

AUGUST 2026

“Optimal Fixed-Terms Contracts under One-Sided Enforcement and Persistent Adverse Selection”

David Martimort and Aggey Simons (Semenov)

Optimal Fixed-Terms Contracts under One-Sided Enforcement and Persistent Adverse Selection

DAVID MARTIMORT

Toulouse School of Economics and Institut Universitaire de France, david.martimort@tse-fr.eu

AGGEY SIMONS (SEMENOV)

University of Ottawa, aggey.simons@uottawa.ca

August 4, 2026

ABSTRACT. We characterize the optimal contract within the class of fixed-terms mechanisms in a repeated buyer-seller relationship with persistent adverse selection and one-sided limited enforcement. The seller reports his type once, and the corresponding transfer and output are then repeated in every period. After receiving the current transfer, the seller may breach, pay an enforceable penalty, and terminate the relationship. In this benchmark, the enforcement problem collapses to a bound on the transfer targeted to the most efficient type. This yields a three-regime characterization. With strong enforcement, the repeated static second-best contract is feasible. With weak (intermediate) enforcement, the top transfer is capped, inducing bunching among efficient types and additional downward distortions. With very weak enforcement, sustaining compliance through public penalties alone would require excessive quantity distortions, and the optimal contract instead leaves strictly positive continuation rents, including for the least efficient type. We interpret the associated distortion as a virtual enforcement cost.

KEYWORDS. Adverse selection, Limited enforcement, Relational contracts, Contract breach.

JEL CLASSIFICATION. D82, D86, K12, C61.

1. INTRODUCTION

In many procurement, construction, and trade relationships, payment must precede at least part of performance. This timing creates a simple but important contractual friction: after receiving payment, the seller may prefer to retain the payment and fail to deliver. We study this problem in a repeated buyer-seller relationship with persistent private information about the seller's marginal cost. The seller reports his type once, at the beginning of the relationship. Conditional on that report, the buyer specifies a transfer and an output that are repeated in every period. In each period, the buyer prepays; the seller then either delivers or keeps the current payment, pays an enforceable penalty, and terminates the relationship. We call this a *fixed-terms mechanism*. It is stationary in the specific sense that the same report-contingent terms are repeated after a single initial report.

The empirical relevance of this friction is clear in environments where formal enforcement is weak. [Antràs and Foley \(2015\)](#) document that cash-in-advance terms are used more intensively in international transactions with weaker contractual enforcement. [Macchiavello and Morjaria \(2015\)](#) show, in the Kenyan rose export sector, that when formal contracts are hard to enforce, trade relies heavily on the value of future relationships and reliability improves as relationships mature. In procurement and construction, retention, performance bonds, and adjudication play related roles by limiting the risk that a contractor is paid and then fails to perform. These examples all point to the same interaction: public enforcement and the value of future trade jointly determine which contractual promises are credible.

The transaction sequence explains the one-sided nature of the enforcement problem. Once the buyer has prepaid, her current obligation has been fulfilled, whereas delivery remains under the seller's control. Following [Martimort et al. \(2017a\)](#), we interpret Π as a reduced-form measure of the maximum expected monetary consequence that public enforcement can impose and collect following non-delivery. Depending on the application, it may represent recoverable damages, forfeited collateral, a collectible claim under a performance bond, or another verifiable sanction. We treat Π as a fixed institutional primitive, common across types, in order to isolate how a given enforcement technology changes the screening menu.

The paper characterizes the optimal fixed-terms mechanism. The key observation is that the seller's enforcement problem collapses to a restriction on the transfer intended for the most efficient type. This reduction yields a sharp three-regime taxonomy. Under *strong enforcement*, the repeated static second-best contract is feasible. Under *weak enforcement*, the transfer to the most efficient type is capped by the penalty, so incentive compatibility propagates that cap into bunching among efficient types and additional downward distortions. Under *very weak enforcement*, compliance could still be sustained by public penalties alone, but only at the cost of quantity distortions that are no longer worthwhile: at the optimum, continuation values instead do the remaining work, and the contract optimally leaves a strictly positive rent even for the least efficient type. In this last regime, the discount factor enters the allocation directly, and the least efficient type receives a positive continuation rent. A purely static screening problem with an exogenous transfer cap cannot generate this bottom enforcement rent. This is the precise sense in which the dynamic content of the model is strongest in the very-weak-enforcement region.

Under full commitment and costless enforcement, the fixed-terms restriction is without loss in the persistent-type environment of [Baron and Besanko \(1984\)](#). With limited enforcement, it is a substantive restriction: it rules out repeated reports, history-dependent quantities and transfers, and the backloading of rewards across dates. We impose it to isolate how post-payment breach affects screening when these intertemporal adjustment margins are unavailable. In that case, enforcement operates through the size of the report-contingent transfer, and the dynamic compliance problem becomes an endogenous restriction on the top of the screening menu. The resulting characterization is therefore a benchmark for relationships governed by fixed terms rather than a characterization of the unrestricted dynamic optimum.

Fixed terms may be relevant when principals must offer common terms to agents who begin relationships at different dates, when arbitrage makes discriminatory prices difficult to sustain, or when administrative and legal considerations favor standardized schedules. In franchising, for example, royalty rates and franchise fees display substantial persistence over time ([Lafontaine and Shaw, 1999](#)). The fixed-terms benchmark also provides a transparent comparison point for richer mechanisms in which transfers, quantities, and continuation values evolve over the life of the relationship.

LITERATURE REVIEW. The paper is related to dynamic mechanism design and relational contracting. Under full commitment and costless enforcement, [Baron and Besanko \(1984\)](#) show that with persistent private information the optimal long-term contract is stationary and coincides with the repetition of the static optimum. Dynamic limited-enforcement models emphasize a different adjustment margin. In [Albuquerque and Hopenhayn \(2004\)](#), the borrower's equity entitlement rises, debt falls, and borrowing constraints relax as the lending relationship matures. More generally, [Ray \(2002\)](#) shows that efficient self-enforcing agreements may contain an initial phase in which payoffs are backloaded before the relationship reaches a mature phase. These models use intertemporal variation in promised payoffs or actions to manage enforcement. Our fixed-terms restriction removes that margin. Enforcement then operates through the cross-sectional screening menu and reduces to a restriction on its highest transfer.

A useful relational benchmark is [Levin \(2003\)](#), who studies self-enforcing contracts with hidden information and i.i.d. types. Under his maintained conditions, stationary contracts are optimal and self-enforcement limits the range of contingent payments. The common force in Levin's model and ours is that limited enforceability compresses feasible transfers and may induce pooling. Here, however, the seller's type is persistent, it is reported only once, and the relevant deviation is a post-payment breach by the seller. Under fixed terms, this produces a cap on the transfer assigned to the most efficient type and, under very weak enforcement, a positive enforcement rent even for the least efficient type.

The closest relational-contracting comparison with persistent private information is [Malcomson \(2016\)](#). Malcomson studies noncontractible effort and shows that, when continuation equilibria following revelation lie on the feasible Pareto frontier, full separation of a continuum of persistent types is impossible. Pooling there is therefore a feasibility implication of dynamic enforcement and optimal continuation. Our mechanism differs in two respects. Output and transfers are formally specified after a single initial report, and the seller's deviation consists of retaining a prepaid transfer and terminating the relationship. The standard Baron-Myerson allocation is recovered under strong enforcement; enforcement-induced bunching appears when the top-transfer restriction binds and the principal optimally compresses the screening menu. The positive rent of the least efficient type under very weak enforcement is also specific to the present post-payment breach problem.

The transaction order distinguishes our model from self-enforcing trade credit. In [Troya-Martinez \(2017\)](#), the supplier delivers first and the retailer later decides how much to repay; the threat of trade suspension disciplines repayment. [Fuchs et al. \(2022\)](#) study an upstream party that advances inventory to a credit-constrained vendor who may abscond with the inventory or its proceeds; the optimal arrangement uses a credit line and gradual growth. These are closely related reverse-timing environments: goods or inventory move first and the downstream party may default on payment. Here payment moves first and the privately informed seller may default on delivery. This reversal places the enforcement restriction directly on the transfer used for screening and generates the three regimes characterized below.

The weak-enforcement region is also related to static screening problems with transfer caps or payment bounds, as in [Thomas \(2002\)](#) and [Gautier and Mitra \(2006\)](#). The difference is that the cap is generated here by the seller's breach option. The paper is also related to [Martimort et al. \(2017b\)](#), which studies stationary contracts under two-sided limited enforcement. Isolating seller breach generates an additional regime in which the least efficient type receives a strictly positive enforcement rent. Finally, [Martimort et al. \(2017a\)](#) study a binary-type model with two-sided limited enforcement in which the unrestricted optimum is non-stationary. That analysis shows how backloading and growing trade can manage enforcement when contractual terms are allowed to vary. The present paper instead provides a complete characterization of the corresponding fixed-terms benchmark.

ORGANIZATION OF THE PAPER. Section 2 presents the model and derives the fixed-terms enforcement constraint. Section 3 states the principal's problem and the main theorem. Section 4 solves the model through a reduction to two auxiliary programs. Section 5 discusses interpretation, comparative statics, and the scope of the fixed-terms benchmark. Proofs are relegated to the Appendix.

2. MODEL

- PREFERENCES AND INFORMATION. We consider a long-term relationship between a buyer (the principal) and a seller (the agent). In each period, the buyer purchases a (non-durable) good q from the seller and pays a transfer t .¹ The buyer and the seller

¹The good q can be interpreted as quantity, or quality in the case of a single unit.

have per-period utility functions given respectively by

$$V(q, t) = S(q) - t, \quad U(q, t, \theta) = t - \theta q,$$

where θ is the seller's marginal cost: the agent's type. Assume that the agent's type is drawn once and for all at the beginning of the relationship. The agent privately learns his cost parameter θ which is drawn from the atomless distribution $F(\cdot)$ on the interval $\Theta = [\underline{\theta}, \bar{\theta}] \subset \mathbb{R}_{++}$ with density $f(\cdot)$. The distribution $F(\cdot)$ is common knowledge. The gross surplus function $S(\cdot)$ is increasing and strictly concave ($S'(\cdot) > 0 > S''(\cdot)$) and satisfies the Inada conditions $S'(0) = +\infty, S(0) = 0$. These assumptions ensure that the first-best surplus is always positive (i.e., $\max_{q \geq 0} S(q) - \bar{\theta}q \geq 0$).

The time horizon is infinite, discrete, and the parties have a common discount factor $\delta \in (0, 1)$.

Output is observable each period. At the beginning of the relationship, the buyer offers a long-term contract to the seller, who accepts or rejects after learning his cost type. A *fixed-terms mechanism* has three features: the seller reports only once, at the beginning of the relationship; the report-contingent transfer and output are the same in every period; and, on the equilibrium path, these terms do not depend on past deliveries or payments. The mechanism is stationary in the sense that the same report-contingent allocation is repeated after the initial report. Under full enforcement with persistent types this restriction is without loss (Baron and Besanko, 1984); under limited enforcement it defines the benchmark studied here. Within this class, any equilibrium allocation can be represented in truthful direct form, so we write the mechanism as $\{t(\hat{\theta}), q(\hat{\theta})\}_{\hat{\theta} \in \Theta}$.

• **TIMING.** The contracting game unfolds as follows:

- (i) The seller learns his private cost parameter θ .
- (ii) The buyer offers a fixed-terms mechanism $\mathcal{C} = \{t(\hat{\theta}), q(\hat{\theta})\}_{\hat{\theta} \in \Theta}$.
- (iii) The seller accepts or rejects. If he accepts, he reports $\hat{\theta}$.
- (iv) In each period τ , the buyer prepays $t(\hat{\theta})$ and the seller then either delivers $q(\hat{\theta})$ or breaches, pays the enforceable penalty $\Pi \geq 0$, and terminates the relationship.

The penalty Π is imposed when non-delivery is observed. It is an institutional primitive rather than a choice variable of the buyer: it measures the maximum expected monetary consequence of breach that can actually be imposed and collected. We take it to be common across reports and types to represent a statutory, bonding, or collateral limit that does not vary with the announced menu; in particular, the seller's true cost is not

verifiable. Allowing the collectible sanction to depend on the contractual stake would constitute a separate extension. This fixed-penalty formulation abstracts from remedies that vary with the contractual stake and isolates the interaction between a common public-enforcement technology and the screening menu.

• INCENTIVE COMPATIBILITY AND PARTICIPATION. Because the buyer commits to a fixed-terms mechanism, we focus on truthful fixed-terms mechanisms $\{t(\hat{\theta}), q(\hat{\theta})\}_{\hat{\theta} \in \Theta}$. We denote the seller's per-period rent by

$$U(\theta) = t(\theta) - \theta q(\theta).$$

His ex post participation constraint is

$$U(\theta) \geq 0 \quad \forall \theta \in \Theta. \quad (2.1)$$

LEMMA 2.1 (Incentive compatibility and monotonicity). *The mechanism $\{t(\hat{\theta}), q(\hat{\theta})\}_{\hat{\theta} \in \Theta}$ is incentive compatible if and only if:*

- (i) *q is non-increasing on Θ ;*
- (ii) *$U(\theta) = t(\theta) - \theta q(\theta)$ satisfies the integral representation*

$$U(\theta) = U(\bar{\theta}) + \int_{\theta}^{\bar{\theta}} q(x) dx \quad \forall \theta \in \Theta. \quad (2.2)$$

In particular, U is then convex, non-increasing and absolutely continuous with $\dot{U}(\theta) = -q(\theta)$ a.e., q is a.e. differentiable with $\dot{q}(\theta) \leq 0$ a.e., and at any point where both q and t are differentiable,

$$\dot{t}(\theta) = \theta \dot{q}(\theta). \quad (2.3)$$

The proof is in the Appendix. The integral form (2.2) is the global incentive-compatibility condition; unlike the almost-everywhere differential condition (2.3), it also controls discontinuities.²

DEFINITION 2.2 (Admissible contracts). A pair (t, q) is *admissible* if: (i) q is piecewise continuous, of bounded variation and non-increasing on Θ ; (ii) the rent $U(\theta) = t(\theta) - \theta q(\theta)$ satisfies the envelope condition (2.2); and (iii) (2.1) is satisfied.

²Since U is continuous, at a discontinuity point θ_0 of q the transfer jumps by $t(\theta_0^+) - t(\theta_0^-) = \theta_0(q(\theta_0^+) - q(\theta_0^-)) \leq 0$.

By Lemma 2.1, conditions (i) and (ii) are equivalent to global incentive compatibility. Admissibility therefore coincides exactly with the set of incentive compatible and individually rational fixed-terms mechanisms.

REMARK 2.3 (Continuity of the optimal schedule). Although admissibility allows for discontinuities, we will see below that optimal quantity schedules are continuous at every interior type, and their endpoint values coincide with the corresponding one-sided limits whenever the enforcement constraint has bite; in the remaining (strong-enforcement) case the endpoint values are payoff-irrelevant, so that the continuous representative of any optimal schedule is itself optimal. Proposition 6.1 in the Appendix states and proves this precisely. The proof reduces both programs to a problem that is strictly concave in q and in which every constraint depends on q only through the linear functional $\underline{\theta} q(\underline{\theta}) + \int_{\underline{\theta}}^{\bar{\theta}} q(x) dx$; a mean-preserving local flattening of any interior jump then preserves all constraints exactly and strictly raises the objective. Thus the structural statements in the main text, such as the existence of a single bunching interval, can be read literally for continuous schedules.

- **ENFORCEMENT CONSTRAINT.** The focus on fixed terms mechanisms makes the timing of a possible breach particularly simple. For any report $\hat{\theta}$, the seller's payoff from complying for τ periods and then breaching is affine in $x = \delta^\tau$. Its maximum is therefore attained at an endpoint: either the seller follows the reported contract forever, or he breaches immediately. The first possibility is already governed by incentive compatibility. Conditional on breach, the optimal breach date is consequently the first period.

The report chosen at breach is equally simple. Because $\Theta \subset \mathbb{R}_{++}$, Lemma 2.1 implies that the transfer is weakly decreasing in the reported cost, possible discontinuities of q being included: for any $\theta_1 < \theta_2$, the envelope condition (2.2) gives

$$t(\theta_1) - t(\theta_2) = \int_{\theta_1}^{\theta_2} q(x) dx + \theta_1 q(\theta_1) - \theta_2 q(\theta_2) \geq \theta_1 (q(\theta_1) - q(\theta_2)) \geq 0,$$

where both inequalities use that q is non-increasing. Hence the largest current transfer is $t(\underline{\theta})$. A seller who breaches therefore reports $\underline{\theta}$, collects the largest transfer, and exits immediately.

Formally, allowing for an arbitrary report and an arbitrary breach date $\tau \in \mathbb{N}_0 \cup \{+\infty\}$, where $\mathbb{N}_0 := \{0, 1, 2, \dots\}$, and where $\tau = +\infty$ means that the seller never breaches

(with the convention $\delta^{+\infty} := 0$), gives

$$\frac{1}{1-\delta} U(\theta) \geq \sup_{\hat{\theta} \in \Theta, \tau \in \mathbb{N}_0 \cup \{+\infty\}} \left\{ \sum_{s=0}^{\tau-1} \delta^s (t(\hat{\theta}) - \theta_q(\hat{\theta})) + \delta^\tau (t(\hat{\theta}) - \Pi) \right\}. \quad (2.4)$$

LEMMA 2.4 (Reduction of enforcement). *An admissible (t, q) is enforceable iff*

$$t(\theta) - \theta_q(\theta) \geq (1 - \delta)(t(\underline{\theta}) - \Pi) \quad \forall \theta \in \Theta. \quad (2.5)$$

Lemma 2.4 formalizes the preceding argument. The endpoint corresponding to following a reported contract forever is already controlled by incentive compatibility. The only additional deviation that must be deterred is immediate breach after claiming the transfer intended for the most efficient type. Consequently, the entire one-sided enforcement problem is summarized by a single restriction involving the top transfer.

When Π is moderate, the right-hand side of (2.5) is positive, so enforceability constrains the set of implementable allocations. Lowering $t(\underline{\theta})$ weakens the seller's gain from a take-the-money-and-run deviation, but incentive compatibility then propagates that compression to other types. Conversely, when Π is large enough, the right-hand side of (2.5) is non-positive and the enforcement constraint is implied by the participation constraint. In that case, the fixed-terms problem collapses to the usual static second-best benchmark.

3. OPTIMAL CONTRACT

The principal's problem is

$$(P): \quad \max_{\{(t, q) \text{ admissible}\}} \int_{\underline{\theta}}^{\bar{\theta}} (S(q(\theta)) - t(\theta)) f(\theta) d\theta \quad \text{subject to (2.5).}$$

Participation is already incorporated into admissibility by Definition 2.2(iii).

We extend the classical second-best screening problem by adding the fixed-terms enforcement constraint (2.5). In the absence of that constraint, or whenever it is slack, the optimal fixed-terms allocation is the repeated static contract of [Baron and Myerson \(1982\)](#). Let $(q^{os}(\theta), t^{os}(\theta))$ denote that benchmark, where

$$S'(q^{os}(\theta)) = \theta + \frac{F(\theta)}{f(\theta)} \quad \forall \theta \in \Theta,$$

and

$$t^{os}(\theta) = \theta q^{os}(\theta) + \int_{\theta}^{\bar{\theta}} q^{os}(x) dx.$$

The term $\theta + F(\theta)/f(\theta)$ is the usual virtual cost.

DEFINITION 3.1. For any $r \geq 0$, define the generalized static output $q^{(r)}(\theta)$ by

$$S'(q^{(r)}(\theta)) = \theta + \frac{F(\theta) + r}{f(\theta)} \quad \forall \theta \in \Theta.$$

Thus $q^{os}(\theta) = q^{(0)}(\theta)$. When $r > 0$, the output rule is more distorted than the standard second-best rule. When the top-transfer restriction binds, r is its shadow value. Since a local change in $q(\theta)$ enters expected surplus with density weight $f(\theta)$, the corresponding pointwise wedge is $r/f(\theta)$. We therefore interpret this term as a *virtual enforcement cost*. A fuller derivation is provided after Theorem 3.5.

ASSUMPTION 3.2. (i) The density $f(\theta)$ is differentiable.

(ii) The function $\frac{F(\theta) + r}{f(\theta)}$ is increasing in θ for every $r \in [0, r^*]$, where $r^* = \frac{1-\delta}{\delta}$.

Assumption 3.2 is stronger than the usual regularity condition from the unconstrained screening problem. A simple sufficient primitive condition is that the density be weakly decreasing, $f'(\theta) \leq 0$, which implies the required monotonicity for every $r \geq 0$.³ We therefore interpret the assumption as a generalized monotonicity condition on the constrained virtual cost, rather than as an ordinary hazard-rate restriction.⁴

Next, we define the output which plays an important role in characterizing the optimal contract.

DEFINITION 3.3. For any $r \geq 0$, let q_r^{pool} denote the unique solution of

$$S'(q_r^{\text{pool}}) = \bar{\theta}(1 + r). \quad (3.1)$$

³Examples include exponential and Pareto distributions, as well as parameterizations of beta, gamma, and Weibull distributions whose densities are weakly decreasing on the relevant support.

⁴This condition does not require the density to be globally decreasing. In particular, Gautier and Mitra (2006) provide examples of distributions with locally increasing density (such as truncated normal or logistic distributions) that satisfy their Assumption 1. These examples, with required restrictions on parameters, are also applicable in our setting.

Define the output $q_r(\theta)$ by

$$q_r(\theta) = \begin{cases} q^{(r)}(\theta_r) & \text{if (3.3) has a solution and } \theta < \theta_r, \\ q^{(r)}(\theta) & \text{if (3.3) has a solution and } \theta_r \leq \theta \leq \bar{\theta}, \\ q_r^{\text{pool}} & \text{if (3.3) has no solution.} \end{cases} \quad (3.2)$$

where $\theta_r \in [\underline{\theta}, \bar{\theta}]$ is the (minimal) solution of

$$r = \frac{F^2(\theta_r)}{\theta_r f(\theta_r) - F(\theta_r)}, \quad (3.3)$$

if it exists; otherwise we set $\theta_r = \bar{\theta}$ only as a notational convention.

When $\underline{\theta} < \theta_r < \bar{\theta}$, the output exhibits bunching for types smaller than θ_r . For types greater than θ_r , the output aligns with the generalized static output (see Figure 1). When $r = 0$, the threshold θ_r is $\underline{\theta}$ and the corresponding output $q_r(\theta)$ is the optimal static contract. If no solution to (3.3) exists, the optimal schedule is completely bunched at q_r^{pool} , rather than at $q^{(r)}(\bar{\theta})$. The two formulas coincide at $r = \hat{r}$, where $\hat{r}$ is defined below.

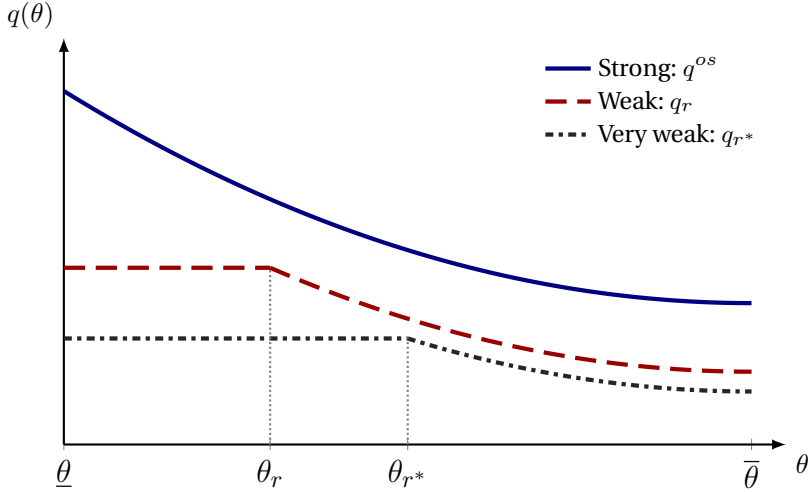


FIGURE 1. Illustrative quantity schedules in the three enforcement regimes

Figure 1 depicts the partial-bunching configuration. In case (ii) of Lemma 3.4, the weak- or very-weak-enforcement schedule may instead be completely pooled. The transfers required to implement the output $q_r(\theta)$ are determined by

$$t_r(\theta) = \theta q_r(\theta) + \int_{\theta}^{\bar{\theta}} q_r(x) dx + \left(t_r(\bar{\theta}) - \bar{\theta} q_r(\bar{\theta}) \right). \quad (3.4)$$

Let $r^* = \frac{1-\delta}{\delta}$.⁵ The following lemma establishes the existence of the output $q_r(\theta)$ for all $r \in [0, r^*]$.

LEMMA 3.4. *Let $r(\theta) := \frac{F^2(\theta)}{\theta f(\theta) - F(\theta)}$ wherever $\theta f(\theta) - F(\theta) > 0$. Exactly one of the following two cases holds:*

- (i) *for every $r \in [0, r^*]$, equation (3.3) admits a unique solution $\theta_r \in [\underline{\theta}, \bar{\theta}]$, which is increasing in r ;*
- (ii) *$\theta f(\theta) - F(\theta) > 0$ on all of Θ and $\hat{r} := r(\bar{\theta}) = \frac{1}{\bar{\theta} f(\bar{\theta}) - 1} < r^*$; then (3.3) admits a unique solution θ_r , increasing in r , if and only if $r \in [0, \hat{r}]$, whereas for every $r \in (\hat{r}, r^*]$ no solution exists and the schedule is completely pooled: $q_r(\theta) = q_r^{\text{pool}}$ for every $\theta \in \Theta$.*

The left panel of Figure 2 depicts case (i). Equation (3.3) defines the threshold θ_r as a function of r whenever a separating region exists. As r increases, the bunching cutoff moves upward; in case (ii) of Lemma 3.4, once r exceeds $\hat{r}$ the separating region disappears and the schedule is completely pooled at q_r^{pool} . In particular, the value $r^* = (1 - \delta)/\delta$ determines the largest generalized distortion relevant for the fixed-terms program.

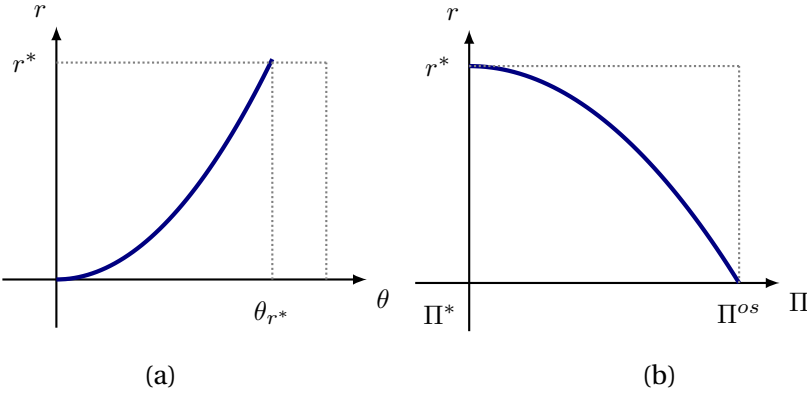


FIGURE 2. r as a function of θ_r and Π

In the right panel of Figure 2, each $r \in [0, r^*]$ corresponds to a penalty $\Pi \in [\Pi^*, \Pi^{os}]$ through

$$\Pi = \underline{\theta} q_r(\underline{\theta}) + \int_{\underline{\theta}}^{\bar{\theta}} q_r(\theta) d\theta. \quad (3.5)$$

⁵The value r^* is the largest shadow value relevant for Program (P^A) .

This expression is simply $t_r(\underline{\theta})$ when the least efficient type earns zero rent. On the complete-pooling branch it reduces to $\Pi(r) = \bar{\theta} q_r^{\text{pool}}$. By Lemma 3.4, $\Pi(r)$ is continuous and strictly decreasing, so its inverse $r = r(\Pi)$ is well defined on $[\Pi^*, \Pi^{os}]$.

Using (3.5), define Π^{os} and Π^* as the penalties associated with $r = 0$ and $r = r^*$ respectively. The threshold Π^{os} is the smallest penalty for which the static second-best contract is enforceable. The threshold Π^* , by contrast, is an optimality threshold rather than a feasibility threshold. For $\Pi < \Pi^*$, enforceable mechanisms leaving zero rent to the least efficient type remain feasible: capping the top transfer at Π , as in Program (P^B) below, always deters breach. Sustaining such a cap, however, would push the shadow value of the enforcement constraint above r^* and hence distort quantities beyond q_{r^*} . At Π^* the principal becomes indifferent between tightening the cap further and relaxing it by conceding continuation rents. Below Π^* , the principal optimally fixes the distortion at r^* and leaves a common enforcement rent $\frac{1-\delta}{\delta}(\Pi^* - \Pi)$ to all types, including the least efficient one. Thus Π^* is the smallest penalty for which the *optimal* contract sustains zero rent at the bottom.

The optimal contract for the agent's enforcement problem can now be described.

THEOREM 3.5. (i) **Strong enforcement.** For $\Pi \geq \Pi^{os}$, the optimal contract involves the infinite repetition of the optimal static contract.

(ii) **Weak enforcement.** For $\Pi^* \leq \Pi \leq \Pi^{os}$, the optimal contract is $(t_r(\theta), q_r(\theta))$, where the parameter r is defined by

$$t_r(\underline{\theta}) = \Pi. \quad (3.6)$$

Here, the least efficient type has zero rent, i.e., $t_r(\bar{\theta}) - \bar{\theta} q_r(\bar{\theta}) = 0$.

(iii) **Very weak enforcement.** For $0 \leq \Pi \leq \Pi^*$, the optimal contract is $(t(\theta), q_{r^*}(\theta))$. The transfers $t(\theta)$ are given by

$$t(\theta) = t_{r^*}(\theta) + \frac{1-\delta}{\delta} (\Pi^* - \Pi).$$

If $\Pi < \Pi^*$, then the least efficient type earns a positive rent,

$$t(\bar{\theta}) - \bar{\theta} q_{r^*}(\bar{\theta}) = \frac{1-\delta}{\delta} (\Pi^* - \Pi) > 0.$$

Under *strong enforcement* ($\Pi \geq \Pi^{os}$), the enforcement constraint does not distort the allocation, and the repeated static second-best contract is feasible and optimal. For

$\Pi > \Pi^{os}$ the constraint is strictly slack; at the boundary $\Pi = \Pi^{os}$ it holds with equality for the least efficient type—since $t^{os}(\underline{\theta}) = \Pi^{os}$ and $U^{os}(\bar{\theta}) = 0$, constraint (2.5) binds at $\bar{\theta}$, but costlessly so, with zero shadow value ($r = 0$). Under *weak enforcement* ($\Pi^* \leq \Pi < \Pi^{os}$), enforceability is equivalent to a cap on the transfer intended for the most efficient type. This cap propagates through incentive compatibility and leads to bunching among efficient types and further downward distortions for the remaining types. If case (ii) of Lemma 3.4 applies and $r(\Pi) \geq \hat{r}$, the bunching interval covers the entire type space. Under *very weak enforcement* ($0 \leq \Pi < \Pi^*$), compliance could still be sustained by public enforcement alone since zero-bottom-rent mechanisms, when capped, remain feasible, but only by distorting quantities beyond q_{r^*} , which is no longer worthwhile. At the optimum, continuation values instead do the remaining work, and the least efficient type earns a strictly positive rent.

STATIC BENCHMARK. The weak-enforcement region is closely related to a static screening problem with a binding upper bound on the top transfer. In that sense, bunching in this region is “static-like.” The additional dynamic content appears most sharply in the very-weak-enforcement region, where continuation values generate an *enforcement rent* for the least efficient type.

WHY BUNCHING ARISES UNDER WEAK ENFORCEMENT. For $\Pi^* \leq \Pi < \Pi^{os}$, the top transfer is capped at $t(\underline{\theta}) = \Pi$. Lowering that transfer reduces the gain from a take-the-money-and-run deviation, but incentive compatibility links nearby types to the same margin. The principal therefore prefers to pool the most efficient types rather than raise the top transfer above the enforceable level.

When the separating region is nonempty, types above the bunching cutoff, $\theta \in [\theta_r, \bar{\theta}]$, follow the generalized static rule. On the complete-pooling branch this region is empty, and the terminal condition (3.1) determines the common quantity instead. In the weak-enforcement regime, the least efficient type earns zero rent, so the top transfer satisfies

$$t(\underline{\theta}) = \underline{\theta}q(\underline{\theta}) + \int_{\underline{\theta}}^{\bar{\theta}} q(x) dx.$$

Let r denote the shadow value of the binding cap $t(\underline{\theta}) = \Pi$. On the separating region, a marginal change in $q(\theta)$ changes the top transfer with unit weight, whereas its contribution to expected surplus is weighted by $f(\theta)$. The pointwise optimality condition is

therefore

$$(S'(q^{(r)}(\theta)) - \theta)f(\theta) = F(\theta) + r.$$

Dividing by $f(\theta)$ gives

$$S'(q^{(r)}(\theta)) = \theta + \frac{F(\theta)}{f(\theta)} + \frac{r}{f(\theta)}.$$

The additional term $r/f(\theta)$ is thus a *virtual enforcement cost*: the common shadow cost of the aggregate transfer restriction is converted into a type-specific marginal distortion through the density weight $f(\theta)$.

RENT DECOMPOSITION. Under strong and weak enforcement, the least efficient type earns zero rent and the seller's rent is purely informational:

$$U(\theta) = \int_{\theta}^{\bar{\theta}} q_r(x) dx.$$

By contrast, when $\Pi < \Pi^*$, the contract must also leave a positive continuation payoff to sustain compliance. The seller's rent then decomposes as

$$U(\theta) = \underbrace{\frac{1-\delta}{\delta}(\Pi^* - \Pi)}_{\text{enforcement rent}} + \underbrace{\int_{\theta}^{\bar{\theta}} q_{r^*}(x) dx}_{\text{information rent}}.$$

The first term is the compliance premium required by weak public enforcement; the second term is the usual information rent.

CONNECTION TO RELATIONAL CONTRACTING. Our analysis connects to [Levin \(2003\)](#), where limited enforceability compresses feasible transfers and can induce pooling. The key difference is that our environment has persistent private information and a one-sided post-payment deviation. Under fixed terms, that specific deviation makes the enforcement problem collapse to a cap on the top transfer; when public enforcement is very weak, continuation values generate an additional bottom rent that is absent from a purely static transfer-cap formulation.

The comparison with [Malcomson \(2016\)](#) further clarifies the source of pooling. In [Malcomson \(2016\)](#)'s relational environment with persistent private information, full separation is generally infeasible under optimal continuation. In our fixed-terms mechanism, the standard screening allocation remains feasible under strong enforcement; bunching emerges only after limited enforcement makes the top-transfer restriction binding.

EXAMPLE 3.6 (Uniform types and quadratic surplus). Suppose that $\Theta = [1, 2]$, the type distribution is uniform and the buyer's gross surplus is $S(q) = 5q - \frac{q^2}{2}$. Assume also that $\delta = \frac{4}{5}$, so that $r^* = \frac{1-\delta}{\delta} = \frac{1}{4}$.⁶ Then the optimal static output is

$$S'(q^{os}(\theta)) = 5 - q^{os}(\theta) = \theta + \frac{F(\theta)}{f(\theta)} = 2\theta - 1,$$

which yields

$$q^{os}(\theta) = 6 - 2\theta, \quad q^{os}(1) = 4, \quad q^{os}(2) = 2.$$

In particular, the transfer to the most efficient type under the repeated static contract is

$$t^{os}(1) = q^{os}(1) + \int_1^2 q^{os}(x) dx = 7.$$

Hence $\Pi^{os} = 7$. For any $r \in [0, r^*]$,

$$S'(q^{(r)}(\theta)) = 5 - q^{(r)}(\theta) = \theta + \frac{F(\theta) + r}{f(\theta)} = 2\theta - 1 + r,$$

so $q^{(r)}(\theta) = 6 - r - 2\theta$.⁷ The bunching cutoff solves (3.3):

$$r = \frac{F^2(\theta_r)}{\theta_r f(\theta_r) - F(\theta_r)} = (\theta_r - 1)^2,$$

that is, $\theta_r = 1 + \sqrt{r}$. Therefore,

$$q_r(\theta) = \begin{cases} 4 - r - 2\sqrt{r}, & 1 \leq \theta < 1 + \sqrt{r}, \\ 6 - r - 2\theta, & 1 + \sqrt{r} \leq \theta \leq 2. \end{cases}$$

Using (3.5), the penalty associated with r is

$$\Pi(r) = t_r(1) = q_r(1) + \int_1^2 q_r(x) dx = 7 - 2\sqrt{r} - 3r.$$

Hence $\Pi^* = \Pi(r^*) = \frac{21}{4}$. This gives three regimes:

(i) *Strong enforcement*: if $\Pi \geq 7$, the optimal contract is the repeated static contract

$$(t^{os}, q^{os}).$$

⁶The quadratic specification is used only on the relevant output range. As shown below, all outputs lie in $[7/4, 4] \subset (0, 5)$, where S is increasing and strictly concave.

⁷On this range the formula is legitimate since $q^{(r)}(\theta) \geq 2 - r^* = \frac{7}{4} > 0$ and $S'(q^{(r)}(\theta)) = 2\theta - 1 + r \in [1, \frac{13}{4}]$ for all $\theta \in [1, 2]$, so the first-order condition indeed characterizes an interior optimum. For large r the expression $6 - r - 2\theta$ would eventually turn negative, but such values never arise: Theorem 3.5 only involves $r \in [0, r^*]$, and in the weak-enforcement regime below, $r(\Pi) \leq r^*$ for all $\Pi \geq \Pi^*$.

(ii) *Weak enforcement*: if $\frac{21}{4} \leq \Pi < 7$, then $r(\Pi)$ is determined by

$$7 - 2\sqrt{r} - 3r = \Pi, \quad r(\Pi) = \left(\frac{\sqrt{22 - 3\Pi} - 1}{3} \right)^2.$$

For example, if $\Pi = 6$, then $r = \frac{1}{9}$ and $\theta_r = \frac{4}{3}$, so the optimal output is

$$q_r(\theta) = \begin{cases} \frac{29}{9}, & 1 \leq \theta < \frac{4}{3}, \\ \frac{53}{9} - 2\theta, & \frac{4}{3} \leq \theta \leq 2. \end{cases}$$

Thus the most efficient types are bunched on $[1, \frac{4}{3})$.

(iii) *Very weak enforcement*: if $\Pi < \frac{21}{4}$, the optimal quantity is q_{r^*} , i.e.,

$$q_{r^*}(\theta) = \begin{cases} \frac{11}{4}, & 1 \leq \theta < \frac{3}{2}, \\ \frac{23}{4} - 2\theta, & \frac{3}{2} \leq \theta \leq 2, \end{cases}$$

and the least efficient type earns a strictly positive rent. For instance, if $\Pi = 5$, then

$$U(2) = \frac{1 - \delta}{\delta}(\Pi^* - \Pi) = \frac{1}{16} > 0.$$

This example highlights the economic content of Theorem 3.5: stronger public enforcement relaxes the cap on the top transfer, reduces bunching, and eventually restores the repeated Baron-Myerson allocation.

4. PROOF OF THEOREM: REDUCTION TO TWO PROGRAMS

Rewriting constraints (2.1) and (2.5) respectively yields

$$t(\theta) - \theta q(\theta) \geq 0 \quad \forall \theta \in \Theta, \quad (4.1)$$

$$t(\theta) - \theta q(\theta) \geq (1 - \delta)(t(\underline{\theta}) - \Pi) \quad \forall \theta \in \Theta. \quad (4.2)$$

Constraint (4.2) includes, on the right-hand side, the transfer for the most efficient type $t(\underline{\theta})$, representing the seller's net gain in the event of breach. This right-hand side becomes non-positive when the net gain falls below the penalty, making the enforcement constraint ineffective. In this case, the problem becomes a second-best problem with a cap Π on transfers. If the net gain from any deviation is less than the penalty, the seller's overall benefit from breaching the contract is non-positive. Therefore, the seller has no incentive to breach as long as ex post rents are non-negative, which is assured if (4.1) holds.

When the seller's net gain from breach exceeds the penalty, the participation constraint follows from the enforcement constraint. The seller can breach, cover the penalty with the gain, and leave the relationship. To prevent breach, the contract must ensure that the seller obtains enough rent from continuation. This leads to two regimes of constraints:

$$(A) : \begin{cases} t(\underline{\theta}) - \Pi \geq 0, \\ t(\theta) - \theta q(\theta) \geq (1 - \delta)(t(\underline{\theta}) - \Pi) \quad \forall \theta \in \Theta, \end{cases}$$

and

$$(B) : \begin{cases} t(\underline{\theta}) - \Pi \leq 0, \\ t(\theta) - \theta q(\theta) \geq 0 \quad \forall \theta \in \Theta. \end{cases}$$

We split the problem (P) into two problems (P^A) and (P^B) , obtained from (P) by replacing (4.1) and (4.2) by systems (A) and (B) respectively.

An optimal solution of (P) is necessarily optimal for either (P^A) or (P^B) . Conversely, among the solutions to (P^A) and (P^B) , the one that provides the highest payoff to the principal is the optimal solution of (P) .

Program (P^A) : Consider first (P^A) . To eliminate $t(\underline{\theta})$ from the right-hand side of (4.2), introduce the adjusted transfer $y(\theta)$:⁸

$$y(\theta) = t(\theta) - (1 - \delta)(t(\underline{\theta}) - \Pi).$$

Then system (A) becomes

$$\begin{cases} y(\underline{\theta}) - \Pi \geq 0, \\ y(\theta) - \theta q(\theta) \geq 0 \quad \forall \theta \in \Theta. \end{cases}$$

Note that the second constraint must bind at $\bar{\theta}$:

$$y(\bar{\theta}) - \bar{\theta}q(\bar{\theta}) = 0. \tag{4.3}$$

Indeed, if $y(\bar{\theta}) - \bar{\theta}q(\bar{\theta}) > 0$, consider the contract $(q(\theta) + \varepsilon, y(\theta))$, for ε such that $y(\bar{\theta}) - \bar{\theta}(q(\bar{\theta}) + \varepsilon) = 0$. This change does not affect feasibility and strictly increases the principal's payoff since $S' > 0$.

⁸Notice that $y(\underline{\theta}) = \delta t(\underline{\theta}) + (1 - \delta)\Pi$ and $t(\underline{\theta}) = \frac{y(\underline{\theta}) - (1 - \delta)\Pi}{\delta}$. Thus $t(\underline{\theta}) = \Pi$ if and only if $y(\underline{\theta}) = \Pi$.

Problem (P^A) can thus be written as

$$\max_{\{y(\cdot), q(\cdot)\} \text{ admissible}} \int_{\underline{\theta}}^{\bar{\theta}} (S(q(\theta)) - y(\theta)) f(\theta) d\theta - \frac{1-\delta}{\delta} (y(\underline{\theta}) - \Pi)$$

subject to (4.3) and

$$y(\underline{\theta}) - \Pi \geq 0. \quad (4.4)$$

Problem (P^A) depends on Π , which enters both the objective and constraint (4.4). We treat (P^A) as an optimal control problem with boundary constraints (4.3)-(4.4) and a scrap value $-\frac{1-\delta}{\delta} (y(\underline{\theta}) - \Pi)$.

LEMMA 4.1. (i) When $\Pi \in [\Pi^*, \Pi^{os}]$, the optimal output for (P^A) is $q_r(\theta)$ with r determined by

$$\underline{\theta} q_r(\underline{\theta}) + \int_{\underline{\theta}}^{\bar{\theta}} q_r(\theta) d\theta = \Pi. \quad (4.5)$$

Furthermore, constraint (4.4) is binding: $y(\underline{\theta}) = \Pi$.

(ii) For $\Pi < \Pi^*$, the optimal output is $q_{r^*}(\theta)$ and (4.4) is slack: $y(\underline{\theta}) > \Pi$.

In Case (i), $t(\underline{\theta}) = \Pi$. The solution lies on the boundary between Programs (P^A) and (P^B) : the penalty pins down the top transfer, but no positive continuation rent is required. Hence the optimal contract does not depend on δ . Case (ii), corresponding to very weak enforcement, arises when $\Pi < \Pi^*$. In this case, the output is fixed at $q_{r^*}(\theta)$. Efficiency cannot be further compromised in favor of enforcement. With such a low penalty, enforcement relies on the discount factor, as reflected in $r^* = (1 - \delta)/\delta$.

Program (P^B) : Now consider

$$(P^B): \max_{\{t(\cdot), q(\cdot)\} \text{ admissible}} \int_{\underline{\theta}}^{\bar{\theta}} (S(q(\theta)) - t(\theta)) f(\theta) d\theta$$

subject to (4.1) and

$$\Pi - t(\underline{\theta}) \geq 0. \quad (4.6)$$

Problem (P^B) is a second-best problem with a cap on transfers. It does not depend on δ . There are two differences between (P^B) and (P^A) : there is no scrap value in the objective of (P^B) , and constraint (4.6) is the reverse of (4.4). The optimal static contract describes the maximum payoff in the presence of the enforcement constraint. Therefore, when $\Pi \geq \Pi^{os}$ the optimal enforcement contract is simply the optimal static contract.

When $\Pi \leq \Pi^{os}$, constraint $\Pi - t(\underline{\theta}) \geq 0$ is binding.⁹ Thus, for $\Pi \in [\Pi^*, \Pi^{os}]$ the optimal outputs and transfers are identical in (P^A) and (P^B) .

LEMMA 4.2. (i) If $\Pi \in [\Pi^*, \Pi^{os}]$, the optimal output is $q_r(\theta)$, where $r \in [0, r^*]$ is defined by (4.5); in addition, (4.6) is binding: $t(\underline{\theta}) = \Pi$.

(ii) If $\Pi > \Pi^{os}$, the optimal output is the static second-best contract, and (4.6) is strict: $t(\underline{\theta}) < \Pi$.

(iii) If $\Pi < \Pi^*$, constraint (4.6) remains binding at any optimum, $t(\underline{\theta}) = \Pi$.¹⁰

COMPARISON BETWEEN (P^A) AND (P^B) : Intuitively, for small Π , the value derived from (P^A) exceeds that of (P^B) , and the reverse holds as Π becomes large. Specifically, for $\Pi \geq \Pi^{os}$, (P^B) has the static second-best contract as its solution. For (P^A) , the constraint on $t(\underline{\theta})$ distorts transfers away from the optimal static contract. The seller's gain in the event of breach, $t(\underline{\theta})$, must at least match the large penalty, leading to an upward distortion of $t^{os}(\underline{\theta})$. Consequently, for $\Pi > \Pi^{os}$, the value of (P^B) exceeds that of (P^A) and the optimal contract is the static second-best.

For $\Pi \in [\Pi^*, \Pi^{os}]$, both (P^A) and (P^B) share the same necessary and sufficient conditions and, therefore, the same solutions. Given that $y(\underline{\theta}) - \Pi$ is binding, the scrap value in the objective of (P^A) is zero. Hence, both programs have the same value.

For any $\Pi < \Pi^*$, the value of (P^B) cannot exceed that of (P^A) . Notably, the argument requires no characterization of the optimum of (P^B) . In the reduced form of Step 1 in the proof of Proposition 6.1 in the Appendix below, (P^B) maximizes $W(q) - u$ subject to $K(q) + u \leq \Pi$ and $u \geq 0$, so its value equals $\sup\{W(q) : K(q) \leq \Pi\}$, with $u = 0$ optimal. Take any monotonically decreasing q with $K(q) \leq \Pi$. Since $K(q^{os}) = t^{os}(\underline{\theta}) = \Pi^{os} > \Pi$ and K is linear, there exists $\lambda \in [0, 1)$ such that the (monotone, admissible) mixture $q_\lambda := \lambda q^{os} + (1 - \lambda)q$ satisfies $K(q_\lambda) = \Pi$. By concavity of W and optimality of q^{os} for W ,

$$W(q_\lambda) \geq \lambda W(q^{os}) + (1 - \lambda)W(q) \geq W(q).$$

⁹Were it slack, the optimum of (P^B) would locally—hence, by strict concavity of the reduced objective (Step 1 in the proof of Proposition 6.1), globally—coincide with the unconstrained static second best, whose top transfer $t^{os}(\underline{\theta}) = \Pi^{os}$ exceeds Π , a contradiction.

¹⁰We do not characterize the optimal output of (P^B) in this range. Along the family (3.5), the cap would require $r > r^*$, outside the range covered by Assumption 3.2. No such characterization is needed for the analysis of (P) .

The pair $(q_\lambda, u = 0)$ is feasible for (P^A) . It satisfies $t(\underline{\theta}) = K(q_\lambda) = \Pi$ and $u = 0 = r^*(K(q_\lambda) - \Pi)$, and it attains the value $W(q_\lambda) \geq W(q)$. Taking the supremum over feasible q shows that the value of (P^A) weakly exceeds that of (P^B) . Therefore the solution to (P^A) is also optimal for (P) .

PROPOSITION 4.3. (i) *For $\Pi < \Pi^*$, the solution to (P^A) specified in Lemma 4.1 is optimal for (P) .*

(ii) *For $\Pi \in [\Pi^*, \Pi^{os}]$, both programs have the same value and yield the same optimal contract.*

(iii) *For $\Pi > \Pi^{os}$, the solution to (P^B) specified in Lemma 4.2 is optimal for (P) .*

5. DISCUSSION

We have characterized the optimal fixed-terms mechanism when the buyer can commit to the announced menu, while the seller retains the option to breach after receiving each current payment. The optimal allocation is pinned down by the interaction between public enforcement, summarized by the breach penalty Π , and private enforcement, summarized by the discounted continuation value, whose weight is governed by δ .

The parameter Π measures the maximum expected monetary consequence that the institutional environment can impose and collect following non-delivery. Court effectiveness, limits on recoverable damages, performance bonds, forfeitable deposits, and collectible collateral may all contribute to this object. Holding Π fixed across types separates the strength of public enforcement from the seller's private cost and from the design of the screening menu. A sanction that varies with output, transfer, or reported type would constitute a separate extension.

The discount factor δ governs the weight placed on the continuation value of repeat trade. In reduced form, a larger continuation value may also reflect reputational capital, switching costs, or other relationship-specific rents. The comparative statics therefore describe how fixed contractual terms respond when either publicly recoverable sanctions or the value of future trade changes.

WHAT THE FIXED-TERMS BENCHMARK ISOLATES. The fixed-terms restriction suppresses the buyer's ability to backload rewards, expand quantities gradually, or condition continuation terms on past performance. In unrestricted dynamic limited-enforcement models, these instruments can relax early no-deviation constraints: promised values increase

and contractual distortions decline as the relationship matures (Ray, 2002, Albuquerque and Hopenhayn, 2004, Martimort et al., 2017a). Our benchmark asks how the screening menu adjusts when those intertemporal instruments are unavailable. The response is cross-sectional. Enforcement first restricts the transfer assigned to the most efficient seller; incentive compatibility then propagates this restriction into bunching and additional output distortions; and, when public recovery is too small, a common continuation rent must be added to the rent of every type. This is the main general lesson of the fixed-terms analysis.

The analysis identifies the optimal allocation within an economically relevant contractual class and provides a benchmark against which the value of backloading, gradual adjustment, or repeated reporting can be assessed.

DECOMPOSITION OF RENTS. As discussed after Theorem 3.5, incentive compatibility implies the envelope condition $\dot{U}(\theta) = -q(\theta)$ a.e., hence

$$U(\theta) = U(\bar{\theta}) + \int_{\theta}^{\bar{\theta}} q(x) dx. \quad (5.1)$$

The regime-dependent term is $U(\bar{\theta})$.

- **Strong and Weak Enforcement** ($\Pi \geq \Pi^*$). In these regimes the least efficient type earns zero rent, $U(\bar{\theta}) = 0$. Thus the agent's rent is purely informational:

$$U(\theta) = \int_{\theta}^{\bar{\theta}} q_r(x) dx,$$

with $r = 0$ under strong enforcement ($\Pi \geq \Pi^{os}$) and $r = r(\Pi) \in [0, r^*(\delta)]$ under weak enforcement ($\Pi^* \leq \Pi \leq \Pi^{os}$). A higher penalty Π relaxes the transfer cap and reduces distortions (lower r), thereby increasing information rents.

- **Very Weak Enforcement** ($\Pi < \Pi^*$). Here a zero-bottom-rent contract could still deter breach by tightening the transfer cap further, but doing so would be suboptimal; the optimal contract instead uses relational incentives. The optimal quantity schedule is $q_{r^*}(\theta)$ with $r^*(\delta) = \frac{1-\delta}{\delta}$, and the principal must leave a strictly positive *enforcement rent* to the least efficient type:

$$U(\bar{\theta}) = \frac{1-\delta}{\delta} (\Pi^* - \Pi) > 0.$$

Using (5.1), the agent's rent decomposes into

$$U(\theta) = \underbrace{\frac{1-\delta}{\delta}(\Pi^* - \Pi)}_{\text{Enforcement rent}} + \underbrace{\int_{\theta}^{\bar{\theta}} q_{r^*}(x) dx}_{\text{Information rent}}.$$

Holding δ fixed, raising Π reduces the enforcement rent at rate $(1-\delta)/\delta$ while leaving q_{r^*} unchanged; thus the agent is strictly worse off as Π increases within this regime, until Π reaches Π^* .

This yields the following comparative statics.¹¹

COROLLARY 5.1 (Institutional improvements and surplus). *Consider the optimal fixed-terms contract at (Π, δ) .*

- (i) **(Output and bunching.)** *For $\Pi \geq \Pi^{os}$, $q = q^{os}$ (no bunching) and is independent of δ . For $\Pi^*(\delta) \leq \Pi \leq \Pi^{os}$, the optimal contract is (t_r, q_r) with $r = r(\Pi)$ decreasing in Π , so bunching weakly decreases with Π . For $\Pi < \Pi^*(\delta)$, the optimal quantity is $q_{r^*}(\delta)$, independent of Π , and $r^*(\delta) = \frac{1-\delta}{\delta}$ decreases in δ , implying less distortion as δ rises. Moreover, $\Pi^*(\delta)$ is increasing in δ .*
- (ii) **(Principal payoff.)** *The principal's expected payoff is weakly increasing in Π and weakly increasing in δ .*
- (iii) **(Agent payoff.)** *Holding δ fixed, the agent's expected payoff is non-monotone in Π : it strictly decreases with Π on $\Pi < \Pi^*(\delta)$ (reduced enforcement rents), and weakly increases with Π on $\Pi^*(\delta) \leq \Pi \leq \Pi^{os}$ (higher information rents), and is constant for $\Pi \geq \Pi^{os}$. Holding Π fixed, the effect of δ on the agent's expected payoff is locally zero in the Strong and Weak regimes (i.e., for changes in δ that do not move the economy across the threshold $\Pi^*(\delta)$), and is generically ambiguous in the Very Weak regime.¹²*

Parts (i)–(ii) follow because increasing Π or δ relaxes the enforcement constraint, expanding the feasible set and allowing the principal to do weakly better. For the agent, (5.1) implies that changes in $U(\bar{\theta})$ shift all rents pointwise. In the very weak regime, $q = q_{r^*}$ is independent of Π while $U(\bar{\theta}) = \frac{1-\delta}{\delta}(\Pi^*(\delta) - \Pi)$, so increasing Π reduces $U(\bar{\theta})$ and hence $U(\theta)$ for all types. In the weak regime, $U(\bar{\theta}) = 0$ and increasing Π lowers r , increasing q_r pointwise and therefore increasing information rents $\int_{\theta}^{\bar{\theta}} q_r(x) dx$.

¹¹Recall that Π^* depends on δ through $r^* = \frac{1-\delta}{\delta}$. We write $\Pi^* = \Pi^*(\delta)$ emphasizing this dependence.

¹²Higher δ improves efficiency but reduces enforcement rents.

6. APPENDIX

PROOF OF LEMMA 2.1. *Necessity.* Incentive compatibility means $U(\theta) = \max_{\hat{\theta} \in \Theta} \{t(\hat{\theta}) - \theta q(\hat{\theta})\}$ for all θ . As the upper envelope of a family of affine functions of θ with slopes $-q(\hat{\theta})$, U is convex, hence locally Lipschitz and absolutely continuous on Θ , and at every differentiability point the envelope theorem gives $\dot{U}(\theta) = -q(\theta)$. Integrating yields (2.2). Convexity of U then implies that $-q$ is non-decreasing, i.e., q is non-increasing; standard incentive arguments give the same conclusion directly, since for $\theta_1 < \theta_2$ adding the two incentive constraints between θ_1 and θ_2 yields $(\theta_2 - \theta_1)(q(\theta_1) - q(\theta_2)) \geq 0$.

Sufficiency. Fix $\theta, \hat{\theta} \in \Theta$ and note that $t(\hat{\theta}) - \theta q(\hat{\theta}) = U(\hat{\theta}) + (\hat{\theta} - \theta)q(\hat{\theta})$. Using (2.2),

$$U(\theta) - (t(\hat{\theta}) - \theta q(\hat{\theta})) = U(\theta) - U(\hat{\theta}) - (\hat{\theta} - \theta)q(\hat{\theta}) = \int_{\hat{\theta}}^{\theta} (q(x) - q(\hat{\theta})) dx \geq 0,$$

where the inequality holds for either order of θ and $\hat{\theta}$ because q is non-increasing. Hence truth-telling is a global best response. Finally, writing $t(\theta) = U(\theta) + \theta q(\theta)$ shows that at a discontinuity point θ_0 of q , continuity of U forces $t(\theta_0^+) - t(\theta_0^-) = \theta_0(q(\theta_0^+) - q(\theta_0^-))$. $\square$

PROOF OF LEMMA 2.4. Multiply both sides of (2.4) by $(1 - \delta)$ and write $U(\hat{\theta}; \theta) := t(\hat{\theta}) - \theta q(\hat{\theta})$. For any fixed report $\hat{\theta}$ and any $\tau \in \mathbb{N}_0 \cup \{+\infty\}$,

$$(1 - \delta) \sum_{s=0}^{\tau-1} \delta^s U(\hat{\theta}; \theta) + (1 - \delta)\delta^\tau (t(\hat{\theta}) - \Pi) = (1 - x)U(\hat{\theta}; \theta) + (1 - \delta)x(t(\hat{\theta}) - \Pi),$$

where $x := \delta^\tau \in (0, 1]$ for finite τ , while $\tau = +\infty$ corresponds to $x = 0$ (perpetual compliance, under the convention $\delta^{+\infty} := 0$). The feasible set of x is thus $X := \{\delta^\tau : \tau \in \mathbb{N}_0\} \cup \{0\} \subset [0, 1]$, which contains both endpoints of $[0, 1]$: $x = 1$ is attained at $\tau = 0$ (immediate breach) and $x = 0$ at $\tau = +\infty$ (no breach). For fixed $\hat{\theta}$ the expression is affine in x , so its supremum over X equals its maximum over $[0, 1]$, attained at an endpoint $x \in \{0, 1\}$ (both feasible) and is worth:

$$\max \left\{ U(\hat{\theta}; \theta), (1 - \delta)(t(\hat{\theta}) - \Pi) \right\}.^{13}$$

Taking now the supremum over $\hat{\theta} \in \Theta$ (attained, as the next two observations show) yields

$$(1 - \delta) \cdot \text{RHS of (2.4)} = \max \left\{ \underbrace{\max_{\hat{\theta}} U(\hat{\theta}; \theta)}_{= U(\theta) \text{ by IC}}, \underbrace{(1 - \delta) \max_{\hat{\theta}} t(\hat{\theta}) - (1 - \delta)\Pi}_{=(1 - \delta) t(\underline{\theta})} \right\}.$$

The equality $\max_{\hat{\theta}} U(\hat{\theta}; \theta) = U(\theta)$ uses truthfulness, which is guaranteed by admissibility (Lemma 2.1 and Definition 2.2). Moreover, t is weakly decreasing in θ , including at any discontinuity point of q : for $\theta_1 < \theta_2$, the envelope condition (2.2) gives

$$\begin{aligned} t(\theta_1) - t(\theta_2) &= \int_{\theta_1}^{\theta_2} q(x) dx + \theta_1 q(\theta_1) - \theta_2 q(\theta_2) \\ &\geq (\theta_2 - \theta_1) q(\theta_2) + \theta_1 q(\theta_1) - \theta_2 q(\theta_2) \\ &= \theta_1 (q(\theta_1) - q(\theta_2)) \geq 0, \end{aligned}$$

since q is non-increasing and $\theta_1 > 0$. Hence $\max_{\hat{\theta}} t(\hat{\theta}) = t(\underline{\theta})$. Therefore, (2.4) is equivalent to

$$U(\theta) \geq \max \left\{ U(\theta), (1 - \delta)(t(\underline{\theta}) - \Pi) \right\} \quad \forall \theta,$$

which is in turn equivalent to (2.5). $\square$

PROOF OF LEMMA 3.4. Equation (3.3) gives $r(\underline{\theta}) = 0$. Wherever $\theta f(\theta) - F(\theta) > 0$, differentiation yields

$$r'(\theta)(\theta f(\theta) - F(\theta)) = 2F(\theta)f(\theta) - r\theta f'(\theta) = 2F(\theta)f(\theta) - \frac{F(\theta)^2 + rF(\theta)}{f(\theta)} f'(\theta).$$

Assumption 3.2 therefore implies $r'(\theta) > 0$ for interior values in the relevant range $r \in [0, r^*]$ (with $r'(\underline{\theta}) = 0$). If the denominator reaches zero before $\bar{\theta}$, then $r(\theta)$ tends to $+\infty$, and every $r \in [0, r^*]$ is reached once. If the denominator remains positive on Θ , the range of the increasing locus is $[0, \hat{r}]$, where

$$\hat{r} = r(\bar{\theta}) = \frac{1}{\bar{\theta} f(\bar{\theta}) - 1}.$$

This gives the two cases stated in the lemma.

¹³Had breach dates been restricted to finite τ , the value at $x = 0$ would only be approached as $\tau \rightarrow \infty$, a supremum rather than a maximum; including $\tau = +\infty$, i.e., the deviation “report $\hat{\theta}$ and comply forever,” closes the feasible set. This endpoint is precisely the deviation controlled by incentive compatibility, so the reduction is unaffected.

At $r = \widehat{r}$, the partial-bunching and complete-bunching formulas coincide because

$$\bar{\theta} + \frac{1 + \widehat{r}}{f(\bar{\theta})} = \bar{\theta}(1 + \widehat{r}).$$

On the complete-bunching branch, $S'(q_r^{\text{pool}}) = \bar{\theta}(1 + r)$, so q_r^{pool} is continuous and decreasing in r , and $\Pi(r) = \bar{\theta}q_r^{\text{pool}}$. Together with the corresponding monotonicity on the partial-bunching branch, this makes $\Pi(r)$ continuous and strictly decreasing on $[0, r^*]$. $\square$

PROOF OF LEMMA 4.1. *Optimality conditions for problem (P^A) .* Using Proposition 6.1, we formulate the problem in the usual optimal-control form. We explicitly incorporate monotonicity by introducing the control variable $z(\theta) = \dot{q}(\theta)$. Problem (P^A) is then formulated as an optimal control problem with state variables $y(\theta)$ and $q(\theta)$ and control $z(\theta)$. The co-state variables corresponding to (6.2) and (6.3) are denoted by $\lambda_1(\theta)$ and $\lambda_2(\theta)$ respectively.

The Hamiltonian is

$$H(y, q, z, \lambda_1, \lambda_2, \theta) = (S(q) - y)f(\theta) + \lambda_1\theta z + \lambda_2 z,$$

which is concave in (y, q, z) for all θ . Let $(y(\theta), q(\theta), z(\theta))$ be an admissible triplet with continuous, a.e. differentiable $y(\theta), q(\theta)$ and piecewise continuous $z(\theta)$. Then $(y(\theta), q(\theta), z(\theta))$ is optimal if and only if there exist continuous, piecewise differentiable co-state variables $(\lambda_1(\theta), \lambda_2(\theta))$ such that the following conditions (see Seierstad and Sydsaeter (1986), pp. 85, 396) are satisfied:

$$z(\theta) \in \arg \max_z H(y(\theta), q(\theta), z, \lambda_1(\theta), \lambda_2(\theta), \theta) \quad \forall \theta, \quad (6.1)$$

$$\dot{y}(\theta) = \theta z(\theta), \quad (6.2)$$

$$\dot{q}(\theta) = z(\theta) \leq 0, \quad (6.3)$$

$$\dot{\lambda}_1(\theta) = f(\theta) \quad \text{a.e.}, \quad (6.4)$$

$$\dot{\lambda}_2(\theta) = -S'(q(\theta))f(\theta) \quad \text{a.e.}, \quad (6.5)$$

$$\lambda_1(\underline{\theta}) = \frac{1 - \delta}{\delta} - \beta(1 - \delta), \quad \lambda_2(\underline{\theta}) = 0, \quad \beta \geq 0 \quad (= 0 \text{ if } y(\underline{\theta}) - \Pi > 0), \quad (6.6)$$

$$\lambda_1(\bar{\theta}) = \gamma, \quad \lambda_2(\bar{\theta}) = -\gamma\bar{\theta}, \quad \gamma \geq 0. \quad (6.7)$$

Denote

$$r = \lambda_1(\underline{\theta}) = \frac{1-\delta}{\delta} - \beta(1-\delta) = \frac{(1-\delta)(1-\beta\delta)}{\delta}. \quad (6.8)$$

Conditions (6.2), (6.3) and (6.4)-(6.6) imply

$$\lambda_1(\theta) = r + F(\theta), \quad (6.9)$$

and

$$\lambda_2(\theta) = - \int_{\underline{\theta}}^{\theta} S'(q(u)) f(u) du. \quad (6.10)$$

Define $\psi(\theta) = \lambda_1(\theta)\theta + \lambda_2(\theta)$. From (6.9)-(6.10) we get

$$\psi(\theta) = (r + F(\theta))\theta - \int_{\underline{\theta}}^{\theta} S'(q(u)) f(u) du.$$

Optimality condition (6.1) yields

$$\psi(\theta)z(\theta) = 0, \quad \psi(\theta) \geq 0 \quad \forall \theta. \quad (6.11)$$

Derivation of r^ .* Note that

$$\psi(\underline{\theta}) = r\underline{\theta} \geq 0. \quad (6.12)$$

Thus $r \geq 0$. From (6.8), the minimal value $r = 0$ corresponds to $\beta = 1/\delta > 0$, and the maximal value $r^* = \frac{1-\delta}{\delta}$ corresponds to $\beta = 0$.

The form of optimal $q(\theta)$. If $\psi(\theta) > 0$ over a non-degenerate interval, then $z(\theta) = 0$ on this interval, rendering both state variables $q(\theta)$ and $y(\theta)$ constant due to (6.2) and (6.3). Conversely, if $\psi(\theta) = 0$ over a non-degenerate interval Θ' it follows that $\dot{\psi}(\theta) = 0$, leading to

$$S'(q(\theta)) = \theta + \frac{F(\theta) + r}{f(\theta)} \quad \text{for all } \theta \in \Theta'.$$

This condition implies that $q(\theta) = q^{(r)}(\theta)$ for all $\theta \in \Theta'$.

We now show that there can be at most one bunching interval. Suppose, to the contrary, that there are two disjoint bunching intervals, denoted $\Theta_1 = [\underline{\theta}, \theta_1)$ and $\Theta_3 = (\theta_2, \bar{\theta}]$.¹⁴

¹⁴Note that θ_1 may coincide with $\underline{\theta}$.

Within the intermediate interval $\Theta_2 = (\theta_1, \theta_2)$, we have $\psi(\theta) = 0$, so $q(\theta) = q^{(r)}(\theta)$ for all $\theta \in \Theta_2$. For all $\theta \in \Theta_1$, $q(\theta) = q^{(r)}(\theta_1) = q_1$, and for all $\theta \in \Theta_3$, $q(\theta) = q^{(r)}(\theta_2) = q_3$. If Θ_3 were nonempty, then $q_3 > q^{(r)}(\theta)$ for all $\theta \in \Theta_3$, and hence $S'(q_3) < S'(q^{(r)}(\theta))$. Thus,

$$\dot{\psi}(\theta) = r + F(\theta) + f(\theta)\theta - S'(q_3)f(\theta) > r + F(\theta) + f(\theta)\theta - S'(q^{(r)}(\theta))f(\theta) = 0$$

for all $\theta \in \Theta_3$. Since $\psi(\theta_2) = 0$ and, by the boundary condition, $\psi(\bar{\theta}) = 0$, this is a contradiction. Thus there may be only one bunching interval $[\underline{\theta}, \theta_1]$.

If bunching covers the whole support, $q(\theta) = Q$ and $z(\theta) = 0$ for every interior type. The terminal condition $\psi(\bar{\theta}) = 0$ gives

$$S'(Q) = \bar{\theta}(1 + r), \quad (6.13)$$

hence $Q = q_r^{\text{pool}}$. In case (ii), for $r \geq \hat{r}$, the monotonicity of the locus implies that the associated switching function is non-negative throughout the interior, so the completely bunched schedule satisfies the optimality conditions. At $r = \hat{r}$, it joins the partial-bunching branch continuously.

Derivation of (3.3) when the separating region is nonempty. The first equation to determine θ_1 and r is $\psi(\theta_1) = 0$:

$$0 = (r + F(\theta_1))\theta_1 - \int_{\underline{\theta}}^{\theta_1} S'(q(u))f(u) du. \quad (6.14)$$

Since $q(u) = q_1 = q^{(r)}(\theta_1)$ for $u \in [\underline{\theta}, \theta_1]$, and $S'(q_1) = \theta_1 + \frac{r+F(\theta_1)}{f(\theta_1)}$, we obtain

$$r\theta_1 = (r + F(\theta_1)) \frac{F(\theta_1)}{f(\theta_1)}, \quad (6.15)$$

which is equivalent to (3.3).

Construction of the optimal contract. We have established that the optimal output is the schedule in Definition 3.2: it has the partial-bunching form depicted in Figure 1 when the separating region is nonempty, and is completely pooled at q_r^{pool} otherwise.

At $\Pi = \Pi^{os}$, we have $r = 0$. Hence (4.5) has the trivial solution $\theta_r = \underline{\theta}$, and the optimal output is the static second-best contract. In this case, $\beta = 1/\delta > 0$, so (4.4) binds and $y(\underline{\theta}) = \Pi^{os}$.

For all $\Pi \in [\Pi^*, \Pi^{os}]$, constraint (4.4) binds so $y(\underline{\theta}) = \Pi$ and $t(\underline{\theta}) = \Pi$. Define r via $\Pi(r) = \underline{\theta}q_r(\underline{\theta}) + \int_{\underline{\theta}}^{\bar{\theta}} q_r(\theta) d\theta = \Pi$. By Lemma 3.4, $\Pi(r)$ is continuous and strictly decreasing, so r is uniquely determined for $\Pi \in [\Pi^*, \Pi^{os}]$.

For $\Pi < \Pi^*$, constraint (4.4) is slack and the solution is the same as at Π^* . $\square$

PROOF OF LEMMA 4.2. Modulo the change of variables, the optimality conditions (6.1)-(6.5) are the same as for (P^A) . The transversality condition (6.6) is replaced by

$$\lambda_1(\underline{\theta}) = \beta'(1 - \delta), \quad \lambda_2(\underline{\theta}) = 0, \quad \beta' \geq 0 \quad (= 0 \text{ if } \Pi - t(\underline{\theta}) > 0).$$

In this case the multiplier in (P^B) is formally unrestricted. The co-state analysis is otherwise identical to that of (P^A) : $\lambda_1(\theta) = r + F(\theta)$ with $r := \beta'(1 - \delta) \geq 0$, and when the cap binds, $t(\underline{\theta}) = \Pi$ determines r through (4.5). Since the mapping $r \mapsto t_r(\underline{\theta})$ in (3.5) is decreasing, with values Π^{os} at $r = 0$ and Π^* at $r = r^*$, the solution of (4.5) satisfies $r \in [0, r^*]$ precisely when $\Pi \in [\Pi^*, \Pi^{os}]$. On this range, Assumption 3.2 guarantees that the generalized virtual cost $\theta + \frac{F(\theta) + r}{f(\theta)}$ is increasing, so the single-bunching schedule q_r is well defined and the sufficiency argument applies. If $r = 0$, then the optimal output is the static second-best contract and the corresponding transfers are $t_0(\theta)$ defined by (3.4) with zero rent for $\bar{\theta}$. If $r > 0$, then the optimal output is $q_r(\theta)$ and transfers are given by (3.4) with zero rent for $\bar{\theta}$.

For $\Pi < \Pi^*$, the cap would correspond to $r > r^*$, outside the range covered by Assumption 3.2. Part 3 of the lemma therefore claims only that the cap binds and makes no claim about the shape of the optimal output. This is sufficient because the comparison in Section 4 directly shows that the value of (P^B) is weakly dominated by that of (P^A) . $\square$

PROPOSITION 6.1 (Continuity). *Under Assumption 3.2: (i) every optimal solution (t, q) of (P) has a quantity schedule that is weakly decreasing on Θ and continuous at every interior type; (ii) the endpoint values $q(\underline{\theta})$ and $q(\bar{\theta})$ can differ from the one-sided limits $q(\underline{\theta}^+)$ and $q(\bar{\theta}^-)$ only in payoff-irrelevant ways—and at $\underline{\theta}$ only in the strong-enforcement regime $\Pi > \Pi^{os}$ —so that replacing them by those limits preserves admissibility, feasibility, and the value of the objective; (iii) consequently, every optimal schedule agrees on $(\underline{\theta}, \bar{\theta})$ with a schedule that is continuous on all of Θ and is itself optimal.*

PROOF OF PROPOSITION 6.1. Weak decrease holds by admissibility. The argument rests on a reduction of both programs to a problem that is strictly concave in the quantity schedule and in which every remaining constraint is linear.

Step 1: Reduction. By Definition 2.2, an admissible pair is parameterized by (q, u) , where q is piecewise continuous and non-increasing, $u := U(\bar{\theta}) \geq 0$, and $t(\theta) = \theta q(\theta) + \int_{\theta}^{\bar{\theta}} q(x) dx + u$. By Fubini's theorem, the buyer's objective equals

$$\int_{\underline{\theta}}^{\bar{\theta}} (S(q(\theta)) - t(\theta)) f(\theta) d\theta = W(q) - u, \quad W(q) := \int_{\underline{\theta}}^{\bar{\theta}} \left(S(q(\theta)) - \left(\theta + \frac{F(\theta)}{f(\theta)} \right) q(\theta) \right) f(\theta) d\theta,$$

and the breach transfer equals $t(\underline{\theta}) = K(q) + u$, where $K(q) := \underline{\theta} q(\underline{\theta}) + \int_{\underline{\theta}}^{\bar{\theta}} q(x) dx$. Since $U(\theta) = u + \int_{\theta}^{\bar{\theta}} q(x) dx$ is non-increasing, the pointwise constraints of systems (A) and (B) can bind only at $\bar{\theta}$. Hence

$$(P^A) : \max_{q, u} W(q) - u \text{ s.t. } K(q) + u \geq \Pi \text{ and } u \geq (1 - \delta)(K(q) + u - \Pi), \text{ i.e., } u \geq r^*(K(q) - \Pi);$$

$$(P^B) : \max_{q, u} W(q) - u \text{ s.t. } K(q) + u \leq \Pi \text{ and } u \geq 0.$$

Three facts follow. First, q enters all constraints only through monotonicity and the *linear* functional $K(q)$. Second, W is strictly concave in q (by strict concavity of S) and the objective is linear in u . Third, the value of q at any single interior type is payoff-irrelevant and is squeezed by monotonicity between the one-sided limits; only $q(\underline{\theta})$ is payoff-relevant as a point value, through the term $\underline{\theta} q(\underline{\theta})$ in $K(q)$.

Step 2: No interior jump. Let (q, u) be optimal for either program and suppose q jumps at $\theta_0 \in (\underline{\theta}, \bar{\theta})$, with $a := q(\theta_0^-) > q(\theta_0^+) =: b$. For $h > 0$ small enough that $[\theta_0 - h, \theta_0 + h] \subset (\underline{\theta}, \bar{\theta})$, let q_h equal the average $c_h := (2h)^{-1} \int_{\theta_0-h}^{\theta_0+h} q(x) dx$ on $[\theta_0 - h, \theta_0 + h]$ and coincide with q elsewhere. Since $q(\theta_0 - h) \rightarrow a$, $q(\theta_0 + h) \rightarrow b$ and $c_h \rightarrow (a + b)/2 \in (b, a)$ as $h \downarrow 0$, the schedule q_h is non-increasing for h small. Moreover $\int q_h = \int q$ and $q_h(\underline{\theta}) = q(\underline{\theta})$, so $K(q_h) = K(q)$: with u unchanged, *every* constraint of the program is preserved, and the transfer re-derived from (2.2) makes the perturbed pair admissible. As for the value, using $\int_{\theta_0-h}^{\theta_0+h} (q_h - q) dx = 0$ together with the continuity of f and of $\theta \mapsto \theta f(\theta) + F(\theta)$,

$$W(q_h) - W(q) = \int_{\theta_0-h}^{\theta_0+h} (S(q_h) - S(q)) f d\theta - \int_{\theta_0-h}^{\theta_0+h} (\theta f + F)(q_h - q) d\theta = 2h f(\theta_0) \Gamma + o(h),$$

where $\Gamma := S\left(\frac{a+b}{2}\right) - \frac{1}{2}(S(a) + S(b)) > 0$ by strict concavity of S and $a > b$. The type-dependent payoff weight $f(\theta_0) > 0$ enters multiplicatively and does not affect the sign. Hence $W(q_h) > W(q)$ for h small, contradicting optimality. Therefore q has no interior jump; and at interior types where the one-sided limits agree, monotonicity forces q to equal them. So q is continuous on $(\underline{\theta}, \bar{\theta})$, proving (i).

Step 3: Continuous schedules attain the supremum. For any admissible (q, u) , the forward averages $q_h(\theta) := h^{-1} \int_{\theta}^{\theta+h} \tilde{q}(x) dx$ (with $\tilde{q}$ extended by $q(\bar{\theta}^-)$ to the right of $\bar{\theta}$) are continuous, non-increasing, and satisfy $q_h \leq q$ pointwise, hence $K(q_h) \leq K(q)$: the pair (q_h, u) remains feasible for (P^B) and its value converges to $W(q) - u$ by bounded convergence. Symmetrically, the backward averages $q^h \geq q$ (with $\tilde{q}$ extended by $q(\underline{\theta})$ to the left of $\underline{\theta}$, so that $q^h(\underline{\theta}) = q(\underline{\theta})$ and $K(q^h) \downarrow K(q)$), paired with $u_h := \max\{u, r^*(K(q^h) - \Pi)\} \rightarrow u$, are feasible for (P^A) with value converging to $W(q) - u$. Hence, in both programs, the supremum over *continuous* admissible schedules equals the supremum over all admissible schedules; in particular, the continuous schedules identified in Lemmas 4.1 and 4.2 are optimal in the full admissible class, not merely among continuous ones.

Step 4: Endpoints. The value $q(\bar{\theta})$ enters neither W , nor K , nor any constraint, so setting $q(\bar{\theta}) = q(\bar{\theta}^-)$ is payoff-irrelevant. At $\underline{\theta}$, monotonicity permits only an upward atom $\alpha := q(\underline{\theta}) - q(\underline{\theta}^+) \geq 0$, whose sole effect is to raise the breach transfer $t(\underline{\theta})$ by $\underline{\theta}\alpha$. In (P^A) , if $K(q) > \Pi$ at the optimum, then $u = r^*(K(q) - \Pi)$ and removing the atom strictly reduces the required continuation rent (recall $r^* > 0$ and $\underline{\theta} > 0$): hence $\alpha = 0$. If instead $t(\underline{\theta}) = \Pi$ binds, an atom forces the interior schedule q° to satisfy $K(q^\circ) < \Pi$; since the maximizer of the strictly concave W over the convex set $\{q : K(q) \leq \Pi\}$ is unique up to a.e. equality and equals q_r with $K(q_r) = \Pi$ (Lemma 4.2 together with Step 3), and since $q^\circ = q_r$ a.e. would imply $K(q^\circ) \geq \Pi$, we obtain $W(q^\circ) < W(q_r)$ strictly: again $\alpha = 0$. The same comparison rules out atoms in (P^B) whenever the cap binds ($\Pi \leq \Pi^{os}$). Only in the strong-enforcement regime $\Pi > \Pi^{os}$, where the cap is slack at the optimum, is a small atom feasible; it is then payoff-irrelevant, and the right-continuous version is optimal with the same value. This proves (ii) and (iii). $\square$

BIBLIOGRAPHY

- ALBUQUERQUE, R. AND H. A. HOPENHAYN (2004): "Optimal lending contracts and firm dynamics," *The Review of Economic Studies*, 71, 285–315.
- ANTRÀS, P. AND C. F. FOLEY (2015): "Poultry in Motion: A Study of International Trade Finance Practices," *Journal of Political Economy*, 123, 853–901.
- BARON, D. P. AND D. BESANKO (1984): "Regulation and Information in a Continuing Relationship," *Information Economics and Policy*, 1, 267–302.
- BARON, D. P. AND R. B. MYERSON (1982): "Regulating a Monopolist with Unknown Costs," *Econometrica*, 50, 911–930.

FUCHS, W., B. GREEN, AND D. LEVINE (2022): “Optimal arrangements for distribution in developing markets: Theory and evidence,” *American Economic Journal: Microeconomics*, 14, 411–450.

GAUTIER, A. AND M. MITRA (2006): “Regulating a monopolist with limited funds,” *Economic Theory*, 27, 705–718.

LAFONTAINE, F. AND K. L. SHAW (1999): “The Dynamics of Franchise Contracting: Evidence from Panel Data,” *Journal of Political Economy*, 107, 1041–1080.

LEVIN, J. (2003): “Relational Incentive Contracts,” *American Economic Review*, 93, 835–857.

MACCHIAVELLO, R. AND A. MORJARIA (2015): “The Value of Relationships: Evidence from a Supply Shock to Kenyan Rose Exports,” *American Economic Review*, 105, 2911–2945.

MALCOMSON, J. (2016): “Relational Incentive Contracts with Persistent Private Information,” *Econometrica*, 84, 317–346.

MARTIMORT, D., A. SEMENOV, AND L. STOLE (2017a): “A Theory of Contracts with Limited Enforcement,” *Review of Economic Studies*, 84, 816–852.

——— (2017b): “Optimal stationary contract with two-sided imperfect enforcement and persistent adverse selection,” *Economics Letters*, 159, 18–22.

RAY, D. (2002): “The time structure of self-enforcing agreements,” *Econometrica*, 70, 547–582.

SEIERSTAD, A. AND K. SYDSAETER (1986): *Optimal control theory with economic applications*, Elsevier North-Holland, Inc.

THOMAS, L. (2002): “Non-linear pricing with budget constraint,” *Economics Letters*, 75, 257–263.

TROYA-MARTINEZ, M. (2017): “Self-enforcing trade credit,” *International Journal of Industrial Organization*, 52, 333–357.