Winter 2018

In Defense of the Billable Hour: A Monitoring Theory of Law Firm Fees

Jonathan H. Choi
University of Minnesota Law School

Follow this and additional works at: https://scholarcommons.sc.edu/sclr

Part of the Law Commons

Recommended Citation
IN DEFENSE OF THE BILLABLE HOUR:
A MONITORING THEORY OF LAW FIRM FEES

Jonathan H. Choi*

I. INTRODUCTION

The billable hour has an image problem. It has become a scapegoat for all the unpleasantries of law firm life: long hours, dull work, cantankerous clients. Associates see the billable hour as the most visible symbol of their bondage, an arch-capitalistic machine that transforms time into dollars and bright-eyed young lawyers into fee-producing zombies.

So it should come as no surprise that lawyers and commentators have long predicted the abolition of the billable hour. It skews incentives, they say; it encourages inefficiency; it’s just plain unpleasant! These calls have accelerated in recent years, in part due to widely publicized law firm misconduct in reporting hours.¹ All kinds of lawyers have joined the crusade,


from in-house counsel to the heads of some of the largest law firms in America.\(^2\)

At the same time, the billable hour has proven remarkably resilient against campaigns for reform. Critics have rallied to abolish the billable hour virtually since its invention.\(^3\) For the past twenty years, legal academics and professionals have routinely predicted the imminent demise of traditional billing practices.\(^4\) Yet, the billable hour is still by far the most popular billing method today.\(^5\)

To resolve this paradox, this Article delves into the history of the billable hour, as well as the economics literature on agency theory. Prior literature has suggested that law firms bill by the hour because they cannot bear the risk of cost overruns, and that the use of the billable hour is determined by the tradeoff between risk allocation and incentive costs.\(^6\) This Article instead proposes that the billable hour is a monitoring tool: a means by which a client can impose discipline on its lawyer and compare the services of different lawyers. This explains the timing of the billable hour’s adoption—it became the standard in the 1970s, just as large, sophisticated in-house legal departments appeared and began to agitate for more cognizable metrics of lawyer performance.

---


3. For an early example, see generally Harry A. Ackley, Let’s Throw Out the Hourly Basis for Fees, 49 A.B.A. J., Jan. 1963, at 76, 76 (arguing that “setting fees on the basis of the amount of time spent on a case is improper”).


6. See infra Section II.A.
Monitoring also explains the pattern of billable hour usage in law firms today. Alternative billing methods—flat fees, contingent fees, etc.—are most common in engagements with individual clients and in engagements that require complex, niche work. What do individuals and those with esoteric legal needs have in common? They are all relatively unfamiliar with the work of the lawyers they hire, which makes it difficult to interpret the hourly bills that such lawyers produce. Moreover, because these clients are less likely to offer repeat business, they have less leverage to keep lawyers in line using the information they obtain from hourly bills. Because these individual clients cannot exploit the key monitoring function of the billable hour, they turn to alternative fee arrangements.

The monitoring theory of billable hours has significant practical implications. It suggests that if in-house legal departments become larger and more sophisticated, they might utilize billable hours more than ever before. This contradicts the standard prediction that more efficient, modern in-house counsel will no longer need the billable hour.\(^7\)

Above all, this Article argues that the billable hour serves a legitimate function that almost all of its skeptics have overlooked. Far from being a bugaboo that distorts incentives and enslaves associates, the billable hour will be the appropriate compensation structure for a wide swath of clients and firms. Critics of the billable hour should acknowledge its uses as well as its abuses.

II. **Monitoring and Agency Theory**

We should begin by dispelling the idea that the hourly bill is a *fait accompli*, paid by clients without question. The reality is more complex—an itemized hourly bill conveys crucial information about the lawyer’s efforts and provides a medium for the client to direct those efforts. This is especially important when firms have repeat interactions with clients, because repeat clients will have greater leverage to push back against bills that they find unreasonable.

Consider the client–firm interaction under ideal conditions for monitoring, where an expert in-house attorney routinely contracts with an outside law firm for legal services. The in-house attorney typically has experience working at a law firm,\(^9\) meaning that she understands both sides

---

7. This Article uses “billable hours,” “fixed fees,” and “contingent fees” as synecdoche for the fee agreements that utilize them.

8. See, e.g., Ribstein, supra note 4, at 770.

of the transaction and is aware of the potential abuses of hourly billing. Moreover, she knows the extent of her legal needs before approaching outside counsel in the first place. Critics sometimes assume the in-house attorney to be powerless in the face of avaricious firm lawyers—but this view vastly understates the sophistication of the modern corporate legal department. The firm and the corporation often have a robust existing relationship that the law firm has worked hard to cultivate. The threat that the client will withdraw future business if the lawyer abuses the hourly billing system gives the client considerable leverage.  

The hourly bill conveys two main pieces of information to the in-house counsel. It tells her the outside lawyer’s hourly rate, and it tells her how the outside firm has spent its time on a matter. The in-house counsel will have an idea of how much time a matter should take, particularly if it is a routine issue and if she has contracted similar work to the law firm before. As a result, she knows what range of charges is reasonable; often, she will specify a ceiling on the number of hours that a law firm may spend on a matter or require that the firm seek authorization before exceeding a certain spend. If she finds the fees unreasonable, she can dispute them. Finally, if she feels the firm has failed to work efficiently, she can simply refuse to hire it again on future matters. 

Particularly in the modern, highly competitive legal market, corporate clients do not hesitate to use all of the above tactics in order to extract the best deals from law firms. For example, law firms are frequently forced to “write down” their final bills when clients complain that the bills are unjustifiably high—anticipating this, many firm partners will also “write off” the hours of their associates even before submitting a bill to a client. In-house counsel can effectively enforce discipline on its outside counsel by virtue of their

---

10. One might wonder why this should be so if contracts with outside counsel are executed at market-clearing rates. One answer is that they are not, and that law firms are paid “efficiency wages” to give them an incentive to do good work and to maintain reputations for efficiency. See generally Efficiency Wage Models of the Labor Market (George A. Akerlof & Janet L. Yellen eds., 1986) [hereinafter Akerlof & Yellen] (exploring efficiency wage models as a means to understanding fluctuations in the labor market). Another answer is that there are fixed costs of acquiring a client relationship, and the profit of corporate work will generally be on repeat business rather than the first matter a law firm works on. Thus, the client will have power over the law firm analogous to that in a hold-up problem because the law firm has already made considerable investments in the client relationship. See generally Faruk Gul, Unobservable Investment and the Hold-Up Problem, 69 ECONOMETRICA 343, 344 (2001) (outlining the tendency of businesses to invest efficiently creating an interaction between unobservable investment and dynamic bargaining).  

repeat relationship and the value of the corporation’s business. Therefore, we should discard any cartoonish vision of an impotent corporate counsel simply yielding to billing abuses by law firms. In-house attorneys actively glean information from hourly bills and use that information to direct the conduct of outside counsel.

We can clarify this point through the lens of agency theory. Agency theory is the study of the relationship between principals (here, clients) and agents (here, lawyers). The principal must trust the agent to represent its interests in the absence of perfect information. However, we assume the agent to be self-interested; thus, the dominant tension in agency theory is the difficulty in encouraging the agent to act as altruistically as possible without imposing unduly burdensome constraints on either party.

In this framework, monitoring is one example of an “agency cost”—a cost of aligning principal and agent incentives or failing to do so. The traditional definition of agency costs is: “the sum of: (1) the monitoring expenditures by the principal, (2) the bonding expenditures by the agent, [and] (3) the residual loss.”

Bonding expenditures are expenditures to promote agent loyalty—for example, forgone opportunities when the agent signs a non-compete contract. The residual loss is the cost imposed on the principal when the agent fails to serve the principal’s interests—for example, the cost of lost productivity when an employee shirks at work. And most important to my theory is monitoring, which “includes more than just measuring or observing the behavior of the agent. It includes efforts on the part of the principal to ‘control’ the behavior of the agent through budget restrictions, compensation policies, operating rules[,] etc.”

It is important to recognize that we cannot entirely eliminate agency costs—we could attempt to reduce residual loss by undertaking costly bonding, for example, or we could accept some amount of residual loss in order to avoid the costs of monitoring, but each case involves a tradeoff.

---

12. Note, however, that clients are not necessarily all corporations. Private equity firms, for example, are highly valued clients with considerable leverage because they bring so much repeat business. David Lat, When $1,000 an Hour Is Not Enough, N.Y. TIMES: DEALBOOK (Oct. 3, 2007, 7:12 AM), http://dealbook.nytimes.com/2007/10/03/when-1000-an-hour-is-not-enough.
15. See id.
16. Id. at 308 n.9.
between two competing values. The central problem of agency theory is to minimize overall agency costs. Monitoring is merely a reasonably inexpensive means to reduce agent misconduct in the context of law firm billing.

To see why, consider the two most popular alternatives to the billable hour: the fixed fee and the contingent fee. The fixed fee is a flat amount negotiated up front, to be paid regardless of the time the lawyer spends on the matter and regardless of the outcome.17 On the other hand, the contingent fee (sometimes called a “success fee”)18 is paid as a function of positive client outcomes—for example, as a percentage of a plaintiff’s recovery at trial.19 Each method brings an idiosyncratic mix of incentive based tradeoffs. Many commentators miss this point, criticizing billable hours for their incentive effects without acknowledging the fundamental impossibility of costlessly aligning lawyer and client objectives.20

Fixed fees, for instance, encourage efficiency but discourage effort. In that respect, they are the inverse of hourly fees—because flat-fee lawyers do not benefit when their clients obtain better outcomes, such lawyers will tend to shirk additional work because such work marginally benefits their clients but not themselves. The degree of their shirking may be limited by their desire to maintain good reputations, by honesty, and by professional pride—but all of these factors should prevent the abuse of billable hours as well, and we know from experience that incentives will always influence practitioners on the margins. Just as billable hours might encourage a litigation team to extend a trial in order to pad its fees, a fixed fee arrangement might encourage a litigation team to settle early in order to avoid additional, uncompensated work. Both approaches present hazards in opposing directions.21

19. 1 BOUVIER LAW DICTIONARY, supra note 17, at 1063.
21. Note that what may be inefficient for clients may be socially efficient, insofar as litigation imposes negative externalities on opposing counsel—who generally must process and respond to every legal action undertaken by the other side—and the legal system—which will provide court time, judge expertise, etc. without being reimbursed for the full extent of trial costs. Thus, as a social policy, we might think it a good thing to encourage lawyers to shirk by settling early, but such a discussion exceeds the scope of this Article. See generally Steven Shavell, The Fundamental Divergence Between the Private and the Social Motive to Use the
Contingent fees might initially seem like an ideal method to align the interests of lawyers and clients—after all, the lawyer will be paid only if her client wins. But upon closer consideration, contingent fees still create significant moral hazards. Consider a case where a lawyer will receive one third of the plaintiff’s winnings at trial. The lawyer has an opportunity to negotiate an early settlement; she could pursue a more favorable one, but she estimates it would take fifty additional hours to do so, in which time she could instead earn $10,000 by pursuing another case. The estimated gain to the client from further work is $24,000, meaning that the estimated gain to the attorney is only $8,000. Based on these facts, a purely self-interested lawyer will choose the easy settlement, to the detriment of her client. So, like the fixed fee, the contingent fee also encourages shirking because the lawyer must bear all marginal costs of her labor while receiving only part of the benefits. Contingent fees provide more of an incentive for attorney effort than fixed fee arrangements, but not as much as the hourly fee. Consequently, they generally align principal-agent incentives better than flat fees. On the other hand, contingent fees are also more complex to negotiate, easier for the agent to manipulate because of their complexity, and more costly in terms of their enforcement.

Finally, billable hours encourage thoroughness but discourage efficiency. In general, lawyers billing by the hour will try to overestimate the extent of legal work and complete matters as slowly as possible. In terms of agency costs, the billable hour system incurs residual losses when the marginal cost of

---

*Legal System, 26 J. Legal Stud. 575, 575–76 (1997) (discussing the expensive nature of the legal system as a social institution being too costly and socially inappropriate).*

22. The lawyer’s expected marginal gain from the settlement is $8,000, which is less than the $10,000 she could make elsewhere.

23. The only way that a lawyer would not have an incentive to shirk would be for her to receive 100% of the gains, perhaps by buying the client’s litigation interest in its entirety. However, the sale of litigation rights is uncommon and of questionable legality. See Binyamin Appelbaum, *Investors Put Money on Lawsuits to Get Payouts*, N.Y. Times (Nov. 14, 2010), https://www.nytimes.com/2010/11/15/business/15lawsuit.html; see generally Winand Emons, *Expertise, Contingent Fees, and Insufficient Attorney Effort*, 20 Int’l Rev. L. & Econ. 21, 21–23 (2000) (providing a discussion on the theories supporting hourly fees being economically efficient).

24. In the context of litigation, a lawyer could theoretically fully align her incentives with a plaintiff if the lawyer were to purchase 100% of the legal claim from the plaintiff. However, litigation finance arrangements of this sort are not generally pursued in the United States—law firms typically lack the liquidity to fund such arrangements, which are typically driven by third parties such as hedge funds, and anti-champery rules prohibit the assignment of lawsuit claims in many jurisdictions. See Jonathan T. Molot, *Litigation Finance: A Market Solution to a Procedural Problem*, 99 Geo. L.J. 65, 95–96 (2010); Anthony J. Sebok, *The Inauthentic Claim*, 64 Vand. L. Rev. 61, 70 (2011); Joanna M. Shepherd, *Ideal Versus Reality in Third-Party Litigation Financing*, 8 J.L. Econ. & Pol’y 593, 594–95 (2012).
of an additional hour exceeds the marginal benefit to the client. In contrast, the residual loss of fixed and contingent fees is suboptimal effort exerted by the lawyer. These residual costs are different in kind rather than degree, which makes comparison between them difficult—but not impossible. A key difference between excessive billing and suboptimal effort is that monitoring can remedy the former but not the latter. Using the information provided in an hourly bill, clients can effectively reduce their residual loss for inefficient work. This is because a lawyer’s hourly bill must explicitly state the amount and nature of the work undertaken. The lawyer cannot manipulate these statements without committing outright fraud—and even then, an expert in-house client can often tell when hours have been inflated.\(^{25}\) In contrast, the quality of legal work is difficult for the client to discern. Legal work is to a large extent a “credence good”\(^{26}\)—its quality will only become apparent long after it has been paid for, if ever. For example, a plaintiff might never know that a flat-fee lawyer could have obtained a larger recovery through trial than through early settlement. In transactional work, even an expert client might not discover an error in due diligence until long after a deal has closed. Thus, clients are unable to monitor the quality of legal services, making it difficult to mitigate the residual losses of fixed fees.

As a monitoring tool, the billable hour is not very costly—in fact, many law firms informally track hours even prior to the rise of the billable hour in the 1970s,\(^{27}\) and firms that do not bill by the hour often require lawyers to internally report hours even on matters that are billed using alternative fees.\(^{28}\) So, the marginal agency cost of the billable hour is relatively small, and for the right client, it can avoid both the incentive problems of alternative fees and the potential for agent inefficiency.

In fact, clients paying hourly fees will benefit from monitoring even before an engagement has begun. Just as they must monitor lawyers during the engagement to prevent opportunistic overwork, they must also monitor lawyers prior to the start of the engagement to avoid opportunistic recommendations of unprofitable work.

\(^{25}\) At the very least, the client should be able to tell when one firm has inflated its hours out of proportion with other firms. It is reasonable to think that the market for outside legal work is subject to a certain amount of “puffery,” whereby all firms will systematically overstate their hours by a certain amount. Clients can counter this by negotiating for lower hourly rates in anticipation of the puffery, or they can just accept puffery as a component of an efficiency wage, Akerlof & Yellen, supra note 10, at 106.

\(^{26}\) The term was first introduced by Michael R. Darby and Edi Karni in their seminal paper, Free Competition and the Optimal Amount of Fraud, 16 J.L. & ECON. 67, 81 (1973).

\(^{27}\) See Jones & Glover, supra note 4, at 294.

\(^{28}\) Wachtell Lipton is one example. See infra Section III.A.
For example, imagine that a plaintiff is deciding whether to pursue a lawsuit with an expected recovery of $1,500. Lawsuits of this type are either high-effort, with expected legal fees of $3,000, or low-effort, with expected fees of $1,000, with a 50% probability of the lawsuit falling into either category. Expert clients (those capable of monitoring) know whether the case is high-effort or low-effort, but non-expert clients do not know. Lawyers can perceive whether a case is high-effort or low-effort, but they do not share this information with the client, creating an information asymmetry.29

Even assuming that lawyers do not bill more hours than necessary, a non-expert client in this hypothetical ought not to pay on an hourly basis. This is because the lawyer will opportunistically accept any case on an hourly basis, even if the case is unprofitable for the client. As a result, the expected cost to sue will be $2,000 (the average of $1,000 for the low-labor matter and $3,000 for the high-labor matter), which is greater than the expected recovery. However, if the client solicits lawyers to take the case on a fixed fee basis for $1,000, then lawyers will only take the lawsuit if they determine it to be a low-effort lawsuit, and the client will have an expected profit of $500 (the $1,500 recovery minus $1,000 in fixed fees). Therefore, the non-expert client will prefer fixed fees in order to force lawyers to differentiate between low- and high-effort cases. This result is a form of bonding—aligning the lawyer’s incentive to select low-effort cases with the client’s interest in not pursuing high-effort cases. However, the expert client need not bond because it can monitor the lawyer’s recommendations—it can select the low-effort cases and only pursue those. The expert client is therefore indifferent between hourly and fixed fees on the basis of case selection—it will instead select fee arrangements on some other basis, perhaps to avoid the agency costs of credence goods, as discussed above.

None of this is to say that monitoring is the only way to minimize agency costs. Clients might instead opt for bonding rather than monitoring—for example, by relying on firm reputation as a guarantee of work quality.30 Or a client could simply tolerate residual losses as less costly than attempts to bond or monitor.

---

29. This hypothetical is a variant on one offered by James D. Dana, Jr. & Kathryn E. Spier, Expertise and Contingent Fees: The Role of Asymmetric Information in Attorney Compensation, 9 J.L. ECON. & ORG. 349, 352–53 (1993).

30. Professional reputations are costly to build and maintain because they require practitioners to accrue reputational capital and to ensure that their own agents (partners and associates at the same firm) do not erode that capital. For a fuller description of the theory of reputational bonding, see Larry E. Ribstein, Ethical Rules, Agency Costs, and Law Firm Structure, 84 VA. L. REV. 1707, 1714–20 (1998).
Given the choice between these strategies, we might wonder what set of tradeoffs help lawyers and clients to select the appropriate billing method. The answer will be useful in two respects: it will help us to understand the advantages and disadvantages of each approach, and it will help us to predict shifts in which firms use which methods. All the methods discussed are used by at least some lawyers in the status quo, although hourly billing predominates. What considerations prevent firms on the margins from adopting hourly billing, and what considerations have led the billable hour to its place of pride in the legal profession?

A. The Risk-Incentives Tradeoff

One popular school of thought in agency theory argues that the equilibrium is determined by the tradeoff between risk appetite and incentives.31 According to this school, the principal is generally better positioned to bear risk, but the agent must also bear some of the risk so that the agent is incentivized to do good work.32 In the words of one leading article attempting to explain the prevalence of hourly billing:

Our model suggests that the optimal contract will be influenced by a balancing of two concerns: efficient risk distribution and limiting “moral hazard”—the moral hazard is the danger that a fixed-fee contract will induce the lawyer to conduct too little work and that an hourly contract will induce excess work. Economic forces will encourage the client and lawyer to choose the contract type that offers the lowest sum of risk costs and costs from moral hazard.33

Using a real-world example, suppose that a construction firm wishes to incentivize its employees to work diligently, without having to constantly monitor their performance. One way that it might do so is to link compensation to output—for example, it could pay its welders ten cents for each rivet they weld. But what if the rate at which welders work depended on factors outside their control? What if, for example, welders worked only half as quickly on a cold day or only a quarter as quickly in the rain? The welders might be uncomfortable having their pay fluctuate by a factor of eight based on random environmental variables. So, in order to attract risk-averse welders to do the same work for uncertain compensation, the construction firm would have to pay a risk premium.

The risk premium is an additional amount of money required to compensate employees for taking on risk. Imagine a welder who is indifferent between a risk-free salary of $X and a risky—perhaps output-based—salary of $Y. If the welder is risk averse, $Y will be greater than $X. The risk premium is thus $Y − $X, quantifying the cost of risk to the welder. A core premise of the risk-incentives model is that the risk premium will generally be greater for the employee than for the employer because the employer has more financial reserves to smooth the impact of output volatility. (In real-world terms, a large construction firm could probably weather a few weeks or even months of delays, but many welders would immediately suffer if their pay were reduced for even a week.) Because the risk premium is greater for the employee, the conventional Coasean move is to assign the risk to the lowest-cost bearer. 34 To do otherwise incurs an efficiency cost, determined by subtracting the employer’s risk premium from the employee’s risk premium. This cost must be weighed against the efficiency gain from aligning principal and agent incentives through compensation.

So, under the risk-incentives model, the construction firm will only attempt to match employee compensation to output until the marginal benefit from doing so equals the marginal risk premium it must pay its workers. Thus, it might institute an hourly wage accompanied by a bonus at the end of the

34. The idea that the law should assign risks to their lowest-cost bearer was famously spearheaded by Guido Calabresi. See generally GUIDO CALABRESI, THE COST OF ACCIDENTS: A LEGAL AND ECONOMIC ANALYSIS (1970) (discussing accident costs and how to effectively reduce them); Guido Calabresi, Some Thoughts on Risk Distribution and the Law of Torts, 70 YALE L.J. 499, 500 (1961) (examining the theoretical justifications of different methods of distributing losses). Calabresi built on Ronald Coase’s massively influential 1960 paper, The Problem of Social Cost. Coase argued that, absent transaction costs, parties should bargain over the assignment of rights in order to efficiently satisfy mutual interests. See R.H. Coase, The Problem of Social Cost, 3 J.L. & ECON. 1 passim (1960). For example, absent other incentive problems, the welder should be willing to take a lower salary in order to avoid having to bear the risk of salary fluctuations, and the construction firm should take the deal.
project based on the number of rivets the welders have completed. The classical example of blended incentive pay and flat pay is the compensation of white-collar employees, who are often paid a combination of a guaranteed salary and a contingent bonus. The idea that compensation packages respond to risk tolerance jives with our anecdotal experience with corporations, where non-executive employees—who one expects to have lower risk tolerances because their total wealth will generally be lower—are mostly paid fixed salaries; in contrast, highly paid executives, who should be better positioned to tolerate swings in compensation, are generally paid a higher percentage of their compensation in the form of incentive pay.35

One might analogize that the law firm is like the welder or the CEO: the client wants to incentivize the lawyers’ performance, but the lawyers (true to type, risk averse) want to place the risk of cost overruns on the client. Consequently, the optimal employment contract will depend on the relative risk tolerances of each party—the greater the relative risk aversion of the law firm, the more likely it is that the law firm will demand an hourly fee (or demand a large risk premium to use fixed fees). The greater the risk aversion of the client, the more likely that the client will demand to pay a fixed fee (or the higher the risk premium the client will be willing to pay). If the client’s and law firm’s risk tolerances are equal, they should be collectively indifferent between different fee arrangements, so long as the party bearing the risk is properly compensated.

This story appeals to legal scholars, who are familiar with the conceptual framework that one ought to assign risk to its lowest cost bearer.36 It provides a theoretically appealing equilibrium that could potentially explain otherwise puzzling real-world variation in billing structures. Most importantly, it comports with the lived experience of lawyers—one major objection from practitioners considering the shift toward alternative fees is that they are uncomfortable taking on the risk that a matter will be more expensive than initially budgeted.37

35. There are other explanations, of course: low-level employees are unlikely to affect overall corporate performance, so performance pay may simply be less effective for them than for top executives. A more nuanced modern account suggests that stock options for non-executives may encourage performance by encouraging workers to monitor the performance of their peers and to enforce group anti-shirking norms. See Yael V. Hochberg & Laura Lindsey, Incentives, Targeting and Firm Performance: An Analysis of Non-Executive Stock Options, 23 REV. FIN. STUD. 4148, 4174–75 (2010).
36. See supra note 34 and accompanying text.
However, the risk–incentives model clearly conflicts with the monitoring model. The risk–incentives tradeoff suggests that as parties become less risk averse—as one would expect because law firms and corporations have become substantially larger over the past forty years—\(^{38}\) that the incentives side will dominate and that output-linked pay (like success fees) will become more common. On the other hand, the monitoring model suggests that as parties become larger and more sophisticated, they will better be able to use hourly fees as a monitoring tool, and as a result, hourly fees will become more common. To resolve this conflict, let us now turn to evidence from practice.

III. BILLABLE HOURS IN PRACTICE

A. Today

The risk–incentives model produces conveniently testable predictions. If workers resist incentive pay because they are risk averse, then it follows that workers in fields with more volatile production functions should resist incentive pay more, and therefore, incentive pay should be the most common in fields where outcomes are least volatile. In fact, the opposite is true. As economist Canice Prendergast discovered in a broad survey of different industries (including sharecroppers, franchisees, and natural gas contractors), there is generally a robust positive relationship between the level of risk attaching to a project and the likelihood that the contractors will use incentive pay.\(^{39}\)

In the context of law firms, the risk–incentives model suggests that the form of compensation should be determined by the relative risk tolerances of the principal (client) and agent (law firm). Large law firms will generally be better able to bear the risk of cost overruns (therefore favoring fixed fees), as well as the risk of outcome uncertainty (favoring contingent fees). On the other hand, large corporate clients will similarly be better able to bear both risks and should prefer hourly fees. Consequently, small law firms serving large corporate clients should use hourly fees, whereas large law firms serving individual clients should use fixed or contingent fees.

Contrary to the predictions of the risk–incentives model, however, the vanguard of the alternative fee movement has been small, boutique law firms engaged in niche work. These law firms often serve large corporate clients


who are better able to bear the risk of overruns than the law firm. If risk-bearing is the motivation for alternative fees, it makes little sense that the early adopters of alternative fees are exactly those firms worst positioned to bear risk.

Consider Wachtell, Lipton, Rosen & Katz, a relatively small New York law firm. Wachtell is one of the most profitable law firms in the United States but has only a single office and 272 lawyers, many fewer than its competitors. Wachtell initially built its reputation on its expertise in mergers and acquisitions, which are often one-off engagements with very large, publicly traded companies. Companies in mergers and acquisitions expect to incur significant transaction costs related to the deal, and they are not particularly sensitive to the risk of overruns. Thus, Wachtell is a good example of a relatively small firm that typically does work for large clients that are better positioned to bear risk. The risk—incentives model predicts that Wachtell will be more likely to use the billable hour than its competitors.

In fact, however, Wachtell is widely known for its alternative billing practices—it typically uses something akin to the success fee, taking its fee as a percentage of the size of the deal, much like an investment bank. This arrangement, which places all of the risk of overruns or failed deals on Wachtell, has proven immensely profitable—yet Wachtell is still the only firm among its peers to primarily use this method.

Another example is the boutique litigation firm of Bartlit Beck Herman Palenchar & Scott LLP. Bartlit Beck is a Chicago- and Denver-based boutique litigation firm with just 83 lawyers, which has vocally been at the forefront

of the alternative billing movement. Bartlit Beck refuses to bill by the hour, employing instead a combination of flat fees and contingent fees. Yet, its typical clients are some of the largest companies in America, including Tyco, Amazon.com, and Hewlett-Packard. Surely these clients ought to bear the risk of trial uncertainty rather than placing it on a small boutique law firm?

The story becomes even more convincing when we consider that Wachtell, Bartlit Beck, and many of the other firms most enthusiastic about alternative billing are similar to some of their peer firms along virtually every axis except for size. Skadden, Arps, Slate, Meagher & Flom, for example, was founded around roughly the same time as Wachtell by a culturally similar group of partners specializing in the same sort of work. The major distinction between these two firms is that Wachtell stayed relatively small, while Skadden expanded into new locations, bigger offices, and new practice areas, such that it is now one of the largest law firms in the world and one of the largest private companies in the United States. Yet, Skadden relies almost exclusively on billable hours. Similarly, Bartlit Beck regularly litigates with and against law firms with much larger litigation practices, both on the plaintiff and defense side—but these competitors almost always bill by the hour.

In fact, the commonality between the best-known alternative billing firms seems not to be size, but specialization. These firms almost always have a particular specialty—merger and acquisition work at Wachtell, trial practice at Bartlit Beck. Certain other firms implementing alternative billing practices have been larger than Wachtell or Bartlit Beck, but these firms have still

50. See Gilson & Mnookin, supra note 43, at 585; see also supra note 42 and accompanying text.
52. See In the News, supra note 47.
tended to practice a particular specialty. Fish & Richardson P.C., for example, an intellectual property boutique with twelve offices and over 350 attorneys, is often cited along with Bartlit Beck and Wachtell as a billing innovator. Both Fish & Richardson and Wachtell count among the hundred largest law firms in America by revenue. So, it appears that size alone is not determinative of propensity for alternative billing. Instead, there seems to be something about specialization that encourages firms to use flat or contingent fees.

If specialization were the only factor related to alternative fee use, we might doubt whether monitoring was the underlying cause. After all, it could be that boutique firms are simply more progressive in general—they might attract attorneys more eager for experimentation, or homogeneity among the work and interests of the firm members might allow for faster adaptation to market changes. Or it could be that lawyers prefer to work under alternative fee arrangements as a matter of personal preference and that only high-end lawyers have the negotiating power to convince clients to allow them to do so.

But even specialization fails to tell the whole story. Superstar firms like Bartlit Beck, Wachtell, and Fish & Richardson are not the only ones to reject hourly billing. Many plaintiffs’ attorneys, whose clients are usually non-lawyer individuals, bill using contingent fees; many solo practitioners use flat fees in their dealings with non-lawyer individual clients in fields such as family law, real estate law, and estate planning. The work of such solo practitioners is relatively much less specialized and much more routine than the work at Wachtell or Bartlit Beck.

The overall picture is therefore something like a bell-shaped graph between specialization and use of the billable hour. At the low end, solo practitioners provide legal services at flat rates to non-expert clients; in the

56. See Contingency Fee (Contingent Fee), 1 BOUVIER LAW DICTIONARY, supra note 17, at 1063.
59. See, e.g., id.
middle, large law firms provide repeat legal advice by the hour to large, expert corporate clients; and at the high end, boutique law firms provide one-off, niche services at flat rates to the same large corporate clients. So, what do urban super-specialist lawyers and small-town generalist lawyers have in common?

The answer depends not on the lawyers but on the clients. The elderly grandfather seeking estate planning advice and the corporate general counsel seeking takeover defense advice have one thing in common: they have little capacity to monitor the work of the lawyer. The grandfather is largely unfamiliar with the law, and therefore, unable to predict how much a matter should cost or to complain if his lawyer returns a list of unreasonable hourly charges. Similarly, because hostile takeovers are rare events in the life of a company (hopefully), the general counsel will have only a rough idea of how much it should cost or what kind of work such a legal matter requires. Moreover, if general counsel has not worked with her outside counsel before, she will have a more difficult time interpreting the idiosyncrasies of her outside firm’s hourly bills, and she cannot use the potential for repeat business as a means to enforce agent discipline.

Contrast these two examples with the more conventional work of a large, diversified law firm. Such a firm will have repeat relationships with its corporate clients, often even working with the same in-house lawyer on a series of different matters. In this case, in-house lawyers have the ability to specify how much they expect the law firm to bill on each matter. The law firm knows that the in-house lawyers will review bills and that they can move their business elsewhere if they are dissatisfied.60

60. Note, however, that there are some exceptions to the general trend that specialized firms are more likely to bill by the hour. The best known among these is the firm of Crowell & Moring LLP, a generalist Washington law firm with 482 lawyers. Crowell & Moring LLP Firm Profile, CHAMBERS & PARTNERS, http://www.chambers.com/profile/organisation/2838?publicationTypeld=2 (last visited Nov. 19, 2018). Crowell & Moring derives a significant portion of its revenue from alternative fees. See Value-Based Billing, CROWELL & MORING, LLP, http://www.crowell.com/Innovation-in-Service/Value-Based-Billing (last visited Nov. 29, 2018). However, Crowell & Moring seems to have developed its approach to billing as a response to requests from its largest clients, rather than as a proactive policy. In particular, AT&T and DuPont have both been at the vanguard of billing reform in recent years, and both are major Crowell & Moring clients. See Case Studies, CROWELL & MORING, LLP https://www.crowell.com/Case-Studies (last visited Oct. 30, 2018); see also Terry Carter, Do It the DuPont Way, 90 A.B.A. J., Apr. 2004, at 27; Kerry Curry, Five Firms Experiment with Alternative Fee Arrangements, TEX. LAWBOOK, Feb. 25, 2014, at 1, http://www.patrickonpricing.com/wp-content/uploads/2014/03/2014_TexasLawbook_Five_Firms_Experiment_Reprint.pdf.

Why should AT&T and DuPont be so enthusiastic about alternative billing if monitoring is the dominant explanation for billable hours? After all, AT&T and DuPont both have sprawling
So, both as an intuitive matter and as a descriptive matter based on the pattern of billable hours used at law firms, monitoring remains the most viable explanation for the prevalence of the billable hour.

**B. The History of the Billable Hour**

Another simple way to investigate the rationale for the billable hour is to investigate its history. As noted earlier, the risk-incentives model predicts that as law firms and corporations get larger, all parties will become less risk averse, the risk premium will decrease, the incentive side of the equation will dominate, and output-based compensation will become more common. In contrast, the monitoring model predicts that as corporations develop larger in-house legal departments, they will be better able to exploit the monitoring function of the hourly fee, and hourly billing will become more common.

Again, the monitoring story prevails. The billable hour became standard in the United States at a remarkable point in the history of legal practice—in the 1970s, when corporate legal departments transformed from one-man back offices to large, sophisticated miniature law firms. The timing of this shift was no coincidence. First-person accounts suggest that corporate counsel actively demanded improved monitoring metrics, and the billable hour was the response.

Prior to the 1970s, lawyers billed their clients primarily through a combination of fixed fees, contingent fees, and an amorphous method known as “value billing,” wherein they would simply hand the client a bill at the

---

legal departments and presumably possess more legal expertise and monitoring capacity than most firms.

One answer is that AT&T and DuPont require only specialty legal services precisely because their own legal departments are so large. Because both companies have already internalized generalist roles to in-house counsel, they only go to firms like Crowell & Moring for unusual work, where monitoring is relatively difficult. For example, much of the work that Crowell & Moring does for both companies is antitrust litigation, which is a large and unusual event in the life of a company. See *Case Studies*, supra. And indeed, much of the impetus for alternative billing at Crowell & Moring seems to have come from its relatively specialized litigation department. See generally ASS’N COUNSEL, supra note 18.

Another (less convincing) answer is that corporate legal departments obtain a kind of prestige from being on the forefront of billable hour reform. Thus, the distribution of billable hour reformers among general counsel might be skewed in the present because billable hours are currently such a hot topic among lawyers and because leaders of the bar will tend to be general counsel at the largest and most prestigious corporations.


end of the matter for “Professional Services Rendered.” Few lawyers itemized their services by the hour or established a written compensation agreement with the client in advance.

This casual approach to fees reflected the attitude of genteel neglect that pre-1970s lawyers took toward business matters. As members of a “learned profession,” lawyers had historically felt uncomfortable haggling with their clients, like mere tradesmen and contractors. They considered themselves advocates and advisors, and as such, just as trustworthy when it came to fee calculation as when they undertook sensitive legal work for their clients. Indeed, many lawyers considered it “unprofessional” to discuss legal fees at all.

Professional pride alone was little protection against billing abuses. However, prior to the explosive growth of in-house legal departments in the 1970s, corporate clients had little capacity to monitor the efforts or the charges of their outside contractors even if they had wanted to. The omertà of the legal bill complemented the nature of the corporate client.

This changed once corporations became more serious and systematic in their selection of outside counsel. First came the transformation of the general counsel’s role itself. Prior to the 1970s, the platonic ideal of the in-house lawyer was the “lawyer-statesman,” who provided ethical advice and acted as the conscience of the corporation. But in the 1970s, the general counsel suddenly metamorphosed from bookish consigliore to businessman in his own right. Legal departments became larger as corporations themselves grew, and they took on greater relative importance in corporations as regulations

the perceived value of services to the client but not necessarily based on an ex post assessment. See ASS’N CORP. COUNSEL., supra note 18, at 17. This Article uses “value billing” exclusively to refer to ex post billing by law firms.

63. Ezra Tom Clark, Jr., Getting Out of the Hourly Rate Quagmire—Other Billing Alternatives, in BEYOND THE BILLABLE HOUR: AN ANTHOLOGY OF ALTERNATIVE BILLING METHODS 183, 183 (Richard C. Reed ed., 1989); Robert E. Hirshon, Law and the Billable Hour, 88 A.B.A. J., Feb. 2002, at 10; see also Donald S. Akins, Client Acceptance of Alternate Pricing, in BEYOND THE BILLABLE HOUR, supra, at 189 (“Few, if any, records were kept to determine the cost of a particular service—but no one seemed to care.”).


65. Clark, supra note 63, at 183; see also Akins, supra note 63, at 189.


68. At this point, the general counsel was almost always a “he.”
became more complex and profuse. In-house lawyers increasingly behaved like businesspeople and increasingly became businesspeople, often crossing from chief legal officer to chief executive officer.

Unsurprisingly, the lawyer-businessman was much less likely than the lawyer-statesman to tolerate the genteel anti-competitiveness of value billing. In the 1960s and 1970s, general counsel became increasingly occupied with what Anthony Kronman calls the “scientization of law,” which emphasized efficiency and evidence-based rational decision making over the intuitive judgments of the lawyer-statesman. General counsel began to demand oversight over the fees that they were charged by their lawyers; moreover, they wanted metrics with which they could compare the quality and cost of work between different firms. The billable hour served all these purposes.

Clients had long complained about the opacity of value billing, but it was only in the 1970s that they gained enough expertise and leverage to do anything about it. Former ABA President Robert Hirshon recalls that as corporate counsel gained in expertise, they began to demand “a tangible, objective instrument for collecting fees,” leading to a “cost-conscious revolt” against value billing. Mary Ann Altman, founder of the legal consulting firm Altman Weil, recounts that “[h]ourly time billing was well received quickly by clients because by using it their bills were based on something tangible that they could understand rather than on a ‘value of services’ concept.” So, the timing of the shift to billable hours was more than a mere coincidence—clients embraced it in order to monitor lawyers’ efforts more closely.

Meanwhile, non-expert clients—especially individuals who rarely required legal assistance, like those attempting to draft wills or navigate the unfamiliar territory of family law—never stopped using fixed fees. As discussed in Section II.A, supra, flat fee schedules are still very common among solo practitioners doing routine work. For non-expert clients, the hourly bill was just another method of payment, and an unintuitive and unfamiliar one at that. Similarly, a few highly specialized law firms never made the switch over to billable hours. Wachtell, for example, was founded in 1965 and has always used a quasi-success fee based on the size of the deal. Its clients would have seen little point in any other system—its work was

69. Lynch, supra note 66, at 1404.
70. See Kronman, supra note 67, at 200–01. The idea of scientization of law was originally conceived by Mirjan R. Damaška in his seminal analysis of the history of evidence law. See Mirjan R. DAMAŠKA, EVIDENCE LAW ADRIFT 33 passim (1997).
71. Hirshon, supra note 63, at 10.
72. Mary Ann Altman, A Perspective—From Value Billing to Time Billing and Back to Value Billing, in BEYOND THE BILLABLE HOUR, supra note 63, at 11.
73. See Gilson & Mnookin, supra note 43, at 585 n.50; The Firm, supra note 48.
highly specialized and difficult to monitor, and its incentive to perform cheaply was small given the low probability of repeat business.

So again, the history of the billable hour supports the monitoring narrative. Hourly bills became the norm exactly when clients acquired the skills necessary to interpret them. Contemporaneous accounts and the pattern of transition confirm that the motivation for the shift was client demand for tangible metrics of lawyer effort.

IV. CONCLUSION

Taken as a whole, the weight of historical evidence and contemporary practice points toward the monitoring model as the more accurate explanation for the prevalence of billable hours. Descriptive correctness aside, this model has important practical, predictive, and prudential implications for the future of law firm practice.

First, as discussed at length above, the model suggests that firms should prefer alternative fee arrangements for complex matters and hourly fees for simple ones. Certain alternative fee advocates have so far suggested the opposite. For example, the Association of Corporate Counsel advises that “[f]ixed fees work best for litigation matters that both the client and law firm are already familiar with, which involve common issues and a common life cycle, reducing variability in litigation costs.”\(^\text{74}\) This perspective focuses on the risk of cost overruns to the law firm, and therefore, reflects the risk–incentives model. In contrast, the monitoring model implies that fixed fees work worst for routine matters, implying that firms and clients should reserve fixed fees for complex matters.\(^\text{75}\)

Second, the model predicts that if clients become more sophisticated, they will prefer the billable hour even more strongly. Some commentators have opined that the growth of in-house legal departments will cause a decline in the use of billable hours.\(^\text{76}\) The monitoring model suggests the opposite: all else equal, increased client expertise should result—and has resulted—in wider use of the billable hour.

In reality, of course, all else is rarely equal. In-house legal departments might grow by “insourcing” the generalist functions that they currently

\(^{74}\) ASS’N CORP. COUNSEL, supra note 18, at 3.

\(^{75}\) Of course, the fact that specialized law firms are more likely to use alternative fees does not prove that they should use alternative fees. This Article only claims that most critics do not recognize the monitoring function of billable hours at all. Because the best alternative to the monitoring model is the risk–incentives model, this Article proposes many commentators vastly overestimate the importance of risk in their policy suggestions.

\(^{76}\) See, e.g., Ribstein, supra note 4, at 770.
outsource to law firms, leaving those law firms with only niche work. If law firms become increasingly specialized, then billable hours should become less common (again, all else equal). This is so because clients will less be able to monitor specialist lawyers, and the billable hour will therefore be less useful for mitigating residual losses. Therefore, the effect of growth in corporate legal departments depends on the character of the change. However, this prediction provides another prospective basis on which to test the validity of the monitoring model.

Finally and most importantly, critics and practitioners should be circumspect about advocating the switch to alternative fees. Billable hours incur residual losses, but so would any principal-agent arrangement. And billable hours have the considerable advantage that experienced principals can mitigate their residual loss by monitoring, which is much more difficult with alternative fees. This is not to endorse any one billing method, since each principal and agent must negotiate to minimize agency costs according to their own circumstances. This Article simply observes that the billable hour was born for a reason, and it argues that the reports of its death have been greatly exaggerated.\(^\text{77}\)

\(^\text{77. Cf. 3 Albert Bigelow Paine, Mark Twain: A Biography 1039 (1912) (“[T]he report of my death has been grossly exaggerated.”).}\)