C Policy Applications

C.1 FCC’s Proposed Bill-Shock Regulation

US cellular-phone customers are typically charged steep penalty fees for exceeding usage allowances, and the variation in usage allowances across calling plans is an essential instrument for encouraging consumers to self-select into different calling plans. On October 14th, 2010, the FCC proposed bill-shock regulation that would require carriers to notify customers, via voice or text alerts, when they are about to exceed plan limits and begin incurring overage charges (FCC 2010). The FCC’s proposed bill-shock regulation had strong support from consumer groups but was opposed by major cellular carriers (see footnote 4 as well as Genachowski (2010), Wyatt (2010), Deloney et al. (2011), and Stross (2011)). A year later, on October 17th, 2011, it was announced that cellular carriers would voluntarily begin providing such usage alerts by April 2013 and in return the FCC would drop bill-shock regulation (CTIA - The Wireless Association 2011).

The price-discrimination model in Section 5 provides an explanation both for carriers’ use of surprise penalty fees on typical cellular contracts and for carriers’ opposition to proposed bill-shock regulation. If one believes that cellular-phone customers are unbiased and the cellular market is sufficiently competitive, then Corollary 2 implies that the FCC’s recent bill-shock agreement will be counterproductive and lower social welfare. Moreover, consumer groups’ strong support for the agreement would be misplaced, since it would harm some consumers. (Their support is nevertheless understandable since the agreement would be unambiguously beneficial to all consumers but for the resulting endogenous price changes predicted by the model.)

There are four caveats to this criticism of the FCC’s bill-shock agreement. First, the magnitude of the effects described by Corollary 2 may be small. Second, while Corollary 2 assumes a fairly-competitive market, the Department of Justice has argued that cellular carriers do have substantial market power. (The Department of Justice made this argument in 2011 when it sued to block AT&T’s proposed acquisition of T-Mobile (Department of Justice 2011).) As illustrated by Corollary 3 of the working paper version (Grubb 2012), sufficient market power means that bill-shock regulation could increase or decrease social surplus when consumers are unbiased. Third, Grubb (2009) and Grubb and Osborne (forthcoming) show that cellular-phone customers have bi-
ased beliefs about their likely usage and in particular are overconfident. Overconfidence is consistent with naive inattention and causes consumers to underestimate the likelihood of paying penalties. Proposition 8 shows that such naivete and resulting bias provide a second reason (in addition to market power) that bill-shock regulation could raise rather than lower social surplus. Finally, a fourth caveat is that the regulation would apply to fees beyond overage charges such as roaming fees which are typically the same across calling plans, and hence not used for price-discrimination purposes or relevant to this theoretical argument. Roaming charges were the target of recently adopted bill-shock regulation in the EU.

In sum, the welfare impact of bill-shock regulation is ambiguous but there is a clear reason that it could be socially harmful. Thus, implementing the FCC’s bill-shock agreement may lower not just firm profits but also lower total welfare and hurt some consumers. To resolve the theoretical ambiguity, complementary empirical work by Grubb and Osborne (forthcoming) estimates a structural demand model of consumers’ contract and calling choices using a panel of cellular billing data. Grubb and Osborne’s (forthcoming) counterfactual simulation predicts bill-shock regulation will lower social surplus by $26 per consumer and consumer surplus by $33 per consumer annually. In contrast, Jiang (2013) predicts that bill-shock regulation will raise social and consumer surplus by $24 per household annually. (The difference between these empirical results may be due to a number of modeling differences. For instance, in contrast to Jiang (2013), Grubb and Osborne (forthcoming) allow consumer beliefs to be biased and model consumers’ endogenous change in calling behavior in response to information in bill-shock alerts.)

C.2 Overdraft Fees

Turning to a second application, consider overdraft fees: in 2009, US bank overdraft fee revenues from ATM and one-time debit-card transactions were $20 billion (Martin, 2010). Prior to the Federal Reserve Board’s adoption of an opt-in rule, Bank of America and other banks charged high (often $35) overdraft fees on debit and ATM transactions without notifying customers at the point of sale. When the Federal Reserve Board proposed opt-in regulation, banks opposed it.1 Nevertheless, effective August 15, 2010 (July 1, 2010 for new accounts) new Federal Reserve Board rules “prohibit financial institutions from charging consumers fees for paying overdrafts on automated teller machine (ATM) and one-time debit-card transactions, unless a consumer consents, computer

1Prior to regulating overdraft fees, the Federal Reserve solicited public comment. Industry commenters sought to undermine the regulation in every possible way. For instance “industry commenters…urged the Board to permit institutions to vary the account terms …for consumers who do not opt in [to overdraft protection]” (Federal Reserve Board, 2009b). Clearly banks wanted to be able to make declining overdraft protection an expensive account feature.
or opts-in, to the overdraft service for those types of transactions” (Federal Reserve Board, 2009a).

In response to opt-in regulation, Bank of America chose to stop offering overdraft protection on debit-card transactions, despite the fact that Bank of America is estimated to have earned $2.2 Billion from ATM and debit-card-transaction overdraft fees in 2009 (Sidel and Fitzpatrick, 2010). Other major banks have been accused of responding with deceptive marketing campaigns to induce opt-in. For instance, customers filed a federal class-action lawsuit against JPMorgan Chase for such bad behavior in August 2010 (Dinzeo, 2010; McCune et al., 2010). More broadly, the Consumer Financial Protection Bureau (CFPB) announced in February 2012 that it will be investigating reports of such misleading marketing (Wyatt, 2012).

The model used throughout the paper is stylized and fits the overdraft application imperfectly. In particular, while quantity is one-dimensional in the model, overdraft fees depend on both dollars spent and the number of transactions. Overdraft fees are only triggered once dollars spent exceed the account balance, but are then charged on a per-transaction basis. Moreover, marginal costs are likely increasing rather than constant. Nevertheless, the model captures important features of the setting.

For instance, consider the model of naive inattention in Section 4. Assume that the distribution of taste-shocks approximates a Bernoulli distribution in which values \( v_t \) are one with probability \( \alpha \) and zero otherwise. We can interpret \( \alpha \) as the probability a consumer wishes to make a purchase with her debit card. Then \( v = 1 \) is her value for making the purchase with her debit card, rather than an alternative such as a credit card.

Do the results in Section 5 imply that the combination of consumer inattention and overdraft fees could be socially valuable by making price discrimination by banks less distortionary? In fact they do not apply. While banks offer different types of checking accounts, prior to the regulation banks typically charged the same overdraft fees on all accounts (e.g. Bank of America (2010)). Thus heterogeneity in expectations of overdraft usage is typically not an important dimension of self-selection across checking accounts.

Neither the benchmark model nor Section 5’s model of price discrimination explain banks’ widespread use of overdraft fees, failure to notify consumers at the point-of-sale, or aversion to opt-in regulation. A more compelling explanation for these facts is that some consumers are naively inattentive and therefore underestimate the incidence of overdraft fees. Given such naivete, Proposition 8 shows that bill-shock regulation should raise social surplus and protect naive consumers. Moreover, as illustrated by Example 1, bill-shock regulation could stiffen competition and shift sufficient surplus away from banks so as to benefit all consumers.

Consumer naivete may also explain high opt-in rates. Analysts initially predicted that opt-
in regulation would dramatically reduce overdraft-fee revenue (Campbell [2009]). In fact, opt-in
rates have been high (75%) and overdraft-fee revenue has been relatively stable (Benoit [2010]).
High opt-in rates for overdraft fees may be due to deceptive marketing practices currently under
investigation (Wyatt [2012]). However, an alternative explanation is that consumers naively believe
that attention is costless. In this case it is weakly dominant to opt-in, in order to have the option
of making an overdraft if it is ever needed in the future. There is no downside because a consumer
with costless attention would never pay an overdraft fee accidentally.

Whether due to deceptive marketing or naivete about inattention, recent experience with opt-
in regulation suggests that it is a poor substitute for bill-shock regulation. The CFPB apparently
believes that the current opt-in regulation is inadequate and has proposed that a “penalty fee box”
should appear on checking account statements detailing overdraft fees charged during the month
(Wyatt [2012]). While this helps alert consumers when they have been charged overdraft fees, it
still falls short of bill-shock regulation which would help them avoid such fees (and associated $40
cups of coffee) at the point of sale.

D Additional Intuition for Appendix A Results

D.1 Lemma 1 and associated first-order conditions

Intuition for Lemma 1: To understand Lemma 1, begin with case (2). The downward incentive
constraint binds because the firm would like to offer the low segment a discounted markup ($\mu_H^* >
\mu_L^*$). The fact that marginal prices are distorted above marginal cost on the low contract follows
from standard price discrimination logic. High types find increases in marginal prices more costly
than do low types because high types make more purchases. Thus raising marginal prices on the
low contract relaxes the downward incentive constraint (discouraging the high type from choosing
the low contract) at the cost of distorting low-types’ allocations downwards. The positive penalty
fee $p_{3L}$ makes the second-period marginal price larger after an initial purchase. This is optimal
because a deviating high type is more likely to purchase in the first period than a low type.

If $\mu_H^* = \mu_L^*$ then firms have no desire to price discriminate, which means that a single marginal
cost contract is optimal. If $\mu_H^* < \mu_L^*$ then optimal pricing follows a similar intuition to that for the
case $\mu_H^* > \mu_L^*$, but distortions are reversed because high-types receive a discount and hence the
upward incentive constraint binds rather than the downward constraint. Interestingly, although
distortions are upwards for high-types rather than downwards for low types, penalty fees are still
positive on the distorted contract. Nevertheless, their role is reversed. When $\mu_H^* > \mu_L^*$, a positive
penalty $p_{3L}$ increases the distortion of marginal price above marginal cost. In contrast, when
μ_H^* < μ_L^*}, a positive penalty fee \( p_{3H} \) reduces the distortion of marginal price below marginal cost.$^2$

**Intuition for optimal marginal prices:** To understand the marginal prices relevant to Lemma 1 that are characterized by equations (A.4)-(A.6) in Appendix A.1, it is helpful to begin by understanding optimal pricing in a simplified model with only one purchase opportunity rather than two. First consider case (2), \( μ_H^* > μ_L^* \), for which the downward incentive constraint binds and the optimal marginal price on the low contract is distorted above marginal cost.$^3$

\[
pl = c + \frac{-∂Π}{∂U_H} \left( F_L(p_L) - F_H(p_L) \right) \left( 1 - β \right) G_L(U_L) f_L(p_L) \tag{D.1}
\]

The distortion away from marginal cost follows from standard price discrimination logic. High types find increases in marginal prices more costly than do low types because high types make more purchases. Thus raising marginal prices on the low contract relaxes the downward incentive constraint at the cost of distorting low-types’ allocations downwards. The optimal distortion in marginal price, given by equation (D.1), follows from a first-order condition which equates the marginal cost of distortions, \((p_L - c)(1 - β)G_L(U_L)f_L(p_L)\), to the marginal benefit of relaxing the constraint, \(-∂Π/∂U_H (F_L(p_L) - F_H(p_L))\). Unpacking the cost of distortion, \((p_L - c)\) is the lost surplus from the marginal foregone purchase and \((1 - β)G_L(U_L)f_L(p_L)\) is the likelihood a consumer is a low-type on the margin. Unpacking the benefit, \((F_L(p_L) - F_H(p_L))\) is the amount by which raising \(p_L\) relaxes the downward incentive constraint, and \(-∂Π/∂U_H\) is the shadow value of relaxing the constraint.

Optimal marginal pricing for the case \(μ_H^* > μ_L^*\) can now be understood by comparing equations (A.4)-(A.6) to equation (D.1). First-period marginal cost in equation (A.4) is adjusted by \(\int_{p_L}^{p_{3L}+p_{3L}} (v - c) f_L(v) dv\), which is the second-period surplus lost when a first-period purchase triggers the penalty fee in period two. Second-period distortions away from marginal cost in equations (A.5)-(A.6) are adjusted by the additional terms \(F_H(v_{3L}^H)/F_L(v_{3L}^L) < 1\) and \((1 - F_H(v_{3L}^H))/ (1 - F_L(v_{3L}^L)) > 1\) respectively. This implies that penalty fee \(p_{3L}\) is positive. Marginal prices are distorted upwards to discourage the high type from choosing the low contract. A positive penalty fee makes the second-period distortion larger after an initial purchase. This is optimal because a deviating high type is more likely to purchase in the first period than a

---

$^2$The reason is as follows. If both types choose the same contract, the penalty fee is always more likely to be paid by the high type. If \(μ_H^* > μ_L^*\), deviations to contract \(L\) by type \(H\) must be dissuaded. Thus the penalty fee on contract \(L\) should be used to increase the distortion, targeting it towards deviating high types but away from low types who actually choose the contract. If \(μ_H^* < μ_L^*\), deviations to contract \(H\) by type \(L\) must be dissuaded. Thus the penalty fee on contract \(H\) should be used to decrease the distortion, targeting it towards deviating low types but away from high types who actually choose the contract.

$^3$If \(G_θ(U_θ) = 1_{U_θ ≥ 0}\) (see footnote 37), then \(-∂Π/∂U_H = β, G_L(U_L) = 1\), and this matches Courty and Li (2000).
D.2 Intuition for Lemma 2

For intuition behind Lemma 2, consider what happens if penalty $p_3$ is increased by one dollar. First, following equations (20)-(21), base marginal charges $p_1 = p_2 = p_\theta$ must be reduced by $(1 - F_\theta (v^s_\theta))$ to keep expected marginal price $v^s_\theta$ constant. This reduces expected variable payments by $(1 - F_\theta (v^s_\theta))^2$ because the extra dollar in the penalty fee is paid with probability $(1 - F_\theta (v^s_\theta))^2$ but the $(1 - F_\theta (v^s_\theta))$ discount on the base marginal charge is paid with probability $2 (1 - F_\theta (v^s_\theta))$. Thus, following equation (22), a second change is that the fixed fee is increased by $(1 - F_\theta (v^s_\theta))^2$ to keep the offered gross utility $U_\theta$ constant. By construction, these changes leave type $\theta$ indifferent. Any other type $\hat{\theta}$ that chooses contract $\theta$ would pay the same increase in the fixed fee but receive a smaller reduction in expected variable payments. If type $\hat{\theta}$ buys with probability $\pi = 1 - F_{\hat{\theta}} (v^s_{\hat{\theta}})$, her expected variable payments are reduced by $2 \pi (1 - F_\theta (v^s_\theta)) - \pi^2$. Notice that this reduction is maximized at $\pi = (1 - F_\theta (v^s_\theta))$. Thus regardless of whether type $\hat{\theta}$ is a higher type and $\pi > (1 - F_\theta (v^s_\theta))$ or a lower type and $\pi < (1 - F_\theta (v^s_\theta))$, increasing the penalty $p_3$ increases total expected payments for her on contract $\theta$.

E Example 3

Example 3 There are $1/2$ low types and $1/2$ high types. Values are distributed $v_l \sim U [0,10]$ for low types and $v_l \sim U [0,15]$ for high types. Marginal cost is $c = 5$. The market is a Hotelling duopoly with transportation costs $\tau_L = 1/6$ and $\tau_H = 1/3$.

Example 3 extends Example 2 by specifying equal proportions of low and high types and placing them in a Hotelling duopoly. Notice that $1/\bar{f} = 10$, and therefore without bill-shock regulation the contracts described in Example 2 satisfy $p_{3\theta} \leq 1/\bar{f}$. Moreover, $\phi_H (1/\bar{f}) = 1/6$, which is the difference $\tau_H - \tau_L = 1/6$. Therefore, without bill-shock regulation equilibrium is symmetric, allocations are efficient, firms charge markups equal to transportation costs in each segment, and contracts described in Example 2 are in the set of equilibrium contracts. (The high contract could also be priced at $\tau_H$ plus marginal cost because the upward incentive constraint does not bind.) With bill-shock regulation, however, equilibrium will involve distortion and inefficiency. Table 2 below compares outcomes with and without regulation. It is useful to compare welfare changes to the size of the transportation cost difference $\Delta \tau = \tau_H - \tau_L = 1/6 \approx 0.17$, which underlies all price-discrimination related distortions. As should be expected, the welfare consequences of bill-shock regulation are less than $\Delta \tau$. Bill-shock regulation lowers utility to low types by about 22% of $\Delta \tau$, 


d low type.
Table 2: Equilibrium Outcomes in Example 3

<table>
<thead>
<tr>
<th></th>
<th>No Regulation</th>
<th>Bill-Shock Regulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathbf{p}<em>L = (p</em>{0L}, p_{1L}, p_{2L}, p_{3L})$</td>
<td>(2.67, 0, 0, 10)</td>
<td>(-0.03, 5.19, 5.19, 0.20)</td>
</tr>
<tr>
<td>$\mathbf{p}<em>H = (p</em>{0H}, p_{1H}, p_{2H}, p_{3H})$</td>
<td>(3.67, 0, 0, 7.5)*</td>
<td>(0.31, 5, 5, 0)</td>
</tr>
<tr>
<td>Contract $L$ calling thresholds</td>
<td>$v_{1L}^s = v_{2L}^s = 5$</td>
<td>$v_{1H}^0$, $p_{2L}$, $p_{2L} + p_{3L}$ approximates 5.28, 5.19, 5.39</td>
</tr>
<tr>
<td>Gross Utility: $U_L, U_H$</td>
<td>2.33, 6.33</td>
<td>2.30, 6.36</td>
</tr>
<tr>
<td>Gross Surplus: $S_L, S_H$</td>
<td>2.50, 6.67</td>
<td>2.49, 6.67</td>
</tr>
<tr>
<td>Markups: $\mu_L, \mu_H$</td>
<td>0.17, 0.33</td>
<td>0.19, 0.31</td>
</tr>
<tr>
<td>Industry Profits</td>
<td>0.25</td>
<td>0.25</td>
</tr>
</tbody>
</table>

*Or (0.33, 5, 5, 0), or one of many other contracts with $v_{1L}^s = v_{2L}^s = 5$ and $U_H = 6.33$.

creates a deadweight-loss of about 5% of $\Delta \tau$ for each low type, and raises utility to high types by about 16% of $\Delta \tau$. Changes in industry profits are smaller.

References


