

The theory of contracts

Oliver Hart and Bengt Holmström

Introduction

The past decade has witnessed a growing interest in contract theories of various kinds. This development is partly a reaction to our rather thorough understanding of the standard theory of perfect competition under complete markets, but more importantly to the resulting realization that this paradigm is insufficient to accommodate a number of important economic phenomena. Studying in more detail the process of contracting – particularly its hazards and imperfections – is a natural way to enrich and amend the idealized competitive model in an attempt to fit the evidence better. At present it is the major alternative to models of imperfect competition; we will comment on its comparative advantage below.

In one sense, contracts provide the foundation for a large part of economic analysis. Any trade – as a *quid pro quo* – must be mediated by some form of contract, whether it be explicit or implicit. In the case of spot trades, however, where the two sides of the transaction occur almost simultaneously, the contractual element is usually downplayed, presumably because it is regarded as trivial (although this need not be the case; see Section 3). In recent years, economists have become much more interested in long-term relationships where a considerable amount of time may elapse between the *quid* and the *quo*. In these circumstances, a contract becomes an essential part of the trading relationship.

Of course, long-term contracts are not new in economics. Contingent commodity trades of the Arrow–Debreu type are examples par excellence of such contracts. What does seem new is the analysis of contracts written

We would like to thank Jonathan Feinstein, Paul Joskow, John Moore, Sherwin Rosen, Jean Tirole, and Andy Weiss for comments on an earlier draft. Financial support from NSF and the Sloan Foundation is gratefully acknowledged.

by and covering a small number of people. That is, there has been a move away from the impersonal Arrow–Debreu market setting where people make trades “with the market,” to a situation where firm A and firm B, or firm C and union D, write a long-term contract. This departure is not without economic significance. Williamson (1985), in particular, has stressed the importance of situations where a small number of parties make investments which are to some extent relationship-specific; that is, once made, they have a much higher value inside the relationship than outside. Given this “lock-in” effect, each party will have some monopoly power *ex post*, although there may be plenty of competition *ex ante* before investments are sunk. Since the parties cannot rely on the market once their relationship is underway, the obvious way for them to regulate (and divide the gains from) trade is via a long-term contract. Until the advent of contract theory, economists did not have the tools to analyze *ex ante* competitive, *ex post* noncompetitive relationships of this type via formal models.

Research on contracts has progressed along several different lines, each with its own particular interests. It may be useful to begin by mentioning some of these directions before outlining the subjects we will concentrate on in this chapter.

One strand of the literature has focused on the internal organization of the firm, viewing the firm itself as a response to failures in the price system. Questions of interest include structuring incentives for members of the firm, allocating decision authority, and choosing decision rules to be implemented by suitable reward structures. Of course, the objective is partly to gain insight into organization theory as such. But perhaps more importantly, one is interested in knowing whether organization theory matters in the aggregate – that is, to what extent the conduct of firms will be different from the assumed profit-maximizing behavior; and, if it differs, what ramifications follow for market outcomes and overall allocations in the economy.

Another prominent line of research has explored the workings of the labor market. A plausible hypothesis is that contingent claims for labor services are limited for reasons of opportunism. This invites innovation of other types of contracts that can be used as substitutes. The research has centered on the structure of optimal bilateral labor contracts (under various assumptions about enforcement opportunities), on the properties contractual equilibria will have, and in particular on whether these equilibria will exhibit the commonly claimed inefficiencies associated with real-world adjustments in employment.

Inspired by the possibility that long-term contracts may embody price and wage sluggishness, a related body of work has explored their macro-

economic implications [see, e.g., Fischer (1977) and Taylor (1980)]. Unlike most contract analysis this literature has taken the form of contracts as given, typically with nominal wage and price rigidities. This is not as satisfactory as working from first principles, but it has made policy analysis quite tractable.

Financial markets offer another arena of substantial potential for contract theoretic studies that is beginning to be recognized. The importance of limited contracting for the emergence of financial services and institutions has been suggested by D. Diamond (1984), Gale and Hellwig (1985), and Townsend (1980). This line of research also offers prospects for a careful modeling of the role of money and the conduct of monetary policy [see Townsend, Chapter 11 this volume, and D. Diamond (1985)].

As the field is progressing, it becomes harder to place models in specific categories. Initially, models of organizational design ignored market forces, or at least treated them in a very primitive fashion. In contrast, the theory of labor contracts started out without consideration for organizational incentives. More recent models, however, treat both incentive and market issues concurrently. Such crossbreeding is fruitful, but it makes the task of organizing this chapter much harder. Since we have been unable to come up with a natural classification that would avoid this problem, we will employ an outline that follows the historical progress rather closely.

We begin in Section 1 with agency theory as a representative paradigm for the organization-theoretic aspects of contracting. From there we go on to labor contracting (Section 2). Finally, we turn to incomplete contracts and the aforementioned lock-in effects (Section 3). This work, representing more recent methodological trends in contract research, has not advanced very far yet, and our discussion will be correspondingly more tentative in nature.

Needless to say, we will not attempt a comprehensive survey of the large number of contractual models that have appeared to date. Some subjects (e.g., models relating contracts to macroeconomic policy) are left out entirely. So are models of financial contracting. Our intention has been to be selective and critical rather than comprehensive. Although we allow ourselves a rather opinionated tone, we trust this chapter still gives a good idea of the general nature of the ongoing research and a reasonably fair assessment of its main contributions.

Despite our selective approach, the chapter has grown very long. In order that it may be more readily digestible, we have written it so that the three parts can be read essentially independently; each part has a concluding section that sums up its major points.

A word about methodology

Most contract theories are based on the assumption that the parties at some initial date (say, zero) design a Pareto optimal (for them) long-term contract. Optimality is not to be understood in a first-best sense, but rather in a constrained or second-best sense. Indeed, informational and other restrictions that force the contract to be second-best are at the heart of the analysis – without them one would quickly be back in the standard Arrow–Debreu paradigm where contractual form is inessential. Because informational constraints will play a particularly important role in the ensuing discussion, let us note right away that throughout we will restrict attention to cases in which informational asymmetries arise only subsequent to contracting. In the typical language of the literature, we will not consider adverse selection models.

The design of a Pareto optimal contract proceeds by maximizing one party's expected utility subject to the other party (or parties) receiving a minimum (reservation) expected utility level. Which party's utility level is taken as a constraint does not usually matter, because most analyses are partial equilibrium. When there is perfect competition *ex ante*, this reservation utility can be interpreted as that party's date-zero opportunity cost determined in the date-zero market for contracts. When *ex ante* competition is imperfect, the parties will presumably bargain over the *ex ante* surplus from the relationship, and so the reservation expected utility levels become endogenous.

The literature has often been cavalier about the determinants of the reservation utility, because valuable insights have emerged from the general characteristics of Pareto optimality alone. On the other hand, the fact that market forces reduce to simple constraints on expected utilities greatly facilitates equilibrium analysis. Equilibration in expected utilities is usually trivial. This gives the contractual approach its main methodological advantage relative to models of imperfect competition, for instance. The analytical core of contract theory is an optimization problem, whereas in imperfect competition it is an equilibrium problem. Methods for solving optimization exercises are substantially more advanced than methods for solving equilibrium problems.

Of course, substituting an optimization analysis for an equilibrium analysis is not always economically meaningful (for instance, we are not implying that imperfect competition should be studied in this way). Indeed, the economic credibility of the contractual approach may be called into question when, as often happens, optimal contracts become monstrous state-contingent prescriptions. How are such contracts written and enforced?

Three responses to this question can be offered. The first one is to appeal to the powers of the judicial system and its ability to enforce certain explicitly agreed-upon contractual terms. The assumption is that sufficient penalties, either pecuniary or nonpecuniary, will be imposed for breach and hence rational parties will not breach. This assumption makes a model internally consistent, but is unsatisfactory on two accounts. It maintains an artificial dichotomy between those contractual provisions that are assumed to be infinitely costly to enforce and those that are assumed to be completely costless to enforce. Also, it often predicts (by assumption) explicit terms that are much more complex than those we observe and in that sense is no answer to what prodded the enforcement question above.

The second response is a pragmatic one: One could argue that qualitative and aggregate features, rather than contractual detail, are the relevant ones for judging the success of a model. In support of this view one can allude to the implicit nature of contracts in the real world; in other words, suggest that equilibrium outcomes in the real world mimic optimal, complex state-contingent contracts despite the relative simplicity of the explicit agreements we observe. The difficulty with this response is that we do not understand well how implicit contracts of this type are sustained as equilibrium phenomena.

Ideally, one would like to know what determines the division between explicit and implicit enforcement of a contract. This leads to the third approach, which is to confront the enforcement issue explicitly by including realistic legal penalties for breach as well as indirect costs that affect equilibrium behavior, for instance through reputational concerns. Whereas much of the extant literature rests on a combination of the first two responses to the enforcement issue, the present trend is toward the more ambitious, but also more satisfactory, third approach. This will be discussed at some length in Section 3.

1 Agency models

1.1 Introduction

Agency relationships are ubiquitous in economic life. Wherever there are gains to specialization there is likely to arise a relationship in which agents act on behalf of a principal, because of comparative advantage. Examples abound: workers supplying labor to a firm, managers acting on behalf of owners, doctors serving patients, lawyers advising clients. The economic value of decision making made on behalf of someone else would easily seem to match the value of individual consumption decisions. In this light

the attention paid to agency problems has been relatively slight. Moreover, there are some less obvious instances of the same formal agency structure – the government taxing its citizens, the monopolist price-discriminating customers, the regulator controlling firms – all of which are substantial problems in their own right.

If agents could costlessly be induced to internalize the principal's objectives, there would be little reason to study agency. Things become interesting only when objectives cannot be automatically aligned. So what is it that prevents inexpensive alignment? The most plausible and commonly offered reason is asymmetric information, which of course ties closely to the source of agency: returns to specialization. The sincerity of a worker's labor input is often hard to verify, leading to problems with shirking. Informational expertise permits managers to pursue goals of their own such as enhanced social status or improved career opportunities. Private information about individual characteristics causes problems for the government in collecting taxes.

Thus, underlying each agency model is an incentive problem caused by some form of asymmetric information. It is common to distinguish models based on the particular information asymmetry involved. We will use the following taxonomy. All models in which the agent has precontractual information we place under the heading of *adverse selection*; except for an occasional reference, we will not deal at all with this category. Our models will assume symmetric information at the time of contracting. Within this category, which we refer to as *moral hazard*, a further distinction is useful: the case where the agent takes unobservable actions, and the case where his actions (but not the contingencies under which they were taken) may be observed. Arrow (1985) has recently suggested the informative names Hidden Action Model and Hidden Information Model for these two subcategories. The worker supplying unobservable effort is the prototypical hidden action case, while the expert manager making observable investment decisions leads to a typical hidden information model.

As will become clear shortly, the hidden action case formally subsumes the hidden information case. (This rationalizes our use of moral hazard as a joint label.) Nevertheless, it is meaningful to keep the two distinct, because they differ in their economic implications as well as in their solution techniques. In this section we focus on the hidden action case. Section 2, on labor contracting, will illustrate the hidden information case.

The general objective of an agency analysis is to characterize the optimal organization response to the incentive problem. Typically, the analysis delivers a second-best reward structure for the agent, based on information that can be included in the contract. Characterizing the optimal incentive scheme is important but not the prime economic purpose. What

is more interesting is the allocational distortions that come with the incentive solution. Although one could often design incentive schemes that induce the agent to behave as if no information asymmetry were present, that is rarely second-best. Instead, some of the costs of the information asymmetry are borne by distortions in decision rules, task assignments, and other costly institutional arrangements. This is what gives the theory its main economic content.

The agency paradigm has indeed been quite successful in shedding light on institutional phenomena that are beyond received microeconomic theory. The second-best nature of incentive efficient solutions admits a host of arrangements that would be inexplicable if information flows were costless. Examples abound in the literature and we could easily use up our allotted space by describing some of them. However, we have chosen not to follow this line, but rather to be more methodologically oriented. Agency models are not without problems, and this is best brought home by going into the details of a generic structure.

We will begin with three different formulations of the agency problem, each with its own merits. Next we go on to discuss a simple version of hidden action that will suffice to sum up the main insights of that type of model. An economic assessment and critique follow, which in turn lead us to a discussion of recent improvement efforts. These include the role of robustness in simplifying incentive schemes and the use of dynamic models to arrive at richer predictions. Finally, we provide a summary of what agency theory has to offer as well as what (in our view) its shortcomings are.

1.2 *Three formulations*

Let A be the set of actions available to the agent and denote a generic element of A by a . Let θ represent a state of nature drawn from a distribution G . The agent's action and the state of nature jointly determine a verifiable outcome $x = x(a, \theta)$ as well as a monetary payoff $\pi = \pi(a, \theta)$. The verifiable outcome x can be a vector and may include π . The monetary payoff belongs to the principal. His problem is to construct a reward scheme $s(x)$ that takes outcomes into payments for the agent.

The principal values money according to the utility function $v(m)$ and the agent according to the utility function $u(m)$. The agent also incurs a cost from taking the action a , which we denote $c(a)$. We assume initially that the agent's cost of action is independent of his wealth, that is, that his total utility is $u(s(x)) - c(a)$. The principal's utility is $v(\pi - s(x))$.

The agent and the principal agree on the distribution G , the technology $x(\cdot, \cdot)$, and the utility and cost functions.

This is the *state-space formulation* of the agency problem as initiated by Wilson (1969), Spence and Zeckhauser (1971), and Ross (1973). Its main advantage is that the technology is presented in what appear to be the most natural terms. Economically, however, it does not lead to a very informative solution.

There is another, equivalent way of looking at the above problem that yields more economic insights. By the choice of a , the agent effectively chooses a distribution over x and π , which can be derived from G via the technology $x(\cdot, \cdot)$. Let us denote the derived distribution $F(\pi, x; a)$ and its density (or mass function) $f(\pi, x; a)$. This *parameterized distribution formulation* was pioneered by Mirrlees (1974, 1976) and further explored in Holmström (1979). For later reference, we state the principal's problem mathematically in parameterized distribution terms. His problem is to

$$\text{Max } \int v(\pi - s(x))f(\pi, x; a) dx \quad \text{over } a \in A, s(\cdot) \in S, \quad (1.1)$$

subject to

$$\int u(s(x))f(\pi, x; a) dx - c(a) \geq \bar{u}, \quad (1.2)$$

$$\int u(s(x))f(\pi, x; a) dx - c(a) \geq \int u(s(x))f(\pi, x; a') dx - c(a'), \quad \forall a' \in A. \quad (1.3)$$

In this program the principal is seen as deciding on the action he wants the agent to implement and picking the least cost incentive scheme that goes along with that action. It is worth noting that because the principal knows the agent (his preferences), he also knows what action the agent will take even though he cannot directly observe it. Constraint (1.3) assures that the incentive scheme is consistent with the action the principal wants the agent to choose, while constraint (1.2) assures the agent a minimum expected utility level \bar{u} , presumably determined in the marketplace.

A solution to the principal's program is not automatically assured; in fact simple examples can be given in which no optimal solution exists. We will encounter a non-existence example shortly, but otherwise we merely assume that a solution exists.¹

The third, most abstract, formulation is the following. Since the agent in effect chooses among alternative distributions, one is naturally led to take the distributions themselves as the actions, dropping the reference

¹ Grossman and Hart (1983a) offer a set of sufficient conditions for existence. A key condition is that the probabilities controlled by the agent are bounded away from zero.

to a . Let p denote a chosen density (or mass) function over π and x , and let P be the set of feasible densities from which the agent can choose. Because the agent can randomize among actions, P can be assumed convex. In the case (π, x) takes on a finite number of values, P is a simplex. The cost function in this formulation is written as $C(p)$, which also will be convex because of randomization.

Of course, the economic interpretation of the agent's action and the incurred cost is obscured in this *general distribution formulation*, but in return one gets a very streamlined model of particular use in understanding the formal structure of the problem.

This way of looking at the principal's problem is also very general. It covers situations where the agent may observe some information about the cost of his actions, or the expected returns from his actions, before actually deciding what to do; in other words, cases of hidden information. To see this, simply note that whatever strategy the agent uses for choosing actions contingent on information he observes, the strategy will in reduced form map into a distribution choice over (π, x) . Thus, ex ante strategic choices are equivalent to distribution choices in some P (properly restricted, of course). Note also that the primitive cost function for actions, $c(a)$, could be stochastic without affecting the general formulation. Taking expectations over costs $c(a)$ would still translate into a cost function $C(p)$, because the agent's utility function is separable.

1.3 *The basic hidden action model*

Much of the general insights obtained from studying hidden action models can be conveyed in the simplest setting, where the agent has only two actions to choose from. For concreteness, let us identify them with working hard, H , and being lazy, L . Also, assume for the moment that x coincides with the monetary payoff to the principal and that the principal is risk-neutral. If the agent works hard, the distribution over x is $f_H(x)$, while if he is lazy, the distribution is $f_L(x)$. In view of this language it is natural to assume that f_H dominates f_L in a first-order stochastic dominance sense; that is, the cumulative distribution functions satisfy $F_H(x) < F_L(x)$, for all x , and the cost of hard work, c_H , is greater than the cost of being lazy, c_L .

Substituting these simplifying assumptions into (1.1)–(1.3) yields a straightforward program that can be easily solved. First, note that if the principal wants to implement L then he should pay the agent a constant, because that yields optimal risk-sharing. The problem therefore assumes interest only if the principal wishes to implement H , because now some risk-sharing benefits have to be sacrificed in order to provide the agent

with the right incentives. Letting λ and μ be the Lagrangian multipliers for constraints (1.2) and (1.3), respectively, we see that the optimal sharing rule must satisfy:

$$1/u'(s(x)) = \lambda + \mu[1 - f_L(x)/f_H(x)] \quad \text{for a.e. } x. \quad (1.4)$$

This is a particular version of Mirrlees's (1974, 1976) formula, analyzed and interpreted further in Holmström (1979). Let us discuss its revealing message.

First, note that if $\mu = 0$ then we have first-best risk sharing ($s(x)$ constant), and the agent picks L in violation of the incentive constraint. Therefore, $\mu > 0$. With μ positive, $s(x)$ will vary with the outcome x , trading off some risk-sharing benefits for incentive provision; more precisely, it will vary with the likelihood ratio $f_L(x)/f_H(x)$. To understand why, a few words on the likelihood ratio are in order.

The likelihood ratio is a concept familiar from statistical inference. It reflects how strongly x signals that the true distribution from which the sample was drawn is f_L rather than f_H . A high likelihood ratio speaks for L and a low for H ; a value of one is the intermediate case in which nothing new is learned from the sample, because it could equally as well have come from either of the two distributions.

The agency problem is not an inference problem in a strict statistical sense; conceptually, the principal is not inferring anything about the agent's action from x , because he already knows what action is being implemented. Yet, the optimal sharing rule reflects precisely the principles of inference. This can be seen even more transparently by rewriting (1.4) formally in terms of a posterior distribution derived from updating a "prior" on H . Let the prior be γ (= probability of H) and denote the posterior $\gamma'(x)$. Then by Bayes's rule and (1.4), we have:

$$1/u'(s(x)) = \lambda + \mu\{(\gamma'(x) - \gamma)/\gamma'(x)(1 - \gamma)\}. \quad (1.4')$$

From (1.4') we see that the agent is punished for outcomes that revise beliefs about H downward ($\gamma'(x) < \gamma$), and rewarded for outcomes that revise beliefs upward. Moreover, the sharing rule is a function of x only through the posterior assessment $\gamma'(x)$; outcomes that lead to the same posterior imply the same reward. As in statistical decision theory, the posterior is a sufficient statistic about the experimental outcome.

The fact that we can interpret the optimal sharing rule in standard statistical terms is important. It is intuitively appealing and it will yield some interesting predictions. At the same time it will reveal the main weakness of the model: As we will see, very few restrictions can be placed on the shape of the sharing rule.

To begin with, consider the issue of monotonicity. One might think that $s(x)$ should always be increasing in x given that f_H stochastically

dominates f_L . Somewhat surprisingly, this is not true in general. The reason is that higher output need not always signal higher effort despite stochastic dominance. For instance, suppose $f_H(x) = f_L(x+1)$ and $f_L(x)$ is not unimodal (say, it has two humps). Then there will exist two values of x such that the higher one has a larger likelihood ratio $f_L(x)/f_H(x)$ than the smaller one, implying that the larger outcome would speak more strongly for a low choice by the agent than the smaller outcome. Just as statistical intuition would suggest, we should pay the agent less in the high outcome state. However, to the extent one thinks this is not descriptive of the economic situation considered, one can add the assumption that the likelihood ratio is monotone in x . Because [from (1.4)] the sharing rule is monotone in the likelihood ratio, this assumption will assure a monotone sharing rule. Not surprisingly, the Monotone Likelihood Ratio Property (MLRP) is a well-known concept from statistics. It was introduced into economics by Milgrom (1981), on which the above discussion is based.

What about other questions concerning the shape of $s(x)$? For example (anticipating an upcoming discussion), are there natural restrictions on the model that yield linear sharing rules? The answer is No. The problem is that the connection between x as physical output and as statistical information is very tenuous. In fact, the physical properties of x are rather irrelevant for the solution; all that matters is the distribution of the posterior (or likelihood ratio) as a function of the agent's action. To highlight the problem, note that x would not even have to be a cardinal measure for its information content to be the same. Since it is the information content of x that determines the shape of the optimal incentive scheme, it is hard to come up with natural economic assumptions that connect the agent's reward in any particular way to the physical measures of x .

There are cases for which linear rules are optimal; in fact, almost any shape of $s(x)$ is consistent with optimality, because output can be endowed with rather arbitrary information content. To illustrate this, suppose we want an optimal rule that is linear. Start with any example with two actions, MLRP and a continuous outcome space. As argued above, the optimal sharing rule will be monotone for such an example; call it $s^*(x)$. Now transform the example by redefining output as $x' = \alpha s^*(x) + \beta$, where α and β are constants to be determined. Because this transformation is monotone, the information content of x' is the same as the information content of x . It follows that the optimal way of implementing H in this revised example is to pay the agent $s(x') = \alpha^{-1}x' - \alpha^{-1}\beta$, which is a linear function of the output x' . With $s(x')$ the agent is paid $s^*(x)$ whenever x' corresponds to x , since we know this is the cheapest way of implementing H . Or, put into statistical terms, this scheme pays the agent the same function of the posterior as the optimal scheme in the initial

example. The role of α and β is to assure that H remains the optimal action to implement in the transformed example.

The same idea can be used to prove the optimality of other shapes as well. Some very weak restrictions apply. For instance, as proved in Grossman and Hart (1983a), $s(x)$ cannot be decreasing everywhere and on average $s(x)$ cannot be increasing too rapidly either. More generally, one can show that $s(x)$ has to satisfy $0 < \int s'(x) f_H(x) dx < 1$, but that is about all. This inability to place natural restrictions on the model that yield commonly observed sharing rules should be contrasted with the theory of risk sharing, where linear schemes (for instance) arise from simple restrictions on preferences alone.

Although the model puts few constraints on the sharing rule, it yields very sharp predictions about the measures that should enter the contract in the first place. To illustrate this, suppose initially that $x = \pi$ and next introduce some other source of information, y , that could potentially be used in the contract. This could be information about the general economic conditions under which the agent operates, or direct monitoring of his performance, or indirect evidence from the performance of agents in stochastically related technologies. When would it be the case that y is valuable in the sense that a contract based on the vector $x = (\pi, y)$ Pareto dominates all contracts based on π alone?

The answer is evident from our earlier discussion and equation (1.4). The additional signal y will necessarily enter an optimal contract if and only if it affects the posterior assessment of what the agent did; or, perhaps more accurately, if and only if y influences the likelihood ratio. Conversely, $s(x)$ will not depend on y precisely when

$$f_L(x)/f_H(x) = h(\pi) \text{ a.e.} \quad (1.5)$$

If (1.5) is true then y will be worthless, but if (1.5) is false then y will have some strictly positive value because $s(x)$ will depend on it. This necessary and sufficient condition can be translated into a more familiar form:

$$f_i(x) = A(x)B_i(\pi) \text{ a.e., } i = L, H. \quad (1.5')$$

In this form the condition says that π is a sufficient statistic for $x = (\pi, y)$. Thus, we have the simple but strong result that y is valuable if and only if it contains some information about the agent's action that is not already in π [Holmström (1979, 1982a) and Shavell (1979)].

This *sufficient statistic condition* underlines again the close analogy between the strategic principal-agent game and classical statistical decision theory, which describes a game against nature. Blackwell's celebrated result, which states that optimal single-person decision rules can be based on sufficient statistics alone, is very similar. Some differences should be

noted, however. First, although (1.5') states that randomization has no value (just as in Blackwell's theorem), this conclusion depends on the separable form of the agent's utility function, as Gjesdal (1982) has shown. [Of course, this randomization could be carried out without an exogenous costly signal, so in this sense it still remains true that y has no value if (1.5') holds.] Second, that any signal with some information about the agent's action has strictly positive value is a fact without counterpart in Blackwell's theorems.

An alternative way of expressing the sufficient statistic condition is to say that it partially orders various information systems [see Grossman and Hart (1983a) and Gjesdal (1982)]. If x and x' are two different information signals (possibly vectors) that can be ordered by Blackwell's notion of informativeness, so that (say) x is more informative than x' , then it is true that x' is not preferred to x . In fact, if the ordering is strict then x is strictly preferred to x' in almost all agency problems. The qualifier "almost all" is needed to take care of exceptional situations in which x' is equal to the optimal sharing rule $s(x)$ for a particular problem, which of course is as much information as one would ever want from x . We leave the qualifier deliberately vague to avoid straying too far from our main course.

The sufficient statistic result gives the model its main predictive content, as we will indicate shortly.²

1.4 *The general case*

Let us consider briefly what happens when one moves beyond the two-action case studied above. Economically, not much new will come out, but it is worth understanding why.

Consider the common case where the agent's action is a continuous, one-dimensional effort variable. The agent's incentive constraint (1.3) is in this case problematic, and it has been standard practice to replace it with the more manageable first-order condition:

² Our discussion of the optimal incentive scheme would not materially change by assuming that the principal is risk-averse; only the left-hand side of equation (1.4) would change to $v'(x-s(x))/u'(s(x))$. We could also have imposed constraints on the agent's wealth, so that $s(x) \geq w$ and (1.4) would remain intact with this constraint effective whenever $s(x) \leq w$ in (1.4). The case of a wealth constraint is of some economic interest, however. If the wealth constraint is binding it may force the agent to receive more than \bar{w} . The economic intuition is that if the agent cannot be punished sufficiently to induce him to choose H , then a bribe – extra rewards for good outcomes – will be the only alternative. These rewards may well lead to slack in (1.2), as Becker and Stigler (1974) first noted. Subsequently, Shapiro and Stiglitz (1984) have used this feature to study the efficiency wage hypothesis, a theory of underemployment arising from the difference between compensation and opportunity cost.

$$\int u(s(x))f_a(x; a) dx - c'(a) = 0. \tag{1.6}$$

Relaxing (1.3) in this way is referred to in the literature as *the first-order approach*. It is easy to proceed to a characterization of the optimal scheme, provided the relaxation embedded in (1.6) is appropriate. The result is as follows:

$$v'(x - s(x))/u'(s(x)) = \lambda + \mu f_a(x; a)/f(x; a) \text{ for a.e. } x. \tag{1.7}$$

Here f_a/f is the continuous counterpart of the likelihood ratio. It is increasing when MLRP holds. Thus, when this characterization is correct, we obtain the same qualitative insights as from the simple two-action case above, including the sufficient statistics results.

Unfortunately, the first-order approach does not always work, in the sense that it will sometimes pick out a scheme that in the end does not satisfy the global incentive constraint (1.3) even though it does satisfy the first-order condition (1.6). Mirrlees (1975) was the first to recognize this dilemma. Subsequently, Grossman and Hart (1983a) and Rogerson (1985b) worked out conditions that ensure the validity of the first-order approach. It is of some interest to understand the resolution, because the issue has received considerable attention.

First, consider a simple extension of the two-action case. Assume the agent controls the following family of distributions:

$$f(x; a) = af_H(x) + (1 - a)f_L(x), \quad a \in [0, 1]. \tag{1.8}$$

In other words, the agent determines by his effort a convex combination of two fixed distributions. This was called the Spanning Condition by Grossman and Hart; we will refer to it as the Linear Distribution Function Condition (LDFC). Note that by randomizing in the two-action model the agent has access to the family described by (1.8).

With LDFC it is evident that the first-order approach is valid. The reason is that no matter what schedule the principal offers to the agent, the first-order condition will coincide with the agent's global incentive constraint (for a fixed action), because the integral in (1.3) is linear in a .

When we treat the general case using the first-order approach, we are effectively taking a linear approximation of the true family of distributions $f(x; a)$ around the particular action (say, a^*) that the principal wants to implement; in other words, we are treating the problem as if the agent were choosing from the hypothetical family:

$$f'(x; \alpha) = f(x; a^*) + \alpha f_a(x; a^*), \quad \alpha \text{ small}, \tag{1.9}$$

using a cost function

$$\tilde{c}(\alpha) = c(a^*) + \alpha c'(a^*).$$

The family in (1.9) is linear in the same sense as LDFC, and there is no problem in obtaining a proper characterization. (Note that since $\int f_a = 0$, f' is a legitimate distribution for small α , provided f_a is bounded.) However, it may be that once we have induced the agent to choose $\alpha = 0$ (i.e., to choose the desired a^*) among the distributions in (1.9), he would actually want to go to another distribution in the true family $\{f(x; a)\}$ that he is controlling. This involves a discrete jump in the effort level and is the source of the potential problem with the first-order approach.

So the question is what distributions we can add to (1.9) and still be assured that the agent would not want to deviate to any of them. Here is the class proposed originally by Mirrlees and later verified by Rogerson (1985b): Assume that $\{f(x; a)\}$ satisfies MLRP and that it additionally satisfies the Convexity of Distribution Function Condition (CDFC):

$$F(x; \lambda a + (1 - \lambda) a') \leq \lambda F(x; a) + (1 - \lambda) F(x; a'), \quad \forall a, a', \lambda \in (0, 1). \quad (1.10)$$

What (1.10) states is that the agent always has an action available yielding a distribution that stochastically dominates the distribution he could achieve by randomizing between the two actions a and a' (in other words, a peculiar sort of diminishing stochastic returns to scale); LDFC is obviously a special case of (1.10).

Now, let us see why this restriction will do the job. The optimal scheme that obtains with the local family of distributions (1.9) is differentiable. From this it follows, using integration by parts, that

$$\int u(s(x)) f(x; a) dx - c(a) = K - \int u'(s(x)) s'(x) F(x; a) dx - c(a), \quad (1.11)$$

where K is an integration constant. Because of MLRP, $s'(x) > 0$ and so (by CDFC) the right-hand side is a concave function in a . Consequently, none of the distributions in the original family will be as appealing to the agent as the action the principal is implementing from the local family (1.9). Hence, $s(x)$ remains optimal in the extended family as well.

This argument is illustrated in Figure 1. The triangle represents the simplex of all distributions in the case where there are only three possible outcomes x_1, x_2 , and x_3 , which we assume for ease of diagramming. One axis measures p_1 , the other p_2 ; the third, $p_3 = 1 - p_1 - p_2$, does not appear in the figure. The curved line CBA is the one-dimensional manifold of distributions $f(x; a)$ [here represented as $\{(p_1(a), p_2(a)) \mid a \in A\}$]; this set is one-dimensional, because the action a is a scalar. Any straight line in the simplex represents a family satisfying LDFC. The shaded region is the set P of all distributions that the agent has access to when randomized strategies are included (cf. the general distribution formulation in

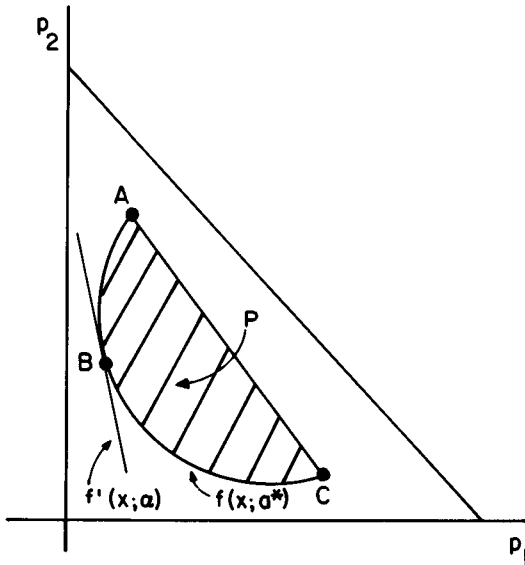


Figure 1

Section 1.2). The figure does not show the cost function or the incentive scheme. With a third dimension measuring costs and rewards, the incentive scheme would be a hyperplane and the cost function a convex manifold in R^3 .

Assume the principal wants to implement the distribution at point B (representing the earlier a^*). Corresponding to the argument above, he starts by designing a cost-minimizing scheme that implements B when the agent's hypothetical alternatives are the distributions along the tangent to B [the tangent represents the distributions in the linear family (1.9)]. This cost-minimizing scheme is characterized by (1.7). Next, CDFC and MLRP assure [using (1.7) and (1.11)] that none of the distributions along the curved line (or in P for that matter) is as attractive to the agent as point B given the scheme in (1.7). Thus, B is indeed implemented in the actual set of feasible distributions P . Without CDFC and MLRP, the agent might want to jump across, for instance to C , when B is being implemented from the tangent set. Then (1.7) would not be valid.

As might be expected, MLRP and CDFC are very restrictive conditions and economically rather peculiar. Particularly, CDFC seems to rule out a number of "natural" families, because few of those we might think

of are closed under convex combinations. For instance, there is no family we know of that satisfies both conditions and is generated from the technology $x = a + \theta$ (or $x = a\theta$). This does not mean that the set of families satisfying both CDFC and MLRP is small. There is an easy way of generating sample families with both properties. Simply start with any two distributions and extend this family by LDFC as in (1.8). If the two initial distributions can be ordered by MLRP, the extended family will have this property. Note that the role of MLRP here is simply to obtain an increasing schedule and not to assure the validity of the first-order approach, which is already guaranteed by LDFC.³

The fact that LDFC appears to be the main instrument for constructing families with CDFC and MLRP leaves open the question of whether there are any interesting cases that do not satisfy LDFC but only CDFC. Except for added convenience in studying examples, this issue is not very interesting either. We already saw how the two-action case was rather rich in generating a variety of optimal incentive schemes. This richness obviously carries over to the LDFC case.

From the preceding discussion one should infer that the first-order approach works in the case where the family of distributions controlled by the agent is one-dimensional in distribution space (LDFC). It also works in cases which are effectively one-dimensional, in the sense that their solution is equivalent to a problem with a one-dimensional family (CDFC plus MLRP). Notice that it is one-dimensionality in distribution space that makes things simpler, not one-dimensionality in the underlying economic variable (effort). Even though effort is taken to be one-dimensional, the curve it traces will generally (when convexified) generate a higher-dimensional P , making matters complex.

What is meant above by “the first-order approach works” also needs a bit of elaboration. Its precise meaning is that the optimal scheme is characterized by (1.7), which is a narrower statement than the claim that one can describe the agent’s choice by first-order conditions. Viewing things in distributional terms, we note that the agent in Figure 1 has two decision variables: p_1 and p_2 . If the cost function over P were strictly convex and the optimal distribution to implement were interior to P (because, say, the cost goes to infinity toward the boundary), then a first-order approach in the traditional sense would work perfectly well. Normally a single first-order condition would not be enough to describe the agent’s behavior, but two would always do. One would then obtain a characteri-

³ Alternatively, of course, one can work with any one-parameter family (for which a solution is known to exist) and then interpret the characterization as referring not to this family necessarily, but to the tangent space of distributions described by (1.9).

zation like (1.7), but with two multipliers μ_1 and μ_2 rather than one. This dilutes the information content of the characterization; the sufficient statistic results will not be as crisp (in particular, optimal incentive schemes may aggregate more than what the earlier sufficient statistic result indicated; see Section 1.6) and statements about monotonicity will be hard to make. Needless to say, when one goes to higher-dimensional cases, the value of a general characterization along these lines quickly disappears.^{4,5}

We conclude that models with a continuous effort variable allow a simple characterization when they look much like the two-action case discussed before. In that case the solution, as far as the optimal reward structure is concerned, exhibits the same features and the same variety. One difference is worth stressing, though. In the two-action model it is difficult to say anything about the agent's choice of action, because it is not determined by a continuous trade-off. One has to compare the solution that implements H with the solution that implements L directly. On the other hand, if effort is a continuous variable and the first-order approach works, then it can be proved (Holmström, 1979) that the optimal level of effort to implement is such that the principal would like to see it go even higher. In other words, in equilibrium we should see principals desiring more effort from their workers. Because this enrichment can be obtained already, by moving from the two-action case to the LDFC case, there appears to be little reason ever to go beyond LDFC in a model that wants to exploit the characterization in (1.7).

1.5 *An intermediate assessment*

The main predictive content of the basic agency model is in the sufficient statistic result, which tells what information should enter into a contract in the first place. Simple as it seems, this result turns out to have quite a bit of economic scope. One trivial implication is that agency relationships create a demand for monitoring. This has generated substantial interest

⁴ Grossman and Hart (1983a) study cases in which the first-order approach may not be applicable. Even with MLRP, incentive schemes need not be monotone. On the other hand, the result that sufficient statistics are sufficient for designing optimal incentive schemes does not depend on the first-order approach. Also, a more informative system (in the Blackwell sense) is strictly better than a less informative one, assuming that the garbling matrix that connects the two systems has full rank. However, signals that provide additional information about the agent's strategy may not be valuable when the first-order approach fails.

⁵ Hidden Information Models, viewed in distribution space, are typically of high dimension, because contingent strategies result in rich distributional choices for the agent (see Section 1.6). This is why the analyses of Hidden Information Models proceed along quite different lines than the analysis of Hidden Action Models.

in the accounting literature and has led to various refinements in predicting the usefulness of different monitoring schemes [for a survey, see Baiman (1982)].

A more significant implication concerns the use of relative performance evaluation [Baiman and Demski (1980), Holmström (1982a)]. Agents who work on tasks that are related – in the sense that one task provides information about the other – should not be compensated solely on individual output, but partly on the output of others. Note that the reason for this (according to the sufficient statistic result) is not that one would like to induce competition for incentive purposes, because if the agents' technologies are not stochastically related then relative performance evaluation is useless at best. Rather, competition is a consequence of the desire to extract information about the circumstances under which the agents performed. This information is used to filter out as much of the exogenous uncertainty as possible, allowing more weight to be placed on individual performance.

A further consequence of the sufficient statistic result is that sometimes aggregate information will do as well as detailed information in relative performance schemes. For instance, if technologies have normal noise then weighted averages of peer performance will suffice as a basis for an optimal scheme. The weights are proportional to the information content of the signals from peers.

Predictions like these accord at least broadly with stylized facts. Relative performance evaluations are commonplace, particularly in the form of prizes (e.g., promotions) awarded to top performers in an organization. Indeed, the labor market as a whole forms a grand incentive structure in which relative evaluations implicitly or explicitly play a dominant role. The literature on rank order tournaments, initiated by Lazear and Rosen (1981), has studied in more detail the performance and design of such contests [see also Green and Stokey (1983) and Nalebuff and Stiglitz (1983)]. We note that the use of rank order as a basis for payment is rarely optimal in the basic agency model; one could usually do better with schemes sensitive to cardinal measures. However, there may be other advantages to rank-order payments not captured by the standard agency model. One reason is that rank is easier to measure in many circumstances. Another argument – suggested by Carmichael (1984), Malcomson (1984a), and Bhattacharya (1983) – is that tournaments provide the principal with incentives to honor promised awards even in cases where legal enforcement is difficult, because performance can be observed but not verified. In tournaments the total amount paid by the principal remains constant and payment should therefore be easy to verify.

Explicit relative performance schemes have recently emerged in executive compensation packages as well. Typically, they relate managerial performance to companies within the industry, which fits the notion that stochastically closer technologies have more value as a basis for optimal rewards. Antle and Smith (1986) have studied more broadly the degree of relative performance evaluation in executive compensation, measuring implicit (as well as explicit) contractual elements. Their statistical tests show that the data in fact exhibit a component of relative compensation, but not to the extent predicted by the basic theory. This seems puzzling at first, but two explanations can be suggested for the evidence. First, executives may be diversifying their portfolio through personal transactions in the market, which do not show up in the data; in fact, Section 1.6 discusses a model with precisely the property that no relative performance payments are necessary because the executive can manufacture them himself. The other, more plausible, reason is that relative performance evaluations distort economic values and thereby decision making (e.g., an executive completely insulated from systematic risk will disregard such risk in evaluating investment decisions). In the one-dimensional agency models normally studied, such decisions are excluded. Including more decision dimensions in the model seems essential for gaining a better fit with the data and a better understanding of the merits of relative performance schemes.

Given that the basic agency model is so general, it is perhaps surprising that it has any predictive value at all. To this can be added the value of having a paradigm within which one may begin considering in more precise terms such subjects as the managerial theory of the firm. Jensen and Meckling's (1976) pioneering work is an example of what insights can be derived from the mere recognition that managers need to be provided with incentives against shirking; another (more explicit) model on the same subject is in Grossman and Hart (1982). Both papers derive the capital structure of the firm from the underlying incentive problems (with opposite hypotheses about the manager's options to dilute the firm's resources). Although these studies beg the question of why capital structure needs to be used for incentive purposes when direct incentive schemes appear to be cheaper, they still open the door for further investigations into a subject that surely is of substantial economic importance.

Let us next turn to the problems with the basic agency model. The main one is its sensitivity to distributional assumptions. It manifests itself in an optimal sharing rule that is complex, responding to the slightest changes in the information content of the outcome x . Such "fine-tuning" appears unrealistic. In the real world incentive schemes do show variety, but not to the degree predicted by the basic theory. Linear or piece-wise linear

schemes, for instance, are used frequently and across a large range of environments.⁶ Their popularity is hardly explained by shared properties of the information technology, as the basic model would have it. It is clear that other technological or organizational features, excluded from the simple model, must be responsible for whatever regularities in shapes we do observe empirically.

Fine-tuned, complex incentive schemes also stand in the way of serious extensions and applications. One can say little about comparative statics properties of the model and it is also hard to introduce additional variables into the analysis. This is a critical drawback, because the unobservable variable in the model (say, effort) is not of primary interest precisely because it cannot be observed. Instead one would be interested in what consequences the agency model has for such observable variables as investment decisions and task assignments, for example. Little has been done in this regard, due to the complexity of the basic solution. [For one attempt that reveals these difficulties, see Lambert (1986).]

Thus casual empiricism, as well as the desire to include decision variables of allocational and aggregate significance, strongly point to a need to refine agency models in the direction of predicting simpler incentive schemes. We turn next to such an effort.

1.6 *Robustness and linear sharing rules*

The prevalence of relatively simple incentive schemes could partly be explained by the costs of writing intricate contracts.⁷ But that is hardly the whole story. A more fundamental reason is that incentive schemes need to perform well across a wider range of circumstances than specified in standard agency models. In other words, incentive schemes need to be robust.

One way of expressing the demand for robustness is to allow the agent a richer set of actions or strategies. Intuitively, the more options the agent has, the more poorly intricate schemes will perform. To give a familiar example: If there is a secondary market for goods, arbitrage will take away all opportunities for price discrimination. Linear schemes are optimal, because they are the only ones that are operational.⁸

⁶ Sharecropping rules are almost exclusively linear, despite great variations in stochastic environments.

⁷ We remind the reader of our discussion of explicit versus implicit incentive schemes in the introduction. Some would argue that real-world schemes are quite complex when viewed as equilibrium phenomena.

⁸ This could be one reason for the prevalence of linear sharing rules in sharecropping. It may also explain why corporate tax schemes are more linear than income tax schemes;

Another elementary example of how added options contribute to simplifications can be given in the context of our basic agency model. We noted that an optimal incentive scheme need not be monotone in general unless MLRP holds. On the other hand, if the agent is allowed free disposal of output, then the only operational schemes are monotone no matter what the stochastic technology looks like. This illustrates the kind of nondistributional considerations that one is led to look for in understanding more universal properties of incentive schemes.

Recently, Holmström and Milgrom (1985) have proposed a simple agency model in which linear schemes are optimal because the agent is assumed to have a rather rich action space. The main idea can best be grasped by describing an example, due to Mirrlees (1974), in which no optimal solution exists. Mirrlees's example has a risk-neutral principal, an agent with unbounded marginal utility for consumption, and a technology with output $x = a + \theta$, where θ is a normally distributed error term with zero mean and a is the agent's labor supply. In other words, the agent controls the mean of a normally distributed output. This technology is the most obvious candidate for an agency analysis, and it is quite a shock to learn that the problem has no solution. The reason is that first-best can be approximated arbitrarily closely by step-function schemes that offer first-best risk sharing (a flat reward) for almost all outcomes except the extremely bad ones for which a severe punishment is applied. This approximation result is in fact easy to understand using the statistical intuition that the basic model offers. The normal technology has a likelihood ratio f_a/f that is unbounded below (it is linear in x). Therefore, very low x values will be very informative about the agent's action, and one can act on that information almost as if it revealed compliance perfectly. The step functions approximate forcing contracts, which are well known to be optimal if there are outcomes that reveal deviations with certainty.⁹

The example is clearly unrealistic and there are ways to patch it (e.g., bound utility or bound the likelihood ratio). But this would be misleading, because the example points to a more fundamental flaw. Step functions come close to first-best only under the unrealistic assumption that one knows exactly the parameters of the problem (utility functions, technology, etc.), and they will generally perform poorly as soon as one introduces slight variations or uncertainty into the model. In other words, the example represents the extreme case of fine-tuning discussed earlier.

Footnote 8 (cont.)

presumably, corporations can circumvent nonlinearities in tax schemes more easily than individuals. (Some would argue that individuals can do a lot of arbitrage as well, making income tax a lot less progressive than it appears.)

⁹ Harris and Raviv (1979) study optimal forcing contracts.

For instance, think of a dynamic context, where the agent is paid (say) at the end of the week, and assume he can observe his own performance during the week so that he can adjust his labor input as a function of the realized path of output. Then step functions will induce a path of effort that will be both erratic and, on average, low (generally, the agent will bide his time to see if there is any need to work at all). In contrast, a linear scheme, which applies the same incentive pressure no matter what the outcome history, will lead to a more uniform choice of effort. This suggests that the optimality of step functions is highly sensitive to the assumption that the agent chooses his labor input only once.¹⁰

This intuition can be made precise by considering a dynamic version of the normal example. Specifically, let the agent control the drift rate μ of a one-dimensional Brownian motion $\{x(t); t \in [0, 1]\}$ over the unit time interval. Formally, the process $x(t)$ is defined as the solution to the stochastic differential equation:

$$dx(t) = \mu(t) dt + \sigma dB(t), \quad t \in [0, 1]. \quad (1.12)$$

Here B is standard Brownian motion (zero drift and unitary variance). Note that the instantaneous variance, $\sigma^2 dt$, is assumed to be constant.

The agent in the model is assumed to have an exponential utility function, and the cost of effort (unlike in our earlier model) is assumed to be independent of the agent's income. In other words, the agent's payoff is:

$$u\left(s(x) - \int c[\mu(t)]\right) = -\exp\left\{-r\left(s(x) - \int c[\mu(t)]\right)\right\} \quad (1.13)$$

as evaluated at the end of the horizon, where $x = x(1)$ is the final position of the process (the profit level at time 1, say), $c(\mu)$ is a convex (instantaneous) cost function, and r is the coefficient of absolute risk aversion. The particular form of the utility function assures that a linear scheme will indeed apply the same incentive pressure over time. In general, income effects would cause distortions.

Notice that if the agent were unable to observe the path $x(t)$, then it would be optimal for him to choose a constant drift rate $\mu(t) = \mu$ [because $c(\cdot)$ is convex], and the end-of-period position x would be normally distributed with mean μ and variance σ . In other words, we would have a model identical to our previously discussed one-period example that had

¹⁰ This can be simply illustrated in the case of a risk-neutral agent, where an infinity of schemes will be first-best. They include a linear scheme with unitary slope as well as the aforementioned step function. However, if the agent receives some noisy information about the technology before choosing his effort, the linear scheme will be uniquely optimal. This idea is used in Laffont and Tirole (1986).

no optimal solution, because step functions approximate first-best. When the agent can observe $x(t)$ and base his choice $\mu(t)$ on the history of the path of $x(t)$ (which we denote x'), the situation is significantly changed. Instead of being constrained to a one-parameter family of outcome distributions, the rich set $\{\mu(x'); t \in [0, 1]\}$ of contingent strategies permits a vastly wider choice. The enormous expansion of the agent's opportunity set limits the principal's options dramatically; in fact, for each strategy that the principal wants to implement there is essentially a unique incentive scheme that he must use, which stands in sharp contrast to the usual flexibility in choice that the principal has in one-dimensional static models.

The one-to-one mapping between strategies and sharing rules makes the model solvable technically (recall the discussion in Section 1.4). The relationship can be written out explicitly, after which it is easy to show that the optimal rule is linear. The interested reader is referred to the original paper (Holmström and Milgrom, 1985) for details.

Intuitively the result can be seen as follows. Consider a discrete version of the Brownian model, one in which the agent controls a Bernoulli process. Because of exponential utility it is easy to see that the optimal compensation scheme, if it could be made contingent on the whole path of periodic outcomes, would be to pay the agent the same bonus each time he has a "success"; the problem is stationary, because there are no income effects. Viewed as an end-of-period payment scheme, this rule pays the agent a constant plus the number of successes times the bonus, which amounts to a linear scheme in end-of-period profits. The Brownian model – being the limit of a Bernoulli process – should therefore be expected to have a linear optimum as well, and indeed it does.

Notice that this line of reasoning shows that the principal need not use the detailed information of the path of the outcome process even if he has access to it. This is a case where an insufficient statistic with respect to the agent's distributional choice (the end-of-period level of profits) is still enough for constructing an optimal rule; in other words, a case where the principal uses more aggregated information than the sufficient statistic results of one-dimensional models would suggest. The reason is that there is no conflict of interest in the timing of effort, only in the aggregate level of effort; hence information about timing is of no value.¹¹

¹¹ We venture the guess that (in multidimensional agency models) additional signals are valuable precisely when they give information about dimensions of choice in which there is a conflict of interest. In one-dimensional models there is a conflict of interest always (by assumption). The result that additional signals have value if they are informative is always true in that case.

The remarkable thing about this model is that by making the incentive problem apparently much more complicated (the rigorous proof that a linear scheme is optimal is nontrivial), it delivers in the end a much simpler solution. In fact, once we know that the optimal incentive scheme is linear it is trivial to solve for its coefficients. A linear scheme will induce the agent to choose a constant level of effort. Therefore we can treat the problem as a static one (cf. the discussion above) in which the agent chooses the mean of a normal distribution, but this time with the constraint that the principal is only allowed to use linear rules. The dynamics rationalizes an ad hoc restriction to linearity in the static model and in the process resolves the nonexistence problem that Mirrlees originally posed!

Computational ease gives the model substantial methodological value. In contrast to general agency models it is easy to conduct comparative statics exercises. More importantly, one can use the model as a building block in studying richer applications of moral hazard. Such applications are further facilitated by the fact that the linearity results extend to situations in which the agent controls the vector of drift rates of a multidimensional Brownian process; or, in static terms, chooses the mean vector of a multivariate normal distribution.

As a brief illustration, let us discuss the effects of agency costs on investment decisions, assuming that investments are made jointly by the principal and the agent. (We cannot let the agent make the choice privately, because that would amount to having him control the variance, which would upset the linearity results.) Suppose there is a collection of projects available for investment. Each project returns $x = \mu + \theta$, where θ is a normally distributed variable with mean m and variance σ^2 and μ is the agent's effort. For a closed-form solution, assume the cost of effort to be quadratic: $c(\mu) = \mu^2/2$. To make the example a bit richer, assume in addition that there is a market index z , normally distributed with variance γ^2 and zero mean, that correlates with x . Then each project can be characterized by the triple (m, σ^2, ρ) , where ρ is the correlation coefficient between z and x .

To determine the best investment one solves first for the optimal incentive scheme and net return to the principal, given a particular project. The optimal scheme is linear in x and z – that is, of the form $s(x, z) = \alpha_1 x + \alpha_2 z + \beta$. The best coefficients are easy to calculate. One finds that the principal should set

$$\alpha_1 = (1 + r\sigma^2(1 - \rho^2))^{-1}, \tag{1.14}$$

$$\alpha_2 = -\alpha_1(\sigma/\gamma)\rho. \tag{1.15}$$

The constant coefficient β is determined by the agent's participation constraint. If he has to be assured a zero certain equivalent, then the principal will be left with an expected net return equal to

$$\pi = m + \frac{1}{2}(1 + r\sigma^2(1 - \rho^2))^{-1}. \quad (1.16)$$

Note that the optimal incentive scheme exhibits relative performance evaluation. The agent is not merely rewarded based on the project outcome x , but also on the market outcome z . The sign of α_2 is the opposite of ρ as one would expect. This is in accordance with the general result that an optimal design should filter out as much uncontrollable risk as possible. Using z as a filter reduces uncontrollable risk by the factor $(1 - \rho^2)$. If x and z happen to be perfectly correlated, all risk can be filtered out and first-best can be achieved. (In first-best, $\alpha_1 = 1$ and $\pi = m + \frac{1}{2}$.)¹²

The best project is the one that maximizes (1.16). Because of the agency problem, we see that project choice depends on the degree of idiosyncratic risk as measured by $\sigma^2(1 - \rho^2)$ (which is the conditional variance of x given z). The price of that risk is a function of the agent's risk aversion (and in general also the cost of effort). There is no price for systematic risk, because the principal is risk-neutral. One could allow a risk-averse principal (with exponential utility) without altering the linearity result; then systematic risk would also enter the decision criterion. But the main point is that, unlike standard portfolio theory, idiosyncratic risk will play a role in investment decisions.

Because idiosyncratic risk carries a price, diversification will generally have value [see Aron (1984) for the same point]. Also, a concern for idiosyncratic risk will give rise to a market portfolio that is more concentrated than under full information. Firms will find value in choosing projects that are more heavily correlated with the market, because that will enable a better incentive design. (This assumes all projects are positively correlated.) Thus, agency costs could amplify aggregate swings in the economy.

This discussion is merely suggestive of what one might be able to do when linear schemes are optimal. It appears that linearity has the potential to take us toward some livelier and more serious economic analyses. [For some other illustrative examples, see the original paper (Holmström and Milgrom, 1985).] On the other hand, the Brownian model is quite

¹² It is worth noting that in this example the agent could privately manufacture the optimal degree of relative performance evaluation by trading in other firms' assets. In other words, the principal could equally well pay the agent based on x alone and leave it up to the agent to filter out uncontrollable risk. (Of course, the agent must not be allowed to trade stock in his own firm.)

special. The technological options are very limited; for instance, the fact that the agent cannot be allowed to make private investment decisions is an unfortunate constraint for applications. The effectiveness of the Brownian model is restricted, because it does not capture the demand for robustness in the most intuitive way. Presumably, one will have to go outside the Bayesian framework and introduce bounded rationality in order to capture the true sense in which incentive schemes need to be robust in the real world.

1.7 *Dynamic extensions*

Dynamic extensions of the basic agency model are of interest for two rather opposite reasons. One has to do with the relevance of the incentive issues portrayed in the static models, the other with the added predictions that might be had from introducing dynamics. In the former category we have theoretical studies that suggest that time may resolve agency problems costlessly. This has been argued both from the perspective of supergames, in which all cooperative gains can be realized between two parties, and in terms of reputation effects created by the market. While we do not concur in either case with the conclusion that incentive problems disappear, it is worth understanding the arguments. They will take us to dynamic models that can expand and sharpen the predictions from the static theory.

The first studies of dynamic agency were those by Radner (1981) and Rubinstein (1979). Both show that in an infinitely repeated version of the basic one-period model, the first-best solution (complete risk-sharing together with correct incentives) can be attained if utilities are not discounted. The analysis does not offer an optimal solution, but rather a class of contracts within which first-best can be reached. These contracts operate like control charts, punishing the agent for a period of time if his aggregate performance falls sufficiently below expectations. Over time, as uncertainty is filtered out by the law of large numbers, the punishments become more severe and the control region tighter. The assumption of no discounting assures that only events in the distant future, where the control is tight and few violations occur, matter.

These models appear to formalize the intuition that in long-term relationships one can cope more effectively with incentive problems, because time permits sharper inferences about true performance.¹³ The fact that

¹³ This interpretation is disputed in Fudenberg et al. (1986). There it is argued that repetition changes the agent's preferences rather than improves monitoring accuracy.

first-best can be achieved is more incidental and a consequence of the unrealistic assumption of no discounting paired with infinite repetition. Even though Radner (1981) has subsequently shown that with some discounting one can still get close to first-best, there is little reason to believe that incentives are costless in reality. The main question then is whether dynamics alters the insights and results from one-period models. In the studies above, as well as in subsequent work by Rogerson (1985a) and Lambert (1983) [see also Roberts (1982) and Townsend (1982)], memory plays a key role, suggesting that an optimal long-term contract might look rather different from a sequence of short-term contracts.

Jumping to such a conclusion is premature, however. The models discussed above assume that the agent cannot borrow and save, in which case long-term contracts substitute in part for self-insurance that would in fact be available to agents (saving is certainly a real option and limited borrowing as well). Could it be true that the gains to long-term contracting identified in the early models are in fact due to restrictions on borrowing and savings?

Recent studies by Allen (1985), Malcomson and Spinnewyn (1985), and Fudenberg et al. (1986) show that this may indeed be the case. More specifically, if one goes to the other extreme and assumes that the agent can access capital markets freely and on the same interest terms as the principal, then long-term contracts will be no better than a sequence of short-term contracts in the (independently) repeated model.

For instance, Allen noted that if there is no discounting, then one can simply appeal to Yaari's (1976) early work on consumption under uncertainty to conclude that a first-best solution can be achieved by having the agent rent the production technology from the principal at a fixed price. The agent, by borrowing or saving, need not be concerned about fluctuations in income, since they can be smoothed out at no cost. In this case self-insurance is perfect and risk carries no premium.

Allen also studies the finite horizon case, but in a pure insurance context [specifically, Townsend's (1982) model], which is simpler than the agency model we have been discussing. Here also he finds that long-term contracts do not dominate short-term contracts. The same results for the agency model are established by Malcomson and Spinnewyn and by Fudenberg et al. These two papers differ in that the former assumes that the agent's borrowing and saving decisions can be verified (and hence his consumption can be controlled contractually), whereas the latter treats these decisions as private to the agent. The basic idea of the argument is very similar, however. The key observation is that long-term contracts can be duplicated by a sequence of short-term contracts by rearranging the pay-

ment stream to the agent without altering its net present value along any realized path. Roughly speaking, the rearrangement works so that the principal clears his balance with the agent in utility terms each period. Because there is a capital market, the timing of payments does not matter. The agent returns to the consumption stream implied by the long-term contract by borrowing and saving appropriately.

Of course, the assumption that the agent can borrow and save freely in the capital market is rather unrealistic. (In addition, the Fudenberg et al. model assumes that the agent can consume negative amounts, which certainly is unrealistic.) Nevertheless, the models do make clear that one should not rush to the conclusion that long-term contracts, at least in repeated settings, have substantial benefits; in some situations, the insights of the one-period models remain unaltered with the introduction of dynamics. More importantly, however, these findings suggest that – because we do observe long-term relationships and long-term contracts – some forces other than income smoothing are likely to be behind the benefits.

There are many potential forces one could think of. Informational linkages between periods are discussed in Fudenberg et al., and some other reasons will be taken up in Section 3. Here we want to stress that when contingences are hard or impossible to verify, so that explicit contracts cannot be easily enforced, long-term relationships are likely to provide major advantages. They can implicitly (via reputation effects) support contracts that may be infeasible to duplicate in short-term relationships. Bull (1985) offers a model of this variety, which we will come back to in Section 3. Radner's and Rubinstein's models are also best interpreted in this fashion; both have self-enforcing equilibria that do not require outside enforcement. Lazear's (1979) model on mandatory retirement is in the same vein. Lazear argues that age-earnings profiles slope upward [as an abundance of empirical evidence corroborates; see, however, Abraham and Farber (1985) for contradicting evidence], because that way incentives for work are maintained over the agent's employment horizon. The implication is that termination of employment should be mandatory, because marginal product will be below pay at later stages in the career. Although the argument needs some refinement, Lazear's model serves well as an illustration of how introducing dynamics can yield additional predictions into the basic agency set-up.

As a related example of reputation modeling, let us consider Fama's (1980) argument that incentive problems, particularly managerial incentive problems, are exaggerated in the agency literature, because in reality time will help alleviate them. His reasoning is different from Radner's and Rubinstein's in that it focuses on the power of the market to police mana-

gerial behavior, rather than on the theory of supergames. Fama coins the term “ex post settling up” for the automatic mechanism by which managers’ market values, and hence their incomes, are adjusted over time in response to realized performance. If there is little or no discounting, then the manager will be held fully responsible for his deeds through his lifetime income stream and, Fama claims, induced to perform in the stockholders’ interest.

Fama’s intuitive argument has been formalized in Holmström (1982b). We will sketch the construction partly to indicate that the first-best result hinges on very special assumptions, but also because the model offers the simplest illustration of reputation formation and suggests some interesting extensions.

Consider a risk-neutral manager who operates in a competitive market for managerial labor. Assume the market can follow the manager’s performance over time by observing his periodic output. At the same time, assume that the manager’s fee cannot be made contingent on output, because enforcing third parties cannot verify the output. Therefore the manager will be paid his expected marginal product in each period.

Obviously, if the world only lasted for one period, the manager would have no incentives to put out extra effort. But if he wishes to stay in the profession longer, matters are different. Prospective employers will follow the manager’s performance and forecast his future potential from past behavior. Logically, this means that there must be some characteristic of the manager that is not fully known to the market and that is being signaled by past performance. For managers, competence or talent is a natural candidate for what is being signaled, though many other alternatives could also be considered.

Let us now see how the uncertainty about the manager’s competence will induce effort even though there is no explicit contract.

In the simplest setting the manager controls a linear technology:

$$x_t = a_t + \eta_t + \theta_t,$$

where x_t is output in period t , a_t is the manager’s effort, η_t is a quantified measure of managerial competence, and θ_t is a driftless i.i.d. sequence of stochastic shocks. Managerial competence progresses over time according to a simple auto-regressive process:

$$\eta_{t+1} = \eta_t + \epsilon_t, \tag{1.17}$$

which is independent of the θ sequence.

In each period the manager will be paid his expected marginal product. This is the sum of his expected competence as assessed on the basis

of past performance and the value of his effort a_t . Because the market is assumed to know the utility function of the manager, it can forecast the manager's choice of a_t .

To find out what the manager will do in equilibrium and what he will be paid, one has to solve a rational expectations equilibrium. This is relatively easy if the shock terms θ and ϵ are normal and the prior on competence is also normal. Then the market will be monitoring a standard normal learning process [see DeGroot (1970)] in which assessments about competence are updated based on a weighted average of present beliefs and the last observation of output. If we denote by m_t the expected value of η_t based on history, the m_t progresses as:

$$m_{t+1} = \alpha_t m_t + (1 - \alpha_t)(x_t - a_t). \quad (1.18)$$

Note that the market in updating beliefs about competence will subtract from output the present level of effort, which it can infer in equilibrium. This filters out time-varying transient effects.

The weights α_t are deterministic functions of time, and converge to some equilibrium value $\alpha \in [0, 1]$ in the long run. The value of α depends on the distribution of the stochastic shock terms. If competence remains constant (i.e., $\text{var}(\epsilon_t) = 0$) then $\alpha = 1$. In general α is smaller the more noisy the competence process is relative to the noise in the output process; that is, the higher the ratio is of signal to noise.

Given that the market updates beliefs according to (1.18) and pays the manager in proportion to m_t each period t , it is easy to calculate the return from managerial effort in period t . In a stationary state the marginal return will be given by

$$k = \beta(1 - \alpha)/(1 - \alpha\beta), \quad (1.19)$$

where β is the manager's discount factor and α is the aforementioned long-run value of the updating weight. From this we can see that if β is close to 1, then marginal returns to effort will be close to 1 both in the manager's objective function and the production technology, so incentives will be right. In general, though, effort will fall short of first-best. It will be lower the lower the discount factor is and the lower the ratio is between the variances of ϵ and θ – that is, the more noise there is in the output process and the less innovation there is in the competence process. This is all in line with intuition. If output is very noisy, returns from effort will be distributed further into the future and will have less value. On the other hand, variation in competence will raise the need to reestablish one's reputation and therefore increase effort. Without (1.17), the manager's effort would converge to zero deterministically.

As in the case of Radner's and Rubinstein's models, the result that first-best can sometimes be achieved is of little interest per se. It requires very special and implausible assumptions, in particular that the manager is risk-neutral and does not discount future payoffs. The main point with the model is rather to illustrate that reputation can indeed enforce an implicit contract of some form, when learning about characteristics is a key factor as it often would seem to be. In the particular example, the implicit contract performs exactly like an explicit contract would (in a world with known competence) if that contract were of the form $s(x) = kx + b$, where k is given in (1.19). It is important to note, however, that when relying on reputation effects (at least as determined in the market) there is little freedom to design the contract in desirable ways.

Wolfson (1985) has conducted an empirical study of the returns to reputation in the market for general partners of oil-drilling ventures. The results conform broadly with the implications of the example. In the market for oil-drilling ventures myopic behavior would dictate that general partners complete fewer wells than limited partners desire (because of the tax code). However, because new ventures come up frequently and new partnerships are formed, one might expect general partners to take into account their reputation and complete more wells than would be optimal in the short run. Indeed, Wolfson finds statistically significant evidence for that to be the case. Reputation is priced in the market much as in the model described. The results correspond to a case where $k < 1$, because Wolfson also finds that residual incentive problems remain and that these are reflected in the price of the shares of limited partners.

These empirical findings give reason to explore further the workings of reputation and learning. The general idea can be pursued in many directions and some interesting work has already been done. Gibbons (1985) has considered what organizations can do to align reputation incentives more closely with true productivity. It is evident from the model described that there need not be a very close relationship, particularly in the early periods, between the returns to reputation for a manager and his present marginal product. Indeed, if we think of young managers in lower positions, their returns from effort may vastly exceed the actual product of what they do, because the future value of being considered competent multiplies in general through enhanced responsibility. One way of coping with the problem, suggested by Gibbons, is to control the flow of information about performance potential so that the initial impact of performance is diminished. Perhaps the phenomenon of young professionals joining larger partnerships before establishing their own firms can be seen as a way of protecting oneself against overly strong reactions by the market if mistakes happen in the early career.

Another paper that elaborates on this simple learning model is Aron (1984). She uses the learning effects to derive a number of implications concerning the correlation between the growth rate of firms, the degree of diversification within firms, and the size of firms.

Although our example supported the common intuition that incentive problems are alleviated by long-term considerations, it is important to stress that this is by no means true universally. In fact, career concerns can themselves be a source of incentive problems. For instance, in Holmström and Ricart-Costa (1986) a model is analyzed in which incongruities in risk-taking between managers and shareholders arise purely because of reputation effects. The reason is that managers look upon investments as experiments that reveal information about their competence, while shareholders of course view them in terms of financial returns. The main point is that there is no reason for a project's human capital return to be closely aligned with its financial return; hence the problem requires explicit incentive alignment. For those who distrust incentive models that rely on effort aversion, such a model provides a new channel for analyzing managerial risk-taking incentives.¹⁴

Finally, we want to mention the work of Murphy (1986) as an example of how dynamics can help discriminate between competing theories of compensation. Murphy compares two hypotheses for why age-earnings profiles tend to be upward-sloping. One is the earlier mentioned model of Lazear. The other theory suggests that the upward slope comes from learning about productivity and the contracting process associated with insurance against that risk [see e.g. Harris and Holmström (1982)]. Murphy argues that if the incentive hypothesis were true then the variance in individual earnings should increase with tenure due to income smoothing. The reverse should be true if the learning hypothesis held, because then the effects of performance information are strongest in the early years. Murphy tests these competing positions on panel data for executive compensation drawn from prospectuses. His results are rather inconclusive, perhaps because both effects are really present. But the main point is that in principle dynamic models allow discrimination that is plainly unavailable from single-period studies.

1.8 *Summary and conclusions*

Despite the length of this section we have covered only a few dimensions of the extensive literature on principal-agent models. Before summing up

¹⁴ A related reputation model concerning risk-taking, which derives very interesting predictions about the nature of debt contracts and credit rating in capital markets, is in D. Diamond (1985).

we want to mention two important omissions. One is the lack of examples of Hidden Information Models, which have played a visible role in the literature, often under the name of Mechanism Design [see Harris and Townsend (1981) and Myerson (1979) for seminal contributions and Green (1985) for a unified look at the field]. We will partly make up for this omission in Section 2, where a model of hidden information is analyzed in connection with labor contracting. The mechanism design approach has been quite successful in explaining a range of institutions that are beyond the scope of standard theory, and it has also offered insights into normative problems such as taxation [Mirrlees (1971)], auction design [Harris and Raviv (1981), Myerson (1981), and Maskin and Riley (1984)], and regulation [Baron and Myerson (1982), Baron and Besanko (1984), and Laffont and Tirole (1986)]. In the same way as the models we have discussed here, these models are plagued by an excessive sensitivity to informational assumptions, which makes it hard to go beyond qualitative conclusions.

The other major omission is that we have not discussed at all the general equilibrium effects from contracting, an area in which Stiglitz has been particularly active. As Stiglitz has noted in a variety of different contexts [see, e.g., Arnott and Stiglitz (1985)], the imperfections of second-best contracts will have external effects that may be important. The general idea, familiar from second-best theory, can be described as follows. In all economies, contracting between two parties will have some equilibrium effect on the rest of the economy. However, in the idealized Arrow-Debreu world, equilibrium occurs at a social optimum and so the impact of marginal changes in a bilateral contract will have zero social costs. In contrast, when we are in a second-best world (for whatever reason), marginal changes in contracts will have a first-order effect on the social welfare function, which is not accounted for by the contracting parties.¹⁵ Perhaps one relevant example would be the consequences of nominal contracts in one part of the economy on the use of indexed contracts in other parts.

Naturally, such externalities could give reason for government intervention. However, one should be careful in making sure that there is an improving policy that acts solely on information that the government has available. As a modeler it is easy to spot improvements, because the modeler sees all the relevant information. But that does not imply automatically that the government can improve things, particularly if the more stringent notions of efficiency that are associated with incomplete infor-

¹⁵ A somewhat different dimension of the same problem appears when a party contracts with many independent agents in a decentralized fashion. This has been recently looked at by Cremer and Riordan (1986), but it deserves much more attention.

mation models are applied. Operational welfare schemes in this sense seem to have been little explored in the literature to date.

To summarize:

1. In reduced form all agency models have the agent choose from a family of distributions over observable variables, such as output. A key simplification in Hidden Action Models is to assume that the agent controls a one-dimensional family of distributions. This leads to a simple and intuitive characterization of an optimal scheme. One-dimensionality does not refer here to any economic variable, like effort, but to the set of distributions that the agent can choose from. Understanding this is important for resolving the confusions associated with the validity of the characterization of the optimal rule, which is sometimes (but misleadingly) referred to as the validity of the first-order approach.

2. The main insight of the basic Hidden Action Model is that the optimal incentive scheme looks like one based on an inference about the agent's action from observable signals. This implies that the optimal scheme is highly sensitive to the information content of the technology that the agent controls, which has only loose ties with the physical properties of that technology. Consequently, fiddling with the information technology will accommodate almost any form of incentive schedule and the theory is really without predictive content in this regard.

What does have some predictive content, however, is the result that a contract should use all relevant information that is available up to a sufficient statistic. Among other things, it leads to statements about the use of relative performance evaluation that seem to match empirical evidence at least broadly.

3. However, the extreme sensitivity to informational variables that comes across from this type of modeling is at odds with reality. Real-world schemes are simpler than the theory would dictate and surprisingly uniform across a wide range of circumstances (e.g., linear schemes are quite common in a variety of situations). The conclusion is that something other than informational issues drives whatever regularities one might observe. One possibility that has recently been suggested is that the usual agency models are overly simplistic and fail to account for the need to have schemes that perform well in a variety of circumstances – that is, schemes that are robust. We gave one example of a model in which robustness issues lead to linear schemes. It seems that research in this direction could have high payoffs in the future.

Another reason why schemes in reality are simpler and less sensitive to environmental differences is that exotic contracts are hard to evaluate in terms of both their implied performance and their value for the parties

involved. This is not something which we addressed, because it seems to fall outside the common Bayesian paradigm; but that is not to say it is unimportant. Research along these lines may also have high payoffs.

4. The common Hidden Action Models are rather weak predictively. One reason is that complex incentive schemes make it hard to say anything about distributional choices. The other reason is that the actions in the model are not observable economic variables. (In this regard Hidden Information Models are more useful, because actions – e.g., levels of investment or employment – are usually observable; see Section 2.) Modeling efforts should be directed more toward including interesting economic quantities that focus on allocational consequences of agency. Robustness arguments that predict simpler schemes should be helpful in this endeavor, as indicated in Section 1.6.

5. Another useful direction for sharpening predictions from agency models is to go to dynamic formulations. These bring to bear time series and panel data that allow discriminations that are impossible to make in static models. Dynamic models also bring attention to reputation effects and long-term explicit and implicit contracting that may well be at the center of real-world incentive problems.

2 **Labor contracts**

One of the first applications of contract theory was to the case of contracts between firms and workers [the seminal papers are by Azariadis (1975), Baily (1974), and Gordon (1974)]. Section 2 is concerned with this work and various extensions, including the introduction of asymmetric information and macroeconomic applications. We begin with the Azariadis–Baily–Gordon model itself. [For an excellent recent survey of labor contract theory with a rather different focus from the present one, see Rosen (1985).]

2.1 *The Azariadis–Baily–Gordon (ABG) model*

The ABG model was developed to explain non-Walrasian employment decisions, particularly layoffs, and to understand deviations between wages and the marginal product of labor. It is based on the idea that a firm offers its risk-averse workers wage and employment insurance via a long-term contract.

The model can be described as follows. Imagine a single firm that has a long-term relationship with a group of workers.¹⁶ Presumably a lock-in

¹⁶ On the empirical importance of such relationships, see Hall (1980).

effect of some sort explains why the relationship should be long-term, although this is not modeled explicitly. To simplify, assume that the relationship lasts two periods. At date 0, the firm and workers sign a contract while employment and production occur at date 1. ABG stress the idea of an implicit contract; we postpone discussion of this until Section 3.4 and rely on the contract being explicit and legally binding.

Let the firm's date-1 revenue be $f(s, L)$, where s represents an exogenous demand or supply shock, and L is total employment at date 1. Assume that the date-0 workforce consists of m identical workers, where m is given.¹⁷ Each worker has an (indirect) von Neumann–Morgenstern utility function $U(I, \ell; p)$, where I represents income or wages received from the firm, ℓ is employment in the firm, and p refers to a vector of consumption goods prices. We shall suppose that p is constant and therefore suppress it in what follows.¹⁸ We assume that $U_I > 0$, $U_\ell < 0$, and U is concave in I and ℓ with $U_{II} < 0$ (i.e., workers are risk-averse). The firm, on the other hand, is supposed to be risk-neutral. We shall assume that ℓ is a continuous variable, in contrast to ABG who suppose that it equals 0 or 1.

In the ABG model, the state s is taken to be publicly observable at date 1, although unknown to both parties at date 0. In this case, a contract can be contingent in the sense of making I and ℓ functions of s : $I = I(s)$, $\ell = \ell(s)$. Since ℓ is smooth and U is concave in ℓ , it is desirable to have work sharing at date 1; that is, $\ell(s) = (L(s)/m)$ (so this version of ABG does not explain layoffs; see, however, Section 2.4B). Therefore, an optimal date-0 contract solves:

$$\text{Max } E_s[f(s, m\ell(s)) - mI(s)] \quad \text{subject to } E_s[U(I(s), \ell(s))] \geq \bar{U}, \quad (2.1)$$

where both expectations are taken with respect to the objective probability distribution of s , which is assumed to be common knowledge at date 0. We are adopting the assumption that the firm gets all the surplus from the contract while the workers are held down to their date-0 reservation expected utility levels \bar{U} . Nothing that follows depends on this ex ante division of the surplus, however.

The solution to (2.1) is very simple. Under the usual interiority assumptions, it is characterized by

$$\frac{\partial f}{\partial L}(s, m\ell(s)) = - \left(\frac{\partial U}{\partial \ell}(I(s), \ell(s)) \Big/ \frac{\partial U}{\partial I}(I(s), \ell(s)) \right) \quad \text{for all } s, \quad (2.2)$$

¹⁷ In a more general model, the size of the workforce would be a choice variable.

¹⁸ Two assumptions are embodied here. First, that p is independent of the shock s hitting the firm; and, second, that the firm and workers are sufficiently small that their actions do not affect prices. We shall maintain both assumptions throughout Sections 2 and 3.

$$\frac{\partial U}{\partial I}(I(s), \ell(s)) = \lambda \quad \text{for all } s, \tag{2.3}$$

$$E_s U(I(s), \ell(s)) = \bar{U}, \tag{2.4}$$

where λ is a Lagrange multiplier. Equation (2.2) tells us that the marginal rate of substitution between consumption and employment equals the marginal rate of transformation in each state; (2.3) tells us that a worker's marginal utility of income is constant across states. It is the condition for optimal insurance between a risk-averse agent and a risk-neutral agent. [Note that (2.3) implies that if $\ell(s_1) = \ell(s_2)$ then $I(s_1) = I(s_2)$; that is, wages vary only if employment does.]

Several observations can be made. First, it follows from (2.2) that employment decisions will be ex post Pareto efficient in each state. Hence to emphasize what is by now well known, the ABG model does not explain inefficient employment levels. Although there was some initial confusion about this result, it is not exactly surprising given that an ex ante optimal contract should exploit all the gains from trade ex post (under symmetric information). Employment levels, however, although efficient, are not generally the same as in a standard Walrasian spot market, where the wage $w(s)$ in state s satisfies

$$\frac{\partial f}{\partial L}(s, m\ell(s)) = w(s) = - \frac{\partial U}{\partial \ell}(w(s)\ell(s), \ell(s)) \Big/ \frac{\partial U}{\partial I}(w(s)\ell(s), \ell(s)). \tag{2.5}$$

The point is that the possibility of income transfers across states permits a divergence between $(I(s)/\ell(s))$ and $w(s) = (\partial f/\partial L)(s, m\ell(s))$. In fact, if labor is a normal good and the Walrasian labor supply is upward-sloping, Rosen (1985) has pointed out that employment will generally vary more in a contractual setting than in a Walrasian spot market.¹⁹

An important special case is where labor causes no disutility for a worker per se, but simply deprives him of outside earning opportunities at date 1. This can be represented by

$$U(I, \ell) = \hat{U}(I + R(\bar{\ell} - \ell)), \tag{2.6}$$

where $\bar{\ell}$ is the worker's total endowment of labor and R is the wage in alternative date-1 employment [in (2.6), labor is neither normal nor inferior]. Equations (2.2) and (2.5) both then become

¹⁹ The reason is the following. In a spot market, a worker's incentive to work hard in a good state where the wage rate is high (the substitution effect) will be offset by his desire to consume a lot of leisure given that his income is high (the income effect); and conversely in a bad state. In a contractual setting, the income effect is reduced in size because the firm provides income insurance across different states of the world.

$$\frac{\partial f}{\partial L}(s, m\ell(s)) = R; \quad (2.7)$$

that is, employment levels will be exactly the same in a contract as in a Walrasian spot market. Equation (2.3), on the other hand, implies that

$$I(s) + R(\bar{\ell} - \ell(s)) = \text{a constant}. \quad (2.8)$$

That is, optimal insurance leads to the equalization of a worker's (real) income across states of the world (relative to the prices p), a very different outcome from what one would see in a spot market.

The ability to explain the divergence between workers' wages and their marginal (revenue) product of labor is the principal achievement of the ABG model. In fact, the model provides a striking explanation of sticky (real) wages or incomes, which is in notable contrast to that provided by, say, disequilibrium theory.²⁰

Let us examine the underlying assumptions of the ABG model. A key assumption is that firms are less risk-averse than workers, and are therefore prepared to act as insurers. To the extent that the shock s is idiosyncratic to the firm (we have essentially assumed this anyway in regarding goods prices p as independent of s), this is reasonable since it is probably easier for a firm's owners to diversify away idiosyncratic profit risk via the stock market than it is for workers to diversify away human capital risk. However, the assumption is less convincing in a macroeconomic setting where firms' shocks are correlated.²¹

Even when the shock is idiosyncratic, it is not obvious that a worker must look to his own firm for insurance. Why not go to an insurance company? In the ABG world, where s is publicly observable, there should be no difficulty in making payments to and from the insurance company conditional on s . However, if the model is complicated, some justification for the firm as insurer can be given.

First, it may be the case that, while s is observable to the firm and the workers, it is not observable to the insurance company. If the insurance company relies on a worker to report s , the worker will, of course, have an incentive to announce an s that maximizes his transfer from the insurance company. Now it is possible that the insurance company can learn s by getting independent reports on it from the firm and the workers, but there is the danger that the firm and workers may collude. The whole pro-

²⁰ We have assumed that the firm is risk-neutral, but the main results generalize to the case of firm risk aversion. In particular, as long as the firm is sufficiently less risk-averse than the workers, workers' incomes will be stabilized relative to the spot market outcome. Note also that equation (2.2) continues to hold when the firm is risk-averse.

²¹ Although the Knightian argument can be made that entrepreneurs are, by self-selection, less risk-averse than workers. For a formalization, see Kihlstrom and Laffont (1979).

cess may involve considerable costs relative to the case of insurance by the firm.

In fact, to provide optimal insurance, it is not necessary that the insurance company observe s , only that it observe wages $I(s)$ and employment $\ell(s)$. However, even if it can observe these variables, new problems arise if some aspect of a worker's performance is unobservable to the insurance company. For example, suppose that to make employment productive it is necessary that a worker exert effort, e . Then the optimal risk-sharing contract would insure a worker's wage subject to the worker exerting effort. If the insurance company – which cannot observe s or e – offers insurance, the worker may exert no effort and claim that his low wage was a result of a bad s . Again this problem is reduced if the firm, which does observe e , acts as insurer.

The reader may wonder how, if s and e are not observable to outsiders such as insurance companies, a contract between the firm and workers making I and ℓ functions of s and e can be enforced. This is an important question, to which two answers can be given. First, it may be the case that the firm and workers each have enough evidence to establish to an outsider what s and e really are; that is, in the event of a dispute between them “the truth will come out” (whereas in a three-party contract involving an insurance company, collusion between the firm and the workers may prevent this). Second, if the contract is implicit rather than explicit, then it may be enforced by reputational considerations; that is, the firm will not deny that the worker exerted effort if he really did, because this would ruin its reputation with future workers (for more on this, see Section 3.4).

2.2 *The possibility of worker quits*

The ABG model is based on the idea that firms insure workers against fluctuations in their real income. This means that workers will receive more than their marginal (revenue) product in some states and less in others. A difficulty that has been raised with this is that a worker may quit in the latter states – that is, simply walk away from the contract. This will, of course, be a problem only if the worker's marginal product outside the firm is comparable to that inside; that is, if the lock-in effect that is responsible for the long-term relationship in the first place is small. If it is small, however, the insurance element of the contract will be put under severe pressure.

To see this, suppose that there is a single worker ($m = 1$) who can work either in the firm ($L = 1$) or outside ($L = 0$). To simplify, assume that the worker's marginal (equals average) product, denoted by s , is the same inside and outside the firm (i.e., there is no lock-in at all), and that the

worker cares only about total income: $U = \hat{U}(I)$ [as in (2.6)]. Then, in order to stop the worker from quitting at date 1, the firm must pay him at least s in every state. However, in order to break even on the worker, the firm cannot pay him more than s . The conclusion is that the firm will pay the worker exactly his marginal product in each state, which is, of course, the spot market solution.

In this extreme case of no lock-in, then, the insurance element is completely destroyed. Holmström (1983) has argued that this conclusion is no longer valid when employment and production take place at more than one date. The argument is the following. In the above example, the firm could provide complete insurance at date 1 and at the same time avoid quits by agreeing to pay the worker $\bar{s} = \text{Max } s$ in every state. Of course, the firm takes a loss on this, but if the worker also has a nonstochastic productivity s_0 at date 0, the firm can offset this loss by paying the worker less than s_0 at date 0. There is a cost of doing this because, assuming that the worker cannot borrow, his consumption path will be more steeply sloped over time than he would like. [If the worker's utility function is $\hat{U}(I_0) + \delta \hat{U}(I_1)$, where $1/\delta - 1$ is the market rate of interest, then the first-best contract would have $I_0 = I_1(s) = \bar{I}$ (say) for all s ; i.e., complete income smoothing.] It is easy to show that when this cost is traded off optimally against the insurance benefit, the outcome is incomplete insurance of the following sort: The firm puts a floor on date-1 income by guaranteeing the worker at least $\hat{s} < \bar{s}$; however, in states where $s > \hat{s}$, the firm agrees to pay the worker his full marginal product s .

One benefit of the quit model is that it provides an explanation of the back-end loading of earnings (the worker gets less than his marginal product at date 0 and at least his marginal product at date 1).²² However, the model is based on a number of fairly strong assumptions. First, it is supposed that, while the firm is bound to the contract, the worker can simply walk away. One may ask why the contract cannot specify either that a worker cannot quit at all, or (less extremely) that a quitting worker must compensate the firm by paying an "exit fee." In answering this question, some have appealed to the idea that the courts will not enforce involuntary servitude of this sort (although note that we are really talking about voluntary servitude insofar as the worker presumably agrees to the contract at date 0). While historically this may have been the case, it is interesting to note that attitudes seem to be changing; the use of exit fees (e.g., repayment of training or transportation costs by leaving workers) seems to be on the increase, with recent indications being that the courts are prepared to enforce them (*New York Times*, October 30, 1985). In

²² There is an obvious parallel between Holmström's model and Becker's (1964) analysis of worker training.

particular, there seems to be a move to apply to labor contracts the basic principle of common law that the victim of a breach of contract is entitled to compensatory damages – that is, to be put in as good a position as if the breach had not occurred. In the quit model, compensatory damages correspond to the quitting worker paying the firm $s - I(s)$ in state s , where $I(s)$ is his wage if employed by the firm. In this case, however, the worker never desires to quit and the first-best $I_0 = I_1(s) = \bar{I}$ can be achieved.²³

The quit model also assumes, like the ABG model, that the firm must provide workers' insurance. We have given some justifications for this above, but they become less plausible when the lock-in effect is small. The reason is that, if s is the amount that the worker can earn inside or outside the firm, the assumption that an insurance company cannot observe s is perhaps less convincing (although there may still be problems in enforcing a contract based on s if s is not "verifiable"; see Section 3). Even if outsiders cannot observe s , the worker could still rely on the firm for insurance by borrowing a fixed amount from a bank and depositing it with the firm; the worker would receive it back only if he did not quit (i.e., the worker could post a bond). Such an arrangement would again achieve the first-best, although it may of course stretch to the limit the assumption that the firm will not default on its part of the contract.²⁴ (Note that this arrangement does involve a form of back-loading.)

One case where these criticisms do not apply is where the worker can simply "disappear." If this is so, then the firm knows that it will never be able to collect any exit fee and no bank will be prepared to lend to the worker. Another reason for the absence of exit fees or bond posting is that the worker may sometimes quit for reasons other than a high alternative wage; for example, work in the firm may become intolerable or the worker may become sick. These states are likely to be bad for the worker and so for reasons of risk-aversion he will be unwilling to forfeit a substantial amount in them. (We are assuming that the reason for quitting is not publicly observable and so the exit fee cannot be made contingent upon it.) Considerations such as these seem likely to lead to a not insignificant complication of the model, however, and it is unclear how robust the back-end loading result is to their introduction.

2.3 *Asymmetric information*

Let us return to the case where all parties to the contract are bound. As we have seen, the ABG model can explain sticky (real) wages or incomes,

²³ It is also worth pointing out that various forms of disguised exit fees may actually be quite common; consider, e.g., non-vested pensions.

²⁴ Note that deposits are used in some contexts; consider, for instance, rental deposits.

but not *ex post* inefficient employment. Because of this, various attempts have been made to enrich the ABG model. An important development has been the introduction of asymmetric information. The first set of models along these lines considered the case where the firm's revenue shock s is observed only by the firm at date 1 [see Calvo and Phelps (1977) and Hall and Lilien (1979)]. This "hidden information" assumption, as Kenneth Arrow has termed it, has force when the party with private information is risk-averse.²⁵ It is this supposition that underlies the models of Azariadis (1983) and Grossman and Hart (1981, 1983b): The firm is identified with its risk-averse manager.

A consequence of managerial risk aversion is that it is no longer optimal for the firm to provide workers with complete income insurance as in the basic ABG model; rather, the manager will now want to obtain some insurance himself. The manager's ability to obtain insurance, however, is limited by his private information. For example, an insurance contract that pays the manager $\$ \alpha > 0$ in state one and taxes the manager $\$ \beta > 0$ in state two cannot be implemented if the insurer must rely on the manager to report which of the two states has occurred (the manager will always report state one). However, the manager's incentive to report the wrong state can be lessened by introducing a production inefficiency: The manager will be less inclined to report state one if, as well as receiving $\$ \alpha$, he must choose a production plan that is inefficient and (moreover) relatively unprofitable if the true state is indeed two. We shall see that the second-best optimal insurance contract includes production inefficiencies of this kind; furthermore, under certain conditions the inefficiencies take the form of underemployment of labor in bad states of the world.

It turns out that the case where the manager and workers must provide each other with insurance is quite complicated to analyze. A considerable simplification is possible, however, if it is supposed that each group can get insurance from a risk-neutral third party (this is, of course, a departure from the idea that the firm has a comparative advantage in insuring the workers; or vice versa). Although the existence of such a third party may at first sight seem farfetched, it can be argued that in the case of a public company the firm's shareholders play this role, acting as a financial wedge between the manager and the workers (moreover, risk-neutrality of the shareholders may be reasonable to the extent that they hold well-diversified portfolios).²⁶

If workers can get insurance from a third party, the long-term contract between the firm and workers becomes much less important, and in fact a simple case (which we follow) is where this contract is ignored altogether,

²⁵ If the party with private information is risk-neutral, the first-best can be achieved by making this party the residual income claimant.

²⁶ The analysis below follows Hart (1983).

with the firm being assumed to make all input purchases in the date-1 spot market. That is, we now focus on a risk-averse manager with private information, who insures himself with a risk-neutral third party and buys all his inputs in the date-1 spot market. (Below we discuss the implications of putting the worker–firm contract back into the analysis, particularly when the firm has a comparative advantage in insuring the workers.)

The main implications of asymmetric information can be understood from the special case where there are only two states of the world [we follow Holmström and Weiss (1985)]. We now interpret f to be the manager's benefit function (measured in dollars). This benefit is supposed to be private; that is, it does not show up in the firm's accounts and so payments cannot be conditioned on it. We write $f = f(s, \mathbf{L})$, where we are now more general in allowing $\mathbf{L} \geq 0$ to be a vector of inputs or managerial decisions. It is assumed that, while s is observed at date 1 only by the manager, \mathbf{L} (which is chosen after s is observed) is publicly observable. It is in fact convenient to regard f as the manager's net benefit in state s after all inputs have been purchased in the date-1 spot market.

Let the two states be $s = s_1, s_2$ with probabilities π_1, π_2 respectively ($\pi_1, \pi_2 > 0, \pi_1 + \pi_2 = 1$). The manager signs a contract with a risk-neutral third party. The contract says that in state $s_i, i = 1, 2$, the third party will pay the manager I_i and the manager must choose \mathbf{L}_i . An optimal contract solves

$$\text{Max } \pi_2 V(f(s_2, \mathbf{L}_2) + I_2) + \pi_1 V(f(s_1, \mathbf{L}_1) + I_1), \quad (2.9)$$

subject to the following constraints:

$$\begin{aligned} f(s_2, \mathbf{L}_2) + I_2 &\geq f(s_2, \mathbf{L}_1) + I_1, \\ f(s_1, \mathbf{L}_1) + I_1 &\geq f(s_1, \mathbf{L}_2) + I_2, \\ \pi_2 I_2 + \pi_1 I_1 &\leq 0. \end{aligned}$$

Here V is the manager's von Neumann–Morgenstern utility function, where $V' > 0, V'' < 0$. The third constraint says that the third party is prepared to participate in the contract (we give the firm all the surplus from the transaction). The first and second constraints are the well-known truth-telling constraints [see, e.g., Myerson (1979)]. Because the third party cannot observe s directly it must rely on the manager to report s . Constraints 1 and 2 say that the manager will report $s = s_2$ when s_2 occurs and $s = s_1$ when s_1 occurs.²⁷

Another interpretation of the contract is that, instead of asking him to report s , the contract gives the manager the choice of the pairs (I_i, \mathbf{L}_i)

²⁷ A more general contract would make the outcome (I_i, \mathbf{L}_i) depend stochastically on the report s_j . Such random contracts are more complicated to analyze and, at least for the two-state case considered here, do not lead to substantially different results. On random schemes, see Maskin and Riley (1984) and Moore (1985).

and (I_2, L_2) . The first and second constraints then say that the manager will choose (I_i, L_i) in state s_i .

We shall assume that s_2 is the good state and s_1 the bad state, in the sense that total benefits are higher in s_2 than in s_1 :

$$f(s_2, L) \geq f(s_1, L) \quad \text{for all } L \geq 0, \text{ with strict inequality if } L \neq 0. \quad (2.10)$$

We suppose also that

$$\begin{aligned} &f(s, L) \text{ is strictly concave in } L, \text{ and the (unique)} \\ &\text{maximizer } L(s) \text{ of } f(s, L) \text{ exists and satisfies} \\ &L(s) \neq 0 \text{ and } L_1^* \equiv L(s_1) \neq L(s_2) \equiv L_2^*. \end{aligned} \quad (2.11)$$

That is, the maximizer of $f(s, L)$ is sensitive to s . Statements (2.10) and (2.11) imply immediately that the first-best – the solution to (2.9) without the truth-telling constraints – cannot be achieved under asymmetric information. The first-best has the property that L is chosen to maximize f in each state and the manager is perfectly insured; that is,

$$f(s_2, L_2) + I_2 = f(s_1, L_1) + I_1, \quad (2.12)$$

where

$$\pi_2 I_2 + \pi_1 I_1 = 0 \quad \text{and} \quad L_2 = L_2^*, \quad L_1 = L_1^*.$$

But given (2.10)–(2.11), this violates the first truth-telling constraint (i.e., the manager will *under-report* s in state s_2).²⁸ This observation suggests that only the first truth-telling constraint will be binding in the solution to the second-best. This turns out to be true, as is proved in the appendix. (It is interesting to note that in the two-state case we can establish this even in the absence of a Spencian single-crossing property on marginal benefit.) It follows that the first-order conditions for (2.9) are

$$\begin{aligned} &\pi_2 V'_2 - \mu \pi_2 + \lambda = 0, \\ &(\pi_2 V'_2 + \lambda) \frac{\partial f}{\partial L_k}(s_2, L_2) = 0 \quad \text{for all } k, \\ &\pi_1 V'_1 - \mu \pi_1 - \lambda = 0, \\ &\pi_1 V'_1 \frac{\partial f}{\partial L_k}(s_1, L_1) - \lambda \frac{\partial f}{\partial L_k}(s_2, L_1) = 0 \quad \text{for all } k, \end{aligned} \quad (2.13)$$

where $V_i \equiv V(f(s_i, L_i) + I_i)$, and similarly V'_i ; $\lambda \geq 0$ is the Lagrange multiplier for the first constraint, $\mu \geq 0$ for the third, and the second constraint has a zero multiplier. In fact, $\lambda > 0$ because $\lambda = 0$ gives us the first-best, which we know violates constraint 1.

²⁸ The first-best could be achieved if the manager were risk-neutral, because in this case no insurance is required at all; i.e., $I_1 = I_2 = 0$ and $L_i = L(s_i)$, $i = 1, 2$, which satisfies the truth-telling constraints.

From the second equation in (2.13), we see that

$$\frac{\partial f}{\partial L_k}(s_2, \mathbf{L}_2) = 0 \quad \text{for all } k \quad (\text{i.e., } \mathbf{L}_2 = \mathbf{L}_2^*), \tag{2.14}$$

while the third and fourth equations imply that

$$\mu\pi_1 \frac{\partial f}{\partial L_k}(s_1, \mathbf{L}_1) + \lambda \left(\frac{\partial f}{\partial L_k}(s_1, \mathbf{L}_1) - \frac{\partial f}{\partial L_k}(s_2, \mathbf{L}_1) \right) = 0 \quad \text{for all } k. \tag{2.15}$$

It follows that $(\partial f/\partial L_k)(s_1, \mathbf{L}_1)$ can be zero only if $(\partial f/\partial L_k)(s_2, \mathbf{L}_1)$ is also zero. Therefore, we cannot have $(\partial f/\partial L_k)(s_1, \mathbf{L}_1) = 0$ for all k because this would imply that $\mathbf{L}_1 = \mathbf{L}_1^*$ maximizes $f(s_2, \mathbf{L})$, which contradicts (2.11). Hence we have established

$$\mathbf{L}_1 \neq \mathbf{L}_1^*. \tag{2.16}$$

Equations (2.14) and (2.16) comprise the main result of this (two-state) asymmetric information model: The optimal second-best contract has efficient production in the good state, but inefficient production in the bad state. The intuition behind this is that if production in the bad state were efficient, an improvement could be made by perturbing \mathbf{L}_1 slightly so as to reduce $f(s_2, \mathbf{L}_1)$; this would have only a second-order effect on $f(s_1, \mathbf{L}_1)$ by the envelope theorem but would relax the truth-telling constraint with a positive multiplier (in contrast, perturbations in \mathbf{L}_2 do not relax this constraint). In fact, we will in general have a distortion in *each* of the firm’s input decisions in the bad state. For (2.15) tell us that

$$\frac{\partial f}{\partial L_k}(s_1, \mathbf{L}_1) = 0 \Rightarrow \frac{\partial f}{\partial L_k}(s_2, \mathbf{L}_1) = 0; \tag{2.17}$$

that is, L_k is undistorted in the second-best only if its marginal product does not depend on s . To put it another way, in general, *the manager’s contract with the third party will constrain every observable dimension of action that the manager takes in the bad state.*

To identify the direction of the distortion in L_k , we must put further restrictions on f . Assume that

$$\frac{\partial f}{\partial L_k}(s_2, \mathbf{L}) > \frac{\partial f}{\partial L_k}(s_1, \mathbf{L}) \quad \text{for all } \mathbf{L}; \tag{2.18}$$

that is, the marginal product of each input is higher in the good state for all \mathbf{L} . Then

$$\frac{\partial f}{\partial L_k}(s_1, \mathbf{L}_1) > 0 \quad \text{for all } k, \tag{2.19}$$

since the term after λ in (2.15) is negative, and so the first term must be positive. So (2.19) tells us that each input L_k is underemployed, given other input choices. It does not necessarily follow that $\mathbf{L}_1 < \mathbf{L}_1^*$, although

this will be so if either (i) L is one-dimensional, (ii) L is two-dimensional and $f_{12} > 0$, or (iii) f is Cobb–Douglas. In these cases, we may conclude [using also (2.14)] that *input use varies more across states in the second-best than in the first-best*.

Unfortunately, these results do not generalize easily to the case of more than two states (although there will still generally be distortions). The reason is that it becomes much harder to know which of the many truth-telling constraints will be binding. One case where progress can be made is when there is only one input and, as in (2.18), the marginal product of this input can be ranked across all states. Then only the downward truth-telling constraints are binding and the underemployment result holds. For a discussion of this case, see Hart (1983).

As we have noted, the above model emphasizes the idea of a risk-averse manager trying to get insurance against fluctuations in his net income.²⁹ In order to maintain the assumption of informational asymmetry it must be supposed that this income is private (it doesn't show up in the firm's accounts). A generalization of the above model would have part of net income observed and part not; for example, $f = f_1 + f_2$, where f_1 is the firm's profit and $-f_2$ represents the manager's effort cost in realizing this profit (or f_2 represents managerial "perks"). The only difference now is that the manager's insurance payment I can be conditioned on f_1 so that f_1 becomes like one of the observable inputs L . This case is analyzed in Holmström and Weiss (1985).

The above model completely deemphasizes the role of the long-term contract between the firm and its workers. This can be reintroduced without significant change if it is supposed that the workers, like the manager, can receive wage insurance from a third party; on this, see Hart (1983). If, however, for reasons discussed in Section 2.1, the manager has a comparative advantage in providing insurance, then matters become more complicated. The reason is that the manager is a "flawed" insurer, even if he is risk-neutral, because he has private information. As Chari (1983) and Green and Kahn (1983) have shown, this leads to a further distortion in production. For example, if $U(I, \ell) = \alpha(I) - \ell$, where ℓ is employment and $\alpha'' < 0$, the solution to (2.2)–(2.3) has $I(s) = \text{constant}$ and $\ell(s)$ increasing when the manager is risk-neutral. This, however, gives the manager an incentive always to report the highest employment state. That is, in the two-state example of this section, the manager now has an incentive to report s_2 when the true state is s_1 . To overcome this, the second-best will have $I(s)$ increasing with $\ell(s)$. In addition, in the two-state example,

²⁹ Some versions of the model assume instead that the manager is risk-neutral but cannot have negative net income [see, e.g., Farmer (1985)]. This amounts to a form of risk aversion, however, since it is equivalent to supposing that negative net income gives the manager a utility of minus infinity.

the optimal second-best contract will have the property that the *second* truth-telling constraint is binding, and (2.2) holds with *equality* in the bad state and the left-hand side of (2.2) is *less* than the right-hand side in the good state. This has been called “overemployment” in the good state, although the inequality in (2.2) does not necessarily imply that $\ell(s_2)$ is higher in the second-best than in the first-best. In fact, this overemployment result holds whenever the manager is risk-neutral, as long as $U(I, \ell)$ has the property that leisure is a normal good [see Chari (1983), Green and Kahn (1983)].

If the manager is risk-averse, this desire to over-report s comes into conflict with the desire to under-report s effect discussed above. To put it another way, the manager’s desire to obtain insurance for himself comes into conflict with his role as insurer for the workers. Which effect “wins” depends in some sense on how risk-averse the manager is in comparison with how normal leisure is [see Cooper (1983)]. One case where there is no conflict is when, as in (2.6), the cost of supplying labor comes entirely from missed outside opportunities; or, more generally, when $U(I, \ell) = \hat{U}(I - g(\ell))$.³⁰ Under these conditions, the overemployment effect disappears and we unambiguously have underemployment [see Azariadis (1983), Grossman and Hart (1981, 1983b)]. As we have noted, another case where underemployment is the outcome is when the workers can obtain income insurance elsewhere.

The fact that some asymmetric information models predict underemployment while others predict overemployment has caused some to conclude that this is not a fruitful approach for analyzing employment distortions. This seems unfortunate for several reasons. First, the models all have the property that there is ex post inefficient employment. This is of considerable interest given that most neoclassical models of the labor market – those, say, that treat it as a spot market or analyze wage–employment decisions as a symmetric information bargaining process – predict ex post efficiency. Second, the underemployment and overemployment models may not be in quite as much conflict as is sometimes thought. Because one model refers to underemployment in bad states and the other to overemployment in good states, both in fact suggest increased employment variability compared to the spot market (the word “suggest” is important here since, as we have noted, “overemployment” refers to the relative size of marginal rates of substitution and transformation rather than to differences in labor). From a macroeconomic point of view, this may be the most important conclusion. Third, the question of whether the

³⁰ In this case, where labor is neither normal nor inferior, the manager’s desire to over-report s vanishes. This follows from the fact that the solution to (2.2)–(2.3) has $I - g(L) = \text{constant}$ when the manager is risk-neutral. Since the manager is the residual income claimant, he has no incentive to misreport s .

overemployment effect is likely to dominate the underemployment effect in a particular context is one that empirical work can shed light on. Most empirical analyses of the labor market find that participation decisions of prime-age males are highly income inelastic [see Killingsworth (1983)]. This suggests that the normality of leisure effect is likely to be very small with respect to significant employment changes that are more than temporary (e.g., severances). To put it another way, outside earning opportunities are likely to swamp leisure as an opportunity cost of labor in such cases, which provides some support for the utility function $\hat{U}(I - g(\ell))$ and for the underemployment effect. On the other hand, the overemployment effect may be more relevant in the case of temporary layoffs or short-run variations in hours.³¹

Finally, mention should be made of a body of literature that considers other asymmetries of information. Some papers have analyzed the case where workers have private information about their opportunity costs [see, e.g., Kahn (1985) and Moore (1985)] while others have studied situations where firms and workers each possess some private information. This last “two-sided” case is very complex, and only limited progress has so far been made in its analysis [see, e.g., d’Aspremont and Gerard-Varet (1979); Riordan (1984); and, particularly, Moore (1984)].

2.4 *Extensions of the labor contract model*

A *Macroeconomic applications*

The original labor contract model was developed with an eye to macroeconomic applications. The discovery that employment decisions are ex post efficient [and, under (2.6), the same as in the Walrasian model] perhaps dampened enthusiasm, but the advent of the asymmetric information models has stimulated some new work in this direction.

A simple way to incorporate the model of Section 2.3 into a macroeconomic setting is to suppose that the economy consists of many identical managerial firms, with perfectly correlated demand or supply shocks s . Given (2.14) and (2.19), this would seem to give us an explanation of why an aggregate down shock would lead to a greater fall in employment in each firm than would be expected in a spot market.

Unfortunately, this is too simple. If all firms reduce employment, one would surely expect this to be observable to workers (and third parties); moreover, since presumably no firm has an influence on aggregate employment and aggregate employment is perfectly correlated with s , the asymmetry of information will disappear if payments are conditioned on this variable.

³¹ Feldstein’s (1976) work suggests that surprisingly many layoffs are in fact temporary.

Two ways of overcoming this problem have been attempted. One is to suppose that the aggregate shock causes a change in the variance of the distribution of s , as well perhaps as its mean [see Grossman, Hart, and Maskin (1983)]. Suppose that there are two states of the economy: one, α_1 , in which the variance of s is very small and the other, α_2 , in which it is large. Consider the special case where the Walrasian aggregate employment levels are the same in α_1 and α_2 . Then, under asymmetric information, a shock that moves the economy from α_1 to α_2 , *even if it is publicly observed* (say, through changes in aggregate employment), will reduce total employment. This is because the asymmetry of information will be (almost) irrelevant in the low-variance state $\alpha = \alpha_1$ (where a firm's profitability can essentially be deduced from macro variables), but will have force in the high-variance state $\alpha = \alpha_2$. Hence aggregate employment will be close to the Walrasian level in α_1 , but will be below the Walrasian level in α_2 ; the latter because low- s firms have lower employment levels under asymmetric information [by (2.19)] while high- s firms do not have higher employment levels [by (2.14)]. Together these arguments yield the conclusion that total employment will be lower in α_2 than in α_1 under asymmetric information. In fact, the same logic generalizes to show that if Walrasian aggregate employment falls when the economy is hit by a variance-increasing shock, this fall will be amplified under asymmetric information.

Farmer (1984) exploits a similar idea. Suppose that a publicly observable macroeconomic shock increases the cost of firms' inputs, for example by raising the real rate of interest. Then, although the distribution of s may not change, firms' net profits fall. If managers have decreasing absolute risk aversion, this is like an increase in managerial risk aversion. This will increase the distortion found in low- s firms (which is a function of risk aversion), without there being offsetting effects in high- s firms [by (2.14)]. Hence an aggregate increase in unemployment will again be amplified under asymmetric information.

A second approach is to suppose that s consists of a component common to all firms and an idiosyncratic component [see Holmström and Weiss (1985)]. The common component will presumably again show up in the aggregate employment figures, and so wages can be conditioned on it. But suppose these figures are published with a lag – after managers learn their s and employment decisions must be made. A low- s manager will then be unsure if his is one of many adversely affected firms (i.e., if there has been an aggregate down shock), or if he is in a minority (i.e., if he has had a bad idiosyncratic shock).³² In the first case, he will be able to reduce the wage rate (with a lag), whereas in the second case he will not (it is not incentive-compatible to allow a firm with a bad idiosyncratic

³² This is of course the same confusion that Lucas (1972) exploited.

shock to reduce wages). A risk-averse manager will put relatively high weight on the second possibility and so will cut back on employment as a second-best way of reducing the wage bill. As above, this is not compensated for by an increase in employment in high- s firms [by (2.14)]. The result can be shown to be greater aggregate employment variability between economywide up shocks and down shocks than would occur in a spot market.

The conclusion that aggregate employment levels can be inefficient raises the question of whether there is a role for government intervention. In a version of the Grossman–Hart–Maskin (1983) model, where s reflects a relative demand shock, it can be shown that a policy that stabilizes demand across different firms can be welfare-improving. This is because demand shifts have an externality effect via their impact on the extent of the asymmetry of information between firms and workers and/or third parties. Because externalities like this seem to be a fairly pervasive feature of asymmetric information/moral hazard models (see Section 1), it seems likely that there will be a role for government intervention in other models too (of course, the usual qualification that the government may require very good information to improve things should be borne in mind). Work on this topic is still in its infancy, however, and general results on the nature of macroeconomic externalities and the way to correct them are not yet available.

B *Involuntary unemployment*

We have focused on whether contract theory can explain ex post inefficient allocations. A related question which has received attention is whether the theory can explain involuntary layoffs. The results here have been rather disappointing.

To understand the issues, let us return to the ABG model, but drop the assumption of work-sharing. Instead we suppose that $\ell = 0$ or 1 for each worker at date 1. A contract will now specify a number of workers $n(s) \leq m$ who should work in state s , a payment $I_e(s)$ to each of these and a payment $I_u(s)$ to each of the laid-off workers. The total wage bill $W(s)$ in state s then equals

$$W(s) = n(s)I_e(s) + (m - n(s))I_u(s). \tag{2.20}$$

Since the firm cares only about the size of this wage bill and not how it is divided, an optimal contract must *in each state* solve:

$$\text{Max} \left[\frac{n(s)}{m} U(I_e(s), 1) + \left(1 - \frac{n(s)}{m}\right) U(I_u(s), 0) \right] \tag{2.21}$$

subject to

$$W(s) = W.$$

Here the maximand is the expected utility of each worker, given that lay-offs are chosen randomly. The first-order conditions for (2.21) are

$$\frac{\partial U}{\partial I}(I_e(s), 1) = \frac{\partial U}{\partial I}(I_u(s), 0); \quad (2.22)$$

that is, retained and laid-off workers should have the same marginal utility of income.³³ It is not difficult to show that this implies that laid-off workers are better off than retained workers if leisure is a normal good, worse off if it is inferior, and equally well off if $U = \hat{U}(I - g(\ell))$; that is, if the demand for leisure is income-inelastic.

Since it is hard to argue empirically that leisure is inferior, this model gives us the perverse result that there will be ex post involuntary retentions. Various attempts have been made to get away from this.³⁴ One approach is to drop the assumption that the utility function $U(I, \ell)$ is publicly known at date 1. For example, suppose that $U(I, \ell) = \hat{U}(I + \tilde{R}(\bar{\ell} - \ell))$, where \tilde{R} , the outside reservation wage at date 1, is a random variable. A simple case is where neither the firm nor the workers know \tilde{R} when the layoff decision is made (but both know its distribution). Under these conditions, Geanakoplos and Ito (1982) have shown that the optimal contract will involve involuntary layoffs only if \hat{U} exhibits increasing absolute risk aversion (which is usually regarded as implausible).

A second case is where \tilde{R} is known to the workers but not to the firm. If workers' \tilde{R} 's are correlated and there are many of them, it is likely that the firm will be able to elicit the common component, and so the natural case to study is where the \tilde{R} 's are independently drawn from a known distribution. In this case, however, Moore (1985) has shown that the utility function \hat{U} gives rise to involuntary *retentions* if $\ell < \bar{\ell}$.³⁵ One disturbing feature of any contract where retention is involuntary is that it gives workers an incentive to be fired. Several papers have built on this, developing models in which involuntary layoffs are part of an incentive scheme to encourage the firm's work force to work hard [see Malcomson (1984b),

³³ This is the wage bill argument of Akerlof and Miyazaki (1980). Note that the conclusion that an optimal contract must satisfy (2.21) generalizes to the case where the firm is risk-averse, because a risk-averse firm also cares only about the size and not about the division of the wage bill in a particular state.

³⁴ Azariadis was able to explain involuntary layoffs in his original (1975) paper, but only by making the arbitrary assumption that layoff pay is zero.

³⁵ A third approach is to focus on the costly search process that laid-off workers must engage in to find a new job [see, e.g., Arnott, Hosios, and Stiglitz (1985)]. It is clear that workers will not be provided with the right incentives to search if they are guaranteed a fixed income level, independently of whether they find new employment. However, since a firm can preserve incentives by giving a departing worker a lump-sum payment, it does not follow from this that laid-off workers will be worse off than retained workers. In fact the results on this are ambiguous.

Hahn (1984), and Eden (1985)]. These models may apply to situations where employers have discretion about whom to lay off, but in practice this appears usually not to be the case – in union contracts, for example, layoffs are almost always by seniority.

It should be emphasized that none of these theories explains involuntary unemployment at the contract date. The reason is that, if there is a competitive labor market at date 0, an optimal contract will have the property that each employed worker's expected utility equals \bar{U} , the market clearing level. In particular, it cannot be an equilibrium for employed workers to receive more than \bar{U} and employment to be rationed, because individual firms could then increase profit by reducing wages, $I(s)$, in each state (without distorting incentives). This conclusion is subject to some qualifications. First, it may be impossible to reduce wages in some states because workers are at the boundary of their consumption set. Second, in models involving worker effort, if a worker's utility function $U(e, I)$ is appropriately nonseparable in effort and income, then a reduction in I may have a sufficiently adverse effect on a worker's desire to work to be unprofitable for the firm [see, e.g., Malcomson (1981); a similar incentive effect underlies much of the efficiency wage literature; see, e.g., Shapiro and Stiglitz (1984)]. In both of these cases, employed workers may receive more than \bar{U} . Third, involuntary unemployment at the contract date is possible in models where there is adverse selection at date 0, an important case that falls outside the scope of this survey [see Weiss (1980) and Stiglitz and Weiss (1981)].

C *Long-term (repeated) contracts*

Labor contracts, whether implicit or explicit, have been regarded as most important in long-term relationships. To formalize these relationships as a "one-shot" situation (as in Sections 2.1–2.3) does not seem very satisfactory. Nevertheless, there are dynamic versions for which the preceding analysis applies essentially intact. A particular example that precisely fits the structure of (2.9) is in Fudenberg et al. (1986).

Consider an infinitely (and independently) repeated version of the one-period model studied above. The manager's utility over a consumption stream $\{c_t\}$ is given by $\Sigma - \delta' \exp(-rc_t)$, where δ is the discount factor and r is the manager's coefficient of absolute risk aversion. The manager can borrow and save freely at the interest rate $(1 - \delta)/\delta$. This is not observed by the principal.

As discussed in Section 1.7, an optimal long-term contract can in this situation be duplicated by a sequence of short-term contracts – with exponential utility and independent shocks a sequence of identical short-term contracts. Note, however, that an optimal one-period contract in the

dynamic model is not the same as in the static model, because the manager can smooth consumption. Instead, the one-period solution in the dynamic case is the same as if the manager worked just once, but consumed forever. (Because there are no income effects and shocks are independent, contracts across periods do not affect each other.) This program can be reduced to the form (2.9) as follows.

Assume the manager consumes after he is paid in the single period he works (this is the reverse of Fudenberg et al., but of no consequence for the decomposition result). Let w_i be the manager's net wealth if s_i occurs in that period [i.e., $w_i = f(s_i, L_i) + I_i$, $i = 1, 2$, in our earlier notation]. With no further income, the manager will consume the interest on his wealth in all future periods; that is, he will consume $(1 - \delta)w_i$ forever. This implies a lifetime utility $V(w_i) = -\exp\{-r(1 - \delta)w_i\}/(1 - \delta)$. Consequently, using this V as the manager's utility function in (2.9), we obtain the optimal short-term contract for the dynamic case.

Notice that the only difference between the static problem and the dynamic (short-term) problem is that the manager's risk-aversion coefficient is smaller in the latter. In the dynamic case the coefficient is $r(1 - \delta)$, while in the static case it is r . The reduction in risk aversion comes from self-insurance in the dynamic model. In the limit, as δ goes to one and there is no discounting of the future, the manager effectively acts in a risk-neutral fashion. One optimal (and first-best) solution in that situation is to rent out the technology to the manager and let him carry all the risk. (Recall our earlier comment on Yaari's work in Section 1.7.)

Since the introduction of dynamics in this example only changes the manager's risk-aversion coefficient, the earlier static analysis applies directly. We conclude that while there will be a smaller allocational distortion in the multi-period situation than in the one-period situation, in both cases the distortion will be qualitatively the same.

It is also worth noting that not all long-run relationships are subject to independent shocks. With serial correlation of the s_i 's, however, the gains from self-insurance may be substantially reduced. For instance, in the extreme case of a single shock that persists forever, there are no self-insurance gains at all, and the optimal long-term contract will be the repeated static contract from (2.9). More generally, with positive correlation, repetition will have a smaller effect in reducing the level of second-best inefficiency than with independent shocks (but see Section 2.4D below).

D *Enforcement of the contract*

The asymmetric information contract models are sometimes criticized on the ground that "while the parties may agree in advance to have unemployment in bad states of the world, they will surely change their mind

once such a state is realized.” To understand this, consider a firm that signs a contract with a single worker. Suppose that $\ell = 0$ or 1, that there are two states $s = s_2$ or $s = s_1$ ($s_2 > s_1$), and that the ex post opportunity cost of labor is zero. An optimal second-best contract might have the property that $\ell = 1$ when $s = s_2$ and $\ell = 0$ when $s = s_1$. But suppose now that $s = s_1$ is realized and the firm lays off the worker. Then the argument goes that the firm and worker will recontract at this stage since they will recognize that there are some unexploited gains from trade (assuming that $s_1 > 0$).

Such recontracting can only make the parties worse off in ex ante terms (assuming it is anticipated) – otherwise the original contract would not have been an optimal one. The question therefore is whether the parties can precommit themselves not to renegotiate; in the static model of Section 2.3, the answer seems to be yes. Presumably there is a last moment at which employment decisions must be made. Let the original contract state that the firm can change its mind about whether to employ the worker up to this last moment. Then any threat to lay off the worker before the last moment is not credible because the worker knows that the firm can costlessly change its mind, while a threat at the last moment is, of course, useless to the firm since by that time it is too late to renegotiate.

The recontracting criticism does not therefore seem to be valid when there is only one employment date. However, it does have force in a dynamic context. Change the above example so that the worker can work or not work on each of T days (but suppose, in contrast to Section 2.4C, that the shock s is the same for all days). The optimal second-best contract might call for the worker to be laid off for $1 < T_1 \leq T$ days in the bad state $s = s_1$. However, it is hard to see what is to stop the parties from renegotiating such a contract after one day of unemployment, given that there are clearly unexploited gains from trade at this point and that the only irreversible decision which has been made concerns the first day’s layoff.³⁶

In future work it would be interesting to investigate the constraints that such renegotiation puts on dynamic contracts.³⁷ We shall return to the issue of ex post renegotiation in Section 3.

³⁶ A similar phenomenon arises in a dynamic bargaining context where a seller would like to commit himself to make a single take-it-or-leave-it offer to a buyer, but cannot do so since he cannot constrain himself not to make a second offer if his first offer is rejected. See, e.g., Fudenberg and Tirole (1983). Note that there is a fundamental difference between all the parties agreeing to tear up the contract and one party repudiating the contract – something which we have implicitly assumed never occurs, e.g. because the resulting damage payment is so large.

³⁷ A start on this has been made by Dewatripont (1985).

2.5 *Summary and conclusions*

There seem to be two major conclusions from the labor contract literature. First, in an optimal contract, there will be systematic discrepancies between wages and the marginal product of labor. Second, under asymmetric information, there will be ex post inefficiencies.

Both these conclusions have important implications for the way we think about labor markets. In almost all empirical work on the labor market, for instance, it is taken for granted that wages measure the opportunity cost of labor, and that firms will be on their demand curves or workers on their supply curves or both. In a contracting framework, as Rosen (1985) has stressed, none of these suppositions is valid. To take another example, it is often assumed that the following is a good model of union behavior: The union chooses the wage rate to maximize the representative worker's utility subject to the constraint that the firm will be on its labor demand curve. According to the contracting framework, however, such behavior is irrational because both the firm and the union can make themselves better off by agreeing on a wage-employment pair that lies on the efficiency frontier.

In view of the strong implications of the labor contract approach, it is important to know how well the theory matches up with the facts. Serious econometric work on this topic is only just beginning, but some interesting papers by Brown and Ashenfelter (1986) and by Card (1985) are already available. These papers test the prediction of the ABG model that ex post employment levels can be explained by opportunity costs rather than actual wages [as in (2.7)]. The results obtained so far suggest little support for this hypothesis, but it is possible that some of the explanatory power of actual wages found by Brown and Ashenfelter and by Card can be traced to asymmetries of information (as in the model of Section 2.3), rather than being a rejection of the optimal contracting approach per se. Unfortunately, testing the asymmetric information contract model directly is a very difficult task, and we are not aware of any attempts so far in this direction.

A much less formal empirical approach, which has been adopted by Oswald (1984), is to examine actual labor contracts to see whether they contain the features that one might expect from the theory. The results here have again been less than favorable to the contracting approach. First, most non-union contracts are surprisingly rudimentary, sometimes consisting of as little as a verbal statement that an employee has a job at a particular (current) wage. Second, union contracts, although frequently lengthy and complex, do not contain a number of the provisions that the theory suggests they should. For example, it is rare to find joint agree-

ments on wages and employment; typically wage rates are specified over the course of the contract, but employment decisions are left to the firm. [Although this is not inconsistent with the model of Section 2.3, in more general asymmetric information models (where, e.g., workers and firms both have private information) an optimal contract will involve joint determination of employment by firms and workers.] Other anomalies are the lack of indexation of wages to retail prices or to variables such as firm employment or firm sales, and the limited provisions for layoff pay.

Of course, one possible escape for the contract theorist is to argue that whatever does not appear in the explicit contract is simply part of an implicit contract (see the introduction). This is akin to the proposition that a theory should be judged by its predictions (e.g., whether employment levels are determined solely by opportunity costs) rather than by its assumptions (whether a particular contractual provision is physically present). Although there is surely something to this idea, it seems a considerable act of faith to rely on the notion of an implicit contract, given that so little is presently known about how implicit agreements are enforced (but see Section 3.4). In fact, in view of the current ignorance about this, it seems curious – and unfortunate – that the whole field often goes under the name of Implicit Labor Contracts.

Given that empirical support for the labor contract model is at present rather limited, the question arises whether the contracting approach is worth pursuing. Not surprisingly, we feel strongly that the answer is yes. The main reason is that there appears to be no serious alternative available for analyzing this class of problems. For example, the wage-setting-union model described above may fit some of the facts better, but it is based on the assumption that the parties fail to exploit all the gains from trade, which (in theoretical terms) seems unacceptable. Rather than abandoning the contracting framework, therefore, it seems desirable to try to modify it so as to make it more realistic, for example by incorporating further moral hazards or asymmetries of information or – and perhaps this is most important – by introducing the costs of writing contracts (see Section 3). It should also be noted that firm/worker relationships are only one application of the contracting framework. In Section 3, we argue that other applications (e.g., to input supply contracts between firms) may in the long run be at least as fruitful, as well perhaps as being more consistent with the facts.

2.6 Appendix

It is easy to show that, if f is continuous, a solution to (2.9) exists. Denote it by $(\hat{L}_1, \hat{L}_2, \hat{I}_1, \hat{I}_2)$. Clearly, $\pi_2 \hat{I}_2 + \pi_1 \hat{I}_1 = 0$. Furthermore, at least

one of the truth-telling constraints must be binding (otherwise a Pareto improvement could be made by moving in the direction of the first-best). We consider three cases.

Case 1 (both truth-telling constraints are binding):

$$\begin{aligned} f(s_2, \hat{\mathbf{L}}_2) + \hat{I}_2 &= f(s_2, \hat{\mathbf{L}}_1) + \hat{I}_1, \\ f(s_1, \hat{\mathbf{L}}_1) + \hat{I}_1 &= f(s_1, \hat{\mathbf{L}}_2) + \hat{I}_2. \end{aligned} \tag{1}$$

In this case the manager is indifferent between the two states. Hence $\hat{I}_1 = \hat{I}_2 = 0$ since, if $\hat{I}_i < \hat{I}_j$, a Pareto improvement could be achieved by replacing the old contract by a new contract $(\hat{\mathbf{L}}_i, \hat{I}_i, \hat{\mathbf{L}}_j, \hat{I}_j)$. But, if $\hat{I}_1 = \hat{I}_2 = 0$, it is optimal to set \mathbf{L}_i equal to its first-best value, \mathbf{L}_i^* , $i = 1, 2$, which contradicts (1). Therefore Case 1 is impossible.

Case 2 (only the second truth-telling constraint is binding):

$$\begin{aligned} f(s_1, \hat{\mathbf{L}}_1) + \hat{I}_1 &= f(s_1, \hat{\mathbf{L}}_2) + \hat{I}_2, \\ f(s_2, \hat{\mathbf{L}}_2) + \hat{I}_2 &> f(s_2, \hat{\mathbf{L}}_1) + \hat{I}_1. \end{aligned} \tag{2}$$

The second inequality, together with (2.10), implies that $f(s_2, \hat{\mathbf{L}}_2) + \hat{I}_2 > f(s_1, \hat{\mathbf{L}}_1) + \hat{I}_1$; that is, the manager prefers s_2 to s_1 . In this case, however, by a standard risk-sharing argument, a Pareto improvement can be made by lowering I_2 and raising I_1 , keeping $\pi_2 I_2 + \pi_1 I_1$ constant (the truth-telling constraints will continue to be satisfied). Hence Case 2 is ruled out.

We are left with Case 3, where only the first truth-telling constraint is binding; this case was analyzed in the text.

3 **Incomplete contracts**

3.1 *The benefits of writing long-term contracts*

The literature on labor contracts focuses on income-shifting as the motivation for a long-term contract; that is, on the gains the parties receive from transferring income from one state of the world or one period to another. In the ABG model, the worker wants to insure his income. This is also the case in the quit model, where in addition the worker wants to smooth his consumption over time. Finally, in the Azariadis–Grossman–Hart model, it is the entrepreneur/manager who desires insurance.

In all these models, the rationale for the contract would disappear if the agents were risk-neutral and faced perfect capital markets. Even if risk aversion and imperfect capital markets are present, the ABG and

Holmström explanations of labor contracts rely on the assumption that firms have a comparative advantage in providing insurance and income-smoothing opportunities to workers.

It is perhaps unfortunate that so much attention has been devoted to “financial” contracts of this type. As we noted in the introduction, a fundamental reason for long-term relationships is the existence of investments that are to some extent party-specific. Although this lock-in effect is often used to motivate the long-term relationship between workers and firms in labor contract models, it then tends to be ignored. Yet this lock-in effect can explain the existence and characteristics of long-term contracts even in the presence of risk neutrality and perfect capital markets. Moreover, in the case (say) of supply contracts involving large firms, risk neutrality and perfect capital markets may be reasonable assumptions in view of the many outside insurance and borrowing/lending opportunities available to such parties.

The importance of a long-term contract when there are relationship-specific investments can be seen from the following example [based on Grout (1984); see also Crawford (1982)]. Let B and S be, respectively, the buyer and seller of (one unit of) an input. Suppose that in order to realize the benefits of the input, B must make an investment a which is specific to S ; for example, B might have to build a plant next to S . Assume that there are just two periods: The investment is made at date 0, while the input is supplied and the benefits are received at date 1. S 's supply cost at date 1 is c , while B 's benefit function is $b(a)$ (all costs and benefits are measured in date-1 dollars).

If no long-term contract is written at date 0, the parties will determine the terms of trade from scratch at date 1. If we assume that neither party has alternative trading partners at date 1 then there is, given B 's sunk investment cost a , a surplus of $b(a) - c$ to be divided up. A simple assumption to make is that the parties split this 50:50 (this is the Nash bargaining solution). That is, the input price p will satisfy $b(a) - p = p - c$. This means that the buyer's overall payoff, net of his investment cost, is

$$b(a) - p - a = \frac{b(a) - c}{2} - a. \quad (3.1)$$

The buyer, anticipating this payoff, will choose a to maximize (3.1), that is, to maximize $\frac{1}{2}b(a) - a$.

This is to be contrasted with the efficient outcome, where a is chosen to maximize total surplus, $b(a) - c - a$. Maximizing (3.1) will lead to underinvestment; in fact, in extreme cases, a will equal zero and trade will not occur at all. The inefficiency arises because the buyer does not receive the full return from his investment – some of this return is appropriated

by the seller in the date-1 bargaining. Note that an up-front payment from S to B at date 0 (to compensate for the share of the surplus that S will receive later) will not help here, because it will only change B 's objective function by a constant (it's like a lump-sum transfer). That is, it redistributes income without affecting real decisions.

Efficiency can be achieved if a long-term contract is written at date 0 specifying the input price p^* in advance. Then B will maximize $b(a) - p^* - a$, yielding the efficient investment level a^* . An alternative method is to specify that the buyer must choose $a = a^*$ (if not, he pays large damages to S); the choice of p can then be left until date 1, with an up-front payment by S being used to compensate B for his investment. The second method presupposes that investment decisions are publicly observable, and so in practice may be more complicated than the first (see Section 3.3).

This example – formalizing arguments contained in Williamson (1975, 1985) and in Klein, Crawford, and Alchian (1978) – illustrates the role of a long-term contract when there are relationship-specific investments. The word “investment” should be interpreted broadly; the same factors will apply whenever one party is forced to pass up an opportunity as a result of a relationship with another party (e.g., A 's “investment” in the relationship with B may be not to lock into C). That is, the crucial element is a sunk cost (direct or opportunity) of some sort. (An effort decision is one example of a sunk cost.) Note that the income-transfer motive for a long-term contract is completely absent here; there is no uncertainty and everything is in present-value terms.

In spite of their importance, the analysis of “real” contracts of this type (“real,” rather than “financial,” because their rationale comes from the existence of real decisions such as investments) is in its infancy. A notable early reference is Becker's (1964) analysis of worker training. More recently, Williamson (1975, 1985) and Klein, Crawford, and Alchian (1978) have emphasized the difficulty of writing contracts that induce efficient relationship-specific investments as an important factor in explaining vertical integration.

Section 3 of this survey will be concerned primarily with the analysis of such real contracts. At this stage, however, it may be useful to summarize the general benefits of writing long-term contracts. We have discussed the income-transfer and “real” motives. Let us note three further benefits. First, if a relationship is repetitive, it may save on transaction costs to decide in advance what actions each party should take rather than to negotiate a succession of short-term contracts. Second, if asymmetries of information arise during the course of the relationship, letting the parties negotiate as they go along may lead to ex post bargaining inefficiencies

[as in, e.g., Fudenberg and Tirole (1983)], which can be avoided by a long-term contract.³⁸ Third, a long-term contract may be useful for screening purposes – for example, a firm may attract a productive worker by offering a high future reward in the event that the worker is successful. [This is an example drawn from the adverse selection literature; see, e.g., Salop and Salop (1976).]

Given the many advantages of long-term contracts, the question that obviously arises is why we don't see more of them, and why those we do see seem often to be limited in scope. To this question we now turn.

3.2 *The costs of writing long-term contracts*

Contract theory is sometimes dismissed because “we don't see the long-term contingent contracts that the theory predicts.” In view of the benefits of long-term contracts, this statement (even if true) needs to be explained.

The first point to make is that there is no shortage of complex long-term contracts in the world. Joskow (1985), for example, in his recent study of transactions between electricity generating plants and mine-mouth coal suppliers, finds that some contracts between the parties extend for fifty years, and a large majority for over ten years. The contractual terms include quality provisions, formulas linking coal prices to costs and prices of substitutes, indexation clauses, and so forth. The contracts are both complicated and sophisticated. Similar findings are contained in Goldberg and Erickson's (1982) study of petroleum coke.

At a much more basic level, a typical contract for personal insurance, with its many conditions and exemption clauses, is not exactly a simple document. Nor for that matter is a typical house rental agreement. On the other hand, as we noted in Section 2, labor contracts are often surprisingly rudimentary, at least in certain respects.

Given that complex long-term contracts are found in some situations but not others, it is natural to explain any observed contract as an outcome of an optimization process in which the relative benefits and costs of additional length and complexity are traded off at the margin. We have

³⁸ For example, suppose that B does not have to make any investment, but that his benefit from the input is stochastic: $b = 10$ with probability $1/2$ and 3 with probability $1/2$. Assume that B learns the exact value of b at date 1 while S does not, that $c = 0$ for sure, and that both parties are risk-neutral. Then if bargaining occurs from scratch at date 1, and S has the power to make take-it-or-leave-it offers, he will set a price of 10 (obviously S will not find it profitable to set a price other than 10 or 3; the price of 10 gives him higher expected profit). But this means that a mutually beneficial trade will not be made in the event $b = 3$. On the other hand, the first-best can be achieved by a long-term contract specifying that the buyer can insist on supply of the input in all circumstances at some predetermined price.

given some indication of the determinants of the benefits of length and complexity. But what about the costs? These are much harder to pin down because they fall under the general heading of “transactions costs,” a notoriously vague and slippery category. Of these, the following seem to be important: (1) the cost to each party of anticipating the various eventualities that may occur during the life of the relationship; (2) the cost of deciding, and reaching an agreement about, how to deal with such eventualities; (3) the cost of writing the contract in a sufficiently clear and unambiguous way so that the terms of the contract can be enforced; and (4) the legal cost of enforcement.

One point to note is that all these costs are present also in the case of short-term contracts, although presumably they are usually smaller. In particular, since the short-term future is more predictable, the first cost is likely to be much reduced, and so possibly is the third. However, it certainly is not the case that there is a sharp division between short-term contracts and long-term contracts, with (as is sometimes supposed) the former being costless and the latter being infinitely costly.

It is also worth emphasizing that, when we talk about the cost of a long-term contract, we are presumably referring to the cost of a “good” long-term contract. There is rarely significant cost or difficulty in writing *some* long-term contract. For example, the parties to an input supply contract could agree on a fixed price and level of supply for the next fifty years; they do not, because (presumably) such a rigid arrangement would be very inefficient.³⁹

Due to the presence of transaction costs, the contracts people write will be incomplete in important respects. The parties will quite rationally leave out many contingencies, taking the point of view that it is better to wait and see what happens than to try to cover a large number of individually unlikely eventualities. Less rationally, the parties will leave out other contingencies that they simply do not anticipate. Instead of writing very long-term contracts the parties will write limited-term contracts, with the intention of renegotiating these when they come to an end. Contracts will often contain clauses that are vague or ambiguous, sometimes fatally so.

³⁹ In some cases, the courts will not enforce such an agreement, taking the point of view that the parties could not really have intended it to apply unchanged for such a long time. A clause to the effect that the parties really do mean what they say should be enough to overcome this difficulty, however. In other cases, it may be impossible to write a binding long-term contract because the identities of some of the parties involved may change. For example, one party may be a government that is in office for a fixed period, and it may be impossible for the government to bind its successors. This latter idea underlies the work of Kydland and Prescott (1977) and of Freixas, Guesnerie, and Tirole (1985).

Anyone familiar with the legal literature on contracts will be aware that almost every contractual dispute that comes before the courts concerns a matter of incompleteness. In fact, incompleteness is probably at least as important empirically as asymmetric information as an explanation for departures from “ideal” Arrow–Debreu contingent contracts. In spite of this, relatively little work has been done on this topic, the reason (presumably) being that an analysis of transaction costs is so complicated. One problem is that the first two transaction costs referred to above are intimately connected to the idea of bounded rationality [as in Simon (1982)], a concept that has not yet been successfully formalized. As a result, perhaps, the few attempts that have been made to analyze incompleteness have concentrated on the third cost, the cost of writing the contract.

One approach, due to Dye (1985), can be described as follows. Suppose that the amount of input q traded between a buyer and seller should be a function of the product price p faced by the buyer: $q = f(p)$. Writing down this function is likely to be costly. Dye measures the costs in terms of how many different values q takes on as p varies; in particular, if $\#\{q \mid q = f(p) \text{ for some } p\} = n$ then the cost of the contract is $(n - 1)c$, where $c > 0$. This means that a noncontingent statement “ $q = 5$ for all p ” has zero cost; the statement “ $q = 5$ for $p \leq 8$, $q = 10$ for $p > 8$ ” has cost c ; and so forth.

The costs Dye is trying to capture are real enough, but the measure used has some drawbacks. It implies, for example, that the statement “ $q = p^{1/2}$ for all p ” has infinite cost if p has infinite domain, and does not distinguish between the cost of a simple function like this and the cost of a much more complicated function. As another example, a simple indexation clause to the effect that the real wage should be constant (i.e., the money wage = λp for some λ) would never be observed because, according to Dye’s measure, it too has infinite cost. In addition, the approach does not tell us how to assess the cost of indirect ways of making q contingent; for example, the contract could specify that the buyer (having observed p) can choose any amount of input q he likes, subject to paying the seller σ for each unit.

There is another way of getting at the cost of including contingent statements. This is to suppose that what is costly is describing the state of the world ω rather than writing a statement per se. That is, suppose that ω cannot be represented simply by a product price, but is very complex and of high dimension – for example, it includes the state of demand, what other firms in the industry are doing, the state of technology, and so on. Many of these components may be quite nebulous. To describe ω ex ante in sufficient detail for an outsider (e.g., the courts) to verify that a partic-

ular state $\omega = \hat{\omega}$ has occurred, and so enforce the contract, may be prohibitively costly. Under these conditions, the contract will have to omit some (in extreme cases, all) references to the underlying state.

Similar to this is the case where what is costly is describing the characteristics of what is traded, or the actions (e.g., investments) the parties must take. For example, suppose that there is only one state of the world, but that q now represents the quality of the item traded rather than the quantity. An ideal contract would give a precise description of q . However, quality may be multidimensional and very difficult to describe unambiguously (and vague statements to the effect that quality should be “good” may be almost meaningless). The result may be that the contract will have to be silent on many aspects of quality and/or actions.

Models of this sort of incompleteness have been investigated by Grossman and Hart (in press) and Hart and Moore (1985) for the case where the state of the world cannot be described, and by Bull (1985) and Grossman and Hart (1986; in press) for the case where quality and/or actions cannot be specified. These models do not rely on any asymmetry of information between the parties. Both parties may recognize that the state of the world is such that the buyer’s benefit is high or the seller’s cost is low, or that the quality of an item is good or bad or that an investment decision is appropriate or not. The difficulty is conveying this information to others; that is, *it is the asymmetry of information between the parties on the one hand, and outsiders (such as the courts) on the other hand, which is the root of the problem.*

To use the jargon, incompleteness arises because states of the world, quality, and actions are *observable* (to the contractual parties) but not *verifiable* (to outsiders). We describe an example of an incomplete contract along these lines in Section 3.3.

3.3 *Incomplete contracts: an example*

We will give an example of an incomplete contract for the case where it is prohibitively costly to specify the quality characteristics of the item to be exchanged or the parties’ investment decisions. Similar problems arise when the state of the world cannot be described. The example is a variant of the models in Grossman and Hart (1986; in press) and in Hart and Moore (1985).

Consider a buyer B who wishes to purchase a unit of input from a seller S . B and S each make a (simultaneous) specific investment at date 0 and trade occurs at date 1. Let I_B, I_S denote (respectively) the investments of B and S , and to simplify assume that each can take on only two values, H or L (high or low). These investments are observable to B and S , but are

not verifiable (they are complex and multidimensional, or represent effort decisions) and hence are noncontractible. We assume that at date 1 the seller can supply either satisfactory input or unsatisfactory input. “Unsatisfactory” input has zero benefit for the buyer and zero cost for the seller (so it’s like not supplying at all). “Satisfactory” input yields benefits and costs that depend on ex ante investments. These are indicated by

$$I_B = H \quad I_B = L$$

| | | |
|-----------|---------|---------|
| $I_S = H$ | (10, 6) | (9, 7) |
| $I_S = L$ | (9, 7) | (6, 10) |

The first component refers to the buyer’s benefit v , and the second to the seller’s cost c . So if $I_S = H$ and $I_B = H$ then $v = 10$ and $c = 6$ (if input is satisfactory). From these gross benefits and costs must be subtracted investment costs, which we assume to be 1.9 if investment is high and zero if it is low (for each party). (All benefits and costs are in date-1 dollars.) Note that there is no uncertainty and so attitudes to risk are irrelevant.

Our assumption is that the characteristics of the input (e.g., whether it is satisfactory) are observable to both parties, but are too complicated to be specified in a contract. The fact that they are observable means that the buyer can be given the option to reject the input at date 1 if he does not like it. This will be important in what follows.

An important feature of the example is that the seller’s investment affects not only the seller’s costs but also the buyer’s benefit and the buyer’s investment affects not only the buyer’s benefit but also the seller’s cost. The idea here is that a better investment by the seller increases the quality of satisfactory input; and a better investment by the buyer reduces the cost of producing satisfactory input – that is, input that can be used by the buyer.

For instance, one can imagine that B is an electricity generating plant and S is a coal mine that the plant sites next to. I_B might refer to the type of coal-burning boiler that the plant installs and I_S to the type of mine the coal supplier develops. By investing in a better boiler, the power plant may be able to burn lower quality coal, thus reducing the seller’s costs, while still increasing its gross (of investment) profit. On the other hand, by developing a good seam, the mine may raise the quality of coal supplied while reducing its variable cost.

The first-best has $I_B = I_S = H$, with total surplus equal to $10 - 6 - 3.8 = .2$ (if $I_B = H$ and $I_S = L$ or vice versa then surplus = .1; if $I_B = I_S = L$ then no trade occurs and surplus is zero). This could be achieved if either investment or quality were contractible as follows. If investment is con-

tractible, an optimal contract would specify that the buyer must set $I_B = H$ and the seller $I_S = H$ and give the buyer the right to accept the input at date 1 at price p_1 or reject it at price p_0 . If $10 > p_1 - p_0 > 6$, the seller will be induced to supply satisfactory input (the gain $p_1 - p_0$ from having the input accepted exceeds the seller's supply cost) and the buyer to accept it (the buyer's benefit exceeds the incremental price $p_1 - p_0$). If, on the other hand, quality is contractible, then the contract could specify that the seller must supply input with the precise characteristics that make it satisfactory when $I_B = I_S = H$. Each party would then have the socially correct investment incentives because, with specific performance, neither party's investment affects the other's payoff (there is no externality).

We now show that the first-best cannot be achieved if investment and quality are both noncontractible. A second-best contract can make price a function of any variable that is verifiable. Investment and quality are not verifiable (nor is v or c), but we shall suppose that whether the item is accepted or rejected by the buyer is verifiable; thus the contract can specify an acceptance price p_1 and a rejection price p_0 . In fact, p_0, p_1 can also be made functions of (verifiable) messages that the buyer and seller send each other, reflecting the investment decisions that both have made [as in Hart and Moore (1985)]. The following argument is unaffected by such messages and so, for simplicity, we ignore them (but see note 44).

Can we sustain the first-best by an appropriate choice of p_0, p_1 ? The seller always has the option of choosing $I_S = L$ and producing an item of unsatisfactory quality, which yields him a net payoff of p_0 . In order to induce him not to do this, we must have

$$p_1 - 6 - 1.9 \geq p_0 \quad (\text{i.e., } p_1 - p_0 \geq 7.9). \quad (3.2)$$

Similarly, the buyer's net payoff must be no less than $-p_0$ since he always has the option of choosing $I_B = L$ and rejecting the input. That is,

$$10 - p_1 - 1.9 \geq -p_0 \quad (\text{i.e., } p_1 - p_0 \leq 8.1). \quad (3.3)$$

Hence $(p_1 - p_0)$ must lie between 7.9 and 8.1.

Now the seller has an additional option. If he expects the buyer to set $I_B = H$, he can choose $I_S = L$ and, given that $8.1 \geq p_1 - p_0 \geq 7.9$, still be confident that trade of satisfactory input will occur under the original contract at date 1 (the buyer will accept satisfactory input since $v = 9 > p_1 - p_0$, while the seller will supply it since $p_1 - p_0 > 7 = c$). But if the seller deviates, his payoff rises from $p_1 - 6 - 1.9$ to $p_1 - 7$. (The example is symmetric and so a similar deviation is also profitable for the buyer.) Hence the equilibrium of $I_B = I_S = H$ will be disrupted.

We see, then, that the first-best cannot be sustained if investment and quality are both noncontractible. The reason is that it will be in the interest

of the seller (respectively, the buyer) to reduce investment because, although this reduces social benefit by lowering the buyer's (seller's) benefit, it increases the seller's (buyer's) own profit. The optimal second-best contract will instead have $I_B = H$, $I_S = L$ (or vice versa), which will be sustained by a pair of prices p_0, p_1 such that $9 > p_1 - p_0 > 7$. Total surplus will be .1 instead of the first-best level of .2.⁴⁰

The conclusion is that inefficiencies can arise in incomplete contracts even though the parties have common information (both observe investments and both observe quality). The particular inefficiency that occurs in the model analyzed is in *ex ante* investments. *Ex post* trade is always efficient relative to these investments since p_1, p_0 can and will be chosen so that $v > p_1 - p_0 > c$; that is, the seller wants to supply and the buyer to receive satisfactory input. The example can be regarded as formalizing the arguments of Williamson (1975, 1985) and of Klein, Crawford, and Alchian (1978) that relationship-specific investments will be distorted due to the impossibility of writing complete contingent contracts – note that this result is achieved without imposing arbitrary restrictions on the form of the permissible contract.⁴¹

The above example can be modified to illustrate an interesting possibility that can arise in an incomplete contract. Suppose we change the $I_S = H$, $I_B = L$ payoffs from (9, 7) to (9, 8.2) and the $I_S = L$, $I_B = H$ payoffs to (7.8, 7). The first-best stays the same. But now it can be sustained *as long as renegotiation of the contract is impossible at date 1*. In particular, choose $p_1 - p_0 = 8$. Then if either the buyer or seller deviates from the first-best, $v \geq p_1 - p_0 \geq c$ will be violated and so the deviating party's profit will fall to p_0 (for the seller) or $-p_0$ (for the buyer).

⁴⁰ It is worth pointing out why we have assumed that both the buyer and seller make investments. If only the buyer (respectively, the seller) invests, the first-best can be achieved by choosing $p_1 - p_0$ between 6 and 7 (9 and 10): Any deviation by the buyer (seller) will then be unprofitable since it will lead to no trade. This argument depends on the assumption of no renegotiation of the contract at date 1, an issue we deal with below. However, even if renegotiation is allowed, the first-best can be achieved with one-sided investment by a contract that fixes p_0 but gives the investing party the power to choose any p_1 he wants. This party then faces the social net benefit function because he extracts all the surplus.

⁴¹ The inefficiency that we have identified may not seem so surprising given that our model resembles that found in the literature on moral hazard in teams [see, e.g., Holmström (1982a)]. In that literature, each agent takes a private action that affects total benefits; in our model, investment decisions have this property. However, there are some differences between the frameworks. First, in our context, the agents observe each other's actions. Second, the externality in investments only materializes in the event that trade occurs, and so the terms of trade can be used to mitigate the externality. In any case, our purpose is not the development of a new model, but rather the application of it to a new context – the analysis of the consequences of incomplete contracting.

However, the first-best may not be sustainable if renegotiation is possible. The previous argument showing that $7.9 \leq p_1 - p_0 \leq 8.1$ still applies. Without loss of generality, set $p_0 = 0$ in the following. Suppose the seller chooses $I_S = L$, while $I_B = H$. Then at date 1, the parties will realize that, because $7.8 = v < p_1$, trade – although mutually beneficial – will not occur under the original contract (the buyer will reject the input). Hence they will presumably lower the price p_1 to lie between 7 and 7.8. But as long as the new price $p'_1 > 7.2$, the seller's net payoff will be higher than if he does not deviate (because his first-best surplus $p_1 - 7.9 < 0.2$). Hence the seller will deviate unless his power to keep p_1 up in the renegotiation is rather limited. (If the parties split the gains from renegotiation 50:50, $p'_1 = 7.4$ and the seller will certainly deviate.)⁴² If the seller is a poor bargainer, however, the buyer will presumably deviate; that is, he will set $I_B = L$, anticipating that $I_S = H$ – the parties will then agree to raise the price p_1 to lie between $c = 8.2$ and $v = 9$, and the buyer's net payoff will rise as long as the new price $p''_1 < 8.8$.

In this modified example, then, the buyer and seller can do better if they can precommit themselves not to renegotiate the contract! We have encountered this possibility before in Section 2.4D, but note that the method proposed there for preventing renegotiation (in the static case) will not work here (that method depended on the worker not knowing what the firm was going to do until the last moment, whereas here both parties will recognize the need for renegotiation as soon as investment decisions are made). Simply putting in renegotiation penalties in the original contract (e.g., the buyer must pay the seller a million dollars if there is renegotiation) is unlikely to be effective because the parties can always agree to rescind the old contract, thereby voiding the penalties [see Schelling (1960)].⁴³

⁴² In fact, Hart and Moore (1985) give an argument that the seller will be strongly advantaged in a renegotiation involving a price decrease, and that $p'_1 = p_0 + 7.8$.

⁴³ The inclusion of a third party in the contract – with the initial two parties promising to pay the third party a large sum of money if they ever renegotiate – also does not overcome the problem because, if there are ex post gains from renegotiation, the third party can be persuaded at date 1 to give up his claim to this large sum in exchange for a side-payment. The inclusion of a third party may help, however, to the extent that it makes renegotiation more costly; e.g., because it is known that the third party will be unavailable at a crucial moment during the renegotiation process.

It should be noted that third parties have uses beyond their ability to make renegotiation more difficult. A third party can act as a financial wedge between the initial contracting parties, so that the amount the seller receives in a particular state (p_1 or p_0) differs from the amount the buyer pays, with the third party making up the difference. Also, whenever actions or states are observable but not verifiable, it may be possible to get the initial parties to reveal their information to outsiders by inducing them to make reports to the third party, with a penalty due if their reports don't match (equilibria other than

If renegotiation cannot be prevented, the condition that there are no ex post Pareto improvements from recontracting must be imposed as a constraint in the original contract [as in Hart and Moore (1985); recall that, in the present context, information is symmetric and so it is clear what a Pareto improvement is]. We have already noted that such a constraint may be important in dynamic asymmetric information labor contract models, and it seems to apply to other contexts too. For instance, a firm may wish to convince a customer that it is not going to reduce its price in the future, but a binding contract to that effect may be infeasible since the firm and customer know that they will agree to rescind it at a later date (an agreement not to raise the price, on the other hand, may not suffer from the same difficulty). Other examples in the same spirit may be found in Schelling's (1960) interesting discussion of the difficulties of making commitments.⁴⁴

Returning to our example, we may illustrate a theory of ownership presented in Grossman and Hart (1986; in press). It is sometimes suggested that when transaction costs prevent the writing of a complete contract, there may be a reason for firm integration [see Williamson (1985)]. Consider the payoffs of our diagram (page 135) and suppose that *B* takes over *S*. The control that *B* thereby gains over *S*'s assets may allow *B* to affect *S*'s costs in various ways, and this may reduce the possibility of opportunistic behavior by *S*. To take a very simple (and contrived) example, suppose that if *S* chooses $I_S = L$ then *B* can take some action α with respect to *S*'s assets at date 1 so as to make *S*'s cost of supplying either satisfactory or unsatisfactory input equal to 9. (In the coal–electricity example, α might refer to the part of the mine's seam the coal is taken out of; note that we now drop the assumption that the cost of supplying unsatisfactory

the truth-telling one may be a problem here). A difficulty with either of these arrangements is that there may be a great incentive for two of the three parties to collude; e.g., one of the initial two parties can deliberately report the wrong information, having (secretly) agreed with the third party to divide up the penalty that will result. If such collusion is possible, it can be shown in the present context that a three-party contract offers no advantage over a two-party one [see Hart and Moore (1985), Eswaran and Kotwal (1984)].

⁴⁴ In our earlier discussion we mentioned, but did not analyze, the possibility that the parties might send (verifiable) messages to each other at date 1, reflecting their jointly observable investment decisions, with the contract specifying how final prices p_1 and p_0 should depend on these messages. It should be noted that the use of such messages does not allow the first-best to be achieved in the original example, at least if renegotiation is possible. This is because if $v = 9$ and $c = 7$ then trade will occur at date 1 at some price (say, p'_1) that depends on the messages sent. Hence to make a deviation from the first-best $I_B = I_S = H$ unprofitable for the buyer, we must have $10 - p_1 - 1.9 \geq 9 - p'_1$, where p_1 is the trading price when $v = 10$ and $c = 6$. On the other hand, to make it unprofitable for the seller, we must have $p_1 - 6 - 1.9 \geq p'_1 - 7$. These inequalities are inconsistent.

input is zero.) Imagine furthermore that this action increases B 's benefit, so that B will indeed take it at date 1 if S chooses L . Then with this extra degree of freedom, the first-best can be achieved. In particular, if $p_1 = p_0 + 6.1$ then $I_S = I_B = H$ is a Nash equilibrium because (by the above reasoning) any seller deviation will be punished, whereas a buyer deviation will induce the seller to supply unsatisfactory input, given that $p_1 < p_0 + 7$.⁴⁵

Note that if action α could be specified in the initial contract then there would be no need for integration: The initial contract would simply say that B has the right to choose α at date 1. Ownership becomes important, however, if (i) α is too complicated to be specified in the date-0 contract and therefore qualifies as a residual right of control; and (ii) residual rights of control over an asset are in the hands of whomever owns that asset. The point is that, under incompleteness, the allocation of residual decision rights matters because the contract cannot specify precisely what each party's obligations are in every state of the world. To the extent that ownership of an asset guarantees residual rights of control over that asset, vertical and lateral integration can be seen as ways of ensuring particular – and presumably efficient – allocations of residual decision rights. [Although in the above example integration increases efficiency, this is in no way a general conclusion. In Grossman and Hart (1986; in press), examples are presented where integration reduces efficiency.]

Before concluding this section, we should emphasize that for reasons of tractability we have confined our attention to incompleteness that is due to a very particular sort of transaction cost. In practice, some of the other transactions costs we have alluded to are likely to be at least as important, if not more so. For example, in the type of model we have analyzed, although the parties cannot describe the state of the world or quality characteristics, they are still supposed to be able to write a contract that is unambiguous and anticipates all eventualities. This is very unrealistic. In practice, a contract might (say) have B agreeing to rent S 's concert hall for a particular price. But suppose S 's hall then burns down. The contract will usually be silent about what is meant to happen under these conditions (there is no hall to rent, but should S pay B damages and if so how much?), and so, in the event of a dispute, the courts will have to fill in the missing provision. (A situation where it becomes impossible or extremely costly to supply a contracted-for good is known in the legal literature as one of “impossibility” or “frustration.”) An analysis of this sort of incompleteness is, though extremely difficult, a very important topic

⁴⁵ This assumes that the contract cannot be renegotiated at date 1. However, even if renegotiation is possible, the buyer's deviation will be unprofitable. This is because the renegotiated price for satisfactory input, p'_1 , will satisfy $p'_1 \geq p_0 + 7$, and hence the buyer's net profit if he deviates is $9 - p'_1 \leq 8.1 - p_1$.

for future research. It is likely to yield a much richer and more realistic view of the way contracts are written and throw light on how courts should assess damages [this latter issue has begun to be analyzed in the law and economics literature; see, e.g., Shavell (1980)].⁴⁶

3.4 *Self-enforcing contracts*

The previous discussion has been concerned with explicit binding contracts that are enforced by outsiders, such as the courts. Even the most casual empiricism tells us that many agreements are not of this type. Although the courts may be there as a last resort (the shadow of the law may therefore be important), these agreements are enforced on a day-to-day basis by custom, good faith, reputation, and so on. Even in the case of a serious dispute, the parties may take great pains to resolve matters themselves rather than go to court. This leads to the notion of a self-enforcing or implicit contract [the importance of informal arrangements like this in business has been stressed by Macaulay (1963) and Ben-Porath (1980), among others].

People often bypass the legal process, presumably because of the transaction costs of using it. The costs of writing a good long-term contract (discussed in Section 3.2) are relevant here. So also is the skill with which the courts resolve contractual disputes. If contracts are incomplete and contain missing provisions as well as vague and ambiguous statements, appropriate enforcement may require abilities and knowledge (what was in the parties' minds?) that many judges and juries do not possess. This means that going to court may be a considerable gamble – and an expensive one at that. (This is an example of the fourth transaction cost noted in Section 3.2.)

Although the notion of implicit or self-enforcing contracts is often invoked, a formal study of such agreements has begun only recently [see, e.g., Bull (1985)], with a considerable stimulus coming from the theory of

⁴⁶ Mention should also be made of a theory of damages developed by Diamond and Maskin (1979). Diamond and Maskin consider a situation where a buyer and seller plan to trade with each other, but recognize that it may be efficient in some states of the world for one of them to trade instead with another party; for instance, the seller may find another buyer with a higher willingness to pay. Under these conditions, the buyer and seller can use the breach damages in their initial contract as a way of extracting surplus from this new party. For example, the bargaining position of a new buyer will be weakened if he must compensate the seller for breaching his contract with the original buyer. (This argument assumes that the new party cannot negotiate ex post with the buyer and seller together to waive the damage payment.) This idea has been used in an interesting paper by Aghion and Bolton (1985) to explain how long-term contracts can deter entry in an industry.

repeated games (cf. the model in Section 1.7). This literature has stressed the role of reputation in “completing” a contract. That is, a party may behave “reasonably,” even if he is not obliged to do so, in order to develop a reputation as a decent and reliable trader. In some instances such reputational effects will operate only within the group of contractual parties – this is sometimes called *internal* enforcement of the contract – while in others the effects will be more pervasive. The latter will be the case when some outsiders to the contract (e.g., other firms in the industry or potential workers for a firm) observe unreasonable behavior by one party and as a result are more reluctant to deal with it in the future. In this case the enforcement is said to be *external* or *market-based*. (The model of Section 1.7 uses the idea of external enforcement.) Note that there may be a tension between this external enforcement and the reasons for the absence of a legally binding contract in the first place – the more people can observe the behavior, the more likely it is to be verifiable.

The distinction between an incomplete contract and a standard asymmetric information contract should be emphasized here. It is the former that allows reputation to operate, since the parties have the same information and can observe whether reasonable behavior is being maintained. In the latter case, it is unclear how reputation can overcome the asymmetry of information between the parties that is the reason for the departure from an Arrow–Debreu contract.

The role of reputation in sustaining a contract can be illustrated using the following model [based on Bull (1985) and Kreps (1984); this is an even simpler model of incomplete contracts than that of the last section]. Assume that a buyer B and a seller S wish to trade an item at date 1 that has value v to the buyer and cost c to the seller, where $v > c$. There are no ex ante investments and the good is homogeneous, so quality is not an issue. Suppose, however, that it is not verifiable whether trade actually occurs. Then a legally binding contract which specifies that the seller must deliver the item and the buyer must pay p , where $v > p > c$, cannot be enforced. The reason is – assuming (as we shall) that simultaneous delivery and payment are infeasible – that if the seller has to deliver first then the buyer can always deny that delivery occurred and refuse payment, while if the buyer has to pay first, the seller can always claim later that he did deliver even when he did not. As a result, if the parties must rely on the courts, a gainful trading opportunity will be missed.

The idea that not even the level of trade is verifiable is extreme, and Bull (1985) in fact makes the more defensible assumption that it is the quality of the good that cannot be verified (in Bull’s model, S is a worker and quality refers to his performance). Bull supposes that quality is observable to the buyer only with a lag, so that take-it-or-leave-it offers of the type considered in Section 3.3 are not feasible. As a result, the seller

always has an incentive to produce minimum quality (which corresponds in the above model to zero output). Making quantity nonverifiable is a cruder but simpler way of capturing the same idea [this is the approach taken in Kreps (1984)].

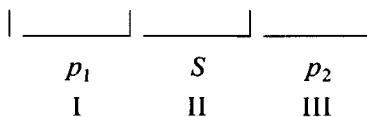
Note that in the above model incompleteness of the contract arises entirely from transaction cost (3), the difficulty of writing and enforcing the contract.

To introduce reputational effects one supposes that this trading relationship is repeated. Bull (1985) and Kreps (1984) follow the supergame literature and assume infinite repetition in order to avoid unraveling problems. This approach, as is well known, suffers from a number of difficulties. First, the assumption of infinite (or, in some versions, potentially infinite) life is hard to swallow. Second, “reasonable” behavior (i.e., trade) is sustained by the threat that if one party behaves unreasonably so will the other party from then on. Although this threat is credible (more precisely, subgame perfect), it is unclear why the parties could not decide to continue to trade after a deviation; that is, to “let bygones be bygones.” [See Farrell (1984); this is another example where the ability to renegotiate *ex post* hurts the parties *ex ante*.]

It would seem that a preferable approach is to assume that the relationship has finite length, but introduce asymmetric information, as in Kreps and Wilson (1982) and in Milgrom and Roberts (1982). The following is based on some very preliminary work that we have undertaken along these lines.

Suppose that there are two types of buyers in the population, honest and dishonest. Honest buyers will always honor any agreement or promise that they have made, whereas dishonest ones will do so only if this is profitable. A buyer knows his own type, but others do not. It is common knowledge that the fraction of honest buyers in the population is π ($0 < \pi < 1$). In contrast, all sellers are known to be dishonest. All agents are risk-neutral.

Assume for simplicity that a single buyer and seller are matched at date 0, with neither having any alternative trading partners at this date or in the future (we are here departing from the *ex ante* perfect competition story that we have maintained for most of the chapter). Consider first the one-period case. Then a date-0 agreement can be represented as follows:



The interpretation is that the buyer promises to pay the seller p_1 before date 1 (stage I); in return, the seller promises to supply the item at date 1

(stage II); and in return for this, the buyer promises to make a further payment of p_2 (stage III).

We should mention one further assumption. Honest buyers, although they never breach an agreement first, are presumed to feel under no obligation to fulfill the terms of an agreement that has already been broken by a seller (interestingly, this theory of buyer psychology has parallels in common law). Note that if a buyer ever breaks an agreement first, he reveals himself to be dishonest, with the consequence that no further self-enforcing agreement with the seller is possible and hence trade ceases.

What is an optimal agreement? Consider the preceding diagram. The seller knows that he will receive p_2 only with probability π since a dishonest buyer will default at the last stage. Since the seller is himself dishonest, he will supply at stage II only if it is profitable for him to do so; that is, only if

$$\pi p_2 - c \geq 0. \tag{3.4}$$

Assume for simplicity that the seller has all the bargaining power at date 0 (nothing that follows depends on this). Then the seller will wish to maximize his overall payoff

$$p_1 + \pi p_2 - c, \tag{3.5}$$

subject to (3.4), which makes it credible that he will supply at stage II, and subject also to the constraint that he does not discourage an honest buyer from participating in the agreement at date 0. Because [with (3.4) satisfied] buyers know for certain that they will receive the item, this last condition is

$$v - p_1 - p_2 \geq 0. \tag{3.6}$$

Note that a dishonest buyer's payoff $v - p_1$ is always higher than an honest buyer's payoff given in (3.6), so there is no way to screen out dishonest buyers. In the language of asymmetric information models, the equilibrium is a pooling one.

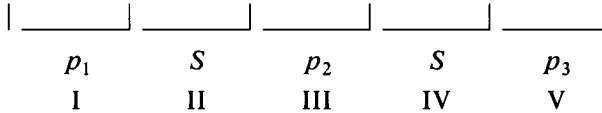
Because the seller's payoff is increasing in p_1 , (3.6) will hold with equality (the buyer gets no surplus). (More generally, changes in p_1 simply redistribute surplus between the two parties without changing either's incentive to breach.) If we substitute for p_1 in (3.5), the seller's payoff becomes $v - p_2(1 - \pi) - c$, which - when maximized subject to (3.4) - yields the solution $p_2 = c/\pi$. The maximized net payoff is

$$v - \frac{c}{\pi} \tag{3.7}$$

which is less than the first-best level, $v - c$.

We see then that the conditions for trade are more stringent in the absence of a binding contract. If $c/\pi > v > c$, there are gains from trade that will not be realized in a one-period relationship.

Suppose now that the relationship is repeated. Consider a two-period version of the above and assume no discounting. Now the following diagram applies:



That is, the agreement says that the buyer pays, the seller supplies the first time, the buyer pays more, the seller supplies a second time, and the buyer makes a final payment. Rather than solving for the optimal arrangement, we shall simply show that the seller can do better than in the one-period case. Let $p_3 = c/\pi$, $p_2 = c$, and $p_1 = 2v - c - c/\pi$. Then (i) the seller will supply at stage IV (if matters have gone that far), knowing that he will receive p_3 with probability π ; (ii) both honest and dishonest buyers will pay p_2 at stage III, the latter because (at a cost of c) they thereby ensure supply worth $v > c$ at stage IV; (iii) the seller will supply at stage II because this gives him a net payoff of $p_2 + \pi p_3 - 2c \geq 0$, while if he does not the arrangement is over and his payoff is zero; (iv) an honest buyer is prepared to participate since his surplus is nonnegative (actually, zero).

The seller's overall expected net payoff is

$$p_1 + p_2 + \pi p_3 - 2c = 2v - c - \frac{c}{\pi}, \tag{3.8}$$

which exceeds twice the one-period payoff. Hence trade is more likely to take place in a two-period relationship than in a one-period one. In fact, it can be shown that the above is an optimal two-period agreement.

Repetition improves things by allowing the honest buyer to pay less the second time (stage III) than the third time (stage V). That is, the arrangement *back-loads* payments. This is acceptable to the seller because he knows that even a dishonest buyer will not default at stage III since he has a large stake in the arrangement continuing. To put it another way, the dishonest buyer does not want to reveal his dishonesty at too early a stage.

The same arrangement can be used when there are more than two periods: The buyer promises to pay c at every stage except the last, when he pays (c/π) . In fact, the per-period surplus of the seller from such an arrangement converges to the first-best level $(v - c)$ as the number of periods tends to infinity (assuming no discounting, of course).

Although the above analysis is extremely provisional and sketchy, we can draw some tentative conclusions about the role of reputation and indicate some directions for further research. First, the notion of a psychic cost of breaking an agreement seems to be a useful – and not unrealistic – basis for a theory of self-enforcing contracts. It is obviously desirable to drop the assumption that some agents are completely honest and others completely dishonest, and assume instead that the typical trader has a finite psychic cost of breaking an agreement, where this cost is distributed in the population in a known way. In other words, everybody has their price, but this price varies. Preliminary work along these lines suggests that the above results generalize; in particular, repetition makes it easier to sustain a self-enforcing agreement.

Of course, asymmetries of information about psychic costs are not the only possible basis for a theory of reputation. For example, the buyer and seller could have private information about v and c , and might choose their trading strategies to influence perceptions about the values of these variables. A theory of self-enforcing contracts should ideally generate results which are not that sensitive to where the asymmetry of information is placed. However, the work of Fudenberg and Maskin (1986) in a related context suggests that this may be a difficult goal to achieve.⁴⁷

There are a number of other natural directions in which to take the model. One is to introduce trade with other parties. For example, the seller may trade with a succession of buyers rather than a single one. The extent to which repetition increases per-period surplus in this case depends on whether new buyers observe the past broken promises of the seller. (This determines the degree to which external enforcement operates; more generally, a new buyer may observe that default occurred in the past but be unsure about who was responsible for it.) If new buyers do not observe past broken promises then repetition achieves nothing, which gives a very strong prediction of the possible benefits of a long-term relationship between a fixed buyer and seller. Even if past broken promises are observed perfectly, it appears that, *ceteris paribus*, a single long-term agreement may be superior to a succession of short-term ones. The reason is that in the latter case the constraint is imposed that each party must receive nonnegative surplus over *their* term of the relationship, whereas in the former case there is only the single constraint that surplus must be nonnegative over the whole term [see Bull (1985) and Kreps (1984)].

Probably the most important extension is to introduce incompleteness

⁴⁷ The role of uncertainty about v and c in determining reputation has been investigated by Thomas and Worrall (1984).

due to other sorts of transaction costs, such as the bounded rationality costs (1) and (2) discussed in Section 3.2. The problem is that the same factors that make it difficult to anticipate and plan for eventualities in a formal contract apply also to informal arrangements. That is, an informal arrangement is also likely to contain many missing provisions. But then the question arises, what constitutes “reasonable” or “desirable” behavior (in terms of building a reputation) with regard to states or actions that were not discussed *ex ante*? Custom, among other things, is likely to be important under these conditions: Behavior will be reasonable or desirable to the extent that it is generally regarded as such [for a good discussion of this, see Kreps (1984)]. This raises many new and interesting (as well as extremely difficult) questions.

Even though our analysis of reputation is very preliminary, it can throw some light on the ABG implicit contract model. There the firm insures the workers against fluctuations in their marginal product of labor. Uncertainty and risk aversion will obviously complicate the analysis of self-enforcing agreements considerably, but the above results suggest that a long-term agreement that stabilizes the workers’ net income may be sustainable even in the absence of a binding contract, particularly if trade is repeated. Moreover, this can be so even if the marginal product of labor is (perfectly) correlated over time (in the above model it is constant), suggesting that an implicit contract may be sustained also for the asymmetric case studied in Section 2.3 (correlation of the marginal product is important because, in its absence, the asymmetry of information may disappear asymptotically; see Section 2.4C). With strong correlation, however, the conditions for an implicit contract will be more stringent, since a firm that has had a bad draw – and knows that this is permanent – will have a stronger incentive to breach [see Newbery and Stiglitz (1983)]. More generally, the fact that a contract must be self-enforcing will impose constraints on the form that it can take. An analysis of the precise conditions under which implicit contracts can be sustained (and of their resulting characteristics), when there is risk aversion and asymmetric information, is an interesting and important topic for future research.

3.5 *Summary and conclusions*

The vast majority of the theoretical work on contracts to date has been concerned with what might be called complete contracts. In this context, a complete contract is one that specifies each party’s obligations in every conceivable eventuality, rather than a contract that is fully contingent in the Arrow–Debreu sense. According to this terminology, the asymmetric

information labor contracts of Section 2.3 are just as complete as the symmetric information ones of Section 2.1.

In reality it is usually impossible to lay down each party's obligations completely and unambiguously in advance, and so most actual contracts are seriously incomplete. In Section 3, we have tried to indicate some of the complications of such incompleteness. Among other things, we have seen that incompleteness can lead to departures from the first-best even when there are no asymmetries of information among the contracting parties (and, moreover, the parties are risk-neutral).

More important perhaps than this is the fact that incompleteness raises new and difficult questions about how the behavior of the contracting parties is determined. To the extent that incomplete contracts do not specify the parties' actions fully (i.e., they contain "gaps"), additional theories are required to tell us how these gaps are filled in. Among other things, outside influences such as custom or reputation may become important under these conditions. In addition, outsiders such as the courts (or arbitrators) may have a role to play in filling in missing provisions of the contract and resolving ambiguities, rather than in simply enforcing an existing agreement. Incompleteness can also throw light on the importance of the allocation of decision rights or rights of control. If it is too costly to state precisely how a particular asset is to be used in every state of the world, it may be efficient simply to give one party "control" of the asset, in the sense that he is entitled to do what he likes with it, subject perhaps to some explicit (contractible) limitations.

Although the importance of incompleteness is very well recognized by lawyers, as well as by those working in law and economics, it is only beginning to be appreciated by economic theorists. It is to be hoped that work in the next few years will lead to significant advances in our formal understanding of this phenomenon. Unfortunately, progress is unlikely to be easy since many aspects of incompleteness are intimately connected with the notion of bounded rationality, a satisfactory formalization of which does not yet exist.

As a final illustration of the importance of incompleteness, consider the following question. Why do parties frequently write a limited term contract, with the intention of renegotiating this when it comes to an end, rather than writing a single contract that extends over the whole length of their relationship? In a complete contract framework such behavior cannot be advantageous, because the parties could just as well calculate what will happen when the contract expires and include this as part of the original contract. It is to be hoped that future work on incomplete contracts will allow this very basic question to be answered.

References

- Abraham, K., and H. Farber. 1985. "Job Duration, Seniority, and Earnings." Mimeo, Massachusetts Institute of Technology.
- Aghion, P., and P. Bolton. 1985. "Entry-Prevention Through Contracts with Customers." Mimeo, Massachusetts Institute of Technology.
- Akerlof, G., and H. Miyazaki. 1980. "The Wage Bill Argument Meets the Implicit Contract Theory of Unemployment." *Review of Economic Studies* 47: 321-38.
- Allen, F. 1985. "Repeated Principal-Agent Relationships with Lending and Borrowing." *Economic Letters* 17: 27-31.
- Antle, R., and A. Smith. 1986. "An Empirical Investigation of the Relative Performance Evaluation of Corporate Executives." *Journal of Accounting Research* 24: 1-39.
- Arnott, R., and J. Stiglitz. 1985. "Labor Turnover, Wage Structures, and Moral Hazard: The Inefficiency of Competitive Markets." *Journal of Labor Economics* 3: 434-62.
- Arnott, R., A. Hosios, and J. Stiglitz. 1985. "Implicit Contracts, Labor Mobility and Unemployment." Mimeo, Queen's University, Canada.
- Aron, D. 1984. "Ability, Moral Hazard, and Firm Diversification, Parts I and II." Mimeo, University of Chicago.
- Arrow, K. 1985. "The Economics of Agency." In J. Pratt and R. Zeckhauser (eds.), *Principals and Agents: The Structure of Business*, pp. 37-51. Boston: Harvard Business School Press.
- Azariadis, C. 1975. "Implicit Contracts and Underemployment Equilibria." *Journal of Political Economy* 83: 1183-1202.
1983. "Employment with Asymmetric Information." *Quarterly Journal of Economics (Supplement)* 98: 157-73.
- Baily, M. 1974. "Wages and Employment Under Uncertain Demand." *Review of Economic Studies* 41: 37-50.
- Baiman, S. 1982. "Agency Research in Managerial Accounting: A Survey." *Journal of Accounting Literature* 1: 154-213.
- Baiman, S., and J. Demski. 1980. "Economically Optimal Performance Evaluation and Control Systems." *Supplement to Journal of Accounting Research* 18: 184-220.
- Baron, D., and D. Besanko. 1984. "Regulation, Asymmetric Information, and Auditing." *Rand Journal of Economics* 15: 447-70.
- Baron, D., and R. Myerson. 1982. "Regulating a Monopolist with Unknown Costs." *Econometrica* 50: 911-30.
- Becker, G. 1964. *Human Capital*. New York: Columbia University Press.
- Becker, G., and G. Stigler. 1974. "Law Enforcement, Malfeasance and Compensation of Enforcers." *Journal of Legal Studies* 3: 1-18.
- Ben-Porath, Y. 1980. "The F-Connection: Families, Friends, and Firms and the Organization of Exchange." *Population and Development Review* 6: 130.

- Bhattacharya, S. 1983. "Tournaments and Incentives: Heterogeneity and Essentiality." Research Paper No. 695, Graduate School of Business, Stanford University.
- Brown, J., and O. Ashenfelter. 1986. "Testing the Efficiency of Employment Contracts." *Journal of Political Economy* 94: 540–87.
- Bull, C. 1985. "The Existence of Self-Enforcing Implicit Contracts." C. V. Starr Center, New York University.
- Calvo, G., and E. Phelps. 1977. "Employment Contingent Wage Contracts." *Journal of Monetary Economics* 5: 160–8.
- Card, D. 1985. "Efficient Contracts and Costs of Adjustment: Short-Run Employment Determination for Airline Mechanics." Mimeo, Princeton University.
- Carmichael, L. 1984. "Firm-Specific Human Capital and Promotion Ladders." *Rand Journal of Economics* 14: 251–8.
- Chari, V. 1983. "Involuntary Unemployment and Implicit Contracts." *Quarterly Journal of Economics (Supplement)* 98: 107–22.
- Cooper, R. 1983. "A Note on Overemployment/Underemployment in Labor Contracts Under Asymmetric Information." *Economic Letters* 12: 81–7.
- Crawford, V. 1982. "Long-Term Relationships Governed by Short-Term Contracts." ICERD DP, London School of Economics.
- Cremer, J., and M. Riordan. 1986. "On Governing Multilateral Transactions with Bilateral Contracts." Discussion Paper No. 134, Studies in Industry Economics, Stanford University.
- D'Aspremont, C., and L. Gerard-Varet. 1979. "Incentives and Incomplete Information." *Journal of Public Economics* 11: 25–45.
- DeGroot, M. 1970. *Optimal Statistical Decisions*. New York: McGraw-Hill Book Company.
- Dewatripont, M. 1985. "Renegotiation and Information Revelation Over Time: The Case of Optimal Labor Contracts." Mimeo, Harvard University.
- Diamond, D. 1984. "Financial Intermediation and Delegated Monitoring." *Review of Economic Studies* 51: 393–414.
1985. "Reputation Acquisition in Debt Markets." Mimeo, University of Chicago.
- Diamond, P., and E. Maskin. 1979. "An Equilibrium Analysis of Search and Breach of Contract, I: Steady States." *Bell Journal of Economics* 10: 282–316.
- Dye, R. 1985. "Costly Contract Contingencies." *International Economic Review* 26: 233–50.
- Eden, B. 1985. "Labor Contracts, Enforcement and Fluctuations in Aggregate Employment: The Case of No Severance Payments." Working Paper, University of Iowa.
- Eswaran, M., and A. Kotwal. 1984. "The Moral Hazard of Budget-Breaking." *Rand Journal of Economics* 15: 578–81.
- Fama, E. 1980. "Agency Problems and the Theory of the Firm." *Journal of Political Economy* 88: 288–307.
- Farmer, R. 1984. "A New Theory of Aggregate Supply." *American Economic Review* 74: 920–30.

1985. "Implicit Contracts with Asymmetric Information and Bankruptcy: The Effect of Interest Rates on Layoffs." *Review of Economic Studies* 52: 427–42.
- Farrell, J. 1984. "Renegotiation-Proof Equilibrium in Repeated Games." Mimeo, Massachusetts Institute of Technology.
- Feldstein, M. 1976. "Temporary Layoffs in the Theory of Unemployment." *Journal of Political Economy* 84: 937–57.
- Fischer, S. 1977. "Long Term Contracts, Rational Expectations, and the Optimal Money Supply Rule." *Journal of Political Economy* 85: 191–205.
- Freixas, X., R. Guesnerie, and J. Tirole. 1985. "Planning Under Incomplete Information and the Ratchet Effect." *Review of Economic Studies* 52: 173–92.
- Fudenberg, D., B. Holmström, and P. Milgrom. 1986. "Repeated Moral Hazard with Borrowing and Saving." Draft, School of Organization and Management, Yale University.
- Fudenberg, D., and E. Maskin. 1986. "The Folk Theorem in Repeated Games with Discounting and with Incomplete Information." *Econometrica* 54: 533–54.
- Fudenberg, D., and J. Tirole. 1983. "Sequential Bargaining with Asymmetric Information." *Review of Economic Studies* 50: 221–48.
- Gale, D., and M. Hellwig. 1985. "Incentive Compatible Debt Contracts: the One Period Problem." *Review of Economic Studies* 52: 647–64.
- Geanakoplos, J., and T. Ito. 1982. "On Implicit Contracts and Involuntary Unemployment." Cowles Foundation Discussion Paper #640, Yale University.
- Gibbons, R. 1985. "Essays on Labor Markets and Internal Organization." Unpublished dissertation, Stanford University.
- Gjesdal, F. 1982. "Information and Incentives: The Agency Information Problem." *Review of Economic Studies* 49: 373–90.
- Goldberg, V., and J. Erickson. 1982. "Long-Term Contracts for Petroleum Coke." Department of Economics Working Paper Series No. 206, University of California–Davis.
- Gordon, D. 1974. "A Neo-Classical Theory of Keynesian Unemployment." *Economic Inquiry* 12: 431–59.
- Green, J. 1985. "Differential Information, the Market and Incentive Compatibility." In K. Arrow and S. Honkapohja (eds.), *Frontiers of Economics*, pp. 178–99. Oxford: Basil Blackwell.
- Green, J., and C. Kahn. 1983. "Wage Employment Contracts." *Quarterly Journal of Economics (Supplement)* 98: 173–88.
- Green, J., and N. Stokey. 1983. "A Comparison of Tournaments and Contracts." *Journal of Political Economy* 91: 349–64.
- Grossman, S., and O. Hart. 1981. "Implicit Contracts, Moral Hazard and Unemployment." *American Economic Review (Papers and Proceedings)* 71: 301–8.
1982. "Corporate Financial Structure and Managerial Incentives." In J. McCall (ed.), *The Economics of Information and Uncertainty*, pp. 107–37. Chicago: University of Chicago Press.
- 1983a. "An Analysis of the Principal–Agent Problem." *Econometrica* 51: 7–45.
- 1983b. "Implicit Contracts Under Asymmetric Information." *Quarterly Journal of Economics (Supplement)* 71: 123–57.

1986. "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration." *Journal of Political Economy* 94: 691-719.
- In press. "Vertical Integration and the Distribution of Property Rights." In A. Razin and E. Sadka (eds.), *Economic Policy in Theory and Practice*. New York: Oxford University Press.
- Grossman, S., O. Hart, and E. Maskin. 1983. "Unemployment with Observable Aggregate Shocks." *Journal of Political Economy* 91: 907-28.
- Grout, P. 1984. "Investment and Wages in the Absence of Binding Contracts: A Nash Bargaining Approach." *Econometrica* 52: 449-60.
- Hahn, F. 1984. "Implicit Contracts and Involuntary Unemployment." Discussion Paper No. 71, Cambridge University.
- Hall, R. 1980. "Employment Fluctuations and Wage Rigidity." *Brookings Papers on Economic Activity* 1: 91-123.
- Hall, R., and D. Lilien. 1979. "Efficient Wage Bargains Under Uncertain Supply and Demand." *American Economic Review* 69: 868-79.
- Harris, M., and B. Holmström. 1982. "A Theory of Wage Dynamics." *Review of Economic Studies* 49: 315-33.
- Harris, M., and A. Raviv. 1979. "Optimal Incentive Contracts with Imperfect Information." *Journal of Economic Theory* 20: 231-59.
1981. "Allocation Mechanisms and the Design of Auctions." *Econometrica* 49: 1477-99.
- Harris, M., and R. Townsend. 1981. "Resource Allocation Under Asymmetric Information." *Econometrica* 49: 33-64.
- Hart, O. 1983. "Optimal Labour Contracts Under Asymmetric Information: An Introduction." *Review of Economic Studies* 50: 3-35.
- Hart, O., and J. Moore. 1985. "Incomplete Contracts and Renegotiation." Working Paper, London School of Economics.
- Holmström, B. 1979. "Moral Hazard and Observability." *Bell Journal of Economics* 10: 74-91.
- 1982a. "Moral Hazard in Teams." *Bell Journal of Economics* 13: 324-40.
- 1982b. "Managerial Incentive Problems - A Dynamic Perspective." In *Essays in Economics and Management in Honor of Lars Wahlbeck*. Helsinki: Swedish School of Economics.
1983. "Equilibrium Long-Term Labor Contracts." *Quarterly Journal of Economics (Supplement)* 98: 23-54.
- Holmström, B., and P. Milgrom. 1985. "Aggregation and Linearity in the Provision of Intertemporal Incentives." Cowles Discussion Paper No. 742.
- Holmström, B., and J. Ricart-Costa. 1986. "Managerial Incentives and Capital Management." *Quarterly Journal of Economics* 101: 835-60.
- Holmström, B., and L. Weiss. 1985. "Managerial Incentives, Investment and Aggregate Implications: Scale Effects." *Review of Economic Studies* 52: 403-26.
- Jensen, M., and W. Meckling. 1976. "Theory of the Firm: Managerial Behavior, Agency Costs, and Capital Structure." *Journal of Financial Economics* 3: 305-60.
- Joskow, P. 1985. "Vertical Integration and Long-Term Contracts." *Journal of Law, Economics and Organization* 1: 33-80.

- Kahn, C. 1985. "Optimal Severance Pay with Incomplete Information." *Journal of Political Economy* 93: 435–51.
- Kihlstrom, R., and J. Laffont. 1979. "A General Equilibrium Entrepreneurial Theory of Firm Formation Based on Risk Aversion." *Journal of Political Economy* 87: 719–48.
- Killingsworth, M. 1983. *Labor Supply*. Cambridge University Press.
- Klein, B., R. Crawford, and A. Alchian. 1978. "Vertical Integration, Appropriable Rents and the Competitive Contracting Process." *Journal of Law and Economics* 21: 297–326.
- Kreps, D. 1984. "Corporate Culture and Economic Theory." Mimeo, Stanford University.
- Kreps, D., and R. Wilson. 1982. "Reputation and Imperfect Information." *Journal of Economic Theory* 27: 253–79.
- Kydland, F., and E. Prescott. 1977. "Rules Rather Than Discretion: The Inconsistency of Optimal Plans." *Journal of Political Economy* 85: 473–92.
- Laffont, J-J., and E. Maskin. 1982. "Theory of Incentives: An Overview." In W. Hildenbrand (ed.), *Advances in Economic Theory*, pp. 31–94. Cambridge: Cambridge University Press.
- Laffont, J-J., and J. Tirole. 1986. "Using Cost Observation to Regulate Firms." *Journal of Political Economy* 94: 614–41.
- Lambert, R. 1983. "Long-Term Contracting and Moral Hazard." *Bell Journal of Economics* 14: 441–52.
1986. "Executive Effort and Selection of Risky Projects." *Rand Journal of Economics* 16: 77–88.
- Lazear, E. 1979. "Why is There Mandatory Retirement?" *Journal of Political Economy* 87: 1261–84.
- Lazear, E., and S. Rosen. 1981. "Rank-Order Tournaments as Optimum Labor Contracts." *Journal of Political Economy* 89: 841–64.
- Lucas, R. 1972. "Expectations and the Neutrality of Money." *Journal of Economic Theory* 4: 103–24.
- Macaulay, S. 1963. "Non-Contractual Relations in Business: A Preliminary Study." *American Sociological Review* 28: 55–67.
- Malcomson, J. 1981. "Unemployment and the Efficiency Wage Hypothesis." *Economic Journal* 91: 848–66.
- 1984a. "Rank-Order Contracts for a Principal with Many Agents." Working Paper No. 8405, Universite Catholique de Louvain.
- 1984b. "Work Incentives, Hierarchy, and Internal Labor Markets." *Journal of Political Economy* 92: 486–507.
- Malcomson, J., and F. Spinnewyn. 1985. "The Multiperiod Principal Agent Problem." Discussion Paper No. 8511, University of Southampton.
- Maskin, E., and J. Riley. 1984. "Optimal Auctions with Risk-Averse Buyers." *Econometrica* 52: 1473–1518.
- Milgrom, P. 1981. "Good News and Bad News: Representation Theorems and Applications." *Bell Journal of Economics* 12: 380–91.
- Milgrom, P., and J. Roberts. 1982. "Predation, Reputation and Entry Deterrence." *Journal of Economic Theory* 27: 280–312.

- Mirrlees, J. 1971. "An Exploration in the Theory of Optimum Income Taxation." *Review of Economic Studies* 38: 175–208.
1974. "Notes on Welfare Economics, Information and Uncertainty." In M. Balch, D. McFadden, and S. Wu (eds.), *Essays in Economic Behavior Under Uncertainty*, pp. 243–58. Amsterdam: North-Holland.
1975. "The Theory of Moral Hazard and Unobservable Behavior – Part I." Mimeo, Nuffield College, Oxford.
1976. "The Optimal Structure of Authority and Incentives Within an Organization." *Bell Journal of Economics* 7: 105–31.
- Moore, J. 1984. "Contracting Between Two Parties With Private Information." Working Paper, London School of Economics.
1985. "Optimal Labour Contracts When Workers Have a Variety of Privately Observed Reservation Wages." *Review of Economic Studies* 52: 37–67.
- Murphy, K. 1986. "Incentives, Learning and Compensation: A Theoretical and Empirical Investigation of Managerial Labor Contracts." *Rand Journal of Economics* 17: 59–76.
- Myerson, R. 1979. "Incentive Compatibility and the Bargaining Problem." *Econometrica* 47: 61–74.
1981. "Optimal Auction Design." *Mathematics of Operations Research* 6: 58–73.
- Nalebuff, B., and J. Stiglitz. 1983. "Prizes and Incentives: Towards a General Theory of Compensation and Competition." *Bell Journal of Economics* 13: 21–43.
- Newbery, D., and J. Stiglitz. 1983. "Wage Rigidity, Implicit Contracts and Economic Efficiency: Are Market Wages Too Flexible?" Economic Theory Discussion Paper 68, Cambridge University.
- New York Times*. 1985. "The Two-Tier Wage Impact." October 30.
- Oswald, A. 1984. "Efficient Contracts Are on the Labour Demand Curve: Theory and Facts." Mimeo, Oxford University.
- Radner, R. 1981. "Monitoring Cooperative Agreements in a Repeated Principal-Agent Relationship." *Econometrica* 49: 1127–48.
1985. "Repeated Principal-Agent Games with Discounting." *Econometrica* 53: 1173–98.
- Riordan, M. 1984. "Uncertainty, Asymmetric Information and Bilateral Contracts." *Review of Economic Studies* 51: 83–93.
- Roberts, K. 1982. "Long-Term Contracts." Mimeo, Warwick University.
- Rogerson, W. 1985a. "Repeated Moral Hazard." *Econometrica* 53: 69–76.
- 1985b. "The First-Order Approach to Principal-Agent Problems." *Econometrica* 53: 1357–68.
- Rosen, S. 1985. "Implicit Contracts: A Survey." *Journal of Economic Literature* 23: 1144–75.
- Ross, S. 1973. "The Economic Theory of Agency: The Principal's Problem." *American Economic Review* 63: 134–9.
- Rubinstein, A. 1979. "Offenses That May Have Been Committed by Accident – An Optimal Policy of Retribution." In S. Brams, A. Shotter, and G. Schwödinger (eds.), *Applied Game Theory*, pp. 406–13. Wurtzburg: Physica-Verlag.

- Salop, J., and S. Salop. 1976. "Self-Selection in the Labor Market." *Quarterly Journal of Economics* 90: 619–27.
- Schelling, T. 1960. *The Strategy of Conflict*. Harvard University Press.
- Shapiro, C., and J. Stiglitz. 1984. "Equilibrium Unemployment as a Worker Incentive Device." *American Economic Review* 74: 433–44.
- Shavell, S. 1979. "Risk Sharing and Incentives in the Principal and Agent Relationship." *Bell Journal of Economics* 10: 55–73.
1980. "Damage Measures for Breach of Contract." *Bell Journal of Economics* 11: 466–90.
- Simon, H. 1982. *Models of Bounded Rationality*. MIT Press.
- Spence, M., and R. Zeckhauser. 1971. "Insurance, Information and Individual Action." *American Economic Review (Papers and Proceedings)* 61: 380–7.
- Stiglitz, J., and A. Weiss. 1981. "Credit Rationing in Markets with Imperfect Information." *American Economic Review* 71: 393–410.
- Taylor, J. 1980. "Aggregate Dynamics and Staggered Contracts." *Journal of Political Economy* 88: 1–24.
- Thomas, J., and T. Worrall. 1984. "Self-Enforcing Wage Contracts." Mimeo, University of Cambridge.
- Townsend, R. 1980. "Models of Money With Spatially Separated Agents." In J. Kareken and N. Wallace (eds.), *Models of Monetary Economics*. Minneapolis: Federal Reserve Bank.
1982. "Optimal Multiperiod Contracts and the Gain from Enduring Relationships Under Private Information." *Journal of Political Economy* 90: 1166–86.
- Weiss, A. 1980. "Job Queues and Layoffs in Labor Markets with Flexible Wages." *Journal of Political Economy* 88: 526–38.
- Williamson, O. 1975. *Markets and Hierarchies: Analysis and Antitrust Implications*. New York: Free Press.
1985. *The Economic Institutions of Capitalism*. New York: Free Press.
- Wilson, R. 1969. "The Structure of Incentives for Decentralization under Uncertainty." *La Decision* 171.
- Wolfson, M. 1985. "Empirical Evidence of Incentive Problems and Their Mitigation in Oil and Tax Shelter Programs." In J. Pratt and R. Zeckhauser (eds.), *Principals and Agents: The Structure of Business*, pp. 101–25. Boston: Harvard Business School Press.
- Yaari, M. 1976. "A Law of Large Numbers in the Theory of Consumer's Choice Under Uncertainty." *Journal of Economic Theory* 12: 202–17.

