DISCUSSION PAPER SERIES

DP15251

PARTICIPATION CONSTRAINTS IN DISCONTINUOUS ADVERSE SELECTION MODELS

David Martimort and Lars Stole

INDUSTRIAL ORGANIZATION

ISSN 0265-8003

PARTICIPATION CONSTRAINTS IN DISCONTINUOUS ADVERSE SELECTION MODELS

David Martimort and Lars Stole

Discussion Paper DP15251
Published 04 September 2020
Submitted 17 August 2020

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

PARTICIPATION CONSTRAINTS IN DISCONTINUOUS ADVERSE SELECTION MODELS

Abstract

We present a set of necessary and sufficient conditions for a class of optimal control problems with pure state constraints for which the objective function is linear in the state variable but the objective function is only required to be upper semi-continuous in the control variable. We apply those conditions to a number of economic environments in contract theory where discontinuities in objectives prevail. Examples of applications include nonlinear pricing of digital goods, nonlinear pricing under competitive threat, and common agency models of regulation.

JEL Classification: D82

Keywords: Optimal Control, non-smooth optimization, convex analysis, type-dependent participation constraints, Principal-Agent Models

David Martimort - david.martimort@psemail.eu
Paris School of Economics and CEPR

Lars Stole - lars.stole@chicagobooth.edu
University of Chicago Booth School of Business

Working paper

PARTICIPATION CONSTRAINTS IN DISCONTINUOUS ADVERSE SELECTION MODELS[1]

August 6, 2020

David Martimort and Lars Stole

Abstract. We present a set of necessary and sufficient conditions for a class of optimal control problems with pure state constraints for which the objective function is linear in the state variable but the objective function is only required to be upper semi-continuous in the control variable. We apply those conditions to a number of economic environments in contract theory where discontinuities in objectives prevail. Examples of applications include nonlinear pricing of digital goods, nonlinear pricing under competitive threat, and common agency models of regulation.

Keywords. Optimal control, non-smooth optimization, convex analysis, type-dependent participation constraints, principal-agent models.

1. INTRODUCTION

The textbook treatment of optimal screening contracts typically takes the agent's outside option as a fixed constant, independent of type.[1] More complex settings which allow for competition by rival principals, non-trivial ownership rights on productive assets and type-dependent fixed costs require a departure from this restrictive assumption. Lewis and Sappington (1989a) initiated the seminal study of screening contracts in this more general setting by constructing the solution to a class of optimal control problems with type-dependent participation constraints. This class of problems was further enriched by Maggi and Rodriguez (1995) with the most general statement of the problem and its solution culminating in the analysis offered by Jullien (2000).

These techniques have allowed modelers to apply the optimal contracting paradigm to more general economic contexts, unveiling new features of optimal contracts. Applications have spread through many fields including the design of nonlinear prices under the threat of bypass (Curien, Jullien and Rey, 1998; Biglaiser and Mezzetti, 1993), competition in nonlinear prices (Martimort and Stole, 2009; Stole, 1995, 2003; Calzolari and DeNicolo, 2013), trade policy in open economies (Brainard and Martimort, 1996), regulation of privately-owned monopolies (Caillaud, 1990), and optimal contracting under liability constraints (Ollier and Thomas, 2013) or in dynamic contracting environments (Deb and Said, 2015), to name a few examples.

Unfortunately, the existing techniques also have their own limits. In particular, the need for tractability has led authors to restrict their analysis to economic environments which

[1]We are especially thankful to John Birge for many helpful discussions. We also thank Simon Board, the Editor in charge, three referees for very useful comments and suggestions, and seminar participants at the 2019 Summer Meetings of the North American Econometric Society (Seattle) and the European Econometric Society (Manchester). A less general version of the main theorem in this paper appeared in an earlier, unpublished note, *"Necessary and sufficient conditions for non-smooth linear-state optimal control problems"* (2009). The present paper provides the more general result along with a geometric intuition for its proof and several relevant applications. The usual disclaimer applies.

[a]Paris School of Economics-EHESS, david.martimort@psemail.eu.
[b]University of Chicago Booth School of Business, lars.stole@chicagobooth.edu.

[1]See Laffont and Martimort (2002, Chapter 3) for instance.

are sufficiently *smooth.* In many circumstances, such as when firms face non-trivial fixed costs or sunk investments, the environment is inherently discontinuous. In other settings, such as when principals compete against one another using payment schedules, equilibria may emerge which exhibit discontinuities in each player's payoff function. In those cases, we are left uncertain about the consequences of such discontinuities for optimal contracts and the generality of results when environments (and equilibria) are not assumed to be smooth *a priori.* Important economic insights may have gone unnoticed because of our restricted attention. Developing the required techniques for *non-smooth* environments and showing how they apply in practice are the purposes of this paper.

First, we present a set of necessary and sufficient conditions for a class of optimal control problems with pure state constraints for which the objective function is linear in the state variable but the objective function may exhibit kinks and discontinuities in the control variable. Second, we apply these techniques to quite natural contracting environments where existing techniques are inadequate or have made implicit restrictions on contracting possibilities that have come unnoticed. Examples of applications include nonlinear pricing of digital goods, nonlinear pricing under competitive threat, common agency models of regulation and multi-unit auctions with externalities.

Section 2 presents our main result: A theorem that characterizes solutions to optimal control problems in which the objective function is only restricted to be upper semi-continuous. This theorem builds on earlier work by Vinter and Zheng (2000), but it refines its application to the case of quasi-linear objectives which is prevalent in contract theory. While Vinter and Zheng (2000) focuses on necessary conditions for optimality, we prove that these conditions are also sufficient in our context. We also discuss to which extent this theorem extends the existing literature and especially the work by Jullien (2000) under much weaker conditions. Section 3 develops various applications of our framework, deriving for each of those economic insights that would not have been available without the use of new techniques. The Appendix contains not only the proof of our Theorem and of various results related to our economic applications but also some brief overview of optimization techniques in non-smooth environments, providing more intuition for our main theorem.

2. THE THEOREM

We will consider control problems in which the state variable, u, is restricted to be an absolutely continuous function on the interval $\Theta = [\underline{\theta}, \overline{\theta}]$; $AC(\Theta, \mathbb{R})$ denotes the feasible set of such functions. As a motivation, in the context of principal-agent models, the state variable is typically the agent's information rent as a function of his type. The requirement of absolute continuity then follows from standard incentive compatibility arguments.[2] We will also restrict attention to problems in which that state variable must satisfy a non-negativity constraint:

$$(2.1) \qquad u(\theta) \geq 0 \quad \forall \theta \in \Theta \equiv [\underline{\theta}, \overline{\theta}].$$

Using again our motivation of principal-agent problems, the non-negativity constraint (2.1) corresponds to a participation constraint that requires the agent to accept an offer

[2]See Milgrom and Segal (2002) and Carbajal and Ely (2013). As an example, if the agent's utility is continuously differentiable in type with a uniformly-bounded derivative, then the agent's indirect utility function is necessarily Lipschitz continuous (and therefore absolutely continuous).

rather than take an outside option, which is here normalized at zero.[3] When the state variable u is both absolutely continuous and non-negative, it is said *admissible*.

We are interested in the following pure-state control program:

$$(\mathcal{P}) : \text{Maximize}_{u \in AC(\Theta,\mathbb{R})} \int_{\underline{\theta}}^{\overline{\theta}} (s(\theta, \dot{u}(\theta)) - u(\theta) f(\theta))\, d\theta \text{ subject to (2.1).}$$

Regularity. We only assume that $s(\theta, v)$ is an upper semi-continuous function of v for all θ, bounded from above, and that $f(\theta)$ is a positive, bounded function giving rise to an absolutely-continuous definite integral $F(\theta) \equiv \int_{\underline{\theta}}^{\theta} f(\theta) d\theta$. Without loss of generality, we normalize f such that $F(\overline{\theta}) = 1$. In the principal-agent context, this allows us to interpret F as a continuous probability distribution and f as its associated density. We also make a minimal technical assumption that $s(\cdot, \cdot)$ is $\mathcal{L} \times \mathcal{B}$-measurable, where $\mathcal{L}$ denotes the set of Lebesgue measurable subsets of Θ and $\mathcal{B}$ is the set of Borel measurable subsets of $\mathbb{R}$.

Linearity. We define the integrand for program $(\mathcal{P})$ as $L(\theta, u, \dot{u}) \equiv s(\theta, \dot{u}) - uf(\theta)$. We should make clear that the key restriction we have placed on $(\mathcal{P})$ is that, for any θ and $\dot{u}$, the maximand is a linear function of the state variable u. As we will see below in our applications, this linearity is found in a number of economic problems, especially in contract theory where agents are risk-neutral and payoffs are linear in money. The function $s(\theta, \dot{u})/f(\theta)$ can there be viewed as a surplus function while u is the share of this surplus that is captured by the agent – his information rent.

The linearity restriction is the primary source of many sharp results in the analysis that follows, including the ability for us to relax the continuity of s, the characterization of the solution by means of a simple and handy generalized gradient condition, and the fact that necessary conditions for optimality are also sufficient. Indeed, non-smooth techniques are particularly useful if one can find the solutions of such control problems as pointwise optima. This is where the assumption of linearity of the maximand in u provides purchase. Linearity allows such transformation; a point which is already well-known from principal-agent screening models in quasi-linear and smooth environments (Myerson, 1981, Lemma 3; Baron and Myerson, 1982, Lemma 2; Laffont and Martimort, 2002, Chapter 3 for a textbook treatment).[4] Things are somewhat more complicated with a state-dependent reservation utility but the basic intuition remains. Linearity allows us to separate incentive and participation concerns[5] from the possible non-smoothness of the surplus function that is handled by taking the super-differential of the concave envelope for $s(\theta, v)$.

Comparison with complete information models. To better isolate the role of non-smoothness, consider the case where the integrand is reduced to $L(\theta, u, v) \equiv s(\theta, v)$.

[3]Readers accustomed with the extant literature will certainly have noticed that even in the case of a non-zero outside option, the formulation of the participation constraint as (2.1) comes from a re-normalization. Our examples below repeatedly illustrate this point.

[4]When quasi-linearity is not assumed, this familiar trick no longer works as, for instance, in the well-known model of the optimal taxation due to Mirrlees (1971).

[5]Incentive and participation concerns are captured by the term $F(\theta) - \overline{\gamma}(\theta)$ in the optimality condition (2.5) below.

In the contract theory applications below, it would correspond to a scenario with complete information on the agent's preferences so that the principal can fully extract the agent's rent. The optimization problem so constructed can be solved pointwise by means of standard techniques for non-smooth problems.[6] Any solution $v^*(\theta)$ must satisfy the following pair of conditions,

$$0 \in \partial\overline{co}(s)(\theta, v^*(\theta)) \text{ and } \overline{co}(s)(\theta, v^*(\theta)) = s(\theta, v^*(\theta)), \tag{2.2}$$

where $\overline{co}(s)(\theta, v)$ is the concave majorization of the function s evaluated at (θ, v) and $\partial\overline{co}(s)(\theta, v)$ is the set of gradients of s at (θ, v).[7] The solution to our original program $(\mathcal{P})$ will differ from v^* in (2.2) as a result of the addition of the linear term $-uf(\theta)$ in the Lagrangean. Condition (2.5) below indicates how the solution needs to be modified. As familiar from the principal-agent literature, $u(\theta)$ will have to maximize a virtual surplus function obtained by compounding the impact of participation and incentive constraints.

Heuristic Approach. Following Jullien (2000), a heuristic treatment of problem $(\mathcal{P})$ would consist in first adding a Lagrange multiplier $\mu(d\theta)$ to the participation constraint (2.1) with the complementarity slackness condition

$$\mu(d\theta) = 0 \quad \text{if } \theta \in \{\tilde{\theta} : \quad u(\tilde{\theta}) > 0\}.$$

Second, we could then form a Lagrangean as

$$\int_{\underline{\theta}}^{\overline{\theta}} (s(\theta, \dot{u}(\theta)) - u(\theta)f(\theta))d\theta + \int_{\underline{\theta}}^{\overline{\theta}} u(\theta)\mu(d\theta).$$

Third, with a simple integration by parts, we would write the integrand as

$$\int_{\underline{\theta}}^{\overline{\theta}} (s(\theta, \dot{u}(\theta)) + (F(\theta) - \overline{\gamma}(\theta))\dot{u}(\theta))\, d\theta$$

where the adjoint function $\overline{\gamma}(\theta) = \int_{[\underline{\theta},\theta)} \mu(d\tilde{\theta})$, is right-continuous, strictly increasing at points where (2.1) is binding, and satisfies the boundary conditions $\overline{\gamma}(0) = 0, \overline{\gamma}(1) = 1$. Those facts allow us to identify $\overline{\gamma}$ with a distribution function. Finally, pointwise optimization would give us the optimality condition

$$\dot{\overline{u}}(\theta) \in \arg\max_v s(\theta, v) + (F(\theta) - \overline{\gamma}(\theta))v.$$

Two difficulties come with this heuristic approach. The first one is merely technical and puts conditions on the Lagrange multiplier. Indeed, integrating by parts requires that $\mu(d\theta)$ lies in the dual space of non-negative functions in $AC(\Theta, \mathbb{R})$; i.e., $\mu(d\theta)$ must be the "derivative" of a function of bounded variation. The second difficulty is that, once we proceed to pointwise optimization, we may have to deal with a non-smooth objective since $s(\theta, v)$ is only required to be upper semi-continuous. The optimality conditions have to be expressed by means of tools imported from non-smooth, convex analysis.

Main Result. We now present our main result for this class of problems.

[6]See Section 3.3 below for details.

[7]Appendix A provides a brief discussion of non-smooth, convex analysis. Because we are focused on maximization, our tools rely on concave majorizations (i.e., the minimal concave envelope of a function) rather than convex minorizations. Likewise, we are interested in the set of gradients of a concave function (super-differentials) rather than the set of gradients of a convex function (sub-differentials).

Theorem 1 *$\overline{u}$ is a solution to program $(\mathcal{P})$ if and only if $\overline{u}$ is admissible and there exists a probability measure μ defined over the Borel subsets of Θ with an associated adjoint function, $\overline{\gamma}:\Theta\to[0,1]$, defined by $\overline{\gamma}(\underline{\theta})=0$ and if $\theta>\underline{\theta}$,*

$$\overline{\gamma}(\theta)=\int_{[\underline{\theta},\theta)}\mu(d\tilde{\theta}),$$

such that the following conditions are satisfied:

$$\int_{\underline{\theta}}^{\overline{\theta}}\overline{u}(\tilde{\theta})\mu(d\tilde{\theta})=0, \tag{2.3}$$

$$\overline{co}(s)(\theta,\dot{\overline{u}}(\theta))=s(\theta,\dot{\overline{u}}(\theta)) \text{ for a.e. } \theta\in\Theta, \tag{2.4}$$

$$0\in F(\theta)-\overline{\gamma}(\theta)+\partial\overline{co}(s)(\theta,\dot{\overline{u}}(\theta)) \text{ for a.e. } \theta\in\Theta. \tag{2.5}$$

The conditions in Theorem 1 are similar to those in Theorem 1 in Jullien (2000). In both theorems, necessary and sufficient conditions are stated in terms of a probability measure which serves to express a "complementary slackness condition" (2.3) and an optimality condition (2.5). Moreover, both theorems use a similar condition to establish the continuity of $\dot{\overline{u}}(\theta)$ in the solution to $(\mathcal{P})$. Yet, the present Theorem requires substantially weaker assumptions on the primitive function s. To import results in Seierstadt and Sydsaeter (1987, Theorems 2 and 3, Chap. 5), who provide necessary and sufficient conditions satisfied by an optimal path for an optimal control problem with pure state constraints, Jullien (2000) requires the stronger hypothesis that $s(\theta,v)$ is twice continuously differentiable in v. Our contribution is to demonstrate the force and the broader validity of these conditions for problems with integrands that are only upper semi-continuous.

As in Jullien (2000), the mesure $\mu(d\theta)$ stems for the shadow cost of the participation constraint (2.1) around θ. The adjoint function $\overline{\gamma}(\theta)$ can thus be interpreted as the sum of these shadow costs for all inframarginal types. It is thus non-decreasing and constant on any open interval where the participation constraint is slack. Replacing the right-hand side of (2.1) uniformly by $\epsilon<0$ for all $\tilde{\theta}\le\theta$ would relax the optimization problem and increase its value by $\overline{\gamma}(\theta)\epsilon$.

The adjoint function $\overline{\gamma}$ so constructed is right-continuous. Since the probability measure μ may have mass points where the participation constraint begins to bind, $\overline{\gamma}(\theta)$ may have upward jumps at such points. This possibility may only arise at a countable number of points since an increasing function is almost everywhere differentiable. Hence, the participation constraint is only binding on a countable number of segments; a convenient fact that allows one to isolate the relevant segments.

Smoothness. We now investigate under which conditions the optimal solution remains continuously differentiable. In principal-agent models, it often means that the underlying control, say output, is continuous.

PROPOSITION 1 *If*

$$\mathcal{V}(\theta, \sigma) \equiv \arg\max_{v \in \mathbb{R}} \; s(\theta, v) + (F(\theta) - \sigma)v \tag{2.6}$$

is single-valued and continuous over the domain $(\theta, \sigma) \in \Theta \times [0, 1]$*, then the solution* $\overline{u}$ *to* $(\mathcal{P})$ *is continuously differentiable.*

That $\mathcal{V}(\theta, \sigma)$ is single-valued and continuous is implied by the strict concavity of $s(\theta, \cdot)$. It is also implied by the weaker condition in Jullien (2000, Assumption 2) that $s(\theta, v) - (\sigma - F(\theta))v$ is strictly quasi-concave in v for any $\sigma \in [0, 1]$. Together with the stronger hypothesis that $s(\theta, v)$ is twice continuously differentiable in v, Lemma 7 in Jullien (2000, p. 32) then provides a smooth version of (2.5) by means of a first-order condition, namely

$$F(\theta) - \overline{\gamma}(\theta) + \frac{\partial s}{\partial v}(\theta, \dot{\overline{u}}(\theta)) = 0 \text{ a.e.}.$$ [8]

Proposition 1 is more general since it allows for the possibility that $s(\theta, v)$ fails to be continuous at its maximum.

That $\overline{u}$ is continuously differentiable captures the fact that often in applications the optimal control, say output in a principal-agent context, is itself continuous. Examples abound in the literature where continuity is not optimal if the virtual surplus is not strictly concave. Models of bypass and regulation under the threat of entry due to Caillaud (1990), Laffont and Tirole (1990) and Currien, Jullien and Rey (1995), as an illustration, feature discontinuities in the optimal control because the bypass technology entails a fixed cost. In those papers, the authors have dealt with discontinuities "by hand" using details specific to the setting. Specifically, the authors first anticipate where the binding participation constraints lie (sometimes a complex task in itself), they next reconstruct the agent's profile of information rent and, finally, the principal's non-concave virtual surplus, before proceeding to optimization. This guess-and-check approach, although useful, provides little guidance when non-concavities arise precisely for types where the participation constraint binds.[9]

Section 3.1 analyzes a nonlinear pricing with a fixed cost of building extra capacity to accommodate customers' needs that illustrates such intricacy. Avoiding paying this fixed cost requires giving up extra rent to loyal customers since otherwise, they would switch to a competitive fringe. The fixed-cost discontinuity thus determines the subset of types for which the participation constraint is binding.

Continuing on that front, Section 3.2 presents a model of nonlinear prices under the threat of competition by a fringe. A key property of this model is that the discontinuity in the surplus function is now inherited from properties of the agent's information rent when the latter can split his purchases between the incumbent and the fringe. The discontinuity is thus endogenously derived from demand considerations.

[8] Galbraith and Vinter (2004) provide also alternative conditions ensuring Lipschitz-continuity of the optimal control.

[9] Another source of possible discontinuities comes when the types distribution has mass points; an assumption that we have ruled out for simplicity. See Lewis and Sappington (1993) and Cremer, Khalil and Rochet (1998) for applications to information gathering where a mass point of agents remains uninformed; Hellwig (2010) provides a more general treatment.

We should warn readers that, in non-smooth contexts, the solution $\overline{u}$ is unlikely to remain continuously differentiable. Jumps in the underlying control occur. This feature is nicely illustrated in the analysis of the discontinuous equilibria of a common agency game that is developed in Section 3.3 below.

A More Primitive Statement of the Problem. Principal-agent problems with type-dependent participation constraints as studied in the pathbreaking works of Lewis and Sappington (1989a), Maggi and Rodriguez (1995) and Jullien (2000) are often expressed under the form $(\mathcal{P}')$ below so as to make the nature of the agent's outside option, $\hat{U}(\theta)$, and its associated participation constraint more explicit:

$$(\mathcal{P}') : \text{Maximize}_{U\in AC(\Theta,\mathbb{R}),q} \int_{\underline{\theta}}^{\overline{\theta}} \left(\tilde{s}(\theta, q(\theta)) - U(\theta)\right) f(\theta)d\theta$$
$$\text{subject to } \dot{U}(\theta) = g(q(\theta), \theta) \text{ a.e., and } U(\theta) \geq \hat{U}(\theta) \text{ for all } \theta \in \Theta.$$

The control variable q, which is assumed to be measurable, is generally interpreted as a quantity vector that belongs to a feasible set $\mathcal{Q} \subseteq \mathbb{R}^k$.[10] The primitive surplus function $\tilde{s}(\theta, q)$ is defined over $\mathcal{Q}$.[11] The differential equation that defines $\dot{U}(\theta)$ immediately follows from the well-known envelope condition for incentive compatibility. The participation constraint $U(\theta) \geq \hat{U}(\theta)$ allows that the agent's outside option may vary by type.

The program $(\mathcal{P}')$ can easily be transformed into the canonical program $(\mathcal{P})$ we explore if $\hat{U}(\cdot)$ is differentiable almost everywhere. To illustrate, set $u(\theta) = U(\theta) - \hat{U}(\theta)$ and set

$$s(\theta, v) = \max_{q\in\mathcal{Q}} \left\{\tilde{s}(\theta, q)f(\theta) \text{ s.t. } v = g(q, \theta) - \dot{\hat{U}}(\theta)\right\}.$$

This reduction is particularly easy when g is itself a bijection between q and $\dot{u}$ for all θ, which implies that the control q can be expressed as a function of (v, θ), namely $q(\theta) = g_q^{-1}(v + \dot{\hat{U}}(\theta), \theta)$. In that case, substitution yields

$$s(\theta, v) = \tilde{s}(\theta, g_q^{-1}(v + \dot{\hat{U}}(\theta), \theta)))f(\theta).$$

While it is possible that the function $s(\theta, v)$, obtained as a maximum over all controls that generate the same derivative $\dot{u}(\theta)$, may be *"smoother"* than $\tilde{s}(\theta, q)$, it will typically fall short of twice continuous differentiability as required in Jullien (2000).

3. APPLICATIONS

This section shows the broad applicability of our approach by highlighting applications that requires non-smooth analysis. Our first application (Section 3.1) deals with nonlinear pricing of a digital good under the threat of competition by a low-quality fringe. Because the provider of a high-quality good needs to build capacity for extra services, the cost function is discontinuous. We show that, under the threat of competition, a monopolist may choose to leave rents to avoid building the extra capacity; thereby stressing how the

[10] Incentive compatibility requires on top that some monotonicity conditions also hold (for instance, $g(q(\theta), \theta)$ may have to be non-decreasing in θ). Those conditions can often be handled by extension of the state variable and adding a new non-negative control.

[11] Note that the domain of definition of q can be easily included into the objective to fit with the formalism of Theorem 1 if we set $\tilde{s}(\theta, q(\theta)) = -\infty$ for $q \notin \mathcal{Q}$.

cost discontinuity interacts with the characterization of the set of types with a binding participation constraint.

Our second example, developed in Section 3.2, is a different model of nonlinear pricing where a buyer may split his purchases between an incumbent firm and a fringe. The fringe has limited capacity but sells a perfect substitute to the incumbent's product. We show that this possibility introduces a discontinuity in the surplus function simply because the buyer's rent has a different slope whether he purchases or not from the fringe.

Finally, in Section 3.3, we turn to a model of intrinsic common agency with countervailing incentives, generalizing a well-known model due to Lewis and Sappington (1989b). We show existence of different classes of discontinuous equilibria that solve non-smooth self-generating problems. Those equilibria are obtained with non-smooth contracts and output discontinuities may arise at points where the participation constraints is binding.

3.1. *Nonlinear Pricing by a Digital Firm*

Pricing for digital products and other technology-based services is notoriously complex. Variable increases in demand for these products are fulfilled by adding blocks of computing or network infrastructure. Although marginal costs are essentially zero, additional blocks of service typically require significant fixed costs. Such fixed costs introduce discontinuities for which the tools developed above can be usefully applied.

Because of this technical difficulty, the screening literature in such environment is sparse. Spulber (1993) and Thomas (2001) have studied nonlinear pricing when a monopolist faces a fixed capacity constraint in more general contexts. On top of the usual information distortions, the optimal consumption profile depends on the shadow cost of the capacity constraint; a result we quickly review below. For digital products, Huang and Sundararajan (2011) follow a similar path and determine the set of types who are bunched at the capacity level. Yet, nothing is known on how a monopolist would design nonlinear pricing when facing a competitive fringe. To illustrate, suppose that there is no cost of building extra services, so the logic of continuous screening models under the threat of competition applies (e.g., Champsaur and Rochet, 1989; Stole, 1995). The ability of the monopolist to screen consumers and extract their information rent by reducing consumption levels is limited by their possibility to switch to the fringe. With a fixed cost for additional capacity, a tension appears between, on the one hand, charging high prices to extract information rents and avoid additional capacity costs and, on the other hand, inducing more consumers to switch from the fringe.

Technology and Demand. A monopolist sells a digital product. To supply more than one unit, an extra block of capacity must be built. The firm's cost function has thus an upward jump discontinuity at $q = 1^+$, namely $C(q) = k\delta_{q>1}$. Let the set of feasible outputs be $\mathcal{Q} = [0, \overline{Q}]$ where $\overline{Q}$ is finite but sufficiently large to ensure interior solutions under all the circumstances we consider below.

On the demand side, there is one consumer with a valuation for q units of services equal to

$$(S - \theta)q - \frac{q^2}{2}. \tag{3.1}$$

The demand shock θ is a non-observable heterogeneity parameter, uniformly distributed on $\Theta = [\underline{\theta}, \overline{\theta}]$ with $\overline{\theta} - \underline{\theta} = 1$. The parameter S reflects the quality of the service.

Monopoly. Under complete information, the monopoly extracts all consumer's surplus. Because of the cost discontinuity, the bilateral surplus so obtained is

$$(S-\theta)q - \frac{q^2}{2} - k\delta_{q>1}.$$

This surplus function is also discontinuous with a downward jump at $q = 1^+$. Its concave envelope thus is a linear segment for $q \in [1, 1+\sqrt{2k}]$. Following the approach for solving non-smooth problems that is outlined in Appendix A, we find that the complete information consumption level $q^{fb}(\theta)$ satisfies

$$q^{fb}(\theta) = \begin{cases} 1 & \text{if } \theta \in [S-1-\sqrt{2k}, S-1] \\ S-\theta & \text{otherwise.} \end{cases}$$

Intermediate types in $[S-1-\sqrt{2k}, S-1]$ are thus bunched together at $q^{fb}(\theta) = 1$.[12] Consumption is discontinuous and jumps downwards at $\theta_0 = (S-1-\sqrt{2k})^-$.

We now turn to the case where θ is private information for the consumer. Let $U(\theta)$ be the *informational rent* (indirect utility) when offered a nonlinear price $T(q)$[13]

$$U(\theta) \equiv \max_{q\in\mathcal{Q}} \, (S-\theta)q - \frac{q^2}{2} - T(q).$$

$U(\theta)$ is absolutely continuous and convex as the minimum of linear functions of θ. It is thus almost everywhere differentiable with a derivative (wherever it exists) given by the envelope condition

$$\dot{U}(\theta) = -q(\theta).^{14} \tag{3.2}$$

This envelope condition in tandem with the convexity of U are necessary and sufficient conditions for the implementability of $q(\cdot)$ using the tariff $T(q(\theta)) = (S-\theta)q(\theta) - \frac{1}{2}q(\theta)^2 - U(\theta)$.[15] For the pure monopoly setting, standard techniques can be used to derive the optimal consumption levels under asymmetric information. Knowing that the participation constraint $U(\theta) \geq 0$ necessarily binds for the worst type $\overline{\theta}$, it is routine to reconstruct and maximize pointwise a virtual surplus that is expressed as

$$(S - 2\theta + \underline{\theta})q - \frac{q^2}{2} - k\delta_{q>1}.$$

Taking again the concave envelope of this function, we replicate our earlier findings *mutatis mutandis* to get the following expression of the monopoly solution as

$$q^m(\theta) = q^{fb}(2\theta - \underline{\theta}).^{16} \tag{3.3}$$

[12] For simplicity, we assume that $\underline{\theta} < S-1-\sqrt{2k} < S-1 < \overline{\theta}$, an assumption that ensures that a strict subset of types chooses $q=1$.

[13] From the Taxation Principle (Rochet, 1987), focusing on nonlinear tariffs is without loss of generality in the space of non-stochastic mechanisms.

[15] Because $U(\theta)$ is convex, the consumption level $q(\theta)$ is a non-increasing selection within the best-response correspondence $\arg\max_{q\in\mathcal{Q}} \, (S-\theta)q - \frac{q^2}{2} - T(q)$. This monotonicity condition is first neglected and then verified ex post for the solution to the relaxed program.

Because of the non-concavity of the virtual surplus, the monopoly quantity is again discontinuous. As this example clearly shows, sometimes standard techniques may suffice. Unfortunately, it is no longer the case when participation constraints are more complex and the discontinuity of the surplus function precisely determines where these constraints are binding. This is the scenario we now investigate.

COMPETITIVE FRINGE. We now assume that a competitive fringe, without capacity constraint, can supply a low-quality version of digital good at zero marginal cost. Specifically, we assume that the intercept of the inverse demand function for the low-quality good is $\hat{S} < S$ with $\Delta = S - \hat{S} > 0$, and the consumption of the two versions of the good is mutually exclusive. A consumer who chooses the low-quality good would thus buy $\hat{q}(\theta) = \hat{S} - \theta$ units and obtain a net surplus of

$$\hat{U}(\theta) = \max_{q \in \mathcal{Q}} (\hat{S} - \theta)q - \frac{q^2}{2} \equiv \frac{(\hat{S} - \theta)^2}{2}.$$[17]

Adopting our canonical form, we set $u(\theta) = U(\theta) - \hat{U}(\theta)$. The customer chooses to buy exclusively from the monopolist when the participation constraint (2.1) holds. With this notation, the envelope condition for incentive compatibility (3.2) becomes

$$\text{(3.4)} \qquad \dot{u}(\theta) = \hat{q}(\theta) - q(\theta), \text{ a.e.}$$

Expressed in terms of $v = \dot{u}$, the surplus function becomes

$$s(\theta, v) = (S - \theta)(\hat{q}(\theta) - v) - \frac{1}{2}(\hat{q}(\theta) - v)^2 - k\delta_{\hat{q}(\theta) - v > 1}.$$

Note that the downward discontinuity implies that the concave envelope of $s(\theta, v)$ is linear over some range. We are now equipped to derive the optimal solution.

Smooth case: $k = 0$. To emphasize the difference between our techniques and those used in the smooth scenario studied by Jullien (2000), we start with the familiar case where there is no capacity constraint. The surplus function $s(\theta, v)$ is therefore smooth and strictly concave in v. Our Theorem 1 then takes he same form as Theorem 1 in Jullien (2000). The adjoint function $\overline{\gamma}_0(\theta) = \int_{\underline{\theta}}^{\theta} \mu_0(ds)$ and the optimal consumption level $\overline{q}_0(\theta) = \hat{q}(\theta) - \dot{\overline{u}}_0(\theta)$ must satisfy the following first-order condition for a smooth problem:

$$\text{(3.5)} \qquad \overline{\gamma}_0(\theta) - \theta + \underline{\theta} = \frac{\partial s}{\partial v}(\theta, \dot{\overline{u}}_0(\theta)) \Leftrightarrow \overline{q}_0(\theta) = S - 2\theta + \underline{\theta} + \overline{\gamma}_0(\theta).$$

When there is competition with a fringe, strong downwards distortions of consumption for the lowest types are no longer so attractive for the monopolist since these types could switch to the fringe.

PROPOSITION 2 *Suppose that $k = 0$ and let $\tilde{\theta} = \underline{\theta} + \Delta < \overline{\theta}$. The following optimal consumption levels and adjoint functions satisfy the necessary and sufficient conditions for optimality of Theorem 1:*

$$\text{(3.6)} \qquad \overline{q}_0(\theta) = \max\{q^m(\theta); \hat{q}(\theta)\},$$

[16] We assume that $q^{fb}(2\overline{\theta} - \underline{\theta}) > 0$ to ensure consumption is positive for all types.

[17] We assume $\underline{\theta} < \hat{S} - 1 < \overline{\theta}$ to ensure that choice of the competitive fringe product results in positive consumption.

$$(3.7) \qquad \overline{\gamma}_0(\theta) = \max\{\theta - \tilde{\theta}; 0\}.$$

The participation constraint (2.1) is binding on $[\tilde{\theta}, \overline{\theta}]$*, slack elsewhere.*

The presence of the fringe limits the ability of the monopolist to price discriminate. Screening distortions are less pronounced than in a pure monopoly setting, a result that echoes findings in Champsaur and Rochet (1989), Stole (1995, 2003), and Calzolari and DeNicolo (2015) among many others.

The intuition for the shape of the solution can be borrowed from the work of Maggi and Rodriguez (1995). In a smooth environment, those authors have proposed a typology of the various patterns of binding participation constraints that may arise at optimal contracts in structured environments. They argue that whether participation constraints bind on a whole interval or only at the extreme of the types set depends on the relative convexity of the rent profile that the monopolist would like to implement vis-à-vis the reservation payoff. Due to the lower quality of the alternative that it offers, buying from the fringe provides a type-dependent reservation payoff that is less convex than what the monopolist would like to implement without competition. The participation constraint is then binding on a whole interval (here including the type with the lowest valuation).

Observe that the measure μ_0 is absolutely continuous with respect to the Lebesgue measure on $[\tilde{\theta}, \overline{\theta})$ but it has a mass point at $\overline{\theta}$, namely $\mu(\{\overline{\theta}\}) = \Delta < 1$. Finally, $\tilde{\theta}$ is defined by a "smooth-pasting condition" $q^m(\tilde{\theta}) = \hat{q}(\tilde{\theta})$; output is continuous at $\tilde{\theta}$ and $\overline{u}_0$ is differentiable at that point.

Discontinuity: $k > 0$*.* The surplus function $s(\theta, v)$ has a downward jump. Theorem 1 in Jullien (2000) now no longer applies while Theorem 1 still does. The optimality condition (2.5) must thus be expressed in terms of a subdifferential:

$$(3.8) \qquad \overline{\gamma}(\theta) - (\theta - \underline{\theta}) \in \partial_v \overline{\text{co}}(s)(\theta, \dot{\overline{u}}(\theta)).$$

To streamline exposition and limit the number of possible cases to study, we impose the condition

$$(3.9) \qquad \hat{q}(\tilde{\theta}) > 1 > \hat{q}(\overline{\theta}).$$

This condition ensures that the monopolist has to incur some extra cost k if it offers the same consumption levels everywhere on a right-neighborhood of $\tilde{\theta}$. To avoid paying such cost, the monopolist would like to have a bunch of types beyond $\tilde{\theta}$ still consuming only one unit. Yet, this objective conflicts with the possibility that those types would switch to the fringe. To avoid paying k, the monopolist must redistribute part of these gains to customers under the form of extra information rent beyond what they receive from the fringe. The market is now segmented with four different connected subsets of types. This pattern corresponds to an adjoint function $\overline{\gamma}$ which is obtained by slightly modifying $\overline{\gamma}_0$ (as explained in (3.11) below) according to our economic intuition.

Proposition 3 *Suppose that* $k > 0$ *but small enough, and that Condition (3.9) holds. The following optimal consumption levels and adjoint functions satisfy the necessary and* *sufficient conditions for optimality of Theorem 1:*

$$(3.10)\quad \overline{q}(\theta)=\begin{cases} q^m(\theta) & \text{if } \theta\in[\underline{\theta},\tilde{\theta}),\\ \hat{q}(\theta) & \text{if } \theta\in[\tilde{\theta},\theta_1) \text{ and } \theta\in[\theta_3,\overline{\theta}),\\ 1 & \text{if } \theta\in[\theta_1,\theta_2),\\ q^m(\theta)+\sqrt{k}+\theta_1-\tilde{\theta} & \text{if } \theta\in[\theta_2,\theta_3);\end{cases}$$

$$(3.11)\quad \overline{\gamma}(\theta)=\begin{cases} 0 & \text{if } \theta\in[\underline{\theta},\tilde{\theta}),\\ \theta-\tilde{\theta} & \text{if } \theta\in[\tilde{\theta},\theta_1) \text{ and } \theta\in[\theta_3,\overline{\theta}),\\ \sqrt{k}+\theta_1-\tilde{\theta} & \text{if } \theta\in[\theta_1,\theta_2);\end{cases}$$

where θ_1, θ_2, θ_3 are such that $\tilde{\theta}<\theta_1<\theta_2<\theta_3<\overline{\theta}$ and

$$(3.12)\quad \hat{q}(\theta_1)=1+(\sqrt{2}-1)\sqrt{k}, \hat{q}(\theta_2)=1+\left(\frac{\sqrt{2}}{2}-1\right)\sqrt{k}, \hat{q}(\theta_3)=1+(\sqrt{2}-2)\sqrt{k}.$$

The participation constraint (2.1) is binding on $[\tilde{\theta},\theta_1]\cup[\theta_3,\overline{\theta}]$ and slack elsewhere.

The optimal consumption remains equal to $\overline{q}_0(\theta)=\max\{q^m(\theta),\hat{q}(\theta)\}$ for the bottom interval $[\underline{\theta},\theta_1)$, i.e., for those types with the highest consumption levels for which the incumbent finds it worth to incur the extra capacity. There is still a downward jump in consumption at θ_1. Types in the right-neighborhood $[\theta_1,\theta_2)$ are all bunched together and consume only one unit from the monopolist although they would like to get more from the fringe. To save on the cost of the extra capacity, the monopolist gives to those types a price discount to compensate for the constrained consumption and, as a result, those types get a payoff above what they would get with the fringe. Yet, types with a lower valuation consume less than one unit of service up to the point where the lowest valuations in the interval $[\theta_3,\overline{\theta}]$ again consume the same amount as with the fringe.

Again, some intuition for this pattern of rents can be grasped from the typology offered by Maggi and Rodriguez (1995) although here the set of types with binding outside options is significantly more complex than in their earlier study. Indeed, the fixed cost now forces a constant consumption at $q=1$ for a bunch of intermediate types. The rent profile that the monopolist would like to implement becomes less convex than the outside option for such types. We know from Maggi and Rodriguez (1995) that, in such contexts, the participation constraint may be slack for intermediate types. In contrast with Maggi and Rodriguez (1995) who focus on the rather pure cases where the reservation payoff is uniformly more (or less) convex than the monopoly's profile, the comparison is here reversed for the very highest types. In sum, the participation constraint is now slack for such intermediate types but it now binds on two intervals around that area.

The measure μ is again absolutely continuous with respect to the Lebesgue measure on the interior of the set of the types, namely, $[\tilde{\theta},\theta_1)\cup[\theta,\overline{\theta})$, where (2.1) is binding. It has again a mass point at $\overline{\theta}$ still given by $\mu(\{\overline{\theta}\})=\Delta$. Now, to cope with the consumption discontinuity at θ_1^- and the fact that a right-neighborhood of that type receives rent beyond what it gets from the fringe, μ has also another charge at that discontinuity point θ_1, namely $\mu(\{\theta_1\})=\sqrt{k}$.

3.2. *Nonlinear Pricing with Split Purchases*

The previous model can be slightly modified to cover other competitive scenarios. Instead of exclusively buying from one seller as in Section 3.1, consumers may have access to a competitive fringe which sells a perfect substitute to the incumbent's product and they might split accordingly their purchases. This scenario is relevant for a number of important contracting environments where previous analysis have instead been developed under the assumption of exclusive contracting. To illustrate, regulated monopolies can be subject to competition by entrants with alternative bypass technologies. Earlier works (Caillaud, 1990; Laffont and Tirole, 1990; Currien, Jullien and Rey, 1995), ruled out the possibility that consumers may also purchase from entrants. Recent technological advances in telecommunication services, broadcasting and internet practices suggest that the theory of bypass should now also account for the possibility of split purchases between an incumbent operator and entrants on some segments of its services.

Incumbent manufacturers may also design exclusive contracts with their retailers to prevent their representing alternative suppliers. In Aghion and Bolton (1987)'s theory of contracts as a barrier to entry, retailers buy entirely from the entrant if they switch thanks to an assumption of unit demand while, more recently, Calzolari and DeNicolo (2015) have analyzed how such exclusivity clause endogenously arises at equilibrium.[18] Yet, the existing Antitrust bans on exclusivity arrangements that prevail in some countries make it important to study how vertical contracts are modified when retailers may also sell products from rivals.

Lastly, competition in financial and insurance markets has often been modeled, following Rotschild and Stiglitz (1976) in a context with exclusive contracts while often exclusivity clauses can hardly be enforced as noticed by Attar et al. (2011). A trader may buy a financial asset from one broker and then turn to another one to complete his portfolio.

In these settings, standard screening distortions are limited by the buyer's ability to switch to the fringe. As we will see in the model below, such possibility introduces discontinuities in the virtual surplus. Instead of coming from the cost function as in Section 3.1, discontinuities come from the demand side and, more precisely, from how the buyer responds to changes in contractual terms with a dominant firm.

Model. A dominant firm produces a good at constant marginal cost $c \geq 0$. On the demand side, consumers still have preferences given by (3.1) where θ, uniformly distributed on $[\underline{\theta}, \overline{\theta}]$ (with $\overline{\theta} - \underline{\theta} = 1$), is again private information. Importing *mutatis mutandis* our earlier findings of Section 3.1, the monopoly solution would consist in offering

$$q^m(\theta) = S - 2\theta + \underline{\theta} - c.$$

A competitive fringe sells a perfect substitute to the incumbent's product at price $p > c$ (let $\Delta = p - c > 0$). Entry would thus be inefficient in a first-best world. We denote by q_0 the quantity bought from the fringe. The fringe has a capacity constraint $k > 0$ and thus $q_0 \in \mathcal{Q}_0 = [0, k]$. The key difference with Section 3.1 is that consumers can now always purchase from the fringe if the incumbent charges a price which is too high. Formally, if

[18] Choné and Linnemer (2016) and Martimort, Pouyet and Stole (2018) also investigate the possibility of split purchases in related contexts.

a consumer with type θ chooses q units from the incumbent, his indirect utility function becomes

$$v(\theta, q) = \max_{q_0 \in \mathcal{Q}_0} (S - \theta)(q + q_0) - \frac{(q + q_0)^2}{2} - pq_0.$$

Define the highest level of consumption from the incumbent firm which induces consumption of k from the competitive fringe as

$$\hat{q}(\theta) = S - \theta - p - k,$$

which we assume to be positive. For $q \leq \hat{q}(\theta)$, the consumer purchases $q_0(\theta) = k$ units from the competitive fringe, for $q \geq \hat{q}(\theta) + k$, the consumer purchases $q_0(\theta) = 0$ from the competitive fringe. Straightforward computations yield

$$v(\theta, q) = \begin{cases} (S - \theta)(q + k) - \frac{(q+k)^2}{2} - pk & \text{if } q \leq \hat{q}(\theta), \\ \frac{(\hat{q}(\theta)+k)^2}{2} + \Delta q & \text{if } q \in [\hat{q}(\theta), \hat{q}(\theta) + k], \\ (S - \theta)q - \frac{q^2}{2} & \text{if } q \geq \hat{q}(\theta) + k. \end{cases}$$ [19]

In particular, had the consumer not bought from the incumbent, i.e., $q = 0$, he would consume up to capacity from the fringe and get a payoff worth

$$\hat{U}(\theta) \equiv v(\theta, 0) = \hat{q}(\theta)k + \frac{k^2}{2}.$$ [20]

For future reference, we will assume that

$$q^m(\overline{\theta}) > k. \tag{3.13}$$

This condition ensures that the incumbent firm still wants to serve the type with the lowest possible demand $\overline{\theta}$ even when that type consumes up to the fringe's capacity.

A DISCONTINUOUS SURPLUS FUNCTION. When he instead buys q units from the dominant firm which charges a nonlinear price $T(q)$, a consumer with type θ gets

$$U(\theta) = \max_{q \in \mathcal{Q}} v(\theta, q) - T(q).$$

To import our general formalism, we again introduce the state variable $u(\theta) = U(\theta) - \hat{U}(\theta)$. The participation constraint takes its usual form (2.1) while the standard envelope condition for incentive compatibility becomes

$$\dot{u}(\theta) = \begin{cases} -q(\theta) & \text{if } q(\theta) < \hat{q}(\theta), \\ -\hat{q}(\theta) & \text{if } q(\theta) \in [\hat{q}(\theta), \hat{q}(\theta) + k], \\ -q(\theta) + k & \text{if } q(\theta) > \hat{q}(\theta) + k. \end{cases} \tag{3.14}$$

[19] To better understand the shape of the indirect utility function $v(\theta, q)$, we may think of the consumer as only choosing in between consuming 0 or k units from the fringe. His indirect utility function would be the (non-concave) maximum of two concave functions because of a discontinuous jump in the corresponding choice. The possibility of consuming any arbitrary amount within the interval $\mathcal{Q}_0$ concavifies this indirect utility function and introduces a linear piece for intermediate consumption levels from the dominant firm.

[20] Since $v(\overline{\theta}) < 0$, this right-hand side remains positive.

This relationship actually shows that $\dot{u}$, taken as a function of q, is not bijective.[21] This phenomenon captures the role of competition from the fringe. If, to facilitate rent extraction, the dominant firm is willing to slightly reduce the buyer's consumption $q(\theta)$ when it takes values in $(\hat{q}(\theta), \hat{q}(\theta)+k)$, the buyer can always consume more from the fringe.

The dominant firm wants to maximize the bilateral surplus of contracting with the buyer net of the rent this buyer gets above his reservation payoff obtained when purchasing exclusively from the fringe. This net surplus is thus

$$\tilde{s}(\theta, q(\theta)) - u(\theta) = v(\theta, q(\theta)) - v(\theta, 0) - cq(\theta) - u(\theta).$$

Replacing $q(\theta)$ by its expression in terms of $\dot{u}(\theta)$ (from (3.14), when $\dot{u}(\theta) \neq -\hat{q}\theta)$) whenever such inversion is possible leads, straightforward computations (detailed in the Proof of Proposition 4) allow us to express the gross surplus as simply

$$s(\theta, v) = (-\hat{q}(\theta) - \Delta)v - \frac{v^2}{2} + k\Delta\delta_{v \leq -\hat{q}(\theta)}. \tag{3.15}$$

This surplus function is upper semi-continuous, has a downward jump discontinuity at $v = -\hat{q}(\theta)$, and is maximized at $v_1(\theta) = -\hat{q}(\theta) - \Delta < -\hat{q}(\theta)$. This downward jump captures the fact that, when consumption from the dominant firm is too low (which means $v(\theta) = -\dot{u}(\theta)$), the bilateral surplus between the dominant firm and the customer diminishes by $k\Delta$, reflecting the opportunity cost of purchasing k units from the fringe.

Proposition 4 *Let $\tilde{\theta} = \underline{\theta} + \Delta$ and $\tilde{\theta}_0 = \tilde{\theta} + \sqrt{2k\Delta}$. Suppose that $\tilde{\theta}_0 \leq \overline{\theta}$ and (3.13) holds. The following optimal consumption levels and adjoint functions satisfy the necessary and sufficient conditions for optimality of Theorem 1:*

$$\overline{q}(\theta) = \begin{cases} q^m(\theta) & \text{if } \theta \in [\underline{\theta}, \tilde{\theta}], \\ \text{arbitrary} \in [\hat{q}(\theta), \hat{q}(\theta)+k] & \text{if } \theta \in (\tilde{\theta}, \tilde{\theta}_0), \\ q^m(\theta) - k & \text{if } \theta \in [\tilde{\theta}_0, \overline{\theta}]. \end{cases} \tag{3.16}$$

$\overline{\gamma}(\theta)$ has a mass point at $\overline{\theta}$, $\mu(\{\overline{\theta}\}) = 1$.

The participation constraint (2.1) is binding at $\overline{\theta}$ only.

Types who are the most eager to buy (i.e., $\theta \in [\underline{\theta}, \tilde{\theta})$) are not tempted to switch to the fringe. They consume the same downward distorted monopoly quantity $q^m(\theta)$ as what the incumbent firm would offer if the fringe was absent. This quantity nevertheless remains large enough to make it unattractive for the buyer to switch. For intermediate types in $[\tilde{\theta}, \tilde{\theta}_0)$, the monopolist loses some of his ability to screen. Any attempt to reduce consumption from the incumbent is entirely compensated by the buyer consuming more from the fringe. The incumbent's sales are indeterminate but screening distortions are reduced to avoid switching. Finally, types in $[\tilde{\theta}_0, \overline{\theta}]$ react to the downward screening distortion offered by the incumbent by consuming up to capacity from the fringe.

Although the discontinuity of $s(\theta, v)$ bears some similarity with that analyzed in Section 3.1, the characterization of the solution is rather different thanks to the different nature

[21] Carbajal and Ely (2016) present an interesting model of behavioral consumers having loss aversion that has similar features.

of the participation constraint. Over the range $(\tilde{\theta}, \tilde{\theta}_0)$, the solution entails $\overline{u}(\theta) = -\hat{q}(\theta)$ independently of the incumbent's sales; which explains the indeterminacy in those sales. Because of the downward discontinuity jump of $s(\theta, v)$ at $v = -\hat{q}(\theta)$, $\overline{co}(s)(\theta, v)$ has a kink at that point and entails a flat segment for $v \in [-\hat{q}(\tilde{\theta}_0), v_1(\tilde{\theta}_0)]$ where $v_1(\tilde{\theta}_0) = -\hat{q}(\tilde{\theta}_0) + \sqrt{2k\Delta}$. At type $\tilde{\theta}_0$, the incumbent is thus indifferent between choosing $\dot{\overline{u}}(\tilde{\theta}_0) = -\hat{q}(\tilde{\theta}_0)$ and moving up to $\dot{\overline{u}}(\tilde{\theta}_0) = v_1(\tilde{\theta}_0)$, reducing his sales to $q^m(\tilde{\theta}_0) - k$ and allowing the buyer to purchase from the fringe up to capacity.

3.3. *Common Agency with Countervailing Incentives*

The characterization of all equilibrium allocations of common agency games under asymmetric information is an important question. In such games, principals offer contingent transfer schedules to a common agent who, upon acceptance and based on his private information, chooses a decision on their behalf. Earlier works in the field have often restricted the analysis to the case where principals compete through differentiable schedules (Laffont and Tirole, 1990; Martimort and Semenov, 2008, Martimort and Stole, 2009, Calzolari and DeNicolo, 2013, among others). Of course, this restriction is akin to an equilibrium refinement and modelers might want to know if the economic predictions of the smooth equilibria so selected carry over to other non-smooth equilibria.

Recently, progresses on this front have been accomplished by Martimort, Semenov and Stole (2017) for intrinsic common agency games and Martimort and Stole (2018) for delegated common agency games.[22] The first of these papers proposes necessary and sufficient conditions for an allocation to arise at an equilibrium of an intrinsic common agency games when principals can offer upper semi-continuous tariffs. The analysis is based on two important steps. First, Martimort, Semenov and Stole (2017) observe that an intrinsic common agency game is an aggregate game (Martimort and Stole, 2012). In a quasi-linear environment, its equilibria are thus also solutions to a *self-generating problem* obtained by summing the objectives of all principals. Second, the pointwise maximization of the virtual surplus associated to this self-generating problem characterizes equilibrium decisions. Of course, the possibility that a given principal offers a discontinuous schedule makes this step of the analysis a non-smooth optimization problem. Computing best responses would *a priori* require the use of the control techniques we develop in this paper. Fortunately, the analysis in Martimort, Semenov and Stole (2017) is simplified by the fact that all principals agree on whoever type is the least efficient one. The participation constraint is thus known to be binding for this type. From there, it is easy to recover the rent profile, continue with the standard Myersonian integration by parts, compute the corresponding virtual surplus and proceed to its optimization.

Yet, in a number of contexts this direct approach may not suffice, especially when countervailing incentives arise and participation constraints may be binding on the interior of the types set. To illustrate, regulated firms may differ not only through marginal costs but also because of their fixed costs, low marginal costs implying also large upfront investment. Firms are thus torn in between their incentives to exaggerate either marginal or fixed costs as in Lewis and Sappington (1989a). Also, an employee may be endowed with some initial resources of a given size to produce a task on his principal's behalf.

[22]Common agency is intrinsic when the agent is forced to contract with all principals or not contract at all. Regulation is an example in order It is delegated when the agent keeps the right to select any subset of offers. This scenario is relevant for lobbying games.

Producing less than this capacity for the principal allows that agent to enjoy profits from unused resources on alternative markets as in Lewis and Sappington (1989b). We now investigate such model.

Model. An agent produces q units of good or services at a cost

$$\theta(q-k)$$

where $k>0$ is a fixed capacity. Any production $q>k$ is costly to produce while excess capacity can be sold on another market in case the requested quantity q is less than k. The benefits of doing so is in that case $\theta(k-q)$. Let $\mathcal{Q}=[0,\overline{Q}]$ be the set of possible outputs with $\overline{Q}$ being large enough to ensure interior solutions. For simplicity, we again assume that the agent's type θ is uniformly distributed on $\Theta=[\underline{\theta},\overline{\theta}]$ with $\overline{\theta}=\underline{\theta}+1$.

There are n principals contracting simultaneously with the agent. Common agency is intrinsic; the agent must accept all such contracts if he chooses to produce at all. His outside opportunity gives a payoff of zero. As an example, we may think of several regulatory agencies that indeed contract independently with a firm to regulate different aspects of its performances. Being active on the market requires to abide to all regulations. Principal P_i enjoys a gross surplus $S_i(q)$ when q units are produced with S_i being continuously differentiable, increasing and strictly concave. To regulate the agent, P_i may offer any upper-semi continuous nonlinear price $T_i(q)$ defined over $\mathcal{Q}$.[23] Her net payoff becomes

$$S_i(q)-T_i(q).$$

For future reference, we denote the aggregate surplus and aggregate nonlinear price (resp. aggregate nonlinear price of all principal except i) respectively as $\mathcal{S}(q)=\sum_{i=1}^n S_i(q)$ and $\mathcal{T}(q)=\sum_{i=1}^n T_i(q)$ (resp. $\mathcal{T}_{-i}(q)=\sum_{j\neq i}T_j(q)$).

For reasons that will appear clearer below, we will assume that

$$\underline{\theta}<\mathcal{S}'(k)<\overline{\theta}. \tag{3.17}$$

This condition simply means that it is efficient to produce more (resp. less) than the existing capacity k when costs are low (resp. high) enough. This requirement is thus the raison d'être for the countervailing incentives that arise in this context.

Equilibrium Characterization. When accepting an array of offers $\mathbf{T}=(T_1,...T_n)$, the agent gets a payoff $u(\theta)$ and chooses a quantity $q(\theta)$ defined as

$$u(\theta)=\max_{q\in\mathcal{Q}}\mathcal{T}(q)-\theta(q-k) \text{ and } q(\theta)=\arg\max_{q\in\mathcal{Q}}\mathcal{T}(q)-\theta(q-k) \tag{3.18}$$

As the maximum of linear functions of θ, $u(\theta)$ is convex, so it is a.e. differentiable and $q(\theta)$ is non-increasing. At any point of differentiability, the envelope condition for incentive compatibility thus writes as

$$\dot{u}(\theta)=-q(\theta)+k. \tag{3.19}$$

[23]From Peters (2001) and Martimort and Stole (2002), there is no loss of generality in considering this strategy space in our setting.

A (pure strategy) Nash equilibrium is an array $\overline{\mathbf{T}} = (\overline{T}_1, ..., \overline{T}_n)$ of nonlinear upper semi-continuous schedules which are best responses to each other. Importantly, intrinsic common agency games are actually aggregate games (Martimort and Stole, 2012). This property is key to get a compact description of Nash-equilibrium conditions. Indeed, each principal can always undo whatever aggregate nonlinear price $\overline{\mathcal{T}_{-i}}(q)$ was offered by others at equilibrium and behave as if he was himself offering the aggregate price $\mathcal{T}(q)$. At equilibrium, the so called *Aggregate Concurrence Principle*[24] applies and all principals end up with objectives that are aligned with those of a *surrogate principal*. That surrogate principal would indeed solve the following *self-generating problem*:[25]

$$(\mathcal{P}): \quad \max_{q(\theta)\in\mathcal{Q}, u(\theta), T} \int_{\underline{\theta}}^{\overline{\theta}} [\mathcal{S}(q(\theta)) - \theta(q(\theta)-k) + (n-1)(\overline{\mathcal{T}}(q(\theta)) - \theta(q(\theta)-k)) - nu(\theta)]d\theta,$$

subject to (2.1), (3.18) and (3.19).

What is remarkable here is that the maximand depends on an endogenous equilibrium object, the aggregate schedule $\overline{\mathcal{T}}$. Its only known property is that it is upper semi-continuous. Differentiability would be nice because it would help to characterize equilibrium allocations by means of a first-order condition but imposing this requirement a priori too restrictive and may leave unexplored other non-smooth equilibrium allocations of interest.

To see how, we proceed as above and write this optimization in its canonical form by introducing the following surplus function:

$$(3.20) \quad ns(\theta, v) = \mathcal{S}(k - v) + \theta v + (n-1)(\overline{\mathcal{T}}(k - v) + \theta v).$$

The optimality conditions (2.4) and (2.5) should hold, with the latter being expressed as

$$(3.21) \quad \overline{\gamma}(\theta) - \theta + \underline{\theta} \in \partial\overline{co}(s)(\theta, \dot{\overline{u}}(\theta)) \text{ a.e. } \theta \in \Theta$$

where the adjoint function $\overline{\gamma}(\theta)$ is non-decreasing, right-continuous, constant on all intervals where (2.1) is slack and such that $\overline{\gamma}(\underline{\theta}) = 0$ and $\overline{\gamma}(\overline{\theta}) = 1$.

To this condition, we must also add the equilibrium requirement that the agent also chooses the same quantity than what the surrogate principal suggests. Still with our notations, this condition writes as:

$$(3.22) \quad 0 \in \partial_v \overline{co}(\overline{\mathcal{T}}(k - \dot{\overline{u}}(\theta)) + \theta\dot{\overline{u}}(\theta)).$$

• *The "almost" smooth equilibrium.* Had $\overline{\mathcal{T}}$ been continuously differentiable, at least almost everywhere, conditions (3.21) and (3.22) would respectively become

$$(3.23) \quad \mathcal{S}'(\overline{q}^m(\theta)) - \theta + (n-1)(\overline{\mathcal{T}}'(\overline{q}^m(\theta)) - \theta) = n(\theta - \underline{\theta} - \overline{\gamma}(\theta)),$$

$$(3.24) \quad \overline{\mathcal{T}}'(\overline{q}^m(\theta)) = \theta \text{ a.e. } \theta \in \Theta.$$

[24]Martimort and Stole (2012).
[25]See the Appendix for details.

Simplifying (3.23) then yields

$$\mathcal{S}'(\overline{q}^m(\theta)) = \theta + n(\theta - \underline{\theta} - \overline{\gamma}(\theta)) \text{ a.e. } \theta \in \Theta. \tag{3.25}$$

With smoothness, this condition follows from applying an envelope condition on the pointwise maximization of the surrogate principal's objective taking into account the fact that part of this objective, namely $(n-1)(\overline{\mathcal{T}}(q(\theta)) - \theta(q(\theta) - k))$ is already maximized by the agent's choice $q(\theta)$ at equilibrium. Below, we will investigate how to proceed without such envelope condition in non-smooth problems.

We now exhibit a pair $(\overline{u}^m(\theta), \overline{\gamma}^m(\theta))$ that satisfies all necessary conditions for optimality and invoke our sufficiency result to complete the analysis. Consider thus the following adjoint function

$$\overline{\gamma}^m(\theta) = \begin{cases} 0 & \text{if } \theta \in [\underline{\theta}, \theta_1^n], \\ \theta - \underline{\theta} + \frac{1}{n}(\theta - \mathcal{S}'(k)) & \text{if } \theta \in [\theta_1^n, \theta_2^n], \\ 1 & \text{if } \theta \in [\theta_2^n, \overline{\theta}] \end{cases} \tag{3.26}$$

where θ_1^n and θ_2^n, whose existence within $(\underline{\theta}, \overline{\theta})$ follows from (3.17), are respectively defined as

$$\mathcal{S}'(k) = \theta_1^n + n(\theta_1^n - \underline{\theta}) = \theta_2^n + n(\theta_2^n - \overline{\theta}).$$

Clearly, $\overline{\gamma}^m(\theta)$ as defined in (3.26) is strictly increasing on $[\theta_1^n, \theta_2^n]$, which is consistent with (2.1) being binding on this interval.

Observe also that $\overline{q}^m(\theta)$ is non-increasing as requested by incentive compatibility and that it satisfies

$$\overline{q}^m(\theta) \begin{cases} > k & \text{if } \theta \in [\underline{\theta}, \theta_1^n), \\ = k & \text{if } \theta \in [\theta_1^n, \theta_2^n], \\ < k & \text{if } \theta \in (\theta_2^n, \overline{\theta}]. \end{cases} \tag{3.27}$$

In other words, $\overline{q}$ entails some bunching on $[\theta_1^n, \theta_2^n]$. From (3.27), $\overline{u}^m(\theta)$ is decreasing (resp. increasing) and positive on $[\underline{\theta}, \theta_1^n)$ (resp. on $(\theta_2^n, \overline{\theta}]$) and zero on the bunching area $[\theta_1^n, \theta_2^n]$.

The last step to fully characterize the equilibrium is to find the aggregate nonlinear price that implements this rent profile. A simple duality argument[26] shows that the following schedule succeeds:

$$\overline{\mathcal{T}}^m(q) = \min_{\theta \in \Theta} \overline{u}^m(\theta) + \theta(q - k). \tag{3.28}$$

It is straightforward to check that $\overline{\mathcal{T}}^m$ so constructed is concave as the minimum of linear functions of q, continuously differentiable almost everywhere, except at $q = k$ where it only admits a subdifferential $\partial\overline{\mathcal{T}}^m(k) = [\theta_1^n, \theta_2^n]$.[27] All types in the bunching area $[\theta_1^n, \theta_2^n]$ choose output k at that kink.

[26]See Martimort, Semenov and Stole (2017) for more details.

[27]Our use of the word "subdifferential" for a concave function is somewhat abusive but it is also self-explanatory.

Finally, the equilibrium tariffs offered by each principal, $\overline{T}_j^m$, are then designed so that all principals' objectives are in fact aligned with those of the surrogate principal:

$$\frac{1}{n}\left(\mathcal{S}(q) - \overline{\mathcal{T}}^m(q)\right) = S_j(q) - \overline{T}_j^m(q) \quad \forall q \in \mathcal{Q}. \tag{3.29}$$

[28]

This condition implies that all principals agree on inducing the same equilibrium output even if each of them may entertain an individual deviation.

We collect the preceding arguments into the following proposition.

PROPOSITION 5 *In the continuous-allocation equilibrium of the intrinsic common agency game with countervailing incentives, the allocation* $\overline{q}^m(\theta) = k$ *for all* $\theta \in [\theta_1^n, \theta_2^n]$*, where* θ_1^n *and* θ_2^n *are characterized by,*

$$\mathcal{S}'(k) = \theta_1^n + n(\theta_1^n - \underline{\theta}) = \theta_2^n + n(\theta_2^n - \overline{\theta}),$$

and $\overline{q}^m(\theta)$ *satisfies*

$$\mathcal{S}'(\overline{q}^m(\theta)) - \theta = n(\theta - \underline{\theta}) > 0, \quad \forall \theta \in [0, \theta_1^n],$$

and

$$\mathcal{S}'(\overline{q}^m(\theta)) - \theta = n(\theta - \overline{\theta}) < 0, \quad \forall \theta \in [\theta_2^n, \overline{\theta}].$$

If we assume that the aggregate surplus of the principals is fixed independent of n, $\mathcal{S}$, and the principals have symmetric preferences with $S_i(q) = \frac{1}{n}\mathcal{S}(q)$, then we can write the boundary types for the bunching region as a function of n:

$$\theta_1^n = \frac{1}{n+1}\mathcal{S}'(k) + \frac{n}{n+1}\underline{\theta} \quad \in (\underline{\theta}, \overline{\theta}),$$

$$\theta_2^n = \frac{1}{n+1}\mathcal{S}'(k) + \frac{n}{n+1}\overline{\theta} \quad \in (\underline{\theta}, \overline{\theta}),$$

where the inclusions follow from (3.17). We conclude that θ_1^n is decreasing in n and θ_2^n is increasing in n, implying an every expanding region of bunching, as illustrated in Figure 3.3.

Following Lewis and Sappington (1989b), this equilibrium outcome can be interpreted as an inflexible rule over the bunching area. This bunching area increases with the number of principals. There is less flexibility as more non-cooperating principals are involved. The intuition is the following. In this common agency context, each principal has to pay the cost of inducing information revelation from the agent. The information rent is thus counted n times (a feature that appears in the expression of the surrogate principal's objective). Distortions are magnified. It in turn exacerbates countervailing incentives making it more attractive to implement an inflexible rule.[29]

[28]See the Appendix for details.

[29]This result is also reminiscent of the work of Mezzetti (1997). His work differs from ours because this author assumes that principals are selling differentiated products through their agent. His model is thus one of *private common agency,* in contrast with ours where all principals control the same variable; a setting with *public common agency.* Moreover, Mezzetti (1997) does not look for equilibria with discontinuous output schedules while the use of our methodology unveils their existence in our setting as we see below.

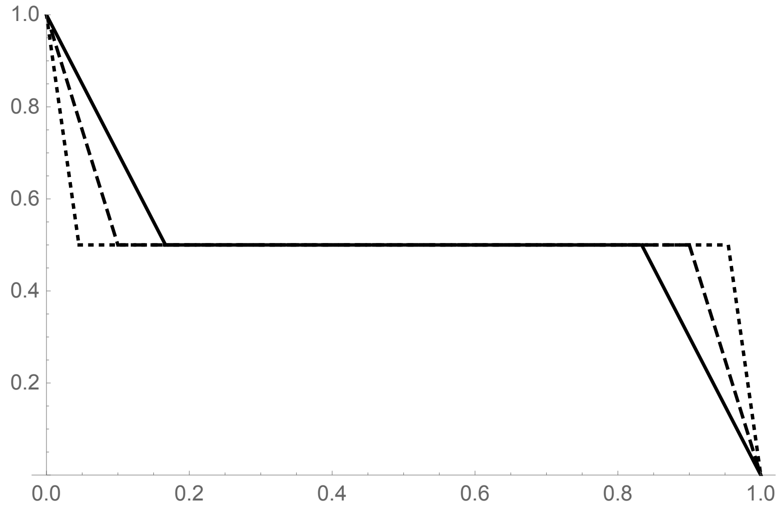

FIGURE 1.— Graph of equilibrium allocation for the case of $n = 2$ (solid line), $n = 4$ (dashed line), and $n = 10$ (doted line), assuming $\mathcal{S}(q) = \left(q - \frac{1}{2}q^2\right)$, $[\underline{\theta}, \overline{\theta}] = [0, 1]$, and $k = \frac{1}{2}$.

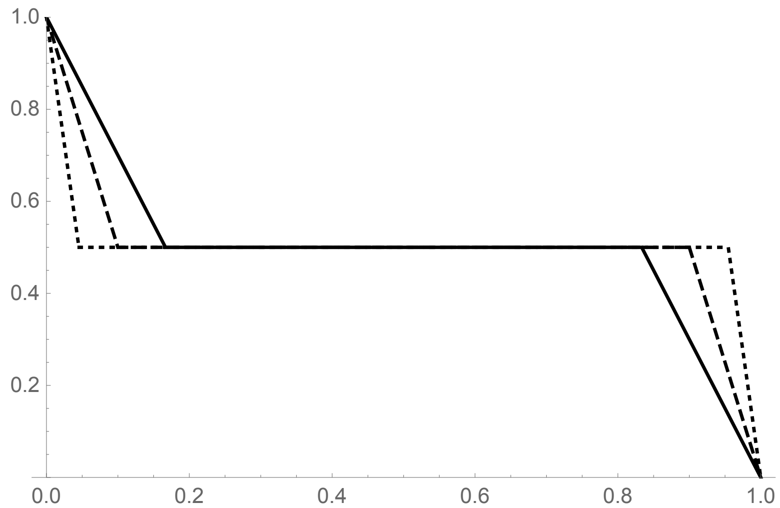

FIGURE 2.— Graph of equilibrium allocation for the case of $n = 2$ (solid line) and $n = 4$ (dashed line) assuming $\mathcal{S}(q) = n\left(q - \frac{1}{2}q^2\right)$, $[\underline{\theta}, \overline{\theta}] = [0, 1]$, and $k = \frac{3}{4}$.

•*Non-smooth equilibria.* The characterization of such equilibria has been made in Martimort, Semenov and Stole (2017) for the case where the participation constraint is binding at one extreme of the type space. The recipe of that paper allows one to reconstruct discontinuous equilibria starting from the continuous allocation $\overline{q}^m$ and repeatedly removing single output intervals from $\overline{q}^m(\Theta)$. In a sense the allocation $(\overline{u}^m, \overline{q}^m)$ is *maximal* given the equilibrium range is as large as possible. Indeed, we can apply this methodology *mutatis mutandis* here also when we restrict such removal to domains where the participation constraint is non-binding.

Constructing such equilibria is rather easy following the approach in Martimort, Semenov and Stole (2017). Indeed, let us take two types (θ_3, θ_4) such that $\underline{\theta} < \theta_3 < \theta_4 < \theta_1^n$ and thus

(3.30) $\quad \overline{q}^m(\underline{\theta}) > \overline{q}^m(\theta_3) > \overline{q}^m(\theta_4) > k.$[30]

Notice that $\overline{\gamma}^m(\theta) = 0$ for all $\theta \in [\underline{\theta}, \theta_1^n]$ and that $\overline{q}^m(\theta_3) > \overline{q}^m(\theta_4)$. Define then θ_0 such that

$$\mathcal{S}(\overline{q}^m(\theta_3)) - (\theta_0 + n(\theta_0 - \underline{\theta}))\overline{q}^m(\theta_3) = \mathcal{S}(\overline{q}^m(\theta_4)) - (\theta_0 + n(\theta_0 - \underline{\theta}))\overline{q}^m(\theta_4). \tag{3.31}$$

It can be easily checked that there exists a unique such type θ_0.

We now construct an equilibrium whose output range $\overline{q}(\Theta)$ is the non-connected set $q^m(\Theta)/(\overline{q}^m(\theta_4), \overline{q}^m(\theta_3))$. At the discontinuity of the output profile, it must be that type θ_0 is indifferent between choosing $\overline{q}^m(\theta_3)$ and $\overline{q}^m(\theta_4)$ which means that the equilibrium aggregate tariff must also satisfy the following incentive constraint:

$$\overline{\mathcal{T}}(\overline{q}^m(\theta_3)) - \theta_0(\overline{q}^m(\theta_3) - k) = \overline{\mathcal{T}}(\overline{q}^m(\theta_4)) - \theta_0(\overline{q}^m(\theta_4) - k). \tag{3.32}$$

Taken in tandem, (3.31) and (3.32) also show that the surrogate principal's objective function is kept constant at type θ_0 when moving across the discontinuity from $\overline{q}^m(\theta_3)$ to $\overline{q}^m(\theta_4)$.

Proposition 6 *Fix $(\theta_0, \theta_3, \theta_4) \in \Theta$ defined as in (3.30) and (3.31). The following output levels and the adjoint function $\overline{\gamma}^m(\theta)$ satisfy the necessary and sufficient conditions for optimality of Theorem 1 when applied to the self-generating problem $(\mathcal{P})$:*

$$\overline{q}(\theta) = \begin{cases} \overline{q}^m(\theta) & \text{if } \theta \in [\underline{\theta}, \theta_3] \cup [\theta_4, \overline{\theta}], \\ \overline{q}^m(\theta_3) & \text{if } \theta \in [\theta_3, \theta_0], \\ \overline{q}^m(\theta_4) & \text{if } \theta \in (\theta_0, \theta_4]. \end{cases} \tag{3.33}$$

An equilibrium aggregate tariff that implements the above output profile is:

$$\overline{\mathcal{T}}(q) = \begin{cases} \overline{\mathcal{T}}^m(q) & \text{if } q \in [\overline{q}^m(\overline{\theta}), \overline{q}^m(\theta_4)) \cup (\overline{q}^m(\theta_3), \overline{q}^m(\underline{\theta})], \\ \overline{\mathcal{T}}^m(\overline{q}^m(\theta_4)) + \theta_0(q - \overline{q}^m(\theta_4)) & \text{if } q \in [\overline{q}^m(\theta_4), \overline{q}^m(\theta_3)]. \end{cases} \tag{3.34}$$

The participation constraint (2.1) is binding on $[\theta_1^n, \theta_2^n]$ and slack elsewhere.

The strength of our approach is to use necessary and sufficient conditions for a solution to the self-generating problem that do not impose any regularity property on the equilibrium aggregate price $\overline{\mathcal{T}}$ that enters the maximand, except upper semi-continuity. Close inspection of those formula nevertheless shows that the discontinuity in the output profile that arises at θ_0 does not necessarily imply a discontinuity of this schedule. Indeed, although $\overline{\mathcal{T}}(q)$ as defined in (3.34) is concave and thus almost everywhere differentiable, it is not so either at $q = k$ or at $q = \overline{q}^m(\theta_3)$ and $q = \overline{q}^m(\theta_4)$. In fact, $\overline{\mathcal{T}}(q)$ only admits sub-differentials there. Yet, this schedule remains continuous.

Our approach has nevertheless some broader appeal. Consider the alternative nonlinear price:

$$\overline{\mathcal{T}}_1(q) = \begin{cases} \overline{\mathcal{T}}^m(q) & \text{if } q \in [\overline{q}^m(\overline{\theta}), \overline{q}^m(\theta_4)] \cup [\overline{q}^m(\theta_3), \overline{q}^m(\underline{\theta})], \\ -\infty & \text{if } q \in (\overline{q}^m(\theta_4), \overline{q}^m(\theta_3)). \end{cases} \tag{3.35}$$

[30] The same construction could be entertained with the polar assumption $\overline{q}^m(\overline{\theta}) < \overline{q}^m(\theta_4) < \overline{q}^m(\theta_3) < k$. We leave details to the reader.

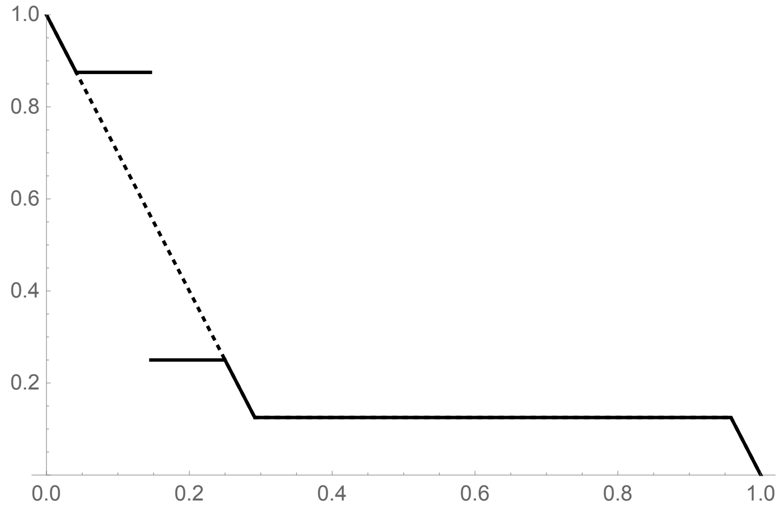

FIGURE 3.— Graph of a discontinuous equilibrium allocation with discontinuity between $\theta_3 = \frac{1}{24}$ and $\theta_4 = \frac{6}{24}$, assuming $n = 2$, $\mathcal{S}(q) = \left(q - \frac{1}{2}q^2\right)$, $[\underline{\theta}, \overline{\theta}] = [0, 1]$, and $k = \frac{1}{8}$.

These tariffs, as suggested in Martimort, Semenov and Stole (2017) have *by construction* the nice property to preclude any principal to deviate by inducing the agent to choose an output in the discontinuity gap $(\overline{q}^m(\theta_4), \overline{q}^m(\theta_3))$ since it would imply infinite punishments from other principals. The tariff $\overline{\mathcal{T}}_1(q)$ is clearly upper semi-continuous but it obviously fails to be concave. Observe that nevertheless $\overline{co}(\overline{\mathcal{T}}_1) = \overline{\mathcal{T}}$. It follows from this property that both the agent's maximization problem and the self-generating problem still have for solution $\overline{q}(\theta)$ as defined in (3.33). Yet, even though the self-generating problem constructed with $\overline{\mathcal{T}}_1(q)$ has a non-smooth maximand, its solution can still be obtained with our approach. This remark shows in passing that different aggregates may induce the same solutions to the self-generating problem and thus support the same equilibrium outcome.

Let us turn now to the characterization of another class of discontinuous equilibria where the discontinuity in the equilibrium output schedule will arise at a point θ_0 where the participation constraint is binding. If such equilibria exist, they are necessarily of a very different nature than those described in Proposition 6. Indeed, the possibility of an output discontinuity at such point means that the right- and the left-derivatives of $\overline{u}$ are different, so θ_0 is necessarily an isolated point where (2.1) is binding. This remark suggests that we may look for an adjoint function $\overline{\gamma}(\theta)$ that would have an upward jump at such discontinuity point θ_0:

$$(3.36) \quad \overline{\gamma}(\theta) = \begin{cases} 0 & \text{if } \theta \in [\underline{\theta}, \theta_0), \\ 1 & \text{f } \theta \in [\theta_0, \overline{\theta}]. \end{cases}$$

To describe such equilibrium, we now proceed as above. We again choose θ_3 and θ_4 on the separating range of $\overline{q}^m$ but now satisfying

$$(3.37) \quad \overline{q}^m(\underline{\theta}) > \overline{q}^m(\theta_3) > k > \overline{q}^m(\theta_4) > \overline{q}^m(\overline{\theta}).$$

Since $\overline{\gamma}$ is right-continuous at the discontinuity point θ_0, we have $\overline{\gamma}(\theta_0) = 1$ and θ_0 is defined so that

$$(3.38) \quad \mathcal{S}(\overline{q}^m(\theta_3)) - (\theta_0 + n(\theta_0 - \overline{\theta}))\overline{q}^m(\theta_3) = \mathcal{S}(\overline{q}^m(\theta_4)) - (\theta_0 + n(\theta_0 - \overline{\theta}))\overline{q}^m(\theta_4).$$

Taken in tandem with the incentive constraint of that type θ_0 that still writes as (3.32), these conditions again imply that the surrogate principal's surplus function remains constant throughout the discontinuity at θ_0.

PROPOSITION 7 *Fix* $(\theta_0, \theta_3, \theta_4) \in \Theta$ *defined as in (3.37) and (3.38). The output levels and the adjoint function* $\overline{\gamma}^m(\theta)$ *as defined respectively in (3.33) and (3.36) satisfy the necessary and sufficient conditions for optimality of Theorem 1 when applied to the self-generating problem* $(\mathcal{P})$*. An equilibrium aggregate tariff that implements the above output profile is:*

$$\overline{\mathcal{T}}(q) = \begin{cases} \overline{\mathcal{T}}^m(q) - \overline{\mathcal{T}}^m(\overline{q}^m(\theta_3)) + \theta_0(\overline{q}^m(\theta_3) - k) & \text{if } q \in (\overline{q}^m(\theta_3), \overline{q}^m(\underline{\theta})], \\ \theta_0(q-k) & \text{if } q \in [\overline{q}^m(\theta_4), \overline{q}^m(\theta_3)], \\ \overline{\mathcal{T}}^m(q) - \overline{\mathcal{T}}^m(\overline{q}^m(\theta_4)) + \theta_0(\overline{q}^m(\theta_4) - k) & \text{if } q \in (\overline{q}^m(\theta_4), \overline{q}^m(\underline{\theta})]. \end{cases} \tag{3.39}$$

The participation constraint (2.1) is binding only at θ_0 *and slack elsewhere.*

REFERENCES

AGHION, P. AND P. BOLTON (1987). Contracts as a Barrier to Entry, *The American Economic Review,* **77** 388–401.

ATTAR, A., T. MARIOTTI AND F. SALANIÉ (2011). Nonexclusive Competition in the Market for Lemons. *Econometrica,* **79** 1869–1891.

BARON, D. AND R. MYERSON (1982). Regulating a Monopolist with Unknown Costs. *Econometrica,* **50** 911–930.

BIGLAISER, G. AND C. MEZZETTI (1993). Principals Competing for an Agent in the Presence of Adverse Selection and Moral Hazard. *Journal of Economic Theory,* **61** 302–330.

BRAINARD, S. AND D. MARTIMORT (1997). Strategic Trade Policy with Incompletely Informed Policymakers. *Journal of International Economics,* **42** 33–65.

CAILLAUD, B. (1990). Regulation, Competition, and Asymmetric Information. *Journal of Economic Theory,* **52** 87–110.

CALZOLARI, G. AND V. DENICOLO (2013). Competition with Exclusive Contracts and Market-Share Discounts. *The American Economic Review,* **103** 2384–2411.

CALZOLARI, G. AND V. DENICOLO (2015). Exclusive Contracts and Market Dominance. *American Economic Review,* **105** 3321–51.

CARBAJAL, J. AND J. ELY (2013). Mechanism Design without Revenue Equivalence. *Journal of Economic Theory,* **148** 104-133.

CARBAJAL, J. AND J. ELY (2016). A Model of Price Discrimination under Loss Aversion and State-Contingent Reference Points. *Theoretical Economics,* **11** 455-485.

CHAMPSAUR, P. AND J.C. ROCHET (1989). Multiproduct Duopolists. *Econometrica* **57** 533–557.

CHONÉ, P. AND L. LINNEMER (2015). Nonlinear Pricing and Exclusion: I. Buyer Opportunism, *The RAND Journal of Economics,* **46** 217–240.

CREMER, J., F. KHALIL AND J.C. ROCHET (1998). Contracts and Productive Information Gathering. *Games and Economic Behavior,* **25** 174–193.

CURIEN, N., B. JULLIEN AND P. REY (1998). Pricing Regulation under Bypass Competition. *The RAND Journal of Economics,* **29** 259–279.

DEB, R. AND M. SAID (2015). Dynamic Screening with Limited Commitment. *Journal of Economic Theory,* **159** 891–928.

Ferrera, J. (2014). *An Introduction to Nonsmooth Analysis.* Academic Press.

Galbraith, G and R. Vinter (2004). Regularity of Optimal Controls for State-Constrained Problems. *Journal of Global Optimization* **28** 305–317.

Hellwig, M. (2010). Incentive Problems with Unidimensional Hidden Characteristics: A Unified Approach. *Econometrica,* **78** 1201–1237.

Huang, K. and A. Sundararajan (2011). Pricing Digital Goods: Discontinuous Costs and Shared Infrastructure. *Information Systems Research,* **22** 721–738.

Jullien, B. (2000). Participation Constraints in Adverse Selection Models. *Journal of Economic Theory* **93** 1–47.

Laffont, J. J. and J. Tirole (1990). Optimal Bypass and Cream Skimming. *The American Economic Review,* **80** 1042–1061.

Laffont, J. J. and J. Tirole (1991). Privatization and Incentives. *Journal of Law, Economics and Organization,* **7** 84–105.

Laffont, J.J. and D. Martimort (2002). *The Theory of Incentives: The Principal-Agent Model.* Princeton University Press.

Lewis, T. and D. Sappington (1989a). Countervailing Incentives in Agency Problems. *Journal of Economic Theory,* **49** 294–313.

Lewis, T. and D. Sappington (1989b). Inflexible Rules in Incentive Problems. *The American Economic Review,***79** 69–84.

Lewis, T. and D. Sappington (1993). Ignorance in Agency Problems. *Journal of Economic Theory,* **61** 169–183.

Maggi, G. and A. Rodriguez-Clare (1995). On Countervailing Incentives. *Journal of Economic Theory.* **66** 238–263.

Martimort, D. and A. Semenov (2008). Ideological Uncertainty and Lobbying Competition. *Journal of Public Economics*, **92** 456–481.

Martimort, D., J. Pouyet and L. Stole (2018). Screening Contracts as a Barrier to Entry, Mimeo PSE.

Martimort, D., A. Semenov and L. Stole (2017). A Complete Characterization of Equilibria in an Intrinsic Common Agency Screening Game. *Theoretical Economics.* **13** 1151–1189.

Martimort, D. and L. Stole (2002). The Revelation and Delegation Principles in Common Agency Games. *Econometrica.* **70** 1659–1673.

Martimort, D. and L. Stole (2009). Market Participation in Delegated and Intrinsic Common Agency Games. *The RAND Journal of Economics*, **40** 78–102.

Martimort, D. and L. Stole (2012). Representing Equilibrium Aggregates in Aggregate Games with Applications to Common Agency. *Games and Economic Behavior,* **76** 753–772.

Martimort, D. and L. Stole (2015). Menu Auctions and Influence Games. Mimeo Chicago Booth School of Business.

Mezzetti, C. (1997). Common Agency with Horizontally Differentiated Principals. *The RAND Journal of Economics,* **28** 323–345.

Milgrom, P. and I. Segal (2002). Envelope Theorems for Arbitrary Choice Sets. *Econometrica,* **70** 583–601.

Mirrlees, J. (1971). An Exploration in the Theory of Optimum Income Taxation. *The Review of Economic Studies,* **38** 175–208.

Mussa, M. and S. Rosen (1978). Monopoly and Product Quality. *Journal of Economic Theory,* **18** 301–317.

Myerson, R. (1981). Optimal Auction Design. *Mathematics of Operations Research.* **6** 58–73.

Ollier, S. and L. Thomas (2013). Ex Post Participation Constraint in a PrincipalAgent Model with Adverse Selection and Moral Hazard. *Journal of Economic Theory,* **148** 2383–2403.

Peters, M. (2001). Common Agency and the Revelation Principle. *Econometrica.* **69** 1349–1372.

Rochet, J. C. (1985). The Taxation Principle and Multi-Time Hamilton-Jacobi Equations. *Journal of Mathematical Economics.* **14** 113–128.

Rochet, J. C. (1987). A Necessary and Sufficient Condition for Rationalizability in a Quasi-Linear Context. *Journal of Mathematical Economics.* **16** 191–200.

Rockafellar, T. and R. Wets (2004). *Variational Analysis.* Berlin: Springer.

Rothschild, M. and J. Stiglitz (1976). Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information. *The Quarterly Journal of Economics,* **90** 629–649.

Royden, H.L. (1988). *Real Analysis.* New York: Prentice Hall.

Seierstad, A. and K. Sydsaeter (1987). *Optimal Control Theory with Economic Applications.* Amsterdam: North Holland.

Spulber, D. (1993). Monopoly Pricing of Capacity Usage under Asymmetric Information. *Journal of Industrial Economics* **41** 241–257.

Stole, L. (1995). Nonlinear Pricing and Oligopoly. *Journal of Economics and Management Strategy,* **4** 529–562.

Stole, L. (2003). Price Discrimination and Imperfect Competition. In *Handbook of Industrial Organization,* M. Armstrong and R. Porter eds., **3** 34–47.

Thomas, L. (2001). Cost Functions in Non-Linear Pricing. *Economics Letters* **72** 53–59.

Vinter, R. (2000). *Optimal Control.* Boston: Birkhauser.

Vinter, R. and H. Zheng (1998). Necessary Conditions for Optimal Control Problem with State Constraints. *Transactions of the American Mathematical Society* **350** 1181–1204.

APPENDIX A: NON-SMOOTH OPTIMIZATION

This Appendix briefly reminds the reader about necessary and sufficient conditions for the general problem of maximizing an upper semi-continuous function, $h : \mathbb{R} \to \mathbb{R}$, over a compact set $X \subset \mathbb{R}$.

A generalization of the first-order condition for smooth, concave optimization programs can indeed be obtained for general upper semi-continuous programs by introducing a few concepts from non-smooth convex analysis. The basic idea is that any solution to the original upper semi-continuous program must lie on the minimal concave envelope or *concavification* of the objective. Consider, for example, the upper semi-continuous function graphed in Figure 3.3 in bold.

Figure 3.3:

This function is defined over the real line, but the restricted domain of interest is $X = [\underline{x}, \overline{x}]$. The minimal-concave envelope over this domain is depicted by the dashed lines in the graph. Notice its value is negative infinity outside of $[\underline{x}, \overline{x}]$. Obviously, the maximum of *this* concave envelope is a solution to the original program. More generally, in the case in which there is a continuum of solutions (i.e., the maximum is achieved on a horizontal component of the majorization), there exist two solutions to the original program – the endpoints of the majorization. It is, in this sense, without loss to convert an upper semi-continuous program over a compact set into a concave (but possibly non differentiable) program over the same set. Formally, we will denote $\overline{co}_X(h)$ to refer to the concavification of an objective function, h, over a domain, X,[31] and $\overline{co}_X(h)(x)$ to refer to the value of this envelope evaluated at x.[32]

Having reached the conclusion that we may focus on the concave envelope of the program, we can now import the generalized notion of derivative from convex analysis. Formally, we will define a set of gradients at any point to be all those vectors which "support" the graph at the given point, and we refer to this set-valued notion of derivative as the generalized gradient or the *super-differential*, denoted $\partial h(x)$ when applied to a concave function h at point x.[33] Where h is differentiable, the super-differential is single-valued and corresponds to the gradient. If h exhibits a kink and $X \subseteq \mathbb{R}$, the super-differential is an interval of gradients with endpoints corresponding to the left- and right-side derivatives at the point. More generally, if $X \subseteq \mathbb{R}^n$, then

$$\partial h(x) = \{\tau \in \mathbb{R}^n \mid h(y) \leq h(x) + \langle \tau, y - x \rangle \; \forall \; y \in \mathbb{R}^n\}.$$

Using this generalization of gradient, we can now state the necessary and sufficient conditions for x^* to be a maximum of an upper semi-continuous function, h, over some given domain $x \in X$: if x^* is a solution to the maximization program, then the following first-order condition must be satisfied:

$$\text{(A1)} \qquad 0 \in \partial \overline{co}_X(h)(x^*).$$

Furthermore, if x^* satisfies (A1) and the envelope coincides with h at x^*, i.e.,

$$\text{(A2)} \qquad \overline{co}_X(h)(x^*) = h(x^*)$$

then x^* solves the maximization program.

These conditions can be further tightened when a component of the objective function is affine. To this end, suppose that $h = g + f$ where g is affine (and slightly abusing notations, let us write $g(x) = gx$). Well-known identities from convex analysis give us:

$$\overline{co}_X(h)(x) = gx + \overline{co}_X(f)(x) \text{ and } \partial \overline{co}_X(h)(x) = g + \partial \overline{co}_X(f)(x).$$

Thus, the linear linear part of the objective can be factored out and the "first-order" necessary and sufficient condition for the optimality of x^* reduces to

$$-g \in \partial \overline{co}_X(f)(x^*).$$

This property will be repeatedly used throughout our analysis, first, to derive generalized first-order conditions for our infinite-dimensional optimal control problem and, second, to tackle applications in contract theory where such a decomposition is frequently available.

[31] When $X = \mathbb{R}$, we simplify notations and omit the subscript.

[32] In the non-smooth optimization literature, often one considers the minimal concave envelope of h over the real line instead of some domain X, but in this case with a penalty function, $\Psi_X(x)$ which equals 0 for $x \in X$ and $-\infty$ for $x \notin X$. Thus, in our notation, $\overline{co}_X(h) = \overline{co}_{\mathbb{R}}(h + \Psi_X)$.

[33] In our context of maximizing a concave function, it is perhaps more accurate to say that the graph of a concave function "supports" its gradients. Nonetheless, we use the term "support" from convex analysis given it is evocative and familiar. The term *sub-differential* is the parallel notion of super-differential when applied to convex functions. When we refer to the generalized gradient of a function that is understood to be convex, we will abuse notation slightly by again using the notation $\partial h(x)$, where it is understood that when h is convex, then

$$\partial h(x) = \{\tau \in \mathbb{R}^n \mid h(y) \geq h(x) + \langle \tau, y - x \rangle \; \forall \; y \in \mathbb{R}^n\}.$$

See Ferrera (2014) for an introduction to non-smooth analysis and an in-depth discussion of super- and sub-differentials.

APPENDIX B: PROOF OF THEOREM 1

Preliminaries for Non-Smooth Analysis. We draw heavily from Vinter and Zheng (1998) in the following presentation. A complete treatment can be found in the monograph of Vinter (2000). Theorem 3 from Vinter and Zheng (1998) appears as Theorem 10.2.1 in Vinter (2000).

Take a closed set $A \subseteq \mathbb{R}^n$ and a point $x \in A$. A vector $r \in \mathbb{R}^n$ is a *limiting normal* to A at x if there exists a sequence $(x_i, r_i) \to (x, r)$ with $x_i \in A$ and a constant $M \geq 0$ such that for each i in the sequence $r_i \cdot (x_i - x) \leq M||x_i - x||^2$, where $|| \cdot ||$ denotes Euclidean distance. The cone of limiting normal vectors to A at x is denoted $N_A(x)$. Given a lower semi-continuous function $g : \mathbb{R} \to \mathbb{R} \cup \{+\infty\}$ and a point $x \in \mathbb{R}$ such that $g(x) < +\infty$, the *limiting sub-differential* of g at x is defined as

$$\partial g(x) \equiv \{\xi \,|\, (\xi, -1) \in N_{\text{epi}\{g\}}(x, g(x))\},$$

where epi$\{g\}$ is the *epigraph* of the function g defined as

$$\text{epi}\{g\} \equiv \{(x, \alpha) \in \mathbb{R} \times \mathbb{R} \,|\, \alpha \geq g(x)\}.$$

The *asymptotic limiting sub-differential* of g at x, written $\partial^\infty g(x)$, is defined as

$$\partial^\infty g(x) \equiv \{\xi \,|\, (\xi, 0) \in N_{\text{epi}\{g\}}(x, g(x))\}.$$

Finally, we define:

$$\partial_x^> h(t, x) \equiv \overline{co}\{\lim_i \xi_i \,|\, \exists t_i \to t, x_i \to x \;\text{ s.t. } h(t_i, x_i) > 0 \text{ and } \xi_i \in \partial_x h(t_i, x_i) \;\forall i\}.$$

Two results from non-smooth analysis (Vinter, 2000, Propositions 4.3.3 and 4.3.4) that we use are (1) $\partial^\infty g(x) = \{0\}$ if g is Lipschitz continuous and (2) for any x such that $g(x)$ is finite,

$$N_{\text{epi}\{g\}}(x, g(x)) = \{(\xi d, -\xi) \,|\, \xi > 0\,, d \in \partial g(x)\} \cup \{\partial^\infty g(x) \times \{0\}\}.$$

A *local maximizer* of $\Lambda(x)$ is a feasible arc, $\overline{x}$, which maximizes $\Lambda(x)$ over all feasible arcs $x \in AC(\Theta, \mathbb{R}_+)$ within an ε neighborhood of $\overline{x}$, $||\overline{x} - x||_{ac} \leq \varepsilon$ where we denote the norm on the space of absolutely continuous functions by $||x||_{ac} \equiv ||x(\underline{\theta})|| + \int_{\underline{\theta}}^{\overline{\theta}} ||\dot{x}(\theta)||d\theta$. A *local minimizer* is defined analogously.

Necessity. First, and for completeness, we reproduce here Theorem 3 of Vinter and Zheng (1998) which provides necessary conditions for solutions to the following minimization program:

$$(\mathcal{P}') : \qquad \text{Minimize } J(x) \equiv \int_{\underline{\theta}}^{\overline{\theta}} L(\theta, x(\theta), \dot{x}(\theta))d\theta$$

$$\text{subject to } x \in AC(\Theta, \mathbb{R}) \text{ and } h(\theta, x(\theta)) \leq 0 \text{ for all } \theta \in \Theta \equiv [\underline{\theta}, \overline{\theta}].$$ [34]

We will prove necessity for Theorem 1 by specializing this Theorem, exploiting the fact that our integrand in Λ is a linear function of x and $h(\theta, x) = -x$.

Theorem B.1 *(Vinter and Zheng, 1998, Theorem 3) Let $\overline{x}$ be local minimizer for $(\mathcal{P}')$ in $AC(\Theta, \mathbb{R})$ such that $J(\overline{x}) < +\infty$. Assume that the following hypotheses are satisfied:*

H_1. $L(\cdot, x, \cdot)$ is $\mathcal{L} \times \mathcal{B}$ measurable for each x and $L(\theta, \cdot, \cdot)$ is lower semi-continuous for a.e. $\theta \in \Theta$.

[34] We specialize their theorem to our present problem in which the range of $x(\theta)$ is one-dimensional and there is no endpoint cost function.

H_2. For every $N > 0$ there exists $\delta > 0$ and $k \in L^1$ such that

$$||L(\theta, x', v) - L(\theta, x, v)|| \leq k(\theta)||x' - x||, \quad L(\theta, \overline{x}(\theta), v) \geq -k(\theta)$$

for a.e. $\theta \in \Theta$, for all $x, x' \in \overline{x}(\theta) + \delta B$ and $v \in \dot{\overline{x}}(\theta) + NB$, where B is a unit Euclidean ball.

H_3. h is upper semi-continuous near $(\theta, \overline{x}(\theta))$ for all $\theta \in \Theta$, and there exists a constant k_h such that

$$||h(\theta, x') - h(\theta, x)|| \leq k_h||x' - x||$$

for all $\theta \in \Theta$ and all $x', x \in \overline{x}(\theta) + \delta B$.

Then there exist an arc $p \in AC(\Theta, \mathbb{R})$, a constant $\lambda \geq 0$, a non-negative measure μ on the Borel subsets of Θ and a μ-integrable function $\zeta : \Theta \to \mathbb{R}$, such that

(i). $\lambda + \max_{\theta \in \Theta} |p(\theta)| + \int_{\underline{\theta}}^{\overline{\theta}} \mu(d\tilde{\theta}) = K > 0$ *(where K is an arbitrary normalization constant),*[35]

(ii).

$$\dot{p}(\theta) \in \overline{co}\left\{\eta \,|\, (\eta, p(\theta) + \int_{[\underline{\theta}, \theta)} \zeta(\tilde{\theta})\mu(d\tilde{\theta}), -\lambda) \in N_{\text{epi}\{L(\theta,\cdot,\cdot)\}}(\overline{x}(\theta), \dot{\overline{x}}(\theta), L(\theta, \overline{x}(\theta), \dot{\overline{x}}(\theta)))\right\} \text{ a.e.,}$$

(iii).

$$p(\underline{\theta}) = p(\overline{\theta}) - \int_{\underline{\theta}}^{\overline{\theta}} \zeta(\tilde{\theta})\mu(d\tilde{\theta}) = 0,$$

(iv).

$$\left(p(\theta) + \int_{[\underline{\theta}, \theta)} \zeta(\tilde{\theta})\mu(d\tilde{\theta})\right)\dot{\overline{x}}(\theta) - \lambda L(\theta, \overline{x}(\theta), \dot{\overline{x}}(\theta)) \in \arg\max_{v \in \mathbb{R}} \left(p(\theta) + \int_{[\underline{\theta}, \theta)} \zeta(\tilde{\theta})\mu(d\tilde{\theta})\right) v - \lambda L(\theta, \overline{x}(\theta), v),$$

(v). $\zeta(\theta) \in \partial_x^{>} h(\theta, \overline{x}(\theta))$ *μ-a.e. and* $\text{supp}\{\mu\} \subseteq \{\theta \,|\, h(\theta, \overline{x}(\theta)) = 0\}$.

We apply this result to our setting by substituting $xf(\theta) - s(\theta, v)$ in program $(\mathcal{P})$ in place of $L(\theta, x, v)$ and thereby converting the maximization functional Λ in program $(\mathcal{P})$ to the minimization functional J in program $(\mathcal{P}')$. We complete the transformation by requiring that $h(\theta, x) = -x$, and that $L(\theta, x, v)$ is a linear function of x for any (θ, v).

First, we verify that hypotheses H_1-H_3 are satisfied for our program $(\mathcal{P})$. Because $s(\theta, \cdot)$ is upper semi-continuous and $\mathcal{B}$-measurable, and because $L(\theta, x, v)$ is linear in x, H_1 is satisfied. H_2 requires that $L(\theta, \cdot, v)$ is Lipschitz continuous, which is trivial given that L is linear in x with coefficient $f(\theta)$. Because the transformed program has $h(\theta, x) = -x$, h is a continuous linear function of x and thus H_3 is also satisfied.

Next, we specialize the conclusions of Vinter and Zheng (1998) by making use of the additional restrictions on $L(\cdot)$ and $h(\cdot)$. We present this in the following Lemma.

[35]We choose to state the Theorem using $K > 0$ as an arbitrary normalization rather than $K = 1$, which is the normalization chosen in Vinter and Zheng (1998). Later, by setting $K = 3$, we will succeed in normalizing μ to a probability measure which is a more familiar object.

LEMMA B.1 *Suppose that $L(\theta,x,v)$ is a linear function of x and that $h(\theta,x)=-x$. Then the conclusions (i)-(v) of Theorem B.1 imply*

(a). $\lambda+\max_{\theta\in\Theta}|p(\theta)|+\int_{\underline{\theta}}^{\overline{\theta}}\mu(d\tilde{\theta})=K$,

(b). $\dot{p}(\theta)=\lambda f(\theta)$ *a.e.*,

(c). $p(\underline{\theta})=p(\overline{\theta})+\int_{\underline{\theta}}^{\overline{\theta}}\zeta(\tilde{\theta})\mu(d\tilde{\theta})=0$

(d). $\dot{\overline{x}}(\theta)\in\arg\max_{v\in\mathbb{R}}\left(p(\theta)+\int_{[\underline{\theta},\theta)}\zeta(\tilde{\theta})\mu(d\tilde{\theta})\right)v+\lambda s(\theta,v)$, *a.e.*,

(e). $\zeta(\theta)=-1$ *μ-a.e. and* $\operatorname{supp}\{\mu\}\subseteq\{\theta\,|\,\overline{u}(\theta)=0\}$.

PROOF OF LEMMA B.1: Implications (i) and (a) are identical. Implication (ii) requires

$$\dot{p}(\theta)\in\overline{co}\left\{\eta\mid\left(\eta,p(\theta)+\int_{[\underline{\theta},\theta)}\zeta(\tilde{\theta})\mu(d\tilde{\theta}),-\lambda\right)\in N_{\text{epi}(L(\theta,\cdot,\cdot))}\left(\overline{x}(\theta),\dot{\overline{x}}(\theta),L(\theta,\overline{x}(\theta),\dot{\overline{x}}(\theta))\right)\right\},a.e.$$

Because $L(\theta,\overline{x}(\theta),\dot{\overline{x}}(\theta))=f(\theta)\overline{x}(\theta)-s(\theta,\dot{\overline{x}}(\theta))$ is finite, the limiting normal cone in the above expression can be written as

$$\begin{aligned}N_{\text{epi}(L(\theta,\cdot,\cdot))}&\left(\overline{x}(\theta),\dot{\overline{x}}(\theta),L(\theta,\overline{x}(\theta),\dot{\overline{x}}(\theta))\right)\\&=\left\{(\xi d_1,\xi d_2,-\xi)\,|\,\xi>0,\,(d_1,d_2)\in\partial\left(f(\theta)\overline{x}(\theta)-s(\theta,\dot{\overline{x}}(\theta))\right)\right\}\\&\qquad\bigcup\left\{\partial^\infty\left(f(\theta)\overline{x}(\theta)-s(\theta,\dot{\overline{x}}(\theta))\right)\times\{0\}\right\}.\end{aligned}$$

Using the fact that $L(\cdot)$ is additively separable in x and $\dot{x}$ yields (Rockafellar and Wets (2004, Proposition 10.5))

$$\begin{aligned}\partial\left(f(\theta)\overline{x}(\theta)-s(\theta,\dot{\overline{x}}(\theta))\right)&=\partial\left(f(\theta)\overline{x}(\theta)\right)\times\partial\left(-s(\theta,\dot{\overline{x}}(\theta))\right)\\&=\{f(\theta)\}\times\partial\left(-s(\theta,\dot{\overline{x}}(\theta))\right)\end{aligned}$$

and

$$\begin{aligned}\partial^\infty\left(f(\theta)\overline{x}(\theta)-s(\theta,\dot{\overline{x}}(\theta))\right)&\subseteq\partial^\infty\left(f(\theta)\overline{x}(\theta)\right)\times\partial^\infty\left(-s(\theta,\dot{\overline{x}}(\theta))\right)\\&=\{0\}\times\partial^\infty\left(-s(\theta,\dot{\overline{x}}(\theta))\right),\end{aligned}$$

where the last equality uses the fact that a linear function is Lipschitz continuous and hence $\partial^\infty(f(\theta)\overline{u}(\theta))=\{0\}$. Substituting these sub-differentials into the expression for the limiting normal cone, we have a simple inclusion:

$$\begin{aligned}N_{\text{epi}(L(\theta,\cdot,\cdot))}\left(\overline{x}(\theta),\dot{\overline{x}}(\theta),L(\theta,\overline{x}(\theta),\dot{\overline{x}}(\theta))\right)&\subseteq\left\{(\xi f(\theta),\xi d_2,-\xi)\,|\,\xi>0,\,d_2\in\partial\left(-s(\theta,\dot{\overline{x}}(\theta))\right)\right\}\\&\qquad\bigcup\left\{\{0\}\times\partial^\infty\left(-s(\theta,\dot{\overline{x}}(\theta))\right)\times\{0\}\right\}.\end{aligned}$$

This simplifies again to the inclusion

$$\begin{aligned}N_{\text{epi}(L(\theta,\cdot,\cdot))}&\left(\overline{x}(\theta),\dot{\overline{x}}(\theta),\overline{L}(\theta,\overline{x}(\theta),\dot{\overline{x}}(\theta))\right)\\&\subseteq\left\{(\xi f(\theta),\xi d_2,-\xi)\,|\,\xi\geq0,\,d_2\in\partial\left(-s(\theta,\dot{\overline{x}}(\theta))\right)\bigcup\partial^\infty\left(-s(\theta,\dot{\overline{x}}(\theta))\right)\right\}.\end{aligned}$$

The key point to note is that any vector in the limiting normal cone must point in the same direction in the $(\overline{x},\overline{L})$ plane, regardless of d_2. Returning to implication (ii), we see that any point η in the given convex hull must satisfy $(\eta,\cdot,-\lambda)=(\xi f(\theta),\cdot,-\xi)$ for some $\xi\geq0$, and hence the convex hull reduces to $\{\lambda f(\theta)\}$. We conclude that implication (ii) simplifies to implication (b) given that $L(\cdot)$ is both additively separable and linear in x.

Implication (iii) is identical to implication (c).

Using the transformation $L(\theta, x, v) = xf(\theta) - s(\theta, v)$, implication (iv) simplifies to implication (d). Lastly, the fact that $h(\theta, x) = -x$ yields $\partial_x h(\theta, \overline{u}(\theta)) = \partial_x^{>} h(\theta, \overline{x}(\theta)) = \{-1\}$. Thus, implication (v) simplifies to $\zeta(\theta) = -1$ μ-a.e. and

(B1) $\quad supp\{\mu\} \subseteq \{\theta \,|\, \overline{u}(\theta) = 0\}.$

This is implication (e). *Q.E.D.*

An immediate inspection of conditions (a)-(e) suggest further simplifications by combining these conditions. Conditions (b) and (c) jointly yield

$$p(\theta) = \lambda F(\theta).$$

Because $p(\overline{\theta}) = \lambda$ and $\zeta(\theta) = -1$ a.e. with respect to μ, condition (c) also implies

$$\int_{\underline{\theta}}^{\overline{\theta}} \mu(d\tilde{\theta}) = \lambda.$$

Because we also have $\max_{\theta\in\Theta} |p(\theta)| = \lambda$, condition (a) implies $\lambda > 0$ and in particular $\lambda = \frac{K}{3}$. Because the choice of K is arbitrary, we choose $K = 3$ as a normalization, yielding $\lambda = 1$ and $\int_{\underline{\theta}}^{\overline{\theta}} \mu(d\tilde{\theta}) = 1$. Thus, up to this normalization, μ is a probability measure on Θ. Defining now $\overline{\gamma}(\theta) = \int_{[\underline{\theta},\theta)} \mu(d\tilde{\theta})$, the implication in (d) is therefore

(B2) $\quad \dot{\overline{x}}(\theta) \in \arg\max_{v\in\mathbb{R}} \; s(\theta, v) + (F(\theta) - \overline{\gamma}(\theta))\, v, \; a.e..$

This condition can finally be expressed as (2.4) and (2.5) of Theorem 1. Lastly, implication of (e) delivers the complementary slackness condition (2.3). We have therefore proven the necessity of the conditions in Theorem 1.

Sufficiency. Sufficiency is proven by generalizing Arrow's *Sufficiency Theorem* to non-smooth optimal control problems and specializing the theorem to the case in which the objective integrand is a linear function of x. We adapt the argument of Arrow's *Sufficiency Theorem* using the approach of Seierstad and Sydsaeter (1987) but relaxing their continuity and smoothness assumptions. The regularity of the optimal solution follows from arguments involving the necessary conditions.

Let x be any admissible arc satisfying thus $x \in AC(\Theta, \mathbb{R})$ and $x(\theta) \geq 0$ for all $\theta \in \Theta$. Define

$$\Delta = \int_{\underline{\theta}}^{\overline{\theta}} \left\{ \left(s(\theta, \dot{\overline{x}}(\theta)) - \overline{x}(\theta) f(\theta)\right) - \left(s(\theta, \dot{x}(\theta)) - x(\theta) f(\theta)\right) \right\} d\theta.$$

We will demonstrate that, under conditions (B1) and (B2) of Theorem 1, $\Delta \geq 0$.

To this end, it is useful to define the Hamiltonian for program (P) with $\overline{\gamma}(\theta)$ being the adjoint equation which satisfies conditions(B1) and (B2):

$$H(\theta, x, v) \equiv s(\theta, v) - xf(\theta) - (\overline{\gamma}(\theta) - F(\theta))v.$$

Note that $\overline{\gamma}(\theta)$ is defined for $\theta \in (\underline{\theta}, \overline{\theta}]$ and thus $H(\cdot)$ inherits the same domain. Nonetheless, because μ is not part of expression of Δ and F is absolutely continuous, we can ignore the point $\underline{\theta}$ in the integral and conclude that

$$\Delta = \int_{(\underline{\theta}, \overline{\theta}]} \left(H(\theta, \overline{x}(\theta), \dot{\overline{x}}(\theta)) - H(\theta, x(\theta), \dot{x}(\theta))\right) d\theta + \int_{\underline{\theta}}^{\overline{\theta}} (F(\theta) - \overline{\gamma}(\theta)) \left(\dot{x}(\theta) - \dot{\overline{x}}(\theta)\right) d\theta.$$

Define the optimized Hamiltonian as

$$\hat{H}(\theta, x) \equiv \sup_{v \in \mathbb{R}} H(\theta, x, v).$$

Because $\overline{\gamma}(\theta) - F(\theta)$ is bounded on $(\underline{\theta}, \overline{\theta}]$ and $s(\theta, \cdot)$ is bounded from above by assumption, $\hat{H}(\cdot)$ must be finite. Condition (B2) implies that

$$\hat{H}(\theta, \overline{x}(\theta)) = H(\theta, \overline{x}(\theta), \dot{\overline{x}}(\theta))$$

and for any admissible $x \in AC(\Theta; \mathbb{R}_+)$,

$$\hat{H}(\theta, x(\theta)) \geq H(\theta, x(\theta), \dot{x}(\theta)).$$

Combining these facts, we obtain

$$\begin{aligned} H(\theta, \overline{x}(\theta), \dot{\overline{x}}(\theta)) - H(\theta, x(\theta), \dot{x}(\theta)) &\geq \hat{H}(\theta, \overline{x}(\theta)) - \hat{H}(\theta, x(\theta)) \\ &= f(\theta)(x(\theta) - \overline{x}(\theta)). \end{aligned}$$

The last statement relies fundamentally on the linearity of $H(\cdot)$ in x. Substituting into the previous statement for Δ, we have

$$\begin{aligned} \Delta &\geq \int_{(\underline{\theta}, \overline{\theta}]} f(\theta)(x(\theta) - \overline{x}(\theta)) d\theta + \int_{\underline{\theta}}^{\overline{\theta}} (F(\theta) - \overline{\gamma}(\theta)) \left(\dot{x}(\theta) - \dot{\overline{x}}(\theta)\right) d\theta \\ &= \int_{\underline{\theta}}^{\overline{\theta}} \left(f(\theta)(x(\theta) - \overline{x}(\theta)) + F(\theta) \left(\dot{x}(\theta) - \dot{\overline{x}}(\theta)\right)\right) d\theta - \int_{(\underline{\theta}, \overline{\theta}]} \overline{\gamma}(\theta) \left(\dot{x}(\theta) - \dot{\overline{x}}(\theta)\right) d\theta \\ &= \int_{\underline{\theta}}^{\overline{\theta}} \frac{d}{d\theta}[F(\theta)(x(\theta) - \overline{x}(\theta))] d\theta - \int_{(\underline{\theta}, \overline{\theta}]} \overline{\gamma}(\theta) \left(\dot{x}(\theta) - \dot{\overline{x}}(\theta)\right) d\theta \\ &= (x(1) - \overline{x}(1)) - \int_{(\underline{\theta}, \overline{\theta}]} \overline{\gamma}(\theta) \left(\dot{x}(\theta) - \dot{\overline{x}}(\theta)\right) d\theta. \end{aligned}$$

It follows that $\Delta \geq 0$ if

$$(x(1) - \overline{x}(1)) - \int_{(\underline{\theta}, \overline{\theta}]} \overline{\gamma}(\theta) \left(\dot{x}(\theta) - \dot{\overline{x}}(\theta)\right) d\theta \geq 0.$$

If $\overline{\gamma}$ were absolutely continuous, we would be able to integrate the second term by parts and reach such a conclusion. Because $\overline{\gamma}$ has possibly countable upward discontinuities, we must proceed more carefully. Note that $\overline{\gamma}$ is non-decreasing on $(\underline{\theta}, \overline{\theta}]$ with at most a countable number of upward jump discontinuities. $\overline{\gamma}$ is thus the sum of a countable number of singular jump functions plus a measure $d\mu(\theta)$ which is absolute continuous with respect to the Lebesgue measure and thus write as $d\mu(\theta) = \nu(\theta)d\theta$.[36] Denote the set of jump discontinuities by $\{\tau_1, \tau_2, \ldots\}$, a possibly infinite but countable set. Let $\mathcal{I}$ be the index set of τ_i and let define the size of the jump discontinuity at any τ_i by $\Delta\mu(\tau_i) = \overline{\gamma}(\tau_i^+) - \overline{\gamma}(\tau_i) > 0$. We thus write:

$$\overline{\gamma}(\theta) = \sum_{\tau_i \leq \theta, i \in \mathcal{I}} \Delta\mu(\tau_i) + \int_{\underline{\theta}}^{\theta} \nu(\theta) d\theta.$$

[36]Royden (1988).

Since $\overline{\gamma}$ is absolutely continuous outside the discontinuities, we can integrate by parts between any pair of discontinuities. Also note that at any such upward jump point, τ, $\overline{\gamma}$ is left- and right-continuous with $\overline{\gamma}(\tau) < \overline{\gamma}(\tau^+)$ and (by condition (B1)) we have $\overline{x}(\tau^+) = 0$.

Between any two points τ_i and τ_{i+1}, we know

$$\begin{aligned}\int_{(\tau_i,\tau_{i+1}]} \overline{\gamma}(\theta)\left(\dot{u}(\theta) - \dot{\overline{v}}(\theta)\right) d\theta \\ &= \overline{\gamma}(\theta)(x(\theta) - \overline{x}(\theta))|_{t=\tau_i^+}^{\tau_{i+1}} - \int_{(\tau_i,\tau_{i+1})} (x(\theta) - \overline{x}(\theta))\nu(\theta)d\theta \\ &= \overline{\gamma}(\tau_{i+1})(x(\tau_{i+1}) - \overline{x}(\tau_{i+1})) - \overline{\gamma}(\tau_i^+)(x(\tau_i) - \overline{x}(\tau_i)) \\ &\quad - \int_{(\tau_i,\tau_{i+1})} (x(\theta) - \overline{x}(\theta))\nu(\theta)d\theta.\end{aligned}$$

The second equality above uses the fact that x and $\overline{x}$ are continuous on Θ.

Then we may write

$$\begin{aligned}&\int_{(\underline{\theta},\overline{\theta}]} \overline{\gamma}(\theta)\left(\dot{x}(\theta) - \dot{\overline{x}}(\theta)\right) d\theta \\ &= \sum_{i\in\mathcal{I}} \overline{\gamma}(\tau_{i+1})(x(\tau_{i+1}) - \overline{x}(\tau_{i+1})) - (\Delta\mu(\tau_i) + \overline{\gamma}(\tau_i))(x(\tau_i) - \overline{x}(\tau_i)) \\ &- \int_{(\tau_i,\tau_{i+1})} (x(\theta) - \overline{x}(\theta))\nu(\theta)d\theta \\ &\qquad = (x(1) - \overline{x}(1)) - \sum_{i\in\mathcal{I}} \Delta\mu(\tau_i)(x(\tau_i) - \overline{x}(\tau_i)) - \int_{(\tau_i,\tau_{i+1})} (x(\theta) - \overline{x}(\theta))\nu(\theta)d\theta.\end{aligned}$$

By complementary slackness in condition (B1), we know $\overline{x}(\theta)\nu(\theta) = 0$ and at any jump point τ_i we must have $\overline{x}(\tau_i) = 0$. Thus,

$$\int_{(\underline{\theta},\overline{\theta}]} \overline{\gamma}(\theta)\left(\dot{x}(\theta) - \dot{\overline{x}}(\theta)\right) d\theta = (x(1) - \overline{x}(1)) - \sum_{i\in\mathcal{I}} \Delta\mu(\tau_i)x(\tau_i) - \int_{(\tau_i,\tau_{i+1})} x(\theta)\nu(\theta)d\theta.$$

We deduce

$$\begin{aligned}\Delta &\geq (x(1) - \overline{x}(1)) - \int_{(\underline{\theta},\overline{\theta}]} \overline{\gamma}(\theta)\left(\dot{x}(\theta) - \dot{\overline{x}}(\theta)\right) d\theta \\ &= \sum_{i\in\mathcal{I}} \Delta\mu(\tau_i)x(\tau_i) + \int_{(\tau_i,\tau_{i+1})} x(\theta)\nu(\theta)d\theta.\end{aligned}$$

Because $x(\theta) \geq 0$, μ is a non-negative measure, and jump discontinuities $\Delta\mu(\tau_i)$ are positive, we conclude that $\Delta \geq 0$ as claimed. We have proven that conditions (B1) and (B2) are sufficient for a solution.

Proof of Proposition 1: Smoothness of the Solution $\overline{x}$. We add the hypothesis that

$$\mathcal{V}(\theta, \sigma) \equiv \underset{v\in\mathbb{R}}{\arg\max}\ s(\theta, v) + (F(\theta) - \sigma)v$$

is single-valued and continuous for $(\theta, \sigma) \in \Theta \times [0, 1]$. It follows that $\mathcal{V}(\theta, \sigma)$ is non-increasing in σ and from condition (B2), that $\dot{\overline{x}}(\theta) = v(\theta, \overline{\gamma}(\theta))$ a.e.

Suppose to the contrary that $\dot{\overline{x}}$ is discontinuous at some point $\tau \in \Theta$. Initially, suppose that Condition (B2) is extended to hold for all $\theta \in \Theta$ rather than for a.e. $\theta \in (\underline{\theta}, \overline{\theta}]$; call this Condition (B2'). Condition (B2') and the additional hypothesis that $\mathcal{V}(\theta, \sigma)$ is continuous in (θ, σ) jointly imply that $\dot{\overline{x}}(\theta)$ is discontinuous at τ only if $\overline{\gamma}$ is also discontinuous at τ. Any discontinuity in $\overline{\gamma}$, however, must be an upward jump, $\overline{\gamma}(\tau^+) - \overline{\gamma}(\tau) > 0$, implying that $\dot{\overline{x}}(\theta)$ must jump downwards. Complementary slackness (Condition (B1), however, imposes that $\overline{x}(\tau) = 0$, with the implication that a downward discontinuity at τ would violate the state constraint $u(\theta) \geq 0$ in the neighborhood to the immediate right of τ. Hence, continuity must hold for all points $\theta \in [\underline{\theta}, \theta)$ under Condition (B2'). Furthermore, because $\overline{\gamma}$ is left continuous at $t = 1$, no jump in $\dot{\overline{x}}(\theta)$ is possible at this endpoint. We conclude that Condition (B2') implies that $\dot{\overline{x}}(\theta)$ is continuous for all $\theta \in \Theta$. The weaker Condition (B2) allows $\dot{\overline{x}}(\theta)$ to violate the maximization condition on sets of measure zero, including at $\theta = \underline{\theta}$. But such violations have no effect on the solution $\overline{x}$ which is absolutely continuous. Thus, $\overline{x}$ is smooth as posited.

APPENDIX C: PROOFS FOR APPLICATIONS

Proofs of Propositions 2 and 3: First, we define $\tilde{\theta}$, θ_1, θ_1 and θ_3 such that

(C1) $$q^m(\tilde{\theta}) = \hat{q}(\tilde{\theta}) \Leftrightarrow \tilde{\theta} = \underline{\theta} + \Delta,$$

(C2) $$\hat{q}(\theta_1) = 1 + \sqrt{2k} - \gamma_1,$$

(C3) $$\hat{q}(\theta_2) = 1 + \frac{\sqrt{2k}}{2} - \gamma_1,$$

(C4) $$\hat{q}(\theta_3) = 1 + \sqrt{2k} - 2\gamma_1,$$

where

(C5) $$\gamma_1 = \sqrt{k}.$$

Observe that, for k small enough, $\tilde{\theta} < \theta_1 < \overline{\theta}$ when Condition (3.9) is satisfied. It is also immediate to check that $\theta_1 < \theta_2 < \theta_3$ always holds. Moreover, for $k = 0$, we have

(C6) $$\theta_1 = \theta_2 = \theta_3$$

so that, $\theta_3 < \overline{\theta}$ also holds for k small enough and Condition (3.9) is satisfied.

Starting from the expression of $s(\theta, v)$, we now compute

(C7) $$\overline{\text{co}}(s)(\theta, v) =$$

$$\begin{cases} (S-\theta)(\hat{q}(\theta) - v) - \frac{1}{2}(\hat{q}(\theta) - v)^2 & \text{if } v \geq v_2(\theta), \\ (S-\theta)(\hat{q}(\theta) - v_1(\theta)) - \frac{1}{2}(\hat{q}(\theta) - v_1(\theta))^2 - k - (\Delta + v_1(\theta))(v - v_1(\theta)) & \text{if } v \in [v_1(\theta), v_2(\theta)], \\ (S-\theta)(\hat{q}(\theta) - v) - \frac{1}{2}(\hat{q}(\theta) - v)^2 - k & \text{if } v < v_1(\theta) \end{cases}$$

where $v_1(\theta) = \hat{q}(\theta) - 1 - \sqrt{2k}$ and $v_2(\theta) = \hat{q}(\theta) - 1$.

From there, we obtain the expression of the sub-differential

$$\partial_v \overline{\text{co}}(s)(\theta, v) = \begin{cases} -\Delta - v & \text{if } v \leq v_1(\theta) \text{ and } v \geq v_2(\theta), \\ -\Delta - v_1(\theta) & \text{if } v \in [v_1(\theta), v_2(\theta)], \\ [-\Delta - v_2(\theta), -\Delta - v_1(\theta)] & \text{if } v = v_2(\theta). \end{cases}$$

We now check that the pair $(\dot{\overline{u}} = \hat{q} - \overline{q}, \overline{\gamma})$ where $\overline{q}$ is defined in (3.10) and $\overline{\gamma}$ in (3.11) satisfies the necessary and sufficient conditions for optimality of Theorem 1. The corresponding conjecture is that (2.1) is binding on $[\tilde{\theta}, \theta_1] \cup [\theta_3, \overline{\theta}]$ and slack elsewhere. The new adjoint function $\overline{\gamma}$ is thus similar to $\overline{\gamma}_0$ on $[\underline{\theta}, \theta_1)$ but it might also have an upward jump at the point θ_1. This jump is followed by a plateau.

Inserting this conjecture into (3.8) and using (C7) yields the following conditions.

- If $\theta \in [\underline{\theta}, \tilde{\theta})$, (2.1) slack,

  (C8) $\qquad \overline{\gamma}(\theta) = 0$

  and thus

  (C9) $\qquad -\theta + \underline{\theta} = \partial_v \overline{\text{co}}(s)(\theta, \dot{\overline{u}}(\theta)) = -\Delta - \dot{\overline{u}}(\theta) \Leftrightarrow \overline{q}(\theta) = q^m(\theta).$

  For $\partial_v \overline{\text{co}}(s)(\theta, \dot{\overline{u}}(\theta))$ to be equal at $-\Delta - \dot{\overline{u}}(\theta)$, a sufficient condition is that $\dot{\overline{u}}(\theta) \leq v_1(\theta)$ or

  (C10) $\qquad -\theta + \underline{\theta} + \Delta \geq -v_1(\theta) = -\hat{S} + \theta + 1 + \sqrt{2k} \Leftrightarrow q^m(\theta) \geq 1 + \sqrt{2k}.$

  Observe that $\theta \leq \tilde{\theta}$ means $q^m(\theta) \geq \hat{q}(\theta)$ and thus (C10) holds when $\tilde{\theta} \leq \theta_1$.

- If $\theta \in [\tilde{\theta}, \theta_1]$, (2.1) is binding and, by differentiating, we get $\dot{\overline{u}}(\theta) = 0$ on $(\tilde{\theta}, \theta_1)$. Take now

  (C11) $\qquad \overline{\gamma}(\theta) = \theta - \tilde{\theta}$

  which is positive and non-decreasing. Inserting into (3.8), we obtain

  $$0 \in \partial_v \overline{\text{co}}(s)(\theta, 0).$$

  If $\overline{\text{co}}(s)(\theta, 0) = -\Delta - \dot{\overline{u}}(\theta)$, this condition becomes

  (C12) $\qquad \overline{q}(\theta) = \hat{q}(\theta).$

  A sufficient condition for writing $\partial_v \overline{\text{co}}(s)(\theta, \dot{\overline{u}}(\theta)) = -\Delta - \dot{\overline{u}}(\theta)$ as above is that

  $$\dot{\overline{u}}(\theta) = 0 \leq v_1(\theta) \Leftrightarrow \hat{q}(\theta) \geq 1 + \sqrt{2k}$$

  which means $\theta \leq \theta_1$. Because (2.1) is binding at θ_1, we allow for some positive charge there and we denote

  $$\overline{\gamma}(\theta_1) = \theta_1 - \tilde{\theta} + \gamma_1.$$

- If $\theta \in (\theta_1, \theta_3)$, (2.1) is slack and

  (C13) $\qquad \overline{\gamma}(\theta) = \theta_1 - \tilde{\theta} + \gamma_1$

  on this interval. Inserting into (3.8), we obtain:

  (C14) $\qquad \theta_1 - \tilde{\theta} + \gamma_1 - \theta + \underline{\theta} \in \partial_v \overline{\text{co}}(s)(\theta, \dot{\overline{u}}(\theta)).$

  We now consider two sub-cases.

– For $\theta \in (\theta_1, \theta_2]$, we have $\partial_v \overline{\text{co}}(s)(\theta, \dot{\overline{u}}(\theta)) \in [-\Delta - v_2(\theta), -\Delta - v_1(\theta)]$ which means

(C15) $\quad \dot{\overline{u}}(\theta) = v_2(\theta) \Leftrightarrow \overline{q}(\theta) = 1.$

Condition (C14) becomes

$$\theta_1 - \tilde{\theta} + \gamma_1 - \theta + \underline{\theta} \in [-\Delta - v_2(\theta), -\Delta - v_1(\theta)] \Leftrightarrow \theta - \theta_1 - \gamma_1 \in [v_1(\theta), v_2(\theta)]$$

or $\theta \in [\theta_1, \theta_2]$, which holds for this sub-case.

– For $\theta \in (\theta_2, \theta_3]$, we have $\partial_v \overline{\text{co}}(s)(\theta, \dot{\overline{u}}(\theta)) = -\Delta - \dot{\overline{u}}(\theta)$. Inserting into (C14) yields

(C16) $\quad \theta_1 - \tilde{\theta} + \gamma_1 - \theta + \underline{\theta} = -\Delta - \dot{\overline{u}}(\theta) \Leftrightarrow \overline{q}(\theta) = q^m(\theta) + \theta_1 - \tilde{\theta} + \gamma_1.$

A sufficient condition for writing $\partial_v \overline{\text{co}}(s)(\theta, \dot{\overline{u}}(\theta)) = -\Delta - \dot{\overline{u}}(\theta)$ as above is that $\dot{\overline{u}}(\theta) \geq v_2(\theta)$ which amounts to

$$\theta_1 - \theta \leq -v_2(\theta)$$

or $\theta \geq \theta_2$ which holds for this sub-case.

- If $\theta \in [\theta_3, \overline{\theta}]$, (2.1) is binding and, by differentiating, we get $\dot{\overline{u}}(\theta) = 0$ on $(\theta_3, \overline{\theta})$. From (3.8), we obtain:

$$\overline{\gamma}(\theta) - \theta + \underline{\theta} = \partial_v \overline{\text{co}}(s)(\theta, 0) = -\Delta.$$

Thus, on $[\theta_3, \overline{\theta}]$, we have:

(C17) $\quad \overline{\gamma}(\theta) = \theta - \underline{\theta}.$

The parameter γ_1 is chosen so that $\overline{u}(\theta_1) = \overline{u}(\theta_3) = 0$. This condition writes also as

$$0 = \int_{\theta_1}^{\theta_3} \dot{\overline{u}}(\theta) d\theta = \int_{\theta_1}^{\theta_3} (\hat{q}(\theta) - \overline{q}(\theta)) d\theta = 0.$$

Integrating by parts, we get

$$0 = \int_{\theta_1}^{\theta_3} (\dot{\hat{q}}(\theta) - \dot{\overline{q}}(\theta))(\theta - \theta_1) d\theta = 0.$$

We rewrite this condition as

$$0 = \int_{\theta_1}^{\theta_2} -(\theta - \theta_1) d\theta + \int_{\theta_2}^{\theta_3} (\theta - \theta_1) d\theta$$

or

$$2(\theta_2 - \theta_1)^2 = (\theta_3 - \theta_1)^2.$$

Geometrically, this condition just says that the algebraic area between the curves $\hat{q}$ and $\overline{q}$ is zero over $[\theta_1, \theta_3]$. Using the definitions (C2), (C3), and (C4) yields

$$\gamma_1^2 = k$$

or (C5).

Summarizing our previous findings in (C9), (C12), (C15) and (C16) yields the expression of $\overline{q}$ in (3.10). On the other hand, (C8), (C11) and (C17) yields the expression of $\overline{\gamma}$ (3.11).

Observe that Propositon 2 immediately follows from the general case when observing that (C6) holds when $k = 0$.

Q.E.D.

PROOF OF PROPOSITION 4: Standard arguments (see footnote 2) establish that $U(\theta)$ so defined is absolutely continuous and thus a.e. differentiable with

$$\text{(C18)} \qquad \dot{U}(\theta) = \begin{cases} -q(\theta) - k & \text{if } q(\theta) < \hat{q}(\theta), \\ -\hat{q}(\theta) - k & \text{if } q(\theta) \in (\hat{q}(\theta), \hat{q}(\theta) + k), \\ -q(\theta) & \text{if } q(\theta) > \hat{q}(\theta) + k. \end{cases}$$ [37]

From this and the fact that $\dot{\hat{U}}(\theta) = -k$, we get (3.14).

We can express $\tilde{s}(\theta, q)$ as

$$\tilde{s}(\theta, q) = \begin{cases} (S - \theta - k)q - \frac{q^2}{2} - cq & \text{if } q \leq \hat{q}(\theta), \\ \frac{(\hat{q}(\theta)+k)^2}{2} + \Delta q - \left((S - \theta)k - \frac{k^2}{2} - pk\right) & \text{if } q \in [\hat{q}(\theta), \hat{q}(\theta) + k], \\ (S - \theta - c)q - \frac{q^2}{2} - \left((S - \theta)k - \frac{k^2}{2} - pk\right) & \text{if } q \geq \hat{q}(\theta) + k. \end{cases}$$

Observing that $\hat{q}(\theta) + k = S - \theta - p$ and simplifying yields

$$\tilde{s}(\theta, q) = \begin{cases} (\hat{q}(\theta) + \Delta)q - \frac{q^2}{2} & \text{if } q \leq \hat{q}(\theta), \\ \frac{\hat{q}(\theta)^2}{2} + \Delta q & \text{if } q \in [\hat{q}(\theta), \hat{q}(\theta) + k], \\ (S - \theta - c)q - \frac{q^2}{2} - \left((S - \theta - p)k - \frac{k^2}{2}\right) & \text{if } q \geq \hat{q}(\theta) + k. \end{cases}$$

Expressing $q(\theta)$ in terms of $\dot{u}(\theta)$ over the different intervals yields

$$\text{(C19)} \qquad q(\theta) \begin{cases} = -\dot{u}(\theta) & \text{if } \dot{u}(\theta) \geq -\hat{q}(\theta), \\ \in [\hat{q}(\theta), \hat{q}(\theta) + k] & \text{if } \dot{u}(\theta) = -\hat{q}(\theta), \\ -\dot{u}(\theta) + k & \text{if } \dot{u}(\theta) \leq -\hat{q}(\theta). \end{cases}$$

Inserting these expressions of $q(\theta)$ into the expression of $s(\theta, v)$ yields (3.15).

From there, we now compute

$$\overline{co}(s)(\theta, v) =$$

$$\begin{cases} (-\hat{q}(\theta) - \Delta)v - \frac{v^2}{2} & \text{if } v \geq v_2(\theta), \\ -(\sqrt{2\Delta k} + \Delta)(v - v_2(\theta)) + (-\hat{q}(\theta) - \Delta)v_2(\theta) - \frac{v_2(\theta)^2}{2} & \text{if } v \in [-\hat{q}(\theta), v_2(\theta)], \\ (-\hat{q}(\theta) - \Delta)v - \frac{v^2}{2} + \Delta k & \text{if } v < -\hat{q}(\theta) \end{cases}$$

where $v_2(\theta) = \sqrt{2\Delta k} - \hat{q}(\theta)$.

This yields the following expression of the sub-differential for $\overline{co}(s)(\theta, v)$:

$$\partial_v \overline{co}(s)(\theta, v) = \begin{cases} -\hat{q}(\theta) - \Delta - v & \text{if } v \geq v_2(\theta), \\ -\sqrt{2\Delta k} - \Delta & \text{if } v \in (-\hat{q}(\theta), v_2(\theta)], \\ [-\sqrt{2\Delta k} - \Delta, -\Delta] & \text{if } v = -\hat{q}(\theta), \\ -\hat{q}(\theta) - \Delta - v & \text{if } v < -\hat{q}(\theta). \end{cases}$$

With a uniform distribution, the optimality condition (2.5) becomes

$$\overline{\gamma}(\theta) - \theta + \underline{\theta} \in \partial_v \overline{\text{co}}(s)(\theta, \dot{\overline{u}}(\theta))$$

or

$$\text{(C20)} \quad \overline{\gamma}(\theta) - \theta + \underline{\theta} \begin{cases} = -\hat{q}(\theta) - \Delta - \dot{u}(\theta) & \text{if } \dot{u}(\theta) \geq v_2(\theta), \\ = -\sqrt{2\Delta k} - \Delta & \text{if } \dot{u}(\theta) \in [-\hat{q}(\theta), v_2(\theta)], \\ \in [-\sqrt{2\Delta k} - \Delta, -\Delta] & \text{if } \dot{u}(\theta) = -\hat{q}(\theta), \\ = -\hat{q}(\theta) - \Delta - \dot{u}(\theta) & \text{if } \dot{u}(\theta) < -\hat{q}(\theta). \end{cases}$$

We conjecture a solution $(\overline{u}(\theta), \overline{\gamma}(\theta))$ such that (2.1) binds at $\overline{\theta}$ only and thus $\mu(\{\overline{\theta}\}) > 0$ with $\overline{\gamma}(\theta) = 0$ on $[\underline{\theta}, \overline{\theta})$. Thanks to the sufficiency part of our Theorem, we only check that this solution satisfies the necessary conditions for optimality.

- On the interval $[\underline{\theta}, \tilde{\theta})$, this conjecture implies $\overline{\gamma}(\theta) = 0$. Inserting into (C20) yields

  $$-\theta + \underline{\theta} = \partial_v \overline{co}(s)(\theta, \dot{\overline{u}}(\theta)) = -\hat{q}(\theta) - \Delta - \dot{\overline{u}}(\theta).$$

  Because $\theta \leq \tilde{\theta} = \underline{\theta} + \Delta$, we thus have

  $$\dot{\overline{u}}(\theta) = -\hat{q}(\theta) + \theta - \tilde{\theta} < -\hat{q}(\theta).$$

  From (C19), we deduce that

  $$-\overline{q}(\theta) + k = \dot{\overline{u}}(\theta) = -(S - \theta - p - k) + \theta - \tilde{\theta}$$

  and thus

  $$\overline{q}(\theta) = S - 2\theta + \underline{\theta} - c = q^m(\theta).$$

- On the interval $[\tilde{\theta}, \tilde{\theta}_0]$, we have

  $$\dot{\overline{u}}(\theta) = -\hat{q}(\theta).$$

  Indeed, imposing our conjecture $\overline{\gamma}(\theta) = 0$ on (C20) yields

  $$-\theta + \underline{\theta} \in \partial_v \overline{co}(s)(\theta, -\hat{q}(\theta)) = [-\sqrt{2\Delta k} - \Delta, -\Delta] \Leftrightarrow \theta \in [\tilde{\theta}, \tilde{\theta}_0].$$

  From (C19), we deduce that

  $$\overline{q}(\theta) \in [S - \theta - p - k, S - \theta - p] = [\hat{q}(\theta), \hat{q}(\theta) + k].$$

- On the interval $[\tilde{\theta}_0, \overline{\theta})$, our conjecture is $\overline{\gamma}(\theta) = 0$. Inserting into (C20) yields

  $$-\theta + \underline{\theta} = \partial_v \overline{co}(s)(\theta, \dot{\overline{u}}(\theta)) = -\hat{q}(\theta) - \dot{\overline{u}}(\theta) - \Delta.$$

  Because $\theta \geq \tilde{\theta}_0 \geq \tilde{\theta}$, we have

  $$\dot{\overline{u}}(\theta) = -\hat{q}(\theta) + \theta - \tilde{\theta} > -\hat{q}(\theta) + \theta - \tilde{\theta}_0 \geq -\hat{q}(\theta).$$

  From (C19), we deduce that

  $$-\overline{q}(\theta) = \dot{\overline{u}}(\theta) = -(S - \theta - p - k) + \theta - \tilde{\theta}$$

  and thus

  $$\overline{q}(\theta) = S - 2\theta + \underline{\theta} - c - k = q^m(\theta) - k.$$

Gathering all the above findings yields (3.16). The condition (3.13) ensures that the incumbent firm still wants to serve the type with the lowest possible demand $\overline{\theta}$ even when that type consumes up to the fringe's capacity. It implies that $\overline{u}(\theta)$ is everywhere decreasing, consistently with (2.1) being binding at $\overline{\theta}$ only. Hence, $\mu(\{\overline{\theta}\}) = 1$ as required. *Q.E.D.*

Proof of Propositions 6 and 7: We first explain in some details why all equilibria are fully characterized as solutions to self-generating problems.

The Aggregate Concurrence Principle. Putting aside the monotonicity condition for $q(\theta)$ that is checked *ex post* on the solution, we may write principal P_i's optimization problem at a Nash equilibrium, taking as given the aggregate offers made by other principals $\overline{\mathcal{T}}_{-i}$, as:

$$(\mathcal{P}_i): \qquad \max_{q(\theta)\in\mathcal{Q},u(\theta)} \int_{\underline{\theta}}^{\overline{\theta}} [S_i(q(\theta)) - \theta(q(\theta) - K) + \overline{\mathcal{T}}_{-i}(q(\theta)) - u(\theta)]d\theta,$$

subject to (2.1) and (3.19).

Any equilibrium comprises an allocation $(\overline{u}(\theta), \overline{q}(\theta))$ which is a solution to all problems $(\mathcal{P}_i)$ above. It is necessarily also a solution to the *self-generating problem* $(\mathcal{P})$ which is simply obtained by taking the sum of the corresponding maximands. In particular, $\overline{\mathcal{T}}$ implements the allocation $(\overline{u}(\theta), \overline{q}(\theta))$.

Reciprocally, consider a solution $(\overline{u}(\theta), \overline{q}(\theta))$ to the self-generating problem $(\mathcal{P})$ which is implemented by the aggregate schedule $\overline{\mathcal{T}}$. We show that there exists an array $\overline{\mathbf{T}} = (\overline{T}_1, ..., \overline{T}_n)$ of schedules whose aggregate is $\overline{\mathcal{T}}$ and which are best responses to each other. Consider for all j, the tariff $\overline{T}_j$ defined as

$$\text{(C21)} \qquad \frac{1}{n}\left(\mathcal{S}(q) - \overline{\mathcal{T}}(q)\right) = S_j(q) - \overline{T}_j(q) \quad \forall q \in \mathcal{Q}.$$

First, those tariffs' aggregate is clearly $\overline{\mathcal{T}}$. Second, inserting the expressions of $\overline{T}_j$ (for $j \neq i$) into the maximand of $(\mathcal{P}_i)$, P_i's payoff of implementing an allocation $(u(\theta), q(\theta))$ at a best response to $\overline{\mathbf{T}}_{-i}$ writes as

$$\frac{1}{n}\left(\mathcal{S}(q) - \theta(q(\theta) - K) + (n-1)(\overline{\mathcal{T}}(q(\theta)) - \theta(q(\theta) - K)) - nu(\theta)\right).$$

Hence, $(\mathcal{P}_i)$ takes the same form as $(\mathcal{P})$ up to a multiplicative factor $\frac{1}{n}$. P_i's objective has thus been aligned with that of the surrogate principal. As a result, P_i chooses to implement the same allocation $(\overline{u}(\theta), \overline{q}(\theta))$ as this surrogate principal. Given that $(\overline{u}(\theta), \overline{q}(\theta))$ is a solution to the *self-generating problem* $(\mathcal{P})$, $(\overline{u}(\theta), \overline{q}(\theta))$ is implemented by the aggregate $\overline{\mathcal{T}}$. It means that, P_i could be as well off offering a tariff T_i such that the aggregate remains $\overline{\mathcal{T}}$. In other words, P_i's best-response correspondance contains a tariff T_i that satisfies

$$\overline{T}_i(q) = \overline{\mathcal{T}}(q) - \sum_{j\neq i} \overline{T}_j(q)$$

Because $\overline{T}_j$ for $j \neq i$ is given by (C21), we have

$$\overline{T}_i(q) = \overline{\mathcal{T}}(q) - \sum_{j\neq i}\left(S_j(q) - \frac{1}{n}\left(\mathcal{S}(q) - \overline{\mathcal{T}}(q)\right)\right) = S_i(q) - \frac{1}{n}\left(\mathcal{S}(q) - \overline{\mathcal{T}}(q)\right).$$

Thus (C21) holds also for P_i.

Discontinuities at points where (2.1) does not bind. Consider the output profile as in (3.33).[38] By means of the incentive compatibility condition (3.19) and the binding participation constraint (2.1) over the interval $[\theta_1^n, \theta_2^n]$, we can thus reconstruct a rent profile $\overline{u}(\theta)$ which is identical to $\overline{u}^m(\theta)$ over $[\theta_4, \overline{\theta}]$ but (possibly) differs on $[\underline{\theta}, \theta_4]$. More precisely, we get:

$$\text{(C22)} \quad \overline{u}(\theta) = \begin{cases} \overline{u}^m(\theta) & \text{if } \theta \in [\theta_4, \overline{\theta}], \\ \overline{u}^m(\theta_3) + \int_\theta^{\theta_3} (\overline{q}(\tilde{\theta}) - k) d\tilde{\theta} & \text{if } \theta \in [\underline{\theta}, \theta_4]. \end{cases}$$

From (3.33) and (3.19), $\overline{u}^m$ is linear on the segments $[\theta_3, \theta_0]$ and $[\theta_0, \theta_4]$ with respective slopes $k - \overline{q}^m(\theta_3)$ and $k - \overline{q}^m(\theta_4)$. $\overline{u}^m$ has thus a kink at θ_0 with a subdifferential there defined as $\partial \overline{u}^m(\theta_0) = [k - \overline{q}^m(\theta_3), k - \overline{q}^m(\theta_3)]$.

By means of a duality argument similar to (3.28), we can now also reconstruct an aggregate tariff as

$$\text{(C23)} \quad \overline{\mathcal{T}}(q) = \min_{\theta \in \Theta} \overline{u}(\theta) + \theta(q - k).$$

Tedious computations show that $\overline{\mathcal{T}}(q)$ satisfies (3.34). In particular, observe that θ_0 is indifferent between choosing $\overline{q}^m(\theta_3)$ and $\overline{q}^m(\theta_4)$ which means that (3.32) holds. Taken in tandem, (3.31) and (3.32) show that the surrogate principal's objective function is kept constant at type θ_0 when moving from $\overline{q}^m(\theta_3)$ to $\overline{q}^m(\theta_4)$.

Let us check that we have defined an equilibrium. To do so, we must verify that $\overline{u}$ as defined in (C22) is a solution to the self-generating problem where $\overline{\mathcal{T}}$ satisfies (3.35). To this end, we define the surplus function again as in (3.20) and check that the following necessary conditions (2.4) and (2.5) hold.

We first compute

$$n\overline{co}(s)(\theta, v) =$$

$$\begin{cases} \mathcal{S}(k-v) + \theta v + (n-1)(\overline{\mathcal{T}}^m(k-v) + \theta v) \text{ if } k - v \in [\overline{q}^m(\overline{\theta}), \overline{q}^m(\theta_4)) \cup (\overline{q}^m(\theta_3), \overline{q}^m(\underline{\theta})], \\ \mathcal{S}(k-v) + \theta v + (n-1)(\overline{\mathcal{T}}^m(\overline{q}^m(\theta_4)) + \theta_0(k - v - \overline{q}^m(\theta_4)) + \theta v) \text{ if } k - v \in [\overline{q}^m(\theta_4), \overline{q}^m(\theta_3)) \end{cases}$$

and thus

$$n\partial_v \overline{co}(s)(\theta, v) = \begin{cases} -\mathcal{S}'(k-v) + \theta \text{ if } k - v \in [\overline{q}^m(\overline{\theta}), \overline{q}^m(\theta_4)) \cup (\overline{q}^m(\theta_3), \overline{q}^m(\underline{\theta})], \\ -\mathcal{S}'(k-v) + \theta + (n-1)(\theta - \theta_0) \text{ if } k - v \in (\overline{q}^m(\theta_4), \overline{q}^m(\theta_3)), \\ [(n-1)(\theta_4 - \theta_0) - n(\theta_4 - \underline{\theta}), -n(\theta_4 - \underline{\theta})] \text{ if } k - v = \overline{q}^m(\theta_4), \\ [-n(\theta_3 - \underline{\theta}), (n-1)(\theta_3 - \theta_0) - n(\theta_3 - \underline{\theta})] \text{ if } k - v = \overline{q}^m(\theta_3). \end{cases}$$

First, we notice that the output profile (3.33) satisfies (2.4) as requested by the necessary conditions.

Second, we rewrite (2.5) as

$$\text{(C24)} \quad \overline{\gamma}^m(\theta) - (\theta - \underline{\theta}) \in \partial\overline{co}(s)(\theta, \dot{\overline{u}}(\theta)) \ \text{ a.e. } \theta \in \Theta$$

where the adjoint function $\overline{\gamma}^m(\theta)$ is the same as in the "smooth" scenario, i.e., satisfies (3.26).

For points where $\partial_v \overline{co}(s)(\theta, \dot{\overline{u}}(\theta))$ is single valued, the optimality condition (C24) is the same as in the "smooth" scenario and $\overline{q}^m(\theta)$ is thus the solution on $[\underline{\theta}, \theta_3] \cup [\theta_4, \overline{\theta}]$.

[38]Notice that we may have alternatively selected $\overline{q}^m(\theta_0) = \overline{q}^m(\theta_4)$ without changing any result.

If $\theta \in [\theta_3, \theta_0]$, $\overline{q}^m(\theta_3) = k - \dot{\overline{u}}(\theta)$ is the solution since then $\gamma^m(\theta) = 0$ and (C24), when evaluated at such point, becomes

$$-n(\theta - \underline{\theta}) \in [(n-1)(\theta_3 - \theta_0) - n(\theta_3 - \underline{\theta}), -n(\theta_3 - \underline{\theta})] \Leftrightarrow \theta \in [\theta_3, \theta_0]$$

which holds true.

A similar reasoning shows that, if $\theta \in (\theta_0, \theta_4]$, $\overline{q}^m(\theta_4) = k - \dot{\overline{u}}(\theta)$ is also the solution since

$$-n(\theta - \underline{\theta}) \in [(n-1)(\theta_4 - \theta_0) - n(\theta_4 - \underline{\theta}), -n(\theta_4 - \underline{\theta})] \Leftrightarrow \theta \in [\theta_0, \theta_4].$$

This ends the proof.

Discontinuities at points where (2.1) does bind. The equilibrium output is again given by (3.33). By means of the incentive compatibility condition (3.19) and the binding participation constraint (2.1) at θ_0, i.e.,

(C25) $\quad \overline{u}(\theta_0) = 0,$

we can thus reconstruct a rent profile $\overline{u}(\theta)$ as

(C26) $\quad \overline{u}(\theta) = \int_{\theta_0}^{\theta} (k - \overline{q}(\tilde{\theta}))d\tilde{\theta}.$

It is immediate to check that

(C27) $$\overline{u}(\theta) = \begin{cases} (k - \overline{q}^m(\theta_4))(\theta - \theta_0) & \text{if } \theta \in [\theta_0, \theta_4], \\ \overline{u}^m(\theta) - \overline{u}^m(\theta_4) + (k - \overline{q}^m(\theta_4))(\theta_4 - \theta_0) & \text{if } \theta \in [\theta_4, \overline{\theta}], \\ \overline{u}^m(\theta) - \overline{u}^m(\theta_3) + (\overline{q}^m(\theta_3) - k)(\theta_0 - \theta_3) & \text{if } \theta \in [\underline{\theta}, \theta_3], \\ (\overline{q}^m(\theta_3) - k)(\theta_0 - \theta) & \text{if } \theta \in [\theta_3, \theta_0]. \end{cases}$$

Observing that $\theta_3 < \theta_1^n$ and $\theta_4 > \theta_2^n$ so that $\overline{q}^m(\theta_3) > k > \overline{q}^m(\theta_4)$ as requested by our assumption (3.37), implies $\overline{u}^m(\theta_3) = \int_{\theta_1^n}^{\theta}(k - \overline{q}(\tilde{\theta}))d\tilde{\theta}$ and $\overline{u}^m(\theta_4) = \int_{\theta_2^n}^{\theta}(k - \overline{q}(\tilde{\theta}))d\tilde{\theta}$.

It follows from (C27) that $\overline{u}$ is linear on $[\theta_3, \theta_0]$ and $[\theta_0, \theta_4]$ with respective slopes $\overline{q}^m(\theta_3)$ and $\overline{q}^m(\theta_4)$. Since $\overline{q}^m(\theta_3) > k > \overline{q}^m(\theta_4)$ by (3.37), $\overline{u}$ has thus a kink at θ_0 with a subdifferential there being defined as $\partial\overline{u}(\theta_0) = [k - \overline{q}^m(\theta_3), k - \overline{q}^m(\theta_4)]$. Observe that $0 \in \partial\overline{u}(\theta_0)$.

By means of a duality argument similar to (3.28), we can now also reconstruct an aggregate tariff as

(C28) $\quad \overline{\mathcal{T}}(q) = \min_{\theta \in \Theta} \overline{u}(\theta) + \theta(q - k).$

Tedious computations show that $\overline{\mathcal{T}}(q)$ satisfies (3.39). In particular, $\overline{\mathcal{T}}(q)$ is linear over $[\overline{q}^m(\theta_4), \overline{q}^m(\theta_3)]$. We also observe that θ_0 is indifferent over all $q \in [\overline{q}^m(\theta_3), \overline{q}^m(\theta_4)])$, making always zero profit. In particular, (3.32) holds.

Let us check that we have defined an equilibrium. To do so, we must verify that $\overline{u}$ as defined in (C27) is a solution to the self-generating problem where $\overline{\mathcal{T}}$ satisfies (3.39). It means checking that both (2.4) and (2.5) hold where the adjoint function $\overline{\gamma}^m(\theta)$ is now given by (3.36).

We first compute

$$n\overline{co}(s)(\theta, v) =$$

$$\begin{cases} \mathcal{S}(k-v)+\theta v+(n-1)(\overline{\mathcal{T}}^m(k-v)-\overline{\mathcal{T}}^m(\overline{q}^m(\theta_3))+\theta_0(\overline{q}^m(\theta_3)-k)+\theta v) \text{ if } k-v\in(\overline{q}^m(\theta_3),\overline{q}^m(\underline{\theta})], \\ \mathcal{S}(k-v)+\theta v+(n-1)(\theta-\theta_0)v \text{ if } k-v\in[\overline{q}^m(\theta_4),\overline{q}^m(\theta_3)], \\ \mathcal{S}(k-v)+\theta v+(n-1)(\overline{\mathcal{T}}^m(q)-\overline{\mathcal{T}}^m(\overline{q}^m(\theta_4))+\theta_0(\overline{q}^m(\theta_4)-k)+\theta v) \text{ if } k-v\in[\overline{q}^m(\overline{\theta}),\overline{q}^m(\theta_4)). \end{cases}$$

We thus obtain

$$n\partial_v\overline{co}(s)(\theta,v)=\begin{cases} -\mathcal{S}'(k-v)+\theta \text{ if } k-v\in[\overline{q}^m(\overline{\theta}),\overline{q}^m(\theta_4))\cup(\overline{q}^m(\theta_3),\overline{q}^m(\underline{\theta})], \\ -\mathcal{S}'(k-v)+\theta+(n-1)(\theta-\theta_0) \text{ if } k-v\in(\overline{q}^m(\theta_4),\overline{q}^m(\theta_3)) \\ [-\mathcal{S}'(\overline{q}^m(\theta_3))+\theta+(n-1)(\theta-\theta_0),-\mathcal{S}'(\overline{q}^m(\theta_3))+\theta] \text{ if } k-v=\overline{q}^m(\theta_3), \\ [-\mathcal{S}'(\overline{q}^m(\theta_4))+\theta,-\mathcal{S}'(\overline{q}^m(\theta_4))+\theta+(n-1)(\theta-\theta_0)] \text{ if } k-v=\overline{q}^m(\theta_4). \end{cases}$$

First, we again notice that the output profile (3.33) satisfies (2.4) as requested by the necessary conditions.

Second, (2.5) can now be expressed as

(C29) $\overline{\gamma}^m(\theta)-(\theta-\underline{\theta})\in\partial\overline{co}(s)(\theta,\dot{\overline{u}}(\theta))$ a.e. $\theta\in\Theta$

where the adjoint function $\overline{\gamma}^m(\theta)$ is the same as in the "smooth" scenario, i.e., satisfies (3.26).

For θ such that $k-\dot{\overline{u}}(\theta)\in[\overline{q}^m(\overline{\theta}),\overline{q}^m(\theta_4))\cup(\overline{q}^m(\theta_3),\overline{q}^m(\underline{\theta})]$, we have $n\partial_v\overline{co}(s)(\theta,\dot{\overline{u}}(\theta))=-\mathcal{S}'(k-\dot{\overline{u}}(\theta))+\theta$. The optimality condition (C29) is thus the same as in the "smooth" scenario and $\overline{q}^m(\theta)=k-\dot{\overline{u}}^m(\theta)$ is the solution on $[\underline{\theta},\theta_3]\cup[\theta_4,\overline{\theta}]$.

If $\theta\in[\theta_3,\theta_0)$, $\overline{q}^m(\theta_3)=k-\dot{\overline{u}}(\theta)$ is the solution since then $\overline{\gamma}^m(\theta)=0$,

$$n\partial_v\overline{co}(s)(\theta,k-\overline{q}^m(\theta_3))=\left[-\mathcal{S}'(\overline{q}^m(\theta_3))+\theta+(n-1)(\theta-\theta_0),-\mathcal{S}'(\overline{q}^m(\theta_3))+\theta\right]$$

and (C29), when evaluated at such point, becomes

$$-n(\theta-\underline{\theta})\in[-\theta_3-n(\theta_3-\underline{\theta})+\theta+(n-1)(\theta-\theta_0)-\theta_3-n(\theta_3-\underline{\theta})+\theta]$$

which holds true when $\theta\in[\theta_3,\theta_0)$.

If $\theta\in(\theta_0,\theta_4]$, $\overline{q}^m(\theta_4)=k-\dot{\overline{u}}(\theta)$ is the solution since then $\gamma^m(\theta)=1$ and

$$n\partial_v\overline{co}(s)(\theta,k-\overline{q}^m(\theta_4))=\left[-\mathcal{S}'(\overline{q}^m(\theta_4))+\theta,-\mathcal{S}'(\overline{q}^m(\theta_4))+\theta+(n-1)(\theta-\theta_0)\right]$$

and (C29) when evaluated at such point becomes

$$-n(\theta-\overline{\theta})\in[-\theta_4-n(\theta_4-\overline{\theta})+\theta,-\theta_4-n(\theta_4-\overline{\theta})+\theta+(n-1)(\theta-\theta_0)]$$

which again holds true when $\theta\in(\theta_0,\theta_4]$. This ends the proof.

Q.E.D.