

# Online Appendix for *Consumer Inattention and Bill-Shock Regulation*

## B No-Free-Lunch Results

To solve the firm's problem in both attentive and inattentive cases I proceed in two steps. First, I fix a perceived utility  $U^*$  to be offered and solve for the optimal price vector and allocation which implements  $U^*$  subject to the no-free-lunch (NFL) constraint. The price vector and allocation determine expected surplus  $S$  and true expected-utility  $U$ . Hence the first step derives an optimal markup,  $\mu(U^*)$ , to be charged as a function of  $U^*$ . The second step is to choose the perceived expected-utility  $U^*$  which maximizes profits,  $\Pi = G(U^*)\mu(U^*)$ , subject to feasibility under NFL.

### B.1 Attentive Case

When consumers are attentive, the firm finds it optimal to induce the efficient allocation and charge a nonnegative penalty fee.

**Lemma 2** *Given Bernoulli taste shocks, attentive consumers who underestimate demand ( $\alpha' < \alpha$ ),  $c \in (0, 1)$ , and the NFL constraint, firms set prices which induce the efficient allocation: consumers buy if and only if  $v_t = 1$ . Moreover firms charge a nonnegative penalty fee  $p_3 \geq 0$ .*

Following Lemma 2 and conditional on a level of perceived utility  $U^*$  to be offered, the firm's problem reduces to choosing the vector of prices which maximize the expected markup  $\mu$  subject to a set of constraints. For the attentive case, it is useful to omit the penalty fee  $p_3$  and work with the price vector  $\{p_0, p_1, p_2, p_4\}$  where  $p_4 = p_2 + p_3$  is the marginal price for a second-period purchase conditional on a first-period purchase. The constraints are the NFL constraints, the incentive (IC) constraints (that the efficient allocation be incentive compatible), and the offered-utility constraint that  $U^*$  is in fact as specified. Note that the two NFL constraints  $p_0 + p_2 \geq 0$  and  $p_0 + p_1 + p_4 \geq 0$  are redundant given the IC constraints  $p_2, p_4 \geq 0$ . Thus, the firm's problem is:

$$\max_{p_0, p_1, p_2, p_4} (S^{FB} - U) \text{ such that}$$

1. NFL:  $p_0 \geq 0, p_0 + p_1 \geq 0,$
2. IC:  $p_2, p_4, v_1^* \in [0, 1], v_1^* = p_1 + \alpha'(p_4 - p_2),$
3. Perceived expected utility equals  $U^*$

where the true and perceived expected utilities are:

$$U = v_0 - p_0 + 2\alpha - \alpha(p_1 + p_2 + \alpha(p_4 - p_2)), \quad (28)$$

$$U^* = v_0 - p_0 + 2\alpha' - \alpha'(p_1 + p_2 + \alpha'(p_4 - p_2)). \quad (29)$$

Proposition 14 characterizes the solution as a function of  $U^*$ :

**Proposition 14** *Given Bernoulli taste shocks, attentive consumers who underestimate demand ( $\alpha' < \alpha$ ),  $c \in (0, 1)$ , and the NFL constraint: Firms offer the first-best allocation and charge nonnegative penalty fees. Consumers are not exploited:  $U \geq 0$ . Moreover, there are four qualitative pricing regions as a function of the offered perceived expected-utility  $U^*$ : (1)  $U^* \in [0, v_0]$ , (2)  $U^* \in [v_0, v_0 + \alpha']$ , (3)  $U^* \in [v_0 + \alpha', v_0 + \alpha'(2 - \alpha')]$ , (4)  $U^* \in [v_0 + \alpha'(2 - \alpha'), v_0 + 2\alpha']$ . ( $U^* > v_0 + 2\alpha'$  is not feasible given the NFL constraint.) Prices in each region are:*

Region	$p_0$	$p_1$	$p_2$	$p_3$
1	$v_0 - U^*$	1	1	0
2	0	$1 - \alpha'(1 - p_2)$	$1 - (U^* - v_0)/\alpha'$	$1 - p_2$
3	0	$2 - \alpha' - (U^* - v_0)/\alpha'$	0	1
4	0	0	0	$2/\alpha' - (U^* - v_0)/\alpha'^2$

Corresponding markups for each pricing region are:

Region	Markup	$-d\mu/dU^*$
1	$2\alpha(1 - c) - (U^* - v_0)$	1
2	$2\alpha(1 - c) - (\alpha/\alpha')(1 - \alpha + \alpha')(U^* - v_0)$	$(\alpha/\alpha')(1 - \alpha + \alpha')$
3	$2\alpha(1 - c) + \alpha(\alpha - \alpha') - (\alpha/\alpha')(U^* - v_0)$	$(\alpha/\alpha')$
4	$2\alpha(1 - c) + 2(\alpha/\alpha')(\alpha - \alpha') - (\alpha/\alpha')^2(U^* - v_0)$	$(\alpha/\alpha')^2$

Importantly, consumer price sensitivity is declining across the four regions, as measured by the increasing cost of raising perceived-expected utility  $-d\mu/dU^*$ .

## B.2 Inattentive case

Given Bernoulli taste shocks, an inattentive consumer's strategy is described by the pair  $\{b_0, b_1\}$ . These are the probabilities of purchase conditional on realizing  $v_t = 0$  or  $v_t = 1$  respectively:  $b_0 = \Pr(q_t(v_t = 0) = 1)$  and  $b_1 = \Pr(q_t(v_t = 1) = 1)$ . A consumer's perceived expected-utility  $U^*$

is given by equation (30) as a function of prices and the strategy  $\{b_0, b_1\}$ :

$$U^*(b_0, b_1) = -p_0 + v_0 + 2(1 - \alpha') b_0 (-\bar{p}) + 2\alpha' b_1 (1 - \bar{p}) - ((1 - \alpha') b_0 + \alpha' b_1)^2 p_3. \quad (30)$$

Firm profits, as a function of prices, perceived expected-utility  $U^*$ , and the allocation  $\{b_0, b_1\}$  are given by equation (31):

$$\Pi = G(U^*) \left( p_0 + 2((1 - \alpha) b_0 + \alpha b_1) (\bar{p} - c) + ((1 - \alpha) b_0 + \alpha b_1)^2 p_3 \right). \quad (31)$$

The first result is that it will be optimal for firms to set prices which induce the efficient allocation  $\{b_0, b_1\} = \{0, 1\}$ .

**Lemma 3** *Given Bernoulli taste shocks, inattentive consumers who underestimate demand ( $\alpha' < \alpha$ ),  $c \in (0, 1)$ , and the NFL constraint, firms set prices which induce the efficient allocation: consumers buy if and only if  $v_t = 1$ .*

To induce the efficient allocation, the firm must set expected marginal price conditional on a purchase to be between zero and one:  $0 \leq \bar{p} + \alpha' p_3 \leq 1$ . Applying Lemma 3, the firm's problem can thus be reduced to the following:<sup>38</sup>

$$\max_{U^*, \bar{p}, p_3} G(U^*) (p_0 + 2\alpha (\bar{p} - c) + \alpha^2 p_3)$$

such that

1. NFL:  $p_0 \geq 0, p_0 + \bar{p} \geq 0, p_0 + 2\bar{p} + p_3 \geq 0,$
2. IC:  $0 \leq \bar{p} + \alpha' p_3 \leq 1,$
3. Fixed Fee:  $p_0 = -U^* + v_0 + 2\alpha' (1 - \bar{p}) - \alpha'^2 p_3.$

Proposition 15 characterizes optimal prices given a fixed perceived-expected-utility  $U^*$ . For low utility offers, the NFL constraint  $p_0 + \bar{p} \geq 0$  and the IC constraint  $\bar{p} + \alpha' p_3 \leq 1$  both bind. For medium utility offers, the two NFL constraints  $p_0 \geq 0$  and  $p_0 + \bar{p} \geq 0$  both bind. Higher utility offers above  $v_0 + 2\alpha'$  are not feasible given the NFL constraint.

**Proposition 15** *Given Bernoulli taste shocks, inattentive consumers who underestimate demand ( $\alpha' < \alpha$ ),  $c \in (0, 1)$ , and the NFL constraint: Firms offer the first-best allocation and charge nonnegative surprise-penalty-fees, preferring not to disclose them at the point of sale (a strict*

---

<sup>38</sup>It is strictly optimal to set prices symmetrically,  $p_1 = p_2 = \bar{p}$ , since keeping  $\bar{p}$  constant but setting  $p_1 < p_2$  would tighten the NFL constraint  $p_0 + p_1 \geq 0$  without otherwise effecting consumer incentives or firm profits. Similarly, setting  $p_2 < p_1$  would tighten the NFL constraint  $p_0 + p_2 \geq 0$ .

preference for offers  $U^*$  below  $v_0 + \alpha'(2 - \alpha')$ ). Moreover, there are two qualitative pricing regions as a function of the offered perceived expected-utility  $U^*$ : (1) Conditional on offering  $U^* \in [0, v_0 + \alpha']$ , optimal prices and markups are

$$p_1 = p_2 = -p_0 = -\frac{v_0 + \alpha' - U^*}{1 - \alpha'}, p_3 = \frac{v_0 + 1 - U^*}{(1 - \alpha')\alpha'}. \quad (32)$$

$$\mu(U^*) = 2\alpha(1 - c) + Y - (1 + Y)(U^* - v_0). \quad (33)$$

(2) Conditional on offering  $U^* \in [v_0 + \alpha', v_0 + 2\alpha']$ , optimal prices and markups are:

$$p_1 = p_2 = p_0 = 0, p_3 = (2\alpha' + v_0 - U^*)/\alpha'^2. \quad (34)$$

$$\mu(U^*) = 2\alpha(1 - c) + 2(\alpha/\alpha')(\alpha - \alpha') - (\alpha/\alpha')^2(U^* - v_0). \quad (35)$$

(Offering  $U^* > v_0 + 2\alpha'$  is not feasible under NFL.) Importantly, consumer price-sensitivity declines across the two regions, as measured by the increasing cost of raising perceived-expected utility  $-d\mu/dU^*$ :

Region	1	2
$-d\mu/dU^*$	$-(1 + Y) <$	$(\alpha/\alpha')^2$

## C Banning Penalty Fees

In the benchmark model (as well as the model of biased beliefs in Section 5) firms optimally offer attentive consumers two-part tariffs with zero penalty fees. In this case, bill-shock regulation and banning penalty fees have the same effect on market outcomes because inattentive consumers behave as attentive consumers do when marginal prices are constant and therefore not subject to uncertainty. Proposition 3 shows that when firms price discriminate penalty fees are charged even to attentive consumers. Thus pricing resulting from a penalty fee ban (PFB) would differ than that characterized by Proposition 3. Instead pricing would be as follows:

**Proposition 16** *Whether or not consumers are attentive, given a PFB, low and high demand types,  $U_s + \frac{G_s(U_s)}{g_s(U_s)}$  increasing and  $c > 0$ , there are three cases:*

(1) *If  $\mu_L^* = \mu_H^*$ , then a single marginal cost contract,  $P(\mathbf{q}) = p_0 + c(q_1 + q_2)$ , is offered and both types receive the first-best allocation.*

(2) *If  $\mu_H^* > \mu_L^*$ , then the high type receives the first-best allocation via marginal-cost pricing ( $p_{3H} = 0, p_{1H} = p_{2H} = c$ ), while the low type's allocation is distorted downwards below first best ( $\bar{p}_L > c$ ). The downward incentive constraint is binding,  $U_H = U_{HL}$ , while the upward incentive*

constraint can be ignored. The marginal price  $\bar{p}_L$  satisfies:

$$\bar{p}_L = c + \frac{-\partial\Pi}{\partial U_H} \frac{F_L(\bar{p}_L) - F_H(\bar{p}_L)}{(1-\beta)G_L(U_L)f_L(\bar{p}_L)}.$$

(3) If  $\mu_H^* < \mu_L^*$ , then the low type receives the first-best allocation via marginal-cost pricing, while the high type's allocation is distorted upwards above first best. The upward incentive constraint is binding,  $U_L = U_{LH}$ , while the downward incentive constraint can be ignored. The marginal price  $\bar{p}_H$  satisfies:

$$\bar{p}_H = c - \frac{-\partial\Pi}{\partial U_L} \frac{F_L(\bar{p}_H) - F_H(\bar{p}_H)}{\beta G_H(U_H)f_H(\bar{p}_H)}.$$

The proof of Proposition 16 is similar to the proof of Proposition 3. However, the restriction that marginal price be constant results in a single first-order condition, rather than three. This can be obtained from the first-order condition for  $v_{1s}$  in Proposition 3 by substituting in  $v_s = v_{ss} = \bar{p}_s$  and  $p_{3s} = 0$ . Prices which result from a PFB are qualitatively different than those described by Proposition 3 because they do not involve penalty fees. Nevertheless, they share the important feature that they involve allocative distortions whenever  $\mu_L^* \neq \mu_H^*$ . As a result, Proposition 4 and Corollary 2 are equally true under a PFB as under bill-shock regulation.

Propositions 14 and 12 show that attentive but biased consumers are charged penalty fees even without a price discrimination motive if the NFL constraint is binding. Thus a PFB leads to qualitatively different pricing than bill-shock regulation. Under a PFB, the analog of Propositions 14 and 12 is below:

**Proposition 17** *Whether or not consumers are attentive, given a PFB, Bernoulli taste shocks, demand underestimation ( $\alpha' < \alpha$ ),  $c \in (0, 1)$ , and the NFL constraint: Firms offer the first best allocation, consumers are not exploited ( $U \geq 0$ ), and there are two qualitative pricing regions:*

Region	$p_0$	$\bar{p}$	$\mu(U^*)$
$U^* \in [0, v_0]$	$v_0 - U^*$	1	$2\alpha(1-c) - (U^* - v_0)$
$U^* \in [v_0, v_0 + 2\alpha']$	0	$1 - (U^* - v_0)/2\alpha'$	$2\alpha(1-c) - (\alpha/\alpha')(U^* - v_0)$

*Given duopoly on a uniform Hotelling line and base-good value  $v_0$  sufficiently large for strict full-market-coverage, there are two competitive regions over which markups are proportional to*

$\tau$ . *Markups are constant between regions.*

$\tau_{\min}$	$\tau_{\max}$	markup $\mu$	Competition
$2\alpha(1-c)$		$\tau$	fixed fee
$2\alpha'(1-c)$	$2\alpha(1-c)$	$2\alpha(1-c)$	boundary
0	$2\alpha'(1-c)$	$(\alpha/\alpha')\tau$	marginal price

*Duopoly profits equal the markup and consumers' true expected utility is  $U = S^{FB} - \mu \geq 0$ .*

The proof of Proposition 17 is similar to that of Propositions 14 and 12 but simpler. Firms first cut fixed fees but, when these are zero, begin cutting marginal fees to which consumers are less price-sensitive. Thus, a PFB leads to one of two pricing regions, one in which firms compete on fixed fees and duopoly markups are  $\tau$  and one in which firms compete on marginal price and duopoly markups are  $(\alpha/\alpha')\tau$ . A PFB may lead to smaller or larger markup reductions than bill-shock regulation but the analog of Corollary 4 for a PFB is similar. The important differences are that a PFB always leads to a strict markup reduction for  $\tau > 0$  sufficiently small, but the severe bias condition for this to hold at all  $\tau > 0$  is stronger.

**Corollary 5** *Assume duopoly on a uniform Hotelling line, the NFL constraint, Bernoulli taste shocks, inattentive consumers who underestimate demand ( $\alpha' < \alpha$ ), and  $c \in [0, 1)$ . Let  $0 < \tau < (2/3)v_0$ . The market will be fully covered and allocations will be first best with or without a PFB.*

1. *For fixed  $\tau > 0$ , if bias is sufficiently large ( $\alpha'/\alpha$  is sufficiently small) then all consumers are exploited. A PFB increases competition, strictly reduces markups, and eliminates consumer exploitation. For fixed bias, if  $\tau > 0$  is sufficiently small, a PFB increases competition and strictly reduces markups.*
2. *If bias is severe ( $\alpha'/\alpha < 1/2$ ) then a PFB strictly reduces markups for all  $\tau > 0$*
3. *If bias is mild ( $\alpha'/\alpha > 1/2$ ) then a PFB effects markups as described for severe bias except for intermediate transportation costs within the interval  $\tau \in [\tau_1, \tau_2]$ , where  $\tau_1 = (\alpha - 2c\alpha')$  and  $\tau_2 = 2\alpha(1-c)/(1+Y)$ . For  $\tau \in (\tau_1, \tau_2)$ , a PFB strictly increases markups.*

Note that proofs of results presented in this appendix are available from the author upon request.

## D Proofs

### D.1 Derivation of equation (4)

Given  $v_2^* = p_2 + q_1 p_3$ , the expected gross utility from choosing first-period threshold  $v_1^*$  is:

$$U(v_1^*) = v_0 - p_0 + \int_{v_1^*}^1 \left( v_1 - p_1 + \int_{p_2 + p_3}^1 (v_2 - p_2 - p_3) f(v_2) dv_2 \right) f(v_1) dv_1 + F(v_1^*) \int_{p_2}^1 (v_2 - p_2) f(v_2) dv_2.$$

The first-order condition,

$$\frac{dU}{dv_1^*} = f(v_1^*) \left( -v_1^* + p_1 + \int_{p_2}^{p_2 + p_3} (v_2 - p_2) f(v_2) dv_2 + (1 - F(p_2 + p_3)) p_3 \right) = 0,$$

yields equation (4). Moreover, this identifies the global maximum since for  $v_1^* > p_1 + \int_{p_2}^{p_2 + p_3} (v_2 - p_2) f(v_2) dv_2 + (1 - F(p_2 + p_3)) p_3$ ,  $\frac{dU}{dv_1^*} < 0$  and vice-versa.

### D.2 Proof of Proposition 1

A feasible inattentive strategy is a function  $b(v_t)$  which describes a purchase probability for each valuation  $v_t$  to be implemented at all  $t > 0$  independently of date or past usage. Assume that at the contracting stage a consumer plans to take strategy  $b^*$  but later considers a one time deviation to strategy  $b$ . At the planning stage, the consumer chooses  $b^*$  to maximize  $U(b^*, b^*)$ :

$$U(b^*, b^*) = v_0 - p_0 + 2 \int_0^1 \left( v - \frac{p_1 + p_2}{2} \right) b^*(v) dF(v) - p_3 \left( \int_0^1 b^*(v) dF(v) \right)^2.$$

The plan is time consistent if, when considering a one time deviation to strategy  $b$  at the implementation stage, the resulting payoff  $U(b^*, b)$  is maximized at  $b = b^*$ .

$$\begin{aligned} U(b^*, b) &= v_0 - p_0 + \int_0^1 \left( v - \frac{p_1 + p_2}{2} \right) b^*(v) dF(v) \\ &\quad + \int_0^1 \left( v - \frac{p_1 + p_2}{2} \right) b(v) dF(v) - p_3 \left( \int_0^1 b^*(v) dF(v) \right) \left( \int_0^1 b(v) dF(v) \right). \end{aligned}$$

Inspection of the first-order conditions for point-wise maximization at the planning and implementation stages,

$$\frac{dU(b^*, b)}{db(v)} = \frac{1}{2} \frac{dU(b^*, b^*)}{db^*(v)} = f(v) \left( v - \frac{p_1 + p_2}{2} - p_3 \int_0^1 b^*(v) dF(v) \right),$$

shows that the optimal strategy at the planning stage is a threshold strategy satisfying equation (6) and that it is time consistent. A nonnegative penalty fee is sufficient for  $\frac{d}{dv^*} \left( \frac{1}{f(v^*)} \frac{dU(v^*)}{dv^*} \right) = -2(1 + f(v^*)p_3)$  to be strictly negative, which in turn is a sufficient second-order condition for the consumer's maximization problem.

### D.3 Proof of Proposition 2

Firm  $i$ 's profits can be written as  $\Pi^i = G(U^i; U^{-i})(S^i - U^i)$ . For any fixed utility offer  $U^i$ , profits are maximized by choosing marginal prices  $p_1^i$ ,  $p_2^i$ , and  $p_3^i$  to achieve first-best gross surplus, while adjusting the fixed fee  $p_0^i$  to keep  $U^i$  constant. This is true independent of regulation. The offered gross utility  $U^i$  is set via the fixed fee  $p_0^i$  to balance rent extraction versus participation, as in a basic monopoly pricing problem. Firms' gross utility offers are independent of regulation, which implies that matching between firms and consumers and transportation costs are also independent of regulation.

Given attentive consumers and continuously-distributed taste-shocks,  $p_1^i = p_2^i = c$  and  $p_3^i = 0$  are the unique marginal prices which achieve  $S^{FB}$ . Given inattentive consumers and continuously-distributed taste-shocks, any marginal prices which implement  $v^* = c$  are optimal. These include all marginal prices which satisfy  $p_3^i \geq 0$  and equation (6) at  $c = v^*$  because equation (6) is sufficient as well as necessary for incentive compatibility given  $p_3^i \geq 0$ .

### D.4 Proof of Proposition 3

Note, in the proof I write the firm's problem as a choice of marginal prices  $p_{2s}$  and  $p_{4s}$  rather than  $p_{2s}$  and  $p_{3s}$ , where  $p_{4s} = p_{2s} + p_{3s}$ .

**Case I:** The result for  $\mu_H^* = \mu_L^*$  follows because the optimal solution when both IC constraints are relaxed is a single marginal-cost contract with markup  $\mu_L^*$ .

**Case II.** Assume  $\mu_H^* > \mu_L^*$ . Relax the upward incentive constraint  $U_L \geq U_{LH}$  (IC-L).

(1) **Claim:** IC-L slack implies marginal cost pricing ( $v_H = p_{2H} = c$  and  $p_{3H} = 0$ ) and first-best allocation for the high type. **Proof:** Suppose not. Then setting  $\{v_H, p_{2H}, p_{4H}\}$  equal to  $\{c, c, c\}$  while keeping  $U_H$  constant keeps IC-H and participation unaffected without violating IC-L since it has been relaxed. However, it increases surplus and hence profit from type  $H$  - a contradiction.

(2) **Claim:** The downward incentive constraint  $U_H \geq U_{HL}$  (IC-H) binds with equality:  $U_H = U_{HL}$ . **Proof:** Suppose that IC-H were slack. Then there would be marginal cost pricing  $P_s(\mathbf{q}) = p_{0s} + c(q_1 + q_2)$  such that  $U_H = S_H^{FB} - p_{0H}$  and  $U_{HL} = S_H^{FB} - p_{0L}$ . Thus IC-H would be equivalent to  $S_H^{FB} - p_{0H} \geq S_H^{FB} - p_{0L}$ , which implies  $S_H^{FB} - p_{0H} \geq S_L^{FB} - p_{0L}$  and hence  $\mu_H^* \leq \mu_L^*$  at optimal offers  $\{U_H, U_L\}$ . This is a contradiction so IC-H must bind. Moreover, IC-H will bind with equality

given either ZOOM (where  $\partial\Pi/\partial U_H = -\beta$  for all  $U_H > 0$ ) or HOO (where the decreasing marginal revenue assumption,  $U_s + \frac{G_s(U_s)}{g_s(U_s)}$  increasing, implies profits are quasi-concave in  $U_H$ ).

(3) The downward incentive constraint  $U_H \geq U_{HL}$  (IC-H) is convenient to re-express as  $U_H \geq U_L + (U_{HL} - U_L)$ . Let  $Z = (U_{HL} - U_L)$ . Integrating by parts, equation (8) reduces to:

$$U_{s\hat{s}} = v_0 - p_{0\hat{s}} + \int_{v_{s\hat{s}}}^1 (v - p_{1\hat{s}}) dF_s(v) + F_s(v_{s\hat{s}}) \int_{p_{2\hat{s}}}^1 (1 - F_s(v)) dv + (1 - F_s(v_{s\hat{s}})) \int_{p_{4\hat{s}}}^1 (1 - F_s(v)) dv. \quad (36)$$

Thus the expression for  $Z$  can be re-written as:

$$\begin{aligned} Z &= \int_{v_{HL}}^1 (v - p_{1L}) f_H(v) dv - \int_{v_L}^1 (v - p_{1L}) f_L(v) dv \\ &+ \int_{p_{2L}}^1 (F_H(v_{HL})(1 - F_H(v)) - F_L(v_L)(1 - F_L(v))) dv \\ &+ \int_{p_{4L}}^1 ((1 - F_H(v_{HL}))(1 - F_H(v)) - (1 - F_L(v_L))(1 - F_L(v))) dv \end{aligned} \quad (37)$$

where from equation (7) evaluated at  $\hat{s} = s = L$ ,

$$p_{1L} = v_L - \int_{p_{2L}}^{p_{4L}} (1 - F_L(v)) dv, \quad (38)$$

and substituting this into equation (7) evaluated at  $\{s, \hat{s}\} = \{H, L\}$ ,

$$v_{HL} = v_L + \int_{p_{2L}}^{p_{4L}} (F_L(v) - F_H(v)) dv. \quad (39)$$

Given (1) and (2), the firm's problem can be reduced to:

$$\max_{U_L, v_L, p_{2L}, p_{4L}} \left\{ \begin{array}{l} (1 - \beta) G_L(U_L) (S_L(v_L, p_{2L}, p_{4L}) - U_L) \\ + \beta G_H(U_H(U_L, v_L, p_{2L}, p_{4L})) (S_H^{FB} - U_H(U_L, v_L, p_{2L}, p_{4L})) \end{array} \right\}$$

where  $U_H(U_L, v_L, p_{2L}, p_{4L}) = U_L + Z(v_L, p_{2L}, p_{4L})$  and  $Z$  is characterized by equations (37)-(39).

For the remainder of the proof, I suppress subscript "L" from marginal prices  $p_{1L}$ ,  $p_{2L}$ , and  $p_{4L}$ .

The first-order condition for any  $x \in \{v_L, p_2, p_4\}$  is:

$$\frac{d\Pi}{dx} = \frac{\partial\Pi}{\partial S_L} \frac{dS_L}{dx} + \frac{\partial\Pi}{\partial U_H} \frac{dU_H}{dx} = 0.$$

By the envelope condition,  $\partial U_H / \partial v_{HL} = 0$  and hence for any  $x \in \{v_L, p_2, p_4\}$

$$\frac{dU_H}{dx} = \frac{\partial U_H}{\partial x} + \frac{\partial U_H}{\partial p_1} \frac{dp_1}{dx} + \frac{\partial U_H}{\partial v_{HL}} \frac{dv_{HL}}{dx} = \frac{\partial Z}{\partial x} + \frac{\partial Z}{\partial p_1} \frac{dp_1}{dx}.$$

The derivatives in the first term of the first-order condition are:  $\partial \Pi / \partial S_L = (1 - \beta) G_L(U_L)$ ,

$$\frac{dS_L}{dv_L} = f_L(v_L) \left( \int_{p_2}^{p_4} (v - c) f_L(v) dv - (v_L - c) \right),$$

$dS_L/dp_2 = -F_L(v_L)(p_2 - c)f_L(p_2)$ , and  $dS_L/dp_4 = -(1 - F_L(v_L))(p_4 - c)f_L(p_4)$ . Components of the second term in the first-order condition are:  $dp_1/dv_L = 1$ ,  $dp_1/dp_2 = (1 - F_L(p_2))$ ,  $dp_1/dp_4 = -(1 - F_L(p_4))$ ,  $\partial Z / \partial p_1 = (F_H(v_{HL}) - F_L(v_L))$ ,

$$\frac{dU_H}{dp_2} = \frac{\partial Z}{\partial p_2} + \frac{\partial Z}{\partial p_1} \frac{dp_1}{dp_2} = F_H(v_{HL})(F_H(p_2) - F_L(p_2)),$$

$$\frac{dU_H}{dp_4} = \frac{\partial Z}{\partial p_4} + \frac{\partial Z}{\partial p_1} \frac{dp_1}{dp_4} = (1 - F_H(v_{HL}))(F_H(p_4) - F_L(p_4)),$$

and finally since by the envelope condition  $\partial Z / \partial v_L = 0$ ,

$$\frac{dU_H}{dv_L} = \frac{\partial Z}{\partial v_L} + \frac{\partial Z}{\partial p_1} \frac{dp_1}{dv_L} = \frac{\partial Z}{\partial p_1} = (F_H(v_{HL}) - F_L(v_L)).$$

Putting all these pieces together gives the first-order conditions

$$\frac{d\Pi}{dv_L} = (1 - \beta) G_L(U_L) f_L(v_L) \left( \int_{p_2}^{p_4} (v - c) f_L(v) dv - (v_L - c) \right) + \frac{\partial \Pi}{\partial U_H} (F_H(v_{HL}) - F_L(v_L)) = 0$$

$$\frac{d\Pi}{dp_2} = -(1 - \beta) G_L(U_L) F_L(v_L) (p_2 - c) f_L(p_2) + \frac{\partial \Pi}{\partial U_H} F_H(v_{HL}) (F_H(p_2) - F_L(p_2)) = 0 \quad (40)$$

$$\frac{d\Pi}{dp_4} = -(1 - \beta) G_L(U_L) (1 - F_L(v_L)) (p_4 - c) f_L(p_4) + \frac{\partial \Pi}{\partial U_H} (1 - F_H(v_{HL})) (F_H(p_4) - F_L(p_4)) = 0 \quad (41)$$

which can be rearranged to derive equations (17)-(19).

**(4) Claim:**  $v_L, p_2, p_4 > c$  and  $p_4 > p_2$ .

**(a) Claim:**  $p_2, p_4 > c$ : **Proof:** The fact that IC-H is binding implies that  $\partial \Pi / \partial U_H < 0$ . By inspection of equations (18)-(19), first-order stochastic dominance (FOSD) implies  $p_2 \geq c$  and  $p_4 \geq c$ . Moreover,  $p_2 \neq c$  and  $p_4 \neq c$  because FOSD is strict at  $c$ . Therefore  $p_2, p_4 > 0$ .

**(b) Claim:**  $p_4 > p_2$ . **Proof:** Suppose not and  $p_2 \geq p_4$ . If  $p_2 = p_4$  then  $v_{HL} = v_L$  by equation (39). Given  $p_2 = p_4$  and  $v_{HL} = v_L$  equations (18)-(19) imply  $F_H(v_L) = F_L(v_L)$ . Hence equation (17) implies  $v_L = c$ . However this contradicts  $F_H(v_L) = F_L(v_L)$  given strict FOSD at  $c$ . Therefore

$p_2 \neq p_4$ .

If  $p_2 > p_4$  then  $v_{HL} < v_L$  by equation (39) and FOSD. (Note that the inequality is strict because equations (18)-(19) and  $p_2, p_4 > c$  imply that  $F_H(p_2) < F_L(p_2)$  and  $F_H(p_4) < F_L(p_4)$ . Therefore, by continuity of  $F_H$  and  $F_L$ ,  $\int_{p_4}^{p_2} (F_L(v) - F_H(v)) dv > 0$ .) Given  $v_{HL} < v_L$ ,  $F_H$  strictly increasing and FOSD imply  $F_H(v_{HL}) < F_H(v_L) \leq F_L(v_L)$ . Therefore it holds that  $\frac{1-F_H(v_{HL})}{1-F_L(v_L)} > 1 > \frac{F_H(v_{HL})}{F_L(v_L)}$ , which in turn implies that  $\int_a^b \frac{d\Pi}{dp_4} > 0$  follows from  $\int_a^b \frac{d\Pi}{dp_2} \geq 0$  for any  $b > a$  given equations (40)-(41). Now the fact that  $p_2 > p_4$  is optimal implies that  $\int_{p_4}^{p_2} \frac{d\Pi}{dp_2} \geq 0$  so it must also be true that  $\int_{p_4}^{p_2} \frac{d\Pi}{dp_4} > 0$ , contradicting optimality of  $p_4$ . Therefore  $p_4 > p_2$ .

**(c) Claim:**  $v_L > c$ . **Proof:** Given  $F_L(v_L) - F_H(v_{HL}) \geq 0$  and  $p_4 > p_2$ , equation (17) implies that  $v_L > c$ . Thus, it is sufficient to show that  $F_L(v_L) - F_H(v_{HL}) \geq 0$ . Suppose not and  $F_L(v_L) - F_H(v_{HL}) < 0$ . Then  $\frac{1-F_H(v_{HL})}{1-F_L(v_L)} < 1 < \frac{F_H(v_{HL})}{F_L(v_L)}$  and, by a similar comparison of derivatives in equations (40)-(41) as made above in part (b), it follows that  $p_2 > p_4$ , which is a contradiction.

**(5) Claim:** IC-L is satisfied. **Proof:** The final step is to show that the relaxed IC-L constraint is satisfied. This follows from the fact that quantities are monotonic in the ex ante signal:  $q_{t,H}(v^t) \geq q_{t,L}(v^t)$ . To show that IC-L is satisfied, it is sufficient to show that

$$U_H - U_{HL} + U_L(v_{HL}) - U_{LH} \geq 0, \quad (42)$$

where by  $U_L(v_{HL})$  I mean the expected gross utility of type  $L$  who chooses contract  $L$  but uses the optimal first-period threshold of a deviating high type. Because  $U_H = U_{HL}$ , this implies  $U_L(v_{HL}) \geq U_{LH}$  and therefore, as  $U_L \geq U_L(v_{HL})$ , that IC-L is satisfied.

I substitute for utilities in equation (42) using equation (8) and adjusting for the different threshold in the case of  $U_L(v_{HL})$ . After several lines of algebra and integration by parts, this yields the equivalent inequality

$$\int_{p_2}^{p_4} \left( \begin{array}{c} (1 - F_H(v_{HL}))(1 - F_H(v)) \\ -(1 - F_L(v_{HL}))(1 - F_L(v)) \end{array} \right) dv + \int_c^{p_2} (F_L(v) - F_H(v)) dv \geq 0,$$

which is satisfied (strictly) by FOSD and  $p_4 > p_2 > c$ .

**Case III.** The result for  $\mu_H^* < \mu_L^*$  follows by a symmetric argument, where I start by relaxing IC-H and showing that IC-L must bind with equality.

## D.5 Proof of Proposition 4

**Preliminary result:** A useful result not included in Proposition 3 is that  $\mu_H^{*A} > \mu_L^{*A}$  implies  $\mu_H^A > \mu_L^A$ : Suppose not and  $\mu_H^{*A} > \mu_L^{*A}$  but  $\mu_H^A \leq \mu_L^A$ . By Proposition 3,  $S_L^A < S_L^{FB}$ . Therefore,

a strictly profitable deviation would be to offer a single marginal-cost contract with markup  $\mu = (\max\{\mu_H^A, \mu_L^{*A}\} + \min\{\mu_L^A, \mu_H^{*A}\})/2$ . Ignoring the IC constraints, raising  $S_L^A$  to  $S_L^{FB}$  holding markups fixed strictly raises market share in segment  $L$ . Then changing markups to  $\mu$  moves  $\mu_s^A$  weakly closer to  $\mu_s^*$  for  $s \in \{L, H\}$ , thereby weakly raising profits in each segment as allocations are first best and profits are quasi-concave in  $U_s$ . Finally, the deviation contract is trivially incentive compatible. Therefore  $\mu_H^A > \mu_L^A$ . By symmetric argument,  $\mu_H^{*A} < \mu_L^{*A}$  implies  $\mu_H^A < \mu_L^A$ .

**Proof of proposition:** Assuming  $\tau_L$  and  $\tau_H$  are sufficiently small for strict full-market-coverage in equilibrium ensures that every customer strictly prefers to purchase from one of the two firms. In this case, firm A's best response utility offer  $U_s^A$  is always within an open interval for which residual demand from consumers of type  $s$  is  $G_s(U_s^A) = \frac{1}{2\tau_s}(U_s^A - U_s^B + \tau_s)$ . Note that firm A's profits are linear in firm B's offer  $U_s^B$  and hence firm A's expected profits and best response depend only on firm B's expected offer  $E[U_s^B] = \bar{U}_s^B$ . Given  $G_s(U_s^A) = \frac{1}{2\tau_s}(U_s^A - \bar{U}_s^B + \tau_s)$ , the definition of  $\mu_s^*$  implies that  $\mu_s^{*A} = (\bar{\mu}_s^B + \tau_s)/2 + (S_s^{FB} - \bar{S}_s^B)/2$ .

(1)  $\tau_H = \tau_L = \tau$ : (a) In the proposed equilibrium,  $\mu_s^A = \mu_s^B = \tau$  and  $S_s^A = S_s^B = S_s^{FB}$ . Thus, given  $\mu_s^{*A} = (\bar{\mu}_s^B + \tau_s)/2 + (S_s^{FB} - \bar{S}_s^B)/2$ , it follows that  $\mu_s^{*A} = \mu_s^{*B} = \tau$ . Therefore Proposition 3 implies firm offers are best responses to each other.

(b) It remains to show that the proposed equilibrium is unique. I claim that in any equilibrium it holds for both firms that  $\mu_L^* = \mu_H^* = \mu^*$ . Then by Proposition 3, both firms offer a marginal-cost contract with fixed fee  $\mu^*$  and first-best allocations. Therefore, given  $\mu_s^{*A} = (\bar{\mu}_s^B + \tau_s)/2 + (S_s^{FB} - \bar{S}_s^B)/2$ , unconstrained optimal markups satisfy  $\mu_s^{*A} = (\mu^* + \tau)/2$ . Together,  $\mu_s^{*A} = (\mu^* + \tau)/2$  and  $\mu_s^{*A} = \mu^*$  imply  $\mu^* = \tau$  and equilibrium coincides with the proposal. Below I show that  $\mu_H^* \leq \mu_L^*$  for both firms by assuming that  $\mu_H^{*A} > \mu_L^{*A}$  and deriving a contradiction. An (omitted) symmetric argument shows that  $\mu_H^* \geq \mu_L^*$ . Thus my claim  $\mu_L^* = \mu_H^* = \mu^*$  and uniqueness hold.

Proof of claim  $\mu_H^* \leq \mu_L^*$ : Assume that  $\mu_H^{*A} > \mu_L^{*A}$ . For all offers in A's best response, it holds that  $\mu_H^A > \mu_L^A$  and that (following the proof of Proposition 3) IC-H will bind while IC-L is slack so that  $-\partial\Pi^A/\partial U_H^A = \partial\Pi^A/\partial U_L^A > 0$ . Moreover, these inequalities hold in expectation if A uses a mixed strategy:  $\bar{\mu}_H^A > \bar{\mu}_L^A$  and  $-E[\partial\Pi^A/\partial U_H^A] = E[\partial\Pi^A/\partial U_L^A] > 0$ . As  $E[\partial\Pi^A/\partial U_H^A] = (\bar{\mu}_H^A - (\bar{U}_H^A - \bar{U}_H^B) - \tau)/2\tau$  and  $E[\partial\Pi^A/\partial U_L^A] = (\bar{\mu}_L^A - (\bar{U}_L^A - \bar{U}_L^B) - \tau)/2\tau$ , the preceding inequalities imply that  $(\bar{U}_H^A - \bar{U}_H^B) > (\bar{U}_L^A - \bar{U}_L^B)$ . Therefore on average firm A's market share in the high segment is strictly higher than its market share in the low segment. By symmetry  $\mu_H^{*B} > \mu_L^{*B}$  would imply firm B would also have a higher share in segment  $H$  than segment  $L$ , contradicting full-market-coverage in both segments. Therefore  $\mu_H^{*B} \leq \mu_L^{*B}$ .

Given  $\mu_H^{*A} > \mu_L^{*A}$ , IC-H is binding for firm A's offers and high types are indifferent between firm A's low and high contracts. Given  $\mu_H^{*A} > \mu_L^{*A}$  and  $\mu_H^{*B} \leq \mu_L^{*B}$ , Proposition 3 implies that

$q_L^A < q_L^{FB} = q_L^B$ . Consider the location distance  $x$  from firm A at which low types are indifferent between low contracts offered by firms A and B. At this point high types must strictly prefer firm B's low contract to firm A's contract. This follows because both low and high types view the difference in fixed fees and transportation costs between the two low contracts equally but high types value the reduction in marginal price at firm B strictly more than low types. Because high types are indifferent between firm A's two contracts they must choose firm B's contract at location  $x$ . This contradicts firm A having a larger average market share in segment  $H$  than segment  $L$ . Therefore  $\mu_H^* \leq \mu_L^*$ .

(2) If  $\tau_H \neq \tau_L$ , then all equilibria are inefficient: Suppose not, and in equilibrium allocations are efficient. Then  $p_{3s}^B = 0$  and  $p_{1s}^B = p_{2s}^B = c$ . Incentive compatibility implies  $p_{0L}^B = p_{0H}^B$  and therefore  $\mu_H^B = \mu_L^B = \mu^B$ . These statements hold for any offer in B's mixed strategy. Thus  $\mu_s^{*A} = \frac{1}{2}(\bar{\mu}^B + \tau_s)$ , which implies  $\mu_L^{*A} \neq \mu_H^{*A}$ , and by Proposition 3 A's best response includes an inefficient contract.

(3a) Suppose  $\tau_L > \tau_H$ . In a symmetric equilibrium, either  $\mu_L^* = \mu_H^*$ ,  $\mu_L^* < \mu_H^*$ , or  $\mu_L^* > \mu_H^*$  must hold for both firms. Part (2) rules out  $\mu_L^* = \mu_H^*$  if  $\tau_L > \tau_H$ . Suppose that  $\mu_L^* < \mu_H^*$  for both firms. Then, by the same argument used in part (1),  $\bar{\mu}_H^A > \bar{\mu}_L^A$  and  $-E[\partial\Pi^A/\partial U_H^A] = E[\partial\Pi^A/\partial U_L^A] > 0$ . In a symmetric equilibrium  $\bar{U}_s^A = \bar{U}_s^B$  so  $E[\partial\Pi^A/\partial U_H^A] = (\bar{\mu}_H^A - \tau_H)/2\tau_H$  and  $E[\partial\Pi^A/\partial U_L^A] = (\bar{\mu}_L^A - \tau_L)/2\tau_L$  and hence these inequalities imply  $\tau_L < \bar{\mu}_L^A < \bar{\mu}_H^A < \tau_H$ , which contradicts  $\tau_L > \tau_H$ . Therefore  $\mu_L^* > \mu_H^*$  for both firms and Proposition 3 implies the result.

(3b)  $\tau_H > \tau_L$  follows a symmetric argument.

## D.6 Proof of Proposition 5

Solving equation (11) at  $\hat{s} = s$  for  $p_{0s}$  yields

$$p_{0s} = -U_s + v_0 + 2 \int_{v_s}^1 v dF_s(v) - 2\bar{p}_s(1 - F_s(v_s)) - p_{3s}(1 - F_s(v_s))^2. \quad (43)$$

Similarly, solving equation (10) at  $\hat{s} = s$  for  $\bar{p}_s$  yields

$$\bar{p}_s = v_s - p_{3s}(1 - F_s(v_s)). \quad (44)$$

Substituting equations (43)-(44) into equation (11) yields

$$U_{s\hat{s}} = U_{\hat{s}} + 2 \int_{v_{s\hat{s}}}^1 (v - v_{\hat{s}}) dF_s(v) - 2 \int_{v_{\hat{s}}}^1 (v - v_{\hat{s}}) dF_{\hat{s}}(v) - p_{3\hat{s}}(F_{\hat{s}}(v_{\hat{s}}) - F_s(v_{s\hat{s}}))^2. \quad (45)$$

By the envelope condition:

$$\frac{d}{dp_{3\hat{s}}}U_{s\hat{s}} = \frac{\partial}{\partial p_{3\hat{s}}}U_{s\hat{s}} = -(F_{\hat{s}}(v_{\hat{s}}) - F_s(v_{s\hat{s}}))^2 \leq 0. \quad (46)$$

## D.7 Proof of Proposition 6

Proposition 6 is stated for either of two restrictions: (1)  $p_{3s} \leq p^{\max}$  or (2)  $p_{3s} \leq v_s / (1 - F_s(v_s))$ . Both can be written as  $p_{3s} \leq h(v_s)$  for some  $h(v_s)$  that is strictly positive and nondecreasing. All but the last step of the proof work with the restrictions in this general form.

**Step I.** First consider half the parameter space,  $\mu_H^* \geq \mu_L^*$ , and relax IC-L:

By equation (45), IC-H is given by equation (47):

$$U_H \geq U_{HL} = U_L + 2 \int_{v_{HL}}^1 (v - v_L) dF_H(v) - 2 \int_{v_L}^1 (v - v_L) dF_L(v) - p_{3L} (F_L(v_L) - F_H(v_{HL}))^2, \quad (47)$$

where substituting  $\bar{p}_L$  from equation (44) into equation (10) uniquely defines  $v_{HL}$  as

$$v_{HL} = v_L + p_{3L} (F_L(v_L) - F_H(v_{HL})). \quad (48)$$

There are two cases: either (1) IC-H is slack or (2) IC-H binds.

**Case (1)**, IC-H is slack:

(a) Show that IC-L is satisfied given  $p_{3H} \geq 0$ : Since IC-L is relaxed by increasing  $p_{3H}$ , it is sufficient to check at  $p_{3H} = 0$ . If both IC-L and IC-H are slack, then  $v_L = v_H = c$  and at  $p_{3H} = 0$ ,  $P_H(\mathbf{q}) = p_{0H} + c(q_1 + q_2)$ , so that  $U_H = S_H^{FB} - p_{0H}$  and  $U_{LH} = S_L^{FB} - p_{0H}$ . Thus IC-L,  $U_L \geq U_{LH}$ , is equivalent to  $S_H^{FB} - U_H \geq S_L^{FB} - U_L$ , or  $\mu_H^* \geq \mu_L^*$  at optimal offer  $\{U_H, U_L\}$  which is satisfied by assumption.

(b) Substituting  $v_L = c$  into equation (47), gives

$$U_H \geq U_{HL} = (U_L - S_L^{FB}) + 2 \int_{v_{HL}}^1 (v - c) dF_H(v) - p_{3L} (F_L(c) - F_H(v_{HL}))^2.$$

Noting that  $2 \int_{v_{HL}}^1 (v - c) dF_H(v) = S_H^{FB} - 2 \int_c^{v_{HL}} (v - c) dF_H(v)$  and, by definition at optimal utility offers,  $(S_H^{FB} - \hat{U}_H) - (S_L^{FB} - \hat{U}_L) = \mu_H^* - \mu_L^*$ , IC-H simplifies to:

$$(\mu_H^* - \mu_L^*) \leq 2 \int_c^{v_{HL}} (v - c) dF_H(v) + p_{3L} (F_L(c) - F_H(v_{HL}))^2. \quad (49)$$

Further, substituting  $v_L = c$  into equation (48) yields  $v_{HL} = c + p_{3L} (F_L(c) - F_H(v_{HL}))$ . Finally, substituting  $p_{3L} (F_L(c) - F_H(v_{HL})) = v_{HL} - c$  into the right side of equation (49) shows IC-H is

equivalent to  $(\mu_H^* - \mu_L^*) \leq X_H$  when  $p_{3L}$  is set to the maximum  $h_L(c)$ .

**Case (2)**, IC-H binds: If  $\mu_H^* - \mu_L^* > X_H$ , then IC-H cannot be relaxed. Moreover, it will bind with equality given either ZOOM (where  $\partial\Pi/\partial U_H = -\beta$  for all  $U_H > 0$ ) or HOO (where decreasing marginal revenue assumption,  $U_s + \frac{G_s(U_s)}{g_s(U_s)}$  increasing, implies profits are quasi-concave in  $U_s$ ). Therefore  $p_{3L} = h_L(v_L)$

(a) Show  $v_H = c$ , derive FOC for  $v_L$ , and show  $v_L > c$ : **(i)**  $v_H = c$ : Suppose  $v_H \neq c$ . It would strictly increase profits to set  $v_H = c$  while holding  $U_H$  fixed. Doing so strictly increases profits from  $H$  types without affecting IC-H and by assumption IC-L is relaxed. **(ii)** FOC for  $v_L$ : The profit maximization problem is

$$\begin{aligned} & \max_{U_L, v_L} ((1 - \beta) G_L(U_L)(S_L - U_L) + \beta G_H(U_H)(S_H^{FB} - U_H)) \\ \text{s.t. } U_H &= U_L + 2 \int_{v_{HL}}^1 (v - v_L) dF_H(v) - 2 \int_{v_L}^1 (v - v_L) dF_L(v) - p_{3L}(F_L(v_L) - F_H(v_{HL}))^2 \\ S_L &= 2 \int_{v_L}^1 (v - c) dF_L(v), v_{HL} = v_L + p_{3L}(F_L(v_L) - F_H(v_{HL})), \text{ and } p_{3L} = h(v_L). \end{aligned}$$

By the envelope condition,  $\partial U_H / \partial v_{HL} = 0$  so the FOC for  $v_L$  is:

$$\frac{d\Pi}{dv_L} = \frac{\partial\Pi}{\partial S_L} \frac{dS_L}{dv_L} + \frac{\partial\Pi}{\partial U_H} \left( \frac{\partial U_H}{\partial v_L} + \frac{\partial U_H}{\partial p_{3L}} h'(v_L) \right)$$

Taking derivatives and substituting equation (46) for  $\frac{\partial U_H}{\partial p_{3L}} = \frac{\partial U_{HL}}{\partial p_{3L}}$  gives:

$$\begin{aligned} \frac{d\Pi}{dv_L} &= -2(1 - \beta) G_L(U_L)(v_L - c) f_L(v_L) \\ &\quad - \frac{\partial\Pi}{\partial U_H} \left[ 2(F_L(v_L) - F_H(v_{HL}))(1 + p_{3L}f_L(v_L)) + (F_L(v_L) - F_H(v_{HL}))^2 h'(v_L) \right]. \end{aligned}$$

The FOC  $\frac{d\Pi}{dv_L} = 0$  simplifies to equation (24), or for nonnegative marginal prices,  $h_s(v_s) = v_s / (1 - F_s(v_s))$ , to:

$$v_L = c + \frac{\beta}{1 - \beta} \frac{F_L(v_L) - F_H(v_{HL})}{f_L(v_L)} \frac{-\partial\Pi/\partial U_H}{\beta G_L(U_L)} (1 + p_{3L}f_L(v_L)) \left( 1 + \frac{1}{2} \frac{(F_L(v_L) - F_H(v_{HL}))}{(1 - F_L(v_L))} \right).$$

(iii)  $v_L > c$ : Note that  $F_L(v_L) - F_H(v_{HL}) > 0$ . (Suppose not. Then by equation (48) and  $p_{3L} = h_L(v_L) > 0$ ,  $v_{HL} \leq v_L$ . Combined with FOSD, this implies  $0 \leq F_L(v_L) - F_H(v_L) \leq F_L(v_L) - F_H(v_{HL}) \leq 0$ . Then, by equation (24),  $v_L = c$ , which yields a contradiction as  $F_L(c) > F_H(c)$  by assumption.) As IC-H is binding,  $\frac{\partial\Pi}{\partial U_H} < 0$ . Therefore, equation (24) and  $F_L(v_L) - F_H(v_{HL}) > 0$  imply  $v_L > c$ .

(b) Show that IC-L is satisfied given  $p_{3H} \geq 0$ : Suppose that  $\{U_L, v_L, p_{3L}, U_H, v_H, p_{3H}\}$  is the relaxed solution, with IC-H binding so that equation (47) holds with equality. Now consider the alternative contract menu  $\{U_L, v_L, \hat{p}_{3L}, \hat{U}_H, v_H, \hat{p}_{3H}\}$  with  $\hat{p}_{3L} = \hat{p}_{3H} = 0$  and

$$\hat{U}_H = U_L + 2 \int_{v_{HL}}^1 (v - v_L) dF_H(v) - 2 \int_{v_L}^1 (v - v_L) dF_L(v) > U_H$$

(preserving IC-H with equality). In this case IC-H equality, FOSD, and  $v_L \geq v_H$  imply IC-L. This follows standard logic: high types are willing to pay more ex ante for a decrease in marginal price than are low types. If high types are just indifferent to the upgrades, then low types won't find it worthwhile. Next move back to the original contract menu, in two steps. First adjust the  $p_{3s}$  keeping  $\hat{U}_H$  fixed. We know that this relaxes the IC constraints, so it is still IC. Second decrease  $U_H$  back to IC-H binding. The decrease in  $U_H$  relaxes IC-L still further, so it is still satisfied.

**Step II.** Now consider the other half of the parameter space,  $\mu_H^* \leq \mu_L^*$ . The results follow by a nearly symmetrical argument. The only important difference is that for  $\mu_H^* - \mu_L^* < -X_L$ , the first-order condition for  $v_H$ ,

$$v_H = c - \frac{1 - \beta}{\beta} \frac{F_L(v_{LH}) - F_H(v_H)}{f_H(v_H)} \frac{-\partial \Pi / \partial U_L}{(1 - \beta) G_H(U_H)} \left( (1 + p_{3H} f_H(v_H)) - \frac{1}{2} (F_L(v_{LH}) - F_H(v_H)) h'_H(v_H) \right),$$

may call for  $v_H > c$  if  $h'_H(v_H)$  is sufficiently positive, which would violate the relaxed IC-H condition. However, the proposition is stated for  $h_H(v_H) = p^{\max}$  or for  $h_H(v_H) = \frac{v_H}{1 - F_H(v_H)}$  rather than for general  $h_H(v_H)$ . In the former case there is no issue, since  $h'_H(v_H) = 0$ . In the latter case, there is an additional step to show that  $v_H < c$ . Given  $p_{3H} = h_H(v_H) = \frac{v_H}{1 - F_H(v_H)}$ , the first-order condition for  $v_H$  can be re-written as

$$v_H = c - \frac{1 - \beta}{\beta} \frac{F_L(v_{LH}) - F_H(v_H)}{f_H(v_H)} \frac{-\partial \Pi / \partial U_L}{(1 - \beta) G_H(U_H)} (1 + p_{3H} f_H(v_H)) \left( 1 - \frac{1}{2} \frac{(F_L(v_{LH}) - F_H(v_H))}{(1 - F_H(v_H))} \right).$$

In this form, it is apparent by inspection that  $v_H < c$ , despite  $h' > 0$ .

## D.8 Proof of Corollary 1

By assumption (ZOOM),  $G_s(U_s) = 1$  if  $U_s \geq 0$  and  $G_s(U_s) = 0$  otherwise. Thus  $\mu_H^* = S_H^{FB} > S_L^{FB} = \mu_L^*$  and either case (1) or (2) of Proposition 6 holds. Suppose that case (1) holds because  $\mu_H^* - \mu_L^* \leq X_H$ . Then Proposition 6 implies that neither IC-L nor IC-H bind,  $U_L = U_H = 0$  and  $v_L = v_H = c$ . To show a contradiction, notice that the high type can mimic the low type's contract choice and purchase probability by choosing contract  $L$  and a threshold  $v_{HL}$  such that  $F_H(v_{HL}) = F_L(c)$ . By assumption,  $F_H(c) < F_L(c)$ , which implies that  $v_{HL} > c$ . By mimicking

the low type, the high type makes the same expected payments and the same number of purchases as the low type, but at FOSD higher valuations. Thus  $U_{HL} > U_L$ , which given  $U_L = U_H$  implies that  $U_{HL} > U_H$ . This violates IC-H, a contradiction. Therefore  $\mu_H^* - \mu_L^* > X_H$ , and the corollary follows from case (2) of Proposition 6.

## D.9 Proof of Proposition 7

(1) Show proposed equilibrium exists by construction: Impose  $p_{3s} \leq h_s(v_s) = v_s / (1 - F_s(v_s))$ . Assume that each firm offers  $p_{3s} = h_s(c)$ ,  $v_L = v_H = c$ , and  $U_s = S_s^{FB} - \tau_s$ . In this case,  $U_s = \hat{U}_s$  and  $\mu_s = \mu_s^* = \tau_s$ . As a result,  $(\mu_H^* - \mu_L^*) = \tau(H - L)$ . For  $\tau$  sufficiently small, this satisfies the condition for first-best allocations in Proposition 6, which verifies that the proposed offers are best responses. If the constraint  $p_{3s} \leq h_s(v_s)$  were relaxed (no such constraint was imposed in the proposition) this would still be an equilibrium.

(2) Given  $\tau(H - L) \leq X_H$ , show that no other symmetric equilibrium exists: In a symmetric equilibrium, either (a)  $(\mu_H^* - \mu_L^*) \in [-X_L, X_H]$ , (b)  $(\mu_H^* - \mu_L^*) < -X_L$ , or (c)  $(\mu_H^* - \mu_L^*) > X_H$  must hold for both firms. Given (a), the proposed equilibrium is unique up to penalty fees. A symmetric equilibrium is ruled out in case (b) by a similar argument to the proof of Proposition 4 part 3. I rule out a symmetric equilibrium in case (c) by showing that there would exist a profitable deviation:

Suppose a symmetric equilibrium satisfied  $(\mu_H^* - \mu_L^*) > X_H$ . Then by Proposition 6, IC-H binds and IC-L is slack. At any offer in firm A's mixed strategy,

$$\frac{G_H^A(U_H^A)}{g_H^A(U_H^A)} - \frac{G_L^A(U_L^A)}{g_L^A(U_L^A)} = (U_H^A - \bar{U}_H^B + \tau_H) - (U_L^A - \bar{U}_L^B + \tau_L).$$

By symmetry,  $\bar{U}_H^A = \bar{U}_H^B$  so on average it holds that  $\frac{G_H^A}{g_H^A} - \frac{G_L^A}{g_L^A} = (\tau_H - \tau_L)$ . Therefore there exists an offer in A's mixed strategy such that  $\frac{G_H^A}{g_H^A} - \frac{G_L^A}{g_L^A} \leq (\tau_H - \tau_L)$ . Consider this offer. I construct a profitable menu deviation in three steps, ignoring IC-H until the end. (i) Change  $U_s^A$  to the unconstrained optimum at current  $S_s^A$ , which increases profits. This means lowering  $U_H^A$  and raising  $U_L^A$ , and relaxing IC-L. (ii) Change  $S_L^A$  to  $S_L^{FB}$  for type L, which increases profits and does not affect IC-L. (We already have  $S_H^A = S_H^{FB}$  by  $(\mu_H^* - \mu_L^*) > X_H$ .) (iii) Change  $U_L^A$  to  $\hat{U}_L^A$ , which increases profits now that  $S_L^A = S_L^{FB}$ . (We already have  $U_H^A = \hat{U}_H^A$  from step (i).) Decreasing marginal revenue implies that the change in  $U_L^A$  to  $\hat{U}_L^A$  represents an increase in  $U_L^A$  because it follows an increase in  $S_L^A$ . Thus the change relaxes IC-L.

The new contract has strictly higher profits and still satisfies IC-L. Moreover, it has a lower  $U_H^A$  and higher  $U_L^A$  and therefore a lower value of  $\frac{G_H^A}{g_H^A} - \frac{G_L^A}{g_L^A}$ . The new contract menu offers unconstrained

optimal markups and first-best allocations for which  $\mu_s^{*A} = \frac{G_s^A(\hat{U}_s^A)}{g_s^A(\hat{U}_s^A)}$ . As a result

$$\mu_H^{*A} - \mu_L^{*A} = \frac{G_H^A(\hat{U}_H^A)}{g_H^A(\hat{U}_H^A)} - \frac{G_L^A(\hat{U}_L^A)}{g_L^A(\hat{U}_L^A)} \leq \frac{G_H^A(U_H^A)}{g_H^A(U_H^A)} - \frac{G_L^A(U_L^A)}{g_L^A(U_L^A)} \leq \tau(H - L)$$

(where  $U_s^A$  is the original utility offer, and  $\hat{U}_s^A$  is the unconstrained optimal utility offer used in the new menu) and by Proposition 6, IC-H is satisfied for sufficiently small  $\tau$ . Thus this deviation was strictly profitable for firm A, and the proposed contract menus cannot have been part of an equilibrium.

## D.10 Proof of Corollary 2

(1) Total welfare result: With bill-shock regulation, equilibrium pricing matches the attentive case, and Proposition 4 implies that allocations are inefficient in all equilibria for any  $\tau > 0$  (sufficiently small for strict full-market-coverage). In contrast, without bill-shock regulation, Proposition 7 shows that allocations are efficient for sufficiently small  $\tau > 0$ . Moreover, without bill-shock regulation, Proposition 7 shows equilibrium is symmetric so transportation costs are minimized. Thus bill-shock regulation strictly reduces welfare.

(2) Distributional result: Without bill-shock regulation, Proposition 7 implies IC-L and IC-H are slack, meaning  $\frac{\partial \Pi}{\partial U_L} = -\frac{\partial \Pi}{\partial U_H} = 0$ . With bill-shock regulation, Proposition 4 implies that in any symmetric equilibrium IC-H binds and  $\frac{\partial \Pi}{\partial U_L} = -\frac{\partial \Pi}{\partial U_H} > 0$ . Using superscript “BSR” to denote outcomes under bill-shock regulation, this implies high types win,  $U_H^{BSR} > S_H^{FB} - \tau_H = \hat{U}_H$ , but low types lose,  $U_L^{BSR} < S_L^{BSR} - \tau_L < S_L^{FB} - \tau_L = \hat{U}_L$ . Firms still split both segments equally, but now make less on high types  $S_H^{FB} - U_H^{BSR} < \tau_H$  and more on low types  $S_L^{BSR} - U_L^{BSR} > \tau_L$ . On average firms lose money. The first-order condition under bill-shock regulation ( $\frac{\partial \Pi}{\partial U_L} = -\frac{\partial \Pi}{\partial U_H} > 0$ ) and symmetry ( $G_s/g_s = \tau_s$ ) imply that

$$\frac{1}{2} (S_L^{BSR} - U_L^{BSR} - \tau_L) (1 - \beta) = -\frac{\tau_L}{\tau_H} \frac{1}{2} (S_H^{FB} - U_H^{BSR} - \tau_H) \beta < -\frac{1}{2} (S_H^{FB} - U_H^{BSR} - \tau_H) \beta.$$

The inequality shows that the profit gain on low types (LHS) is less than the profit loss on high types (RHS).

## D.11 Proof of Proposition 8

Consumer perceived and true expected gross utilities are

$$U^* = v_0 - p_0 + \int_{v_1}^1 (v - p_1) dF^*(v) + F^*(v_1) \int_{p_2}^1 (1 - F^*(v)) dv + (1 - F^*(v_1)) \int_{p_4}^1 (1 - F^*(v)) dv$$

$$U = v_0 - p_0 + \int_{v_1}^1 (v - p_1) dF(v) + F(v_1) \int_{p_2}^1 (1 - F(v)) dv + (1 - F(v_1)) \int_{p_4}^1 (1 - F(v)) dv,$$

(the analogs of equation (36)), where  $v_1 = p_1 + \int_{p_2}^{p_4} (1 - F^*(v)) dv$  (the analog of equation (5)).

Expected gross surplus is

$$S = v_0 + \int_{v_1}^1 (v - c) dF(v) + F(v_1) \int_{p_2}^1 (v - c) dF(v) + (1 - F(v_1)) \int_{p_4}^1 (v - c) dF(v).$$

The firm's problem can be written as

$$\max_{U^*, v_1, p_2, p_4} G(U^*) (S(v_1, p_2, p_4) - U^* + \Delta(v_1, p_2, p_4)),$$

where  $p_1 = v_1 - \int_{p_2}^{p_4} (1 - F^*(v)) dv$  and  $\Delta = U^* - U$  is the difference between perceived and true expected utility:

$$\begin{aligned} \Delta &= \int_{v_1}^1 (v - p_1) (f^*(v) - f(v)) dv \\ &\quad + F^*(v_1) \int_{p_2}^1 (1 - F^*(v)) dv + (1 - F^*(v_1)) \int_{p_4}^1 (1 - F^*(v)) dv \\ &\quad - F(v_1) \int_{p_2}^1 (1 - F(v)) dv - (1 - F(v_1)) \int_{p_4}^1 (1 - F(v)) dv \end{aligned}$$

First-order conditions for  $x \in \{v_1, p_2, p_4\}$  have the form:

$$\frac{d\Pi}{dx} = G(U^*) \left( \frac{\partial S}{\partial x} + \frac{\partial \Delta}{\partial x} + \frac{\partial \Delta}{\partial p_1} \frac{dp_1}{dx} \right) = 0.$$

Components of the derivatives  $d\Pi/dp_2$  and  $d\Pi/dp_4$  are  $dp_1/dp_2 = (1 - F^*(p_2))$ ,  $dp_1/dp_4 = -(1 - F^*(p_4))$ ,  $\partial\Delta/\partial p_1 = (F^*(v_1) - F(v_1))$ ,  $\partial\Delta/\partial p_2 = F(v_1)(1 - F(p_2)) - F^*(v_1)(1 - F^*(p_2))$ ,  $\partial\Delta/\partial p_4 = (1 - F(v_1))(1 - F(p_4)) - (1 - F^*(v_1))(1 - F^*(p_4))$ ,  $\partial S/\partial p_2 = -F(v_1)(p_2 - c)f(p_2)$ ,  $\partial S/\partial p_4 = -(1 - F(v_1))(p_4 - c)f(p_4)$ . Combining these pieces and canceling terms yields:

$$\frac{d\Pi}{dp_2} = G(U^*) F(v_1) (-(p_2 - c)f(p_2) + F^*(p_2) - F(p_2)),$$

$$\frac{d\Pi}{dp_4} = G(U^*) (1 - F(v_1)) (-(p_4 - c)f(p_4) + F^*(p_4) - F(p_4)).$$

Comparing the derivatives  $d\Pi/dp_2$  and  $d\Pi/dp_4$  shows that at optimal prices  $p_2 = p_4$  and hence  $v_1 = p_1$ . At  $p_2 = p_4$  (and hence  $v_1 = p_1$ ), components of the derivative  $d\Pi/dv_1$  reduce to  $dp_1/dv_1 =$

1,  $\partial\Delta/\partial v_1 = 0$ , and  $\partial S/\partial v_1 = -(v_1 - c) f(v_1)$ . Thus,

$$\left. \frac{d\Pi}{dv_1} \right|_{p_2=p_4} = G(U^*) (-(v_1 - c) f(v_1) + (F^*(v_1) - F(v_1))).$$

As a result, at optimal prices,  $v_1 = p_1 = p_2 = p_4 = p$ ,  $p$  satisfies the first-order condition

$$p = c + \frac{F^*(p) - F(p)}{f(p)},$$

and profits are given by equation (13). FOSD,  $F^*(p) \geq F(p)$ , implies  $p \geq c$ . The assumption  $F^*(c) > F(c)$  implies  $p > c$ .

Consumers are not exploited because the participation constraint ensures that perceived net utility is positive ( $U^* - x \geq 0$ ) and  $F^* \geq F$  implies that true net utility is higher ( $U \geq U^*$ ) so is also positive ( $U - x \geq 0$ ). This is easiest to see at equilibrium prices ( $p_1 = v_1 = p_4 = p_2 = p$ ) for which

$$U - U^* = 2 \int_p^1 (F^*(v) - F(v)) dv \geq 0.$$

## D.12 Proof of Proposition 9

**The firm's problem:** Perceived and true expected gross utilities are:

$$U^* = v_0 - p_0 + 2 \int_{v^*}^1 v dF^*(v) - (p_1 + p_2) (1 - F^*(v^*)) - p_3 (1 - F^*(v^*))^2, \quad (50)$$

$$U = v_0 - p_0 + 2 \int_{v^*}^1 v dF(v) - (p_1 + p_2) (1 - F(v^*)) - p_3 (1 - F(v^*))^2. \quad (51)$$

Substituting equation (15) and  $\bar{p} = v^* - p_3 (1 - F^*(v^*))$  into equation (51), yields true expected gross utility as a function of  $U^*$ ,  $v^*$ , and  $p_3$ :

$$U = U^* + 2 \int_{v^*}^1 \left( \frac{F^*(v) - F(v)}{f(v)} \right) f(v) dv - p_3 (F^*(v^*) - F(v^*))^2. \quad (52)$$

The firm's profit function in equation (16) is then obtained by substituting  $S = v_0 + \int_{v^*}^1 (v - c) dF(v)$  and equation (52) into the expression  $\Pi = G(U^*) (S - U)$ .

**The proof:** The market structure assumption (ZOOM or Hotelling duopoly with strict full-market-coverage) implies that regulation does not effect the extensive margin. The same consumers contract with the same firms before and after regulation. Thus the welfare consequences of regulation (and the remainder of the proof) depend solely on consumption distortions at the intensive margin.

Let  $v^A$  be the optimal marginal price with attentive consumers characterized in Proposition 8. Note that the proof relies on the fact that the firm's profit function in the inattentive case (equation (16)) differs from that in the attentive case (equation (13)) only by the additional term  $p_3 (F^*(v^*) - F(v^*))^2$ .

**Claim (1)** For  $\gamma, c \geq 0$  sufficiently small, it holds that:  $v^* > v^A > c$ . **Proof:** First note that given FOSS, the sign of cross partial derivative  $\partial^2 \Pi / \partial p_3 \partial v^*$  equals the sign of  $(f^*(v^*) - f(v^*))$ :

$$\frac{\partial^2 \Pi}{\partial p_3 \partial v^*} = 2(F^*(v^*) - F(v^*))(f^*(v^*) - f(v^*)).$$

Given  $F < F^*$  for all  $v \in (0, 1)$ , there is an interval  $[0, x)$  for which  $f^* > f$ . (In many natural cases  $f^*$  will cross  $f$  once from above at  $x$ ). Profits are strictly super modular in  $p_3$  and  $v^*$  ( $\partial^2 \Pi / \partial p_3 \partial v^* > 0$ ) over the interval  $(0, x)$ . Moreover,  $x$  is independent of  $\gamma$ . The limit of  $v^A$  as  $\gamma$  approaches zero is  $c$ . Therefore, if  $c < x$  then sufficiently small  $\gamma$  implies  $v^A < x$ . Given the constraint  $p_3 \leq p^{\max}$ , strict super modularity on  $(0, x)$  and  $v^A < x$  together imply  $v^* > v^A$ . This follows (from Edlin and Shannon (1998)) because the change in the firm's maximization problem from attentive to inattentive customers is identical to the change when customers are inattentive but  $p_3$  exogenously increases from zero to  $p^{\max}$ . Note that the result continues to hold with the constraint  $p_3 \leq h(v^*)$  if  $h(v^*)$  is nondecreasing. If  $h(v^*)$  is strictly increasing then the constraint simply creates an additional incentive to raise  $v^*$  when consumers are inattentive: to relax the constraint on  $p_3$ .

**Claim (2)** For  $\gamma \geq 0$  sufficiently small,  $\exists c \in (0, 1)$  such that  $v^A > v^* \geq c$ . **Proof:** Let  $\lambda(x) = (\hat{F}(x) - F(x))/f(x)$ . (Note that  $\gamma\lambda(x) = (F^*(x) - F(x))/f(x)$ .) For sufficiently small  $\gamma$ ,  $\gamma\lambda'(x) < 1$ . In this case the firm's profit function in the attentive case is strictly quasi-concave in  $v^A$  and the first-order condition which characterizes the attentive solution,  $v^A = c + \gamma\lambda(v^A)$ , has a unique solution. Also, by the implicit function theorem,  $v^A$  is a continuous increasing function of  $c$ :  $dv^A/dc = (1 - \gamma\lambda'(x))^{-1} > 0$ . Moreover it varies from  $v^A(c=0) = 0$  to  $v^A(c=1) = 1$ . The inverse is  $c(v^A) = v^A - \gamma\lambda(v^A)$ .

Define  $x = \arg \max_v \{F^*(v) - F(v)\}$  to be the set of values which maximize disagreement. Define  $x^* = \sup \{x\}$  as the largest point in the set. Note that  $x^* < 1$ . Let  $c^* = x^* - \gamma\lambda(x^*)$  be the marginal cost for which  $v^A(c^*) = x^*$ , which exists by the argument in the preceding paragraph.

The inattentive solution is the same as the attentive solution at  $c^*$ , since disagreement is already maximized at  $x^*$ . Thus inattention does not change the distortion at  $c^*$ . However, for marginal costs  $c$  in a neighborhood above  $c^*$ , where  $v^A(c)$  is slightly above  $x^*$ , the inattentive solution will be between  $x^*$  and the attentive solution. Hence  $c^* < c < x^* < v^*(c) < v^A(c)$  for  $c$  in a neighborhood

above  $c^*$ . It is clear that there is a local maximum to the inattentive problem between  $x^*$  and  $v^A(c)$ . Reducing  $v^*$  below  $v^A$  initially has a second-order negative effect on the first term in the firm's profit function but a first-order positive effect on the disagreement term. As  $v^*$  approaches  $x^*$ , the sign of the effects are unchanged but the orders are reversed. There cannot be a global maximum below  $x^*$ , as any  $v^* < x^*$  is dominated by  $x^*$  for both terms in the profit function. Given the assumption that disagreement ( $F^*(v) - F(v)$ ) has finitely many peaks (expressed in the text as the densities crossing finitely many times), I can always take  $c > c^*$  close enough to  $c^*$  such that disagreement is larger at  $v^A(c)$  than any higher  $v$ . In this case  $v^* = v^A(c)$  dominates any higher choice of  $v^*$  for both terms in the profit function ruling out global maxima above  $v^A(c)$ .

**Claim (3)** For  $p^{\max}$  sufficiently large and  $c$  near 1, it holds that  $v^* < c \leq v^A$  and inattentive overconsumption is socially worse than attentive underconsumption. **Proof:** For  $c = 1$ , it holds that  $v^A = 1$  and allocations are first best with attentive customers. However, fix any  $v^* \in (0, 1)$  and for  $p^{\max}$  sufficiently large,

$$2 \int_{v^*}^1 \left( v - 1 - \frac{F^*(v) - F(v)}{f(v)} \right) f(v) dv + p_3 (F^*(v^*) - F(v^*))^2 > 0,$$

which implies with inattentive customers there is overconsumption ( $v^* < 1$ ) and total welfare is strictly lower. By continuity, overconsumption ( $v^* < c$ ) and the social welfare ranking are the same for  $c$  in a neighborhood around  $c = 1$ .

**Claim (4)** Banning penalty fees has an identical effect to bill-shock regulation. **Proof:** The claim follows because optimal pricing to attentive consumers derived in Proposition 8 features constant marginal prices. Moreover, inattentive consumers behave the same as attentive consumers when marginal price is constant.

### D.13 Proof of Proposition 10

**(1) ZOOM result:** Given ZOOM,  $U^* = 0$  and  $G(U^*) = 1$ . **(a)** Attentive case: By Proposition 8, consumer surplus is at least zero and profits are bounded above by  $S^{FB}$ . **(b)** Inattentive case: By assumption, there exists some  $v \in (0, 1)$  for which  $F^*(v) - F(v) = k > 0$ . By equation (16), evaluated at  $p_3 = p^{\max}$ , inattentive profits are at least a constant plus  $p^{\max} k^2$ , which is larger than  $S^{FB}$  for  $p^{\max}$  sufficiently large. As gross (and hence net) consumer surplus is bounded above by  $S^{FB} - \Pi$ , this implies exploitation of inattentive consumers. Comparing the two cases gives the result.

**(2) Duopoly result:** If there is strict full-market-coverage when firms set marginal prices optimally and charge markup  $\tau$ , then this is the equilibrium. It is therefore sufficient to show

that  $\tau < (2/3) \left( v_0 + \int_c^1 (v - c) f^*(v) dv \right)$  and  $\mu = \tau$  imply  $U^* > \tau/2$  and hence strict full-market-coverage in both attentive and inattentive cases. For the attentive case: Integrating by parts, the markup in equation (13) can be rewritten as

$$\mu = v_0 - U^* + \int_p^1 (v - c) f^*(v) + (p - c) (F^*(p) - F(p)).$$

Setting the markup equal to  $\tau$ , the condition  $U^* > \tau/2$  is therefore equivalent to

$$v_0 + \int_p^1 (v - c) f^*(v) + (p - c) (F^*(p) - F(p)) > 3\tau/2,$$

for which  $\tau < (2/3) \left( v_0 + \int_c^1 (v - c) f^*(v) dv \right)$  is a sufficient condition because  $F^*(p) \geq F(p)$  by FOSD and  $p \geq c$  at the optimal  $p$  characterized in Proposition 8. Comparing equations (13) and (16) shows that  $U^*$  is higher in the inattentive case for the same markup, so strict full-market-coverage also holds.

#### D.14 Proof of Proposition 11

First-best gross surplus is  $v_0$  when the add-on has no social value. An upper bound on gross (and hence net) consumer surplus is therefore  $S^{FB} - \Pi = v_0 - \Pi$ . Profits above  $v_0$  therefore correspond to consumer exploitation.

Consider the case without bill-shock regulation. Given  $\tau > 0$ , fix some  $U^*$  such that  $G(U^*) > 0$ . By assumption, there exists some  $v \in (0, 1)$  for which  $F^*(v) - F(v) = k > 0$ . By equation (16), evaluated at  $p_3 = p^{\max}$ , inattentive profits are at least a constant plus  $p^{\max} k^2 G(U^*)$ , which is larger than  $v_0$  for  $p^{\max}$  sufficiently large. Thus firms will find it profitable to sell add-ons (rather than only the base good) and consumers will be exploited. In contrast, Proposition 8 implies that there will be no operating add-on market or exploitation given bill-shock regulation.

#### D.15 Proof of Lemma 1

Follows from Lemmas 2 and 3 which are stated in Appendix B and proved in Online Appendix E.

#### D.16 Proof of Proposition 14

(1) Since the firm's objective and constraints are linear in prices, it is possible to solve the firm's problem as follows. Begin by setting the fixed fee sufficiently high that perceived expected utility is too low and setting marginal charges as high as possible under incentive constraints:  $p_1 = p_2 = p_4 = 1$ . This contract is a two-part tariff with no penalty fees ( $p_3 = 0$ ) which satisfies IC

and NFL constraints but offers too little perceived expected utility. Thus the firm must reduce prices. For  $U^* = v_0 + 2\alpha'$ , there is only one feasible option, which is to set all prices to zero:  $p_0 = p_1 = p_2 = p_4 = 0$ . (Higher perceived utility offers are not feasible under NFL.) However, for lower perceived utility offers the firm has discretion over which prices to reduce and by how much.

The firm would like to raise the perceived expected utility as cheaply as possible. That is, the firm would like to begin by reducing prices which will reduce the expected markup least for a given increase in perceived expected utility. If consumers were unbiased, it would not matter which prices were reduced. In all cases a price reduction that raised perceived expected utility by a dollar would lower the expected markup by a dollar as well. However, because consumers underestimate their demand this is no longer true. Instead consumers are more sensitive to some prices than others.

I calculate the *bang-for-the-buck* of independently decreasing price  $p_n$  as  $\gamma_n \equiv -\frac{dU}{dp_n} / \frac{dU^*}{dp_n}$ , which measures the decrease in markup for a one dollar increase in  $U^*$  due to decreasing  $p_n$ . Computing derivatives from equations (28)-(29), these bang-for-the-buck coefficients are:  $\gamma_0 = -1$ ,  $\gamma_1 = -\alpha/\alpha'$ ,  $\gamma_2 = -\alpha(1-\alpha)/(\alpha'(1-\alpha'))$ , and  $\gamma_4 = -\alpha^2/\alpha'^2$ . These can be ranked as follows:

$$\gamma_0, \gamma_2 > \gamma_1 > \gamma_4.$$

Note that the ranking of  $\gamma_0$  and  $\gamma_2$  depends on whether  $(\alpha + \alpha')$  is less than or greater than one. Reducing the price with a higher bang for the buck (e.g.  $p_1$  has a higher bang-for-the-buck than  $p_4$ ) will result in a smaller cut in markup for the same increase in perceived expected-utility  $U^*$ . Thus, absent a binding constraint, the firm will always reduce the price with the highest bang-for-the-buck first.

The ranking  $\gamma_0 > \gamma_1 > \gamma_4$  is relatively straight forward. There is no confusion about the fixed fee - consumers and the firm both agree that a dollar reduction in  $p_0$  is a transfer of a dollar from firm to consumers. However, consumers underestimate their likelihood of purchase in the first period by a factor  $\alpha'/\alpha$ . Thus they undervalue reductions in  $p_1$  by the same factor, and the expected markup falls at rate  $\alpha/\alpha'$  faster than perceived utility rises. The problem is compounded for  $p_4$ , since consumers underestimate the chance of making two purchases by  $(\alpha'/\alpha)^2$ . On the other hand, the problem is mitigated for  $p_2$ . While consumers underestimate the chance of demanding a unit in the second period, they overestimate the chance that  $p_2$  is the relevant second-period price because they underestimate the likelihood of an initial purchase triggering a penalty fee. Hence  $\gamma_2 > \gamma_1 > \gamma_4$ .

An important point is that  $p_2$  has a higher bang-for-the-buck than  $p_1$  and  $p_4$  (and also than  $p_0$  if  $\alpha + \alpha' > 1$ ). However it cannot be independently reduced before  $p_1$  and  $p_4$  without violating the

incentive constraint  $v_1^* \leq 1$ . Either  $p_1$  or  $p_4$  must be reduced at the same time to maintain incentive compatibility. Lowering  $p_2$  maintains  $v_1^* = 1$  if  $p_4$  is reduced equally or  $p_1$  is reduced a proportion  $\alpha'$  as much (or a convex combination of such reductions in both  $p_1$  and  $p_4$ ). Reducing  $p_1$  in tandem with  $p_2$  has the highest bang-for-the-buck of these options because  $p_1$  has higher bang-for-the-buck than  $p_4$  and need be reduced only at rate  $\alpha' < 1$ . The bang-for-the-buck of simultaneously reducing  $p_2$  and  $p_1$  while maintaining  $v_1^* = 1$  is

$$\gamma_{12} = -\frac{dU/dp_2 + \alpha' dU/dp_1}{dU^*/dp_2 + \alpha' dU^*/dp_1} = -(\alpha/\alpha') (1 - (\alpha - \alpha')).$$

Now the relevant bang-for-the-buck coefficients can be completely ranked:

$$\gamma_0 > \gamma_{12} > \gamma_1 > \gamma_4. \quad (53)$$

This leads to the following conclusion. It is optimal for the firm to reduce prices in the following order (stopping as soon as the prescribed  $U^*$  is achieved): (1) First reduce the fixed fee  $p_0$  until  $U^*$  is achieved or  $p_0 = 0$ . (2) Second, reduce  $p_2$  and  $p_1$  simultaneously such that  $p_1 = 1 - \alpha'(1 - p_2)$  until  $U^*$  is achieved or  $p_2 = 0$ . (3) Third, reduce  $p_1$  until  $U^*$  is achieved or  $p_1 = 0$  (at which point the NFL constraint  $p_0 + p_1 \geq 0$  binds because  $p_0 = 0$ ). (4) Finally the firm should reduce  $p_4$  until  $U^*$  is achieved. This procedure stops at the optimal contract conditional on  $U^*$  and leads to the four qualitative pricing regions in the proposition as a function of  $U^*$ .

(2) The boundary values of  $U^*$  between qualitative pricing regions are found by evaluating equation (29) at the boundary prices. For instance, between region (2) and (3) the boundary prices are  $p_0 = p_2 = 0$ ,  $p_4 = 1$ ,  $p_3 = 1 - \alpha'$ . At these prices,  $U^* = v_0 + 2\alpha'(1 - \alpha')$ . Within a region, all prices but one are at a boundary. For instance in region (2) all other prices are a function of  $p_2$ . Plugging these prices into equation (29) and inverting for  $p_2$  yields the expression  $p_2 = 1 - (U^* - v_0)/\alpha'$ . In this way, the details of all five pricing regions presented in the proposition are derived.

(3) The derived prices in the proposition can be plugged into equation (28) to find  $U$ . Substituting these expressions along with  $S^{FB} = v_0 + 2\alpha(1 - c)$  provide the markups ( $S^{FB} - U$ ) given in the second table of the proposition and differentiation gives  $-d\mu/dU^*$ .

(4) The nonexploitation result  $U \geq 0$  follows by brute force by calculating  $U$  in each region and comparing to zero. However it follows more directly by noting first that  $U$  is increasing in  $U^*$  and second that  $U = U^*$  in region 1 when  $U^* = 0$ .

## D.17 Proof of Proposition 12

By assuming  $\tau$  sufficiently small or  $v_0$  sufficiently large for strict full-market-coverage, I ensure that firm A's best response utility offer  $U_s^A$  is always within an open interval for which residual demand is  $G(U^A, U^B) = \frac{1}{2\tau}(U^A - U^B + \tau) > 0$ . Profits are

$$\Pi^A = G(U^A, U^B) \mu(U^A)$$

where  $\mu(U^A)$  is the markup derived in Proposition 14. Note that firm A's profits are linear in firm B's offer  $U^B$  and hence firm A's expected profits and best response depend only on firm B's expected offer  $\bar{U}^B$ . Firm A's profit function is strictly concave (with kinks at the boundaries between pricing regions) in  $U^A$ . Hence equilibrium is in pure strategies and firm A's best response is a continuous function of  $U^B$ . Strict concavity holds because, away from the kinks,  $d^2\Pi^A/dU^{A2} = g(U^A, U^B) d\mu/dU^A < 0$ , and at the kink  $d\Pi^A/dU^A$  decreases. This follows since

$$\frac{d\Pi^A}{dU^A} = g(U^A, U^B) \mu(U^A) + G(U^A, U^B) \frac{d\mu}{dU^A},$$

and while  $G(U^A, U^B)$  and  $\mu(U^A)$  are continuous and nonnegative,  $d\mu/dU^A$  decreases at kink points as shown in equation (53). The slope  $d\mu/dU^A$  decreases precisely because firms order price cuts from highest bang-for-the buck to lowest.

The optimal  $U^A$  either solves the first-order condition  $\mu(U^A) = -(U^A - U^B + \tau) d\mu/dU^A$ , or is located at a kink at the boundary between pricing regions. In the attentive case, there are seven sub-cases corresponding to the four pricing regions and three kinks. Substituting  $\mu(U^A)$  and  $d\mu/dU^A$  from the markup table in Proposition 14, the first-order conditions for the four pricing regions are:

Table 1

Region	First Order Condition
1	$U^A = \frac{1}{2}(S^{FB} + U^B - \tau)$
2	$U^A = \frac{1}{2}(U^B - \tau + v_0) + \alpha'(1 - c)/(1 - \alpha + \alpha')$
3	$U^A = \frac{1}{2}(U^B - \tau + v_0) + \frac{1}{2}\alpha'(\alpha - \alpha') + \alpha'(1 - c)$
4	$U^A = \frac{1}{2}(U^B - \tau + v_0) + (\alpha'/\alpha)(\alpha - \alpha') + (\alpha'/\alpha)\alpha'(1 - c)$

Denote the boundaries between pricing regions as 1/2, 2/3, and 3/4. At these boundaries, offered utility is: Boundary 1/2,  $U^A = v_0$ ; Boundary 2/3,  $U^A = v_0 + 2\alpha'(1 - \alpha')$ ; Boundary 3/4,  $U^A = v_0 + \alpha'(2 - \alpha')$ .

By inspection, the best response by A has slope  $dU^A/dU^B$  of either zero (at a boundary point between pricing regions) or  $1/2$  (within a pricing region where a first-order condition holds). Since  $dU^A/dU^B \in [0, 1)$  (and as already noted  $U^A(U^B)$  is continuous), there is a unique equilibrium, which is in symmetric pure strategies. (This is true for both attentive and inattentive cases.)

Each first-order condition in Table 1 has a corresponding symmetric solution for the offered  $U^*$  and corresponding markup. Each is relevant for the range of transportation costs for which the solution  $U^*$  actually lies within the relevant pricing region. These are given in Table 2 below:

Table 2

Region	Symmetric Solution, $U^* =$	Relevant Range
1	$v_0 - \tau + 2\alpha(1 - c)$	$\tau > 2\alpha(1 - c)$
2	$v_0 - \tau + 2\alpha'(1 - c) / (1 - \alpha + \alpha')$	$\frac{2\alpha'(1-c)}{(1-\alpha+\alpha')} - \alpha' \leq \tau \leq \frac{2\alpha'(1-c)}{(1-\alpha+\alpha')}$
3	$v_0 - \tau + \alpha'(\alpha - \alpha') + 2\alpha'(1 - c)$	$\alpha'(\alpha - 2c) \leq \tau \leq \alpha'(\alpha - 2c) + \alpha'(1 - \alpha')$
4	$v_0 - \tau + (\alpha'/\alpha)(\alpha - c\alpha')$	$0 \leq \tau \leq \alpha'^2 - 2\alpha'^2c/\alpha$

Plugging the values of  $U^*$  derived in Table 2 into the markup table in Proposition 14 yields markups as a function of transportation cost as described in the proposition. Markups are constant between pricing regions. Denote these boundary regions  $1/2$ ,  $2/3$ , and  $3/4$  respectively. In these regions offered utilities and markups (derived from substituting offered utilities into the markup table in Proposition 14) are given by Table 3:

Table 3

Region	$U^* =$	$\mu =$	Relevant Range
1/2	$v_0$	$2\alpha(1 - c)$	$\frac{2\alpha'(1-c)}{(1-\alpha+\alpha')} \leq \tau \leq 2\alpha(1 - c)$
2/3	$v_0 + 2\alpha'(1 - \alpha')$	$2\alpha(1 - c) - \alpha(1 - \alpha + \alpha')$	$\alpha'(\alpha - 2c) + \alpha'(1 - \alpha') \leq \tau \leq \frac{2\alpha'(1-c)}{(1-\alpha+\alpha')} - \alpha'$
3/4	$v_0 + \alpha'(2 - \alpha')$	$2\alpha(1 - c) - \alpha(2 - \alpha)$	$\alpha'^2 - 2\alpha'^2c/\alpha \leq \tau \leq \alpha'(\alpha - 2c)$

(54)

## D.18 Proof Proposition 15

I begin by solving the firm's problem assuming that the firm chooses to keep penalty fees a surprise. The final step is to show that this is optimal.

NFL says prices can be no lower than  $p_0 = p_1 = p_2 = p_3 = 0$ , and hence offered perceived utility  $U^*$  can be no higher than  $v_0 + 2\alpha'$ . Optimal pricing need only be characterized for  $U^* \in [0, v_0 + 2\alpha']$ . By Lemma 3, the firm will induce the efficient allocation,  $b_0 = 0$ ,  $b_1 = 1$ . As usual, profits are

$\Pi = G(U^*) \mu(U^*)$ . In this case, markups and fixed fees are:

$$\mu(U^*) = -U^* + v_0 + 2((\alpha - \alpha')\bar{p} + \alpha' - \alpha c) + (\alpha^2 - \alpha'^2)p_3,$$

$$p_0 = -U^* + v_0 + 2\alpha'(1 - \bar{p}) - \alpha'^2 p_3.$$

Incentive compatibility requires that the expected marginal price be between zero and one:  $0 \leq \bar{p} + \alpha' p_3 \leq 1$ , or alternatively that the penalty fee be between:  $-\bar{p}/\alpha' \leq p_3 \leq (1 - \bar{p})/\alpha'$ . The three NFL constraints, (a)  $p_0 \geq 0$ , (b)  $p_0 + \bar{p} \geq 0$ , and (c)  $p_0 + 2\bar{p} + p_3 \geq 0$  are:

$$\frac{U^* - v_0 - 2\alpha'(1 - \bar{p}) - 2\bar{p}}{1 - \alpha'^2} \leq p_3 \leq \frac{2\alpha'(1 - \bar{p}) + v_0 - U^*}{\alpha'^2} + \min\left\{0, \frac{\bar{p}}{\alpha'^2}\right\}.$$

**Deriving optimal prices and markups:** There are two cases to consider.

**Case I,**  $U^* < \alpha' + v_0$ : Impose the NFL upper bound  $p_3 \leq (2\alpha'(1 - \bar{p}) + \bar{p} + v_0 - U^*)/\alpha'^2$  and the IC upper bound  $p_3 \leq (1 - \bar{p})/\alpha'$ , but relax the other three constraints. At  $\bar{p} = \bar{p}^* = -\frac{\alpha' - (U^* - v_0)}{1 - \alpha'}$ , both constraints are the same and the optimal penalty fee would be the upper bound  $p_3 = (1 - \bar{p}^*)/\alpha' = \frac{1 - (U^* - v_0)}{(1 - \alpha')\alpha'}$ . For  $\bar{p} > \bar{p}^*$ , the IC upper bound is tighter and the optimal penalty is  $p_3 = (1 - \bar{p})/\alpha'$ . In this case, profits are,

$$\Pi = G(U^*) \left( -U^* + v_0 + 2((\alpha - \alpha')\bar{p} + \alpha' - \alpha c) + (\alpha^2 - (\alpha')^2)(1 - \bar{p})/\alpha' \right),$$

and  $d\Pi/d\bar{p} = -G(U^*)(\alpha - \alpha')^2/\alpha' < 0$ , so  $\bar{p}^*$  dominates any  $\bar{p} > \bar{p}^*$ . For  $\bar{p} < \bar{p}^*$ , the NFL upper bound is binding, and as I show below in Case II,  $d\Pi/d\bar{p} > 0$  so  $\bar{p}^*$  dominates any  $\bar{p} < \bar{p}^*$ . Thus the optimal prices are those given by equation (32) in the proposition. The assumption  $U^* < v_0 + \alpha'$  ensures  $\bar{p}^*$  is negative, and hence the alternative NFL upper bound is satisfied. Substituting for prices, the NFL lower bound reduces to  $U^* \leq v_0 + \alpha' + 1$ , which is satisfied given  $U^* < v_0 + \alpha'$ . The IC lower bound is always satisfied when the upper bound is satisfied with equality. Substituting for prices gives the markup in equation (33).

**Case II,**  $U^* \in [v_0 + \alpha', v_0 + 2\alpha']$ : Relax the incentive constraint and the NFL lower bound on the penalty fee. As profits are increasing in both  $\bar{p}$  and  $p_3$ , for any fixed  $\bar{p}$ , the penalty fee  $p_3$  will be set at the NFL upper bound. If  $\bar{p} \geq 0$ , this implies  $p_3 = (2\alpha'(1 - \bar{p}) + v_0 - U^*)/\alpha'^2$ ,

$$\Pi = G(U^*) \left( 2\alpha(\bar{p} - c) + 2\alpha^2(1 - \bar{p})/\alpha' - (\alpha/\alpha')^2(U^* - v_0) \right),$$

and  $d\Pi/d\bar{p} = -2\alpha(\alpha - \alpha')/\alpha' < 0$ . Thus  $\bar{p} = 0$  dominates any  $\bar{p} > 0$ . If  $\bar{p} \leq 0$ , this implies

$$p_3 = (2\alpha' (1 - \bar{p}) + \bar{p} + v_0 - U^*) / \alpha'^2,$$

$$\Pi = G(U^*) \left( 2\alpha (\bar{p} - c) - \bar{p} + (2\alpha' (1 - \bar{p}) + \bar{p} - (U^* - v_0)) (\alpha/\alpha')^2 \right),$$

and  $d\Pi/d\bar{p} = (\alpha^2 (1 - 2\alpha') - \alpha'^2 (1 - 2\alpha)) / \alpha'^2 > 0$ . Thus  $\bar{p} = 0$  dominates any  $\bar{p} < 0$ . As a result, optimal prices are  $\bar{p} = p_0 = 0$  and  $p_3 = (2\alpha' + v_0 - U^*) / \alpha'^2$ . Substituting for prices, the IC constraint is equivalent to the assumption  $U^* \in [v_0 + \alpha', v_0 + 2\alpha']$  and hence is satisfied. Similarly, the NFL lower bound is equivalent to  $U^* \leq v_0 + 2\alpha'$  and so is satisfied. Substituting for prices gives the markup in equation (35).

**Nondisclosure result:** Comparing the markups derived above to those in Proposition 14 shows that markups are weakly higher in the inattentive case for all  $U^*$ . For  $U^* \in [v_0 + \alpha' (2 - \alpha'), v_0 + 2\alpha']$  the contracts and markups are identical. For  $U^* \in [0, v_0 + \alpha' (2 - \alpha'))$ , the markup up is strictly higher in the inattentive case. To show the latter, denote the inattentive markup by  $\mu^I$  and the attentive markup by  $\mu^A$ . For  $U^* \in [0, v_0]$ ,

$$\mu^I - \mu^A = \frac{(\alpha - \alpha')^2}{\alpha' (1 - \alpha')} (1 - (U^* - v_0)),$$

which is strictly positive because  $(U^* - v_0) < 0$ . For  $U^* \in [v_0, v_0 + \alpha']$ ,

$$\mu^I - \mu^A = \frac{(\alpha - \alpha')^2}{\alpha' (1 - \alpha')} (1 - (U^* - v_0)) + \frac{1}{\alpha'} (1 - \alpha) (\alpha - \alpha') (U^* - v_0),$$

which is strictly positive because  $(U^* - v_0) \in [0, 1)$ . For  $U^* \in [v_0 + \alpha', v_0 + \alpha' (2 - \alpha'))$ ,

$$\mu^I - \mu^A = \alpha (\alpha - \alpha') (\alpha' (2 - \alpha') + v_0 - U^*) / \alpha'^2,$$

which is strictly positive because  $U^* < v_0 + \alpha' (2 - \alpha')$ .

This comparison shows that firms always weakly prefer to keep penalty fees a surprise and do so strictly for  $U^* \in [0, v_0 + \alpha' (2 - \alpha'))$ .

## D.19 Proof of Proposition 13

The argument closely follows that of Proposition 12. The difference is that there are two pricing regions rather than four, and the associated first-order conditions are: (1) For  $U^A < v_0 + \alpha'$ ,

$$U^A = \frac{1}{2} (U^B + v_0 - \tau) + \frac{2\alpha (1 - c) + Y}{2(1 + Y)}.$$

(2) For  $U^A > v_0 + \alpha'$ ,

$$U^A = \frac{1}{2} (U^B + v_0 - \tau) + \alpha' (1 - c\alpha'/\alpha).$$

Otherwise profits are maximized at a kink for  $U^A = v_0 + \alpha'$ .

The corresponding symmetric solutions to the first-order conditions are:

Table 4

Region	Symmetric Solution, $U^* =$	$\mu =$	Relevant Range
1	$v_0 - \tau + (2\alpha(1 - c) + Y) / (1 + Y)$	$(1 + Y)\tau$	$\tau > (2\alpha(1 - c) + Y) / (1 + Y) - \alpha'$
2	$v_0 - \tau + 2\alpha'(1 - c\alpha'/\alpha)$	$(\alpha/\alpha')^2\tau$	$0 \leq \tau \leq (\alpha'/\alpha)(\alpha - 2c\alpha')$

(55)

For intermediate values of  $\tau \in [(\alpha'/\alpha)(\alpha - 2c\alpha'), (2\alpha(1 - c) + Y) / (1 + Y) - \alpha']$ ,

$$U^* = v_0 + \alpha' \quad (56)$$

and

$$\mu = (\alpha/\alpha')(\alpha - 2c\alpha'). \quad (57)$$

Markups are derived by plugging the values of  $U^*$  for the relevant regions into equations (33)-(35).

The preference for surprise penalty fees follows from Proposition 15. The condition for a strict preference,  $U^* < v_0 + \alpha'(2 - \alpha')$ , corresponds to region 2, and therefore (substituting  $U^*$  from region 2) that  $\tau > (\alpha')^2(\alpha - 2c)/\alpha$ .

## D.20 Proof of Corollary 4

**(1) Full-market-coverage result:** A sufficient condition for strict full-market-coverage is that the equilibrium offered utilities  $U^*$  characterized in the proofs of Propositions 12 and 13 under the assumption of full-market-coverage satisfy  $U^* > \tau/2$ . Inspection of Tables 2, 3, and 4 and equation (56) show that for all levels of  $\tau$ ,  $U^* \geq v_0 - \tau$ . Note that  $v_0 - \tau > \tau/2$  is equivalent to  $\tau < (2/3)v_0$  which is true by assumption. Therefore there is strict full-market-coverage in equilibrium.

**(2) Sufficiently large bias result** ( $\alpha'/\alpha$  is sufficiently small): First, consider the inattentive case. By Proposition 13, the minimum equilibrium markup with full-market-coverage is  $(1 + Y)\tau$ . Taking  $\alpha'/\alpha$  small implies taking  $\alpha'$  to zero. Since  $\lim_{\alpha' \rightarrow 0} Y = \infty$ , taking  $\alpha'$  to zero holding  $\tau > 0$  fixed implies that the lower bound on markup,  $(1 + Y)\tau$ , tends to infinity. Since all served consumers are exploited whenever the markup exceeds first best surplus of  $S^{FB} = v_0 + 2\alpha(1 - c)$ , this implies all consumers are exploited for sufficiently large bias. This exploitation must be eliminated by bill-shock regulation because Proposition 14 guarantees that attentive consumers are not

exploited. Since there is still full-market-coverage this must be due to a strict reduction in markup.

**(3) Severe and mild bias results:** Denote the inattentive markup for competitive region 1/2 given by equation (57) as  $\mu_{1/2}^I$ , and extend the notation (superscript “I” for inattentive, or “A” for attentive and subscript for competitive region) for other markups as well. I now rank pairs of the five markups  $\mu_{1/2}^I$ ,  $\mu_{2/3}^A$ ,  $\mu_{1/2}^A$ ,  $\mu_1^I$ , and  $\mu_2^A$  which are given by Table 3, equation (57), and Propositions 12 and 13. First,  $\mu_{1/2}^I > \mu_{2/3}^A$ , because

$$\mu_{1/2}^I - \mu_{2/3}^A = (\alpha/\alpha') (1 - \alpha') (\alpha - \alpha') > 0.$$

Second,  $\mu_{1/2}^I > \mu_{1/2}^A$  given  $\alpha'/\alpha < 1/2$  but  $\mu_{1/2}^I < \mu_{1/2}^A$  given  $\alpha'/\alpha > 1/2$ . This follows because

$$\mu_{1/2}^I - \mu_{1/2}^A = (\alpha/\alpha') (\alpha - 2\alpha').$$

Third,  $\mu_1^I > \mu_2^A$  given  $\alpha'/\alpha < (2\alpha - 1)/\alpha^2$  but  $\mu_1^I < \mu_2^A$  given  $\alpha'/\alpha > (2\alpha - 1)/\alpha^2$ . This follows because

$$\mu_1^I - \mu_2^A = (1 + Y) \tau - (\alpha/\alpha') (1 - \alpha + \alpha') \tau,$$

which can be simplified to

$$\mu_1^I - \mu_2^A = \frac{\alpha - \alpha'}{\alpha' (1 - \alpha')} (2\alpha - 1 - \alpha\alpha') \tau.$$

Putting the last two markup rankings together implies that, given severe bias,  $\mu_{1/2}^I > \mu_{1/2}^A$  or  $\mu_1^I > \mu_2^A$ . Similarly, given mild bias,  $\mu_{1/2}^I < \mu_{1/2}^A$  and  $\mu_1^I < \mu_2^A$ .

Comparing equilibrium markups derived in Propositions 12 and 13, for attentive and inattentive cases respectively, shows that markups coincide for  $\tau \leq (\alpha'^2/\alpha) (\alpha - 2c)$ . (Inattentive region 2 extends to higher  $\tau$  than does attentive region 4 since  $(\alpha'^2/\alpha) (\alpha - 2c) < (\alpha'/\alpha) (\alpha - 2c\alpha')$ .) For  $\tau$  slightly above this range, inattentive markups are strictly higher.

Given  $\mu_{1/2}^I > \mu_{2/3}^A$  and  $\mu_1^I > \mu_2^A$ , severe bias ( $\mu_{1/2}^I > \mu_{1/2}^A$  or  $\mu_1^I > \mu_2^A$ ) is sufficient to ensure that inattentive and attentive markups never cross again (inattentive markup remains strictly higher) for  $\tau > (\alpha'^2/\alpha) (\alpha - 2c)$ . (Note that for  $c \geq \alpha/2$ ,  $(\alpha'^2/\alpha) (\alpha - 2c) \leq 0$ , so  $\tau > (\alpha'^2/\alpha) (\alpha - 2c)$  is implied by  $\tau > 0$ .) For mild bias, the two markups will intersect at some  $\tau > (\alpha'^2/\alpha) (\alpha - 2c)$ . The ranking  $\mu_{2/3}^A < \mu_{1/2}^I < \mu_{1/2}^A$  implies that the first intersection will be at  $\tau_1$ , where  $\mu_{1/2}^I = \mu_{1/2}^A$ . The ranking  $\mu_2^A > \mu_1^I > \mu_1^A$  implies that the second intersection will be at  $\tau_2$ , where  $\mu_1^I = \mu_1^A$ . Taking the expressions for these markups from Table 3, equation (57), and Propositions 12 and 13

and solving for  $\tau_1$  and  $\tau_2$  yields:

$$\{\tau_1, \tau_2\} = \left\{ \frac{1 - \alpha + \alpha'}{\alpha - 2c\alpha'}, \frac{2\alpha(1 - c)}{1 + Y} \right\}.$$

For mild bias, the attentive markup is strictly higher in the interval  $(\tau_1, \tau_2)$ .

## E Additional Proofs

### E.1 Proof of Lemma 2

Given Bernoulli taste shocks, an attentive consumer's strategy is described by the tuple  $\{b_0, b_1, b_{10}, b_{11}, b_{00}, b_{01}\}$ .

The pair  $\{b_0, b_1\}$  describe the probabilities of first-period purchase conditional on realizing  $v_1 = 0$  or  $v_1 = 1$  respectively. Following a first-period purchase, the pair  $\{b_{10}, b_{11}\}$  describe the probabilities of second-period purchase conditional on a realized value of  $v_2 = 0$  or  $v_2 = 1$  respectively. The pair  $\{b_{00}, b_{01}\}$  describe the corresponding second-period purchase probabilities conditional on no purchase in period 1. Let  $p_4 = p_2 + p_3$ . Incentive compatibility constraints are straight forward in the second period. For instance,  $b_{01} = 0$  requires  $p_4 \geq 1$ ,  $b_{01} \in (0, 1)$  requires  $p_4 = 1$ , and  $b_{01} = 1$  requires  $p_4 \leq 1$ . In the first period, purchases are made only if  $v_1 \geq v_1^*$  where

$$v_1^* = p_1 + (1 - \beta) (\max\{0, -p_2\} - \max\{0, -p_4\}) + \beta (\max\{0, 1 - p_2\} - \max\{0, 1 - p_4\}). \quad (58)$$

The expression simplifies substantially if  $p_2, p_4 \in [0, 1]$  (as is shown to be optimal below) in which case

$$v_1^* = p_1 + \alpha' (p_4 - p_2). \quad (59)$$

Then surplus and an attentive consumer's true and perceived expected-utilities are given by equations (60)-(62) as a function of prices and the strategy:

$$\begin{aligned} S &= v_0 + (1 - \alpha) b_0 (-c) + \alpha b_1 (1 - c) \\ &\quad + (1 - (1 - \alpha) b_0 - \alpha b_1) ((1 - \alpha) b_{00} (-c) + \alpha b_{01} (1 - c)) \\ &\quad + ((1 - \alpha) b_0 + \alpha b_1) ((1 - \alpha) b_{10} (-c) + \alpha b_{11} (1 - c)) \end{aligned} \quad (60)$$

$$\begin{aligned} U &= v_0 - p_0 + (1 - \alpha) b_0 (-p_1) + \alpha b_1 (1 - p_1) \\ &\quad + (1 - (1 - \alpha) b_0 - \alpha b_1) ((1 - \alpha) b_{00} (-p_2) + \alpha b_{01} (1 - p_2)) \\ &\quad + ((1 - \alpha) b_0 + \alpha b_1) ((1 - \alpha) b_{10} (-p_4) + \alpha b_{11} (1 - p_4)) \end{aligned} \quad (61)$$

$$\begin{aligned}
U^* &= v_0 - p_0 + (1 - \alpha') b_0 (-p_1) + \alpha' b_1 (1 - p_1) \\
&+ (1 - (1 - \alpha') b_0 - \alpha' b_1) ((1 - \alpha') b_{00} (-p_2) + \alpha' b_{01} (1 - p_2)) \\
&+ ((1 - \alpha') b_0 + \alpha' b_1) ((1 - \alpha') b_{10} (-p_4) + \alpha' b_{11} (1 - p_4))
\end{aligned} \tag{62}$$

Firm profits are  $\Pi = G(U^*)(S - U^* + \Delta)$ , where the perception gap  $\Delta = U^* - U$  is the difference between perceived and true expected utility.

Consider the firm maximizing profits by choosing the prices, and in cases of consumer indifference, the allocation.

**Claim (1)** It is optimal for the firm to induce efficient allocations in the second period and charge prices  $p_2, p_4 \in [0, 1]$ .

**Claim (1a)** It is optimal for the firm to induce  $b_{01} = b_{11} = 1$  and charge  $p_2, p_4 \leq 1$ . **Proof:** Suppose  $b_{11} < 1$ . Then incentive compatibility implies  $p_4 \geq 1$ . Consider first reducing  $p_4$  to  $p'_4 = 1$  if it happens to be higher and second changing the allocation to  $b'_{11} = 1$  while keeping all else fixed. If  $p_4 > 1$  then  $b_{10} = b_{11} = 0$  and the initial reduction in  $p_4$  maintains incentive compatibility of the allocation and satisfies NFL constraints without changing payoffs. Increasing  $b_{11}$  to 1 changes neither  $U^*$  nor  $U$  since consumers are indifferent to purchasing at a price equal to value. However it does increase surplus and does so strictly if there is positive probability of first-period purchase. Thus this is a profitable deviation. A similar argument applies to  $b_{01}$  and  $p_2$ .

**Claim (1b)** It is optimal for the firm to induce  $b_{10} = 0$  and charge  $p_4 \geq 0$ . **Proof:** Suppose  $b_{10} > 0$ . Then incentive compatibility requires  $b_{11} = 1$  and  $p_4 \leq 0$ . Consider the following changes: First, if  $p_4 < 0$  then increase  $p_4$  to  $p'_4 = 0$  and reduce  $p_1$  by the same amount. Second reduce  $b_{10}$  to 0.

The joint price change keeps  $v_1^*$  constant, and hence maintains incentive compatibility of the first-period allocation (see equation (58)). Moreover, if  $p_4 < 0$  then  $b_{10} = 1$  and the joint price change does not affect payoffs because the consumer pays  $p_4$  if and only if she pays  $p_1$  and the sum is constant. The NFL constraints involving  $p_1$  and  $p_4$  are still satisfied. First  $p_0 + p_1 + p_4 \geq 0$  is still satisfied because  $p_1 + p_4$  was held constant. Second, since  $p'_4 = 0$ , this implies the other constraint  $p_0 + p_1 \geq 0$  holds. Thus the joint price change maintains NFL constraints, incentive compatibility, and does not affect payoffs.

The reduction of  $b_{10}$  to 0 increases surplus and does so strictly if first-period purchases have positive probability. Moreover, it does not affect perceived consumer payoff  $U^*$  or the perception gap  $\Delta$  since  $p'_4 = 0$ . (Notice  $b_{10}$  only enters  $U$  and  $U^*$  in the product  $b_{10}p_4$ ). Thus profits increase by the same amount as surplus and this is a profitable deviation.

**Claim (1c)** It is optimal for the firm to induce  $b_{00} = 0$  and charge  $p_2 \geq 0$ . **Proof:** Suppose

$b_{00} > 0$ . Then incentive compatibility requires  $b_{01} = 1$  and  $p_2 \leq 0$ . Consider the following changes: First, if  $p_2 < 0$  then increase  $p_2$  to  $p'_2 = 0$  and increase  $p_1$  and reduce  $p_0$  by the same amount:  $p'_1 = p_1 - p_2$ ,  $p'_0 = p_0 + p_2$ . Second reduce  $b_{00}$  to 0.

The joint price change keeps  $v_1^*$  constant, and hence maintains incentive compatibility of the first-period allocation (see equation (58)). Moreover, if  $p_2 < 0$  then  $b_{00} = b_{01} = 1$  and the joint price change does not affect payoffs.<sup>39</sup> The NFL constraints involving  $p_0$ ,  $p_1$ , and  $p_2$  are all still satisfied. First  $p_0 + p_2 \geq 0$  is satisfied because the sum  $p_0 + p_2$  is held constant. Second,  $p_0 \geq 0$  is implied by  $p_0 + p_2 \geq 0$  since  $p'_2 = 0$ . Third,  $p_0 + p_1 \geq 0$  and  $p_0 + p_1 + p_4 \geq 0$  are satisfied because the sum  $p_0 + p_1$  is held constant. Thus the joint price change maintains incentive compatibility of the allocation, satisfies NFL constraints, and does not affect payoffs.

The reduction of  $b_{00}$  to 0 increases surplus and does so strictly if first-period purchases have probability less than 1. Moreover, it does not affect  $U^*$  or the perception gap because  $p'_2 = 0$ . Thus profits increase by the same amount as surplus and this is a profitable deviation.

**Claim (2)** It is optimal for the firm to charge a nonnegative penalty fee:  $p_4 \geq p_2$ . **Proof:** Suppose not and  $p_4 < p_2$ . By (1) we can consider  $0 \leq p_4 < p_2 \leq 1$ . In this case the expression for  $v_1^*$  is given by equation (59). Consider raising  $p_4$  by  $(p_2 - p_4)$  to  $p'_4 = p_2$  and reducing  $p_1$  by  $\alpha'(p_2 - p_4)$  so that  $v_1^*$  is held constant. The allocation remains incentive compatible and there is no change in surplus. Moreover, the price change leaves  $U^*$  constant as the relative sizes of the opposing price changes (the change in  $p_1$  is smaller by factor  $\alpha'$ ) are offset by the relative probabilities they are perceived to be paid (since the second-period allocation is efficient from part 1, the perceived probability  $p_4$  is paid is smaller than that of  $p_1$  by factor  $\alpha'$ .) However the true utility delivered and hence the perception gap both change because the relative likelihood the two prices are paid depends on  $\alpha$  rather than  $\alpha'$ . Plugging in efficient second-period allocations, the perception gap is initially:

$$\Delta = (\alpha - \alpha') \begin{pmatrix} -b_1 + (b_1 - b_0) p_1 - (1 - p_2) \\ + (b_0 + (\alpha + \alpha') (b_1 - b_0)) (p_4 - p_2) \end{pmatrix} \quad (63)$$

After adjusting prices to  $p'_1 = p_1 - \alpha'(p_2 - p_4)$  and  $p'_4 = p_2$ , this becomes

$$\Delta' = (\alpha - \alpha') (-b_1 + (b_1 - b_0) (p_1 - \alpha'(p_2 - p_4)) - (1 - p_2)).$$

---

<sup>39</sup>If the consumer does not buy in the first period she will buy in the second period ( $b_{00} = b_{01} = 1$ ) and pay an additional  $|p_2|$  because  $p'_2 = 0$ . On the other hand, if she does buy in the first period she will still pay an additional  $|p_2|$  due to the increase in  $p_1$ . However both changes are equally offset by the reduction in the fixed fee.

The difference is

$$\Delta' - \Delta = (\alpha - \alpha') (p_2 - p_4) (b_0 + \alpha (b_1 - b_0))$$

which is nonnegative since  $p_2 > p_4$  by assumption and incentive compatibility requires  $b_1 \geq b_0$ . Thus this is a profitable deviation (strictly profitable if there are any purchases in the first period).

**Claim (3)** It is optimal for the firm to induce the efficient allocation in the first period.

**Claim (3a)** It is optimal for the firm to induce  $b_1 = 1$ . **Proof:** Suppose not and  $b_1 < 1$ . Incentive compatibility requires  $b_0 = 0$  and  $v_1^* = p_1 + \alpha' (p_4 - p_2) \geq 1$ . Suppose  $v_1^* > 1$ . Then  $b_0 = b_1 = 0$  and I can reduce  $p_1$  to  $1 - \alpha' (p_4 - p_2) \geq 1 - \alpha' > 0$  so that  $v_1^* = 1$  without disrupting incentive constraints or effecting payoffs or violating NFL constraints. (Constraint  $p_0 + p_1 \geq 0$  is redundant to  $p_0 \geq 0$  since  $p_1$  is positive. Constraint  $p_0 + p_1 + p_4 \geq 0$  is implied by  $p_4 \geq p_2$  from part (2) and  $p_1$  positive since  $p_0 + p_2 + p_1 \geq p_0 + p_2 \geq 0$ .) So I can safely consider  $v_1^* = 1$ . Now consider raising  $b_1$  to 1. Since  $v_1^* = 1$ , the consumer is indifferent and  $U^*$  is unaffected. Surplus is strictly increased. The perception gap is initially described by equation (63). After the increase in  $b_1$ , this becomes

$$\Delta' = (\alpha - \alpha') \left( \begin{array}{c} -1 + (1 - b_0) p_1 - (1 - p_2) \\ + (b_0 + (\alpha + \alpha') (1 - b_0)) (p_4 - p_2) \end{array} \right)$$

$$\begin{aligned} \Delta' &= (\alpha - \alpha') ((1 - b_0) p_1 - 1) - (1 - p_2) \\ &\quad + (p_4 - p_2) (\alpha - \alpha') (b_0 + (\alpha + \alpha') (1 - b_0)) \end{aligned}$$

and the difference is

$$\Delta' - \Delta = (\alpha - \alpha') (1 - b_1) ((\alpha + \alpha') (p_4 - p_2) - (1 - p_1)).$$

Substituting  $p_1 = 1 - \alpha' (p_4 - p_2)$  into this expression yields

$$\Delta' - \Delta = (\alpha - \alpha') (1 - b_1) (p_4 - p_2) \alpha$$

which is nonnegative since  $p_4 \geq p_2$  by part (2). Thus there is a strict increase in profits at least as high as the increase in surplus. Hence  $b_1 < 1$  could not have been optimal.

**Claim (3b)** It is optimal for the firm to induce  $b_0 = 0$ . **Proof:** Suppose not and  $b_0 > 0$ . Incentive compatibility requires  $b_1 = 1$  and  $v_1^* = p_1 + \alpha' (p_4 - p_2) \leq 0$ . Suppose that  $v_1^* < 0$ . Then  $b_0 = b_1 = 1$  and I can increase  $p_1$  to  $-\alpha' (p_4 - p_2)$  such that  $v_1^* = 0$  without disruption incentive compatibility. If I increase  $p_0$  by the same amount, then payoffs ( $U$ ,  $U^*$ , and  $\Pi$ ) remain constant.

Moreover NFL constraints are still satisfied. The constraints involving  $p_2$  and  $p_4$  are redundant since  $p_2, p_4 \geq 0$ . The constraint  $p_0 + p_1 \geq 0$  is unaffected because the sum remains constant. Moreover, it implies  $p_0 \geq 0$  since  $p_1 = -\alpha'(p_4 - p_2)$  is nonpositive by part (2). Thus the joint price change maintains incentive compatibility, NFL constraints, and does not affect payoffs. So I can safely consider  $v_1^* = 1$  and  $p_1 = -\alpha'(p_4 - p_2)$ .

Note that  $p_0 + p_1 \geq 0$  implies  $p_0 \geq -p_1 = \alpha'(p_4 - p_2)$ . Now consider increasing  $p_1$  to zero, increasing  $p_2$  to  $p_4$ , and reducing  $p_0$  by  $\alpha'^2(p_4 - p_2)$ . NFL constraints are all satisfied. The preceding note shows  $p_0 \geq 0$ . This implies  $p_0 + p_1 \geq 0$  since  $p_1$  is now 0. The remaining constraints with  $p_2$  and  $p_4$  are slack since  $p_2 = p_4$  is strictly positive. Increasing  $p_1$  to 0 lowers  $U^*$  by  $\alpha'^2(p_4 - p_2)$ . Increasing  $p_2$  to  $p_4$  lowers  $U^*$  by  $(1 - \alpha')\alpha'(p_4 - p_2)$ . The total reduction is  $\alpha'(p_4 - p_2)$ . Lowering  $p_0$  by  $\alpha'(p_4 - p_2)$  exactly offsets this change so that  $U^*$  is in fact held constant. Note that the change in  $p_1$  and  $p_2$  ensure that  $v_1^*$  remains equal to zero and incentive compatibility is maintained. Surplus is unchanged but profits are effected via the perception gap. Substituting  $b_1 = 1$  and  $p_1 = -\alpha'(p_4 - p_2)$  into equation (63) yields an expression for the initial perception gap:

$$\Delta = -(\alpha - \alpha')(2 - p_2 - (p_4 - p_2)(\alpha + b_0(1 - \alpha))).$$

After the price change, substituting  $p'_1 = 0$ ,  $p'_2 = p_4$ , and  $b_1 = 1$  into equation (63) yields an expression for the new perception gap:

$$\Delta' = (\alpha - \alpha')(-1 - (1 - p_4)) \tag{64}$$

$$\Delta' = -(\alpha - \alpha')(2 - p_4).$$

Thus the difference is

$$\Delta' - \Delta = (\alpha - \alpha')(p_4 - p_2)(1 - (\alpha + (1 - \alpha)b_0)),$$

which is nonnegative. Thus this price change weakly increases profits.

Finally, lower  $b_0$  to 0. This strictly increases surplus, and does not further effect the perception gap  $\Delta' = -(\alpha - \alpha')(2 - p_4)$  because the penalty fee is zero. Thus profits strictly increase. Hence  $b_0 > 0$  was not optimal.

## E.2 Proof of Lemma 3

Solving equation (30) for  $p_0$  yields:

$$p_0 = -U^* + v_0 + 2(1 - \alpha') b_0 (-\bar{p}) + 2\alpha' b_1 (1 - \bar{p}) - ((1 - \alpha') b_0 + \alpha' b_1)^2 p_3.$$

Substituting this for  $p_0$  into equation (31) gives:

$$\Pi = G(U^*) \left( \begin{array}{c} -U^* + v_0 + 2b_0(-(\alpha - \alpha')\bar{p} - (1 - \alpha)c) + 2b_1((\alpha - \alpha')\bar{p} + \alpha' - \alpha c) \\ + \left( ((1 - \alpha)b_0 + \alpha b_1)^2 - ((1 - \alpha')b_0 + \alpha' b_1)^2 \right) p_3 \end{array} \right).$$

There are four alternatives to the efficient allocation to consider:

**Case (1)**  $b_0 = b_1 = 1$ : Profits and the fixed fee are:

$$\Pi_1 = G(U^*) (-U^* + v_0 + 2(\alpha' - c)),$$

$$p_0 = -U^* + v_0 + 2\alpha' - 2\bar{p} - p_3.$$

If  $U^* \leq v_0 + 2\alpha'$ , then this allocation can be implemented without violating the NFL constraint with prices  $p_1 = p_2 = p_3 = 0$  and  $p_0 = -U^* + v_0 + 2\alpha'$ . If  $U^* > v_0 + 2\alpha'$ , then this allocation is not implementable without violating the NFL constraint. This follows from the fact that  $p_0 + 2\bar{p} + p_3 \geq 0$  is equivalent to  $U^* \leq v_0 + 2\alpha'$ . However, the efficient allocation could be implemented with identical prices, also satisfying the NFL constraint for  $U^* \leq v_0 + 2\alpha'$ , but yielding strictly higher profit,

$$\Pi = G(U^*) (-U^* + v_0 + 2(\alpha' - \alpha c)),$$

by saving production cost  $2(1 - \alpha)c$ . Thus  $b_0 = b_1 = 1$  is never optimal.

**Case (2)**  $b_0 = b_1 = 0$ : Profits and the fixed fee are:

$$\Pi_2 = G(U^*) (-U^* + v_0),$$

$$p_0 = -U^* + v_0.$$

If  $U^* \leq v_0$  then this allocation is implementable without violating the NFL with prices  $p_0 = -U^* + v_0$ ,  $p_1 = p_2 = 1$ , and  $p_3 = 0$ . If  $U^* > v_0$ , then this allocation is not implementable without violating the NFL constraint. However, the efficient allocation can be implemented with identical prices, strictly raising profits by  $2\alpha(1 - c)$  from the additional sales. Thus  $b_0 = b_1 = 0$  is never optimal.

**Case (3)**  $b_0 \in (0, 1)$ ,  $b_1 = 1$ : For this allocation to be implemented,  $b_0$  must satisfy first- and second-order conditions of the consumers' problem:

$$\frac{dU^*}{db_0} = -2(1 - \alpha')(\bar{p} + ((1 - \alpha')b_0 + \alpha')p_3) = 0,$$

and

$$\frac{d^2U^*}{db_0^2} = -2(1 - \alpha')^2 p_3 \leq 0.$$

This requires that  $p_3 \geq 0$  and  $\bar{p} = -((1 - \alpha')b_0 + \alpha')p_3$ . At these prices, the three NFL constraints, (a)  $p_0 \geq 0$ , (b)  $p_0 + \bar{p} \geq 0$ , and (c)  $p_0 + 2\bar{p} + p_3 \geq 0$  are:

$$\max \left\{ \frac{U^* - v_0 - 2\alpha'}{((1 - \alpha')b_0 + \alpha')^2}, \frac{U^* - v_0 - 2\alpha'}{(1 - \alpha')^2(1 - b_0)^2} \right\} \leq p_3 \leq \frac{2\alpha' + v_0 - U^*}{(1 - \alpha')(1 - b_0)((1 - \alpha')b_0 + \alpha')}$$

If  $p_3 \geq 0$ , the upper bound on penalty fees can only be satisfied if  $U^* \leq v_0 + 2\alpha'$ , in which case the lower bound is always satisfied. Moreover, profits are increasing in penalty fee  $p_3$ ,

$$\Pi_3 = G(U^*) \left( v_0 - U^* + 2(\alpha' - \alpha c) - 2b_0(1 - \alpha)c + (\alpha - \alpha')^2(1 - b_0)^2 p_3 \right),$$

so the optimal penalty fee satisfies the upper bound with equality:

$$p_3 = \frac{(2\alpha' + v_0 - U^*)}{(1 - \alpha')(1 - b_0)((1 - \alpha')b_0 + \alpha')}.$$

Given these prices, profits are strictly decreasing in  $b_0$ ,

$$\frac{d\Pi_3}{db_0} = G(U^*) \left( -2(1 - \alpha)c - \frac{(\alpha - \alpha')^2(1 - b_0)}{((1 - \alpha')b_0 + \alpha')} p_3 \right) < 0,$$

for all  $p_3 \geq 0$  and hence any NFL implementable allocation with  $b_0 \in (0, 1)$  is always dominated by the efficient allocation.

**Case (4)**  $b_0 = 0$ ,  $b_1 \in (0, 1)$ : For this allocation to be implemented,  $b_1$  must satisfy first- and second-order conditions of the consumers' problem:

$$\frac{dU^*}{db_1} = +2\alpha'(1 - \bar{p}) - 2(\alpha')^2 b_1 p_3 = 0,$$

and

$$\frac{d^2U^*}{db_1^2} = -2(\alpha')^2 p_3 \leq 0.$$

This requires  $p_3 \geq 0$  and  $\bar{p} = 1 - \alpha' b_1 p_3$ . At these prices, the three NFL constraints, (a)  $p_0 \geq 0$ ,

(b)  $p_0 + \bar{p} \geq 0$ , and (c)  $p_0 + 2\bar{p} + p_3 \geq 0$  are:

$$\max \left\{ \frac{U^* - v_0}{\alpha'^2 b_1^2}, \frac{U^* - v_0 - 2}{(1 - \alpha' b_1)^2} \right\} \leq p_3 \leq \frac{1 + v_0 - U^*}{\alpha' b_1 (1 - \alpha' b_1)}$$

All three constraints can be satisfied only if  $U^* \leq v_0 + \alpha' b_1$ . (This is equivalent to  $\frac{U^* - v_0}{\alpha'^2 b_1^2} \leq \frac{1 + v_0 - U^*}{\alpha' b_1 (1 - \alpha' b_1)}$ , while  $\frac{U^* - v_0 - 2}{(1 - \alpha' b_1)^2} \leq \frac{1 + v_0 - U^*}{\alpha' b_1 (1 - \alpha' b_1)}$  is equivalent to the weaker condition  $U^* \leq 1 + v_0 + \alpha' b_1$ .) Otherwise, this allocation is not implementable without violating NFL. Profits are strictly increasing in  $p_3$ ,

$$\Pi_4 = G(U^*) \left( -U^* + v_0 + 2b_1\alpha(1 - c) + b_1^2 (\alpha - \alpha')^2 p_3 \right),$$

so the optimal penalty fee  $p_3$  will equal the upper bound:

$$p_3 = \frac{1 + v_0 - U^*}{\alpha' b_1 (1 - \alpha' b_1)}.$$

Given these prices, profits are strictly increasing in  $b_1$ ,

$$\frac{d\Pi_4}{db_1} = G(U^*) \left( 2\alpha(1 - c) + \frac{b_1(\alpha - \alpha')^2}{1 - \alpha' b_1} p_3 \right) > 0,$$

so any NFL implementable allocation with  $b_1 \in (0, 1)$  is dominated by the efficient allocation.