Public Contracting in Delegated Agency Games

Martimort, David and Stole, Lars

Paris School of Economics, University of Chicago

17 August 2011

Working paper

PUBLIC CONTRACTING IN DELEGATED AGENCY GAMES[1]

August 17, 2011

David Martimort and Lars Stole

We study games of public delegated common agency under asymmetric information. Using tools from non-smooth analysis and optimal control, we derive best responses and characterize equilibria (both continuous and discontinuous) using *self-generating* optimization programs of which any equilibrium allocation must be a solution. Special attention is given to common agency games in which each principal's payoff is a linear function of the agent's action. In such games the self-generating optimization program reduces to the maximization of the principals' "aggregate" virtual surplus in which the agent's marginal valuation is replaced by a confluence of "virtual" valuations that reflect common agency problems. One noteworthy subset of equilibrium allocations are "virtually truthful" which are the incomplete-information generalization of Bernheim and Whinston's (1986) "truthful" equilibria. Virtually-truthful equilibria are simple to calculate and illustrate two distinct sources of equilibrium distortion: inefficient contracting by a given coalition of active principals and inefficient participation (insufficient activity) by principals. Our results are illustrated by means of two games: a public goods game in which each player simultaneously offers a menu contract to a common provider of the public good in order to induce greater supply, and a lobbying game between conflicting interest groups in which each group offers a menu of contributions to a common political decision-maker in an attempt to influence policymaking.

Keywords: Common agency, asymmetric information, menu auctions, delegated contracting games, public goods, lobbying.

1. INTRODUCTION

This paper considers a large class of multi-principal, common-agent games in which (i) the agent has private information at the time of contracting, (ii) any of the agent's actions may be contracted on by any principal (i.e., contracting variables are *public*), and (iii) the agent is free to accept any subset of contract offers from the principals (i.e., the degree of common agency is endogenous and *delegated* to the agent). Two motivating examples which we consider are the private provision of public goods and influence activities by lobbyists. In the public goods example, the principals are a set of citizens, each of who may offer a contract to a common, privately-informed producer to make a payment as a function of the total public good produced; all of the citizens agree that more public good is better, but the strength of each citizen's preference for the public good may differ. In the lobbying example, the set of principals correspond to lobbyists with (possibly opposite) preferences over a politician's decision; each

[1]We thank seminar participants in Zürich, Toulouse, the 5th CSEF-IGIER Symposium in Economic Theory in Capri, ESEM-Barcelona 2009, Bocconi University of Milan, Princeton, University of Pennsylvania, Paris School of Economics, Bologne, University College of London, Ohio State University, Montpellier, the CEPR Applied Industrial Organization 2010 and the Cowles Foundation Economic Theory Summer 2011 Conference for their helpful remarks. We are especially thankful to David Rahman for comments and to John Birge, Philippe Jehiel, Stephano Lovo and Aggey Semenov for many helpful discussions. Some results and applications have been taken from our previous paper, "Common Agency Games with Common Screening Devices," (2005).

[a]Paris School of Economics-EHESS, david.martimort@parisschoolofeconomics.eu

[b]University of Chicago Booth School of Business,lars.stole@chicagobooth.edu

lobbyist promises contributions to the politician as a function of the latter's publicly observed vote. As we will see, whether the principals' preferences are congruent (e.g., preferences for public goods) or in opposition (e.g., lobbyists on the left and right of a policy issue) will determine the character of the equilibrium distortions.

To obtain a characterization of the equilibria for this class of games, there are four conceptual and technical difficulties that must be overcome. First, even if the agent's non-participation option yields a type-independent payoff, the agent's participation constraint vis-à-vis principal i will typically be type dependent if the agent's best alternative is to contract with a subset of the remaining principals. Moreover, the value of the partial-contracting option is an endogenous object. Second, the set of principals who are actively influencing a particular agent type (i.e., making positive payments to that agent type in equilibrium) is also an endogenous object that must be constructed as part of the equilibrium. As we will see, it is not uncommon for each principal's activity set (i.e., the set of types to whom positive transfers are made) to be a proper subset of the agent's type space and to differ fundamentally from each other. This subtlety is novel to delegated common agency games with adverse selection. Third, the appropriate strategy spaces in these games are large – in the present case, the set of upper semi-continuous functions from actions to transfers. Unfortunately, standard control-theoretic results require that the principals' objective functions are continuous and differentiable. Because principal i's objective function depends upon the equilibrium transfer functions of the other principals, this smoothness restriction would require each principal to choose a differentiable transfer function. Thus, using standard control techniques in this context is implicitly making an equilibrium refinement that perhaps is unpalatable. Regardless, such a refinement should be made explicitly and its implications thoroughly understood. To this end we import and specialize the modern techniques of non-smooth analysis in optimal control allowing for discontinuous transfers. Lastly, most games in this class have a multiplicity of equilibria, so we would like theorems broadly applicable to the entire equilibrium set or, if this is not possible, theorems applicable to an important subset of equilibria where the equilibrium refinement is carefully considered.

This paper addresses each of these four difficulties and contributes in several dimensions. First, to address type-dependent optimization problems with non-smooth objective functions, we derive a novel set of necessary and sufficient conditions for objectives which require only linearity in the state variable – an assumption satisfied in most contracting problems in which payoffs are quasi-linear in money. This theorem, stated generally and proven in Appendix B, is of independent interest to common agency games.

Using the non-smooth control theorem, we are able to establish broad necessary conditions for any equilibrium outcome. This is achieved by noting that our class of common agency games is a special case of an aggregate game with infinite-dimensional strategy spaces. Using the *principle of aggregate concurrence*, a recent result from the aggregate games literature (Martimort and Stole 2011), we are able to state several useful necessary conditions which apply to the set of *all* equilibrium allocations (i.e., all equilibrium actions and agent-rent profiles). Indeed, given the activity sets of the principals, these necessary conditions provide sharp predictions.

To address the endogeneity of the principals' activity sets, we place additional structure on the payoffs of players – in short, we require that each principal's marginal

return to the agent's action is constant. Within this class of common agency games, we are able to demonstrate equilibrium existence and characterize the distortions caused by non-cooperative contracting. This class of equilibria includes discontinuous equilibria as well as a subclass of equilibria which we call *virtually truthful*. For this subclass, we provide a complete characterization of equilibria. We find that these equilibria share remarkable similarities with the Lindahl-Samuelson allocations from public finance and, remarkably, are solutions to a virtual-surplus maximization problem in the spirit of Myerson's (1981) optimal auction program. These equilibrium outcomes are extremely simple to calculate and comparative statics are immediately available. Moreover, as private information becomes less disperse the limiting equilibrium allocation is the *truthful* equilibrium allocation of Bernheim and Whinston (1986a). In this sense, virtually-truthful allocations are generalizations of the complete-information concept. We apply these results to our motivating examples of public goods games and lobbying games to explore the role of congruence and conflict in the principals' preferences.

Literature review

There is an existing literature on common agency games with public contracting variables. Bernheim and Whinston (1986a) initiated the study of such games (referred to as "menu auctions" in that paper) and their results about the existence and properties of truthful equilibria have been successfully applied to many contexts, including international trade and lobbying.[1]. There is also a less developed literature that has looked at intrinsic common agency games with public contracting variable in the context of moral hazard. Bernheim and Whinston (1986b) provided the first application in this framework. Later papers by Holmström and Milgrom (1989) and Dixit (1996) extended these insights to moral hazard games with linear contracts; more recently, Martimort and Stole (2011) extend the model of Innes (1990) to a common agency framework, proving equilibrium existence and characterizing the competitive distortions.

The literature on common agency games with adverse selection is almost exclusively limited to the case of *private contracting*.[2] In a private contracting setting – i.e., a game in which each principal is allowed to contract over a mutually exclusive subset of agent actions – Martimort and Stole (2009a) provided a general analysis of competition with nonlinear prices under both delegated and intrinsic common agency with endogenous activity sets. This is the appropriate class of models for games in which each principal controls a specific screening variable (e.g., the quantity he sells to the buyer) but does not observe the agent's actions on other dimensions (e.g., the quantity purchased from rivals).[3] The focus in the present paper, however, is public agency environments where each principal has access to the same screening variables. A second difference is that, in Martimort and Stole (2009a), manufacturers rank the agent's types the same way, with the agent having the highest valuation for both goods being

[1] Grossman, Helpman, Dixit, et al. (1997)

[2] A notable exception is Martimort (1996). Martimort (2007) introduces the distinction of private and public common agency and previews the analysis in the present paper.

[3] Ivaldi and Martimort (1994) and Calzolari and Scarpa (2008) are earlier studies of delegated common agency games with private contracting but focus *a priori* on cases where all types are served.

the most attractive for both manufacturers.[4] The present analysis is more general and allows for principals having conflicting preferences such as in our lobbying game. Biais, Martimort and Rochet (2000) analyzed a model of competing market-makers on financial markets with traders privately informed on their willingness to buy or sell assets in a common value environment with private agency. Because of symmetry, all market-makers have similar activity sets with a bid-ask spread such that traders having a mild preference for trading do not trade under asymmetric information.

Organization

Section 2 presents our model of delegated common agency under asymmetric information and some preliminary results characterizing each principal's set of incentive feasible allocations and constructing each principal's best response as a solution to a control problem using non-smooth analysis. Necessary and sufficient conditions are presented in Section 3. The implications of these conditions for equilibria are characterized in Section 4. Section 5 specializes to the case in which the principals' preferences are linear in the agent's action. This allows us to obtain closed form solutions for the activity sets and to derive finer properties of equilibria. In Section 6, we apply our findings to public good provision and lobbying games to illustrate the nature of equilibrium distortions. In the former, we obtain an elegant recasting of the Lindahl-Samuelson condition for public goods. In the latter, we derive a simple condition characterizing which principals will exert exclusive influence over an agent type and whether or not the principals' spheres of influence will overlap. Section 7 concludes and highlights several avenues for further research. The proofs of the main results are collected in Appendix A, together with a brief review of some useful results from convex analysis and non-smooth optimal control. A key theorem giving necessary and sufficient conditions for non-smooth optimal control with linear state variables is stated and proven in Appendix B.

2. THE MODEL

2.1. *Public delegated agency games*

Consider n principals indexed with the subscript $i \in \{1, .., n\} = N$. Each principal offers a contribution schedule to a common agent who chooses the level of a public good on their behalf. The set of feasible outputs is $\mathcal{Q} \equiv [q_{\min}, q_{\max}]$. Under *public delegated common agency*, the agent may also choose to contract with any subset of principals.[5]

2.2. *The strategic setting*

Each principal i offers a contract $t_i \in \mathcal{T}$ to the agent, which is assumed to be an upper semicontinuous function that maps outputs into non-negative monetary

[4] Mezetti (1997) provided a model with conflicting and differentiated principals but his focus was on an intrinsic common agency setting, putting aside the complete characterization of the activity sets.

[5] This distinguishes our analysis from the case of *intrinsic common agency* (Martimort and Stole, 2011).

payments.[6] Because common agency is delegated and the agent would never accept a contract stipulating a negative payment for an output he would like to choose, it is without loss of generality to require nonnegative contracts. More precisely, any equilibrium allocation supported by transfers that are not everywhere nonnegative is also an equilibrium allocation for some transfers which are nonnegative everywhere.

2.3. *Preferences*

All players have quasi-linear preferences defined over outputs and payments. The agent's direct benefit from choosing $q \in \mathcal{Q}$ is denoted as

$$S_0(q) - \theta q,$$

where $S_0 : \mathcal{Q} \to \mathbb{R}$ and $\theta \in \Theta \equiv [\underline{\theta}, \overline{\theta}]$. Principal i's direct benefit from the choice of $q \in \mathcal{Q}$ is denoted as $S_i(q)$, where again $S_i : \mathcal{Q} \to \mathbb{R}$. Throughout, the following assumptions hold:

- The benefit functions $S_0, \ldots, S_i, i \in N$, are upper semi-continuous.
- The agent has private information about his type, $\theta \in \Theta$, whose distribution is described by the absolutely continuous cumulative distribution function, $F : \Theta \to [0,1]$, and bounded density function, $f : \Theta \to \mathbb{R}_{++}$. In addition, the distribution of types satisfies a "generalized monotone hazard rate" property; i.e.,
$$\frac{F(\theta) - \kappa}{f(\theta)} \text{ is strictly increasing over } \Theta \text{ for any } \kappa \in [0,1].$$
- Any first-best optimal allocation,
$$q^{FB}(\theta) \in \arg\max_{q \in \mathcal{Q}} \sum_{i=0}^{n} S_i(q) - \theta q,$$
is **minimally** separating; i.e., $q^{FB}(\overline{\theta}) < q^{FB}(\underline{\theta})$.

2.4. *Timing and equilibrium*

The delegated common agency game unfolds as follows. First, the agent learns his private information θ. Second, principals non-cooperatively offer contracts to the agent. Third, the agent accepts some set of the offers, chooses q and receives the corresponding payments $t_i(q)$ required by the accepted offers.

We study the set of pure-strategy Nash equilibria to this game. Formally, we define equilibrium follows where we use an overline " $\bar{}$ " to denote an equilibrium component.

DEFINITION 1 *An* **equilibrium** *is a profile of principals' contribution schedules,* $\overline{\mathbf{t}} = \{\overline{t}_1, \ldots, \overline{t}_n\}$*, and an agent's output strategy,* $\overline{q}_0(\theta|\mathbf{t})$*, such that the following properties hold.*

[6] "Contracts on contracts" are not allowed because, for instance, principals j's offers ($j \neq i$) are non-observable by principal i. On the impact of such self-referencing contracts, see Peters and Szentes (2011).

1. *Given any profile of contributions* $\mathbf{t} \in \mathcal{T}^n$, $\overline{q}_0(\theta|\mathbf{t})$ *maximizes the agent's payoff:*

$$\overline{q}_0(\theta|\mathbf{t}) \in \arg\max_{q\in\mathcal{Q}} S_0(q) - \theta q + \sum_{i\in N} t_i(q).$$

2. $\overline{t}_i$ *maximizes principal* i*'s expected payoff given* $\overline{\mathbf{t}}_{-i}$ *and* $\overline{q}_0$*:*

$$\overline{t}_i \in \arg\max_{t_i\in\mathcal{T}} \int_\Theta (S_i(\overline{q}_0(\theta|t_i,\overline{\mathbf{t}}_{-i})) - t_i(\overline{q}_0(\theta|t_i,\overline{\mathbf{t}}_{-i})))f(\theta)d\theta.$$

2.5. *Incentive feasible set*

We now characterize the set of incentive feasible allocations available to principal i. Define the aggregate transfer functions $T(q) \equiv \sum_{j\in N} t_i(q)$ and $T_{-i}(q) \equiv \sum_{j\in N\setminus\{i\}} t_j(q)$.

DEFINITION 2 *A rent-output profile* (U,q), $U:\Theta\to\mathbb{R}_+$ *and* $q:\Theta\to\mathcal{Q}$, *is* **implementable** *by principal* i *given* T_{-i} *if and only if there exists a contribution schedule* $t_i \in \mathcal{T}$ *such that*

$$U(\theta) = \max_{q\in\mathcal{Q}} \; S_0(q) - \theta q + t_i(q) + T_{-i}(q),$$

$$q(\theta) \in \arg\max_{q\in\mathcal{Q}} \; S_0(q) - \theta q + t_i(q) + T_{-i}(q).$$

We define the agent's surplus from contracting with all principals but i for a given equilibrium profile of aggregate transfers, $\overline{T}_{-i}$, as

$$\overline{U}_{-i}(\theta) \equiv \max_{q\in\mathcal{Q}} \; S_0(q) - \theta q + \overline{T}_{-i}(q).$$

Implementability can also be characterized in the dual space of indirect utilities:

LEMMA 1 *A rent-output profile* (U,q) *is implementable by principal* i *given an aggregate equilibrium transfer profile,* $\overline{T}_{-i}$*, if and only if*

$$U(\theta) \geq \overline{U}_{-i}(\theta) \text{ for all } \theta\in\Theta, \tag{2.1}$$

$$-q(\theta) \in \partial U(\theta) \text{ for all } \theta\in\Theta, \tag{2.2}$$

$$U \text{ is convex.} \tag{2.3}$$

The notation "∂" in (2.2) denotes the subdifferential operator of a convex function. Because U is convex, it is differentiable almost everywhere; at all such smooth points the subdifferential is simply the gradient of U and we have $\dot{U}(\theta) = -q(\theta)$.[7]

[7]Condition (2.2) holds also for the rent/output profile $(\overline{U}_{-i}(\theta), \overline{q}_{-i}(\theta))$ since it is itself implementable (when P_i offers a null contribution) and for the status quo profile $(U_0(\theta), q_0(\theta))$ had no principal offered any contract. This property is called *"homogeneity"* in Jullien (2000).

2.6. *Activity sets*

Because contributions are non-negative, the agent is weakly better off accepting all principals' offers and (2.1) is satisfied. One important aspect of our analysis is to determine precisely the subset of types where such contribution is indeed null.

DEFINITION 3 *Principal i's* **equilibrium activity set** *is defined as*

$$\overline{\Theta}_i \equiv \left\{\theta \in \Theta \,|\, U(\theta) > \overline{U}_{-i}(\theta)\right\}.$$

The set of active principals for θ is given by the **equilibrium activity map**

$$\overline{\alpha}(\theta) \equiv \left\{i \in N \,|\, U(\theta) > \overline{U}_{-i}(\theta)\right\}.$$

Note that $\overline{\Theta}_i$ is open in Θ and principal i's contribution is necessarily positive on his activity set. We define the complement of $\overline{\Theta}_i$ as $\overline{\Theta}_i^c$, i.e., the subset of types where the participation constraint (2.1) is binding. The activity mapping $\overline{\alpha}$ contains a complete description of all equilibrium activity sets.

2.7. *A non-smooth control problem*

Principal i looks for an allocation that, at a best response, maximizes the joint payoff of the bilateral coalition he forms with the agent,

$$S_i(q) + S_0(q) + \sum_{j \neq i} t_j(q) - \theta q,$$

minus the information rent, $U(\theta)$, that the type-θ agent retains. Of course, the ability of principal i to extract this rent is limited by the agent's option to contract with other principals. More formally, we can state the relevant program for principal i when facing the profile $\mathbf{t}_{-i}$ as:

$$(\mathcal{P}_i): \max_{\substack{U \in AC(\Theta,\mathbb{R}) \\ q \in \mathcal{L}(\Theta,\mathcal{Q})}} \int_\Theta \left(S_i(q(\theta)) + S_0(q(\theta)) + \sum_{j \neq i} t_j(q(\theta)) - \theta q(\theta) - U(\theta) \right) f(\theta) d\theta$$

$$\text{subject to (2.1), (2.2), (2.3),}$$

where $AC(\Theta, \mathbb{R})$ denotes the set of all absolutely continuous, real functions with domain Θ and $\mathcal{L}(\Theta, \mathcal{Q})$ is the space of measurable functions from Θ to $\mathcal{Q}$.

3. OPTIMALITY CONDITIONS FOR BEST RESPONSES

We rely on a technical contribution which characterizes the solutions for non-smooth optimal control problems with objectives that are linear in the state variable. This result and its lengthy proof is presented in the Appendix as Theorem B.1. In short, Theorem B.1 demonstrates that the necessary and sufficient conditions for programs with twice continuously differentiable objectives that appear in Jullien (2000) generalize to programs with integrands which are simply upper semi-continuous. We specialize the implications of Theorem B.1 to our setting in the following Theorem.

THEOREM 1 *Given the profile of transfers offered by rival principals, $\mathbf{t}_{-i}$, the rent-ouput profile $(\overline{U},\overline{q})$ is a solution to $(\mathcal{P}_i)$ if and only if $(\overline{U},\overline{q})$ satisfies (2.1), (2.2), (2.3) and there exists a measure μ_i (possibly with mass points) defined over the Borel subsets of Θ with an associated adjoint function, $\overline{M}_i:\Theta\to[0,1]$,[8] defined by $\overline{M}_i(\underline{\theta})=0$ and for $\theta>\underline{\theta}$,*

$$\overline{M}_i(\theta)\equiv\int_{[\underline{\theta},\theta)}\mu_i(d\theta),$$

such that the following two conditions are satisfied:

$$\text{(3.1)}\qquad \text{supp}\left\{\mu_i\right\}\subseteq\overline{\Theta_i}^c\equiv\left\{\theta\,|\,\overline{U}(\theta)=\overline{U}_{-i}(\theta)\right\},$$

$$\text{(3.2)}\qquad \overline{q}(\theta)\in\arg\max_{q\in\mathcal{Q}}\ S_i(q)+S_0(q)+\sum_{j\neq i}t_j(q)+\left(\frac{\overline{M}_i(\theta)-F(\theta)}{f(\theta)}-\theta\right)q,\ a.e.\,\theta\in\Theta.$$

The conditions above for a best response do not yet characterize an equilibrium. To do so, we need to derive more precise implications from these conditions. One would like to describe both the equilibrium allocation and the corresponding adjoint functions with only minimal reference to the n-tuple of equilibrium transfers which implements them. Section 4 makes progress towards the first objective with necessary conditions for equilibrium allocations being derived with respect to the aggregate transfers and not their exact distribution among principals. Section 5 specializes the principals' preferences to derive even sharper predictions on adjoint functions and equilibrium allocations.

4. EQUILIBRIA

4.1. *General structure of equilibria*

A remarkable implication of Theorem 1 that has immediate economic content can be found by aggregating the optimality conditions (3.2) across all principals.[9]

THEOREM 2 *In any equilibrium of the delegated agency game with activity sets represented by $\overline{\alpha}$ and the principals' adjoint profile given by $\{\overline{M}_1,\ldots,\overline{M}_n\}$, the allocation satisfies the two conditions*

$$\text{(4.1)}\qquad \overline{q}(\theta)\in\arg\max_{q\in\mathcal{Q}}\ S_0(q)-\theta q+\sum_{i=1}^{n}S_i(q)+\sum_{i=1}^{n}\left(\frac{\overline{M}_i(\theta)-F(\theta)}{f(\theta)}\right)q$$
$$+(n-1)\left(S_0(q)-\theta q+\overline{T}(q)\right),\ a.e.;$$

$$\text{(4.2)}\qquad \overline{q}(\theta)\in\arg\max_{q\in\mathcal{Q}}\ S_0(q)-\theta q+\sum_{i\in\overline{\alpha}(\theta)}S_i(q)+\sum_{i\in\overline{\alpha}(\theta)}\left(\frac{\overline{M}_i(\theta)-F(\theta)}{f(\theta)}\right)q$$
$$+(|\overline{\alpha}(\theta)|-1)\left(S_0(q)-\theta q+\overline{T}(q)\right),\ a.e.$$

[8] $\overline{M}_i$ is piecewise absolutely continuous, but might have jumps wherever μ_i has mass points.

[9] More generally, there is an aggregate concurrence property underlying the result in Theorem 2. This idea is developed in Martimort and Stole (2011) and applied to other contracting games.

REMARKS: The objective function in (4.1) is obtained by summing over all principals the optimality conditions (3.2). Proceeding similarly but over active principals only, we obtain a more precise requirement in (4.2). The difference between the two is that inactive principals do not change the agent's decision; in other words, their objectives are aligned with that of the agent given the contracts of active principals. This congruence is captured by the joint requirement that an equilibrium allocation satisfies (4.1) and (4.2).

Turning to this latter condition, the last term

$$\left(|\overline{\alpha}(\theta)| - 1\right)\left(S_0(q) - \theta q + \overline{T}(q)\right),$$

is a multiple of the agent's equilibrium maximand, which of course is maximized for any incentive compatible allocation. If the transfers are continuously differentiable around the equilibrium allocation, then the presence of this term in the maximand is immaterial by the the agent's first-order condition. We consider more subtle arguments that allow for non-smooth equilibria below.

For the present, consider the first terms in (4.2)

$$S_0(q) - \theta q + \sum_{i \in \overline{\alpha}(\theta)} S_i(q) + \sum_{i \in \overline{\alpha}(\theta)} \left(\frac{\overline{M}_i(\theta) - F(\theta)}{f(\theta)} \right) q.$$

This simplified objective provides a remarkable intuition for the equilibrium allocation: it is *as if* the allocation $\overline{q}(\theta)$ is chosen to maximize the collective surplus of the agent and the principals in the active coalition $\overline{\alpha}(\theta) \subseteq N$, less an information rent term. This term, moreover, is a pointwise sum of the agent's bilateral information rents vis-à-vis each active principal which captures the idea that agency problems are compounded under delegated agency. If we define the average adjoint function across the coalition as

$$\overline{M}_S(\theta) \equiv \frac{1}{|S|} \sum_{i \in S} \overline{M}_i(\theta),$$

then the aggregate information rent term is $|\overline{\alpha}(\theta)|$ times larger than the rent term for the average principal:

$$\sum_{i \in \overline{\alpha}(\theta)} \left(\frac{\overline{M}_i(\theta) - F(\theta)}{f(\theta)} \right) q = |\overline{\alpha}(\theta)| \left(\frac{\overline{M}_{\overline{\alpha}(\theta)}(\theta) - F(\theta)}{f(\theta)} \right) q.$$

In particular, if each principal's adjoint function $\overline{M}_i$ is identical within the coalition, which is the case if all principals do agree on the identity of the "worst" agent's type from their point of view, then the information rent term of the coalition is exactly $|\overline{\alpha}(\theta)|$ times larger than the individual terms. Hence, all else equal, an $|\overline{\alpha}(\theta)|$-fold larger distortion is introduced. In this sense, the distortion is very similar to the double-marginalization which arises in complementary vertical relationships.

While the preceding remarks are economically interesting, we must emphasize that $\overline{\alpha}$ is an endogenous object, so the necessary condition (4.2) is more insightful than practical.[10] Nevertheless, Section 5 makes significant progress toward an understanding of (4.2) by restricting attention to the case where principals have linear preferences.

[10]Remarkably, under intrinsic common agency, $\overline{\alpha}(\theta) \equiv N$, and the necessary condition above can be shown under mild conditions to also be a sufficient condition for equilibrium. (See Martimort and Stole, 2011.)

DEFINITION 4 *For any solution $(\overline{q}, \overline{U})$ to program $(\mathcal{P}_i)$, an adjoint function $\overline{M}_i$ that satisfies the conditions of Theorem 1 is said to* **support** $(\overline{q}, \overline{U})$.

We can build upon Theorem 1 to discern more properties about the adjoint functions $\overline{M}_i$ which support $(\overline{q}, \overline{U})$.

PROPOSITION 1 *For any optimal solution $(\overline{q}, \overline{U})$ to program $(\mathcal{P}_i)$, there exists an adjoint function $\overline{M}_i : \Theta \to [0,1]$ which supports $(\overline{q}, \overline{U})$ with the following properties:*

(i). $\overline{M}_i$ is constant over any connected interval of $\overline{\Theta}_i$;

(ii). if $S_i : \mathcal{Q} \to \mathbb{R}$ is differentiable and concave, then over any open interval of $\overline{\Theta}_i^c$ for which $\overline{q}_{-i}(\cdot)$ is strictly decreasing,

$$\overline{M}_i(\theta) = F(\theta) - S_i'(\overline{q}_{-i}(\theta))f(\theta); \tag{4.3}$$

(iii). $\overline{M}_i$ is continuous on $(\underline{\theta}, \overline{\theta})$.

REMARKS:

• The constancy of $\overline{M}_i$ on intervals of $\overline{\Theta}_i$ is implied by the complementary slackness condition (3.1). Equation (4.3) is a local marginal condition for satisfying (3.2); it is a necessary condition if S_i is concave. The continuity of the adjoint function $\overline{M}_i$ follows from manipulating (3.1) and (3.2) and the fact that the agent's reservation utility, $\overline{U}_{-i}$, is a convex function of θ.

• If we conjecture that the equilibrium allocation is strictly decreasing, Proposition 1 implies that each $\overline{M}_i$ is completely characterized once $\overline{\alpha}$ is known. If we also believe that each principal's activity set is a single interval of Θ, then constructing such an equilibrium is as simple as searching for the endpoints of each principal's activity set and checking that the resulting $\overline{M}_i$ constructions from Proposition 1 satisfy the optimality conditions with respect to $(\overline{q}, \overline{U})$ in Theorem 1.

For any equilibrium allocation function, denote the set of equilibrium actions by $\overline{\mathcal{Q}} \equiv \{q \in \mathcal{Q} \,|\, \exists \theta \in \Theta,\, q = \overline{q}(\theta)\}$.

DEFINITION 5 *Characteristics of an allocation, $\overline{q}$:*

- *An allocation is* **continuous** *if $\overline{q} : \Theta \to \mathcal{Q}$ is continuous;*
- *An allocation is* **separating** *if $\overline{q}$ is non-constant and strictly decreasing for all θ for which $\overline{q}(\theta) \in \text{int}(\overline{\mathcal{Q}})$.*

REMARKS: Continuous allocations may fail to be separating if there are interior bunching points. Such a possibility is not pathological and arises, for example, when S_0 is strictly concave but not everywhere differentiable.[11] Separating allocations may also fail to be continuous; discontinuities arise, for example, if all benefit functions $S_0, S_1, \ldots, S_n$ are linear in q and the solution is of a "bang-bang" nature.[12]

[11] S_0 can be strictly concave but non-differentiable in very reasonable settings. For example, if the agent is producing a public good and S_0 is a cost-minimizing mixture of two cost functions, each with a fixed component and a strictly convex and differentiable variable-cost component, then S_0 will be strictly concave and exhibit a kink.

[12] The definition of separating equilibrium is weaker than the requirement that $\overline{q}$ is strictly decreasing; a separating allocation may exhibit bunching on the boundaries of $\mathcal{Q}$ as long as there is minimal separation.

4.2. *Continuous equilibrium allocations*

Strict concavity of the agent's objective function (benefit function plus transfers) is the key condition for continuity as shown below. Note that this condition does not require that transfer or benefit functions be differentiable.

PROPOSITION 2 *The equilibrium allocation, $\overline{q} : \Theta \to \mathcal{Q}$, is continuous if and only if for any $i \in N$ the function $W_i \equiv S_0(q) + S_i(q) + \overline{T}_{-i}(q)$ is strictly concave or (equivalently) $S_0 + \overline{T}$ is strictly concave:*

$$\overline{q} \in C^0(\Theta) \iff W_i, \text{ strictly concave in } \mathcal{Q} \iff S_0 + \overline{T}, \text{ strictly concave in } \mathcal{Q},$$

where $C^0(\Theta)$ denotes the space of continuous functions on Θ. In any equilibrium in which $\overline{q}$ is continuous at an interior boundary of principal i's activity set, i.e., $\hat{\theta}_i \in (\underline{\theta}, \overline{\theta})$, the following properties hold:

$$\overline{U}(\hat{\theta}_i) = \overline{U}_{-i}(\hat{\theta}_i) \text{ and } \dot{\overline{U}}(\hat{\theta}_i) = \dot{\overline{U}}_{-i}(\hat{\theta}_i) = -\overline{q}(\hat{\theta}_i). \tag{4.4}$$

Conditions (4.4) are "smooth-pasting" conditions that apply to the agent's equilibrium payoff on the boundary of an activity set which is interior to the type set. From an economic point of view, these properties require that whenever a principal starts to be active at an interior point, he does so by changing only marginally the agent's decision.

4.3. *Separating equilibrium allocations*

PROPOSITION 3 *Assume that the benefit functions $S_0, S_1, \ldots, S_n$ are continuously differentiable. If $\overline{q}$ is a separating equilibrium allocation, then for each $i \in N$, the transfer function $\overline{t}_i$ is differentiable over int$(\overline{Q})$. Furthermore, for the corresponding activity map $\overline{\alpha}$ and adjoint functions, $\{\overline{M}_i, \ldots, \overline{M}_n\}$,*

$$S_0'(\overline{q}(\theta)) + \overline{T}'(\overline{q}(\theta)) = \theta \text{ for all } \theta \in \Theta, \overline{q}(\theta) \in \operatorname{int}\overline{\mathcal{Q}}, \tag{4.5}$$

$$\overline{t}_i'(\overline{q}(\theta)) = S_i'(\overline{q}(\theta)) + \frac{\overline{M}_i(\theta) - F(\theta)}{f(\theta)}, \text{ a.e. } \theta \in \Theta, \overline{q}(\theta) \in \operatorname{int}\overline{\mathcal{Q}}, i \in \overline{\alpha}(\theta), \tag{4.6}$$

$$S_0'(\overline{q}(\theta)) + \sum_{j \in \overline{\alpha}(\theta)} \left\{ S_j'(\overline{q}(\theta)) + \frac{\overline{M}_i(\theta) - F(\theta)}{f(\theta)} \right\} = \theta \text{ a.e. } \theta \in \Theta. \tag{4.7}$$

The economics behind these results is straightforward and reduces to understanding how the standard rent-efficiency trade-off of the screening literature is modified in a competitive screening environment. Intuitively, when an agent with type θ behaves like a less (resp. more) efficient type $\theta + d\theta$ (resp. $\theta - d\theta$), he produces the same amount at a lower (resp. higher) cost. Mitigating these incentives require the principals to distort output downwards for all types except the most (resp. least) efficient one. Those "incentive" distortions are captured by the usual hazard rate term $F(\theta)/f(\theta)$ (resp. $F(\theta) - 1/f(\theta)$) that discounts each principal's marginal contribution

below (resp. above) his marginal valuation. Because the type-dependent participation constraints that are specific to the delegated agency game limit the ability of any principal to extract rent and reduce outputs, other "participation" distortions must now be taken into account. Those distortions increase marginal contributions on any activity sets. They are captured by the new non-negative term $\overline{M}_i(\theta)/f(\theta)$.

When the measure μ_i does not put too much mass on types less than θ, we have $\overline{M}_i(\theta) < F(\theta)$. This corresponds to the case where principal i finds it relatively cheap to buy types below θ. The driving force behind output distortions are then the incentive distortions which tend to reduce output and principal i's marginal contribution. When instead μ_i puts enough mass on types lower than θ, i.e., principal i finds it too costly to induce participation from those types less than θ and prefers to stay inactive. We have $\overline{M}_i(\theta) > F(\theta)$ and output distortions are mitigated to ensure participation.

The distortions induced by active principals are compounded altogether at equilibrium as it can be shown on Equation (4.7). Importantly, not all distortions may go in the same direction. This will be illustrated by means of examples in Section 5 below.

PROPOSITION 4 *Assume that the benefit functions $S_0, S_1, \ldots, S_n$ are continuously differentiable. If $\overline{q}$ is a separating equilibrium allocation, then for each $i \in N$, the transfer function $\overline{t}_i$ satisfies the following conditions at any interior boundary of principal i's activity set $\hat{\theta}_i \in (\underline{\theta}, \overline{\theta})$ such that $q(\hat{\theta}_i) \in int(\overline{Q})$*

$$\overline{t}_i(\overline{q}(\hat{\theta}_i)) = \overline{t}_i'(\overline{q}(\hat{\theta}_i)) = 0. \tag{4.8}$$

The "smooth-pasting" conditions (4.8) follow up on (4.4). They imply that principal i's non-negative transfer function is locally convex around any point $\overline{q}_i(\hat{\theta}_i)$ which is in the interior of the range of equilibrium outputs.

5. GAMES WITH LINEAR PREFERENCES

5.1. *Equilibria*

Although Proposition 3 describes equilibrium outputs, it says little on who the active principals are on a given set of types. To do so, we first need to describe the adjoint functions $\overline{M}_i$. This would generally be a daunting task because the condition (4.3) depends on $\overline{q}_{-i}$ which is itself an equilibrium construction. Luckily, sharp predictions can be achieved when principals have linear preferences since those adjoints take simple expressions.

Assume that principals' benefit functions are linear: $S_i(q) = s_i q$, $s_i \neq 0$. We partition principals into two sets, $\mathcal{A} = \{i \in N \mid s_i > 0\}$ and $\mathcal{B} = N \setminus \mathcal{A}$. Principals in $\mathcal{A}$ like the agent's action whereas those in $\mathcal{B}$ dislike it. For $i \in \mathcal{A}$, define $\hat{\theta}_i$ as the unique value such that $F(\hat{\theta}_i) - s_i f(\hat{\theta}_i) = 0$, if it exists, and $\hat{\theta}_i = \overline{\theta}$ otherwise. Similarly, for $i \in \mathcal{B}$, define $\hat{\theta}_i$ as the unique value such that $F(\hat{\theta}_i) - s_i f(\hat{\theta}_i) = 1$, if it exists, and $\hat{\theta}_i = \underline{\theta}$ otherwise.

LEMMA 2 *Suppose that $S_i(q) = s_i q$ for $i \in N$. The following adjoint-function profile supports any non-degenerate equilibrium allocation, $\overline{q}$:*

$$\overline{M}_i(\theta) = \max\{F(\theta) - s_i f(\theta), 0\}, \quad i \in \mathcal{A}, \quad \text{and } \overline{M}_i(\theta) = \min\{F(\theta) - s_i f(\theta), 1\}, \quad i \in \mathcal{B}.$$

It is now useful to define the following "virtual" marginal valuations of the principals and their corresponding aggregate

$$\beta_i(\theta) = \max\left\{s_i - \frac{F(\theta)}{f(\theta)}, 0\right\}, \quad i \in \mathcal{A}, \quad \beta_i(\theta) = \min\left\{s_i + \frac{1-F(\theta)}{f(\theta)}, 0\right\}, \quad i \in \mathcal{B},$$

$$\beta(\theta) \equiv \sum_{i\in N} \beta_i(\theta) = \sum_{i\in\mathcal{A}} \max\left\{s_i - \frac{F(\theta)}{f(\theta)}, 0\right\} + \sum_{i\in\mathcal{B}} \min\left\{s_i + \frac{1-F(\theta)}{f(\theta)}, 0\right\}.$$

Note that β is continuous and weakly decreasing under our assumption on the generalized monotonicity of the hazard rate.

Lemma 2, in tandem with the necessary condition (4.1) in Theorem 2, allows us to state a remarkably simple necessary condition for any equilibrium allocation and to provide sufficient condition.

THEOREM 3 *Suppose that $S_i(q) = s_i q$ for $i \in N$ and that S_0 is concave. If $\overline{q}$ is an equilibrium allocation with range $\overline{\mathcal{Q}} = \overline{q}(\Theta)$, then for almost every $\theta \in \Theta$,*

$$\overline{q}(\theta) \in \arg\max_{q\in\overline{\mathcal{Q}}} S_0(q) + (\beta(\theta) - \theta)q. \tag{5.1}$$

Moreover, if for all $\theta \in \Theta$

$$\overline{q}(\theta) \in \arg\max_{q\in\mathcal{Q}} S_0(q) + (\beta(\theta) - \theta)q, \tag{5.2}$$

then $\overline{q}$ is an equilibrium allocation. Such an equilibrium allocation always exists.

The strength of Theorem 3 comes from the fact that (5.1) and (5.2) provide clear characterizations of equilibrium allocations. In comparison with condition (4.1), this is a tremendous simplification: any endogenous variables related to activity sets and equilibrium transfers have now disappeared.

The difference between conditions (5.1) and (5.2) are subtle, but significant. Condition (5.1) restricts optimality over the set of *equilibrium actions*, $\overline{\mathcal{Q}}$, and thus gaps are a priori possible in the range of $\overline{q}$.[13] The condition in (5.2) is stronger than (5.1) since it requires optimality over the whole set of possible actions, $\mathcal{Q}$. When, for instance, S_0 is strictly concave, this restriction, which is akin to an equilibrium refinement, impies that $\overline{q}$ is indeed continuous. We show below that the benefit of such restriction is that it allows one to easily reconstruct equilibrium transfers.

To better understand the implications of those findings, we now give more detailed implications of Theorem 3.

PROPOSITION 5 *Suppose that $S_i(q) = s_i q$ for $i \in N$ and that S_0 is concave.*

[13] The description of the equilibria above bears strong similarities with the literature on mechanism design without transfers in monopolistic screening environments. See Holmström (1984), Melumad and Shibano (1991), Alonso and Matoushek (2008), and Martimort and Semenov (2006) among others. Everything happens as if the equilibrium output $\overline{q}$ and its range $\overline{Q}$ were chosen by a surrogate principal who aggregates the behavior of all principals and maximizes their aggregate virtual surplus as in (5.1).

1. *At any point θ where $\overline{q}$ and $S_0(q)$ are both differentiable (resp. in θ and q), we have*

$$\dot{\overline{q}}(\theta)\left(S_0'(\overline{q}(\theta)) + \beta(\theta) - \theta\right) = 0. \tag{5.3}$$

2. *At any point θ where $\overline{q}(\theta)$ is discontinuous, we have*

$$\Big[S_0(q) + (\beta(\theta) - \theta)q\Big]_{\overline{q}(\theta^+)}^{\overline{q}(\theta^-)} = 0. \tag{5.4}$$

Moreover, the allocation $\overline{q}(\theta)$ exhibits bunching on both the right and left-neighborhoods of such θ.

When $\overline{q}$ is strictly decreasing and continuous, it can be readily identified with the *modified Lindahl-Samuelson* output q^{LS}. That allocation satisfies the following *Lindahl-Samuelson conditions* except that it applies to the principals' "virtual" marginal valuations to account for informational problems:

$$S_0'(q^{LS}(\theta)) + \sum_{i\in\mathcal{A}} \max\left\{s_i - \frac{F(\theta)}{f(\theta)}, 0\right\} + \sum_{i\in\mathcal{B}} \min\left\{s_i + \frac{1-F(\theta)}{f(\theta)}, 0\right\} = \theta. \tag{5.5}$$

On the other hand, discontinuities of $\overline{q}$, if any, have a very simple structure. Across any such discontinuity the aggregate virtual surplus of principals remains constant as shown in (5.4).

In the sequel, we first focus on equilibria that satisfy (5.3), the so-called *virtually-truthful equilibria* (Section 5.2). We then describe how those equilibria can be modified to support discontinuities (Section 5.3).

5.2. *Virtually-truthful equilibria*

Equilibria which satisfy (5.2) have an appealing property that each principal offers her "virtual" marginal valuation for the agent's marginal action. As we will demonstrate, as the agent's private information becomes less heterogeneous, such equilibria converge in a strong sense to the *truthful* equilibria of the complete-information menu auction game in Bernheim and Whinston (1986). Viewed in this light, the equilibria satisfying (5.2) are a generalization of truthful equilibria to the case of incomplete information, and we thus refer to them henceforth as *virtually-truthful* allocations:

DEFINITION 6 *An allocation is* **virtually truthful** *if it satisfies (5.2) for every $\theta \in \Theta$.*

To make the previous remarks precise, recall from Bernheim and Whinston (1986) that "truthful" schedules in the complete information game with linear surplus functions correspond to schedules in the form of $t_i(q) = \max\{0, s_i q - C_i\}$ where the constant C_i is principal i's equilibrium payoff.[14] Such truthful schemes make the agent a residual claimant for the choice of the decision. Remarkably, Bernheim and Whinston

[14] Bernheim and Whinston (1986) and Lebreton and Laussel (2001) show that those constants might not necessarily be unique.

(1986) demonstrated that truthful equilibria always exist and that they exhibit no equilibrium inefficiencies (e.g., no problems with free riding, etc.).

In the present setting of incomplete information, the equilibrium schedules which give rise to an allocation satisfying (5.2) are not truthful. To characterize these schedules, we need to define an inverse correspondence for the equilibrium allocation $\overline{q}$. To this end, define respectively the corresponding value function $V : \Theta \to \mathbb{R}$ and its dual conjugate[15] $V^* : \mathcal{Q} \to \mathbb{R}$ as

$$V(\theta) \equiv \max_{q \in \mathcal{Q}} \ (\beta(\theta) - \theta)\, q + S_0(q) \ \text{ and } \ V^*(q) \equiv \max_{\theta \in \Theta} \ (\beta(\theta) - \theta)\, q - V(\theta).$$

Note that V^* is convex and therefore the subdifferential $\partial V^*(q)$ is a non-decreasing correspondence mapping from $\mathcal{Q}$ onto Θ. Moreover, $\overline{q}(\theta)$ satisfies (5.2) if and only if[16] $\overline{q}(\theta) \in \partial V(\theta)$ or, equivalently, $\theta \in \partial V^*(\overline{q}(\theta))$. Because $\partial V^*(q)$ is single-valued at almost every $q \in \mathcal{Q}$, we will abuse notation slightly and equate the subdifferential with its subgradient at all such points and denote this subgradient by the more evocative notation $\vartheta(q) \equiv \partial V^*(q)$. In this sense, $\vartheta(q)$ can be viewed as the inverse mapping of the allocation $\overline{q}(\theta)$.

With this notation, we can construct the following marginal transfer functions for all $q \in \mathcal{Q}$ at which ϑ is single-valued:

$$\overline{t}_i'(q) = \max\left\{ s_i - \frac{F(\vartheta(q))}{f(\vartheta(q))}, 0 \right\}, \ i \in \mathcal{A}, \ \text{ and } \overline{t}_i'(q) = \min\left\{ s_i + \frac{1 - F(\vartheta(q))}{f(\vartheta(q))}, 0 \right\}, \ i \in \mathcal{B}.$$

Because marginal transfers are continuous for a.e. $q \in \mathcal{Q}$, we can integrate these expressions and recover the equilibrium schedule $\overline{t}_i$ up to some constants

$$\overline{t}_i(q) = C_i + \int_{\overline{q}(\hat{\theta}_i)}^{q} \max\left\{ s_i - \frac{F(\vartheta(q))}{f(\vartheta(q))}, 0 \right\} dq, \ i \in \mathcal{A}, \tag{5.6}$$

$$\overline{t}_i(q) = C_i + \int_{\overline{q}(\hat{\theta}_i)}^{q} \min\left\{ s_i + \frac{1 - F(\vartheta(q))}{f(\vartheta(q))}, 0 \right\} dq, \ i \in \mathcal{B} \tag{5.7}$$

where $\hat{\theta}_i$ is defined as above. Principals for $i \in \mathcal{A}$ are active on all $[\underline{\theta}, \hat{\theta}_i)$ whereas principals for $i \in \mathcal{B}$ are so on all $(\hat{\theta}_i, \overline{\theta}]$.

These schedules are strictly convex over their positive range because ϑ is nonincreasing and the hazard rates are monotonic.[17] Moreover, when $\hat{\theta}_i$ is interior, we have $\overline{t}_i(\overline{q}(\hat{\theta}_i)) = 0$ and the constant C_i is zero. These schedules are therefore smooth at such points with a zero derivative $\overline{t}_i'(\overline{q}(\hat{\theta}_i)) = 0$. These "smooth-pasting" and convexity properties stand in contrast with truthful schedules which exhibit kinks and are only linear over their positive range. The intuition for those differences is that, under asymmetric information, the equilibrium schedule is much more constrained around any point. Indeed, such a schedule now has to go through equilibrium points corresponding to nearby types. This "extra information" implies smooth-pasting.

[15] Appendix A recaps tools from convex analysis that are helpful to understand this construction.
[16] See Lemma 5 in Appendix B.
[17] See the proof of Proposition 9 below for details.

We now return to our earlier claim that these virtually-truthful schedules in the incomplete-information game converge in a sense to the truthful schedules of Bernheim and Whinston (1986) in the complete-information game. For any allocation $\overline{q}$ which satisfies (5.2) everywhere on Θ, consider the equilibrium schedule for some principal i, which we take to be $i \in \mathcal{A}$ without loss of generality. For any absolutely continuous distribution of θ satisfying the monotone-hazard rate property, we have already demonstrated that the unique equilibrium marginal transfer for principal i is given by

$$\overline{t}_i'(\overline{q}(\theta)) = \max\left\{s_i - \frac{F(\theta)}{f(\theta)}, 0\right\}.$$

Consider a class of absolutely-continuous distributions of θ in which each distribution has the same mean and each satisfies the generalized monotone-hazard rate property. Suppose that this class of distributions is second-order stochastically ordered, and let $\kappa \in [0, \infty)$ be an index representing a mean-preserving contraction of F around the common mean. An arbitrary distribution in this class is denoted $F(\theta, \kappa)$ and $\lim_{\kappa\to\infty} F(\theta, \kappa)$ is a Dirac distribution centered on the common mean. Using our previous construction of marginal transfers, we may thus write principal i's marginal transfer for a given distribution κ as

$$\overline{\tau}_i(\theta, \kappa) \equiv \max\left\{s_i - \frac{F(\theta, \kappa)}{f(\theta, \kappa)}, 0\right\}.$$

The following proposition makes precise the sense in which equilibria satisfying (5.2) are a generalization of truthful equilibria to the case of incomplete information.

PROPOSITION 6 *The expected marginal transfer of principal i in the complete-information limit game is principal i's marginal benefit:*

$$\lim_{\kappa\to\infty} \int_\Theta \overline{\tau}_i(\theta, \kappa) f(\theta, \kappa) d\theta = s_i. \tag{5.8}$$

To conclude this section, we provide a complete characterization of virtually-truthful equilibria for games with linear preferences. We will apply this result in Section 6 to a few economic applications to illustrate the value of our framework.

PROPOSITION 7 *Suppose that $S_i(q) = s_i q$ for $i \in N$.*

1. *If $S_0 : \mathcal{Q} \to \mathbb{R}$ is linear, then there exists a type, $\hat{\theta} \in [\underline{\theta}, \overline{\theta}]$, such that every virtually-truthful equilibrium allocation satisfies*

$$\overline{q}(\theta) = \begin{cases} q_{\max} & \theta < \hat{\theta} \\ q_{\min} & \theta > \hat{\theta}. \end{cases} \tag{5.9}$$

2. *If $S_0 : \mathcal{Q} \to \mathbb{R}$ is strictly concave, then a unique virtually-truthful equilibrium allocation exists and it is continuous.*
3. *If $S_0 : \mathcal{Q} \to \mathbb{R}$ is strictly concave and differentiable, then a unique virtually-truthful equilibrium allocation exists and it is both continuous and separating.*

5.3. *Discontinuous equilibria*

Turning now to the analysis of discontinuous equilibria, we specialize our model to the case of two principals.

PROPOSITION 8 *Suppose that $S_i(q) = s_i q$ for $i \in N = \{1,2\}$ and $S_0 : \mathcal{Q} \to \mathbb{R}$ is strictly concave and differentiable. Then, for any virtually-truthful equilibrium $\overline{q}(\theta)$ and any θ_0 where both principals are active in that equilibrium, one can construct nearby equilibria exhibiting a discontinuity at θ_0 (i.e., $\overline{q}^d(\theta_0^+) < \overline{q}^d(\theta_0^-)$).*

To sustain those equilibria, principals design their contracts with "non-serious" out-of-equilibrium offers. For instance, principal 2 stipulates non-negative payments for outputs within the discontinuity gap $[\overline{q}^d(\theta_0^+), \overline{q}^d(\theta_0^-)]$ which are such that principal 1 is just indifferent to inducing the agent with type θ_0 to produce any output within that range. This makes it possible to sustain the discontinuity in the agent's choice.[18]

6. APPLICATIONS

6.1. *Voluntary provision of a public good*

In such setting, all principals/contributors have congruent objectives and enjoy more of the public good being produced. To illustrate the richness of equilibrium patterns compatible with Proposition 7, we nevertheless distinguish two cases.

6.1.1. *Public goods game in which S_0 is strictly concave.*

Consider the public goods game with n principals ordered from $s_1 \geq ... \geq s_n > 0$ ($i \in N = \{1,..,n\}$), and assume that $S_0(q) = -\frac{1}{2}q^2$ so that there are decreasing returns to scale. We also assume that $\mathcal{Q} = [0, q_{\max}]$ with $q_{\max}$ sufficiently large (e.g., $q_{\max} \geq -\underline{\theta} + \sum_{i\in N} s_i$) and $\Theta = [0, \bar{\theta}]$.

PROPOSITION 9 *In the public goods game with decreasing returns to scale $S_0(q) = -\frac{1}{2}q^2$, the virtually-truthful equilibrium allocation is*

$$\overline{q}(\theta) = \max\left\{\sum_{i\in N} \max\left\{s_i - \frac{F(\theta)}{f(\theta)}, 0\right\} - \theta, 0\right\}. \tag{6.1}$$

EXAMPLE: For $N = \{1,2\}$, the efficient output is $q^{FB}(\theta) = \max\{s_1 + s_2 - \theta, 0\}$ whereas the optimal output that principal 1 would implement if he was alone to contract with the agent is $q_1^*(\theta) = \max\{s_1 - \theta - \frac{F(\theta)}{f(\theta)}, 0\}$. It entails only one screening distortion. The equilibrium output of the delegated agency game exhibits two such screening distortions when principal 2 finds it also worth to intervene:

$$\overline{q}(\theta) = \max\left\{s_1 + s_2 - \theta - 2\frac{F(\theta)}{f(\theta)}, s_1 - \theta - \frac{F(\theta)}{f(\theta)}, 0\right\},$$

This example nicely illustrates two sources of inefficiency in delegated common agency games. First, there are the usual extra *two-fold* incentive distortion that

[18] The above construction could be replicated to sustain equilibria with multiple discontinuities.

already arises in intrinsic common agency games.[19] Second, another distortion comes from limited participation; which can be interpreted as an extreme form of free-riding. Indeed, it is straightforward to observe that there exists an non-empty interval for θ such that $s_2 \leq \frac{F(\theta)}{f(\theta)} \leq s_1$ where only principal 1 is active under asymmetric information. Under complete information, instead, both principals would be active on that interval.

Another interpretation of such limited participation is that, under asymmetric information, exclusive contracting can emerge endogenously even if exclusivity clause cannot be enforced. This is so even if, under complete information, both principals would like to contract with the agent.[20]

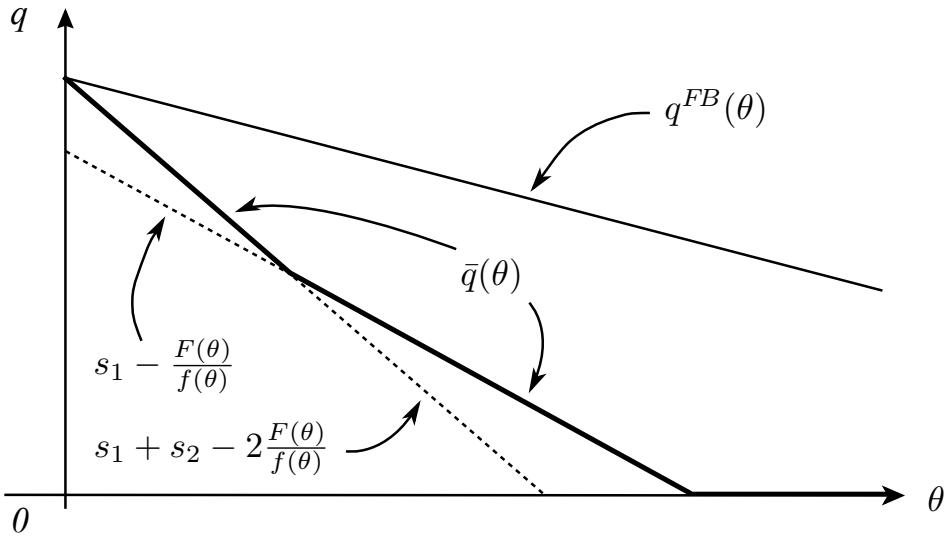

Figure 1: Voluntary provision of a public good.

6.1.2. *Public goods game with linear payoffs* $q \in [0,1]$.

Suppose now that $\mathcal{Q} = [0,1]$ and $S_0(q) = 0$, i.e., the agent ha not value for the good. For simplicity, assume that $\sum_{i \in N} s_i \geq \underline{\theta}$ (i.e., the first best has some public good provision) and that, for each principal i, there is a unique solution, $\hat{\theta}_i$, satisfying $F(\hat{\theta}_i) - s_i f(\hat{\theta}_i) = 0$. The latter implies $\overline{\theta} \geq \hat{\theta}_1 \geq \cdots \geq \hat{\theta}_n \geq \underline{\theta}$. Applying Theorem 3, the set of virtually-truthful allocations is the solution set to (5.2). Under our simplifying assumptions, there is a unique $\hat{\theta}$ such that

$$\sum_{i \in N} \max \left\{ s_i - \frac{F(\hat{\theta})}{f(\hat{\theta})}, 0 \right\} = \hat{\theta}.$$

The following conclusion follows largely from the results in Section 5.

[19] See Martimort and Stole (2011).

[20] This finding is reminiscent of the insights developed by Bernheim and Whinston (1998) in an I.O. context. Those authors showed that exclusive dealing in marketing practices arises when the agency costs of a common representation are too large compared with those under an exclusive representation even though under complete information, common agency is at least always weakly optimal. In our context, those agency costs are due to the agent's private information and not to hidden actions.

PROPOSITION 10 *In the public goods game with linear preferences, a virtually-truthful equilibrium allocation satisfies*

$$\overline{q}(\theta) = \begin{cases} 1 & \theta < \hat{\theta} \\ 0 & \theta > \hat{\theta}. \end{cases} \tag{6.2}$$

Virtually-truthful schedules are linear in q

$$\overline{t}_i(q) = \max\left\{ s_i - \frac{F(\hat{\theta})}{f(\hat{\theta})}, 0 \right\} q, \quad q \in [0,1],\ i \in N.^{21} \tag{6.3}$$

6.2. *Influence and Lobbying Games*

Consider now two competing interest groups having instead conflicting preferences $S_1(q) = -S_2(q) = q$. For instance, principal 1 enjoys higher tax rates q whereas principal 2 prefers lower tax rates. The decision-maker (agent) has some ideal policy he would like to pursue in the absence of any influence by lobbying groups. To model these intrinsic preferences, let assume that $S_0(q) = -\frac{q^2}{2}$ where $q \in \mathcal{Q} = [-q_{max}, q_{max}]$ with q_{max} being large enough to ensure interior solutions. Assume also that the agent's ideal point $q_0(\theta) = -\theta$ is symmetrically distributed over $[-\delta, \delta]$ with $\delta < 1$.[22] Choosing this bliss point gives a status quo payoff $U_0(\theta) = \frac{\theta^2}{2}$ to the agent.[23]

SEGMENTED MARKETS FOR INFLUENCE. Still relying on the general methodology developed in Theorem 3 and Proposition 7, we obtain:

PROPOSITION 11 *The virtually-truthful equilibrium allocation of the lobbying game for $s_1 > 0 > s_2$ is*

$$\overline{q}(\theta) = \max\left\{ s_1 - \frac{F(\theta)}{f(\theta)}, 0 \right\} + \min\left\{ s_2 + \frac{1 - F(\theta)}{f(\theta)}, 0 \right\} - \theta. \tag{6.4}$$

If θ is uniformly distributed, the activity sets of the principals are

$$\overline{\Theta}_1 = [-\delta, \min\{s_1 - \delta, \delta\}) \quad \textit{and} \quad \overline{\Theta}_2 = (\max\{\delta + s_2, -\delta\}, \delta]. \tag{6.5}$$

If type heterogeneity is small relative to the strength of the principals' preferences,

$$\delta < \frac{s_1 + |s_2|}{2},$$

then the principals commonly influence a positive measure of intermediate-type agents; otherwise, each principal has a separate domain of influence.

[21] Similar models have been studied by Martimort (1996) and Lebreton and Salanié (2003). These papers have principals (regulatory bodies in Martimort (1996) and lobbying groups in Lebreton and Salanié (2003)) competing with single transfers –a positive payment if and only if the agent implements the decision they prefer – in a context where the choice of policy is binary. Proposition 10 highlights conditions under which similar results also hold when policy choices are continuous and no restriction is a priori imposed on contribution schedules.

[22] Since principals are symmetrically biased in opposite directions, they would just agree on letting the agent choose his status quo policy had they cooperated.

[23] We could mirror the analysis in Section 6.1.2 and entertain the possibility that competing interest groups are opposed on whether to undertake a discrete policy (allowing free trade or not, allowing some drugs or not, etc...). We leave to the reader to develop such straightforward extension.

The transfers which implement this equilibrium allocation can be found by inverting $\overline{q}(\theta)$ and substituting this expression for $\vartheta(q)$ into (5.6)-(5.7). The lobbying model shows that only decision-makers with mild preferences receive contributions from both interest groups. Unchallenged influence only arises endogenously for the decision-makers who are the most "ideologically" oriented.[24]

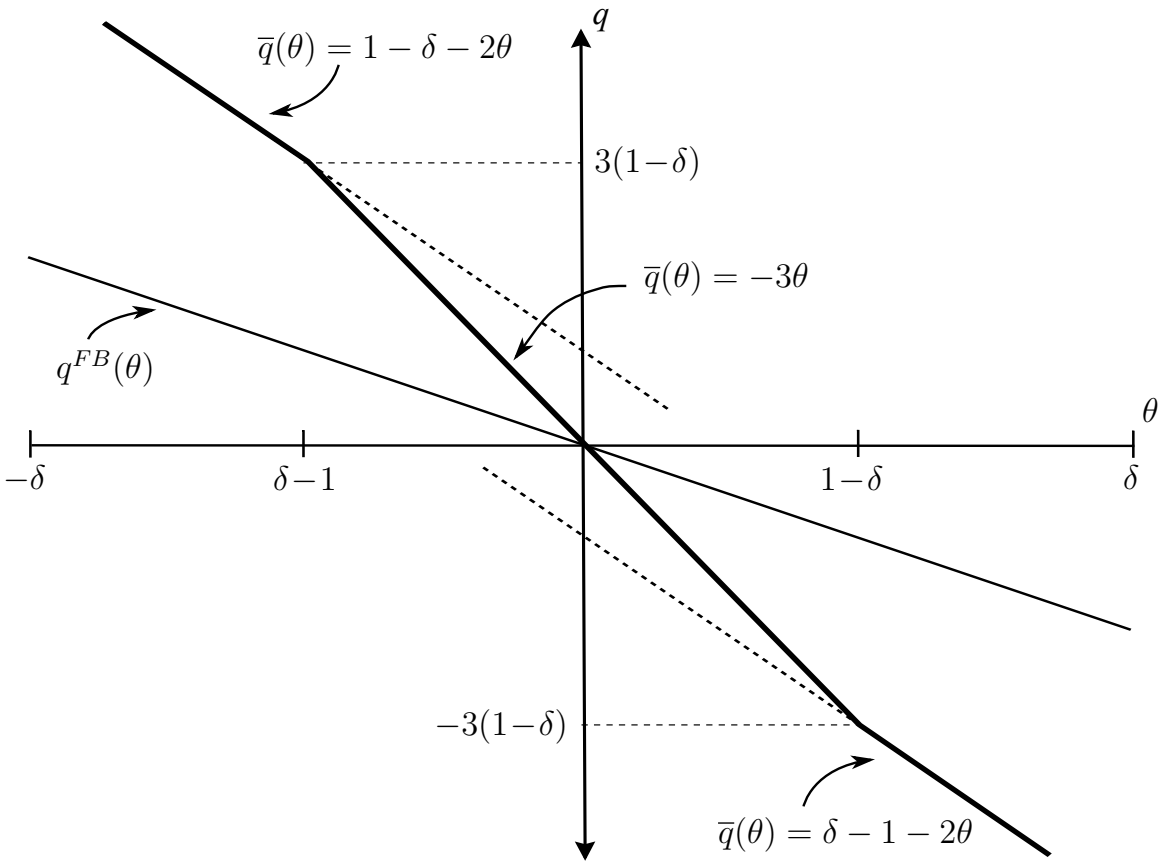

Figure 2: Virtually-truthful equilibrium allocation in a symmetric lobbying game, $s_1 = -s_2 = 1$, with $\delta \in (\frac{1}{2}, 1)$. Common influence arises for types $\theta \in (\delta - 1, 1 - \delta)$.

Discontinuous equilibria in the lobbying game. In that specific lobbying context, we can establish a version of Proposition 8. Taking a uniform and symmetric distribution around zero, we modify the "virtually-truthful" allocation by introducing a discontinuity at zero and provide an upper bound on how strong this discontinuity can be.

Proposition 12 *Suppose that* $s_1 = -s_2 = 1$ *and* $S_0(q) = -\frac{q^2}{2}$ *and that* θ *is uniformly distributed on* $\Theta = [-\delta, \delta]$ *with* $\delta < \frac{1}{2}$. *There exists a continuum of equilibria with a discontinuity at* $\theta_0 = 0$ *and such that* $\overline{q}^d(\theta_0^-) = -\overline{q}^d(\theta_0^+) = q_0$ *and*

[24] Martimort and Semenov (2008) derive further results on the patterns of contributions in a lobbying game. That latter paper did not derive the continuity of equilibrium output as we do here and focused on slightly different objective functions.

$q_0 \in (0, (1-\delta)\sqrt{3}]$. *The corresponding allocation is then*

$$\overline{q}^d(\theta) = \begin{cases} \overline{q}(\theta) \text{ for } \theta \in [-\delta, -q_0/3] \bigcup [q_0/3, \delta], \\ q_0 \text{ for } \theta \in [-q_0/3, 0), \\ -q_0 \text{ for } \theta \in (0, q_0/3]. \end{cases} \tag{6.6}$$

Aggregate welfare in those discontinuous equilibria is lower than the "virtually-truthful" allocation $\overline{q}(\theta) = -3\theta$.

A COMPETITIVE NONLINEAR PRICING REINTERPRETATION. Interestingly, the lobbying model of this section can be transposed *mutatis mutandis* in an I.O. context to study how a consumer having private information on his most preferred bundle mixes between two goods marketed by two competing sellers. Suppose that this consumer wants to acquire one unit of an homogenous good and is located at a point $\theta \in [0, 1]$ on a unit line, one seller being located at each extreme. The consumer has a valuation v for the good and incurs a quadratic transportation cost $-\frac{1}{2}(q-\theta)^2$ when moving away from his "ideal mix" where he consumes $q_0(\theta) = \theta$ from principal 1 and $1 - q_0(\theta)$ from principal 2. Up to some normalizations, the consumer and the sellers' profits are similar to those of the lobbying model above when the marginal cost is the same for both sellers and equal to one. Our previous results can be reinterpreted as giving conditions under which a share of the market is always covered by both sellers. Unless exclusivity is exogenously imposed, mixed bundling arises and market-sharing by means of exclusive contracts cannot be an equilibrium. Hoernig and Valletti (2010) have independently derived a similar insight but, at the outset, restricted their analysis to smooth tariffs. In particular, their analysis does not allow for discontinuous tariff; a restriction which may be justified by the fact that aggregate welfare is greater with the virtually-truthful and continuous allocation they focus on.

7. CONCLUSION

This paper has developed a methodology for solving public delegated common agency games under asymmetric information. In a nutshell, the basic economic insights of this research is that the well-known rent-efficiency trade-off must now significantly be modified to account for the impact of competition among principals. First, the compounded output distortions that arise at equilibrium reflect whether principals have conflicting or congruent objectives. Moreover, those distortions can all be captured by observing that an equilibrium allocation must maximize the "virtual aggregate" surplus of principals which may lead to simple modified Lindahl-Samuelson conditions. Second, contributions for some equilibria are no longer truthful but instead "virtually truthful" to account for the agency costs that each principals incur. Third, a rich pattern of activity sets that reflects the congruence or conflict between principals may emerge at equilibrium and we illustrated those facts by means of examples.

Other specific settings of economic interest (trade, regulation, multi-unit auctions, common representation on retailer markets, etc...) that have already been deeply studied by means of the common agency methodology in a world of complete information would benefit from a serious consideration of agency problems using our methodology. Inefficient representations of interest groups in trade negotiations, endogenous

limited entry in auction contexts, exclusive dealing agreements on retailing markets are, among others, interesting and important economic issues that can be explained by a careful study of the activity sets of some principals in specific contexts.

Beyond, our paper also suggests a few alleys for more theoretical works. First, the techniques developed in this paper might also be useful in private common agency games as well, i.e., when principals rely on different screening variables to control the agent. The ultimate objectives of such investigation should be to describe patterns of market coverage where either exclusive or multiple purchases endogenously arise on different subset of the type space. Such investigation is particularly important since students of market competition have generally found hard to reconcile market data with existing models of competition in nonlinear prices.[25]

Second, some applied settings may require to develop a framework where principals share some common screening devices but keep others private. For instance, one may think of specific games between competing manufacturers dealing with the same retailers and contracting on some commonly observed price downstream but keeping their sales of intermediary goods secret. These settings lie somewhere in between the case of public delegated agency games and the case of private agency games. Extending our methodology to semi-public environments may be important.

We plan to investigate such extensions in future research.

REFERENCES

Aliprantis, C. and K. Border (1990). *Infinite Dimensional Analysis*, 2nd ed. Berlin: Springer.

Alonso, N. and R. Matouschek (2008). Optimal Delegation. *Review of Economic Studies* **75** 259–293.

Bagnoli, C. and T. Bergstrom (2005). Log-Concave Probability and its Applications. *Economic Theory* **26** 445–469.

Bergstrom, T., L. Blume and H. Varian (1986). On the Private Provision of Public Goods. *Journal of Public Economics* **29** 25–49.

Bernheim, D. and M. Whinston (1986a). Menu Auctions, Resource Allocations and Economic Influence. *Quarterly Journal of Economics* **101** 1–31.

Bernheim, D. and M. Whinston (1986b). Common Agency. *Econometrica* **54** 923–942.

Bernheim, D. and M. Whinston (1998). Exclusive Dealing. *Journal of Political Economy* **106** 64–103.

Biais, B., D. Martimort and J.C. Rochet (2000). Competing Mechanisms in a Common Value Environment. *Econometrica* **68** 799–837.

Calzolari, G. and V. De Nicolo (2010). Competition With Exclusive Contracts and Market-Share Discounts. CEPR Discussion Paper 7613.

Calzolari, G. and C. Scarpa (2008). Delegated and Intrinsic Common Agency with Full Market Participation. Unpublished manuscript, University of Bologna.

Champsaur, P. and J.C. Rochet (1989). Multiproduct Duopolists. *Econometrica* **57** 533–557.

[25]Such models either impose exclusivity clauses (Stole, 1995, Rochet and Stole, 2002) or restrict the analysis to cases where a buyer jointly buys from different sellers (Ivaldi and Martimort 1994, Martimort and Stole, 2009a, Calzolari and De Nicolo, 2010). Stole (2007) surveys this literature.

Chiesa, G. and V. Denicolo (2009). Trading with a Common Agent under Complete Information: A Characterization of Nash Equilibria. *Journal of Economic Theory* **144** 296–311.

Clarke, F. (1990). *Optimization and Nonsmooth Analysis*. Philadelphia: SIAM.

Dixit, A., G. Grossman and E. Helpman (1997). Common Agency and Coordination: General Theory and an Application to Government Decision-Making. *Journal of Political Economy* **105** 752–769.

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

Grossman, G. and E. Helpman (1994). Protection for Sale. *American Economic Review* **84** 833–850.

Hoernig, S. and T. Valletti (2010). When Two-Part Tariffs Are Not Enough: Mixing With Nonlinear Pricing. Unpublished manuscript, CEPR Discussion Paper 7720.

Holmström, B. (1984). On the Theory of Delegation, in *Bayesian Models in Economic Theory,* ed. by M. Boyer and R. Kihlstrom, Elsevier Science B. V.

Innes, R., (1990). Limited liability and incentive contracting with ex-ante action choices," *Journal of Economic Theory* **52** 47–65.

Ivaldi, M. and D. Martimort (1994). Competition under Nonlinear Pricing. *Annales d'Economie et de Statistique* **34** 71–114.

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

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

Laussel, D. and M. Lebreton (1998). Efficient Private Production of Public Goods under Common Agency. *Games and Economic Behavior* **25** 194–218.

Laussel, D. and M. Lebreton (2001). Conflict and Cooperation: The Structure of Equilibrium Payoffs in Common Agency. *Journal of Economic Theory* **100** 93–128.

Lebreton, M. and F. Salanié (2003). Lobbying under Political Uncertainty. *Journal of Public Economics* **87** 2589–2610.

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

Loewen, P. and T. Rockafellar (1995). New Necessary Conditions for the Generalized Bolza Problem. *SIAM Journal of Control and Oprimization* **49** 294–313.

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

Martimort, D. (1992). Multi-Principaux avec Anti-Selection. *Annales d'Economie et de Statistiques* **28** 1–38.

Martimort, D. (1996). The Multi-principal Nature of Government. *European Economic Review* **40** 673–685.

Martimort, D. (2007). Multicontracting Mechanism Design. *Advances in Economic Theory, Proceedings of the 2005 World Congress of the Econometric Society*, ed. by R. Blundell, W. Newey and T. Person. Cambridge: Cambridge University Press.

Martimort, D. and A. Semenov (2006). Continuity in Mechanism Design without Transfers. *Economic Letters*, **93** 182–189.

Martimort, D. and A. Semenov (2008). Ideological Uncertainty and Lobbying Competition.

Journal of Public Economics **92** 456–481.

Martimort, D. and L. Stole (2002). The Revelation and Delegation Principles in Common Agency Games. *Econometrica* **70** 1659–1674.

Martimort, D. and L. Stole (2003). Contractual Externalities and Common Agency Equilibria. *Advances in Theoretical Economics* **3**(1), Article 4, http://www.bepress.com/bejte.

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

Martimort, D. and L. Stole (2009b). Selecting Equilibria in Common Agency Games. *Journal of Economic Theory* **144** 604–634.

Martimort, D. and L. Stole (2011). Aggregate Representations of Aggregate Games with Applications to Common Agency Games. Unpublished manuscript, University of Chicago.

Melumad, N. and T. Shibano (1991). Communication in Settings with no Transfers. *RAND Journal of Economics* **22** 173–198.

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

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

Peters, M. and B. Szentes (2011). Definable and Contractible Contracts. Unpublished manuscript, University College of London, forthcoming *Econometrica*.

Rochet, J.C. and L. Stole (2002). Nonlinear Pricing with Random Participation. *Review of Economic Studies* **69** 277–311.

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

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.

Stole, L. (1991). Mechanism Design under Common Agency. Unpublished manuscript, University of Chicago.

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

Stole, L. (2007). Price Discrimination in Competitive Environments. *Handbook of Industrial Organization, Volume 3*, ed. by Armstrong, M. and Porter, R. Amsterdam: North Holland, 2221–2299.

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

Vinter, R. and H. Zheng (1995). Necessary Conditions for Optimal Control Problem with State Constraints. *Transactions of the American Mathematical Society* **350** 1181–1204.

APPENDIX A: PROOFS OF MAIN RESULTS

7.1. *Preliminaries*

For completeness, we review a few definitions and facts from convex analysis, specializing them for our present purposes in Lemmas 3-6. The reader familiar with convex analysis may wish to skip to Section 7.2. We make repeated references to the volume by Rockafellar and Wets (2004) (hereafter RW). Equipped with Lemmas 3-6, we shall more efficiently prove the results presented in the main text.

Let $W : \mathcal{Q} \to \mathbb{R}$ be any finite-valued, upper semi-continuous function defined over a compact domain. For any $\sigma \in \mathbb{R}$, the value function $V : \mathbb{R} \to \mathbb{R}$

$$V(\sigma) \equiv \max_{q \in \mathcal{Q}} W(q) + \sigma q$$

is well defined. Because it is the pointwise maximum of a family of affine functions, it is a convex, proper, and lower semi-continuous (RW, 8.13.). Given our assumptions on boundedness and compactness, for any $\sigma \in \mathbb{R}$, $V(\sigma)$ is finite-valued (and hence proper with effective domain $\mathbb{R}$). It follows that V is a continuous function on $\mathbb{R}$.

Because V is a proper convex function, for any point in its domain we can define its **subdifferential**

$$\partial V(\sigma) \equiv \{y \in \mathbb{R} \,|\, V(s) \geq V(\sigma) + y(s - \sigma) \text{ for all } s\}.^{26}$$

Because V is a proper, convex function, the left and right derivatives $V'_+(\sigma)$ and $V'_-(\sigma)$ exist and satisfy $V'_+(\sigma) \geq V'_-(\sigma)$, and the subdifferential has the simple representation (RW, 8.52):

$$\partial V(\sigma) = \left\{y \in \mathbb{R} \,|\, V'_+(\sigma) \geq y \geq V'_-(\sigma)\right\}.$$

Subdifferentials of convex functions have several other important properties:

LEMMA 3 *Let $V : \mathbb{R} \to \mathbb{R}$ be a proper, convex function and $\partial V : \mathbb{R} \to \mathbb{R}$ its subdifferential:*

- *∂V is* **monotone** *(RW, 12.1, 12.17), i.e.,*

  $$(x_0 - x_1)(y_0 - y_1) \geq 0 \text{ whenever } y_0 \in \partial V(x_0),\ y_1 \in \partial V(x_1).$$

  Furthermore, V is strictly convex on its domain if and only if ∂V is **strictly monotone***:*

  $$(x_0 - x_1)(y_0 - y_1) > 0 \text{ whenever } y_0 \in \partial V(x_0),\ y_1 \in \partial V(x_1).$$

- *The graph of ∂V is closed and convex valued (RW, 12.8).*
- *∂V is almost everywhere single-valued and differentiable, i.e., V is a.e. twice differentiable, (RW-12.66).*
- *∂V is continuous at any point σ where it is single-valued $\partial V(\sigma) = \{V'(\sigma)\}$ (RW-12.63).*

Define the **dual conjugate**[27] of V as the value of the dual program

$$V^*(q) \equiv \max_{\sigma \in \mathbb{R}} \sigma q - V(\sigma).$$

Here $V^* : \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ is a possibly extended-valued function (i.e., its effective domain may be a strict subset of $\mathbb{R}$). Nonetheless, like the value function V, conjugates are necessarily proper, convex, lower semi-continuous functions, and they are

[26] When V is a proper *concave* function, for any point in its domain we can define its **superdifferential**

$$\partial_* V(\sigma) \equiv \{y \in \mathbb{R} \,|\, V(s) \leq V(\sigma) + y \cdot (s - \sigma) \text{ for all } s\}.$$

Of course $\partial_* V = -\partial(-V)$.

[27] This is also known as the Legendre-Fenchel transform.

continuous on their domains. As such, the subdifferential of V^* exists and has the properties listed in Lemma 3. Also note that by construction, $-V^*(q) \geq W(q)$ for all $q \in \mathcal{Q}$.

We now illustrate some of the important duality properties of V, V^* and their subdifferentials. First, the subdifferential allows us to generalize Fermat's first-order rule. If $\overline{q} \in \arg\max_{q\in\mathcal{Q}} W(q) + \sigma q$, it follows that $\overline{q} \in \partial V(\sigma)$. (RW, 10.1.) If $\overline{q} \in \partial V(\sigma)$, it follows that $\overline{q} \in \arg\max_{q\in\mathbb{R}} -V^*(q) + \sigma q$, and hence

$$-V^*(\overline{q}) + \sigma\overline{q} \geq W(q) + \sigma q, \ \forall\, q \in \mathbb{R}.$$

Hence, if $\overline{q} \in V(\sigma)$ and $\overline{q} \in \mathcal{Q}$, then $\overline{q}$ is also a solution to the primal program $\max_{q\in\mathcal{Q}} W(q) + \sigma q$. We have the following statement of these facts.

LEMMA 4 *For all $\sigma \in \mathbb{R}$,*

$$\partial V(\sigma) \cap \mathcal{Q} \quad \subseteq \quad \arg\max_{q\in\mathcal{Q}} \ W(q) + \sigma q \quad \subseteq \quad \partial V(\sigma).$$

Second, there is a useful connection between the subdifferential correspondences of the conjugate pair, V and V^*.

LEMMA 5 *(RW, 11.3)*

- *Each subdifferential is the inverse correspondence of the other: $\partial V^* = (\partial V)^{-1}$, $\partial V = (\partial V^*)^{-1}$. Indeed,*

  (A1) $\quad \overline{q} \in \partial V(\overline{\sigma}) \iff \overline{\sigma} \in \partial V^*(\overline{q}) \iff V(\overline{\sigma}) + V^*(\overline{q}) = \overline{\sigma}\overline{q},$

  whereas $V(\sigma) + V^*(q) \geq q\sigma$ for all $(q, \sigma) \in \mathbb{R}^2$ (Fenchel's inequality).
- *Moreover,*

  (A2) $\quad \partial V(\overline{\sigma}) = \arg\max_{q\in\mathbb{R}} \ q\overline{\sigma} - V^*(q), \qquad \partial V^*(\overline{q}) = \arg\max_{\sigma\in\mathbb{R}} \ \overline{q}\sigma - V(\sigma).$

In what follows, we will say that any (q, σ) which satisfies $\overline{q} \in \partial V(\overline{\sigma})$ (or the equivalent conditions in (A1)) is a **dual pair**.

One final property that we will exploit below is the duality between the differentiability of one value function and the strict convexity of its conjugate pair.

LEMMA 6 *(RW, 11.8, 11.13.) For any proper, lower semi-continuous, convex function V and its conjugate V^*,*

$$\partial V(\sigma) = \{q\} \iff V^*(q) \ \textit{differentiable},$$

$$\partial V^*(q) = \{\sigma\} \iff V(\sigma) \ \textit{differentiable}.$$

Moreover, $\partial V(\sigma)$ (respectively, ∂V^*) is single-valued at q if and only if V (respectively, V^*) is locally strictly convex.

In the proof of the theorems from the main text, we will make heavy use of Lemmas 5 and 6, applied to two different choices of W. When $W \equiv S_0 + t_i + \overline{T}_{-i}$, we define the agent's value function, its dual conjugate and its dual argument respectively as

(A3) $\quad V_0(\sigma) \equiv \max_{q\in\mathcal{Q}} \ S_0(q) + t_i(q) + \overline{T}_{-i}(q) + \sigma q,$

$$\text{(A4)} \qquad V_0^*(q) \equiv \max_{\sigma \in \mathbb{R}} \sigma q - V_0(\sigma),$$

$$\text{(A5)} \qquad \overline{\sigma}_0(\theta) \equiv -\theta.$$

Those quantities will allow us to describe the agent's optimal behavior. Alternatively, to describe principal i's optimal behavior, we apply the previous definitions and Lemmas for the case where $W \equiv W_i \equiv S_0 + S_i + \overline{T}_{-i}$. The value function V_i and its conjugate V_i^* then become

$$\text{(A6)} \qquad V_i(\sigma) \equiv \max_{q \in \mathcal{Q}} S_0(q) + S_i(q) + \overline{T}_{-i}(q) + \sigma q,$$

$$\text{(A7)} \qquad V_i^*(q) \equiv \max_{\sigma \in \mathbb{R}} \sigma q - V_i(\sigma)$$

In this case, we may define the associated dual variable defined as

$$\text{(A8)} \qquad \overline{\sigma}_i(\theta) \equiv \frac{\overline{M}_i(\theta) - F(\theta)}{f(\theta)} - \theta.$$

where $\overline{M}_i(\theta)$ is a non-decreasing left-continuous function.

7.2. *Proofs of statements in the main text*

Proof of Lemma 1: Necessity: Because $t_i \geq 0$, $U(\theta) \geq \overline{U}_{-i}(\theta)$. By definition, $U(\theta) = V_0(-\theta)$ and hence it is convex from Lemma 4. Moreover, $-q(\theta) \in \partial U(\theta)$.

Sufficiency: Given that $U(\theta) \geq \overline{U}_{-i}(\theta)$, the offer is acceptable. U is convex, decreasing, and it is supported by $-q(\theta)$ at each $\theta \in \Theta$. Given that $q(\theta)$ has domain $\mathcal{Q}$, Lemma 4 implies that $q(\theta)$ is optimal for the agent. Hence, (q, U) is implementable. *Q.E.D.*

Proof of Theorem 1: Step 1: The relaxed program. Consider the relaxed program, $(\mathcal{P}_i^r)$, that ignores the convexity constraint (2.3):

$$(\mathcal{P}_i^r) : \max_{(U \in AC, q \in \mathcal{Q}^\Theta)} \int_\Theta \left(S_i(q(\theta)) + S_0(q(\theta)) + \sum_{j \neq i} t_j(q(\theta)) - \theta q(\theta) - U(\theta) \right) f(\theta) d\theta$$

$$\text{subject to (2.1), (2.2).}$$

We apply Theorem B.1 (which is stated and proved in Appendix B) and conclude that for any given profile of transfers $\mathbf{t}_{-i}$ offered by rival principals the rent-output profile $(\overline{U}, \overline{q})$ is a solution to $(\mathcal{P}_i^r)$ if and only if $(\overline{U}, \overline{q})$ satisfies (2.1), (2.2) and there exists a probability measure μ_i defined over the Borel subsets of Θ with an associated adjoint function, $\overline{M}_i : \Theta \to [0, 1]$, defined by $\overline{M}_i(\underline{\theta}) = 0$ and for $\theta > \underline{\theta}$,

$$\overline{M}_i(\theta) \equiv \int_{[\underline{\theta}, \theta)} \mu_i(d\theta),$$

such that the following two conditions are satisfied:

$$\text{(A9)} \qquad supp\left\{\mu_i\right\} \subseteq \left\{\theta \,|\, \overline{U}(\theta) = \overline{U}_{-i}(\theta)\right\},$$

$$\text{(A10)} \quad \overline{q}(\theta) \in \arg\max_{q\in\mathcal{Q}} \; S_i(q) + S_0(q) + \sum_{j\neq i} t_j(q) + \left(\frac{\overline{M}_i(\theta) - F(\theta)}{f(\theta)} - \theta\right) q, \; a.e.$$

Step 2: Convexity of "relaxed" solution: To demonstrate that the solution $\overline{U}$ to the relaxed program is convex it is sufficient to show that there is a non-increasing selection within the best-response correspondence defined by (A10). If so, the conditions of the "relaxed" Theorem apply to the more constrained program obtained by appending (2.3) to (2.1) and (2.2).

Define the value function $V_i(\sigma)$ and $\overline{\sigma}_i(\theta)$ as in (A6) and (A8) respectively. On principal i's activity set, $\overline{\Theta}_i$, derived in the relaxed program, $\overline{M}_i$ is constant for all connected intervals (by Equation (A9)). By assumption, $(F(\theta)-\kappa)/f(\theta)$ is increasing for any $\kappa \in [0,1]$, and so $\overline{\sigma}_i$ defined above is strictly decreasing on $\overline{\Theta}_i$. Thus for all θ in any connected subset of $\overline{\Theta}_i$, the correspondence $\partial V_i(\overline{\sigma}_i(\theta))$ is monotone decreasing. (Lemma 3.)

Restating the optimality condition (3.2) in Theorem 1 as

$$\overline{q}(\theta) \in \arg\max_{q\in Q} \; W_i(q) + \sigma(\theta)q, \; a.e.,$$

with $V_i(\sigma(\theta)) = W_i(\overline{q}(\theta)) + \sigma(\theta)\overline{q}(\theta)$, a.e., we obtain

Lemma 7 *Suppose that $\overline{q}$ is feasible, $\overline{q}(\Theta) \subseteq \mathcal{Q}$, and that $W = S_0 + S_i + \overline{T}_{-i}$. Then*

$$\text{(A11)} \quad (3.2) \iff \overline{q}(\theta) \in \partial V(\overline{\sigma}(\theta)), \; a.e. \iff \overline{\sigma}(\theta) \in \partial V^*(\overline{q}(\theta)), \; a.e.$$

Thus, any solution to the relaxed program satisfies $\overline{q}(\theta) \in \partial V_i(\overline{\sigma}_i(\theta))$ a.e. (Condition (A11)). Because $\partial V_i(\overline{\sigma}_i(\theta))$ is monotone decreasing in θ, we can find a non-increasing solution $\tilde{q}(\theta) \in \partial V_i(\overline{\sigma}_i(\theta))$ for every $\theta \in \overline{\Theta}_i$. Hence, one can choose $\overline{q}$ that solves the relaxed program, and is non-increasing on $\overline{\Theta}_i$.

On any open interval of $\overline{\Theta}_i^c$, we know that $\overline{U}(\theta) = \overline{U}_{-i}(\theta)$, and thus $\overline{q}(\theta) = \overline{q}_{-i}(\theta)$, *a.e.*. Because the latter is non-increasing, we conclude that one can choose $\overline{q}$ that solves the relaxed program and is non-increasing on the interior of $\overline{\Theta}_i^c$.

What remains to be shown is that $\overline{q}$ is non-increasing at any boundary point of $\overline{\Theta}_i^c$. Let $\hat{\theta}$ be such a point. Suppose that $\overline{M}_i$ is continuous at $\hat{\theta}$, so $\overline{\sigma}_i$ is also. On either side of $\hat{\theta}$, our previous argument established that $\overline{q}$ was non-increasing, and therefore $\overline{\sigma}_i$ is non-increasing. Thus, $\partial V_i(\overline{\sigma}(\theta))$ is monotone decreasing in θ and, from the first inclusion in (A11), $\overline{q}$ cannot jump upwards at $\hat{\theta}$. Suppose instead that $\overline{M}_i$ jumps upwards at $\hat{\theta}$. Because $\overline{M}_i$ is left-continuous and μ_i can only have an atom at a point in $\hat{\Theta}_i^c$, it follows that the adjoint function must jump upwards from $\overline{M}_i(\hat{\theta})$ to $\overline{M}_i(\hat{\theta}^+)$. For the neighborhood $[\hat{\theta}, \hat{\theta}+\varepsilon)$, $\overline{U}(\theta) = \overline{U}_{-i}(\theta)$. For the neighborhood $(\hat{\theta}-\varepsilon, \hat{\theta})$, $\overline{U}(\theta) > \overline{U}_{-i}(\theta)$. Because $\overline{U}_{-i}$ is convex and minorizes $\overline{U}$ at $\hat{\theta}$, it must be that $\overline{U}$ is locally convex at $\hat{\theta}$. But then Lemma 1 implies that $\overline{q}$ cannot jump upwards. It follows that any solution that is non-increasing on $\overline{\Theta}_i$ and $\text{int}(\overline{\Theta}_i^c)$ must be non-increasing over all Θ. Hence, $\overline{q}$ is a solution to the relaxed program that is non-increasing. *Q.E.D.*

PROOF OF THEOREM 2: The condition in (3.2) holds for almost every $\theta \in \Theta$ and for any $i \in N$. Thus, fixing θ, we may add up all maximands in (3.2) for $i \in \overline{\alpha}(\theta)$. It follows that $\overline{q}(\theta)$ also maximizes this aggregate maximand in (4.1) almost everywhere.

We obtain (4.2) by summing only over active principals, expanding terms and noting that $\sum_{i\in\overline{\alpha}(\theta)} \sum_{j\in N\setminus\{i\}} \bar{t}_j(q) = (|\overline{\alpha}(\theta)| - 1)T(q) + \sum_{j\notin\overline{\alpha}(\theta)} \bar{t}_j(q)$, we get:

$$\overline{q}(\theta) \in \arg\max_{q\in\mathcal{Q}} \; S_0(q) - \theta q + \sum_{i\in\overline{\alpha}(\theta)} S_i(q) + \sum_{i\notin\overline{\alpha}(\theta)} \bar{t}_i(q) + \sum_{i\in\overline{\alpha}(\theta)} \left(\frac{\overline{M}_i(\theta) - F(\theta)}{f(\theta)}\right) q \\ + (|\overline{\alpha}(\theta)| - 1)\left(S_0(q) - \theta q + \overline{T}(q)\right), \; a.e.$$

Because $\sum_{i\notin\overline{\alpha}(\theta)} \bar{t}_i(\overline{q}(\theta)) = 0$ by definition and $\sum_{i\notin\overline{\alpha}(\theta)} \bar{t}_i(q) \geq 0$, we get the result. *Q.E.D.*

To prove Proposition 1, we will need the following Lemma.

LEMMA 8 *Consider any interval of bunching, $[\theta_0, \theta_1]$, over which $\overline{q}(\theta) = \hat{q}$ and such that $\hat{\theta} \in [\theta_0, \theta_1] \bigcap \overline{\Theta}_i^c$. Then, $[\theta_0, \theta_1] \subseteq \overline{\Theta}_i^c$.*

PROOF OF LEMMA 8: Since $\hat{\theta} \in \overline{\Theta}_i^c$, $\overline{U}(\hat{\theta}) = \overline{U}_{-i}(\hat{\theta})$. Because $\overline{U}(\theta) \geq \overline{U}_{-i}(\theta)$ for all $\theta \in \Theta$ and $\overline{U}_{-i}$ is convex, it follows that $\overline{U}$ must be locally more convex than $\overline{U}_{-i}$ around $\hat{\theta}$. But $\overline{q}(\theta) = \hat{q}$ over $[\theta_0, \theta_1]$ implies that $\overline{U}$ is an affine function on $[\theta_0, \theta_1]$ and so, necessarily $\overline{U}(\theta) = \overline{U}_{-i}(\theta)$ over this whole interval. Finally, $[\theta_0, \theta_1] \subseteq \overline{\Theta}_i^c$. *Q.E.D.*

PROOF OF PROPOSITION 1:
Implication 1: The complementary slackness condition (3.1) in Theorem 1 requires that $\mu_i(\theta) = 0$ for all $\theta \in \overline{\Theta}_i$. Hence, $\overline{M}_i$ must be constant on any connected interval of $\overline{\Theta}_i$.
Implication 2: Here, we again refer to the value function and dual variable defined in (A6) and (A8) and the optimality conditions (A11).

Consider now an arbitrary point $\hat{\theta} \in \text{int}(\overline{\Theta}_i^c)$ around which $\overline{q}_{-i}$ is strictly decreasing. By definition, we have $U(\hat{\theta}) = \overline{U}_{-i}(\hat{\theta})$, and thus $\partial U(\hat{\theta}) = \partial\overline{U}_{-i}(\hat{\theta})$ for such $\hat{\theta}$ and one may as well select within those subgradient correspondences $\overline{q}(\hat{\theta}) = \overline{q}_{-i}(\hat{\theta})$. Furthermore, incentive compatibility requires that

$$\overline{q}(\theta) = \overline{q}_{-i}(\theta) \in \arg\max_{q\in\mathcal{Q}} \; S_0(q) + T_{-i}(q) - \theta q$$

for all $\theta \in \overline{\Theta}_i^c$. Because S_i is differentiable and concave,

$$\overline{q}(\theta) \in \arg\max_{q\in\mathcal{Q}} \; S_i(q) - S_i'(\overline{q}_{-i}(\theta))q$$

for all $\theta \in \overline{\Theta}_i^c$. Adding these two programs together yields

$$\overline{q}(\theta) \in \arg\max_{q\in\mathcal{Q}} \; S_0(q) + T_{-i}(q) - \theta q + S_i(q) - S_i'(\overline{q}_{-i}(\theta))q, \; \forall\theta \in \overline{\Theta}_i^c.$$

This is equivalent to the statement

$$\overline{q}(\theta) = \overline{q}_{-i}(\theta) \in \partial V_i(\overline{\sigma}_0(\theta)), \; \forall\theta \in \overline{\Theta}_i^c,$$

where $\overline{\sigma}_0(\theta) \equiv -(\theta + S_i'(\overline{q}_{-i}(\theta))$. We thus have the joint requirement that

$$\overline{q}(\theta) \in \partial V_i(\overline{\sigma}_0(\theta)) \cap \partial V_i(\overline{\sigma}_i(\theta)),\ a.e.\, \theta \in \overline{\Theta}_i^c.$$

Lemma 5 implies the equivalent statement

$$\overline{\sigma}_i(\theta), \overline{\sigma}_0(\theta) \in \partial V_i^*(\overline{q}_{-i}(\theta)),\ a.e.\, \theta \in \overline{\Theta}_i^c.$$

By hypothesis, $\overline{q}_{-i}$ is strictly decreasing in a neighborhood of $\hat{\theta}$ that is contained in $\overline{\Theta}_i^c$, and so Lemma 5 also implies that $\partial V_i^*(\overline{q}_{-i}(\theta))$ is single valued in the same neighborhood. Hence, $\overline{\sigma}_i(\theta) = \overline{\sigma}_0(\theta)$, *a.e.* in this neighborhood. This, in turn, can only be true if $\overline{M}_i(\theta) = F(\theta) - S_i'(\overline{q}_{-i}(\theta))f(\theta)$, *a.e.* in the neighborhood of $\hat{\theta} \in \text{int}(\overline{\Theta}_i^c)$. Because $\overline{M}_i$ is a non-decreasing function, $\overline{M}_i(\theta) = F(\theta) - S_i'(\overline{q}_{-i}(\theta))f(\theta)$ must hold for all θ in the open neighborhood of $\hat{\theta}$. This is the second implication.

Implications 1 and 2 of the Proposition must hold for *any* adjoint function which supports $(\overline{q}, \overline{U})$. We lastly prove that for any adjoint function that supports $(\overline{q}, \overline{U})$, there is a continuous adjoint function which does so as well. That is, there may be an equivalence class of adjoint functions which support $(\overline{q}, \overline{U})$; our continuity statement is that this class always contains at least one continuous member.

Implication 3: Because $\overline{M}_i$ is constant over $\overline{\Theta}_i$, any upward discontinuity point must lie in $\overline{\Theta}_i^c$. Because the density f is continuous, $\overline{M}_i$ is upward discontinuous if and only if the associated construction $\overline{\sigma}_i$ is upward discontinuous.

Consider an arbitrary upward jump at $\hat{\theta} \in \overline{\Theta}_i^c \cap (\underline{\theta}, \overline{\theta})$ and recall our construction of the proper convex value function, V_i, and the requirement of (3.2) that $\overline{q}(\theta) \in \partial V_i(\overline{\sigma}_i(\theta))$, *a.e.*. Because V_i is convex, ∂V_i is a monotone correspondence. Because $\overline{q}(\theta)$ is a non-increasing selection which lies in $\partial V_i(\overline{\sigma}_i(\theta))$ for a.e. θ, $\overline{\sigma}_i$ must be strictly decreasing over any interval for which $\overline{q}$ is strictly decreasing. In particular, $\overline{\sigma}_i$ cannot jump upwards (and hence $\overline{M}_i$ cannot jump upwards) over any region for which $\overline{q}$ is strictly decreasing from Lemma 3.

What remains is the possibility that $\overline{M}_i$ jumps upward on an inactive region of bunching $[\theta_0, \theta_1]$. From Lemma 8, we may as well take $[\theta_0, \theta_1]$ as the maximal interval of bunching over Θ_i^c.

Suppose now that a discontinuity point of $\overline{M}_i$ arises at some $\hat{\theta}$ which is not an interior point of $[\theta_0, \theta_1]$. Consider first the case $\hat{\theta} = \theta_0$. Because we assume that $\hat{\theta} \in (\underline{\theta}, \overline{\theta})$, it must be that $\hat{\theta} = \theta_0 > \underline{\theta}$. The local convexity of $\overline{U} - \overline{U}_{-i}$ implies that

$$\text{(A12)} \quad \overline{q}(\theta_0^-) \geq \hat{q} = \overline{q}_{-i}(\theta_0^+) = \overline{q}(\theta_0^+),$$

where the first inequality follows from the fact that $\bar{U}$ is convex, the second from the fact that there is bunching over $[\theta_0, \theta_0 + \varepsilon)$ for $\varepsilon > 0$ sufficiently small and principal i is inactive there. Suppose that θ_0 is a point of discontinuity of $\overline{q}$. Then, (A12) implies that $\overline{q}$ must have a downward jump at θ_0. But this is a contradiction with the fact that $\overline{M}_i$ (and therefore $\overline{\sigma}_i$) has an upward jump at θ_0. Hence, $\overline{q}(\theta_0^-) = \overline{q}(\theta_0^+)$ and $\overline{q}$ is continuous at θ_0. This in turn implies that that $\overline{M}_i$ is also right-continuous at θ_0.

The same kind of arguments eliminates the possibility of an upward jump for $\overline{M}_i$ at $\hat{\theta} = \theta_1 < \overline{\theta}$.

Suppose that $\tilde{M}_i$ supports $(\overline{q}, \overline{U})$ but that it has a discontinuity at an interior point of a bunching interval $[\theta_0, \theta_1] \subseteq \overline{\Theta}_i^c$. We now argue that for any such $\tilde{M}_i$, there exists

a modified $\overline{M}_i$ which is continuous and also supports $(\overline{q}, \overline{U})$. Let $\overline{M}_i = \tilde{M}_i$ for all θ except over the interior of bunching regions on which $\tilde{M}_i$ is discontinuous. Over these intervals, let $\overline{M}_i$ be a strictly increasing, continuous approximation to $\tilde{M}_i$ from below. Such an approximation exists and can be made arbitrarily close to $\tilde{M}_i$ so as to satisfy

$$\overline{\sigma}_i([\theta_0, \theta_1]) = \tilde{\sigma}_i([\theta_0, \theta_1]).$$

Because it is strictly increasing on $\overline{\Theta}_i^c$, the modified $\overline{M}_i$ satisfies (3.1) in Theorem 1. Because $\tilde{\sigma}_i([\theta_0, \theta_1]) = \partial V_i^*(\hat{q})$, we have, as required by (3.2),

$$\overline{\sigma}_i([\theta_0, \theta_1]) = \partial V_i^*(\hat{q}).$$

Q.E.D.

Proof of Proposition 2:

First claim: We say that a function f is *effectively strictly concave* (e.s.c) if its minimal upper concave envelope is strictly concave (or alternatively, if convex biconjugate $(-f)^{**}$) is strictly convex). We first prove that the continuity of $\overline{q}$ is equivalent to the effective strict concavity of $W_i \equiv S_0 + S_i + \overline{T}_{-i}$, $i \in N$ and the effective strict concavity of $W_0 \equiv S_0 + \overline{T}$. To this end, we again refer to the value function and dual variable defined in (A6) and (A8) and the optimality conditions (A11). By Proposition 1, $\overline{M}_i$ is continuous. It follows that if $\overline{q}$ is continuous at θ and $\overline{q}(\theta) \in \partial V_i(\overline{\sigma}_i(\theta))$, then $\partial V_i(\overline{\sigma}_i(\theta))$ must be single-valued because it is monotone by Lemma 3.

Lemma 6 establishes that the subdifferential of a convex value function is single-valued if and only if its conjugate is locally strictly convex, which is equivalent to the requirement that $-V^*(q)$ is strictly concave at $q = \overline{q}(\theta)$. This is also equivalent to the requirement that $W_i = S_0 + S_i + \overline{T}_{-i}$ is effectively strictly concave. Such an argument must be true for all $i \in N$. A similar argument also applies to $W_0 = S_0 + \overline{T}$. Hence,

$$\overline{q} \in C^0(\Theta) \iff -V_i^* \text{ strictly concave on } \mathcal{Q} \iff (\overline{U})^* \text{ strictly convex on } \mathcal{Q}.$$

Having established the equivalence relationships using the notion of effective strict concavity, we need only demonstrate that a function which is both effectively strictly concave and upper semi-continuous must itself be strictly concave. To this end, we record the following lemma.

Lemma 9 *$(-f)^{**}$ strictly convex, f u.s.c. $\Longrightarrow$ f is strictly concave.*

Proof of Lemma 9: By definition epi$(-f)^{**}$ is the smallest, closed convex set which contains epi$(-f)$. Because f is u.s.c., the epigraph of $-f$ is also closed. If these sets are equal, then $-f$ is strictly convex (and f is strictly concave). If these sets are different, then epi$(-f)$ must be non-convex over some region and therefore epi$(-f)^{**}$ must have a linear segment over this region. Hence, $(-f)^{**}$ is not strictly convex (i.e., f is not effectively strictly concave) as supposed. *Q.E.D.*

Second claim: Suppose that $\overline{q}$ is continuous at $\hat{\theta}_i$ which is on the boundary of principal i's activity set $\overline{\Theta}_i$ and interior (i.e., $\hat{\theta}_i \in \text{bd}\,\overline{\Theta}_i^c \bigcap \text{int}(\Theta)$). At such $\hat{\theta}_i$ we have

$$\overline{U}(\hat{\theta}_i) = \overline{U}_{-i}(\hat{\theta}_i)$$

where

$$\overline{U}(\hat{\theta}_i) = S_0(\overline{q}(\hat{\theta}_i)) - \hat{\theta}_i\overline{q}(\hat{\theta}_i) + \overline{T}_{-i}(\overline{q}(\hat{\theta}_i)) + \overline{t}_i(\overline{q}(\hat{\theta}_i))$$

and

$$\overline{U}_{-i}(\hat{\theta}_i) = S_0(\overline{q}_{-i}(\hat{\theta}_i)) - \hat{\theta}_i\overline{q}_{-i}(\hat{\theta}_i) + \overline{T}_{-i}(\overline{q}_{-i}(\hat{\theta}_i)).$$

By hypothesis, $\overline{q}$ is continuous at $\hat{\theta}_i$ which is interior, and so $\overline{U}(\theta)$ is differentiable at $\hat{\theta}_i$ with, from Lemma 1, a derivative equal to $-\overline{q}(\hat{\theta}_i)$. Because $\overline{U}$ is smooth and minorized by the convex function $\overline{U}_{-i}$ at $\hat{\theta}_i$, it follows also that $\overline{U}_{-i}(\theta)$ is itself differentiable at $\hat{\theta}_i$ and (4.4) holds. *Q.E.D.*

PROOF OF PROPOSITION 3: We begin by proving that if $\overline{q}$ is separating, then each $\overline{t}_i$ is differentiable on $\text{int}(\overline{\mathcal{Q}})$. To this end, we again refer to the value function and dual variable defined in (A6) and (A8) and the optimality conditions (A11).

Because $\overline{q}$ is separating, there exists an interval $[\theta_0, \theta_1] \subseteq \Theta$ on which $\overline{q}$ is strictly decreasing. For all θ in this interval $\overline{q}(\theta) = \partial V_i(\overline{\sigma}_i(\theta))$. Using Lemma 6, this implies that V_i^* is smooth for all q in the interior of $\overline{\mathcal{Q}}$. Because $W_i \equiv S_0 + S_i + \overline{T}_{-i}$ coincides with $-V_i^*$ on $\overline{\mathcal{Q}}$, W_i must also be differentiable on the interior of this set. Because S_0 and S_i are differentiable by hypothesis, we conclude that $\overline{T}_{-i}$ is differentiable for $q \in \text{int}(\overline{\mathcal{Q}})$. Because this is true for every $i \in N$, it follows that $\overline{t}_i$ is differentiable on $\text{int}(\overline{\mathcal{Q}})$.

Condition (4.5) follows from differentiability and incentive compatibility. Conditions (4.6) and (4.7) follow from differentiability and Theorem 2. *Q.E.D.*

PROOF OF PROPOSITION 4: First, $t(\overline{q}(\hat{\theta}_i)) = 0$ immediately follows from the first condition in (4.4).

Second, from the Proof of Proposition 3, we know that $\overline{t}_i$ is differentiable on $\text{int}(\overline{Q})$ under the posited assumptions. At $\hat{\theta}_i$, we already know from the proof of Proposition 2 that $\overline{q}(\hat{\theta}_i) = \overline{q}_{-i}(\hat{\theta}_i)$. The output $\overline{q}(\hat{\theta}_i) = \overline{q}_{-i}(\hat{\theta}_i) \in \text{int}(\overline{Q})$ therefore satisfies the two first-order conditions:

$$S_0'(\overline{q}(\hat{\theta}_i)) + \overline{T}'(\overline{q}(\hat{\theta}_i)) = S_0'(\overline{q}(\hat{\theta}_i)) + \overline{T}'_{-i}(\overline{q}(\hat{\theta}_i)) = \theta.$$

From this, it immediately follows that $\overline{t}_i'(\overline{q}(\hat{\theta}_i)) = 0$. *Q.E.D.*

The proof of Lemma 2 uses Lemma 10 below.

LEMMA 10 *For $i \in \mathcal{A}$, the function $\theta \mapsto F(\theta) - s_i f(\theta)$ is strictly increasing for $\theta \in (\hat{\theta}_i, \overline{\theta}]$, and $F(\theta) - s_i f(\theta) > 0$ if and only if $\theta \in (\hat{\theta}_i, \overline{\theta}]$. For $i \in \mathcal{B}$, the function $\theta \mapsto F(\theta) - s_i f(\theta)$ is strictly increasing for $\theta \in [\underline{\theta}, \hat{\theta}_i)$, and $F(\theta) - s_i f(\theta) > 1$ if and only if $\theta \in (\hat{\theta}_i, \overline{\theta}]$.*

PROOF OF LEMMA 10: Suppose $i \in \mathcal{A}$. For $\theta = \underline{\theta}$, $F(\theta) - s_i f(\theta) < 0$. If $\hat{\theta}_i = \overline{\theta}$, then $F(\theta) - s_i f(\theta) \leq 0$ over all Θ. If $\hat{\theta}_i \in (\underline{\theta}, \overline{\theta})$, then $F(\hat{\theta}_i) - s_i f(\hat{\theta}_i) = 0$. Because F/f is increasing, $F - s_i f > 0$ if and only if $\theta > \hat{\theta}_i$. The function $F(\theta) - s_i f(\theta)$

is increasing if $\frac{1}{s_i} > \frac{f'(\theta)}{f(\theta)}$. The generalized monotone-hazard rate condition requires $\frac{f'(\theta)}{f(\theta)} < \frac{f(\theta)}{F(\theta)}$, and so $F - s_i f$ is increasing if $\frac{f(\theta)}{F(\theta)} < \frac{1}{s_i}$, which is true for all $\theta > \hat{\theta}_i$.

Take $i \in \mathcal{B}$. For $\hat{\theta}_i = \underline{\theta}$, the function $F(\theta) - s_i f(\theta) > 1$ for all $\theta \in \Theta$. Suppose instead that $\hat{\theta}_i \in (\underline{\theta}, \overline{\theta})$. Then $F(\hat{\theta}_i) - s_i f(\hat{\theta}_i) = 1$ and the generalized monotone hazard-rate condition implies $F - s_i f > 1$ if and only if $\theta > \hat{\theta}_i$. The function $F - s_i f$ is increasing if $\frac{f'(\theta)}{f(\theta)} > \frac{1}{s_i}$. The generalized monotone-hazard rate condition also requires $\frac{f'(\theta)}{f(\theta)} > \frac{f(\theta)}{F(\theta)-1}$, and so $F - s_i f$ is increasing if $\frac{F(\theta)-1}{f(\theta)} < s_i$, which is true for all $\theta < \hat{\theta}_i$. *Q.E.D.*

PROOF OF LEMMA 2: Take now $i \in \mathcal{A}$. Lemma 10 established that $F(\theta) - s_i f(\theta) \leq 0$ for all $\theta \in [\underline{\theta}, \hat{\theta}_i]$, and so for $\theta \leq \hat{\theta}_i$ the fact that $\overline{M}_i(\theta) \geq 0$ implies $s_i + \frac{\overline{M}_i(\theta) - F(\theta)}{f(\theta)} \geq 0$ with all these inequalities being strict on $[\underline{\theta}, \hat{\theta}_i)$. The definitions of $\overline{q}_{-i}$ and $\overline{q}(\theta)$ (from Theorem 1) require

$$\text{(A13)} \quad \overline{q}_{-i}(\theta) \in \arg\max_{q \in \mathcal{Q}} S_0(q) + \overline{T}_{-i}(q) - \theta q$$

$$\text{(A14)} \quad \overline{q}(\theta) \in \arg\max_{q \in \mathcal{Q}} S_0(q) + \overline{T}_{-i}(q) - \theta q + \left(s_i + \frac{\overline{M}_i(\theta) - F(\theta)}{f(\theta)} \right) q$$

Taken together (A13) and (A14) yields $\overline{q}(\theta) \geq \overline{q}_{-i}(\theta)$ for all $\theta \leq \hat{\theta}_i$.

Taking the concave hull of the maximands in (A13) and (A14),[28] we get respectively

$$\text{(A15)} \quad \overline{q}_{-i}(\theta) \in \arg\max_{q \in \mathbb{R}} \overline{\text{co}} \left(S_0(q) + \overline{T}_{-i}(q) - \theta q + \Psi_{\mathcal{Q}}(q) \right),$$

$$\text{(A16)} \quad \overline{q}(\theta) \in \arg\max_{q \in \mathbb{R}} \overline{\text{co}} \left(S_0(q) + \overline{T}_{-i}(q) - \theta q + \Psi_{\mathcal{Q}}(q) + \left(s_i + \frac{\overline{M}_i(\theta) - F(\theta)}{f(\theta)} \right) q \right)$$

where $\Psi_{\mathcal{Q}}(q) = \begin{cases} 0 & \text{if } q \in \mathcal{Q} \\ +\infty & \text{otherwise} \end{cases}$. Because of the linearity in q of the second-bracketed term, we get

$$\overline{\text{co}} \left(S_0(q) + \overline{T}_{-i}(q) - \theta q + \Psi_{\mathcal{Q}}(q) + \left(s_i + \frac{\overline{M}_i(\theta) - F(\theta)}{f(\theta)} \right) q \right) = \overline{\text{co}} \left(S_0(q) + \overline{T}_{-i}(q) - \theta q + \Psi_{\mathcal{Q}}(q) \right) + \left(s_i + \frac{\overline{M}_i(\theta) - F(\theta)}{f(\theta)} \right) q.$$

Hence, (A16) becomes

$$\text{(A17)} \quad \overline{q}(\theta) \in \arg\max_{q \in \mathbb{R}} \overline{\text{co}} \left(S_0(q) + \overline{T}_{-i}(q) - \theta q + \Psi_{\mathcal{Q}}(q) \right) + \left(s_i + \frac{\overline{M}_i(\theta) - F(\theta)}{f(\theta)} \right) q.$$

[28] Let f be a function. $\overline{\text{co}} f$ is the lowest concave function that majorizes f.

Taking the subdifferential forms of (A15) and (A17), we get respectively

$$0 \in \partial_* \overline{\text{co}}(S_0(q) + \overline{T}_{-i}(q) - \theta q + \Psi_{\mathcal{Q}}(q))_{\overline{q}_{-i}(\theta)}, \tag{A18}$$

$$0 \in \partial_* \left(\overline{\text{co}} \left(S_0(q) + \overline{T}_{-i}(q) - \theta q + \Psi_{\mathcal{Q}}(q) \right) + \left(s_i + \frac{\overline{M}_i(\theta) - F(\theta)}{f(\theta)} \right) q \right)_{\overline{q}(\theta)}$$

$$= \partial_* \overline{\text{co}} \left(S_0(q) + \overline{T}_{-i}(q) - \theta q \right)_{\overline{q}(\theta)} + s_i + \frac{\overline{M}_i(\theta) - F(\theta)}{f(\theta)} \tag{A19}$$

where the last equality follows from the "Sum Rule" of subdifferential calculus and the fact that $\left(s_i + \frac{\overline{M}_i(\theta) - F(\theta)}{f(\theta)} \right) q$ is differentiable in q. Taken together (A18) and (A19) show that $\overline{q}_{-i}(\theta) \neq \overline{q}(\theta)$ for all $\theta < \hat{\theta}_i$ and thus $\overline{q}(\theta) > \overline{q}_{-i}(\theta)$ on $[\underline{\theta}, \hat{\theta}_i)$. It follows that principal i is active on that interval and $\overline{M}_i(\theta) = 0$ on $[\underline{\theta}, \hat{\theta}_i)$.

Because $\overline{M}_i$ is left-continuous at $\hat{\theta}_i$, $\overline{M}_i(\hat{\theta}_i) = 0$ and it follows immediately that $\overline{q}(\hat{\theta}_i) = \overline{q}_{-i}(\hat{\theta}_i)$.

Lastly, consider the region $(\hat{\theta}_i, \overline{\theta}]$. The pair $(\overline{q}, \overline{M})$ such that $\overline{q}(\theta) = \overline{q}_{-i}(\theta)$ and $\overline{M}_i(\theta) = F(\theta) - s_i f(\theta)$ satisfies the necessary conditions of Theorem 1.

We conclude that the construction $\overline{M}_i(\theta) = \max\{F(\theta) - s_i f(\theta), 0\}$ for $i \in \mathcal{A}$ supports $\overline{q}$ over Θ as required. A similar argument also establishes that, for $i \in \mathcal{B}$, the construction $\overline{M}_i(\theta) = \min\{F(\theta) - s_i f(\theta), 1\}$ supports $\overline{q}$ over Θ. *Q.E.D.*

PROOF OF THEOREM 3:

NECESSITY. Proposition 2 in tandem with Theorem 2, implies that, for almost every θ, the allocation satisfies

$$\overline{q}(\theta) \in \arg\max_{q \in \mathcal{Q}} \; S_0(q) + (\beta(\theta) - \theta)q + (n-1)(S_0(q) - \theta q + \overline{T}(q)), \tag{A20}$$

where $\overline{T}$ implements $(\overline{q}, \overline{U})$.

Denote

$$\mathcal{V}(\theta) = \max_{q \in \mathcal{Q}} \; S_0(q) + (\beta(\theta) - \theta)q + (n-1)(S_0(q) - \theta q + \overline{T}(q)).$$

By definition, we have

$$\mathcal{V}(\theta) = S_0(\overline{q}(\theta)) + (\beta(\theta) - \theta)\overline{q}(\theta) + (n-1)\bar{U}(\theta).$$

Moreover, $\mathcal{V}$ is absolutely continuous as a value function. It is thus almost everywhere differentiable and for any pair (θ, θ'), we get:

$$\mathcal{V}(\theta) - \mathcal{V}(\theta') = \int_{\theta'}^{\theta} (\beta'(x) - n)q(x)dx.$$

But $\bar{U}$ is itself absolutely continuous and such that

$$\bar{U}(\theta) - \bar{U}(\theta') = -\int_{\theta'}^{\theta} q(x)dx.$$

It follows that $S_0(\overline{q}(\theta)) + (\beta(\theta) - \theta)\overline{q}(\theta)$ is itself absolutely continuous with:

$$S_0(\overline{q}(\theta)) + (\beta(\theta) - \theta)\overline{q}(\theta) - [S_0(\overline{q}(\theta')) + (\beta(\theta') - \theta')\overline{q}(\theta')] \\ = \mathcal{V}(\theta) - \mathcal{V}(\theta') - (n-1)\left[\bar{U}(\theta) - \bar{U}(\theta')\right]$$

or more simply

$$\text{(A21)} \quad S_0(\overline{q}(\theta)) + (\beta(\theta) - \theta)\overline{q}(\theta) - [S_0(\overline{q}(\theta')) + (\beta(\theta') - \theta')\overline{q}(\theta')] = \int_{\theta'}^{\theta} (\beta'(x) - 1)q(x)dx.$$

We may conclude

$$S_0(\overline{q}(\theta)) + (\beta(\theta) - \theta)\overline{q}(\theta) - [S_0(\overline{q}(\theta')) + (\beta(\theta) - \theta)\overline{q}(\theta')] = \int_{\theta'}^{\theta} (\beta'(x) - 1)(q(x) - q(\theta'))dx \geq 0$$

where the last inequality follows from the fact that β and $\overline{q}$ are both weakly non-increasing. Hence, condition (5.1) holds.

Sufficiency. The proof of sufficiency proceeds by construction. The allocation $\overline{q}$ satisfies (5.2) if and only if

$$\overline{q}(\theta) \in \arg\max_{q \in \mathbb{R}} (\beta(\theta) - \theta)q + S_0(q) + \Psi_{\mathcal{Q}}(q).$$

Because the program is concave, the necessary condition is also sufficient for characterizing such a solution. Consider the agent's response to the virtually truthful schedules which are defined in Equations (5.6) and (5.7) in the text. The agent solves

$$\max_{q \in \mathcal{Q}} S_0(q) - \theta q + \overline{T}(q) = \max_{q \mathbb{R}} S_0(q) - \theta q + \overline{T}(q) + \Psi_{\mathcal{Q}}(q).$$

Because $\overline{T}$ is differentiable and S_0 is concave, the local necessary and sufficient condition for an optimal choice is

$$\theta - \overline{T}'(\overline{q}(\theta)) \in \partial_* \left(S_0(\overline{q}(\theta) + \Psi_{\mathcal{Q}}(\overline{q}(\theta))\right).$$

Because $\overline{q}$ is non-decreasing, this local condition is also sufficient for implementability. By construction, $\beta(\theta) = \overline{T}'(\overline{q}(\theta))$, and so we have the condition

$$\theta - \beta(\theta) \in \partial_* \left(S_0(\overline{q}(\theta)) + \Psi_{\mathcal{Q}}(\overline{q}(\theta))\right).$$

But this is the condition for a solution to (5.2), so we conclude that $\overline{T}$ implements $\overline{q}$.

Consider principal i's program in (3.2) where $i \in \mathcal{A}$ and $\overline{M}_i$ is as in Proposition 2:

$$\max_{q \in \mathbb{R}} \left(S_0(q) + \overline{T}(q) + \Psi_{\mathcal{Q}}(q)\right) - \bar{t}_i(q) - \theta q + \max\left\{s_i - \frac{F(\theta)}{f(\theta)}, 0\right\} q.$$

We replace the first parenthetical expression with its concave hull $\overline{\text{co}}\left(S_0(q) + \overline{T}(q) + \Psi_{\mathcal{Q}}(q)\right)$. Note that because $\overline{T}$ implements $\overline{q}$, it follows that for all $q \in \mathcal{Q}$

$$\overline{\text{co}}\left(S_0(q) + \overline{T}(q) + \Psi_{\mathcal{Q}}(q)\right) \geq S_0(q) + \overline{T}(q) + \Psi_{\mathcal{Q}}(q),$$

with equality for any $q \in \overline{\mathcal{Q}}$. This substitution increases the desirability of selecting $q \notin \overline{\mathcal{Q}}$ without changing the value of the maximand for $q \in \overline{\mathcal{Q}}$. Hence, if $\overline{q}$ is a solution to this modified program, it is a solution to the original program. Thus, using $\beta_i(\theta)$ from above, we consider the following concave problem

$$\max_{q \in \mathbb{R}} \overline{\text{co}}\left(S_0(q) + \overline{T}(q) + \Psi_{\mathcal{Q}}(q)\right) - \bar{t}_i(q) + (\beta_i(\theta) - \theta)q.$$

The necessary and sufficient condition for optimality is thus

$$\theta - \beta_i(\theta) + \bar{t}_i'(\overline{q}(\theta)) \in \partial_* \overline{\text{co}}\left(S_0(\overline{q}(\theta)) + \overline{T}(\overline{q}(\theta)) + \Psi_{\mathcal{Q}}(\overline{q}(\theta))\right).$$

By construction, $\beta_i(\theta) = \bar{t}_i'(\overline{q}(\theta))$, and so we have the simpler requirement that

$$\theta \in \partial_* \overline{\text{co}}\left(S_0(q) + \overline{T}(q) + \Psi_{\mathcal{Q}}(q)\right),$$

which holds because $\overline{q}$ is implemented by $\overline{T}$. Thus, the allocation $\overline{q}$ satisfies the optimality condition (3.2) in Theorem 1, using the adjoint function $\overline{M}_i$ from Proposition 2. By construction, principal i's transfer function is positive if and only if $\overline{M}_i(\theta)$ is constant, and so the complementary slackness condition (3.1) is also satisfied. We conclude that it is optimal for principal i to implement the allocation $\overline{q}$ given the contract aggregate $\overline{T}_{-i}$. The transfer which accomplishes this is exactly the constructed $\bar{t}_i$ above.

A similar argument applies to any principal $i \in \mathcal{B}$.

Existence. Given that S_0 is upper semi-continuous and $\mathcal{Q}$ is compact, a solution to (5.2) always exists. *Q.E.D.*

Proof of Proposition 5:

Step 1: From the proof of Theorem 3, we know that $S_0(\overline{q}(\theta)) + (\beta(\theta) - \theta)\overline{q}(\theta)$ is itself absolutely continuous and thus a.e. differentiable. Using (A21) and differentiating yields (5.3).

Step 2: Let θ be a point of discontinuity of $\overline{q}$. We know that $\overline{q}(\theta^-) > \overline{q}(\theta^+)$ by incentive compatibility. To sustain an equilibrium allocation with such discontinuity, the agent with type θ must be indifferent between choosing either $\overline{q}(\theta^-)$ or $\overline{q}(\theta^+)$, i.e.,

$$\text{(A22)} \quad \left[S_0(q) + \overline{T}(q) - \theta q\right]_{\overline{q}(\theta^+)}^{\overline{q}(\theta^-)} = 0.$$

Using the optimality conditions (A20) at such discontinuity point, it must also be that:

$$\text{(A23)} \quad \left[(\beta(\theta) - \theta)q + S_0(q) + (n-1)(S_0(q) + \overline{T}(q) - \theta q)\right]_{\overline{q}(\theta^+)}^{\overline{q}(\theta^-)} = 0.$$

Inserting (A22) into (A23) yields (5.4). Because $\overline{q}$ is a.e. differentiable, such discontinuity point is isolated and $\overline{q}$ is continuous and differentiable both on a right- and a left-neighborhood of θ. Using Item [1.] in those intervals, we deduce that $\dot{\overline{q}}(\theta) = 0$ on such neighborhood. *Q.E.D.*

PROOF OF PROPOSITION 6: Because the distribution of θ converges in probability to μ as $\kappa \to \infty$, we have for any $\varepsilon > 0$ and for every $\kappa \geq K(\varepsilon)$, $F(\mu+\varepsilon,\kappa) - F(\mu-\varepsilon,\kappa) \geq 1-\varepsilon$. Because $s_i \geq \overline{\tau}_i(\theta,\kappa)$,

$$\int_{\Theta} (s_i - \overline{\tau}_i(\theta,\kappa))\, f(\theta,\kappa) d\theta \geq 0.$$

We want to show that this integral converges to zero. To this end, note

$$\begin{aligned}
\int_{\Theta} (s_i - \overline{\tau}_i(\theta,\kappa))\, f(\theta,\kappa) d\theta &= \int_{\underline{\theta}}^{\overline{\theta}} \min\left\{ \frac{F(\theta,\kappa)}{f(\theta,\kappa)}, s_i \right\} f(\theta,\kappa) d\theta \\
&\leq \int_{\underline{\theta}}^{\mu-\varepsilon} F(\theta,\kappa) d\theta + \int_{\mu-\varepsilon}^{\mu+\varepsilon} F(\theta,\kappa) d\theta + \int_{\mu+\varepsilon}^{\overline{\theta}} s_i f(\theta,\kappa) d\theta \\
&\leq (\mu - \varepsilon - \underline{\theta})\varepsilon + 2\varepsilon + s_i\varepsilon.
\end{aligned}$$

Because ε can be made arbitrarily small as $\kappa \to \infty$, we may conclude that (5.8) holds. *Q.E.D.*

PROOF OF PROPOSITION 7: Theorem 3 requires that (5.2) holds for a separating equilibrium. If S_0 is linear, then the objective in the maximand of (5.2) is linear in q. The maximum of a linear functional on a compact set (in this case, an interval $\mathcal{Q} = [q_{\min}, q_{\max}]$) is an extreme point if the linear function is not identically zero. The construction of $\hat{\theta}$ is such that for $\theta < \hat{\theta}$, the maximand is a strictly increasing linear function and for $\theta > \hat{\theta}$, it is a strictly decreasing linear function. Equation (5.9) follows. When $\hat{\theta}$ is interior, it satisfies

$$\beta(\hat{\theta}) = \hat{\theta}.$$

If S_0 is strictly concave, then the maximand of (5.2) is also strictly concave. It follows that there is a unique $\overline{q}(\theta)$ pointwise in θ and $\overline{q}$ is a continuous function. If S_0 is strictly concave and differentiable, $\overline{q}$ must be strictly decreasing in the interior of Θ by Lemma 6, and hence it is separating. *Q.E.D.*

PROOF OF PROPOSITION 8: First, we denote by $\overline{t}_i^d(q)$ principal i's equilibrium schedule in a discontinuous equilibrium and by $\overline{T}^d(q)$ the aggregate. Similar notations are used for outputs. The proof is in several steps.

Step 1: For a discontinuous equilibrium, $\overline{Q}$ is a non-connected set. To simplify the analysis, we construct an equilibrium with a unique discontinuity so that $\overline{Q}$ is made of two disconnected intervals. Generalizations to more points of discontinuity are immediate. Let also this discontinuity be at θ_0 and denote $\overline{q}^d(\theta_0^+)$ and $\overline{q}^d(\theta_0^-)$ the right- and left-hand side limits of $\overline{q}^d$ at θ_0. Incentive compatibility implies that $\overline{q}^d(\theta_0^+) < \overline{q}^d(\theta_0^-)$.[29]

On each connected subset $\overline{Q}_2 = [\overline{q}(\bar{\theta}), \overline{q}^d(\theta_0^+)]$ and $\overline{Q}_1 = [\overline{q}^d(\theta_0^-), \overline{q}(\underline{\theta})]$, Proposition 7 holds. Several facts immediately follow.

[29] The choice of having either $\overline{q}(\theta_0) = \overline{q}^d(\theta_0^+)$ or and $\overline{q}(\theta_0) = \overline{q}^d(\theta_0^-)$ is arbitrary.

Fact 1. $\overline{q}:\Theta\to\mathcal{Q}$ is not only continuous but also separating on the interiors $\text{int}(\overline{Q}_j)$ $(j=1,2)$. Hence, we know that $\vartheta(q)$ is unambiguously defined on $\text{int}(\overline{Q}_j)$. Moreover, for any θ such that $\overline{q}(\theta)\in\text{int}(\overline{\mathcal{Q}}_j)$, $\overline{q}$ solves

$$\text{(A24)}\quad \overline{q}(\theta)\in\arg\max_{q\in\overline{\mathcal{Q}}_j} S_0(q)+(\beta(\theta)-\theta)q.$$

For any $\overline{q}(\theta)\in\text{int}(\overline{\mathcal{Q}}_j)$, strict concavity of the above maximand yields

$$\text{(A25)}\quad S_0'(\overline{q}(\theta))+\beta(\theta)-\theta=0$$

which also implicitly defines $\vartheta(q)$.

Fact 2. From *Fact 1*, it follows that the equilibrium tariffs have the following margins:

$$\text{(A26)}\quad \overline{t}_i^{d'}(q)=\beta_i(\vartheta(q))\quad\text{for any } q\in\text{int}(\overline{Q}_j).$$

Fact 3. By construction $\overline{q}^d(\theta_0^-)\in\overline{Q}_1$ and $\overline{q}^d(\theta_0^+)\in\overline{Q}_2$. There exists $\theta_1<\theta_0$ such that $\theta_1=\vartheta(\overline{q}^d(\theta_0^-))$ and $\theta_2>\theta_0$ such that $\theta_2=\vartheta(\overline{q}^d(\theta_0^+))$.

Step 2: At such discontinuity θ_0, (A23) holds and the "aggregate objective" specified in (A20) is constant across this discontinuity.

Step 3: Let us now take a continuous equilibrium say $\overline{q}$ which is virtually-truthful. We modify this equilibrium to construct a discontinuous equilibrium $\overline{q}^d$ having a discontinuity at a given θ_0. We build the equilibrium tariffs to obtain that principal i's virtual surplus defined as

$$\tilde{W}_i(\theta,q)=(\beta_i(\theta)-\theta)q+S_0(q)+\overline{t}_{-i}^d(q)$$

remains *linear* over the interval $[\overline{q}^d(\theta_0^+),\overline{q}^d(\theta_0^-)]$.[30] This is this expression which is maximized pointwise at a best response as shown in Theorem 1 once one takes the expressions of adjoint functions coming from Lemma 2.

Equilibrium tariffs. Let principal $-i$'s schedule over $[\overline{q}^d(\theta_0^+),\overline{q}^d(\theta_0^-)]$ be

$$\text{(A27)}\quad \overline{t}_{-i}^d(q)=\overline{t}_{-i}^d(\overline{q}^d(\theta_0^-))-[(\beta_i(\theta_0)-\theta_0)q+S_0(q)]_{\overline{q}^d(\theta_0^-)}^q\quad\forall q\in[\overline{q}^d(\theta_0^-),\overline{q}^d(\theta_0^+)].$$

On $\overline{Q}_1$ and $\overline{Q}_2$ respectively, the tariff is respectively defined as

$$\text{(A28)}\quad \overline{t}_{-i}^d(q)=\overline{t}_{-i}^d(\overline{q}^d(\theta_0^+))+\int_{\overline{q}^d(\theta_0^+)}^q\beta_i(\vartheta(z))dz$$

and

$$\text{(A29)}\quad \overline{t}_{-i}^d(q)=\overline{t}_{-i}^d(\overline{q}^d(\theta_0^-))+\int_{\overline{q}^d(\theta_0^-)}^q\beta_i(\vartheta(z))dz.$$

[30] The construction is easier when $n=2$ but could be generalized for $n>2$ at the cost of some notational burden.

where $\overline{t}^d_{-i}(\overline{q}^d(\theta_0^+))$ is defined from (A27). Note that $\overline{t}^d_{-i}(q)$ so constructed is continuous not only on $[\overline{q}^d(\theta_0^+), \overline{q}^d(\theta_0^-)]$ but everywhere on $\overline{Q}$. Also $\overline{t}^d_{-i}(q)$ is differentiable on $(\overline{q}^d(\theta_0^+), \overline{q}^d(\theta_0^-))$ and on the interior of $\mathcal{Q}_1$ and $\mathcal{Q}_2$. Moreover, we have

$$\lim_{q \to \overline{q}^d(\theta_0^-)^-} \overline{t}^{d'}_{-i}(q) = -(\beta_i(\theta_0) - \theta_0) - S_0'(\overline{q}^d(\theta_0^-)) = \beta_{-i}(\theta_1) + \beta_i(\theta_1) - \theta_1 - (\beta_i(\theta_0) - \theta_0)$$

$$\geq \beta_{-i}(\theta_1) = \overline{t}^{d'}_{-i}(\overline{q}^d(\theta_0^-))$$

$$\lim_{q \to \overline{q}^d(\theta_0^+)^+} \overline{t}^{d'}_{-i}(q) = -(\beta_i(\theta_0) - \theta_0) - S_0'(\overline{q}^d(\theta_0^+)) = \beta_{-i}(\theta_2) + \beta_i(\theta_2) - \theta_2 - (\beta_i(\theta_0) - \theta_0)$$

$$\leq \beta_{-i}(\theta_2) = \overline{t}^{d'}_{-i}(\overline{q}^d(\theta_0^+))$$

where the two above inequalities follow from the fact that $\beta_i(\cdot)$ is weakly decreasing.

Principal i's best response. We are now ready to compute for any $\theta \in \Theta$,

$$\overline{q}^d(\theta) \in \arg\max_{q \in \mathcal{Q}} \tilde{W}_i(\theta, q).$$

• Consider $\theta \in [\underline{\theta}, \theta_1]$ and observe then that

$$\overline{q}(\theta) = \arg\max_{q \in \overline{Q}_1} \tilde{W}_i(\theta, q). \tag{A30}$$

Moreover, for $q \in \overline{Q}_2$, we have

$$\frac{\partial \tilde{W}_i}{\partial q}(\theta, q) = \beta_i(\theta) - \theta + S_0'(q) + \beta_{-i}(\vartheta(q)). \tag{A31}$$

Because the generalized monotone hazard rate property holds, we have $\beta_i(\theta) - \theta \geq \beta_i(\vartheta(q)) - \vartheta(q)$ for $\theta \leq \vartheta(q)$ and $q \leq \overline{q}(\theta_0^+)$. Inserting into (A31) yields

$$\frac{\partial \tilde{W}_i}{\partial q}(\theta_0, q) \geq S_0'(q) + \beta_1(\vartheta(q)) + \beta_2(\vartheta(q)) - \vartheta(q) = 0 \text{ for } q \leq \overline{q}^d(\theta_0^-).$$

Hence, we get

$$\overline{q}(\theta_0^+) = \arg\max_{q \in \overline{Q}_2} \tilde{W}_i(\theta, q). \tag{A32}$$

Finally, consider $q \in [\overline{q}^d(\theta_0^+), \overline{q}^d(\theta_0^-)]$ and observe that

$$\frac{\partial \tilde{W}_i}{\partial q}(\theta, q) = \beta_i(\theta) - \theta - (\beta_i(\theta_0) - \theta_0) \geq 0$$

where the last inequality follows from $\beta_i(\cdot)$ being weakly decreasing and $\theta \leq \theta_1 < \theta_0$. Hence, we get:

$$\overline{q}(\theta_0^-) = \arg\max_{q \in \overline{Q}_2} \tilde{W}_i(\theta, q). \tag{A33}$$

Putting together (A30), (A32) and (A33), we finally obtain:

$$\overline{q}^d(\theta) = \overline{q}(\theta). \tag{A34}$$

A similar argument shows again that, for any $\theta \in [\theta_2, \overline{\theta}]$, (A34) holds.
• Consider $\theta \in [\theta_1, \theta_0)$. The same arguments as above show that $\tilde{W}_i(\theta, q)$ is increasing over $\overline{Q}_2 \bigcup [\overline{q}^d(\theta_0^+), \overline{q}^d(\theta_0^-)]$. Moreover, for $q \in \overline{Q}_1$ (i.e., for $q \geq \overline{q}^d(\theta_0^-)$), we have

$$\frac{\partial \tilde{W}_i}{\partial q}(\theta, q) = \beta_i(\theta) - \theta + S_0'(q) + \beta_{-i}(\vartheta(q)). \tag{A35}$$

Because the generalized monotone hazard rate property holds, we have $\beta_i(\theta) - \theta \leq \beta_i(\vartheta(q)) - \vartheta(q)$ for $\theta \geq \theta_1 \geq \vartheta(q)$. Inserting into (A35) yields

$$\frac{\partial \tilde{W}_i}{\partial q}(\theta, q) \leq S_0'(q) + \beta_1(\vartheta(q)) + \beta_2(\vartheta(q)) - \vartheta(q) = 0 \text{ for } q \geq \overline{q}^d(\theta_0^-) = \overline{q}(\theta_1).$$

Hence, we get:

$$\overline{q}(\theta_0^-) = \arg\max_{q \in \overline{Q}_2} \tilde{W}_i(\theta, q) = \arg\max_{q \in \overline{Q}} \tilde{W}_i(\theta, q). \tag{A36}$$

A similar argument shows that, for any $\theta \in (\theta_0, \theta_2]$, we have:

$$\overline{q}(\theta_0^+) = \arg\max_{q \in \overline{Q}} \tilde{W}_i(\theta, q). \tag{A37}$$

• Consider now θ_0. By construction, each principal is indifferent between inducing $\overline{q}^d(\theta_0^+)$ and $\overline{q}^d(\theta_0^-)$. Indeed, the construction of $\bar{t}_{-i}(q)$ in (A27) is such that principal i's virtual surplus is constant over the whole interval $[\overline{q}^d(\theta_0^+), \overline{q}^d(\theta_0^-)]$, $\tilde{W}_i(\theta_0, q) = \tilde{W}_i(\theta_0)$ for any $q \in [\overline{q}^d(\theta_0^+), \overline{q}^d(\theta_0^-)]$ where

$$\begin{aligned}\tilde{W}_i(\theta_0) &= (\beta_i(\theta_0) - \theta_0)\overline{q}^d(\theta_0^+) + S_0(\overline{q}^d(\theta_0^+)) + \bar{t}^d_{-i}(\overline{q}^d(\theta_0^+)) \\ &= (\beta_i(\theta_0) - \theta_0)\overline{q}^d(\theta_0^-) + S_0(\overline{q}^d(\theta_0^-)) + \bar{t}^d_{-i}(\overline{q}^d(\theta_0^-)).\end{aligned}$$

Consider now $q \geq \overline{q}^d(\theta_0^-)$, we have:

$$\frac{\partial \tilde{W}_i}{\partial q}(\theta_0, q) = \beta_i(\theta_0) - \theta_0 + S_0'(q) + \beta_{-i}(\vartheta(q)). \tag{A38}$$

Because, again, the generalized monotone hazard rate property holds, we have $\beta_i(\theta_0) - \theta_0 \leq \beta_i(\vartheta(q)) - \vartheta(q)$ for $\theta_0 \geq \theta_1 \geq \vartheta(q)$. Inserting into (A38) yields

$$\frac{\partial \tilde{W}_i}{\partial q}(\theta_0, q) \leq S_0'(q) + \beta_1(\vartheta(q)) + \beta_2(\vartheta(q)) - \vartheta(q) = 0 \text{ for } q \geq \overline{q}^d(\theta_0^-) = \overline{q}(\theta_1). \tag{A39}$$

A symmetric argument shows also that

$$\frac{\partial \tilde{W}_i}{\partial q}(\theta_0, q) \geq S_0'(q) + \beta_1(\vartheta(q)) + \beta_2(\vartheta(q)) - \vartheta(q) = 0 \text{ for } q \leq \overline{q}^d(\theta_0^+) = \overline{q}(\theta_2). \tag{A40}$$

Putting together (A39) and (A40) shows that

$$[\overline{q}^d(\theta_0^+), \overline{q}^d(\theta_0^-)] = \arg\max_{q\in\mathcal{Q}} \tilde{W}_i(\theta, q).$$

Agent's behavior. We are now ready to compute for any $\theta \in \Theta$,

$$\overline{q}^d(\theta) \in \arg\max_{q\in\mathcal{Q}} S_0(q) - \theta q + \overline{T}^d(q).$$

- The first obvious observation is that, for $\theta \in [\underline{\theta}, \theta_1] \bigcup [\theta_2, \overline{\theta}]$, we have $\overline{q}^d(\theta) = \overline{q}(\theta)$.
- For $\theta \in [\theta_1, \theta_0)$, we have $\overline{q}^d(\theta_0^-) = \arg\max_{q\in\mathcal{Q}} S_0(q) - \theta q + \overline{T}^d(q)$, whereas, for $\theta \in (\theta_0, \theta_2]$, we have instead $\overline{q}^d(\theta_0^+) = \arg\max_{q\in\mathcal{Q}} S_0(q) - \theta q + \overline{T}^d(q)$.
- Finally, for $\theta = \theta_0$, the definitions of $\bar{t}_i(q)$ given in (A27) imply that the agent is indifferent between $\overline{q}^d(\theta_0^-)$ and $\overline{q}^d(\theta_0^+)$. Indeed, we have

$$\left[S_0(q) - \theta_0 q + \overline{T}^d(q)\right]_{\overline{q}^d(\theta_0^+)}^{\overline{q}^d(\theta_0^-)} = [S_0(q) + (\beta(\theta_0) - \theta_0)q]_{\overline{q}^d(\theta_0^+)}^{\overline{q}^d(\theta_0^-)} = 0$$

where the last equality follows from (5.4).

Moreover, it is easy to check that any other output choice gives a lower payoff to the agent.

Non-negativeness of the tariffs. This is where the assumption that the discontinuity is small enough matters. Starting from a continuous equilibrium where both principals are active at θ_0, the modifications of the transfers in (A27) are still non-negative. It is enough for that that the discontinuity jump $\overline{q}^d(\theta_0^-) - \overline{q}^d(\theta_0^+)$ be kept small enough. *Q.E.D.*

PROOF OF PROPOSITION 9: Equation (6.1) describes the unique solution to (5.2). *Q.E.D.*

PROOF OF PROPOSITION 10: Equation (6.2) follows from (5.9). Consider the transfers $\bar{t}_i(q) = \max\left\{s_i - \frac{F(\hat{\theta})}{f(\hat{\theta})}, 0\right\} q$, $i \in N$. An agent will choose $q = 1$ only if $\sum_{i\in N} \bar{t}_i(1) \geq \theta$, or $\sum_{i\in N} \max\left\{s_i - \frac{F(\hat{\theta})}{f(\hat{\theta})}, 0\right\} \geq \theta$, and will choose $q = 0$ if the inequality fails. Given the definition of $\hat{\theta}$, $\overline{q}$ in (6.2) is implementable for these transfers.

Now consider principal i's program, taking the transfers of principals $j \neq i$ as given. We need to demonstrate

$$\overline{q}(\theta) \in \arg\max_{q\in[0,1]} \left(s_i + \left(\sum_{j\neq i} \max\left\{s_j - \frac{F(\hat{\theta})}{f(\hat{\theta})}, 0\right\} - \theta\right) + \frac{\overline{M}_i(\theta) - F(\theta)}{f(\theta)}\right) q.$$

Since $\overline{M}_i(\theta) = \max\{F(\theta) - f(\theta)s_j, 0\}$, we have $s_i + \frac{\overline{M}_i(\theta) - F(\theta)}{f(\theta)} = \max\left\{s_i - \frac{F(\theta)}{f(\theta)}, 0\right\}$ and the solution to the above program is (6.2). *Q.E.D.*

PROOF OF PROPOSITION 11: Equation (6.4) follows immediately as the unique solution to (5.2) for the case of $n = 2$. When θ is uniformly distributed, $\frac{F(\theta)}{f(\theta)} = \theta + \delta$, and $\frac{1-F(\theta)}{f(\theta)} = \theta - \delta$. Define then

$$\begin{aligned}
\overline{q}_1(\theta) &\equiv s_1 - \frac{F(\theta)}{f(\theta)} - \theta = s_1 - \delta - 2\theta, \\
\overline{q}_2(\theta) &\equiv s_2 + \frac{1-F(\theta)}{f(\theta)} - \theta = s_2 + \delta - 2\theta, \\
\overline{q}_{12}(\theta) &\equiv s_1 + s_2 - \frac{F(\theta)}{f(\theta)} + \frac{1-F(\theta)}{f(\theta)} - \theta = s_1 + s_2 - 3\theta.
\end{aligned}$$

Let θ_1 be the solution to $\overline{q}_1(\theta_1) = \overline{q}_{12}(\theta_1)$ if such a root exists and $\theta_1 = -\delta$ otherwise. Thus, $\theta_1 = \max\{-\delta, s_2 + \delta\}$. Similarly, define θ_2 by $\overline{q}_2(\theta_2) = \overline{q}_{12}(\theta_2)$ if such a root exists and $\theta_2 = \delta$ otherwise. Thus, $\theta_2 = \min\{s_1 - \delta, \delta\}$. It follows that principal 1 is active over the set $\overline{\Theta}_1 = [-\delta, \theta_2)$ and principal 2 is active over $\overline{\Theta}_2 = (\theta_1, \delta]$. These sets overlap if and only if $\theta_1 < \theta_2$, i.e., $\max\{-\delta, s_2 + \delta\} < \min\{s_1 - \delta, \delta\}$. This holds if and only if $s_1 - \delta > s_2 + \delta$ or (given that $s_2 < 0$) $\delta < \frac{s_1 + |s_2|}{2}$. *Q.E.D.*

PROOF OF PROPOSITION 12: The proof is in three steps. Note that when $\delta < \frac{1}{2}$, both principals are active in the virtually-truthful equilibrium where $\overline{q} = -3\theta$.

Symmetry of the discontinuity jump. First, observe that any discontinuity must preserve the aggregate virtual surplus of the principals at type $\theta_0 = 0$ where the discontinuity lies. In this lobbying example, this aggregate virtual surplus is:

$$\left(\max\{1 - \delta - \theta_0, 0\} + \min\{-1 + \delta - \theta_0, 0\} - \theta_0\right) q - \frac{q^2}{2} \equiv -\frac{q^2}{2}$$

which is clearly symmetric around the origin. Hence, the discontinuity jumps are of the form $\overline{q}^d(\theta_0^-) = -\overline{q}^d(\theta_0^+) = q_0$.

Discontinuity jump. Proposition 8 constructs such discontinuity on an interval where both principals are active which implies $q_0 \in (0, 3(1-\delta))$. We describe the condition for non-negative transfers below to reduce the relevant interval for q_0.

Non-negative payments. We now provide a condition on q_0 that ensures that the equilibrium transfers $\overline{t}_i^d(q)$ in the discontinuous equilibria remain positive on $[-q_0, q_0]$. To do so, we focus on the tariff $\overline{t}_2^d(q)$ offered by principal 2 and use symmetry to complete the construction. From the proof of Proposition 8 and especially (A27) we know that

(A41)

$$\overline{t}_2^d(q) = \overline{t}_2^d(q_0) - \left[\beta_1(0)q - \frac{q^2}{2}\right]_{q_0}^{q} = \overline{t}_2^d(q_0) + (1-\delta)q_0 - \frac{q_0^2}{2} - (1-\delta)q + \frac{q^2}{2} \quad \text{for all } q \in [-q_0, q_0]$$

since $\beta_1(0) = 1 - \delta$. On the other hand, principal 2 starts to be active at $\theta = -1 + \delta$ and:

$$\overline{t}_2^d(q_0) = \int_{3(1-\delta)}^{q_0} \beta_2(\vartheta(y))dy = \int_{3(1-\delta)}^{q_0} \left(-1 + \delta + \frac{y}{3}\right) dy = -(1-\delta)q_0 + \frac{q_0^2}{6} + \frac{3}{2}(1-\delta)^2.$$

Inserting this expression into (A41) yields

$$\text{(A42)}\quad \overline{t}_2^d(q) = \frac{q^2}{2} - (1-\delta)q - \frac{q_0^2}{3} + \frac{3}{2}(1-\delta)^2 \quad \text{for all } q \in [-q_0, q_0].$$

Observe that

$$\min\{q_0, 1-\delta\} = \arg\min_{q\in[-q_0,q_0]} \overline{t}_2^d(q).$$

Two cases must be studied. When $q_0 < 1-\delta$, we have $\min_{q\in[-q_0,q_0]} \overline{t}_2^d(q) = \overline{t}_2^d(q_0) > 0$. When $q_0 \geq 1-\delta (> -q_0)$, we have instead $\min_{q\in[-q_0,q_0]} \overline{t}_2^d(q) = \overline{t}_2^d(1-\delta) = (1-\delta)^2 - \frac{q_0^2}{3}$ which is non-negative when $q_0 \leq (1-\delta)\sqrt{3}$.

Welfare. Fix $q_0 \in [0, (1-\delta)\sqrt{3}]$ and observe that the expected welfare in the equilibrium with a discontinuity jump q_0 is

$$\mathcal{W}(q_0) = -\frac{1}{4\delta}\left(\int_{-\delta}^{-\frac{q_0}{3}} \overline{q}^2(\theta)d\theta + \int_{-\frac{q_0}{3}}^{\frac{q_0}{3}} q_0^2 d\theta + \int_{\frac{q_0}{3}}^{\delta} \overline{q}^2(\theta)d\theta\right).$$

Optimizing with respect to q_0 immediately yields $\mathcal{W}'(q_0) = -\frac{q_0^2}{3\delta}$ so that $\mathcal{W}(q_0)$ is maximized at $q_0 = 0$, i.e., for the continuous equilibrium. *Q.E.D.*

APPENDIX B: NECESSARY AND SUFFICIENT CONDITIONS FOR NON-SMOOTH OPTIMAL CONTROL WITH A LINEAR STATE VARIABLE

We are interested in the following pure-state control program $(\mathcal{P})$:

$$\begin{gathered}\text{Maximize } \Lambda(x) \equiv \textstyle\int_0^1 \left(S(\theta, u(\theta)) - x(\theta)f(\theta)\right) d\theta \\ \text{subject to } x \in AC(\Theta, \mathbb{R}),\ \dot{x}(\theta) = u(\theta),\ x(\theta) \geq 0 \text{ for all } \theta \in \Theta \equiv [0,1].\end{gathered}$$

The constraints require that the state variable x is a non-negative, absolutely continuous function, $x \in AC(\Theta, \mathbb{R})$. x is said *admissible* if it satisfies these constraints. Note that the integrand $L(\theta, x, u) = (S(\theta, u) - x)f(\theta)$ is linear in x and that the state constraint, $x \geq 0$, is independent of θ. These two restrictions within the class of state-constrained, non-smooth optimal control problems are the source of many sharp results in the analysis that follows.

We assume that $S(\theta, \cdot)$ is an upper-semi continuous function bounded from above and that $f(\theta)$ is a positive and bounded from above function so that $F(\theta) \equiv \int_{[\underline{\theta},\theta]} f(\theta)$ is absolutely continuous. Without loss of generality, we normalize f such that $F(1) = 1$ and interpret F as a continuous probability distribution. Lastly, we assume 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 $\mathbb{R}$. Importantly, we do not assume *a priori* that $S(\theta, \cdot)$ is a continuous function. We present our main result for this class of problems.

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

$$\overline{M}(\theta) \equiv \int_{[\underline{\theta},\theta)} \mu(ds),$$

such that the following two conditions are satisfied:

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

(B2) $\quad \dot{\overline{x}}(\theta) \in \arg\max_{v\in\mathbb{R}} \; S(\theta, v) + (F(\theta) - \overline{M}(\theta))v, \; \textit{for a.e. } \theta \in \Theta.$

Furthermore, if

$$y(\theta, \sigma) \equiv \arg\max_{v\in\mathbb{R}} \; S(\theta, v) + (F(\theta) - \sigma)v$$

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

Remarks:
• Theorem 1 is very similar to 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" (B1) and an optimality condition (B2). Moreover, both theorems use a similar condition to establish the continuity of $\dot{\overline{x}}(\theta)$ in the solution to (P). Jullien's Theorem, however, uses the stronger hypothesis that $S(\cdot)$ is twice continuously differentiable. Our technical contribution is to weaken these hypotheses to requirements of upper semi-continuity. This generalization allows us to apply the necessary and sufficient conditions above to our class of common agency games with upper-semi continuous contract menus.
• The condition that $y(\theta, \sigma)$ 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 $\sigma \in [0, 1]$.
• The adjoint function $\overline{M}(\theta)$. Note in particular that the function $\overline{M}$ is constructed to be *left*-continuous rather than right-continuous.

Proof of Theorem B.1:

Overview. We prove necessity by specializing Theorem 3 from Vinter and Zheng (1998), exploiting fact that our integrand in Λ is a linear function of x and that the state constraint $x(\theta) \geq 0$ is linear and independent of θ. 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. The regularity of the optimal solution follows from arguments involving the necessary conditions. While the proof seems straightforward when viewed from this broad vantage point, a considerable investment in concepts and notation from non-smooth, non-convex analysis is required along the way.

Preliminaries for Non-Smooth Analysis. We first introduce some additional notation. We draw heavily from Vinter and Zheng (1998) in the following presentation.[31]

[31] 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 $A \subseteq \mathbb{R}^k$ and a point $x \in A$. A vector $p \in \mathbb{R}^k$ is a *limiting normal* to A at x if there exists a sequence $(x_i, p_i) \to (x, p)$ and a $K \geq 0$ such that for each i in the sequence $p_i \cdot |x_i - x| \leq K|x_i - x|^2$. The cone of limiting normal vectors to A at x is denoted $N_A(x)$. Given a lower semi-continuous function $g : \mathbb{R}^k \to \mathbb{R} \cup \{+\infty\}$ and a point $x \in \mathbb{R}^k$ such that $g(x) < +\infty$, the *limiting subdifferential* 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

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

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

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

Two results from nonsmooth analysis (e.g., 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,

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

We denote the Euclidean norm in $\mathbb{R}^k$ by $|\cdot|$, and denote the norm on the space of absolutely continuous functions by

$$||x|| \equiv |x(\underline{\theta})| + \int_\Theta |\dot{x}(\theta)| d\theta.$$

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 $\overline{x}$,

$$||\overline{x} - x|| \leq \varepsilon.$$

A *local minimizer* is defined analogously.

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

$$(\mathcal{P}'): \quad \text{Minimize } J(x) \equiv \int_{\underline{\theta}}^{\overline{\theta}} L(\theta, x(\theta), \dot{x}(\theta)) d\theta$$
$$\text{subject to } x \in AC(\Theta, \mathbb{R}) \text{ and } h(\theta, x(\theta)) \leq 0 \text{ for all } \theta \in \Theta \equiv [\underline{\theta}, \overline{\theta}].^{32}$$

Theorem B.2 *(Vinter and Zheng (1998), Theorem 3) Let $\overline{x}$ be a AC local minimizer for $(\mathcal{P}')$ such that $J(\overline{x}) < +\infty$. Assume that the following hypotheses are satisfied:*

H1. $L(\cdot, x, \cdot)$ is $\mathcal{L} \times \mathcal{B}$ measureable for each x and $L(\theta, \cdot, \cdot)$ is lower semi-continuous for a.e. $\theta \in \Theta$.

[32] 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.

H2. *For every $K > 0$ 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 $v \in \dot{\overline{x}}(\theta) + KB$, where B is a unit Euclidean ball.

H3. *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 $p \in AC$, a constant $\lambda \geq 0$, a non-negative measure μ on the Borel subsets of Θ and a μ-integrable function $\gamma : \Theta \to \mathbb{R}$, such that

(i). $\lambda + \max_{\theta \in \Theta} |p(\theta)| + \int_\Theta \mu(ds) = K > 0$ *(where K is an arbitrary normalization constant),*[33]

(ii).

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

(iii).

$$p(\underline{\theta}) = p(\overline{\theta}) - \int_\Theta \gamma(s)\mu(ds) = 0,$$

(iv).

$$\left(p(\theta) + \int_{[\underline{\theta},\theta)} \gamma(s)\mu(ds)\right) \cdot \dot{\overline{x}}(\theta) - \lambda L(\theta, \overline{x}(\theta), \dot{\overline{x}}(\theta)) \geq \left(p(\theta) + \int_{[\underline{\theta},\theta)} \gamma(s)\mu(ds)\right) \cdot v - \lambda L(\theta, \overline{x}(\theta), v)$$

for all $v \in \mathbb{R}$ a.e.,

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

$$\partial_x^> h(\theta, x) \equiv \mathrm{co}\{\lim_i \xi_i \,|\, \exists t_i \to t, x_i \to x \text{ such that } h(\theta, x_i) > 0 \text{ and } \xi_i \in \partial_x h(t_i, x_i) \text{ for all } i\}.$$

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

[33] 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.

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 H1-H3 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, H1 is satisfied. H2 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 functional of x and thus H3 is also satisfied.

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

LEMMA 11 *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.2 imply*

(a). $\lambda + \max_{\theta \in \Theta} |p(\theta)| + \int_\Theta \mu(ds) = K > 0$,

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

(c). $p(\underline{\theta}) = p(\overline{\theta}) + \int_\Theta \gamma(s)\mu(ds) = 0$

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

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

PROOF OF LEMMA 11: Implications (i) and (a) are identical. Implication (ii) requires almost everywhere that

$$\dot{p}(\theta) \in \operatorname{co}\left\{ \eta \mid \left(\eta, p(\theta) + \int_{[S,t)} \gamma(s)\mu(ds), -\lambda \right) \in N_{\operatorname{epi}(L(\theta,\cdot,\cdot))}\left(\overline{x}, \dot{\overline{x}}, L(\theta, \overline{x}, \dot{\overline{x}})\right) \right\}.$$

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

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

Using the fact that $L(\cdot)$ is additively separable in x and $\dot{x}$, a basic chain rule for lower semi-continuous functions (RW, Proposition 10.5) yields

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

and

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

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

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

This simplifies yet again to the inclusion

$$N_{\text{epi}(L(\theta,\cdot,\cdot))}\left(\overline{x},\dot{\overline{x}},\overline{L}\right) \subseteq \left\{(\xi f(\theta),\xi d_2,-\xi)\,|\,\xi\geq 0,\, d_2\in\partial\left(-S(\theta,\dot{\overline{x}}(\theta))\right)\cup\partial^{\infty}\left(-S(\theta,\dot{\overline{x}}(\theta))\right)\right\}.$$

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 $\xi\geq 0$, 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 $\partial_x h(\theta,\overline{x}(\theta))=\partial_x^{\geq}h(\theta,\overline{x}(\theta))=\{-1\}$. Thus, implication (v) simplifies to $\gamma(\theta)=-1$ μ-a.e. and $supp\{\mu\}\subseteq\{t\,|\,\overline{x}(\theta)=0\}$. 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 $\gamma(\theta)=-1$ a.e. with respect to μ, condition (c) also implies

$$\int_{\Theta}\mu(ds)=\lambda.$$

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 $\int_{\Theta}\mu(ds)=1$. Thus, the normalization makes μ a probability measure on Θ. Defining $\overline{M}(\theta)=\int_{[\underline{\theta},\theta)}\mu(ds)$, the implication in (d) is therefore

$$\dot{\overline{x}}(\theta)\in\arg\max_{v\in\mathbb{R}}\; S(\theta,v)+\left(F(\theta)-\overline{M}(\theta)\right)v,\; a.e.,$$

which is condition (B2) of Theorem B.1. Lastly, the implication of (e) delivers the complementary slackness condition (B1). We have therefore proven the necessity of the conditions in Theorem B.1.

SUFFICIENCY: We adapt the argument of Arrow's Sufficiency Theorem using the basic approach of Seierstad and Sydsaeter (1987) but relaxing their continuity and smoothness assumptions.

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

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

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

To this end, it is useful to define the Hamiltonian for program (P) using $M(\theta)-F(\theta)$ as the adjoint equation which satisfies conditions(B1) and (B2):

$$H(\theta, x, v) \equiv S(\theta, v) - x \cdot f(\theta) - (\overline{M}(\theta) - F(\theta)) \cdot v.$$

Note that $\overline{M}(\theta)$ 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

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

Define the optimized Hamiltonian as

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

Because $\overline{M}(\theta) - F(\theta)$ is bounded on $(\underline{\theta}, \overline{\theta}]$ and $S(\theta, \cdot)$ is bounded from above by assumption, we note that $\hat{H}(\cdot)$ 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

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

It follows that $\Delta \geq 0$ if

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

If $\overline{M}$ were absolutely continuous, we would be able to integrate the second term by parts and reach such a conclusion. Because $\overline{M}$ is possibly discontinuous, we must proceed more carefully. Note that $\overline{M}$ is non-decreasing on $(\underline{\theta}, \overline{\theta}]$ with at most a countable number of upward jump discontinuities. Furthermore, $\overline{M}$ is absolutely continuous

elsewhere, allowing us to integrate by parts between any pair of discontinuities. Also note that at any such upward jump point, τ, $\overline{M}$ is left and right continuous with $\overline{M}(\tau) < \overline{M}(\tau^+)$ and (by condition (B1)) we have $\overline{x}(\tau^+) = 0$.

Denote the set of jump discontinuities by $\{\tau_1, \tau_2, \ldots\}$, a possibly infinite set. Let $\mathcal{I}$ be the index set of τ_i. Between any two points τ_i and τ_{i+1}, we know

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

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

Define the size of the jump discontinuity at τ by $d(\tau) = \overline{M}(\tau^+) - \overline{M}(\tau) > 0$. Then we may write

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

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

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

We deduce

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

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

Smoothness of the solution, $\overline{x}$: We add the hypothesis that

$$y(\theta, \sigma) \equiv \arg\max_{v\in\mathbb{R}} \; S(\theta, v) + (F(\theta) - \sigma)v$$

is single-valued and continuous for $(\theta, \sigma) \in \Theta \times [0, 1]$. It follows that $y(\theta, \sigma)$ is non-increasing in σ and from condition (B2), that $\dot{\overline{x}}(\theta) = q(\theta, \overline{M}(\theta))$ 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 $y(\theta, \sigma)$ is continuous in (θ, σ) jointly imply that $\dot{\overline{x}}(\theta)$ is discontinuous at τ only if $\overline{M}$ is also discontinuous at τ. Any discontinuity in $\overline{M}$, however, must be an upward jump, $d(\tau) = \overline{M}(\tau^+) - \overline{M}(\tau) > 0$, 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 $x(\theta) \geq 0$ 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 $\overline{M}$ 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. *Q.E.D.*