Rigidity in bilateral trade with holdup

Rui R. Zhao

Department of Economics, University at Albany, State University of New York

This paper studies bilateral trade in which the seller makes a hidden investment that influences the buyer's hidden valuation. In general it is impossible to implement both first-best efficient trade and efficient investment using budget-balanced trading mechanisms. The paper fully characterizes the constrained efficient contracts. It is shown that the optimal tradeoff between allocative efficiency and incentive provision results in rigidity in trade, the degree of which depends on the seriousness of the holdup problem. Sufficient conditions are also provided for full separation of buyer types to take place in optimal contracts when the holdup problem is not too severe. The seller may overinvest relative to the first best.

KEYWORDS. Bilateral contracting, hidden action and hidden information, holdup problem, nonlinear pricing.

JEL CLASSIFICATION. C72, D20, D82.

1. INTRODUCTION

This paper studies optimal contracting for bilateral trade when a holdup problem is present. It is partly motivated by two apparently conflicting requirements for an optimal trading mechanism. On the one hand, since efficient trade is affected by uncertainties in demand and supply, an optimal contract needs to be *flexible* so as to accommodate changes in the environment. On the other hand, to motivate a party to invest in the relationship the contract should offer sufficient protection for investment return, which typically calls for some *rigid* arrangements so as to minimize opportunistic behavior by the trading partner. In a seminal paper, Rogerson (1992) shows that the conflict between flexibility and rigidity does not exist if investment directly affects only the payoff of the investor: absent an investment externality the holdup problem can be completely resolved and full efficiency can be achieved under very general information structures. This paper instead focuses on the problem in which one party's investment directly affects the payoff of the trading partner so that an externality exists.

Specifically I study a bilateral trade model in which the seller makes a hidden investment before trade that affects the buyer's hidden valuation of the good. The model

generalizes the classic principal-agent model and fits many real-world applications such as supply contracting between a downstream producer and an upstream supplier, procurement contracting between a contractor and government, and the Advance Market Commitment mechanism recently proposed by Barder et al. (2006) as a means to spur R&D investments on vaccines for diseases concentrated in low-income countries.

The efficient mechanism proposed by Rogerson fails to apply to the current setting because of the externality in investment. In fact, there does not exist any budget-balanced mechanism that can implement both efficient trade and efficient investment. The basic reason resembles that of moral hazard in teams, though here the buyer has the unusual action of reporting her type rather than producing the good. With two-sided information asymmetry, a budget-balanced mechanism always leaves some information rent to the buyer and prevents the seller from getting the full *marginal* return on investment, which given efficient trade leads to inefficient investment.

I fully characterize the constrained efficient mechanisms; the results demonstrate an unusually clear relation between rigidity in trade and the severity of the holdup problem in the optimal contract. The analysis also integrates moral hazard and adverse selection and highlights a driving force for optimal contracting that is different from those in the standard models.

In the moral hazard model the primary force that shapes the optimal contract is the tradeoff between risk sharing and incentive provision; with adverse selection, the optimal contract is driven by the tradeoff between allocative efficiency and the need to extract rent from the buyer. In contrast, the agents in the current model are risk neutral towards income, so risk sharing is not a concern; rent extraction per se is also not a concern because lump-sum transfers can be made at ex ante.

Instead, the present model highlights the interaction between incentive provision and allocative efficiency. In one extreme case, the contract can be fully flexible, so that trade is efficient for all types of buyer. But to achieve this allocative efficiency the contract has to give most of the marginal surplus from trade to the buyer due to the need for information revelation, leaving the seller with little incentive to invest. In the opposite case, the contract can be made more rigid, which sacrifices some ex post allocative efficiency but offers a better incentive to the seller. The optimal contract has to strike a balance between these tradeoffs. An important yet familiar result is that it is necessary to distort trade downward from the efficient level; the reason is that reducing trade helps transfer the buyer's rent to the seller and therefore improves the seller's incentive to invest. If not for this incentive reason, downward distortion would be counterproductive as the seller can always be given lump-sum transfers in exchange for efficient trade for all buyer types.

The interaction between incentive provision and allocative efficiency results in some striking features of the optimal contract, which depend in a crucial way on the degree to which the seller's incentive constraint binds at the optimum. First, when the incentive constraint binds to a low degree, i.e. the shadow price on the constraint is small, trade is efficient at the two ends of the type interval. The intuition is that reducing trade for the top or the bottom type shifts a constant amount of rent from the buyer to the seller

for all types, and hence has no effect on the seller's marginal return; then allocative efficiency dictates that trade be efficient for these two types. As the shadow price increases, reducing trade for *low* types has a more significant impact on the seller's marginal return as it punishes the seller for bad performance by reducing the total surplus. Taking advantage of this opportunity for incentive provision, the optimal contract calls for a large reduction in trade for types close to the bottom. Such a move however is in conflict with efficient trade for the lowest type because the incentive compatibility of the buyer requires a non-decreasing trade schedule. The result is *rigidity*: a range of the bottom types must be *bunched* together to receive the same level of trade. The rigidity can be intensified as the shadow price increases further, to the point that a range of the bottom types have to be excluded from trade altogether.

Moreover, if a given level of investment is implemented then a more serious holdup problem due to changes in investment cost will cause trade to be distorted further away from the efficient level for all buyer types. As the degree of the holdup problem goes to zero, the optimal trade schedule becomes fully flexible and efficient. Thus there is a *monotonic* relation between trade distortion (a form of rigidity) and the severity of the seller's incentive constraint.

To be sure, bunching is not a new phenomenon in nonlinear pricing models. What should be stressed here is the reason and the way it occurs in the present model. In standard nonlinear pricing bunching occurs when the type distribution fails to possess certain sorting conditions. What is novel here is that rigidity in trade is a reflection of the seriousness of the holdup problem relative to the effectiveness of investment (see Proposition 3). These factors make bunching especially easy to occur *regardless* of whether the standard sorting conditions are satisfied. Therefore a good way of interpreting the results of this paper is that they identify new, empirically relevant sources of trade distortion, and show that their effects are characterized by the degree of flexibility of the optimal trading schedule.

Relation to the existing literature

The papers closest to the current study are Schmitz (2002a) and a contemporaneous paper by Hori (2006). Schmitz studies a similar model with a hidden action and hidden information but assumes that the good is indivisible and both the investment choice and the buyer's type are binary. He obtains two main results: one is the important observation that full efficiency is impossible under quite reasonable conditions; the other shows that the optimal trade schedule displays the familiar phenomenon of efficiency at the top and downward distortion at the bottom. Hori extends the inefficiency result to the case of continuous investment and type but still with an indivisible good. The model in this paper is similar in spirit but much more general. It adds more realism to the analysis as many bilateral trade relationships involve trading divisible goods and agents can have general utility and cost functions. More importantly, this general treatment helps uncover the fundamental forces that shape optimal contracts in this canonical contracting problem. The payoff from doing so is reflected in the striking features of the optimal contract presented in Section 4.

The existing literature on the holdup problem relies mostly on the methodology of incomplete contracts and the typical treatment assumes both symmetric information and renegotiation, which naturally lead to ex post efficient allocations (see the exceptions below). Therefore unlike the present paper, the existing studies focus exclusively on investment incentives and do not examine the tradeoff between allocative efficiency and incentive provision.[1] An exception is the aforementioned paper by Rogerson (1992), which studies a general version of the holdup problem in a complete contracts framework and shows that full efficiency can be achieved under broad information structures if both investment externalities and renegotiation are ruled out.[2] The current paper differs from Rogerson by focusing on cooperative investment, which was first studied in bilateral trade by Che and Hausch (1999) under the assumption of incomplete contracting.

Investment externalities are common in bilateral relationships, including the classic principal-agent problem where the agent's action affects the profits of the principal. The model in the current paper generalizes the principal-agent model: in addition to the hidden action of the agent the principal *privately* observes how valuable the output is, and on top of that the two parties also choose the level of trade. A related paper along this line is MacLeod (2003), which studies employment contracting with private evaluation. The focus of MacLeod's analysis however is on wage contracts rather than trade; indeed, in his model the trade level is a simple zero-one choice as it concerns only whether or not the agent works for the principal.

The analysis of the current paper is clearly related to the existing literature on nonlinear pricing and screening.[3] One salient feature of the optimal trade schedule in this paper is that it is much harder to achieve full separation of types than in standard screening models: sorting conditions on type distributions are no longer sufficient and rigidity in trade in the form of bunching is an optimal response to the holdup problem. This particular source of trade distortion identified in the current paper differs from those identified elsewhere in the literature. For instance, risk aversion is singled out as a factor by Salanié (1990, 1997). In contrast, the agents in the current model are risk neutral, which leads to a first-best contract in Salanié's model. Some other sources of rigidity in trade include multidimensional types (see Armstrong 1996, Rochet and Choné 1998, Jullien 2000, and Rochet and Stole 2002), countervailing incentives of the agent when reporting private information (Lewis and Sappington 1989), and limited enforcement of contracts (Levin 2003). By identifying a novel source of rigidity in trade, this paper complements these other studies.

The rest of the paper is organized as follows. Section 2 sets out the basic model and proves the inefficiency result. Section 3 formulates the main program. Section 4 characterizes the constrained efficient mechanisms and discusses their many implications. Section 5 offers a numerical algorithm and a computed example. Section 6 illustrates possible *over*investment. Section 7 concludes. All proofs are contained in the appendices.

---

[1]For some earlier work, see Williamson (1979), Grout (1984), and Hart and Moore (1988).

[2]See also Konakayama et al. (1986) and Schmitz (2002b). The former studies a special case of the model of Rogerson; the latter shows that efficiency can still be achieved even if certain types of renegotiation are allowed.

[3]See Mussa and Rosen (1978), Maskin and Riley (1984), and Guesnerie and Laffont (1984).

2. The model

Consider a seller and a buyer who are interested in trading a divisible good. The sequence of events unfolds according to the following time line.

time 0	1	2	3	4
contract signed	seller invests a cost $g(a)$	buyer learns type θ	buyer selects (x,t)	delivery of x cost $c(x)$, transfer t

At date 0, the agents sign a trading contract. The contract specifies a menu of quantity–payment pairs (x,t) from which one pair will be selected in the future. Note that x can also be interpreted as the quality level of the good. At date 1, the seller privately chooses an action or investment level a from some feasible set $\mathscr{A}$ at the cost $g(a)$. The purpose of such an investment is to improve the value of the good to the buyer and hence is cooperative. For instance the investment may create a prototype that demonstrates the functionalities of the good designed exclusively for the buyer. At date 2, after reviewing the prototype the buyer forms a valuation of the good, which is her private information. Ex ante the buyer's valuation type θ is stochastic and is distributed on a set Θ according to a distribution function $F(\theta,a)$. At date 3, the buyer, now having learned her type, chooses from the contract menu a specific quantity–payment pair (x,t). At date 4, the seller produces the requested quantity x at the cost $c(x)$, delivers the goods to the buyer, and collects the payment t from the buyer. Finally the buyer obtains a payoff equal to $u(\theta,x) - t$ and the seller obtains a payoff equal to $t - c(x) - g(a)$.

To summarize, both the seller's investment choice a and the buyer's type θ are their private information respectively; the structure of the model, including the cost functions g, c, the utility function u, and the probability distribution F, is common knowledge. I make the following assumptions on preferences and technology.

ASSUMPTION 1. The set of feasible trade levels is $X = [0,\infty)$ and the set of possible buyer types is $\Theta = [\underline{\theta},\overline{\theta}] \subset \mathbb{R}_+$. For each type θ, the utility function $u(\theta,\cdot): X \longrightarrow \mathbb{R}_+$ is twice continuously differentiable, strictly increasing, and concave, with $u(\theta,0) = 0$. For all $\theta > \theta'$ and $x > 0$, $u(\theta,x) > u(\theta',x)$, i.e. higher type is associated with higher valuation.

ASSUMPTION 2. The feasible investment levels are contained in the set $\mathscr{A} = [\underline{a},\overline{a}] \subset \mathbb{R}_+$. The seller's investment cost function $g : \mathscr{A} \to \mathbb{R}_+$ and production cost function $c : X \to \mathbb{R}_+$ are both twice continuously differentiable, strictly increasing, and convex; moreover, $c(0) = 0$, $\lim_{x\to\infty} c'(x) = \infty$, $g(\underline{a}) = g'(\underline{a}) = g''(\underline{a}) = 0$, $g'(\overline{a}) = \infty$.

An allocation in this environment is an investment choice a and a rule $(x,t): \Theta \to X \times \mathbb{R}$ that specifies a trade $x(\theta)$ and transfer $t(\theta)$ for each buyer type θ. I focus on deterministic contracts. By the Revelation Principle, one can focus on allocations that are incentive compatible, i.e. the buyer has an incentive to report her type truthfully and the seller finds it optimal to choose the assigned investment choice.

Given that the agents' preferences are quasi-linear, the first-best allocations maximize the expected total surplus from trade. Specifically, for each buyer type θ a first-best *efficient trade* $x^E(\theta)$ maximizes the joint surplus $u(\theta,x) - c(x)$; given efficient trade,

a first-best *efficient investment* a^E maximizes the expected joint surplus

$$\int_\Theta [u(\theta, x^E(\theta)) - c(x^E(\theta))] \, dF(\theta, a) - g(a).$$

The first question is whether there exists a transfer schedule $t : \Theta \to \mathbb{R}$ such that the efficient trade $x^E : \Theta \to X$ and efficient investment a^E can be implemented simultaneously, namely whether (x^E, t) and a^E satisfy the buyer's truth-telling constraint

$$u(\theta, x^E(\theta)) - t(\theta) \geqslant u(\theta, x^E(\hat\theta)) - t(\hat\theta) \quad \forall \theta, \hat\theta \in \Theta$$

and the seller's incentive constraint

$$a^E \in \underset{a \in A}{\arg\max} \int_\Theta [t(\theta) - c(x^E(\theta))] \, dF(\theta, a) - g(a).$$

The following proposition shows that the answer in general is negative.

PROPOSITION 1. *Suppose that Assumptions 1–2 are satisfied and a^E is an efficient investment choice in the interior of $\mathcal{A}$. Let $\Theta^* = \{\theta \in \Theta \mid x^E(\theta) > 0\}$. Assume that Θ^* has positive measure that is not concentrated on a single point. Assume also that $F_a(\theta, a^E) \leqslant 0$ for all $\theta \in \Theta^*$ and the inequality is strict on a subset of Θ^* that is of positive measure. Then there does not exist a budget-balanced incentive compatible mechanism that implements a^E and $x^E(\theta)$, for all θ.*

The main hypothesis of the proposition is that F satisfies the first-order stochastic dominance condition $(F_a(\theta, a^E) \leqslant 0)$ for buyer types that really matter (receiving positive trade), which means it is costly to produce (stochastically) better outcomes. This ensures that there is some minimal tension between efficiency and incentive provision; otherwise the seller's incentive is trivial and efficiency clearly is possible. Proposition 1 covers a variety of situations with general distributions of buyer type, which can be continuous, discrete, or mixtures of the two.

The reason for the inefficiency hinges upon the marginal effects of investment on the agents' payoffs. The intuition is as follows. Due to information asymmetry, the buyer necessarily enjoys some information rent if more than one type needs to be served in the efficient allocation. In particular incentive compatibility requires that the buyer's rent be strictly increasing in her type. Then it follows that an increase in the investment level must raise the buyer's payoff because higher investment improves the distribution of buyer's type. At the efficient investment level, the social marginal benefit equals the seller's marginal cost of investment. Since the buyer gets positive marginal benefit from investment, the seller's marginal benefit has to be less than his marginal cost, which of course is not incentive compatible as the seller can increase his payoff by slightly cutting back the investment. In essence, full efficiency cannot be achieved because with a balanced budget the seller cannot receive the full marginal return on investment. This explains also why full efficiency *can* be achieved if only the highest type needs to be

served in the efficient allocation, as found by Schmitz (2002a); in that case the buyer's rent is constantly zero so the seller's objective coincides with the social objective.

Note that the inefficiency result obtained here is for ex ante contracting and does not rely on interim participation constraints. This differs from the Myerson and Satterthwaite (1983) theorem, in which inefficiency is caused by the incompatibility between efficient trade and voluntary participation by all types of the agents.[4]

Also, it should be stressed that three ingredients of the model, asymmetric information, balanced budget, and cooperative investment, are all responsible for the inefficiency, and it is not too hard to find mechanisms that restore full efficiency if one of these assumptions is relaxed.

The next few sections characterize the constrained efficient contracts. For this purpose, the following additional assumptions are made throughout the rest of the analysis.

ASSUMPTION 3. For each $a \in \mathscr{A}$, the distribution function $F(\cdot, a)$ has a continuous density $f(\cdot, a)$ on $[\underline{\theta}, \overline{\theta}]$. Both F and f are continuously differentiable in a, and the partial $f_a(\theta, a)$ is continuous in θ.

ASSUMPTION 4 (Single-Crossing). For all $\theta, \theta' \in \Theta$ with $\theta > \theta'$ and for all $x \in X$, $u_x(\theta, x) \geq u_x(\theta', x)$. In the differentiable case, this condition becomes: for all $\theta \in \Theta$ and for all $x \in X$, $u_{\theta x}(\theta, x) \geq 0$.

ASSUMPTION 5 (FSD). For all $a \in \mathscr{A}$ and all $\theta \in \Theta$, $F_a(\theta, a) \leq 0$, i.e. for $a > a'$, $F(\cdot, a)$ *first-order stochastically dominates* $F(\cdot, a')$.

ASSUMPTION 6 (CDFC). For all $a \in \mathscr{A}$ and for all $\theta \in \Theta$, $F_{aa}(\theta, a) \geq 0$.

Assumption 4 is used in most of the adverse selection literature, which requires that the marginal utility of the buyer be monotonically increasing in type. Assumption 5 is weaker than the following Assumption 7 but is sufficient for some of the results. Assumption 6, called *convexity of distribution function condition,* and also used in the moral hazard literature, roughly speaking requires that the marginal return of investment be diminishing as investment increases.[5]

ASSUMPTION 7 (MLRP). For all $a \in \mathscr{A}$, $(d/d\theta)(f_a(\theta)/f(\theta)) > 0$.

This assumption, the *monotone likelihood ratio property,* is familiar in the moral hazard literature. The specific form given in Assumption 7 is due to Milgrom (1981), who also shows that MLRP implies FSD.

ASSUMPTION 8. The utility function $u(\theta, x)$ satisfies $u_{\theta\theta x} \leqslant 0$ and $u_{\theta xx} \geqslant 0$.[6]

---

[4]Also see Laffont and Maskin (1979). McAfee (1991) and McKelvey and Page (2002) extend the Myerson–Satterthwaite theorem to the case of divisible goods, but without investment.

[5]See Salanié (1997) for further discussions of these assumptions.

[6]This is assumption A8 in Fudenberg and Tirole (1991, page 263).

Assumption 8 is not necessary but simplifies the analysis. In particular, if the function u is separable ($u = \phi(\theta)y(x)$ with $\phi' \geq 0$, $y' \geq 0$ and $y'' \leq 0$), then the analysis still goes through even though $u_{\theta xx} \leq 0$.[7]

3. FORMULATING THE PROBLEM

In this quasi-linear environment, state-independent transfer payments from one agent to the other do not affect either agent's incentive. Therefore a budget-balanced *constrained efficient* (or simply *optimal*) mechanism should maximize the expected total surplus subject to the incentive constraints. The agents' relative bargaining power determines how the surplus should be split between them via lump-sum transfers but it does not affect the "shape" of optimal contract.

Specifically, an optimal contract should solve the following problem:

Program (P1):

$$\max_{x,t,a} V(x,a) \equiv \int_{\underline{\theta}}^{\overline{\theta}} \left[u(\theta, x(\theta)) - c(x(\theta)) \right] f(\theta, a)\, d\theta - g(a) \quad \text{subject to}$$

$$u(\theta, x(\theta)) - t(\theta) \geq u(\theta, x(\hat{\theta})) - t(\hat{\theta}) \quad \forall \theta, \hat{\theta} \in \Theta \qquad (\text{IC}_B)$$

$$a \in \underset{a' \in \mathcal{A}}{\arg\max} \int_{\underline{\theta}}^{\overline{\theta}} \left[t(\theta) - c(x(\theta)) \right] f(\theta, a')\, d\theta - g(a') \qquad (\text{IC}_S)$$

where (IC_B) and (IC_S) are the incentive constraints of the buyer and seller respectively. Note that there is no need to include ex ante participation constraints in the program as they can easily be satisfied by means of an upfront transfer provided that the optimal total surplus exceeds the sum of the agents' reservation utilities.[8]

The combination of hidden information and hidden action in Program (P1) creates some difficulties in solving the problem: the usual integration-by-parts and point-optimization techniques used to solve the standard hidden information models cannot be applied directly. To overcome these difficulties, I follow an approach adapted from Rogerson (1985) and proceed as follows. First, via a series of simplifications of the incentive constraints, Program (P1) is transformed into a relaxed program. Second, the relaxed program is solved using general optimal control methods that directly incorporate the incentive constraints. Finally, I show that a solution to the relaxed program is

---

[7]This is because one can define $\phi \equiv \phi(\theta)$ as type and $y \equiv y(x)$ as trade level, and define a new cost function $C(y) = c(y^{-1}(y))$, which is strictly convex. The model is then transformed into linear form: $u = \phi y$, which satisfies Assumption 8.

[8]As a contrast and for later reference, the problem of a monopolistic seller offering a profit-maximizing contract to the buyer subject to the latter's interim participation constraint is given as follows:

$$\max_{x,t} \int_{\underline{\theta}}^{\overline{\theta}} [t(\theta) - c(x(\theta))] f(\theta, a)\, d\theta \qquad (M)$$

subject to (IC_B) and $u(x(\theta)) - t(\theta) \geq \underline{v}$ for all θ, where $\underline{v}$ is the interim reservation utility of the buyer. This is the relevant problem if the contract is signed *after* investment.

in fact also a solution to the original program. Throughout the paper I focus on con-tracts (x,t) that are piecewise continuously differentiable.[9] This is to be consistent with the assumptions of the main theorems in optimal control theory (e.g. Leonard and Long 1992) on which the subsequent analysis draws heavily.[10]

To derive the relaxed program, I first characterize the buyer's information rent using standard arguments. By incentive compatibility, a type-θ buyer maximizes her utility $u(\theta,x(\hat{\theta}))-t(\hat{\theta})$ by reporting her true type $\hat{\theta}=\theta$, which leads to the first-order condition

$$u_x(\theta,x(\theta))x'(\theta)-t'(\theta)=0. \tag{1}$$

By the envelope theorem, the buyer's rent changes at the rate $u_\theta(\theta,x)$ as the type θ changes. In particular, relative to each θ, every type above θ must receive the same *additional* rent $u_\theta(\theta,x)d\theta$ as type $\theta+d\theta$ does. Given the measure of such types, the extra rent accrued to the types above θ equals $u_\theta(\theta,x(\theta))(1-F(\theta,a))d\theta$. Adding up all these incremental rents over all types yields the buyer's total rent:

$$\int_{\underline{\theta}}^{\overline{\theta}} u_\theta(\theta,x(\theta))(1-F(\theta,a))d\theta.$$

Subtracting the buyer's rent from the total surplus gives the seller's payoff:

$$R(a)\equiv\int_{\underline{\theta}}^{\overline{\theta}}\left\{\left[u(\theta,x(\theta))-c(x(\theta))\right]f(\theta,a)-u_\theta(\theta,x(\theta))(1-F(\theta,a))\right\}d\theta-g(a). \tag{2}$$

The relaxed program can now be defined as follows.

Program (P2):

$$\max_{x,a} V(x,a) \quad \text{subject to}$$

$$\frac{dx}{d\theta}\geqslant 0, x(\theta)\geqslant 0 \quad \forall\theta, \tag{3}$$

$$\frac{dR(a)}{da}\geqslant 0. \tag{4}$$

Appendix B explains in detail how this formulation is derived. The idea can be sum-marized as follows. First, one needs to focus only on the trade schedule x and the in-vestment level a when solving the problem because the transfer schedule $t(\theta)$ is com-pletely pinned down (up to a constant) by the buyer's incentive constraints once the

---

[9] A piecewise C^1 function admits derivatives except at a finite set of points, where left and right deriva-tives exist. Such a function therefore must be continuous.

[10] This treatment is without loss of generality when an optimal contract exists in the piecewise C^1 class: the surplus achievable using an arbitrary implementable contract can be approximated by piecewise C^1 implementable contracts and hence an optimal piecewise C^1 contract is optimal overall. An optimal con-tract can be discontinuous in bilateral trade models in which the good is indivisible and the total surplus is linear in the trade level (e.g. Myerson and Satterthwaite 1983); there an optimal contract does *not* exist within the piecewise C^1 class. I assume that an optimal piecewise C^1 contract exists in the present model. See Guesnerie and Laffont (1984) for a proof of the existence result when there is no investment; the surplus function in the current model satisfies their regularity conditions.

trade schedule is solved. Second, by a standard argument the buyer's incentive constraint (IC_B) can be replaced by the monotonicity condition in (3), as the latter is necessary and sufficient for the former. Third, following Rogerson (1985), the seller's incentive constraint (IC_S) is relaxed to the condition $dR(a)/da \geq 0$. The intuition is that at the seller's optimal choice $a > \underline{a}$, his marginal return from investment is necessarily nonnegative (otherwise he should decrease a). The reason for relaxing (IC_S) to this inequality rather than the usual first-order condition $dR(a)/da = 0$ is to ensure a nonnegative multiplier on the new constraint, which will become clear in the next section. The next lemma shows that one can indeed work with Program (P2) for the nontrivial situation where investment is above the minimum level $\underline{a}$, which is assumed to be the case for the rest of the paper.

LEMMA 1. *Suppose Assumptions 1–6 are satisfied. Then every solution (x^*, a^*) to Program (P2) with $a^* > \underline{a}$ is also a solution to (P1). If such a solution to (P2) exists, then an allocation (x, a), where $a > \underline{a}$, solves (P1) if and only if it solves (P2).*

The logic may be explained as follows. Note that every $a > \underline{a}$ satisfying the seller's original incentive constraint $a \in \arg\max_{a' \in \mathscr{A}} R(a')$ in (P1) also satisfies the first-order condition $dR(a)/da \geq 0$ in (P2) but the reverse is not necessarily true. Therefore if (x^*, a^*) with $a^* > \underline{a}$ solves (P2) and satisfies the constraints of (P1) then it also solves (P1).

4. MAIN RESULTS

In this section I characterize the optimal contract by solving Program (P2). I proceed by identifying the basic forces in the present model that shape the optimal contract. The key for motivating the seller to make the necessary investment is to offer him a payoff $t(\theta) - c(x(\theta))$ that is sufficiently upward-sloping as a function of θ. By (1), the slope of the seller's payoff function can be written as the product of the marginal surplus from trade, $u_x - c'(x)$, and the slope of the trade schedule $x'(\theta)$:

$$t'(\theta) - c'(x)x'(\theta) = (u_x(\theta, x(\theta)) - c'(x))x'(\theta). \tag{5}$$

This result offers a simple intuition for why full efficiency cannot be achieved in this model. Since the marginal surplus $u_x - c'(x)$ equals zero when trade is efficient, it is immediate from the above equation that the seller's payoff is independent of θ if trade is always efficient; then the seller of course has no incentive to invest.

More importantly, observe from (5) that to motivate the seller we need to make the trade schedule steep and keep the marginal surplus high. These considerations result in *downward* distortion in trade, as shown in the following proposition.

PROPOSITION 2. *Suppose Assumptions 1–5 are satisfied. If (x^*, a^*) is a solution to Program (P2), then $x^*(\theta)$ is less than or equal to the efficient trade level, $x^E(\theta)$, for all $\theta \in \Theta$.*

This result may look similar to the downward distortion result in standard nonlinear pricing with hidden information, but the reason is quite different: here it is the moral

hazard constraint that makes any upward distortion counterproductive; hidden information alone does *not* cause any distortion at all.[11] Because of this, the argument for Proposition 2 relies on a *global* approach rather than the typical approach using local first-order conditions. Specifically, by (5) any upward distortion in trade generates a corresponding decreasing segment in the seller's payoff function, which has a negative effect on the seller's incentive. Such a "lump" in the trade schedule can be removed by restoring trade to the efficient level, which relaxes the seller's incentive constraint and also keeps intact the monotonicity constraint on the trade schedule, thus resulting in a genuine increase in total surplus.

Due to the nonstandard constraints, general optimal control methods are needed to solve Program (P2). Below I outline the basic results of the analysis and leave most of the technical details to Appendix C.

I define $\xi = dx/d\theta$ as the control variable. The state variables include trade level x, investment level a, and an auxiliary variable that handles the integral constraint (4). The costate variables associated with the three state variables are respectively μ, γ, and λ.

The variable λ deserves some explanation as it plays an important role in the characterization. It is the multiplier on the constraint $dR(a^*)/da \geqslant 0$. As such it reflects the shadow price of relaxing this constraint. In particular, we can imagine that the constraint is of the form $dR(a^*)/da + \phi \geqslant 0$ where ϕ measures the change in the seller's marginal return from the use of a technology that adds a linear component ϕa to the seller's total return. Of course, $\phi = 0$ in the current problem. But if a technology with $\phi > 0$ is available, at the optimal allocation how much would the social planner be willing to pay for a one-unit increase in ϕ? The answer is λ.

When solving for the optimal contracts, it is useful to think of the solution as being obtained in two stages. In the first stage, an optimal trade schedule x is solved for each given investment level a; in the second stage, the optimal investment level a^* is determined. In Appendix C, I show that the necessary and sufficient conditions for a trade schedule x^* along with the costate variables $\mu(\theta)$ and $\lambda \in \mathbb{R}$ to maximize the expected total surplus while implementing the investment level a^* (or an arbitrary level a by dropping the asterisk) include the law of motion

$$\frac{d\mu}{d\theta} = -\left[u_x(\theta, x^*) - c'(x^*)\right]\left[f(\theta, a^*) + \lambda f_a(\theta, a^*)\right] - \lambda u_{\theta x}(\theta, x^*)F_a(\theta, a^*), \quad (6a)$$

the boundary conditions

$$\mu(\overline{\theta}) = 0, \mu(\underline{\theta})x^*(\underline{\theta}) = 0, \quad (6b)$$

and two complementary slackness conditions

$$\forall \theta \in \Theta, \mu(\theta) \leqslant 0, \frac{dx^*}{d\theta} \geqslant 0, \text{ and } \mu(\theta)\frac{dx^*}{d\theta} = 0 \quad (6c)$$

$$\lambda \geqslant 0, \frac{dR(a^*)}{da} \geqslant 0, \text{ and } \lambda\frac{dR(a^*)}{da} = 0. \quad (6d)$$

---

[11]The optimal contract simply demands the seller to invest at the efficient level and give the buyer the right to make a take-it-or-leave-it offer after her type is realized. The buyer offers to trade at the efficient level.

For the rest of the analysis in this section, I maintain Assumptions 1–4 and 6–8, and also assume $u_x(\theta,0) < \infty$ and $u_{\theta x}(\theta,0) < \infty$. These latter technical assumptions are made to avoid setting an unbounded value for the costate variable μ when $x^* = 0$ and are not as restrictive as they may seem. For instance if $u = \theta y(x)$ with $y'(0) = \infty$ we can redefine y as the quantity and transform u into a linear function.[12]

As it turns out, the characterization of optimal contracts depends crucially on whether and where the monotonicity constraint $dx^*/d\theta \geq 0$ binds. When the monotonicity constraint binds at a type θ, the trade schedule x^* may have to involve *bunching*, i.e. there is a neighborhood $\mathcal{O}$ of θ such that all types in the set $\mathcal{O} \cap \Theta$ receive the same level of trade. If bunching does not occur at θ then the complementary slackness condition (6c) implies that $\mu = 0$ in a neighborhood of θ, which by (6a) leads to the following observation.

LEMMA 2. *If there is no bunching at type θ then optimal trade $x^*(\theta)$ satisfies*

$$\left[u_x(\theta,x^*) - c'(x^*)\right]\left[f(\theta,a^*) + \lambda f_a(\theta,a^*)\right] + \lambda u_{\theta x}(\theta,x^*)F_a(\theta,a^*) = 0. \qquad (7)$$

This equation has the following interpretation. The optimal contract must balance the tradeoff between maximizing total surplus and providing the seller with an adequate incentive to invest. For each type θ, increasing the trade level $x^*(\theta)$ by a small amount dx raises the total surplus by

$$\left[u_x(\theta,x^*) - c'(x^*)\right]f(\theta,a^*)dx.$$

Increasing x also affects the expected total surplus through its impact on the seller's incentive. This can happen in two ways. First, the rent of all buyer types above θ is increased by $u_{\theta x}(1 - F(\theta,a))dx$, which reduces the seller's marginal return by $-u_{\theta x}F_a(\theta,a)dx$. This is called the *rent-shifting effect*. Second, the change in total surplus directly affects the seller's marginal return by $(u_x(\theta,x^*) - c'(x^*))f_a(\theta,a)dx$, which may be called the *surplus effect*. Since λ is the shadow price of relaxing the constraint on the seller's marginal return, the total incentive effect therefore is given by

$$\lambda\left[(u_x(\theta,x^*) - c'(x^*))f_a(\theta,a^*) + u_{\theta x}(\theta,x^*)F_a(\theta,a^*)\right]dx.$$

Equation (7) says that when there is no bunching, slightly adjusting the trade level $x^*(\theta)$ should have zero first-order effect on the optimal value, i.e. the above effects sum to zero.

Equation (7) reinforces the earlier argument that the buyer's information rent causes an efficiency loss only through its effect on the seller's incentive constraint. If the seller's incentive constraint is less stringent, i.e λ is small, then the effect of the information rent is small and vanishes when $\lambda = 0$.

---

[12]The appeal of the Inada condition $u_x(\theta,0) = \infty$ is to ensure an interior solution, which it however fails to deliver in the present model. By (8) or (10), even if $u_x(\theta,0) = \infty$ trade may still be zero as $u_{\theta x}(\theta,0)$ can also be ∞. The reason is that although for a given type the gain from a small amount of trade dx is of a higher order of magnitude, its incentive cost is also proportionally large. Think of $u = \theta\sqrt{x}$, which is equivalent to a linear model.
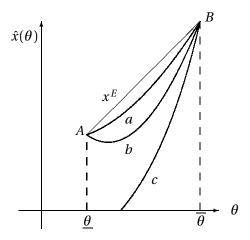
FIGURE 1. Trade schedules

The interaction between the seller's incentive problem and the buyer's self-selection shapes the optimal contract and leads to several features of the optimal trade schedule that are striking when compared with the standard models of screening/self-selection with hidden information.

PROPOSITION 3. *In a solution (x^*, a^*) to Program (P2), trade is efficient at the top: $x^*(\overline{\theta}) = x^E(\overline{\theta})$, and generically either (i) there is no distortion at the bottom: $x^*(\underline{\theta}) = x^E(\underline{\theta})$, or (ii) there is bunching and even exclusion of trade for a range of the bottom types.*

To better understand these features, it is essential to trace out the effects of different values of the shadow price λ on the optimal trade schedule. Consider first the case in which λ is small. Then the solution to (7), denoted by $\hat{x}$, should be close to the efficient schedule x^E and hence should be nondecreasing in θ given that x^E is strictly increasing; this implies that $\hat{x}$ is indeed optimal. Rewrite (7) as

$$u_x(\theta, x) - c'(x) = u_{\theta x}(\theta, x) \cdot K(\theta) \tag{8}$$

where

$$K(\theta) \equiv -\frac{\lambda F_a(\theta)}{f(\theta) + \lambda f_a(\theta)}. \tag{9}$$

The wedge $u_{\theta x} K(\theta)$ between marginal utility u_x and marginal cost $c'(x)$ of course explains why optimal trade generally deviates from the efficient level. Recall the incentive effects of a small reduction in trade for type θ and observe from the above equations that the wedge between u_x and $c'(x)$ is the rent-shifting effect $-u_{\theta x} F_a(\theta)$ multiplied by the term $\lambda/(f + \lambda f_a)$.

Moreover, note that the rent-shifting effect is equal to zero at the two end types $\overline{\theta}$ and $\underline{\theta}$; the reason is that changing the trade level of each end type shifts a *constant* amount of rent from the buyer to the seller for all θ, which of course has no effect on

the seller's marginal return. Absent this effect, the wedge between marginal utility and marginal cost disappears and the optimal trade simply maximizes the weighted surplus $(u(\theta,x)-c(x))(f+\lambda f_a)$. The result is efficient trade as long as $f+\lambda f_a > 0$, which is always true at the top and is true at the bottom as well when λ is small. The outcome, illustrated by curve a in Figure 1, is a trade schedule stretched downward from the efficient curve x^E with its two ends fixed at the efficient points A and B. This logic works fine as long as λ is small.

As the shadow price λ increases, the wedge $K(\theta)$ increases as well as long as it remains positive (below I discuss what happens when it is negative), implying a greater need to reduce the trade level. Intuitively, a higher shadow price makes increasing the seller's marginal return more attractive, thus making it profitable to shift more rent from the buyer to the seller through a reduction in the trade level.

This downward pressure on trade however is not uniform across all buyer types. Specifically, of the two incentive effects brought about by a reduction in the trade level, the rent-shifting effect is always positive but the surplus effect is not. In particular, a decrease in surplus for lower types where $f_a(\theta) < 0$ has a positive effect on the seller's marginal return and hence improves his incentive, but a decrease in surplus for the higher types where $f_a(\theta) > 0$ has a negative effect on the seller's marginal return and thus hurts his incentive. The intuition is that punishing the seller for bad performance is good for his incentive but punishing him for good performance is not (a reduction in surplus is a punishment for the seller because his payoff equals the surplus minus the buyer's rent). Therefore, the two incentive effects of downward trade distortion are mutually *reinforcing* at low types but are *countervailing* at high types. This difference is reflected in the wedge $u_{\theta x} K(\theta)$: the rent-shifting effect $-u_{\theta x} F_a(\theta)$ is amplified at lower types but is subdued at higher types through the factor $f(\theta) + \lambda f_a(\theta)$. Moreover, within the reinforcing region the two effects are still uneven: the rent-shifting effect equals zero at the left end $\underline{\theta}$ and the surplus effect vanishes at the right end where $f_a = 0$. As a result, the combined downward pressure on trade is strongest somewhere in the middle of the reinforcing region. A higher shadow price λ only amplifies this unevenness, making downward distortion especially pronounced in the lower portion of the trade schedule.

As λ increases, this downward force eventually becomes strong enough that the trade curve, with its two ends fixed at the efficient points, becomes (asymmetrically) U-shaped and a portion of it is stretched below the left and *lower* end (see curve b). At that moment, the monotonicity constraint $dx/d\theta \geq 0$ is violated at the lower end of the curve so bunching must occur. It is clear from this process that this type of bunching can occur only at the bottom rather than the top of the type space because the highest type always trades more goods than the other types.

For even more severe incentive problems, part of the trade curve is stretched down to the horizontal axis and bunching occurs at zero quantity, i.e. there is exclusion of trade for a range of the bottom types (curve c). This is easy to see when λ exceeds the threshold $-f(\underline{\theta})/f_a(\underline{\theta})$; then for θ close to $\underline{\theta}$ the left-hand side of (7) is less than zero for all $x \in (0, x^E(\theta))$, which means that any increase in trade up to the efficient level *always* has a negative marginal effect on the objective function. Therefore optimal trade is zero for such bottom types.

Observe that in this process there is no need to distort the trade level of the highest type and it is not optimal to distort the trade level of the lowest type unless bunching is necessary. Furthermore, the fact that the optimal trade schedule x^* meets the strictly increasing efficient schedule x^E at the highest type $\overline{\theta}$ implies that bunching never happens at the top: the curve x^* has to be strictly increasing near the point $\overline{\theta}$ because otherwise that part of x^* is above the efficient curve x^E, which is impossible by Proposition 2.

It should be emphasized that the type of bunching characterized in Proposition 3 does *not* depend on special sorting conditions (or lack thereof) on the probability distribution of the buyer's type, but rather on the strength with which the seller's incentive constraint binds at the optimum. This differs from standard nonlinear pricing, where full separation can be obtained if appropriate sorting conditions are imposed.

Bunching at a nonzero quantity for the lower types amounts to a *minimum purchase commitment*: regardless of the realized type the buyer commits to the purchase of a minimum quantity at a preset price. In practice a minimum purchase commitment can offer the firm some insurance for its R&D investments. But one may wonder whether a more flexible trade schedule can enhance efficiency in trade and also offer better incentive for the firm to invest. The analysis here suggests that although the exact level of the minimum quantity may vary, the qualitative feature is robust as long as the incentive problem is sufficiently severe.

To illustrate the results of Proposition 3, consider the following example adapted from Rogerson (1985).

EXAMPLE 1. The distribution of the buyer's type θ is given by

$$F(\theta, a) = \left(\frac{\theta - \underline{\theta}}{\overline{\theta} - \underline{\theta}} \right)^a .$$

This distribution satisfies MLRP and CDFC. Moreover, since $f_a(\underline{\theta})/f(\underline{\theta}) = -\infty$,[13] in any optimal solution the shadow price λ exceeds $-f(\underline{\theta})/f_a(\underline{\theta}) = 0$. By the preceding analysis in this case a range of the bottom types must be excluded from trade. In contrast, full separation is possible in the standard nonlinear pricing model with this same distribution. ◇

In general, due to the complex nature of the optimal contract it is difficult to obtain conditions on the primitives that foretell the occurrence of bunching; in the next section, I numerically compute a parameterized example to further illustrate the results of Proposition 3.

---

[13]By L'Hôpital's rule,

$$\left. \frac{a f_a(\theta)}{f(\theta)} \right|_{\theta=\underline{\theta}} = \ln\left(\frac{\theta - \underline{\theta}}{\overline{\theta} - \underline{\theta}} \right) \Big/ \left(\frac{\theta - \underline{\theta}}{\overline{\theta} - \underline{\theta}} \right)^{-1} = -\left(\frac{\theta - \underline{\theta}}{\overline{\theta} - \underline{\theta}} \right)^{-3} = -\infty.$$

Once the optimal trade schedule is determined, the transfer schedule can be obtained by integrating out (1):

$$t^*(\theta) = t^*(\underline{\theta}) + \int_{\underline{\theta}}^{\theta} u_x(\theta_1, x^*(\theta_1)) x^{*\prime}(\theta_1) \, d\theta_1$$

where $t^*(\underline{\theta})$ is a constant and should be chosen appropriately to satisfy any ex ante participation constraints. The transfer schedule should have the same bunching regions as the trade schedule. It is steeper than the first-best transfer schedule, at least for higher types. This can be seen from $t^{*\prime}(\theta) = u_x \cdot x^{*\prime}(\theta)$: downward distortion in trade increases the marginal utility u_x and also makes the trade schedule steeper near the highest type, thus raising the slope of the transfer schedule.

The analysis so far has highlighted the bunching of some bottom types due to the seller's incentive problem. It turns out that bunching in other parts of the type space is also hard to rule out. Indeed, compared with standard nonlinear pricing it is much harder to get even limited separation of types in the present model. The reason can be illustrated as follows. In the standard nonlinear pricing model, if one ignores the monotonicity constraint $dx/d\theta \geq 0$ then the trade schedule x^m is determined by the equation[14]

$$u_x - c'(x) = u_{\theta x} \cdot \frac{1}{h(\theta)} \tag{10}$$

where

$$h(\theta) \equiv \frac{f(\theta)}{1 - F(\theta)}$$

is the familiar *hazard rate*. Compared with (8), the difference between the two models lies in the terms $K(\theta)$ and $1/h(\theta)$.

To determine whether bunching is necessary at some point θ, the routine analysis is to totally differentiate (8) and then see if the sign of $dx/d\theta$ can be determined. The result, (C.15) in Appendix C, shows that the solution is nondecreasing at the point θ if $K'(\theta) \leq 1$. By the same token, the schedule x^m is nondecreasing at the point θ if $d(1/h)/d\theta \leq 1$. While this latter condition can be easily guaranteed by an increasing hazard rate, the condition $K'(\theta) \leq 1$ is hard to ensure without knowing the shadow price λ.

To delineate the factors that determine the occurrence of bunching in the present model, I introduce two measures that capture the effects of investment on the type distribution. First, let $S(\theta) = 1 - F(\theta)$ be the *survival* function, which is the probability of (e.g. a machine) surviving at least up to "time" θ.[15] Then the *event density* $f(\theta)$ is the likelihood of the bad event (machine failure) occurring per unit of time. Define the investment *elasticity of survival* as

$$\varepsilon^s = \frac{a}{S(\theta)} \cdot \frac{dS}{da} = \frac{-a F_a(\theta)}{1 - F(\theta)}$$

---

[14]One may derive this from Program (M) in footnote 8.

[15]The dependence of these functions on a is suppressed, which is done whenever no confusion can arise.

and the investment *elasticity of event* as

$$\varepsilon^f = \frac{a}{f(\theta)} \cdot \frac{df}{da} = \frac{a f_a(\theta)}{f(\theta)}.$$

With these measures, one can write

$$K(\theta) = \frac{1}{h(\theta)} \cdot \frac{\lambda \varepsilon^s(\theta)}{a + \lambda \varepsilon^f(\theta)}.$$

As in standard nonlinear pricing, the hazard rate h captures the effect of hidden information. But now there is a new moral hazard effect captured by the second term. This decomposition sheds light on the reason why it is hard to separate different buyer types in the current model. Even if one assumes both an increasing hazard rate and MLRP so that the denominator of $K(\theta)$ is strictly increasing in θ, the bound on $K'(\theta)$ is still not easy to establish because the numerator of $K(\theta)$ is also increasing in θ. Indeed, we have

$$\frac{d\varepsilon^s}{d\theta} = h(\theta)(\varepsilon^s(\theta) - \varepsilon^f(\theta)) \geq 0,$$

because $\varepsilon^s \geqslant \varepsilon^f$ by MLRP.[16] Therefore, it is not surprising that bunching is hard to prevent in the present model.

Nevertheless, the next proposition provides sufficient conditions to rule out any bunching other than that occurring at the bottom of the type distribution as characterized in Proposition 3. Under these conditions there is room for separating different buyer types when the seller's incentive problem is not too severe. A striking feature of these conditions is that they are much more stringent than the typical sufficient conditions for full separation to occur in the standard screening models.

PROPOSITION 4. *Suppose $f' \geqslant 0$, $f'_a \geqslant 0$, $f'' \geqslant 0$, and $f''_a \geqslant 0$.[17] Then an optimal trade schedule takes **one** of the following forms.*

(a) *Full Separation: there is full participation and full separation of all types.*

(b) *Partial Exclusion: there is some $\theta' \in (\underline{\theta}, \overline{\theta})$ such that $x^*(\theta) = 0$ for $\theta \in [\underline{\theta}, \theta']$ and there is participation and full separation for $\theta \in (\theta', \overline{\theta}]$.*

(c) *Partial Pooling: there is some $\theta' \in (\underline{\theta}, \overline{\theta})$ such that $x^*(\theta) = x^*(\underline{\theta}) > 0$ for $\theta \in [\underline{\theta}, \theta']$ and there is full separation for $\theta \in (\theta', \overline{\theta}]$.*

Proposition 4 and Proposition C.1 in the appendix are also useful for computing the optimal contracts, which is important as the model in general is difficult to solve analytically.[18] The results of Proposition 4 provide techniques for identifying the critical point θ' when bunching is present.

---

[16]By MLRP, for all $\theta' > \theta$, $f_a(\theta')f(\theta) \geq f(\theta')f_a(\theta)$. Integrating out θ' over the range $[\theta, \overline{\theta}]$ yields $f(\theta)(-F_a(\theta)) \geq f_a(\theta)(1 - F(\theta))$.

[17]Primes denote (partial) derivatives with respect to θ.

[18]The importance of computation in nonlinear pricing is illustrated by the work of Wilson (1993).
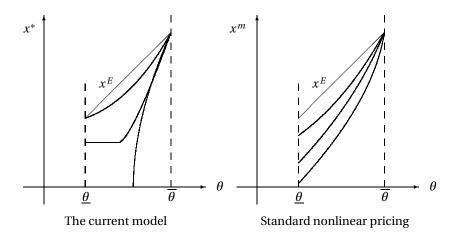
FIGURE 2. Optimal trade schedules

The final result of this section aims to isolate the effect of the seller's incentive constraint on optimal trade. To prepare for this result, suppose that we want to implement some *given* investment level $a > \underline{a}$ while maximizing total surplus. The previous analysis of optimal trade was carried out at the optimal investment level, but the characterizations clearly apply to the current problem of maximizing total surplus while implementing a fixed level of investment. What happens to the optimal trade schedule if the investment cost $g(a)$ changes? An important observation from (8) is that for a given investment level a any change in the investment cost function g affects the optimal trade schedule x^* *only* through its impact on the shadow price λ. Thus, by holding a constant we can focus on changes in optimal trade that are *purely* due to changes in the incentive problem without introducing effects related to changes in distribution $F(\theta, a)$.

PROPOSITION 5. *Consider implementing some given investment level a for fixed functions u, c and F. Let changes in the investment cost function g correspond to shadow prices $\lambda' > \lambda > 0$ in the optima; then the optimal trade schedules satisfy $x^*(\theta; \lambda') \leq x^*(\theta; \lambda)$ for all θ.*

Thus, although the optimal trade level may go up or down following a change in the investment cost, it must change in the same direction across all θ. Also, since the efficient trade schedule x^E is fixed and the constrained-optimal schedule satisfies $x^*(\theta; \lambda) \leq x^E$ for all $\lambda > 0$, this result implies that a more severe incentive problem leads to (weakly) more distortion in trade for *all* buyer types. To the extent that ex post distortion reflects a form of rigidity, Proposition 5 demonstrates a general positive relation between the seriousness of the seller's incentive problem and the rigidity in optimal trade.

5. AN EXAMPLE AND AN ALGORITHM

This section offers an algorithm for numerically solving the model. An example is computed to illustrate the characterization results obtained in the previous section.
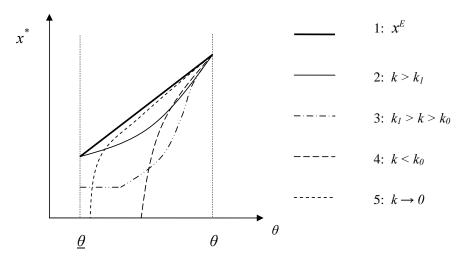
FIGURE 3. Optimal trading schedules in Example 2.

EXAMPLE 2. Assume $F(\theta, a) = a(\theta - \underline{\theta})^2 + (1-a)(\theta - \underline{\theta})$ where $\theta \in [\underline{\theta}, \overline{\theta} = \underline{\theta} + 1]$ and $a \in [0,1]$. Note that the distribution F is a mixture of the uniform distribution and a distribution that stochastically dominates the uniform distribution. The investment a shifts more weight from the uniform distribution to the better distribution. We also assume a linear utility function $u = \theta x$, a quadratic cost function $c(x) = x^2/2$, and an investment cost function $g(a) = k(1/(1-a) - 1 - a)$, where $k > 0$ is a parameter. Note that $g(0) = g'(0) = 0$ and $g(1) = g'(1) = g''(1) = \infty$. The first-best trade is simple: $x^E(\theta) = \theta$. This example satisfies all the assumptions. Moreover, we have

$$f = 2a(\theta - \underline{\theta}) + 1 - a; f_a = 2(\theta - \underline{\theta}) - 1.$$

Therefore $f' \geqslant 0$, $f'' = 0$, $f'_a > 0$, and $f''_a = 0$. Hence Proposition 4 is applicable. $\Diamond$

Even for this simple example, an analytical solution cannot be obtained. The numerical algorithm can be outlined as follows.

- For each pair (a, λ), use conditions (i)–(iv) in Proposition C.1 to derive the trade schedule $x(\theta; a, \lambda)$. Specifically, for $\lambda \leq -f(\underline{\theta})/2f_a(\underline{\theta})$ the solution can be derived using (8). Otherwise, we need to numerically find the critical point θ' given in Lemma C.2 below which bunching occurs.

- Define two nonlinear equations of a and λ by plugging $x(\theta, a, \lambda)$ into the seller's incentive constraint $dR/da = 0$ and the first-order condition (v) in Proposition C.1. This is done by creating a function that calls the function $x(\cdot, a, \lambda)$.

- Solve the two equations for a and λ. The trade schedule $x(\theta, a, \lambda)$ is then determined.

I solve this example for $\underline{\theta} = 0.1$ and for various values of the parameter k. The optimal trade schedules are schematically illustrated in Figure 3.[19] The shape of the optimal trade schedule depends on the value of k in relation to two critical values, $k_0 \approx 0.012$ and $k_1 \approx 0.068$. The results illustrate the conclusions of Propositions 3 and 4 (refer to the proof of Proposition 3 for the cutoff points of λ).

- $k > k_1 \Longrightarrow \lambda \le -f(\underline{\theta})/2f_a(\underline{\theta})$, so there is full participation and full separation with efficient trade at both ends (Curve 2).

- $k_0 < k < k_1 \Longrightarrow -f(\underline{\theta})/2f_a(\underline{\theta}) < \lambda < -f(\underline{\theta})/f_a(\underline{\theta})$, so bunching occurs at the bottom (Curve 3).

- $k < k_0 \Longrightarrow \lambda > -f(\underline{\theta})/f_a(\underline{\theta})$, so there is exclusion of trade (Curves 4 and 5).

- The optimal investment level a^* is inversely related to k.

It is interesting to note that the optimal trade schedule x^* is not monotonic in k. And as $k \downarrow 0$, curve x^* (line 5) is everywhere close to x^E except at the very low end of the type distribution, where x^* drops abruptly to zero. The idea is that as investment cost becomes very low, it is possible to provide an incentive by punishing the seller only in the very bad states; for all other states trade should be approximately efficient.

6. Overinvestment

The holdup problem in this model is illustrated by the following result.

OBSERVATION. *Let (a^*, x^*) be a solution to Program (P2) and let a^e be the investment choice that maximizes expected total surplus given the second-best trade schedule x^*. If both a^e and a^* are in the interior of $\mathscr{A}$ then $a^* < a^e$, i.e. conditional on x^* the seller underinvests.*

The reason is simple and can be sketched as follows. By definition, a^e satisfies

$$\int_\Theta [u(\theta, x^*(\theta)) - c(x^*(\theta))] f_a(\theta | a^e) d\theta - g'(a^e) = 0.$$

But the seller's incentive constraint requires

$$\int_\Theta [u(\theta, x^*(\theta)) - c(x^*(\theta))] f_a(\theta | a^*) d\theta + \int_\Theta u_\theta F_a(\theta | a^*) d\theta = g'(a^*)$$

which implies

$$\int_\Theta [u(\theta, x^*(\theta)) - c(x^*(\theta))] f_a(\theta | a^*) d\theta - g'(a^*) > 0.$$

---

[19]The schedules are not monotonically stacked as predicted by Proposition 5 because investment is not fixed.

By the concavity of the total surplus function, $a^* < a^e$.

However, such a definite relation does not exist between a^* and the first-best investment level a^E, which is determined by

$$g'(a^E) = \int_{\Theta} [u(\theta, x^E(\theta)) - c(x^E(\theta))] f_a(\theta, a^E) \, d\theta.$$

To see why, recall that the second-best investment level a^* is determined by

$$g'(a^*) = \int_{\Theta} [u(\theta, x^*(\theta)) - c(x^*(\theta)) - v(\theta)] f_a(\theta, a^*) \, d\theta$$

where $v(\theta)$ is the rent paid to a type-θ buyer. Notice that the second-best contract creates two opposite effects on the seller's incentive to invest. On the one hand, to elicit truthful revelation a type-θ buyer should be given rent $v(\theta)$, which increases in θ. Holding constant the contract, this effect reduces the seller's marginal return and hence discourages investment. On the other hand, in the second-best contract trade is distorted downward, especially for the low types. This reduces the total surplus and lowers the seller's payoff when types are low, which increases the seller's marginal return and encourages investment. When the second effect is strong enough, the second-best investment level a^* can be greater than the first-best level a^E.

To illustrate this point explicitly, consider a simple version of the model in which the buyer has only two types, $\theta_2 > \theta_1$, and the utility function of a buyer of type θ_i is given by $u(\theta_i, x) = \theta_i x$. Redefining variables if necessary we can take the seller's investment choice as the probability that type θ_2 will occur: $p \in [0, 1]$.[20] The investment cost is given by $g(p)$ with $g(0) = g'(0) = 0$, $g' > 0$, and $g'' > 0$. The cost function $c(x)$ satisfies $c' > 0$, $c'' > 0$, $c(0) = c'(0) = 0$, and $c'(\infty) = \infty$.

A constrained efficient mechanism $((x_i, t_i), p)$ solves the program

$$\max_{x_i, t_i, p} p(\theta_2 x_2 - c(x_2)) + (1 - p)(\theta_1 x_1 - c(x_1)) - g(p) \quad \text{subject to}$$

$$\theta_i x_i - t_i \geq \theta_i x_j - t_j, \quad i, j = 1, 2$$

$$t_2 - t_1 = c(x_2) - c(x_1) + g'(p).$$

Let x_i^E be the first-best efficient trade level for type θ_i buyer and let p^E be the first-best efficient investment. The following proposition characterizes the constrained efficient mechanisms.

PROPOSITION 6. *In an optimal contract, $x_2 = x_2^E$ and $x_1 < x_1^E$. Furthermore, let $A = \theta_2 x_1 - c(x_1)$ and $C = \theta_1 x_1^E - c(x_1^E)$. Then if $A > C$, there is underinvestment, $p < p^E$; if $A = C$, then $p = p^E$; and if $A < C$, there is overinvestment, $p > p^E$.*

---

[20]Given an increasing probability function $p(a)$ and cost function $\hat{g}(a)$, one can define p as effort and a new cost function $g(p) = \hat{g}(a^{-1}(p))$.

EXAMPLE 3. Assume $\theta_2 = 2$, $\theta_1 = 1$, $c(x) = x^2$, and $g(p) = kp^2$ where $k > 0$. Note that the first-best efficient trade levels are $x_i^E = \theta_i/2$ for $i = 1, 2$. It can be shown that there is a cutoff for the parameter k: $k_c \approx 0.5335$ such that if $k \in [0.5, k_c)$, there is *over*investment; if $k > k_c$, there is underinvestment; and if $k = k_c$, the investment level is first-best efficient.   ◊

To be sure, in the standard principal-agent model it is possible that the agent may pick an action level greater than the one he would choose in the absence of the principal. This is already pointed out by Shavell (1979) and Grossman and Hart (1983). However, the reason that "overwork" may arise in the principal-agent model is because the agent is risk averse and the principal can offer the agent partial insurance; if the agent is risk neutral then the first best can be achieved with or without the principal. In contrast, in the current model, despite the fact that both agents are risk neutral towards income the first best is not achievable. The agent overinvests primarily because the screening of buyer types can create favorable marginal returns on investment for the seller. Therefore the sources of distortion in an agent's action are different in the two models.

7. CONCLUSION

This paper has analyzed a bilateral trade problem in which the seller makes hidden investment that influences the buyer's hidden valuation, thus creating a holdup problem. Under broad conditions it is impossible for budget-balanced trading mechanisms to implement both first-best efficient investment and efficient trade. I fully characterize the constrained efficient contracts. Compared with standard nonlinear pricing without a holdup problem, the optimal trading schedule displays rigidity in the form of bunching when the holdup problem is severe. In particular, bunching and exclusion of trade can occur even if the distribution of buyer types satisfies the usual sorting conditions that guarantee full separation in the standard nonlinear pricing models. Moreover, given the investment level, as the holdup problem becomes more severe trade distortion relative to the efficient level becomes larger across the board for all buyer types.

This paper contributes to the literature by considering a rich model of bilateral trade with a holdup problem. It identifies new sources and novel features of trade and investment distortions. As contract theorists are well aware, specific results in contract theory are sensitive to the information and technological structures of the model. By deriving implications of optimal contracting under alternative information and technological structures this paper takes a further step towards a better understanding of their roles in real-world bilateral trade problems.

APPENDIX

A. THE INEFFICIENCY THEOREM

PROOF OF PROPOSITION 1.   Let $(x^E(\theta), t(\theta))$ be a mechanism that implements a^E. Since a^E is an interior point of $\mathscr{A}$, the following first-order condition holds:

$$\int_{\underline{\theta}}^{\overline{\theta}} [u(\theta, x^E(\theta)) - c(x^E(\theta))] \, dF_a(\theta, a^E) = g'(a^E). \tag{A.1}$$

Similarly, incentive compatibility for the seller implies

$$\int_{\underline{\theta}}^{\overline{\theta}} [t(\theta) - c(x^E(\theta))] \, dF_a(\theta, a^E) = g'(a^E). \tag{A.2}$$

Then subtracting (A.2) from (A.1) yields

$$\int_{\underline{\theta}}^{\overline{\theta}} [u(\theta, x^E(\theta)) - t(\theta)] \, dF_a(\theta, a^E) \, d\theta = 0.$$

Namely, the marginal effect of investment on the buyer's expected payoff is equal to zero. However, I show next that the incentive constraints of the buyer imply

$$\int_{\underline{\theta}}^{\overline{\theta}} [u(\theta, x^E(\theta)) - t(\theta)] \, dF_a(\theta, a^E) > 0, \tag{A.3}$$

thus yielding a contradiction.

First, by Assumption 2, $x^E(\theta') > 0$ and $\theta'' > \theta'$ imply $x^E(\theta'') > 0$. Therefore the set Θ^* must be an interval $[\theta_0, \overline{\theta}]$ for some $\theta_0 < \overline{\theta}$ and $x^E(\theta_0) \geq 0$, $x^E(\theta) > 0$ for all $\theta \in (\theta_0, \overline{\theta}]$.

Define $U(\theta) \equiv u(\theta, x^E(\theta)) - t(\theta)$. Then the incentive constraint of the buyer and Assumption 2 imply that for $\theta'' > \theta'$,

$$U(\theta'') \geq u(\theta'', x^E(\theta')) - t(\theta') \geq u(\theta', x^E(\theta')) - t(\theta') = U(\theta'),$$

and the second inequality is strict if $x^E(\theta') > 0$. Therefore the function U is strictly increasing on the set $\Theta^* = [\theta_0, \overline{\theta}]$.

The distribution function $F(\cdot, a^E)$ has at most countably many jump discontinuities at as many mass points. On the set $[\theta_0, \overline{\theta}]$, divide every interval between two adjacent mass points (including the end points $\theta_0, \overline{\theta}$) further into sub-intervals of non-zero measures such that the function $F_a(\cdot, a^E)$ is either strictly increasing, strictly decreasing, or constant on the interior of each sub-interval; treat each mass point θ as a single sub-interval with its "monotonicity" determined by the sign of $F_a(\theta, a^E) - F_a(\theta^-, a^E)$. Denote the sub-intervals by $[\theta_0, \theta_1], \ldots, [\theta_n, \overline{\theta}]$, where $n \leq \infty$.

I prove that for $j = 0, 1, \ldots, n$,

$$\int_{\underline{\theta}}^{\theta_j} U(\theta) \, dF_a(\theta, a^E) \geqslant U(\theta_j) \int_{\underline{\theta}}^{\theta_j} dF_a(\theta, a^E) \tag{A.4}$$

"$>$" if $\exists [\theta_{m-1}, \theta_m], 1 \leq m \leq j$, on which $F_a(\cdot, a^E)$ is not constant.

Since by assumption such an interval $[\theta_{m-1}, \theta_m]$ always exists, in the limit (A.4) implies

$$\int_{\underline{\theta}}^{\overline{\theta}} U(\theta) \, dF_a(\theta, a^E) > U(\overline{\theta}) \int_{\underline{\theta}}^{\overline{\theta}} dF_a(\theta, a^E) = 0,$$

which is (A.3) and we are done.

Equation (A.4) is shown by induction as follows. Consider $j = 0$. If $\theta_0 = \underline{\theta}$ then the result is trivially true. Suppose $\theta_0 > \underline{\theta}$. By definition, $x^E(\theta) = 0$ for all $\theta \in [\underline{\theta}, \theta_0]$; then the incentive constraint of the buyer implies $U(\theta) = U(\theta_0)$ for all $\theta \in [\underline{\theta}, \theta_0]$. It follows that

$$\int_{\underline{\theta}}^{\theta_0} U(\theta) \, d F_a(\theta, a^E) = U(\theta_0) \int_{\underline{\theta}}^{\theta_0} d F_a(\theta, a^E).$$

Now suppose (A.4) holds for $j = 0, 1, \dots$. Consider $j + 1$. Without loss of generality, suppose $F_a(\cdot, a^E)$ is not constant on (θ_j, θ_{j+1}).

Suppose $F_a(\cdot, a^E)$ decreases on (θ_j, θ_{j+1}). Since U is strictly increasing, we have

$$\int_{\theta_j}^{\theta_{j+1}} U(\theta) \, d F_a(\theta, a^E) > U(\theta_{j+1}) \int_{\theta_j}^{\theta_{j+1}} d F_a(\theta, a^E). \tag{A.5}$$

Adding (A.5) to the induction hypothesis (A.4) yields

$$\int_{\underline{\theta}}^{\theta_{j+1}} U(\theta) \, d F_a(\theta, a^E) > U(\theta_j) \int_{\underline{\theta}}^{\theta_j} d F_a(\theta, a^E) + U(\theta_{j+1}) \int_{\theta_j}^{\theta_{j+1}} d F_a(\theta, a^E)$$

$$\geq U(\theta_{j+1}) \int_{\underline{\theta}}^{\theta_{j+1}} d F_a(\theta, a^E)$$

The second inequality follows from $U(\theta_{j+1}) > U(\theta_j)$ and $\int_{\underline{\theta}}^{\theta_j} d F_a(\theta, a^E) \leq 0$.

Next suppose $F_a(\cdot, a^E)$ increases on (θ_j, θ_{j+1}). Since U is strictly increasing, it follows that

$$\int_{\theta_j}^{\theta_{j+1}} U(\theta) \, d F_a(\theta, a^E) > U(\theta_j) \int_{\theta_j}^{\theta_{j+1}} d F_a(\theta, a^E).$$

Adding this to the induction hypothesis (A.4) yields

$$\int_{\underline{\theta}}^{\theta_{j+1}} U(\theta) \, d F_a(\theta, a^E) > U(\theta_j) \int_{\underline{\theta}}^{\theta_{j+1}} d F_a(\theta, a^E) \geq U(\theta_{j+1}) \int_{\underline{\theta}}^{\theta_{j+1}} d F_a(\theta, a^E),$$

where the second inequality follows from $U(\theta_{j+1}) > U(\theta_j)$ and $\int_{\underline{\theta}}^{\theta_{j+1}} d F_a(\theta, a^E) \leq 0$.

The above argument proves (A.4), which in turn implies (A.3) and completes the proof.   □

B. Formulating program (P2)

The following characterization of the buyer's incentive constraint IC_B is standard (see e.g. Guesnerie and Laffont 1984 or Fudenberg and Tirole 1991, Theorem 7.3 for a proof).

LEMMA B.1. *Suppose Assumptions 1–4 are satisfied. A piecewise differentiable schedule (x, t) satisfies (IC$_B$) if and only if $(\partial u/\partial x)(dx/d\theta) = dt/d\theta$ and $dx/d\theta \geq 0$.*

Let $v(\theta) \equiv u(\theta, x(\theta)) - t(\theta)$ be the total utility that a type θ buyer receives when she reports her type truthfully. Then by Lemma B.1 we have

$$\frac{dv}{d\theta} = \frac{\partial u}{\partial \theta} + \frac{\partial u}{\partial x}\frac{dx}{d\theta} - \frac{dt}{d\theta} = \frac{\partial u}{\partial \theta} \quad \forall \theta,$$

which implies

$$v(\theta) = \int_{\underline{\theta}}^{\theta} \frac{\partial u}{\partial \theta}(\tau, x(\tau)) d\tau + v(\underline{\theta}).$$

Then the seller's payoff is given by

$$R(a) = \int_{\underline{\theta}}^{\overline{\theta}} (t(\theta) - c(x(\theta))) d\theta - g(a) = \int_{\underline{\theta}}^{\overline{\theta}} (u(\theta, x(\theta)) - c(x(\theta)) - v(\theta)) d\theta - g(a).$$

Substituting for $v(\theta)$ and using integration by parts, we obtain the expression for $R(a)$ in (2) and arrive at Program (P2) in the text.

PROOF OF LEMMA 1. I first show that the first-order constraint is binding: $R'(a^*) = 0$. Suppose to the contrary that the constraint is not binding. Then with the only active constraints $dx/d\theta \geq 0$ and $x(\theta) \geq 0$, an optimal solution x^* has to coincide with the efficient trade schedule: $x^*(\theta) = x^E(\theta)$ for all θ. It then follows from equation (C.1) that $R'(a) = -g''(a)$. Therefore $R'(a^*) < 0$ for $a^* > \underline{a}$, a contradiction.

Next I show that $R''(a) \leq 0$, for all a. By equation (C.1),

$$R''(a) = -\int_{\underline{\theta}}^{\overline{\theta}} \left[u_x(\theta, x^*(\theta)) - c'(x^*(\theta)) \right] \frac{dx^*}{d\theta} F_{aa}(\theta, a) d\theta - g''(a).$$

By Proposition 2, $x^*(\theta) \leq x^E(\theta)$ for all θ, which implies $u_x(\theta, x^*(\theta)) - c'(x^*(\theta)) \geq 0$ for all θ. Since $dx^*/d\theta \geq 0$, $F_{aa} \geq 0$ (by CDFC), and $g'' \geq 0$, it follows that $R''(a) \leq 0$.

Therefore any point a at which $R'(a) = 0$ is indeed a maximizer of $R(a)$. Therefore (x^*, a^*) satisfies the constraints of (P1), so it solves (P1).

Finally, any other allocation (x, a) with $a > \underline{a}$ that solves (P1) must yield the same surplus as (x^*, a^*) does, so it must also solve (P2) because it is within the choice set of (P2) and (x^*, a^*) solves (P2). □

C. CHARACTERIZATION RESULTS

PROOF OF PROPOSITION 2. First note that by the definition of $R(a)$, we have

$$\frac{dR}{da} = \int_{\underline{\theta}}^{\overline{\theta}} \left\{ \left[u(\theta, x(\theta)) - c(x(\theta)) \right] f_a(\theta, a) + u_\theta(\theta, x(\theta)) F_a(\theta, a) \right\} d\theta - g'(a).$$

Using integration by parts, we have

$$
\frac{dR}{da} = -\int_{\underline{\theta}}^{\overline{\theta}} \left\{ \left[u_x(\theta, x(\theta)) - c_x(x(\theta)) \right] x'(\theta) F_a(\theta, a) \right\} d\theta - g'(a). \tag{C.1}
$$

I prove the result via contradiction. Suppose $x^*(\theta) > x^E(\theta)$ for some θ. Since x^* is continuous,[21] there must exist $\theta' < \theta''$ such that $x^*(\theta) > x^E(\theta)$ for all $\theta \in (\theta', \theta'')$, $x^*(\theta') \geqslant x^E(\theta')$ ("$>$" only if $\theta' = \underline{\theta}$), and $x^*(\theta'') \geqslant x^E(\theta'')$ ("$>$" only if $\theta'' = \overline{\theta}$).

Consider the following alternative trade schedule x^{**} defined by: $x^{**}(\theta) = x^E(\theta)$ for all $\theta \in [\theta', \theta'']$ and $x^{**}(\theta) = x^*(\theta)$ otherwise. Since x^* is non-decreasing it follows that x^{**} is also non-decreasing. Moreover, since $dx^*/d\theta \geqslant 0$, $F_a \leq 0$, and $x^*(\theta) > x^E(\theta)$ for all $\theta \in (\theta', \theta'')$, it follows from (C.1) that replacing x^* with x^{**} weakly relaxes the constraint $dR/da \geqslant 0$. Therefore (x^{**}, a^*) satisfies the constraints of (P2). Note that the expected surplus is higher with this new contract because inefficient trade is replaced by efficient trade on (θ', θ''). This contradicts the optimality of solution (x^*, a^*). □

To reformulate Program (P2) as an optimal control problem, define a new state variable k as follows. For all $\hat{\theta} \in \Theta$, let

$$
k(\hat{\theta}) = \int_{\underline{\theta}}^{\hat{\theta}} \left\{ \left[u(\theta, x) - c(x) \right] f_a(\theta, a^*) + u_\theta(\theta, x) F_a(\theta, a^*) - f(\theta, a) g'(a) \right\} d\theta. \tag{C.2}
$$

With the other two state variables x, a, and the control variable $\xi = dx/d\theta$, Program (P2) can be reformulated as the following optimal control problem.

Program (Q):

$$
\max_{x,k,a,\xi} \int_{\underline{\theta}}^{\overline{\theta}} \left[u(\theta, x(\theta)) - c(x(\theta)) - g(a) \right] f(\theta, a) d\theta \quad \text{subject to}
$$

$$
\dot{x} = \xi
$$
$$
\dot{k} = \left[u(\theta, x(\theta)) - c(x(\theta)) \right] f_a(\theta, a) + u_\theta(\theta, x(\theta)) F_a(\theta, a) - f(\theta, a) g'(a)
$$
$$
\dot{a} = 0
$$
$$
\xi \geqslant 0, x(\underline{\theta}) \geqslant 0, k(\underline{\theta}) = 0, k(\overline{\theta}) \geqslant 0.
$$

Note that the constraint $\dot{x} = \xi(\theta) \geqslant 0$ and boundary condition $x(\underline{\theta}) \geqslant 0$ ensure $x(\theta) \geqslant 0$ for all θ.

Let μ, λ, γ be the costate variables associated with the state variables x, k, a respectively. The Hamiltonian is then given by

$$
H \equiv \mu(\theta)\xi(\theta) + \left[u(\theta, x) - c(x) \right] \left[f(\theta, a) + \lambda f_a(\theta, a) \right]
$$
$$
+ \lambda u_\theta(\theta, x) F_a(\theta, a) - f(\theta, a) \left[g(a) + \lambda g'(a) \right].
$$

The following result characterizes the necessary and sufficient conditions for a solution to (Q).

---

[21] Recall that x^* is piecewise differentiable.

PROPOSITION C.1. *If (x^*, a^*) is a solution to Program (Q) with $a^* > \underline{a}$ then there exist a piecewise C^1 function $\mu : \Theta \to \mathbb{R}$ and a real number λ such that the following conditions hold.*

(i) $d\mu/d\theta = -[u_x(\theta, x^*) - c'(x^*)][f(\theta, a^*) + \lambda f_a(\theta, a^*)] - \lambda u_{\theta x}(\theta, x^*)F_a(\theta, a^*)$

(ii) $\mu(\overline{\theta}) = 0, \mu(\underline{\theta})x^*(\underline{\theta}) = 0$

(iii) *for all* $\theta \in \Theta$, $\mu(\theta) \leqslant 0$, $dx^*/d\theta \geqslant 0$, *and* $\mu(\theta)(dx^*/d\theta) = 0$

(iv) $\lambda \geqslant 0$, $dR(a^*)/da \geqslant 0$, *and* $\lambda(dR(a^*)/da) = 0$

(v) $\int_{\underline{\theta}}^{\overline{\theta}} \partial H(a^*)/\partial a \geqslant 0$; "$= 0$" *if* $a^* < \overline{a}$.

Conversely, if there exist μ and λ such that conditions (i) - (iv) are satisfied then x^* is an optimal path given a^*.

Conditions (i)–(iv) are the same as (6a)-(6d) in the text, reproduced here for convenience. The necessity, which follows from the Pontryagin Maximum Principle, is relatively standard. The sufficiency part, however, is not. The main complication is that the Hamiltonian is not necessarily concave in x when $f(\theta, a^*) + \lambda f_a(\theta, a^*) < 0$. As will be seen below this possibility is not merely a technical nicety but rather has an important relation with the severity of the holdup problem.

PROOF.　By the Pontryagin Maximum Principle, optimal paths x^*, k^*, a^*, and ξ^* satisfy the following necessary conditions.

$$\forall \theta, \xi(\theta) \text{ maximizes } H(\xi) \text{ subject to } \xi \geqslant 0 \tag{C.3}$$

$$\dot{x} = \frac{\partial H}{\partial \mu} \tag{C.4}$$

$$\dot{k} = \frac{\partial H}{\partial \lambda} \tag{C.5}$$

$$\dot{a} = \frac{\partial H}{\partial \gamma} = 0 \tag{C.6}$$

$$\dot{\mu} = -\frac{\partial H}{\partial x} \tag{C.7}$$

$$\dot{\lambda} = -\frac{\partial H}{\partial k} \tag{C.8}$$

$$\dot{\gamma} = -\frac{\partial H}{\partial a} \tag{C.9}$$

$$\mu(\underline{\theta}) \leqslant 0, \mu(\underline{\theta})x(\underline{\theta}) = 0, \mu(\overline{\theta}) = 0 \tag{C.10}$$

$$\lambda(\overline{\theta}) \geqslant 0, k(\overline{\theta}) \geqslant 0, \lambda(\overline{\theta})k(\overline{\theta}) = 0 \tag{C.11}$$

$$\gamma(\underline{\theta}) = \gamma(\overline{\theta}) = 0. \tag{C.12}$$

Equation (C.3) is the optimization condition for the control variable ξ. Equations (C.4) to (C.9) are the laws of motion for x, k, a, μ, λ, and γ respectively. The last three equations are the transversality conditions for μ, λ, and γ respectively.

Since H is linear in ξ, the optimization condition (C.3) implies condition (iii) of the proposition. Conditions (i) and (ii) simply follow from (C.7) and (C.10). Since H does not depend on k, (C.8) implies that λ is constant. Then (C.11) implies (iv). Finally (v) follows from (C.9) and (C.12).

As is standard, conditions (i)–(iv) are sufficient for x^*, k^*, and ξ^* to be optimal given a^* if the Hamiltonian H is concave in x, k, and ξ. A simple inspection tells us that H is indeed concave if $f(\theta, a^*) + \lambda f_a(\theta, a^*) \geqslant 0$. But this is not necessarily true for all θ; so we need to establish sufficiency explicitly.

Let x^*, μ, and λ satisfy conditions (i)–(iv) of the proposition. Suppose $f(\theta, a^*) + \lambda f_a(\theta, a^*) < 0$ for some θ. Then by MLRP there exists $\hat{\theta}$ such that $f(\theta, a^*) + \lambda f_a(\theta, a^*) < 0$ for $\theta < \hat{\theta}$ and $f(\theta, a^*) + \lambda f_a(\theta, a^*) \geqslant 0$ for $\theta \geqslant \hat{\theta}$. By Lemma C.1 below, we have $x^*(\theta) = 0$ for all $\theta \in [\underline{\theta}, \hat{\theta}]$.

Now we show that any piecewise C^1 schedule x that satisfies the constraints of Program (Q) is dominated by x^*. By Proposition 2, we need only consider schedules satisfying $x \leqslant x^E$ (recall that x^E is the first-best schedule). The proof modifies standard arguments for sufficiency when the Hamiltonian H is concave in the state and control variables x, k, and ξ.[22]

Let $V^* = \int_{\underline{\theta}}^{\overline{\theta}} [u(\theta, x^*) - c(x^*(\theta))] f(\theta, a^*) d\theta$ and $V = \int_{\underline{\theta}}^{\overline{\theta}} [u(\theta, x) - c(x(\theta))] f(\theta, a^*) d\theta$ be the alternative expected surpluses. To simplify notation, a^* is suppressed in the density and distribution functions. Let H^* be the Hamiltonian with the starred state and control variables x^*, k^*, and ξ^*, and let H be the Hamiltonian with variables x, k, and ξ. Then, since $x^* = 0$ on $[\underline{\theta}, \hat{\theta}]$, we have

$$V^* - V = \int_{\underline{\theta}}^{\hat{\theta}} -[u(\theta, x) - c(x)] f(\theta) \, d\theta$$

$$+ \int_{\hat{\theta}}^{\overline{\theta}} \left\{ [u(\theta, x^*) - c(x^*)] f(\theta) - [u(\theta, x) - c(x)] f(\theta) \right\} d\theta$$

$$= \int_{\underline{\theta}}^{\hat{\theta}} -[u(\theta, x) - c(x)] f(\theta) \, d\theta + \int_{\hat{\theta}}^{\overline{\theta}} \left\{ (H^* - \mu \dot{x}^* - \lambda \dot{k}^*) - (H - \mu \dot{x} - \lambda \dot{k}) \right\} d\theta$$

which from integration by parts

$$= \int_{\underline{\theta}}^{\hat{\theta}} -[u(\theta, x) - c(x)] f(\theta) \, d\theta + \int_{\hat{\theta}}^{\overline{\theta}} \left\{ (H^* - H) + \dot{\mu} x^* + \dot{\lambda} k^* - (\dot{\mu} x + \dot{\lambda} k) \right\} d\theta$$

$$- \left\{ \mu(\overline{\theta}) x^*(\overline{\theta}) + \lambda k^*(\overline{\theta}) - \mu(\hat{\theta}) x^*(\hat{\theta}) - \lambda k^*(\hat{\theta}) \right\}$$

$$+ \left\{ \mu(\overline{\theta}) x(\overline{\theta}) + \lambda k(\overline{\theta}) - \mu(\hat{\theta}) x(\hat{\theta}) - \lambda k(\hat{\theta}) \right\}.$$

---

[22]See for example Leonard and Long (1992, Theorem 6.5.2 on page 213).

Since $\dot{\lambda} = \mu(\overline{\theta}) = x^*(\hat{\theta}) = k^*(\hat{\theta}) = \lambda k^*(\overline{\theta}) = 0$, $\mu(\hat{\theta}) \leqslant 0$, $x(\hat{\theta}) \geqslant 0$, and $\lambda k(\overline{\theta}) \geqslant 0$, the above

$$\geqslant \int_{\underline{\theta}}^{\hat{\theta}} -\left\{ \left[u(\theta,x) - c(x) \right] f(\theta) + \lambda k(\hat{\theta}) \right\} + \int_{\hat{\theta}}^{\overline{\theta}} \left\{ (H^* - H) + \dot{\mu}(x^* - x) \right\} d\theta$$

which by using the definition of $k(\hat{\theta})$ in (C.2)

$$= \int_{\underline{\theta}}^{\hat{\theta}} -\left\{ \left[u(\theta,x) - c(x) \right] (f(\theta) + \lambda f_a(\theta)) + \lambda u_\theta(\theta,x) F_a(\theta) \right\} d\theta$$

$$+ F(\hat{\theta}) g'(a^*) + \int_{\hat{\theta}}^{\overline{\theta}} \left\{ (H^* - H) + \dot{\mu}(x^* - x) \right\} d\theta.$$

Since the first two terms are nonnegative, by the concavity of H on $[\hat{\theta}, \overline{\theta}]$ we have

$$V^* - V \geqslant \int_{\hat{\theta}}^{\overline{\theta}} \left\{ (H^* - H) - \frac{\partial H^*}{\partial x}(x^* - x) \right\} d\theta \geqslant 0,$$

completing the proof.   □

LEMMA C.1. *If* $\lambda > -f(\underline{\theta})/f_a(\underline{\theta})$ *then* $x^*(\theta) = 0$ *for all* $\theta \in [\underline{\theta}, \hat{\theta}]$, *where* $\hat{\theta}$ *satisfies* $f(\hat{\theta}, a^*) + \lambda f_a(\hat{\theta}, a^*) = 0$.

PROOF. By MLRP, $f(\theta, a^*) + \lambda f_a(\theta, a^*) < 0$ for $\theta < \hat{\theta}$ and $f(\theta, a^*) + \lambda f_a(\theta, a^*) > 0$ for $\theta > \hat{\theta}$. We also have $u_x(\theta, x^*) - c'(x^*) \geq 0$ since $x^* \leq x^E$. It follows that for all $\theta < \hat{\theta}$,

$$\dot{\mu} = -\left[u_x(\theta, x^*) - c'(x^*) \right] \left[f(\theta, a^*) + \lambda f_a(\theta, a^*) \right] - \lambda u_{\theta x}(\theta, x^*) F_a(\theta, a^*) > 0.$$

Since $\mu \leqslant 0$, this implies $\mu(\theta) < 0$ for all $\theta < \hat{\theta}$. Then by the complementary slackness condition (6c), x^* is constant on $[\underline{\theta}, \hat{\theta}]$. Since $x^*(\underline{\theta}) = 0$ by (6b), we have $x^*(\theta) = 0$ for all $\theta \in [\underline{\theta}, \hat{\theta}]$.   □

PROOF OF LEMMA 2.   If there is no bunching at θ, then x^* is strictly increasing on the interval $(\theta - \epsilon, \theta)$ or $(\theta, \theta + \epsilon)$ or both for some small $\epsilon > 0$. By condition (iii) in Proposition C.1, on an interval where $dx^*/d\theta > 0$, costate variable $\mu = 0$, hence $\dot{\mu} = 0$. By condition (i) this implies $-\partial H/\partial x = 0$ on such an interval. By continuity, $x^*(\theta)$ satisfies the same equation.   □

PROOF OF PROPOSITION 3.   To prove the efficiency at the top, let $\hat{x}(\theta)$ be the trade level, if it exists, that solves (7) at types where $f + \lambda f_a > 0$. Note that $\hat{x}(\overline{\theta})$ equals the efficient trade $x^E(\overline{\theta})$. Suppose the optimal trade $x^*(\overline{\theta}) < x^E(\overline{\theta})$. Then by Lemma 2 bunching must occur at $\overline{\theta}$. Let the optimal schedule x^* cross $\hat{x}$ at some $\theta_1 < \overline{\theta}$ so that $x^* < \hat{x}$ at every $\theta \in (\theta_1, \overline{\theta}]$. We show that x^* is not optimal, thus yielding a contradiction. Specifically, on the interval $(\theta_1, \overline{\theta}]$ replace x^* with $\hat{x}$. Then the total expected surplus is clearly increased

since $\hat{x}$ deviates less from the efficient schedule on $(\theta_1, \overline{\theta}]$. The new schedule is still non-decreasing. Moreover, observe from (C.1) that the seller's marginal return is increased because the integrand was initially zero $(dx^*/d\theta = 0)$ but now becomes positive on the interval $(\theta_1, \overline{\theta}]$. Therefore the new schedule satisfies all the constraints and is better than x^*. This contradiction proves $x^*(\overline{\theta}) = x^E(\overline{\theta})$.

Let $\delta = -f(\underline{\theta})/f_a(\underline{\theta})$ be the inverse of the percentage change in $f(\underline{\theta})$ following a small change in investment. Then the rest of the results follow from these three statements:

(a) if $\lambda > \delta$, there is exclusion of trade for a range of the bottom types

(b) if $\delta/2 < \lambda < \delta$, there is bunching at $\underline{\theta}$

(c) if $0 \leqslant \lambda < \delta/2$, either $x^*(\underline{\theta}) = x^E(\underline{\theta})$ or there is bunching at $\underline{\theta}$.

Part (a) follows directly from Lemma C.1.

Part (b): Suppose the statement is not true. Then by Lemma 2, $x^*(\underline{\theta})$ satisfies (8). Totally differentiating (8) with respect to θ, we obtain

$$\left(u_{xx} - c''(x) - K(\theta) \cdot u_{\theta xx} \right) \frac{dx^*}{d\theta} = u_{\theta x} \left(\frac{dK}{d\theta} - 1 \right) + u_{\theta\theta x} \cdot K(\theta). \tag{C.13}$$

Differentiate (9) and use the fact that $F_a(\underline{\theta}) = 0$ to obtain

$$K'(\underline{\theta}) = -\frac{\lambda f_a}{f + \lambda f_a} + \frac{\lambda F_a(f' + \lambda f_a')}{(f + \lambda f_a)^2} = -\frac{\lambda f_a}{f + \lambda f_a}$$

where primes denote derivatives with respect to θ. Then the hypothesis $\delta/2 < \lambda < \delta$ implies

$$K'(\underline{\theta}) > 1.$$

Combining this with the conditions $K(\underline{\theta}) = 0$, $u_{xx} - c''(x) < 0$, and $u_{\theta x} > 0$, we obtain from (C.13) $dx^*(\underline{\theta})/d\theta < 0$, which is a contradiction.

Part (c): If there is no bunching at $\underline{\theta}$ then by Lemma 2, $x^*(\underline{\theta})$ satisfies (8). Since $F_a(\underline{\theta}) = 0$, (8) implies $u_x(\theta, x^*(\underline{\theta})) - c'(x^*(\underline{\theta})) = 0$ and hence $x^*(\underline{\theta}) = x^E(\underline{\theta})$. □

PROOF OF PROPOSITION 5.   Let $\lambda' \geq \lambda > 0$. By Lemma C.1, at points (if any) where $1 + \lambda'(f_a(\theta)/f(\theta)) < 0$, $x^*(\theta; \lambda') = 0 \leq x^*(\theta; \lambda)$. Thus we need only consider points where $1 + \lambda'(f_a(\theta)/f(\theta)) > 0$. Note also that at such points $1 + \lambda(f_a(\theta)/f(\theta)) > 0$.

For a given $\lambda > 0$, at each point where $f(\theta) + \lambda f_a(\theta) > 0$ let $\hat{x}(\theta; \lambda)$ be the maximizer of the Hamiltonian H. Note that when $f(\theta) + \lambda f_a(\theta) > 0$ the Hamiltonian H is strictly concave in x; hence the maximizer $\hat{x}(\theta)$ is either the unique x satisfying $\partial H/\partial x = 0$ or is equal to zero.

We first show that $\hat{x}(\theta; \lambda') \leq \hat{x}(\theta; \lambda)$ for $\lambda' > \lambda > 0$. At points where (8) (i.e. $\partial H/\partial x = 0$) admits the unique solution $\hat{x}(\theta; \lambda)$, differentiate (8) with respect to λ:

$$[u_{xx}(\theta, \hat{x}) - c''(\hat{x}) - K \cdot u_{\theta xx}] \frac{d\hat{x}}{d\lambda} = u_{\theta x} \frac{dK}{d\lambda}.$$

Since $u_{xx}(\theta, \hat{x}) - c''(\hat{x}) < 0$, $K \geq 0$, $u_{\theta xx} \geq 0$, $u_{\theta x} > 0$, and $dK/d\lambda > 0$, we have $d\hat{x}/d\lambda < 0$. If (8) does not admit a solution, so that $\hat{x}(\theta; \lambda) = 0$, then (8) also does not admit a solution for $\lambda' > \lambda$, so $\hat{x}(\theta; \lambda') = 0$. In conclusion, $\hat{x}(\theta; \lambda') \leq \hat{x}(\theta; \lambda)$.

Since by Lemma 2 $x^*(\theta; \lambda) = \hat{x}(\theta; \lambda)$ when there is no bunching at θ, it follows that $x^*(\theta; \lambda') \leq x^*(\theta; \lambda)$ whenever neither schedule involves bunching at θ.

Suppose $x^*(\lambda') > x^*(\lambda)$ at some $\tilde{\theta}$. Then one of the schedules is flat at this point. For types sufficiently close to $\overline{\theta}$, both $x^*(\lambda)$ and $x^*(\lambda')$ are strictly increasing and hence $x^*(\lambda') < x^*(\lambda)$. Therefore $x^*(\lambda')$ must cross $x^*(\lambda)$ at some point $\theta_1 < \overline{\theta}$ with $x^*(\theta; \lambda') < x^*(\theta; \lambda)$ for all $\theta \in (\theta_1, \overline{\theta})$. Moreover, one of the following cases must hold for types below θ_1.

Case (a): $x^*(\lambda')$ crosses $x^*(\lambda)$ at some $\theta_0 < \theta_1$ with $x^*(\lambda') > x^*(\lambda)$ on (θ_0, θ_1).

Case (b): $x^*(\lambda') > x^*(\lambda)$ on interval $[\underline{\theta}, \theta_1)$.

Consider case (a). Since $x^*(\lambda')$ increases at θ_0, by condition (iii) of Proposition C.1 $\mu(\theta_0; \lambda') = 0$. Also, since $x^*(\theta_1; \lambda') = x^*(\theta_1; \lambda) = \hat{x}(\theta_1; \lambda) > \hat{x}(\theta_1; \lambda')$, by the concavity of H we have

$$\dot{\mu}(\theta, \lambda') = -\frac{\partial H(x^*(\theta; \lambda'), \lambda')}{\partial x} > 0$$

for θ less than but near θ_1. Hence $\mu(\theta_1, \lambda') < 0$. Therefore

$$\int_{\theta_0}^{\theta_1} -\frac{\partial H(x^*(\theta; \lambda'), \lambda')}{\partial x} \, d\theta = \mu(\theta_1; \lambda') - \mu(\theta_0; \lambda') < 0.$$

Moreover, for $\theta \in [\theta_0, \theta_1]$, $x^*(\theta; \lambda') \geq x^*(\theta; \lambda)$ and $\lambda' > \lambda$ imply

$$\frac{\partial H(x^*(\theta; \lambda'), \lambda')}{\partial x} < \frac{\partial H(x^*(\theta; \lambda'), \lambda)}{\partial x} \leq \frac{\partial H(x^*(\theta; \lambda), \lambda)}{\partial x}.$$

Combining the above inequalities, we have

$$\mu(\theta_1; \lambda) - \mu(\theta_0; \lambda) = \int_{\theta_0}^{\theta_1} -\frac{\partial H(x^*(\theta; \lambda), \lambda)}{\partial x} \, d\theta < 0$$

which given $\mu(\theta_0; \lambda) \leq 0$ implies $\mu(\theta_1; \lambda) < 0$. However, since $x^*(\lambda)$ increases at θ_1, the complementary slackness condition (iii) implies $\mu(\theta_1; \lambda) = 0$, a contradiction.

For case (b), note that by condition (ii), $x^*(\underline{\theta}; \lambda') > x^*(\underline{\theta}; \lambda) \geq 0$ implies $\mu(\underline{\theta}; \lambda') = 0$. The result then follows from the same argument for case (a) by taking $\theta_0 = \underline{\theta}$.   □

The following lemma is used to prove Proposition 4.

LEMMA C.2. *Let (x^*, a^*) be a solution to Program (Q) with $a^* > \underline{a}$ and let μ and λ be the associated costate variables. Recall that $K(\theta) = -\lambda F_a(\theta)/(f(\theta) + \lambda f_a(\theta))$. If $K'(\theta) < 1$ at each point θ where $f(\theta) + \lambda f_a(\theta) > 0$ then either*

   (i) there is full participation and full separation of all types, or

(ii) there exists some θ', which may equal $\underline{\theta}$, such that there is exclusion of trade for types $\theta \in [\underline{\theta}, \theta']$ and there is full separation and participation for types above θ'. Specifically,

$$f(\underline{\theta}) + \lambda f_a(\underline{\theta}) \begin{cases} > 0 \\ = 0 \\ < 0 \end{cases} \implies \begin{cases} (i) \\ (i) \ or \ (ii) \\ (ii) \end{cases} and \quad x^*(\underline{\theta}) \begin{cases} = x^E(\underline{\theta}) \\ \in [0, x^E(\underline{\theta})] \\ = 0. \end{cases}$$

PROOF.   The solution satisfies the necessary conditions (i) - (iv) in Proposition C.1, which we show imply the desired properties of x^*. First, at points where $f(\theta) + \lambda f_a(\theta) > 0$, consider as before the schedule $\hat{x}(\theta)$ determined by the equation $\partial H / \partial x = 0$ or

$$u_x - c'(x) = u_{\theta x} \cdot K(\theta), \tag{C.14}$$

where $K(\theta)$ is defined in (9) in the text. By the Implicit Function Theorem we can totally differentiate (C.14) with respect to θ and get

$$\left(u_{xx} - c''(x) - K(\theta) \cdot u_{\theta xx} \right) \frac{d\hat{x}}{d\theta} = u_{\theta x} \left(\frac{dK}{d\theta} - 1 \right) + K(\theta) \cdot u_{\theta\theta x}. \tag{C.15}$$

By Assumption 8 and the conditions $u_{xx} - c''(x) \leqslant 0$, $K \geqslant 0$, and $dK/d\theta < 1$, (C.15) implies $d\hat{x}/d\theta > 0$. Therefore if (C.14) is solved by $\hat{x}(\theta') \geqslant 0$ at some point θ' then it is solved by $\hat{x}(\theta) > 0$ for all $\theta > \theta'$ and $\hat{x}$ is strictly increasing at those points.

Case 1: $f(\underline{\theta}) + \lambda f_a(\underline{\theta}) > 0$. Then MLRP implies $f(\theta) + \lambda f_a(\theta) > 0$ for all θ. Therefore $\hat{x}$ is defined for every θ and is strictly increasing. By Lemma 2, if the solution x^* differs from $\hat{x}$ at some point then it is flat around the point. Therefore given that $\hat{x}$ is strictly increasing if $x^*(\theta) < \hat{x}(\theta)$ then x^* is constant on $[\theta, \overline{\theta}]$ and if $x^*(\theta) > \hat{x}(\theta)$ then x^* is constant on $[\underline{\theta}, \theta]$. The former is impossible as it violates efficiency at the top. The latter implies that x^* is greater than efficient trade x^E at points near $\underline{\theta}$, which is impossible by Proposition 2. In conclusion, $x^* = \hat{x}$ and $x^*(\theta) = x^E(\theta)$ for $\theta = \underline{\theta}, \overline{\theta}$.

Case 2: $f(\underline{\theta}) + \lambda f_a(\underline{\theta}) < 0$. Then by MLRP there exists some $\hat{\theta} \in (\underline{\theta}, \overline{\theta})$ below which $K(\theta) < 0$, above which $K(\theta) > 0$, and for which $K(\hat{\theta}) = \infty$. Let $\theta' = \inf\{\theta > \hat{\theta} : \hat{x}(\theta) \geqslant 0 \text{ solves (C.14)}\}$. Then the schedule $\hat{x}$ is strictly increasing on $[\theta', \overline{\theta}]$ with $\hat{x}(\theta') = 0$ and $\hat{x}(\overline{\theta}) = x^E(\overline{\theta})$. We need to show only that $x^* = \hat{x}$ on $[\theta', \overline{\theta}]$, as monotonicity then implies $x^* = 0$ on $[\underline{\theta}, \theta']$. Recall that x^* is flat whenever $x^* \neq \hat{x}$. Hence, if $x^* > \hat{x} \geq 0$ at some $\theta'' \geq \theta'$, then $x^*(\hat{\theta}) = x^*(\theta'') > 0$, which contradicts Lemma C.1. On the other hand, if $x^* < \hat{x}$ at some $\theta'' \geq \theta'$, then x^* is constant on $[\theta'', \overline{\theta}]$, violating efficiency at the top. The desired result then follows.

Case 3: $f(\underline{\theta}) + \lambda f_a(\underline{\theta}) = 0$. The property of the schedule x^* now depends on the value of

$$K(\underline{\theta}) = \lim_{\theta \downarrow \underline{\theta}} - \frac{\lambda F_a(\theta)}{f(\theta) + \lambda f_a(\theta)} = 1 \left. \middle/ \frac{d}{d\theta} \left(\ln \frac{-f_a(\underline{\theta})}{f(\underline{\theta})} \right) \right.$$

and the boundary value $u_x(\underline{\theta}, 0) - c'(0)$. If (C.14) admits a solution $\hat{x}(\underline{\theta}) \geq 0$ at the point $\underline{\theta}$ then $x^* = \hat{x}$ for all θ and *only* in this nongeneric case we may have $x^*(\underline{\theta}) < x^E(\underline{\theta})$ and

no bunching at $\underline{\theta}$. Otherwise, let $\theta' = \inf\{\theta : \hat{x}(\theta) \geqslant 0$ solves (C.14)$\}$; then arguments similar to the ones in Case 2 imply $x^* = 0$ on $[\underline{\theta}, \theta']$ and $x^* = \hat{x}$ on $(\theta', \overline{\theta}]$. This completes the proof. $\qquad\square$

PROOF OF PROPOSITION 4. Again let $\delta = -f(\underline{\theta})/f_a(\underline{\theta})$. Then the proposition follows from the following three statements.

(a) If $0 \leqslant \lambda \leqslant \delta/2$ then there is full participation and separation with $x^*(\underline{\theta}) = x^E(\underline{\theta})$.

(b) If $\lambda > \delta$ then there is some θ' such that $x^* = 0$ on $[\underline{\theta}, \theta']$ and there is participation and full separation on $(\theta', \overline{\theta}]$.

(c) If $\delta/2 < \lambda \leqslant \delta$ then there is some $\theta' \geqslant \underline{\theta}$ such that $x^*(\theta) = x^*(\underline{\theta}) \geqslant 0$ for $\theta \in [\underline{\theta}, \theta']$ and there is full separation on $(\theta', \overline{\theta}]$.

(a) We need only the assumption $f' > 0$ for this part. First note that $f + \lambda f_a > 0$ for all θ. Differentiating $K(\theta) = -\lambda F_a/(f + \lambda f_a)$ yields

$$K'(\theta) = -\frac{\lambda f_a}{f + \lambda f_a} + \frac{\lambda F_a(f' + \lambda f_a')}{(f + \lambda f_a)^2} \tag{C.16}$$

where a prime ($'$) denotes a derivative with respect to θ. By MLRP, we have

$$\frac{d}{d\theta}\left(\frac{f_a}{f}\right) = \frac{f_a' f - f_a f'}{f^2} > 0$$

and hence $f_a' > f_a f'/f$, which together with $f' > 0$ and $f + \lambda f_a > 0$ implies

$$f' + \lambda f_a' > f' + \lambda \frac{f_a f'}{f} = \frac{f'}{f}(f + \lambda f_a) > 0. \tag{C.17}$$

Now by (C.16) and using $F_a \leq 0$ and MLRP, we have

$$K'(\theta) \leqslant -\frac{\lambda f_a}{f + \lambda f_a} \leqslant -\frac{\lambda f_a(\underline{\theta})}{f(\underline{\theta}) + \lambda f_a(\underline{\theta})} \leqslant 1 \quad \forall \theta.$$

The result then follows from Lemma C.2.

(b) Define $A(\theta) = F_a(f' + \lambda f_a') - f_a(f + \lambda f_a)$. Then whenever $f + \lambda f_a \geq 0$, we have

$$A'(\theta) = F_a(f'' + \lambda f_a'') - f_a'(f + \lambda f_a) \leqslant 0. \tag{C.18}$$

From (C.16), we have

$$K'(\theta) = \lambda \frac{F_a(f' + \lambda f_a') - f_a(f + \lambda f_a)}{(f + \lambda f_a)^2} \equiv \frac{\lambda A}{(f + \lambda f_a)^2}. \tag{C.19}$$

Let $\hat{\theta}$ be the point where $f + \lambda f_a = 0$. Since $f' + \lambda f_a' \geq 0$ by (C.17), at the point $\hat{\theta}$ we have

$$A(\hat{\theta}) = \underbrace{F_a(f' + \lambda f_a')}_{\leq 0} - \underbrace{f_a(f + \lambda f_a)}_{=0} \leqslant 0,$$

which together with (C.18) and (C.19) implies $A(\theta) \leq 0$ and $K'(\theta) \leq 0$ for all $\theta \geq \hat{\theta}$. The result then follows from Lemma C.2.

(c) If $\lambda = \delta$ then by the preceding argument, $A(\theta) \leqslant 0$ and hence $K'(\theta) \leq 0$ for all θ. Thus by Lemma C.2 either there is full participation and full separation or there is some θ' below which there is no trade and above which there is participation and full separation.

If $\delta/2 < \lambda < \delta$ then $f(\theta) + \lambda f_a(\theta) > 0$ for all θ, which by (C.18) implies that $A'(\theta) \leq 0$ for all θ. Since $A(\theta)$ and $K'(\theta)$ have the same sign and $A(\underline{\theta}) > 0$, it follows that there is some $\theta^* \in (\underline{\theta}, \overline{\theta})$ such that $K'(\theta) \geq 0$ for $\theta \leq \theta^*$ and $K'(\theta) \leq 0$ for $\theta \geq \theta^*$. Now (C.19) together with $A'(\theta) \leq 0$ and $f' + \lambda f'_a \geq 0$ implies that $K'(\theta)$ decreases on $[\underline{\theta}, \theta^*]$. Since

$$K'(\underline{\theta}) = -\frac{\lambda f_a(\underline{\theta})}{f(\underline{\theta}) + \lambda f_a(\underline{\theta})} > 1,$$

it follows that there exists some $\hat{\theta} \leq \theta^*$ below which $K'(\theta) \geq 1$ and above which $K'(\theta) \leq 1$. Then the pointwise maximizer of the Hamiltonian, $\hat{x}(\theta)$, is decreasing on $[\underline{\theta}, \hat{\theta}]$ and is nondecreasing on $[\hat{\theta}, \overline{\theta}]$. "Ironing" this U-shaped schedule then implies that the optimal solution x^* involves exactly one bunching region at the bottom, which is the desired result.   □

PROOF OF PROPOSITION 6.   First substitute the last constraint into the incentive constraints to obtain

$$\theta_2 x_2 - \theta_2 x_1 \geq c(x_2) - c(x_1) + g'(p) \tag{C.20}$$
$$\theta_1 x_1 - \theta_1 x_2 \geq -c(x_2) + c(x_1) - g'(p). \tag{C.21}$$

I show that in an optimal solution the constraint (C.21) is not binding.

Suppose both (C.21) and (C.20) are binding. Then $x_2 = x_1$ and $p = 0$. The optimal solution has to be $x_2 = x_1 = x_1^E$. However, this solution is dominated by (x_i^E, p^*) where $p^* > 0$ is given by $\theta_2 x_2^E - \theta_2 x_1^E = c(x_2^E) - c(x_1^E) + g'(p^*)$.

Suppose (C.21) is binding and (C.20) is not. Consider two cases. Case 1: $x_1 \neq x_1^E$. One can change x_1 slightly toward x_1^E to some x_1' so that (C.20) is intact and $\theta_1 x_1' - c(x_1') > \theta_1 x_1 - c(x_1)$. Hence constraint (C.21) is relaxed and the value of the objective function is increased, which is impossible. Case 2: $x_1 = x_1^E$. Then the only way for (C.21) to be binding is to have $x_2 = x_1^E$ and $p = 0$, which apparently is not optimal. Thus (C.21) cannot be binding.

Next we show that the optimal solution must have $x_2 = x_2^E$. If not, one can slightly change x_2 toward x_2^E, which increases total surplus without violating any constraint.

Note that (C.20) and (C.21) imply $x_2 \geq x_1$. Therefore the problem becomes

$$\max_{x_1, p} p(\theta_2 x_2^E - c(x_2^E)) + (1 - p)(\theta_1 x_1 - c(x_1)) - g(p) \quad \text{subject to}$$

$$x_2^E \geq x_1$$
$$\theta_2 x_2^E - c(x_2^E) \geq \theta_2 x_1 - c(x_1) + g'(p). \tag{C.22}$$

Let $\lambda \geq 0$ be the multiplier on (C.22). The first-order conditions are

$$(1 - p)(\theta_1 - c'(x_1)) \leq \lambda(\theta_2 - c'(x_1)) \quad (\text{"} = \text{" if } x_1 > 0) \tag{C.23}$$
$$\theta_2 x_2^E - c(x_2^E) - (\theta_1 x_1 - c(x_1)) = g'(p) + \lambda g''(p). \tag{C.24}$$

Suppose $\lambda = 0$. Then by (C.23), $x_1 = x_1^E$. But (C.22) and (C.24) imply that $\theta_2 x_1 - \theta_1 x_1 = 0$, which is impossible. Therefore $\lambda > 0$ (hence (C.22) is binding). Note also that x_2^E must be strictly greater than x_1 for otherwise $\theta_2 - c'(x_1) = 0$, which is inconsistent with (C.23). It follows that $\theta_2 - c'(x_1) > 0$. Since $\lambda > 0$ it follows from (C.23) that $\theta_1 - c'(x_1) > 0$. Thus $x_1 < x_1^E$.

To prove the latter half of the proposition, simply note that p^E and the second-best investment p satisfy the following conditions respectively:

$$\theta_2 x_2^E - c(x_2^E) - [\theta_1 x_1^E - c(x_1^E)] = g'(p^E)$$

$$\theta_2 x_2^E - c(x_2^E) - [\theta_2 x_1 - c(x_1)] = g'(p). \qquad \square$$

References

Armstrong, Mark (1996), "Multiproduct nonlinear pricing." *Econometrica*, 64, 51–75. [88]

Barder, Owen, Michael Kremer, and Heidi Williams (2006), "Advance market commitments: A policy to stimulate investment in vaccines for neglected diseases." *Economists' Voice*, 3, 1–6. [86]

Che, Yeon-Koo and Donald B. Hausch (1999), "Cooperative investments and the value of contracting." *American Economic Review*, 89, 125–147. [88]

Fudenberg, Drew and Jean Tirole (1991), *Game theory*. MIT Press, Cambridge, Massachusetts. [91, 108]

Grossman, Sanford J. and Oliver D. Hart (1983), "An analysis of the principal-agent problem." *Econometrica*, 51, 7–45. [106]

Grout, Paul A. (1984), "Investment and wages in the absence of binding contracts: A Nash bargaining approach." *Econometrica*, 52, 449–460. [88]

Guesnerie, Roger and Jean-Jacques Laffont (1984), "A complete solution to a class of principal-agent problems with an application to the control of a self-managed firm." *Journal of Public Economics*, 25, 329–369. [88, 93, 108]

Hart, Oliver D. and John Moore (1988), "Incomplete contracts and renegotiation." *Econometrica*, 56, 755–785. [88]

Hori, Kazumi (2006), "Inefficiency in a bilateral trading problem with cooperative investment." *B.E. Journals in Theoretical Economics: Contributions to Theoretical Economics*, 6, 1–11. [87]

Jullien, Bruno (2000), "Participation constraints in adverse selection models." *Journal of Economic Theory*, 93, 1–47. [88]

Konakayama, Akira, Toshihide Mitsui, and Shinichi Watanabe (1986), "Efficient contracting with reliance and a damage measure." *Rand Journal of Economics*, 17, 450–457. [88]

Laffont, Jean-Jacques and Eric Maskin (1979), "A differentiable approach to expected utility maximizing mechanisms." In *Aggregation and Revelation of Preferences* (Jean-Jacques Laffont, ed.), 289–308, North-Holland, Amsterdam. [91]

Leonard, Daniel and Ngo Van Long (1992), *Optimal Control Theory and Static Optimization in Economics*. Cambridge University Press, Cambridge. [93, 112]

Levin, Jonathan (2003), "Relational incentive contracts." *American Economic Review*, 93, 835–857. [88]

Lewis, Tracy R. and David E. M. Sappington (1989), "Countervailing incentives in agency problems." *Journal of Economic Theory*, 49, 294–313. [88]

MacLeod, W. Bentley (2003), "Optimal contracting with subjective evaluation." *American Economic Review*, 93, 216–240. [88]

Maskin, Eric and John Riley (1984), "Monopoly with incomplete information." *Rand Journal of Economics*, 15, 171–196. [88]

McAfee, R. Preston (1991), "Efficient allocation with continuous quantities." *Journal of Economic Theory*, 53, 51–74. [91]

McKelvey, Richard D. and Talbot Page (2002), "Status quo bias in bargaining: An extension of the Myerson–Satterthwaite theorem with an application to the Coase theorem." *Journal of Economic Theory*, 107, 336–355. [91]

Milgrom, Paul R. (1981), "Good news and bad news: Representation theorems and applications." *Bell Journal of Economics*, 12, 380–391. [91]

Mussa, Michael and Sherwin Rosen (1978), "Monopoly and product quality." *Journal of Economic Theory*, 18, 301–317. [88]

Myerson, Roger B. and Mark A. Satterthwaite (1983), "Efficient mechanisms for bilateral trading." *Journal of Economic Theory*, 29, 265–281. [91, 93]

Rochet, Jean-Charles and Philippe Choné (1998), "Ironing, sweeping, and multidimensional screening." *Econometrica*, 66, 783–826. [88]

Rochet, Jean-Charles and Lars A. Stole (2002), "Nonlinear pricing with random participation." *Review of Economic Studies*, 69, 277–311. [88]

Rogerson, William P. (1985), "The first-order approach to principal-agent problems." *Econometrica*, 53, 1357–1367. [92, 94, 99]

Rogerson, William P. (1992), "Contractual solutions to the hold-up problem." *Review of Economic Studies*, 59, 777–793. [85, 88]

Salanié, Bernard (1990), "Sélection adverse et aversion pour le risque." *Annales d'Économie et de Statistique*, 131–149. [88]

Salanié, Bernard (1997), *The Economics of Contracts*. MIT Press, Cambridge, Massachusetts. [88, 91]

Schmitz, Patrick W. (2002a), "On the interplay of hidden action and hidden information in simple bilateral trading problems." *Journal of Economic Theory*, 103, 444–460. [87, 91]

Schmitz, Patrick W. (2002b), "Simple contracts, renegotiation under asymmetric information, and the hold-up problem." *European Economic Review*, 46, 169–188. [88]

Shavell, Steven (1979), "On moral hazard and insurance." *Quarterly Journal of Economics*, 93, 541–562. [106]

Williamson, Oliver E. (1979), "Transaction-cost economics: The governance of contractural relations." *Journal of Law and Economics*, 22, 233–261. [88]

Wilson, Robert (1993), *Nonlinear Pricing*. Oxford University Press, New York. [101]