Billing Abuses by the Experts:  
A Game-Theoretic Analysis of Legal Services

Chris Kuo*

School of Management, Boston University, U.S.A.

Billing abuses exist when an expert charges the expert hourly rate but offers quality that does not match the charge. This paper provides a game-theoretic analysis so that, when the quality of an expert’s services is unobservable to consumers before purchase, hourly rate competition can eliminate the profits necessary to induce the expert to offer the quality services. This paper further demonstrates that the threat of business termination by customers is not sufficiently credible to ensure that the experts will produce high quality services. Given the fact that billing abuses always exist, this paper suggests that auditing by the customers is necessary to reduce the overcharged amount, and customers have to audit a high percentage of the billed amount.

Keywords: asymmetric information, sub-game perfect equilibrium, credence service

JEL classification: C7, D82, G22

1 Introduction

Hourly billing in law firm practice has been around for fifty years in the United States and other countries. While it is generally believed that the majority of the attorneys are ethical in their billing practices, many sources have concerns that attorneys routinely overbill their clients. One of the most significant factors that
inflate legal costs is the use of hourly fees that are higher than necessary. Attorneys perform paralegal tasks but bill at attorney hourly rates. In the U.S., the American Bar Association (ABA) has taken official notice of the problems caused by lawyer billing abuses: “One major contributing factor to the negative public opinion of the legal profession appears to be the billing practices of some of its members.” Billing abuses have also made headline news. For example, in Japan, a lawyer was found padding the invoices for a government-funded legal service entity for several years.1 A number of U.S. legal scholars have also documented the abusive lawyer billing practices. Ross (1996) conducted surveys of lawyers and found an upward trend in the percentage of attorneys who admitted to billing abuses and did not consider it to be an ethical problem. His data showed that the percentage of attorneys who believed that “double billing” was unethical fell from 64.7 percent in the 1995-1996 survey to 51.8 percent in the 2006 - 2007 survey. Lerman (1990) documents the tactics used by law firms to overcharge their clients when compensation is based on the hours worked. Terms such as overbilling and abusive billing are almost used interchangeably in the above studies. Corporate and government customers who rarely take the time to effectively monitor their legal bills are particularly attractive targets for abusive billing scams by unethical lawyers.

The fundamental economic issue of the abusive billing practice is information asymmetry between experts and customers, such as lawyer-client or physician-patient relationships. While the lawyer or doctor identifies the services that the customer needs, he has the opportunity to exploit the informational asymmetry by defrauding the customer. There are numerous examples of information asymmetry in our daily lives. For example, a homeowner asks an experienced contractor for a roof repair. The contractor may allow the apprentice to perform the job but still charges the master’s hourly rate. The example in this study focuses on the legal profession, but the discussion can be generalized to many other expert-customer relationships.

How can a customer prevent the overbilling practice? A common method is to closely monitor the charges. The Defense Research Institute published The Recommended Case Handling Guidelines to insurers and legal defense firms as a model that they could follow in promulgating their own billing guidelines (Ream,

Large business entities typically have their billing guidelines that the lawyers have agreed upon, and should abide by (Matturro, 1999, 2000). According to the guidelines, any non-compliant billing practices will result in reductions in the bills.

However, many companies are not capable of auditing the enormous amount of bills they receive and merely review a percentage of them. This paper will show that the hourly rate that a lawyer has agreed upon has already factored the review percentage into it. The higher the percentage of the billed amount that a customer is able to audit, the less likely it is that the agreed hourly rate can be inflated. This suggests that government or corporate clients should be equipped to audit the bills for cost containment purposes. This study also implies that a high percentage of the billed amount should be audited to reduce the overcharged bills.

A customer can terminate the business relationship with the service provider if any non-compliance is deemed to be detrimental. In the absence of any industry self-regulation or third-party enforcement, this self-protection by customers constitutes a threat to service providers. However, this paper will show that the threat of business termination is not sufficient to ensure that a lawyer does not perform tasks that do not require lawyer expertise. A lawyer might mix tasks of paralegals or secretaries into his or her service and still charge the lawyer’s hourly rate. Since neither bill auditing nor termination of the business relationship is sufficient to prevent billing abuses, the results imply that a systematic mechanism such as industry self-regulation or third-party enforcement is needed to correct the problem. Enforcement by legal ethics boards or professional groups, such as the American Bar Association, is recommended to set a higher standard to eliminate billing abuses.

This paper proceeds as follows: Section 2 reviews the related literature. Section 3 presents the model and shows the service choices and overbilling behaviors in equilibrium. Section 4 examines whether the threat of business termination by a customer is able to ensure that the expert focuses on the expert-only service. Finally, Section 5 concludes.

2 Related Literature

Goods or services for which an expert knows more about what a consumer needs than the consumer himself are called credence goods or services (Darby and Karni,
Consumers rely on experts to identify the correct type of service. Legal, medical, repair services and various types of consulting and advisory services belong to this broad category. Pitchik and Schotter (1987) and Wolinsky (1993) analyze the situations in which the expert may defraud the customer, by misrepresenting a low-cost service as a costly one, given that the customer cannot observe the type of service provided. Dulleck and Kerschbamer (2006) demonstrate that if neither verifiability nor liability holds, an expert has an incentive to provide low-quality services but to charge at high-quality rates. McCluskey (2000) shows that a profit-maximizing producer can gain from deceiving consumers by using false quality claims. So repeat-purchase relationships and third-party monitoring are necessary for the production of high-quality credence goods.

The analysis in this paper is also related to the literature on contract theory and reputation. Klein and Leffler (1981) show that when there is a sufficiently high price premium, the future stream of profits from producing the high-quality good is greater than the one-shot cost savings from producing low quality, and so theoretically firms will have the correct incentives. Shapiro (1983) suggests that sellers who sold high-quality products in the past have an incentive to maintain their reputation, because good reputations are rewarded with high prices and high profits in the future. Similarly, Frankel and Schwarz (2009) show that there exists an equilibrium in which experts act truthfully according to the interests of the customers, and they are rewarded with future business. Kranton (2003) shows that the production of high-quality goods may not exist when firms compete for market share. Dana and Fong (2010) model the service choices of oligopolists in an infinitely repeated game. They show that high-quality services can be supported as a perfect Bayesian equilibrium in a tacit collusion. The general conclusion in this literature is that firms have an incentive to produce high-quality products to maintain their reputations with consumers, or that the conditions that produce a temporary increase in profit may induce a firm to provide low-quality goods.
3 Model

Long-term business relationships typically exist between lawyers and their large clients including insurers or governments. This motivates the study to model the relationship in a repeated game. Some studies in the legal service literature also provide the same treatment (Johnston and Waldfogel, 2002; Pecorino and Boening, 2010). We assume there are \( N \geq 2 \) identical lawyers indexed by \( i = 1, \ldots, N \), and many customers, normalized to 1.

The legal service for a customer can be categorized into two types: the type of services that require a lawyer and the service of the paralegal or legal assistant. Assume \( c_l \) is the cost for a lawyer to produce a unit of the service that requires the expertise of a lawyer, and \( c_p \) is the cost for him/her to produce a unit of the service of the paralegal or legal assistant, \( c_p < c_l \). The lawyer can use his time to produce strictly the service that requires a lawyer and incurs \( c^u = c_l \). The lawyer can also use his/her time to produce a mix of the two types of services. Throughout this paper we call the services that require a lawyer, expert-only service, and the mix of two types of services, mixed service. In addition, we will use the terms expert and lawyer interchangeably in order to include a wide range of expert-customer relationships. In the mixed service, the lawyer performs \( \rho \in (0, 1) \) of the service that requires his expertise, and \( (1 - \rho) \) that can be performed by a paralegal or legal assistant. Thus, the cost of producing the mixed service is \( c^m = \rho c_l + (1 - \rho)c_p \). Obviously, \( c^m = c^u \) because \( \rho \) is strictly less than 1.

In our model the lawyer charges the clients a single hourly rate, \( p' \), for a unit of the service in period \( t \), where \( p' > c^u > 0 \). Without loss of generality, the

---

2 The type of work performed by a lawyer includes giving legal advice to clients, representing clients in court and filing legal motions in court. A paralegal is described by the National Federation of Paralegal Associations (NFPA) as someone who has qualified for that title “through education, training or work experience, to perform substantive legal work that requires knowledge of legal concepts and is customarily, but not exclusively, performed by a lawyer.” Tasks performed by a paralegal include case planning, development and management, legal research, interviewing clients, drafting and analyzing legal documents and representing clients before a state or federal administrative agency if that representation is permitted by statute, court rule, or administrative rule. A paralegal is not permitted to practice law, provide legal advice, or represent a client in court.

3 In legal services there are three common ways of billing: contingent, fixed and an hourly-rate basis. Garoupa and Gomez-Pomar (2008) presents an interesting discussion regarding why large law firms prefer hourly fees over contingent fees.
demand of a customer for legal service in each period is normalized to 1. We assume that the hourly rate for a unit of paralegal service is \( p^* \) which is exogenously determined by the labor market of the paralegals. Our study focuses on the general case where the lawyer’s hourly rate is higher than the paralegal’s, \( i.e., \ p' > p^* \).

When a lawyer produces a mixed service, he/she is supposed to charge \( p' \) for \( \rho \) portion of the service and \( p^* \) for the \((1-\rho)\) portion. However, the lawyer can defraud the customers by charging \( p' \) but still producing the mixed service. Since \( p' > p^* \), overbilling always exists when the lawyer performs a mixed work. This is an important feature of our analysis.

Let us turn to the customer side. The knowledge asymmetry between the experts and the customers obviously creates incentives for opportunistic experts. A customer can make some efforts to audit the service. The customer adjusts the billed amount according to the billing guidelines, then pays the adjusted amount. However, due to the enormous invoice volume, a customer can only audit a \( \gamma \in (0,1) \) percentage of the dollars. An aggressive customer may choose to review a higher percentage of the dollars. We assume that \( \gamma \) is exogenous and is common knowledge to both the customers and the experts. If the expert performs the expert-only service and charges the expert rate, there is no reduction. If the expert performs the mixed work, the amount of the reduction for the audited \( \gamma \) portion is \( \gamma(1-\rho)(p' - p^*) \). We assume the customers perform perfect audits for the \( \gamma \) portion. The customers bear the entire auditing cost.

The experts in our game compete for market share. We denote the market share of expert \( i \) in period \( t \) as \( s^t_i \in [0,1] \). The market share \( s^t_i \) is a function of the expert’s hourly rate \( p' \), the rival’s rates in vector form \( \mathbf{p'}_t \) and the past number of customers \( s^{t-1}_i \). The game begins in period 0 with \( s^0 \) consumers assigned to expert \( i \). \( \sum s^t_i = 1 \) in period \( t \). The profit for the expert-only service is \( (p' - c^u)s^t_i(p', \mathbf{p'}_t, s^{t-1}_i) \) in period \( t \), and that of the mixed service is \( (p' - c^u)s^t_i(p', \mathbf{p'}_t, s^{t-1}_i) \). In the repeated game, the continuation payoffs for an expert equal the discounted value of all the periods with a common discount factor \( \delta \in (0,1) \). We define the continuation payoff for expert \( i \) when producing the expert-only service as \( \Pi^e_i(p', \mathbf{p'}_t, s^t_i) \) and \( \Pi^m_i(p', \mathbf{p'}_t, s^t_i) \) for the mixed service. Let \( \overline{s}_i \) denote the expected, or mean, market share for expert \( i \) in period \( t \). If the expert performs the expert-only service at the hourly rate \( p' \) in all periods, the
continuation payoff is:

$$\Pi^i_t(p'_t, P'_t, s^i_t) = (p'_t - c''_t) \cdot \bar{F}_i(p'_t, P'_t, s^{i+1}_t) + \sum_{k=1}^{\zeta} \delta^{i+k}(p'_t - c''_t) \cdot \bar{F}_i(p'_t, P'_t, s^{i+1}_t).$$

(1)

If the expert performs the mixed service, given the fact that the customer audits the $$\gamma$$ portion, the continuation payoff is:

$$\Pi^i_t(p'_t, P'_t, s^i_t) = (p'_t - c''_t) \cdot \bar{F}_i(p'_t, P'_t, s^{i+1}_t)$$

$$- \gamma(p'_t - p''_t)(1 - \rho) \cdot \bar{F}_i(p'_t, P'_t, s^{i+1}_t)$$

$$+ \sum_{k=1}^{\zeta} \delta^{i+k}(p'_t - c''_t) \cdot \bar{F}_i(p'_t, P'_t, s^{i+1}_t)$$

$$- \gamma \sum_{k=1}^{\zeta} \delta^{i+k}(p'_t - p''_t)(1 - \rho) \cdot \bar{F}_i(p'_t, P'_t, s^{i+1}_t).$$

(2)

The second and the fourth terms on the right-hand side of Equation (2) are the reduced amounts by the customer in period $$t$$ and future periods. By contrast, Equation (1) does not have the adjusted amount for overbilling because the expert bills truthfully.

Throughout the paper the discussions focus on subgame-perfect Nash equilibria. A subgame-perfect Nash equilibrium exists when the players’ strategy choices result in a Nash equilibrium in every subgame. That means that each player’s strategy is to optimize the expected payoff at each stage in each period, given the current information and beliefs in each move.

### 3.1 The Incentive Rate for the Expert-Only Service

We first need to know the minimum hourly rate that can motivate the expert to produce the expert-only service. Obviously, the expert must somehow be rewarded if he or she should have an incentive to produce the expert-only service. Otherwise, the expert will attempt to use a mixed service at lower cost. There exists an hourly rate for which the payoffs of the two types of services are the same for the expert in every period. Above this rate, the payoff from performing the expert-only service is strictly higher than that of the mixed service payoff. Below this rate, the payoff from performing the mixed service is higher. We call this rate $$p^*_n$$, the minimum incentive rate, and the rate above, $$p''_n$$, the expert-only rate. At $$p^*_n$$, an expert is indifferent in producing either service. The relationships are formulated in the
following equations:

\[
\Pi_i^H(p^n_i, P^n_i, s') > \Pi_i^H(p^u_i, P^u_i, s') \quad \text{for} \quad p_i' = p_i'' > p_i^*, \quad \Pi_i^H(p_i^*, P_i^*, s') = \Pi_i^H(p_i^*, P_i^*, s') \quad \text{for} \quad p_i' = p_i^*,
\]

(3)

There is a slight difference between the minimum incentive rate and the quality-assuring price discussed in Shapiro (1983) and Allen (1984). The quality-assuring price is the minimum price that motivates a firm to produce high-quality goods or services. By contrast, at the minimum incentive rate in our paper, the expert may provide the expert-only service or the mixed service.

When an expert provides the mixed service but still charges at his/her hourly rate, overbilling occurs. Table 1 summarizes the service types and billing practices if the hourly rate is above, at or below the minimum incentive rate. Because all of the N experts are identical, we state that \( BP_i = BP_i = LH_i = LH_i = c_i = c_i \) for notational simplicity.

<table>
<thead>
<tr>
<th>Service Type</th>
<th>Billing Practice</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expert-only</td>
<td>No overbilling</td>
</tr>
<tr>
<td>Either</td>
<td>May overbill</td>
</tr>
<tr>
<td>Mixed</td>
<td>Overbill</td>
</tr>
</tbody>
</table>

Furthermore, we evaluate the two payoffs in Equations (1) and (2) at \( p_i' = p_i'' \) and \( T \to \infty \). According to Equation (3), we obtain:

\[
p_i^* = p_i' + \frac{1}{\rho} \left( c_i'' - c_i' \right), \quad p_i'' > p_i^*, \quad p_i' > p_i^*.
\]

(4)

The properties of \( p_i^* \) and \( p_i'' \) are described in Proposition 1.

**Proposition 1.** The hourly rate that motivates an expert to provide the expert-only service should be higher than the minimum incentive rate, \( p_i^* \). The minimum incentive rate, and rates above, (i) increase as the cost of producing the expert-only relative to the mixed service increases; and (ii) decrease as the audit percentage for the bills by the customers increases. At \( p_i^* \), the expert will produce either the expert-only or the mixed service.
3.2 Overbilling Choice in Equilibrium

The service choice of an expert will be influenced by the market competition. In this section we shall see that the market competition can eliminate the profits necessary to motivate an expert for the expert-only service. Let us consider a simple framework in which experts compete for market share in a finite repeated game. In the basic framework, expert \( i \) is always hired in period \( t = 1, \ldots, T \). When the expert is discovered producing the mixed work, the customer reduces the overcharged amount to the appropriate amount, but still continues the business relationship. Figure 1 displays the payoffs for expert \( i \) in a single period. The expert sets the hourly rate at \( p'' \), or \( p^s \) or \( p' \in (p^s, p^s) \). In addition, the expert decides whether to perform the expert-only service, \( H \), or the mixed service, \( L \). Table 1 tells us that the expert will always perform the expert-only service if \( p' = p^s \), and the mixed service if \( p' < p^s \). This is shown in Figure 1.

With a finite horizon of \( T \) periods, the experts have an incentive to play non-cooperatively in the last period, and therefore in the next to last period, and so on. So the interactions become the one-shot game, repeated \( T \) times. Denote \( (a) = \Pi_i(p') \), \( (b) = \Pi_i^n(p^s) \), \( (c) = \Pi_i^n(p'') \). The equilibrium hourly rate is the one that provides the maximal payoffs of \( (a), (b), \) and \( (c) \). If the expert chooses to produce the mixed service at \( p' = p^s \), the overcharged amount will be \((1 - \rho)(p^s - p')\). According to Equation (4), the overcharged amount becomes
Proposition 2. In a finite repeated game where experts compete for market share, in the absence of collusive strategies, the equilibrium hourly rate and service choice is one of the following:

1. If \( c_i > \max\{a, b\} \), the equilibrium hourly rate is \( p'' > p^* \), and the expert produces the expert-only service.
2. If \( b > \max\{a, c\} \), the equilibrium hourly rate is \( p^* \) and the expert produces either the expert-only service or the mixed service. The overcharged amount is \( (1/\gamma)(c^* - c^+) \) per unit if the expert provides the mixed service.
3. If \( a > \max\{b, c\} \), the equilibrium hourly rate is \( p' \in (p^*, p^+) \), and the expert always produces the mixed service. The overcharged amount is \( (1-p)(p' - p^*) \) per unit.

Figure 1: Expert \( i \)'s Service Choices

Proposition 2 shows how the service choice of an expert is determined by the equilibrium hourly rate. If the rate in equilibrium is driven to or below the minimum incentive rate, an expert will provide the mixed service and overbilling occurs. In
addition, Proposition 2.2 indicates that the overcharged amount is inversely related to the auditing percentage $\gamma$. This suggests that the customers should audit a high percentage of the bills to attack the problem of overbilling.

4 The Threat of Business Termination

If the mixed service is detected in the continual business relationships, customers can announce that they will not give the expert any future business. Is this strategy a credible threat to warrant that the expert will always provide the high-quality services? This section will show that such a mechanism is still not capable of preventing billing abuses.

It is appropriate to model the above interactions in an infinitely repeated game ($T = \infty$). The timing of events in each period is as follows. In stage 1 the experts set their hourly rates simultaneously. In stage 2 each customer hires an expert and pays the expert’s rate, or chooses not to hire at all. All customers learn the experts’ service choices and can use this information in subsequent periods for future hiring decisions. In stage 3 the experts then choose to produce either the expert-only service or the mixed service. An expert does not produce any service if there are no customers.

The previous literature shows that, when quality is a choice variable, firms have an incentive to produce high quality items to maintain their reputation (Klein and Leffler, 1981; Shapiro, 1983). On the contrary, Kranton (2003) argues that price competition can eliminate the price premium needed to induce firms to produce high-quality goods and services. Kranton (2003) applies the method of Abreu (1988) and asks whether high-quality production on the equilibrium path can be supported by a punishment. This punishment will give a firm the worst payoffs possible in any perfect equilibrium if the firm produces low-quality products. Then by Propositions 4 and 5 in Abreu (1988), if high-quality production cannot be sustained by such a punishment, it cannot be sustained by any punishment.

Our methodological approach in this section is drawn from Kranton (2003). It differs from Kranton (2003) in that our stylized payoff functions enable us to analyze the auditing reductions, which are internalized by the experts while making strategic moves. In the model there are two strategic variables for an expert in each
period: the hourly rate choice and the service choice. The model focuses on non-collusive hourly-rate strategies in which experts cannot tacitly or overtly collude on the hourly rate.\textsuperscript{4} Any deviating expert can only be punished for service choice, and not hourly rate choice. This setting allows the model to examine whether the threat of business termination by customers is sufficient to ensure that the expert offers the expert-only service.

The first step is to find the punishment that can offer the worst payoff possible as suggested by Abreu (1988). As proved by Abreu (1988), if the expert-only service cannot be sustained by the worst-payoff punishment, it cannot be sustained by any other punishment. Consider the following punishment code: Once an expert performs the mixed service, the customer refuses to pay any hourly rate above the cost of the mixed service \(L^c\). There is no point then for the expert to continue to provide the mixed service willingly for this customer, and derives zero profit forever. In other words, the future business relationship between the expert and the customers ceases to exist. All consumers refuse to hire the “notorious” expert in the future periods. All other experts reset their hourly rates and the market share. This punishment gives an expert who performs the mixed service zero profit in any period. Since zero profit is the worst payoff in any subgame perfect equilibrium, this punishment becomes the optimal penal code.

**Lemma 1.** The optimal penal code is the punishment which results in an expert earning zero profits in all future periods \(\tau \geq t + 1\) if the expert produces the mixed service in period \(t\).

When the hourly rate is \(L^c\), expert \(i\) has no incentive to produce the expert-only service. Customers would not hire the expert at an hourly rate higher than \(L^c\) to obtain a negative consumer’s surplus, because customers believe the expert would provide the mixed service in future periods. In addition, if the expert attempts to set the hourly rate above \(L^c\), he/she will offer a strictly lower consumer’s surplus so that there will be no customers. Setting a rate below \(L^c\) gives the expert a negative payoff. Therefore, the expert is locked in this rate for all periods \(\tau \geq t + 1\).

With Lemma 1 established, we now determine under what condition the

\textsuperscript{4}Extensive studies have shown that a wide range of strategic interactions, including non-collusive and collusive behaviors, can occur in an infinitely-repeated games. See Fudenberg and Tirole (1991). In addition, Feuerstein (2005) provides a survey of the recent literature.
production of the expert-only service can or cannot be sustained. Figure 2 presents the sequence of moves of experts and consumers in a single period. Similar to Kranton (2003), the model considers a one-period deviation by an expert in the pricing stage. In period $t$ an expert decides whether to deviate from the expert-only hourly rate $p^n$ to $p^d = (p^d, p^d)$. In period $\tau \geq t+1$ the hourly rate returns to $p^n$. The market share increases to $s^d$ in period $t$, and returns to the mean level $\bar{s}$ in period $\tau \geq t+1$. Customers observe these rates and decide whether to hire (Hire) or not (Don’t). When hired, expert $i$ decides whether to perform the expert-only service, $H$ or the mixed service, $L$. The interaction results in four outcomes: $(g)$ through $(n)$. The outcomes whereby the customers choose not to hire (Don’t) delineate the optimal penal code in Lemma 1. Furthermore, the possible Nash equilibrium at which the expert charges the expert hourly rate $p^n$ and provides the expert-only service is $(n)$. In other words, we will examine whether this Nash equilibrium, on the infinite equilibrium path, can be supported by the off-equilibrium path, which is the optimal penal code.

\[\begin{array}{c|c|c}
\text{Customer} & p^d \in (p^L, p^R) & p^H \\
\hline
\text{Don’t} & 0 & \text{Hire} \\
\text{Hire} & \Pi^E(p^d) & \Pi^H(p^H)
\end{array}\]

\[\begin{array}{c|c|c}
\text{Customer} & p^d \in (p^L, p^R) & p^H \\
\hline
\text{Don’t} & 0 & \text{Hire} \\
\text{Hire} & \Pi^E(p^d) & \Pi^H(p^H)
\end{array}\]

Figure 2: The Threat of Business Termination

The game is solved by backward induction. We compare the payoffs of $(g)$ to $(h)$, and $(m)$ to $(n)$, respectively. Expert $i$ maximizes the payoffs and
makes the service choice accordingly. When the hourly rate is lower than the
minimum incentive rate, the expert has no incentive to produce the expert-only work.
So the possible outcomes are \((g)\) : offering the mixed service, and \((n)\) : offering
the expert-only service, as underlined in Figure 2. We continue to solve the game
backwards to see the expert’s deviation choice in the hourly rate. The continuation
payoffs of \((g)\) and \((n)\) are:

\[
\begin{align*}
(g) : & \quad \Pi^t(p^g) = (p^g - c^g)s^t - \gamma(p^1 - p^g)(1 - \rho)s^t \\
& \quad + \sum_{t=1}^{\infty} \delta^t [(p^g - c^g) s^t - \gamma(p^1 - p^g)(1 - \rho)s^t].
\end{align*}
\]

\[
\begin{align*}
(n) : & \quad \Pi^t(p^n) = (p^n - c^n)s^t + \sum_{t=1}^{\infty} \delta^t [(p^n - c^n) s^t].
\end{align*}
\]

The condition that \((g) > (n)\) is:

\[
(p^g - c^g)s^t - \gamma(p^1 - p^g)(1 - \rho)s^t + \frac{\delta}{1-\delta} [(p^g - c^g) s^t - \gamma(p^1 - p^g)(1 - \rho)s^t] > (p^n - c^n)s^t + \frac{\delta}{1-\delta} (p^n - c^n)s^t,
\]

which is re-arranged as:

\[
\begin{align*}
(p^g - c^g)s^t - \gamma(1 - \rho)(p^1 - p^g)s^t - \gamma \frac{\delta}{1-\delta} (p^n - p^g)(1 - \rho)s^t > (p^n - c^n)s^t - \frac{\delta}{1-\delta} (c^n - c^g)s^t.
\end{align*}
\]

When Equation (5) holds, an expert has the incentive to undercut the hourly
rate from \(p^n\) to \(p^g\) and provides the mixed service. The economic meanings of
the terms in Equation (5) are the following:

<table>
<thead>
<tr>
<th>#</th>
<th>Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1</td>
<td>The profit of the increased market share in period (t) when undercutting the hourly rate from (p^n) to (p^g)</td>
</tr>
<tr>
<td>L2</td>
<td>–the reduction of the caught overcharges in period (t)</td>
</tr>
<tr>
<td>L3</td>
<td>–the present value of the reduction in the caught overcharges for all future periods</td>
</tr>
</tbody>
</table>
Equation (5) characterizes an opportunistic expert calculating the benefits to make the service offer. The profit from the increased market share, subtracting any reduction due to the auditing effort, motivates the expert to deviate in period $t$. Besides, the market share $s^t$ on the left-hand side of Equation (5) carries another interesting meaning: the deviation of the market share from the mean market share measures the responsiveness to the hourly rate deviation. In a market where there is weak customer loyalty or low locked-in costs for customers, a small price cut can result in a large increase in the market share. Therefore, it is relatively easy for an expert to convert customers from the rival by cutting the hourly rate. The lower hourly rate will cause the expert to perform the mixed service.

Lemma 1 and Equation (5) together lead us to conclude that, when experts compete for market share, the equilibrium path of providing the expert-only service and charging the expert-only rate cannot be sustained. This result is summarized in the following proposition:

**Proposition 3.** In the absence of collusive hourly-rate strategies, the threat of business termination does not ensure that an expert will perform the expert-only work if Equation (5) holds.

Since the customer-enforcement mechanism is not capable of preventing billing abuses and less-than-expert services, a systematic mechanism such as industry self-regulation or a third-party enforcement may be needed to regulate the issue.

## 5 Conclusion

Although it is believed that the majority of attorneys practice billing according to ethical standards, many sources have reported that billing abuse is a prevailing issue in legal services. Similar issues exist in many other professions, such as in medical or repair fields in which the service providers have superior knowledge over the
customers. I have modeled the service choice and hourly rate competition on a repeated game. In the model a lawyer can mix in services that do not require the lawyer expertise and still charge a single hourly rate, which presents the opportunity for overbilling. I demonstrated that billing abuses may exist in equilibrium.

There are two mechanisms for the customers to prevent less-than-expert services: auditing and termination of the business relationship. The model showed that the auditing capacity helped the customer to reduce the over-payment, but it did not deter the lawyer from billing abuses. The threat of business termination is, again, not a credible threat to enforce the service quality. If the temporary profit justifies the loss in the long-term profit, the lawyer may choose to charge a lower hourly rate, which results in the less-than-expert service that costs the lawyer less to produce. Since the problem of overbilling always exists, the customers have to enhance the auditing capacity to reduce the overcharge, and have to audit a high percentage of the bills.

References


Defense Research Institute.


