's unobserved actions t—this is a special feature of the linear incentive model.

The agent's certainty equivalent (ACE) corresponding to the effort allocation vector $\mathbf{t}$ under these circumstances is equal to his or her expected earnings minus the cost of effort minus the risk premium:

$$(19)\quad \mathrm{ACE} = \alpha_1 t_1 + \lambda t_2 + \delta t_4 - C(\mathbf{t}) - R(\boldsymbol{\alpha}, \boldsymbol{\lambda}, \boldsymbol{\delta}, \boldsymbol{\sigma}^2).$$

The agent will seek to maximize this by choice of $\mathbf{t}$. Notice that, in this formulation, nonselling activities (t_3) are neither directly measured nor directly rewarded through the three incentive instruments; they are undertaken only to the extent that the agent's marginal cost is negative, which can occur if these activities make selling easier or if the agent simply values customer appreciation.

For notational convenience, let us denote the incentive coefficient in (19) corresponding to activity t_i by a_i and let $\eta_i = 1/\sigma_i^2$ denote the precision of the instrument

[7]An alternative formulation is to set $t_4 = 0$, rather than $\delta = 0$, for an agent who is forbidden to accept commissions. We have investigated one such formulation and found that it leads to similar results.

[8]Since we are interested in *explaining* the attributes of employee agents and independent agents, there is necessarily an ambiguity in our model about what *defines* the two sales methods. Depending on one's viewpoint and the context, one might think that either $\lambda = 1$ or $\delta = 1$ or perhaps $\lambda = \delta = 1$ best defines an independent sales agent. This language problem is not significant for us. What matters is how the instruments get used in the model and whether that usage matches what we observe.

The IRS uses several tests for judging whether a service relationship is one of contracting or of employment, indicating that contractor status is an imprecise term. See footnote 13.

for each activity. Thus, $\eta_1 = 1/\sigma_X^2$, $\eta_2 = 1/\sigma_Y^2$, $\eta_3 = 0$, and $\eta_4 = 1/\sigma_Z^2$. With a slight abuse of notation, we will write $R(\mathbf{a}, \boldsymbol{\eta}) = R(\boldsymbol{\alpha}, \boldsymbol{\lambda}, \boldsymbol{\delta}, \boldsymbol{\sigma}^2)$. Then, assuming an interior optimum, the first-order optimality condition is $a_i = C_i(\mathbf{t})$ for $i = 1, \ldots, 4$ or $\mathbf{a} = \partial \mathbf{C}$. Inverting this system gives the supply function for effort, $\mathbf{t} = \mathbf{t}(\mathbf{a}) = (\partial \mathbf{C})^{-1}\mathbf{a}$.

For the remainder of our analysis, we shall need two extra assumptions:

ASSUMPTION 1: *The different types of inputs are substitutes in the agent's effort supply function:*

$$\partial t_i / \partial a_j < 0 \qquad \text{for all } i \neq j.$$

ASSUMPTION 2: *For all effort inputs i, the agent's effort supply function is supermodular.*

Assumption 1 is an essential part of the intuition, as we will demonstrate shortly. In the differentiable case, Assumption 2 means that $\partial^2 t_i / \partial a_j \partial a_k \geq 0$ for all $j \neq k$ and for all i. Using the effort supply function $(\partial \mathbf{C})^{-1}$, one can express Assumptions 1 and 2 in a seemingly more primitive form, involving only derivatives (first, second, and third) of the cost function. However, doing so would be meaningless, because it would in no way help us to assess the plausibility of the assumptions or test their validity empirically. By contrast, the assumptions above could in principle be tested either experimentally, or by asking the agent the two hypothetical questions: how would the agent respond to an increase in the incentive intensity of an activity (Assumption 1); and how would his supply responsiveness change in that case (Assumption 2)? In the context of our application, it seems quite plausible that Assumption 1 holds. The plausibility of Assumption 2 is harder to assess. Some guidance can be found from the special case in which the cost function is quadratic. In that case, the effort supply function is linear in the vector $\mathbf{a}$ of incentive coefficients, so Assumption 2 is satisfied trivially.

Replacing the incentive coefficients in (18) by the corresponding expressions in terms of $\mathbf{a}$, the total certainty equivalent of the principal and agent is

$$\begin{aligned}(20)\quad &\mathrm{TCE}(\mathbf{a}, \mathbf{p}, \boldsymbol{\eta}) \\ &= \sum_{i=1}^{4} p_i t_i(\mathbf{a}) - C(\mathbf{t}(\mathbf{a})) - R(\mathbf{a}, \boldsymbol{\eta}) \\ &= \sum_{i=1}^{4} p_i t_i(\mathbf{a}) - C(\mathbf{t}(\mathbf{a})) \\ &\quad - \tfrac{1}{2} r \left[a_1^2/\eta_1 + a_2^2/\eta_2 + a_4^2/\eta_4 \right].\end{aligned}$$

PROPOSITION 1: *Suppose Assumptions 1 and 2 are satisfied, that a_3 is fixed at zero, and that p_1, p_2, and p_4 are fixed. Then the function $\mathrm{TCE}(a_1, a_2, a_4, \eta_1, \eta_2, \eta_4, -p_3)$ is supermodular on the domain where $\boldsymbol{\eta} \geq \mathbf{0}$, $p_3 \geq 0$, and $a_j \leq p_j$ for $j = 1, 2, 4$. For the case where C is quadratic, the TCE function is supermodular on the larger domain obtained by dropping the restrictions that $p_j \geq a_j$.*

PROOF:

Recall that the sum of supermodular functions is supermodular. The supermodularity of the $-R(\mathbf{a}, \boldsymbol{\eta})$ term can be verified by calculating all the mixed partial derivatives. The nonzero mixed partial derivatives involving $-p_3$ are $\partial^2 \mathrm{TCE} / \partial(-p_3)\partial a_i = -\partial t_3 / \partial a_i > 0$ for $i = 1, 2, 4$. For the remaining terms, since $a_j = C_j(\mathbf{t}(\mathbf{a}))$ for all $\mathbf{a}$, one may calculate that

$$\begin{aligned}(21)\quad &(\partial / \partial a_j) \sum_{i=1}^{4} p_i t_i(\mathbf{a}) - C(\mathbf{t}(\mathbf{a})) \\ &= \sum_{i=1}^{4} \partial t_i / \partial a_j \left[p_i - C_i(\mathbf{t}(\mathbf{a})) \right] \\ &= \sum_{i=1}^{4} \partial t_i / \partial a_j (p_i - a_i)\end{aligned}$$

and hence, for $j \neq k$:

$$\begin{aligned}(22)\quad &\left(\partial^2 / \partial a_j \partial a_k\right) \sum_{i=1}^{4} p_i t_i(\mathbf{a}) - C(\mathbf{t}(\mathbf{a})) \\ &= -\partial t_k / \partial a_j + \sum_{i=1}^{4} \partial^2 t_i / \partial a_j \partial a_k (p_i - a_i).\end{aligned}$$

Assumption 1 ensures that the first term is positive. Assumption 2 ensures that the second term is positive provided that $p_i - a_i \geq 0$. In case C is quadratic, $\partial^2 t_i / \partial a_j \partial a_k \equiv 0$ for all i, j, and k, regardless of **a**.

Remark: The domain restrictions in Proposition 1 do not limit the applicability of the theory. The precision vector $\boldsymbol{\eta}$ is always positive. Also, because the components of **t** are substitutes, if any $a_i / p_i > 1$, then reducing all such ratios to unity improves the resource allocation (the "concertina theorem"). In a principal–agent model, reducing them has the additional benefit of reducing the risk borne by the agent. So long as setting $a_i = p_i$ for $i = 1, 2, 4$ is in the feasible set, the optimum always lies in the set where TCE is supermodular.

Proposition 1 expresses the fundamental idea behind our theory. Under particular assumptions which assert that the substitution effect ($\partial t_i / \partial a_j < 0$) is either reinforced by the other effects or dominates them, the various kinds of incentives are complementary over the relevant range. By definition, this means that the opportunity cost of raising the incentives for one activity decreases with an increase in the incentives for any of the other activities.

The intuition behind this complementarity is transparent from the proof and merits a verbal interpretation. Since a maximizing agent equates his personal marginal benefit a_i to the marginal cost, the net surplus created by an additional unit of effort devoted to activity i is $p_i - a_i$. Because the activities are substitutes (Assumption 1), the opportunity cost of increasing a_j in terms of reduced performance in some other activity $i \neq j$, is equal to $(p_i - a_i)$ times the reduction in the level of activity i [equation (21)]. If the slope of the agent's response function is a constant, as when costs are quadratic, or if it changes in the right way (Assumption 2), this opportunity cost is reduced when a_i is larger. Therefore, an exogenous increase in a_i leads to an increase in all the other a_j.

That the incentive intensities are complements suggests the intuition that they should move together. When one goes up, the others should too. As we stressed in the Introduction, however, this intuition is incomplete. All the instruments are endogenous variables. Their movements will depend on how the exogenous parameters enter into the problem. Theorem 1 asserts that whenever an exogenous parameter is complementary with *all* the instruments (an increase in that parameter does not reduce any instrument's marginal return) and provided that the instruments themselves are complementary, then the instruments will move together in response to a change in that parameter.

As one can infer from Proposition 1, not all parameters are complementary with all the incentive instruments in our application. It is instructive to study one such case, assuming for simplicity that the cost function is quadratic. Suppose the marginal value of activity 1, p_1, increases. From equation (21) we have $\partial^2 \mathrm{TCE} / \partial p_i \, \partial a_j = \partial t_i / \partial a_j$, which is negative for $j \neq i$, but positive for $j = i$. Consequently, the marginal return p_1 is complementary with a_1, while $-p_1$ is complementary with the other incentives; so TCE is not supermodular in this parameter, and one cannot conclude that the instruments will move together. Indeed, the cross-partial indicates that the first-order response to an increase in p_1 will be to increase the incentives for activity 1 and to *decrease* the incentives for the other activities. The economic intuition is simple. The level of activity 1 should go up, since its marginal value has increased. This can be accomplished either by rewarding this activity or by discouraging other activities. In this case, the optimal response is to do some of both. The conclusion is that the incentive instruments may move in opposite directions in response to changes in p_1 (and p_2 and p_4).

Why does $-p_3$ make the list of parameters that are complementary in Proposition 1? Because there is no incentive instrument a_3 for rewarding this activity. All the mixed partial derivatives with respect to the other instruments are negative, just as we found with p_1. By changing the sign on p_3 these all become positive. In economic terms, the

only way to provide incentives for nonselling activities, is by *not* providing as strong incentives for the other activities. Thus, one reason for integrating the sales force into the firm could be to reduce the incentives for activities that compete with nonselling activities. Weaker work incentives for employees are often suggested as a major cost of integration. In a multitask environment, low-powered incentives can instead be a major benefit in cases where market incentives are too strong.[9]

We have so far assumed that the precision with which direct selling efforts can be measured (η_1) is exogenous. One would normally think that this precision can be varied by increasing or decreasing the intensity of monitoring or by abandoning team selling, for instance. A minor change in the model will accommodate that. Let $K(\eta_1;\gamma)$ represent the cost of improving information about sales performance, where K is nondecreasing and γ is an exogenous cost parameter. The objective function for the contracting problem is then

$$T = \mathrm{TCE}(\mathbf{a},\boldsymbol{\eta},\mathbf{p}) - K(\eta_1;\gamma).$$

If $-K$ is supermodular, so that higher γ corresponds to a lower marginal cost of monitoring performance, it follows that the total surplus function T is supermodular in the arguments $(\mathbf{a},\boldsymbol{\eta},-p_3,\gamma)$.

We can now record the main comparative-statics predictions of our theory.

PROPOSITION 2: *All else equal, an exogenous reduction in p_3 (the "importance of nonselling activities"), an exogenous reduction in the variance of returns on future sales or of outside selling options (increases in η_2 or η_4), or an exogenous decrease in the marginal cost of sales performance measurement (an increase in γ) will lead to increases in the optimal values of α, λ, δ, and η_1, that is, higher commissions, more customer ownership by the sales agent, more freedom to sell for other manufacturers, and more emphasis on direct sales measurement.*

The incentive instruments considered in our formal model do not, of course, exhaust the options open to the firm. An important excluded instrument for the firm is that of job design. The selling and nonselling activities could be separated and assigned to different personnel. That way, incentives for selling could be increased without causing a substitution of effort away from nonselling activities. For an analysis of job design along these lines, see Holmstrom and Milgrom (1991). One could also introduce economies of scope from selling products of other manufacturers. If the extra returns to scope can be represented by an additional term in the objective function, then increased economies of scope will have the same effects as increases in the parameters listed in Proposition 2. If they enter in a different way, one can use the model to analyze in what ways this may be empirically problematic.

B. *Covariations in Cross-Sectional Data*

The preceding discussion certainly is suggestive of the type of covariations in instruments and clustering of incentive attributes that we have been looking for. But as Section II made clear, individual comparative-statics results do not guarantee positive correlations when all the parameters are varied simultaneously or when additional heterogeneity is introduced; that is, systematic patterns may not be evident in aggregate data. In this subsection we will take a look at what types of data can be pooled and what controls need to be introduced to enable us to transform the comparative-statics results of the theory into statistical predictions regarding covariations.

Combining Proposition 2 and Theorem 3, we have the following proposition.

PROPOSITION 3: *Assume that the parameter vector (p_1,p_2,p_4) is fixed. Suppose further that the parameter vector $\boldsymbol{\nu} = (\boldsymbol{\eta},-p_3,\gamma)$ is associated. Then the vector*

[9]See Holmstrom and Tirole (1991) for a further articulation of this point in the context of transfer pricing.

$(\boldsymbol{\alpha}, \boldsymbol{\delta}, \boldsymbol{\lambda}, \boldsymbol{\eta}, -p_3, \gamma)$ *is associated. The same conclusion holds if all the variables are measured with additive errors that are independent of each other and of the exogenous parameters.*

Remark: The assumption that $\boldsymbol{\nu}$ is associated should not be confused with the maintained hypothesis that the stochastic variables ε in the model are independent. Association here refers to the characteristics of the population of agency relationships. While it is possible that the agency relationships are distributed independently along the three dimensions of $\boldsymbol{\nu}$, it is equally plausible that the factors that increase the variance of future sales returns, for instance, also increase the variance of current sales, implying an association between η_1 and η_2.

Proposition 3 provides an example of the type of statistical prediction that the paper's opening paragraph alluded to. According to Proposition 3, high commission rates α, independent contracting ($\lambda = 1$) and agent freedom ($\delta = 1$) should tend to be found together in a cross section of sales regions, and furthermore, they should all tend to rise with increases in the parameters in $\boldsymbol{\nu}$. In particular, all pairs of variables and parameters should exhibit a nonnegative covariance. We wish to stress that, by virtue of Theorem 2, even if we cannot observe all the underlying parameters in $\boldsymbol{\nu}$, the observed subvector of variables is still associated. We can therefore test Proposition 3 in censored data. This lends an advantage to covariation tests relative to tests in which all the parameters in $\boldsymbol{\nu}$ must be observed and included in a regression. In addition, of course, we are not specifying any functional forms, and the restrictions on measurement errors are quite weak.

Testing association may be less natural than testing predictions about conditional expectations. A number of predictions on conditional expectations can be presented. The following proposition focuses on differences between independent contracting and employment.

PROPOSITION 4: *Under the assumptions of Proposition 3 and assuming in addition that the vector of parameters $\boldsymbol{\nu}$ is affiliated:*

(a) $\Pr[\lambda^* = 0 | \gamma]$ *is (weakly) decreasing in* γ,
(b) $\Pr[\lambda^* = 0 | p_3]$ *is (weakly) increasing in* p_3,
(c) $\Pr[\delta^* = 1 | \lambda^* = 1] \geq \Pr[\delta^* = 1 | \lambda^* = 0]$,
(d) $\mathrm{E}[\alpha^* | \lambda^* = 1] \geq \mathrm{E}[\alpha^* | \lambda^* = 0]$.

The proof follows from the supermodularity of T and the definition of affiliation.[10]

According to the proposition, it will be more likely that a sales region is served by an in-house agent, (a) the higher is the marginal cost of evaluating performance and (b) the more important are nonselling activities.[11] Independent sales representatives will on average, (c) enjoy more freedom (right to private business) and (d) work on higher commissions.

The most troublesome aspect of our theory is that the marginal-return parameters (p_1, p_2, p_4) must be assumed constant in the population under study. The same goes for any parameters of the cost function C, which we have simply assumed is the same across agencies. If these potentially problematic parameters can be observed, and if the full vector of parameters is affiliated, then Proposition 3 continues to hold on any subset of the data for which the cost and marginal-return parameters are the same (i.e., we can replace the covariations of Proposition 3 with the corresponding condi-

[10] If the parameter vector $\boldsymbol{\nu}$ is associated but not affiliated, then we can replace (a) and (b) with the weaker statements:

(a′) $\Pr[\lambda^* = 0 | \gamma \geq \bar{\gamma}] \geq \Pr[\lambda^* = 0 | \gamma < \bar{\gamma}]$, for all $\bar{\gamma}$;
(b′) $E[\lambda^* = 0 | p_3 < \bar{p}_3] \leq E[\lambda^* = 0 | p_3 \geq \bar{p}_3]$, for all $\bar{p}_3$.

Also, in (a′) we could have replaced the conditioning variable γ with the precision η_1 of the performance evaluation.

[11] This conclusion does not depend on defining a sales employee as one who does not have customer ownership ($\lambda = 0$). Parts (a) and (b) would be equally true with δ^* replacing λ^*, that is, by defining sales employees as agents who cannot peddle the products of other manufacturers.

tional covariations). Of course, the number of observations in such a subset may be very small, but one can remedy this by performing a joint test of the positivity of all conditional covariances between two variables.

Controlling in this way for variations in marginal returns and costs may in practice be hard, because the relevant parameters can be difficult to observe. If appropriate controls cannot be found, the best we can do is to offer predictions about correlations between a single exogenous variable and a single endogenous variable; in other words, statistical tests of one-parameter comparative-statics results. As an illustration we offer the following proposition.

PROPOSITION 5: *Suppose the marginal-cost parameter γ is independent of all the other parameters of the model. Then λ^* and γ are associated and* $\Pr[\lambda^* = 0|\gamma]$ *is nonincreasing in* γ.

PROOF:

According to Proposition 2, λ^* is nonincreasing in γ. The results then follow from Theorem 3.

We could similarly have concluded that independent agents are paid on average higher commission when sales performance is easier to measure, assuming again that the marginal-cost parameter is independent of the other parameters.

C. *Empirical Evidence*

The predictions of our model fit well with the findings in Anderson and Schmittlein (1984) and Anderson (1985). The main variables explaining the choice between in-house sales agents and independent representatives in a sales region were found to be the "difficulty of evaluating performance" and the "importance of nonselling activities." Both of these increased the probability of direct selling, that is, the use of in-house sales agents, as parts (a) and (b) of Proposition 4 predict. Independent sales representatives worked 100 percent on commission. Though the papers do not report how high these commissions were, one can infer that, since the representatives were paid no salary, the commissions were substantially higher than for sales employees. (Anderson [1985] reports that sales agents were paid by salary "with a small bonus or commission in addition.") It is also evident that sales agents could not sell products of other manufacturers. There is no statement about independent representatives selling exclusively the products of one firm, but it is clear that some did not. Thus the probability of private business is higher for independent agents as in part (c) of Proposition 4. Finally, regarding part (d), Anderson (1985) notes that it was common practice to rotate sales employees from one region to another, reducing such an agent's returns from customer loyalty. Independent agents cannot be rotated and were in fact reported to stay for long periods in the same region.

What about the problems with variations in marginal benefits (B') and marginal costs (C')? The studies controlled for one relevant variable, namely, customer loyalty, which also influenced the choice between independents and employees, but much less than the two variables mentioned above. No variables were used to control for other marginal benefits or marginal costs. However, if the complementarities among incentive instruments are strong, the required discrete and significant shifts in λ and δ may have overwhelmed the contrary direct effects from shifts in marginal benefits and costs.

Though we have tailored our model to fit the context studied by Anderson and Schmittlein (1984) and Anderson (1985), there are other studies that provide indirect support. In two separate studies, Andrea Shepard (1993) and Margaret Slade (1992) have found that the probability that a gas station will be independently owned, rather than owned by an oil company, is higher when the station provides hard-to-monitor extra services. Moreover, independent stations retain a larger fraction of revenue from gasoline sales. Both of these facts accord with the predictions of our theory. Rajiv Lal and V. Srinivasan (1994) report

that the fraction of compensation paid in the form of commissions rather than salary is higher for independent sales agents, affirming another prediction of our theory.

Another example in which the incentives of independent contractors are significantly different from those of employees arises in franchising.[12] Many companies, including McDonald's, for instance, use in-house managers to operate some units and franchisees to operate others (James Brickley and Frederick Dark, 1987). Franchisees typically receive very strong incentives; in McDonald's case they pay 5–10 percent royalties on sales, implying an effective commission rate of more than 90 percent (Alan Krueger, 1991). They also get to keep the value added from selling the unit. By contrast, in-house managers typically receive little or no explicit incentive pay and of course, they receive no portion of the increase in value of the franchise outlet. Thus, there is a strong correlation between incentives for enhancing the market value of the unit and incentives for immediate sales, as our theory suggests.

However, other parts of the franchising evidence do not match our predictions. First, franchisees remain on a tight leash: they cannot sell the products of other firms, and many of their operating rules are determined by the franchisor (one could not tell from visiting a McDonald's whether it is franchised or company-owned). More importantly, while Brickley and Dark (1987) also find that monitoring costs best explain the choice between franchising and company ownership, they find that *the relationship is opposite* to that predicted by our model: the harder it is to monitor a unit (as proxied by the distance to headquarters), the *more likely* it is that the unit is franchised.

How damaging is this for our theory? We believe that the problem lies not in the logic of our theory, but in important differences between the franchising and industrial selling contexts. One difference is the concern for a shared brand-name reputation in franchising, which creates a potential free-rider problem among franchisees. This explains the tight constraints and significant monitoring imposed on franchisees. Also, in the fast-food industry, once one has controlled for food and service quality as required by the shared brand name, the other activities that increase current sales and future value are largely the same, which makes ownership and commissions substitutable rather than complementary incentive instruments. In our model it is essential that ownership and commissions influence competing dimensions of the agent's portfolio of activities, creating an undesirable substitution effect when just one of the incentives is strengthened.

More important than speculating about the possible reasons for the contradictory evidence is the observation that Brickley and Dark (1987) derive a reverse hypothesis regarding the effects of monitoring costs from the same general paradigm that we have used, namely, agency theory. What our analysis shows is that their hypothesis requires assumptions beyond the mere observation that there is an agency problem: a general appeal to "agency costs" is too vague a basis for making testable predictions. Comparing the evidence of industrial selling to that on franchising thus leads to an important lesson: it is essential to tailor agency models to the empirical context they are intended to illuminate. Our modeling framework offers a convenient laboratory for developing such models and for identifying the critical assumptions.

IV. Conclusion

In this paper we have sought to explain why the attributes of an employment relationship differ in so many ways from the attributes of a contractor relationship. The fact that employment and contracting are multifaceted relationships, each characterized by its own distinct set of attributes, ranks as one of the most significant regularities to be explained by a theory of the firm. While it may be tempting for empirical rea-

[12] We are grateful to James Brickley for extensive discussions on this subject.

sons to try to classify relationships based on a single attribute, more complicated criteria are often used in reality. For example, the U.S. Internal Revenue Service uses all of the attributes listed in our model in deciding whether a relationship is one of employment rather than contracting.[13]

A framework that acknowledges the multi-attribute nature of the employment relationship makes it possible to apply modern extensions of price theory to create a more complete and natural explanation of a firm's organizational decisions. Our theory treats the attributes of a contracting relationship just the same as if they were inputs to production. It then explores how changes in a parameter affecting one attribute affect choices of other attributes. Just as one would not try to explain a firm's demand for an input solely in terms of its input price, ignoring technology and the prices of substitutes and complements, one should not try to explain any one particular attribute of the employment relationship without reference to substitute and complement attributes. In our application, an increase in the cost of measuring sales performance acts like an increased input price, leading to the substitution of salary for commissions and to a complementary increase in limitations on outside activities and the frequency of job rotation. Similarly, an increase in the importance of outside activities acts like a change in technology in the classical theory of the firm, leading to changes in the "inputs" (incentive instruments) that increase the effort devoted to nonselling activities.

Many of the ideas that we use in our analysis are closely linked to optimal taxation and to the theory of regulation. In optimal tax theory, the distortions induced by a tax on one commodity can often be attenuated by an accompanying tax on its substitutes or by reduced taxes on its complements. Our idea of balancing incentives among activities within firms has a similar logic. More generally, to the extent that a firm restricts its employees' rights to trade with outsiders, the firm can be analyzed as an economy in its own right, and the principles of optimal taxation and regulation can be applied to its internal transactions.

One insight that emerges from this sort of regulatory analogy is that an optimal incentive structure may require the elimination or muting of incentives which in a market relationship would be too strong. Thus, the use of low-powered incentives within the firm, although sometimes lamented as one of the major disadvantages of internal organization, is also an important vehicle for inspiring cooperation and coordination. In our model, this was illustrated by the result that when "nonselling" activities become more important, the optimal adjustment is to reduce commission rates.

Our model also shows how the effectiveness of low-powered incentives within the firm may be enhanced by simultaneously placing constraints on the employee's freedom to act. Like low-powered incentives, bureaucratic constraints have frequently come under attack as being costly and inefficient. It is notable that the specific criticisms we hear today—that bureaucracy impedes innovation, that it is hopelessly slow, that it limits communication, that it prevents personal growth, and so on—are all problems that have been with bureaucracy from its inception. But we are hardly the first generation to recognize how costly bureaucracy can be. The natural economic hypothesis is that bureaucratic constraints can serve a purpose. To the extent that there is a need for reduced bureaucracy today, it may be because the environment has changed. Our modeling approach identifies some of the environmental changes that could account for the current trend toward greater worker responsibility ("empowerment"), as well as the implications

[13]The IRS "Independent Contractor Checklist" asks, among other things, whether one promotes one's business as "available to perform services for the general public or other businesses"; whether one "provides the necessary tools, supplies, equipment, and materials to perform the work"; whether one "maintains control over the progress of one's work and the methods of accomplishing it"; and whether one is "paid based on completion of the job rather than by the hour." Negative answers tend to disqualify a person from claiming independent contractor status.

such a change is likely to have on the other components of an incentive system.

Although our formal analysis of employment versus independent contracting incorporates several attributes of these relationships, its focus is still quite narrow. We have treated only the case of a single agent with a fixed set of tasks. Important new issues arise when one considers how several workers should be organized to maximize total surplus and how their contractual relationships should be designed. We have touched on some of these multi-agent issues in earlier work (Holmstrom and Milgrom, 1990, 1991), in which we studied the surplus-maximizing degree of teamwork and the optimal grouping of tasks into jobs. Even after jobs are delineated, however, the allocation of jobs among firms still needs to be analyzed. In the case of sales agents studied in this paper, independent agents often work not for themselves, but for other firms. The question is: why?

We have also focused on a limited set of incentive instruments, and one may well ask whether other instruments, such as promotions, implicit incentives, or simple verbal encouragement are not equally important. They probably are, and they would have to be included in any purportedly complete study. As new instruments are added, however, the key according to our theory is to evaluate them not in isolation, but as part of a coherent incentive system.

REFERENCES

Alchian, Armen and Demsetz, Harold. "Production, Information Costs, and Economic Organization." *American Economic Review*, December 1972, *62*(5), pp. 777–95.

Anderson, Erin. "The Salesperson as Outside Agent or Employee: A Transaction Cost Analysis." *Marketing Science*, Summer 1985, *4*(3), pp. 234–54.

Anderson, Erin and Schmittlein, David. "Integration of the Sales Force: An Empirical Examination." *Rand Journal of Economics*, Autumn 1984, *15*(3), pp. 385–95.

Barlow, R.; Bartholemew, D.; Bremner, J. and Brunk, H. *Statistical inference under order restrictions: The theory and application of isotonic regression*. New York: Wiley, 1972.

Brickley, James A. and Dark, Frederick H. "The Choice of Organizational Form: The Case of Franchising." *Journal of Financial Economics*, June 1987, *18*(2), pp. 401–20.

Coase, Ronald. "The Nature of the Firm." *Economica*, November 1937, *4*(4), pp. 386–405; reprinted in *The firm, the market and the law*. Chicago: University of Chicago Press, 1988, pp. 33–55.

Grossman, Sanford and Hart, Oliver. "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration." *Journal of Political Economy*, August 1986, *94*(4), pp. 691–719.

Hart, Oliver and Moore, John. "Property Rights and the Nature of the Firm." *Journal of Political Economy*, December 1990, *98*(6), pp. 1119–58.

Holmstrom, Bengt. "Moral Hazard in Teams." *Bell Journal of Economics*, Autumn 1982, *13*(2), pp. 324–40.

Holmstrom, Bengt and Milgrom, Paul. "Aggregation and Linearity in the Provision of Intertemporal Incentives." *Econometrica*, March 1987, *55*(2), pp. 303–28.

_____. "Regulating Trade Among Agents." *Journal of Institutional and Theoretical Economics*, March 1990, *146*(1), pp. 85–105.

_____. "Multitask Principal–Agent Analyses: Incentive Contracts, Asset Ownership and Job Design." *Journal of Law, Economics and Organization*, 1991, *7*, Special Issue, pp. 24–52.

Holmstrom, Bengt and Tirole, Jean. "Transfer Pricing and Organizational Form." *Journal of Law, Economics and Organization*, Fall 1991, *7*(2), pp. 201–28.

_____. "Market Liquidity and Performance Measurement." *Journal of Political Economy*, August 1993, *101*(4), pp. 678–709.

Klein, Benjamin; Crawford, Robert and Alchian, Armen. "Vertical Integration, Appropriable Rents, and the Competitive Contracting Process." *Journal of Law and Economics*, October 1978, *21*(2), pp. 297–326.

Krueger, Alan. "Ownership, Agency, and Wages: An Examination of Franchising in

the Fast Food Industry." *Quarterly Journal of Economics*, Fall 1991, *106*(1), pp. 75–101.

Lai, Rajiv and Srinivasan, V. "Compensation Plans for Single and Multiproduct Salesforces: An Application of the Holmstrom-Milgrom Model." *Management Science*, 1994 (forthcoming).

Milgrom, Paul and Roberts, John. "The Economics of Modern Manufacturing: Technology, Strategy, and Organization." *American Economic Review*, June 1990, *80*(3), pp. 511–28.

Milgrom, Paul and Shannon, Christina. "Monotone Comparative Statics." *Econometrica*, January 1994, *62*(2), pp. 157–80.

Milgrom, Paul and Weber, Robert. "A Theory of Auctions and Competitive Bidding." *Econometrica*, September 1982, *50*(5), pp. 1089–1122.

Shepard, Andrea. "Contractual Form, Retail Price, and Asset Characteristics in Gasoline Retailing." *Rand Journal of Economics*, Spring 1993, *24*(1), pp. 58–77.

Simon, Herbert. "A Formal Theory of the Employment Relationship." *Econometrica*, July 1951, *19*(3), pp. 293–305.

Slade, Margaret. "Multitask Agency and Organizational Form: An Empirical Assessment." Working Paper, University of British Columbia, 1992.

Topkis, Donald. "Minimizing a Submodular Function on a Lattice." *Operations Research*, March-April 1978, *26*(2), pp. 305–21.

Williamson, Oliver. *The economic institutions of capitalism: Firms, markets, relational contracting*. New York: Free Press, 1985.