View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Research Papers in Economics

Nº 353

September 2008

Documento de Trabajo

ISSN (edición impresa) **0716-7334**

ISSN (edición electrónica) **0717-7593**

The Unexpected Effects of Caps on Non-Economic Damages

Álvaro Bustos
Ronen Avraham

Versión impresa ISSN: 0716-7334
Versión electrónica ISSN: 0717-7593

PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
INSTITUTO DE ECONOMIA

Oficina de Publicaciones
Casilla 76, Correo 17, Santiago
www.economia.puc.cl

THE UNEXPECTED EFFECTS OF CAPS ON NON-ECONOMIC DAMAGES

Álvaro Bustos*
Ronen Avraham**

Documento de Trabajo Nº 353

Santiago, Septiembre 2008

*a-bustos@law.northwestern.edu / a-bustos@kellogg.northwestern.edu
**r-avraham@law.northwestern.edu

INDEX

ABSTRACT 1

1. INTRODUCTION 2

2. BARGAINING MODELS AND TORT REFORM- LITERATURE REVIEW 8

3. THE MODEL 9
3.1 Equilibria 12
3.2 Preliminary Considerations about the Equilibria 16

4. RESULTS 19
4.1 Do Caps Accelerate or Delay Settlement? 19
4.2 Welfare Implications 22
4.2.1 Proportion of Disputes Settled 22
4.2.2 Do Caps Increase or Decrease Litigation Expenses? 23
4.2.3 Do Caps Increase or Decrease Plaintiffs' Recoveries? 26
4.3 Discussion: Which states should enact caps? 27

5. CONCLUSIONS 29

REFERENCES 30

APPENDIX 36

The Unexpected Effects of Caps on Non-Economic Damages

Ronen Avraham and Álvaro Bustos[1]

First Draft: May 2007. This version: September 2008

Abstract:

We study the economic and legal implications of the enactment of caps on non-economic damages on parties in conflict who know that state supreme courts may strike down the caps as unconstitutional within a few years of enactment. We develop a simple screening model where parties have symmetric expectations regarding the probability of a strike down and asymmetric information regarding plaintiff's non-economic harm. Our model makes several surprising predictions: First, caps may *increase* the length of resolution of disputes if the caps are low enough or the probability of a strike down is large enough. Second, although caps always increase the percentage of disputes that are settled out of courts, they do not necessarily save litigation expenses. Third, while caps always reduce the recoveries of plaintiffs with large claims, caps may *increase* recoveries of plaintiffs with low claims compared to their recoveries in states with no caps. We conclude that to increase welfare legislators have to tailor caps to the economic and constitutional circumstances in their state in ways which we characterize in the paper.

Keywords: tort reform, caps on recoveries, length of dispute resolution

JEL classification: K13, K20, K41

[1] Northwestern University School of Law: r-avraham@law.northwestern.edu; Catholic University of Chile, School of Management (currently at Northwestern University, School of Law and Kellogg School of Management: a-bustos@law.northwestern.edu/a-bustos@kellogg.northwestern.edu.) We thank Kenneth Ayote, Wiliam Berry, Lee Epstein, Laura Langer, Emerson Tiller, Max Schanzenbach, Katheryn Spier, Abraham Wickelgren and participants of seminars given at Northwestern Law School, the 2007 American Law and Economic Meetings held at Harvard and the 2007 Conference on Empirical Legal Studies held at NYU for useful comments.

1. Introduction

In the last few decades, dozens of different tort and medical malpractice reforms have been enacted, struck down, and at times, legislatively repealed or *re*enacted (see Avraham [2006a]). Indeed, tort reform is perhaps the foremost legal rights related item on legislative agendas. Interest groups regularly spend hundreds of millions of dollars promoting or opposing reform.[2] Pressure for tort reform is also building on the federal level. No fewer than sixteen bills to federalize various aspects of medical malpractice law were debated in the Congress during the period of Republican control between 1996 and 2006.[3] The most recent bill passed in the Senate was in 2006.[4]

One of the most popular reforms is caps on non-economic damages such as pain and suffering, loss of consortium, etc. By 2007, twenty-six states had enacted some type of cap on non-economic damages. From 1991 to 2007 alone caps on non-economic damages were enacted in 14 states. During this period, such caps were struck down by supreme courts in 5 states. In some states, such as Illinois, Ohio and Wisconsin, caps were struck down by state supreme courts and later reenacted in amended form. Sometimes this cycle repeated itself.[5]

Proponents of caps on non-economic damages argue that these caps will reduce excessive recoveries, expedite settlement, and reduce overall litigation expenses (see Atiya [1980] and Rubin [1993] among others). Proponents of tort reform reason that reducing the uncertainty associated with unlimited jury awards for non-economic

---

[2] Data on interest groups' expenditures on tort reform is available at opensecrets.org. *See* http://www.opensecrets.org/lobbyists/issuesum.asp?txtname=Torts

[3] *See* 104 H.R. 3103 (1996), 104 H.R. 956 (1996), 105 H.R. 1091 (1997), 106 H.R. 2242 (1999), 107 H.R. 2563, 107 S. 812, 107 H.R. 4600 (2002), 108 H.R. 5 (2003),108 S.2061, 108 S. 11 (2003), 108 S. 2207, 108 H.R. 4280 (2004), 109 H.R. 534 (2005), 109 S. 366, 367, 354, 109 H.R. 5, 109 S. 22

[4] S.22, 109th Cong. (2006).

[5] *See* Table 1 in the appendix for more detailed information on states that enacted and struck down caps on non-economic damages. For instance, Illinois enacted caps on noneconomic damages (735 ILCS 5/2-1115.1) effective on March 9, 1995. The reform limited noneconomic damages to $500,000. However, on December 18, 1997, the Illinois Supreme Court affirmed the district court's decision from August 20, 1996 and held that the reform violated the State Constitution (*Best v. Taylor Machine Works, Inc., 689 N.E.2d 1057 (Ill. 1997)*). On August 25, 2005 Illinois enacted again caps on non-economic damages, only to see them struck down on November 13, 2007 by a state trial court. (As of September 2008, the case is still pending before the Illinois Supreme Court.)

Electronic copy available at: http://ssrn.com/abstract=1125182

damages will facilitate out-of-court negotiation (see e.g. Atiya [1980] pp 216). They argue that caps on total recovery incentivize plaintiffs to resolve disputes through less costly out-of-court settlements rather than gamble for big awards from costly trials. Indeed, Watanabe (2006) predicts that reduced uncertainty will shorten the time to settlement.

On the other side, opponents of caps argue that caps often reduce recoveries for the most severely injured plaintiffs, thereby shifting the costs of injuries away from blameworthy parties and onto the most needy tort victims (see Viscusi [1991] pp 107 and [ALI] 1991 pp 219-20). They also argue that caps might dilute defendants' incentives to take optimal care (see Arlen [2000]).

Neither proponents nor opponents of caps on non-economic damages have concerned themselves with the uncertainty surrounding the constitutionality of caps tort reform. Historically, the constitutionality of more than half of the caps on non-economic damages enacted into law met legal challenges on state constitutional grounds within few years of enactment.[6] Indeed, there is some anecdotal evidence that both the size of the caps and their constitutionality are perceived to be important. As for the size of the caps a report by the U.S Department of Health and Human Services claims that there is a "substantial difference" between the impact of caps on non economic damages in "states with meaningful caps" and "states without meaningful caps," where meaningful means that caps that are not larger than $350,000 (see U.S DHHS [2003]). As for a cap's constitutionality, the chairman of the ISMIE Mutual Insurance Company, which provides liability insurance for doctors, has recently argued that the positive impact of tort reform in the states is felt only "after the Supreme Courts in these states have upheld the meaningful reforms." (See Parsons [2005]).

The veil of uncertainty surrounding the constitutionality of reforms between an enactment date and a final ruling by a state's Supreme Court may incentivize litigants

[6] Moreover, in recent years the practice of challenging tort reform in state court has become a much more coordinated. The Center for Constitutional Litigation, PC is a national law firm dedicated to challenging tort reforms in states and federal courts. The Center receives most of its revenues from the national trial lawyers' trade group (called: The American Association of Justice) and from the states' trial lawyer associations. As of December 2007 the Center had forty tort-reform-related pending cases across the United States in which lawyers from the Center were helping trial lawyers nationwide to challenge tort reform (See, Lynne Marek, *A small firm wages a '100 year war' on tort reform: Center is engaged in 40 cases challenging limits on tort claims*, National Law Journal, December 10, 2007).

differently than scholars generally assume. Specifically, expectations of a strike down might delay settlement and consequently increase overall costs.

In this paper we develop an asymmetric information screening model from which we draw inferences about the effect of caps on non-economic damages on length and cost of litigation and on recoveries for different types of plaintiffs. Our model accounts for the size of the caps on non-economic damages as well as for the plaintiff's idiosyncratic non-economic harm and for both parties' (symmetrical) expectations of the eventual strike down of a cap.

In order to study the impact of caps on the time of resolution of disputes and the welfare of the parties, we map the negotiation process over two rounds, each divided into a period of settlement and a period of trial. In addition, we assume two types of plaintiffs: one with high non-economic harm ("high type") and one with low non-economic harm ("low type"). We include a discount factor to account for the cost of delayed resolution. We also consider factors addressing a defendant's settlement and litigation costs. Finally, we include factors addressing the probability that a state's Supreme Court will strike down caps on non-economic damages. These factors are important to a plaintiff's choice to accept early offers or proceed to the final rounds of negotiation.

As a baseline we first characterize the solution of the model when damages are not capped ("Regime NC"). We show that in this Regime the dispute is always solved in the first round of negotiations. As is standard in this type of model, if the probability that the plaintiff is a low type is small enough, the defendant makes a high settlement offer which is accepted by both types of plaintiffs and the dispute is settled immediately (a pooling equilibrium takes place). Alternatively, if the probability that the plaintiff is a low type is large enough, the defendant makes a low settlement offer which is only accepted by the low type plaintiff. The high type plaintiff rejects the offer and litigates immediately (a separating equilibrium takes place).[7]

The novelty in this paper begins once we account for a cap that limits possible recoveries during the first round of negotiations but that may get struck down with a

[7] We assume that courts yield accurate awards. Thus, because high type plaintiffs go straight to litigation, low type plaintiffs cannot gain by mimicking high type plaintiffs (in court, low type plaintiffs only recover their low value claim).

given probability during the second round of negotiations.[8] In that case, not only does the defendant tend to make smaller settlement offers, but also the plaintiff, under certain circumstances, has incentives to reject the initial offer in hope of recovering more—which would happen if the caps were struck down.

Our model identifies the size of the cap *and* the probability of a strike down as the two key factors interacting to drive parties' decisions. If the cap is high enough or the probability of a strike down is low enough, the parties' decisions are equivalent to those in Regime NC with the exception that the maximum non-economic harm that the high type plaintiff can recover is equal to the cap and therefore is lower than her *actua*l non-economic harm.[9] We denote states with caps of this kind as Low Expected Trim Caps (LTC) because the plaintiff should not expect to lose much by accepting the first offer. Essentially, in this situation a plaintiff gains nothing by waiting: The plaintiff's damages do not much exceed the cap, or the probability of strike down is low, so that the expected present value of a second-round resolution is not better than the defendant's first offer.

On the other hand, if the cap is low enough or the probability of a strike down is high enough the high type always waits for the second round of negotiations, and the low types, knowing that, sometimes decide to mimic that decision. We denote states with caps of this kind as High Expected Trim caps (HTC) because the plaintiff expects to lose a high fraction of her recovery if she accepts the first offer.

We start by showing that in Regime LTC, the time to settlement always decreases in comparison to Regime NC, whereas in Regime HTC, that time to settlement may *increase*.[10]

The reason for the first result is straightforward. The pooling and separating equilibria in Regime LTC are essentially similar to those in Regime NC. The critical difference is that more pooling equilibria take place in Regime LTC than in Regime NC

[8] We consider that the strike down happens only during the trial and not during the settlement; otherwise trials would not take place. Explicitly considering that strike-downs take place in both instances would only complicate the model without adding significant value.

[9] The analysis is done for the case in which the cap lies between the values of the low and high type plaintiffs. We do not consider the extreme cases in which the cap is lower than the low type plaintiff or higher than the high type plaintiff because the predictions become trivial.

[10] Indeed, in Avraham & Bustos (2008) we found that the average time-to-settlement in states with Regime NC is 4.08 years, in states with Regime LTC is 3.26 and in states with Regime HTC is 4.4 years. All differences are significant at 1% or less.

because, from the perspective of the defendant, making a "high" settlement offer is cheaper in Regime LTC than in Regime NC.[11] Thus, more disputes are resolved by an initial settlement offer in Regime LTC than in Regime NC.

The reason for the second result is less direct but still uncomplicated. The pooling and separating equilibria in Regime HTC are dissimilar to those in Regime NC, because after the first settlement offer is made some plaintiffs in Regime HTC are incentivised to wait to resolve the case in future rounds of negotiations. Specifically, it is still true that more pooling equilibria take place in Regime HTC than in Regime NC, which could imply that time to settlement decreases in Regime HTC. Yet, in the cases in which a separating equilibrium occurs, many disputes are resolved in the second round of negotiations in Regime HTC instead of in the first round. Determining how an increased number of pooling equilibria balances out with more second-round negotiations depends on the magnitude of the following parameters: the legal cost of settlement, the discount factor, and the probability of a strike down.

We show that if the cost of settlement is small enough or the discount factor and the probability of a strike down are large enough the second effect dominates the first one and the time to settlement is *larger* in Regime HTC than in Regime NC.

In order to show why, we first demonstrate that in the extreme case where the costs of settlement are zero, or instead the discount factor and the expected probability of a strike down are one, disputes are resolved later in Regime HTC than in Regime NC. Under these extreme values any effect that tends to reduce the time to settlement in Regime HTC disappears. Second, we show that the time to settlement in Regime HTC decreases when the cost of settlement becomes larger or when the discount factor and the probability of a strike down become smaller because in all these cases the defendant has stronger incentives to induce a pooling equilibrium, as he expects to pay relatively less than in a separating equilibrium. Thus, the continuity of the parameters suggest that there

[11] This is because a settlement offer is "high" or "low" relative to some maximum award attainable at trial—so that a "high" offer is cheaper under Regime LTC in which the maximum award attainable at trial is capped.

are three thresholds[12] which, if met, make the time to settlement in Regime HTC longer than in Regime NC.

Our model also uncovers two important effects of caps on social welfare. First, it shows that the expected litigation expenses in Regime HTC may be larger than the expected litigation expenses in Regime NC. This follows from the fact that disputes in Regime HTC may take longer. A common complaint about the tort system is that it is inefficient: for every dollar of compensation paid by the defendant only 50 cents go to the plaintiff; the rest is lost as costs (See Avraham [2006b] pp 97). The model suggests that caps might not only fail to improve the efficiency of the system but in fact might make it worse.

Second, the model predicts alterations in plaintiffs' awards. High type plaintiffs are always worse off in a caps regime, because they either recover less or recover later. In contrast, the model shows that some low type plaintiffs may be either worse off or better off under a caps regime. Specifically, some low type plaintiffs who used to mimic high type plaintiffs and consequently recover high type awards under Regime NC will only recover the cap under Regimes LTC or HTC, and thus will be worse off. But some other low types (potentially plaintiffs with frivolous lawsuits)[13] who were sorted into the separating equilibria under Regime NC will fall into the pooling equilibria under Regimes LTC or HTC, thus obtaining a higher award.

We therefore conclude that without tailoring caps to the economic and constitutional environment in the state, state legislators may find that enacted caps might decrease welfare by increasing overall litigation costs and the time to resolving the disputes, by under-compensating the severely injured victims or by over-compensating frivolous plaintiffs.

The rest of the paper is organized as follows. In Section 2, we review the literature. In Section 3 we present the model. In Section 4 we show our main theoretical results. In Section 5 we conclude.

---

[12] Cost is below a certain threshold, discount factor and probability of strike down are above other thresholds.

[13] That plaintiffs with small claims, even plaintiffs with negative expected value, can extract settlements was first observed by Bebchuk (1988) and was widely discussed in the literature,

2. Bargaining Models and Tort Reform- Literature Review

Our paper engages two strands of literature: the literature on bargaining models and the literature on the impact of tort reform.

There is a great deal of theoretical literature on bargaining models of dispute resolution examining why and when parties litigate instead of settle. (See, e.g., the surveys by Daughety [2000] and Spier [2005]). Parties may delay or even forgo settlement, even if symmetrically informed, when the relative structure of their litigation costs makes holding out for trial attractive, such as in Spier (2005) where costs are "lumpy". Parties may also forgo settlement when they do not share a common prior belief as to the likely outcome of a trial. (e.g. Landes [1971], Posner [1973] or Priest and Klein [1984]). Furthermore, one-sided, asymmetric information regarding a defendant's liability or plaintiff's harm may increase the likelihood of a trial. (See, e.g., Bebchuk [1984], Nalebuff [1987], Reinganum and Wilde [1986], Spier [1992] or Sieg [2000]). The same result may occur when there exists two-sided asymmetric information—that is, both parties have information regarding their liability or harm that their adversary does not possess. (see e.g. Schweizer [1989], or Daughety and Reinganum [1994]). None of these models, however, takes into account the existence of caps, their size, or their constitutionality.

Despite the attention tort reform, attracts there are only couple of law and economics models of it. Those that exist usually deal with the impact of tort reform on plaintiffs' recoveries or on physicians' initial behavior (Currie and McLeod [2008], Watanabe [2006]). There also are few empirical studies that explore the impact of tort reform on time to settlement. Babcock and Pogarsky (1999) conducted laboratory experiments demonstrating that caps on jury awards encourage settlement. Kessler (1996) explored the causes of delay in settling automobile accident disputes. He found that reform imposing prejudgment interest, originally designed to reduce delay, in fact increases the time to settlement. Recently, Watanabe (2006), using a structural model approach, found that capping jury awards or eliminating the contingency fee rule significantly shortens the expected time to resolution and lowers the expected total legal costs. Overall, however, there is only little academic consensus about the actual impact of tort reform on various litigation outcomes such as average awards, frequency of litigation and total payments.

(see surveys by Danzon [2000] or Kessler and Rubinfeld [2004]). Again, none of these studies takes into account the possibility that the tort reform will be struck down or the impact of the size of the caps on dispute resolution times or recoveries.

3. The Model

A risk neutral victim (Plaintiff) has a valid claim of x dollars of non-economic harm against a risk neutral negligent wrongdoer (Defendant).[14] While the liability of the wrongdoer is not disputed, there is uncertainty about the amount of the victim's harm. There are two possible types of victims: (1) A victim with high non-economic harm, x_H; and (2) a victim with low non-economic harm, x_L, where $x_H > x_L$. In either case, the defendant cannot observe the plaintiff's actual non-economic harm, x, instead he can only estimate (perhaps based on the observable economic harm) the probability, π, that the plaintiff is a low harm type victim. We assume that π is drawn from a probability distribution with density $f(\pi)$.[15]

In order to capture the possibility of acceleration or delay in the resolution of the conflict between the parties, we map the negotiation process over two rounds, each divided into a period of settlement and a period of trial, overall four periods. In the first period the defendant makes a settlement offer (S_1) that the plaintiff can either accept or reject. If the plaintiff accepts S_1, the game ends there. If the plaintiff rejects it, the parties move to the second period. In the second period the plaintiff either goes to court which would award damages (x_H or x_L) based on the victim's type,[16] or will wait for a new settlement offer in next round of negotiation. In the second round of negotiation (the third period) the defendant makes a new settlement offer (S_3) that the plaintiff, again, can

[14] In a more general formulation, Plaintiff's claim $X = x_o + x_i$ has two components: an observable component, x_o, which represents the economic harm, such as medical bills, loss of income, etc, and an idiosyncratic unobservable component, x_i which represents the non-economic harm such as pain and suffering, mental anguish, etc. For simplicity we normalize the observable components, x_o, to equal zero and focus on the idiosyncratic component, x_i that we denote x.

[15] The probability distribution is not relevant for the characterization of the game played by plaintiff and defendant, but it will be relevant in Section 4 when we compare properties of regimes with and without caps.

[16] The fact that courts can correctly observe plaintiff's true harm is not a strong assumption because it is equivalent to assuming that courts are not systematically biased and get it right, on average. Indeed, this is the same assumption used by Spier (1992) and Watanabe (2006).

either accept or reject. If the plaintiff accepts S_3, the game ends there. If the plaintiff rejects it the parties move to the fourth period. In the fourth period the parties go to court with certainty, and the court would award x_H or x_L according to the victim's damages. The timing of actions is the following:

At t = 1 the defendant makes settlement offer (S_1)

If the plaintiff accepts the offer the dispute ends there

If the plaintiff rejects the offer the parties move to the second period

At t = 2 the plaintiff decides whether to go to court

If the plaintiff decides to go to court the dispute is resolved there

If the plaintiff decides not to go to court the parties move to the third period

At t = 3 the defendant makes settlement offer (S_3)

If the plaintiff accepts the offer the dispute ends there

If the plaintiff rejects the offer the parties move to the fourth period

At t = 4 the parties go to court with certainty and the dispute is resolved there

Settlement negotiations and litigation are costly to both parties. Following other studies, (e.g. Spier, 1992) we normalize the plaintiff's costs to be zero.[17] Hence, we assume that the defendant faces a fixed cost c for each settlement offer associated with the pretrial negotiation (for example, if the plaintiff accepts S_1 the defendant incurs c, but if the plaintiff rejects S_1, waits for a new offer at the third period (S_3) and accepts it, the defendant incurs 2c in nominal terms). In addition, we assume that the defendant incurs a fixed cost k if the case goes to court (either in period 2 or period 4) with $k > c$, and the parties have the same discount factor which we denote δ.

We compare the negotiation behavior and recoveries of the low type plaintiff and the high type plaintiff in a regime with and without caps on non-economic damages. As its name suggests it, a cap on non-economic damages establishes the maximum amount that can be recovered by plaintiffs in courts for their non economic harm. We denote it

[17] This usual practice does not affect the generality of the results because, as will be seen in the derivation of the game's equilibriums, what matters is the difference in the litigation costs faced by the parties.

$x^c \in [x_L, x_H]$. Note that the cap is binding in courts only and does not impose any direct limit on the settlement amount.

As was explained in the introduction, caps are routinely struck down by state supreme courts. From this, it follows that rational agents develop expectations that a strike down may take place—not necessarily in their case—sometime prior to the resolution of their case. In reality, these expectations may even change with time. For simplicity, we assume that both parties share the belief that the cap may get struck down with probability α and that the uncertainty is resolved once and for all at t = 4.[18] Notice that at the beginning of t = 4 there still is uncertainty about the amount that will be recovered but before the trial court makes a decision, the uncertainty is resolved by a ruling or by inaction by the state's high court.[19] Diagram 1 presents the time line in the fourth period.

[Diagram 1 here]

Lastly, we define $\bar{x} = \alpha x_H + (1-\alpha)x^c$ as the expected payment obtained by a high type plaintiff if she goes to trial in the fourth period when caps are in place: If the caps are struck down, the high type plaintiff receives her true valuation, x_H, whereas if the caps are upheld she gets the cap, x^c.

[18] The assumption that both sides have the same beliefs about the probability of a strike down describes reality more accurately as we do not think that, in general, is true that one side has more (or less) information related to the "political desires to eliminate caps" than the other side. The assumptions that beliefs do not evolve through time and are not endogenously determined allow us to measure a first order magnitude of the impact of expectations over the behavior of agents. A rational expectations equilibrium approach would require an extensive description of the role of judges with a consequent deviation in the focus of the paper.

[19] As was mentioned in footnote 6, a special law firm called The Center for Constitutional Litigation, PC routinely challenges tort reforms in states and federal courts. For example, on November 2007 the law firm convinced a trial court in Illinois to strike down a medical malpractice reform enacted in August 2005. As of September 2008 the case is still pending at the Illinois Supreme Court. Lawyers in Illinois follow such cases closely and have been developing expectations regarding the probability of strike down at least from the moment the case was filed in the lower court on November 2006, a little over a year after the enactment of caps reform. See LeBron v. Gottlieb Memorial Hospital, No. 06-L-12109.

3.1 Equilibria

Complete proofs of the equilibria reached in the various regimes (with and without caps) are relegated to Appendix A. Here we summarize the most important characteristics and implications of the equilibria and the parties' strategic behaviors.

Recall that we denote as Regime NC the equilibrium in which there are no caps. When caps are in place, we identify the existence of two types of equilibria. The first equilibrium takes place when the cap is high enough or the expectation of a strike down is low enough. In this equilibrium the present value of the expected payment obtained by a high type plaintiff *is not* significantly trimmed if she decides to settle immediately (first period) instead of waiting for a future resolution of the dispute (third or fourth period). We denote this equilibrium as a regime with caps and low expected trim (Regime LTC).

The second equilibrium takes place when the cap is low enough or the expectation of a strike down is high enough. In this equilibrium the present value of the expected payment obtained by a high type plaintiff *is* significantly trimmed if she decides to settle immediately instead of waiting for a future resolution of the dispute. We denote that equilibrium as a regime with caps and high expected trim (Regime HTC).[20]

REMARK While the model provides a simple way to classify states as Regimes LTC or HTC , as an empirical matter it is not as easy to find proxies for the probability of a strike down, α , and for a high-type claim, x_H . However, by using the political composition of the states' Supreme Courts as a proxy for α (under the assumption that liberal courts are more likely to strike down the reform) and the distribution of awards in a state as the basis for estimating x_H Avraham and Bustos (2008) suggest that, for example, Maine is a state with Regime HTC and North Dakota is a state with Regime LTC.

A common property in the solutions for every type of regime (NC, LTC and HTC) is that there exists a cutoff probability that the plaintiff is a low type victim such that for any probability, π , smaller than this cutoff value, the solution defines a pooling

[20] Obviously, some high (low) trims of the recoveries may take place in Regime LTC (HTC) as the equilibria refer to the average value of trims.

equilibrium where the defendant ends up paying the same amount of money to both types of plaintiffs. For any probability higher than this cutoff, the solution defines a separating equilibrium where, in general, the defendant ends up paying different amounts of money to the high and the low type victims.[21] As will be explained in more detail below, there are different cut-off probabilities for the no-caps regime (π^{NC}), for the regime with caps and high expected trim (π^{HTC}), and for the regime with caps and low expected trim (π^{LTC}). Figure 1 summarizes the most important characteristics of the solutions for the regimes with and without caps which we start detailing next.

[Figure 1 here]

A Regime With No Caps (Regime NC)- When the parties face no caps, there exists a unique perfect Bayesian equilibrium. When there is a low probability that the defendant faces a low type victim, i.e. when $\pi < \pi^{NC}$, the defendant's offer is δx_H and both types of plaintiffs accept it. In that pooling equilibrium the low type plaintiff benefits from defendant's unwillingness to offer δx_L.[22] Conversely, when there is a high probability that the defendant faces a low type victim, i.e. when $\pi > \pi^{NC}$, the defendant offers δx_L. In that separating equilibrium, the low type settles immediately, because waiting will not yield her a higher recovery, whereas the high type will settle in the second period of the first round of negotiation (litigation), because that will yield her a recovery of $\delta x_H > \delta x_L$.

A Regime With Caps- When the parties face caps there still exists a unique perfect Bayesian equilibrium, but there are two important differences from Regime NC. First, when caps are in place the maximum that any plaintiff can recover is not δx_H but δx^c. Second, the incentives of the parties to wait for future periods are changed so that parties

[21] We will see that in some cases of separating equilibrium both types end up recovering the same. Nevertheless, in those cases, the properties of the solution differ from the properties of a pooling equilibrium.

[22] The defendant does not want to incur a second round of negotiations costs, c, and possibly legal costs, k, if the case goes to trial because there are not enough low type victims for the defendant to benefit from separating them out.

may end up resolving their disputes in the second round of negotiation (i.e. in third or fourth period).

In order to distinguish between the high and low expected trim equilibria we define the expected trim in recovery as $\alpha(x_H - x^c)/x^c$. This expression captures the expected nominal disutility (disutility when $\delta = 1$) that the high type plaintiff suffers if she settles in the first period for x^c instead of in the third period for the expected value of $\alpha x_H + (1-\alpha)x^c$. Note that we are not claiming that all the settlements will be equal to x^c. As we will see later, there are cases (depending on the value of π) in which the parties settle for more or less than x^c.

The higher the caps are, or the lower the expectation for a strike down is, the more willing is a plaintiff to settle in the first period. The reason is that a plaintiff expects a small trim by settling in the first period because either the caps are high (so settling now for x^c does not entail a large loss) or the expectation for a strike down is low (so there is not much gained from waiting to the next period).

Appendix A shows that different equilibria take place depending on whether $\alpha(x_H - x^c)/x^c$ is larger or smaller than the cutoff $(1-\delta^2)/\delta^2$ (or equivalently, if x^c is larger or smaller than $\delta^2 \bar{x}$). If $\alpha(x_H - x^c)/x^c < (1-\delta^2)/\delta^2$ we are in Regime LTC. If $\alpha(x_H - x^c)/x^c > (1-\delta^2)/\delta^2$ we are in Regime HTC.[23] Figure 2 shows the set of equilibria for Regimes LTC and HTC in the space (α, x^c). Notice how the separation of cases depends on *both* parameters. For example, when $\alpha = 1$, which means that the probability of a strike down is 1, there exists a cut-off value $\delta^2 x_H$ such that for values of the cap smaller than that cut-off we are in Regime HTC, but for values larger than that cut-off we are in Regime LTC. Also observe that the lower the discount factor is the less attractive it is for the plaintiff to delay the resolution of the dispute and consequently the more likely it is that we are in Regime LTC.

[Figure 2 here]

[23] We do not consider the case: $\alpha(x_H - x^c)/x^c = (1-\delta^2)/\delta^2$ because it does not add to the main discussion. The equilibrium strategies are a mix of the strategies that define the LTC and HTC solutions.

A Regime With Caps and Low Trim (Regime LTC)- When $x^c > \delta^2 \bar{x}$, the results that we obtained in the regime without caps are replicated here only in that x^c replaces x_H. Specifically, when there is a low probability that the defendant faces a low type victim, i.e. when $\pi < \pi^{LTC}$, there is a pooling equilibrium in which both types of victims receive δx^c in the first period. Otherwise, when there is a high probability that the defendant faces a low type victim, i.e. when $\pi > \pi^{LTC}$, there is a separating equilibrium where the low type victim receives δx_L and the high type victim receives δx^c. Like in the case of no caps, the low type recovers in the first period and the high type in the second period, both in the first round of negotiation. The reason the high type is not willing to wait for a third period is simple: The probability of a strike down is not large enough, and/or the caps are large enough relative to her true harm, so the plaintiff does not expect to gain much from waiting.

A Regime With Caps and High Expected Trim (Regime HTC)- When $x^c < \delta^2 \bar{x}$ the analysis becomes more nuanced. As in regimes NC and LTC, it is still true that when there is a low probability that the defendant faces a low type victim (i.e. when $\pi < \pi^{HTC}$), there is a pooling equilibrium in which both types of victims receive $\delta^3 \bar{x}$ in the first period. However, when there is a high probability that the defendant faces a low type victim (i.e. when $\pi > \pi^{HTC}$), things change in two ways compared to the other regimes. First, the high type victim always waits for the second round (third period) settlement offer (S_3). The reason is that when the defendant offers δx_L in the first period, the plaintiff can only recover δx^c in the second period and *both* expressions are smaller than $\delta^3 \bar{x}$ which is what the high type plaintiff expects to recover in the third period. To see that indeed $S_3 = \delta^3 \bar{x}$, notice that the defendant mixes between two offers: with probability $p^D = (\delta^2 \bar{x} - x_L) / \delta^2 (\bar{x} - x_L)$ he offers δx_L in which case the high type plaintiff goes to court and recovers an expected value of $\delta \bar{x}$,[24] or, alternatively, with

[24] The offer δx_L is an option given that the defendant knows that the low type may have mimicked the high type.

probability $1-p^D$ he offers $\delta\bar{x}$ in which case the high type plaintiff accepts it. Second, the low type plaintiff may not only decide to wait in the first period, but also in the second period (the idea is to mimic the high type's decision). The low type settles in the first period for δx_L with probability $1-p^{LP}=1-k(1-\pi)/(\pi(\bar{x}-x_L))$ or settles in the third period for (S_3) with probability p^{LP}.[25] In order to support the mixed strategies equilibrium the expected (and discounted) value of the recovery in the third period is δx_L.

3.2 Preliminary Considerations about the Equilibria

It is straightforward to show that there are more pooling equilibria under Regime LTC than under Regime NC. This is because the amount that the defendant needs to offer to induce plaintiff's immediate acceptance is smaller under the regime with a cap than the regime without a cap (δx^c instead of δx_H). In order to see that formally, note that the expressions determining π^{NC} and π^{LTC} are:

$$\pi^{NC}\delta x_L+(1-\pi^{NC})\delta x_H+(1-\pi^{NC})\delta k+c=\delta x_H+c \quad (1)$$

and

$$\pi^{LTC}\delta x_L+(1-\pi^{LTC})\delta x^C+(1-\pi^{LTC})\delta k+c=\delta x^C+c \quad (2)$$

respectively. The expressions identify the πs that make the defendant indifferent between pooling and separating equilibrium. The right hand side of (1) and (2) corresponds to the cost faced by the defendant in a pooling equilibrium (single settlement payment plus negotiation cost, c) while the left hand side corresponds to the cost faced by

[25] As usual in these cases, there is no equilibrium in which the low type plaintiff plays a pure strategy because if her strategy is always to wait she recovers only δx_L. In this case the defendant induces the high type to settle in the first period by offering her $\delta^3\bar{x}$ and has certainty that in the second period he is facing the low type. But then the low type has no incentives to wait, hence the strategy cannot be an equilibrium. In the same way, if her strategy is never to wait she recovers only δx_L because the defendant will not offer her $\delta^3\bar{x}$. Again, there are incentives to deviate.

the defendant in a separating equilibrium (settlement payment tailored to each type of plaintiff plus the cost of negotiation, c, plus the cost of trial if plaintiff is the high type). As for a given π, the separating equilibrium is more attractive than the pooling equilibrium by $\pi\delta(x_H - x_L) - (1-\pi)\delta k$ under Regime NC (we subtract the left hand side from the right hand side of (1)) but only by $\pi\delta(x^c - x_L) - (1-\pi)\delta k$ under Regime LTC (the same but for (2)). [26] It follows that the set of values of π for which the pooling equilibrium dominates the separating equilibrium is larger under regime LTC than it is under Regime NC (i.e., $\pi^{NC} < \pi^{LTC}$).

The comparison between the set of equilibria generated under Regime HTC and the set of equilibria generated under the other two regimens requires more elaboration. First, note that the identity that determines the value of π^{HTC} is

$$\pi\delta x_L + (1-\pi)\delta^3 \bar{x} + (1-\pi)\delta\left[\frac{\delta c}{1-\pi^*} + p^D\delta^2 k\right] + c = \delta^3 \bar{x} + c \qquad (3)$$

where $\pi^* = k/(k + \bar{x} - x_L)$. Again, (3) establishes the point of indifference for the defendant between pooling and separating equilibrium. If we compare (3) with (1) and (2), we realize that not only the expected recovery of the high type victim has changed, now it is $\delta^3 \bar{x}$ instead of δx_H or δx^c, but also the expected litigation costs faced by the defendant when he offers $S_1 = \delta x_L$. In regimes NC and LTC the expected overall legal expenses were $(1-\pi)\delta k + c$ whereas now, under Regime HTC, they are $(1-\pi)\delta\left[\frac{\delta c}{1-\pi^*} + p^D\delta^2 k\right] + c$. The first expression in the square bracket represents the defendant's negotiation costs in the third period. The second expression in the square bracket represents the litigation costs that might take place in the fourth period.

[26] These are values of π larger than $\pi^{NC} = k/(x_H - x_L + k)$ for Regime NC and larger than $\pi^{LTC} = k/(x^c - x_L + k)$ for Regime LTC.

REMARK Notice that the negotiation costs are always larger under Regime HTC than under Regimes NC or LTC as $(1-\pi)\delta\left[\frac{\delta c}{1-\pi^*}\right]+c>c$. In addition, as $p^D\delta^2<1$, the litigation costs in Regime HTC are always smaller than in Regimes NC or LTC. As we will explain with more detail later, the higher negotiation costs in Regime HTC is what causes this regime to be sometimes more costly than Regime NC.

The former analysis allows us to derive a number of properties. First, π^{HTC} is increasing with c because the higher the litigation expenses are the less attractive it is for the defendant to offer δx_L (that offer may induce the plaintiff to wait and thus generate an extra round of negotiations). Second, π^{HTC} is decreasing with δ and α because the smaller $\delta^3\bar{x}$ is the more attractive it is for the defendant to make the offer that induces the pooling equilibrium.[27] Third, from inspection of (3), we realize that if c = 0 and δ = α = 1 then $\pi^{HTC}=\pi^{NC}$.

Taken together, these properties imply that there are more pooling equilibria under Regime HTC than under Regime NC (π^{NC} does not depend on c, δ or α). In addition, if the settlement costs, c, are large enough there are more pooling equilibria under Regime HTC than under Regime LTC (i.e. $\pi^{HTC}>\pi^{LTC}$). But if the settlement costs, c, are small enough the opposite is true (i.e. $\pi^{HTC}<\pi^{LTC}$). We summarize the former analysis in the next two Lemmas.

Lemma 1: π^{HTC} is strictly increasing in c but strictly decreasing in δ and α.
Proof: See Appendix B.

Lemma 2:

a) For all values of c, δ and α, $\pi^{NC}<\pi^{LTC}$ and $\pi^{NC}<\pi^{HTC}$

[27] At first, this looks counterintuitive as the larger δ or α are, the more the defendant expects to pay in litigation expenses if the plaintiff decides to wait. Nevertheless, δ and α also determine how much the defendant pays the plaintiff if the dispute is resolved in the first period ($\delta^3\bar{x}$), and this last effect dominates the first one. See the proof of proposition 1 in Appendix B for more details.

b) For all values of δ and α, $\pi^{HTC} > \pi^{LTC}$ if $c > \left[\left(\frac{\delta^2 \bar{x} - x_L}{x^c - x_L} - \delta^2 p^D\right)\frac{1-\pi^*}{\delta}\right]k$ but $\pi^{HTC} \in \left[\pi^{NC}, \pi^{LTC}\right]$ if $c \in \left[0, \left(\frac{\delta^2 \bar{x} - x_L}{x^c - x_L} - \delta^2 p^D\right)\frac{1-\pi^*}{\delta}\right]k$.

Proof: Part a) follows from the former analysis. Part b) follows from straightforward algebra.

4. Results

4.1 Do Caps Accelerate or Delay Settlement?

We start by asking whether caps reduce the time required by the parties to solve their disputes after we account for the parties' knowledge that the caps may get struck down some time after they are enacted. In this Section we show that: 1) the expected length of disputes in LTC states is shorter than equivalent disputes in states without caps; 2) the expected length of disputes in HTC states may be longer than equivalent disputes litigated in states without caps. This last possibility is realized when settlement costs, c, are low and expected recoveries, $\delta^3 \bar{x}$, are high.

To start, note that, from the perspective of the social planner who knows that π is drawn from the distribution f(π), the expected lengths of resolution of disputes (number of periods) for regimes NC and LTC are given by

$$\begin{aligned} L^{NC} &= \int_0^{\pi^{NC}} f(x)dx + \int_{\pi^{NC}}^1 \{x + 2(1-x)\} f(x)dx \\ &= 1 + \int_{\pi^{NC}}^{\pi^{LTC}} (1-x) f(x)dx + \int_{\pi^{LTC}}^1 (1-x) f(x)dx \end{aligned} \qquad (4),$$

and

$$\begin{aligned} L^{LTC} &= \int_0^{\pi^{LTC}} f(x)dx + \int_{\pi^{LTC}}^1 \{x + 2(1-x)\} f(x)dx \\ &= 1 + \int_{\pi^{LTC}}^1 (1-x) f(x)dx \end{aligned} \qquad (5),$$

respectively. Both expressions tell us that plaintiffs resolve the dispute in at most two periods. For π smaller than the cutoff, all plaintiffs accept the defendant's first period settlement offer. For π larger than the cutoff, low type plaintiffs resolve the dispute in one period while high types proceed to the second period—litigation. By inspection, we notice that (4) and (5) are dissimilar only in the limits of the integrals (the cutoffs). As we know that $\pi^{LTC} > \pi^{NC}$, it is straightforward to conclude that disputes under Regime LTC are resolved more quickly than are disputes under Regime NC, because equation 5 (Regime LTC) has one fewer term than does equation 4 (Regime NC) (corresponding to the high type plaintiffs who solve the dispute in one period under Regime LTC as opposed to two periods under Regime NC). This is valid for all possible belief distributions, f(π), about the type of plaintiff.

We proceed in the same way to compare the length of dispute resolution under Regime NC and HTC. First, we write the expression for the length of dispute resolution under Regime HTC:

$$L^{HTC} = \int_0^{\pi^{HTC}} f(x)dx + \int_{\pi^{HTC}}^1 \left\{(1-p^{LP}+3p^{LP})x + \left(4p^D + 3(1-p^D)\right)(1-x)\right\} f(x)dx$$
$$= 1 + \int_{\pi^{HTC}}^1 \left(\frac{2(\bar{x}-x_L+k)}{\bar{x}-x_L} + \frac{\delta^2\bar{x}-x_L}{\delta^2(\bar{x}-x_L)}\right)(1-x)f(x)dx \qquad (6),$$

The expression tells us that while in a pooling equilibrium (π smaller than the cutoff) all the plaintiffs solve the dispute in one period, in a separating equilibrium (π larger than the cutoff) the low type plaintiffs settle their dispute in one period with probability $1-p^{LP}$ but in three periods with probability p^{LP}, and high types either settle in period 3 with probability $1-p^D$ or go to trial in period 4 with probability p^D.

Then, if we rewrite (4) as

$$L^{NC} = 1 + \int_{\pi^{NC}}^{\pi^{HTC}} (1-x)f(x)dx + \int_{\pi^{HTC}}^1 (1-x)f(x)dx \qquad (4'),$$

we notice that there are two dissimilarities between (4') and (6) which push the length of resolution of disputes in different directions. First, the difference in the limit of the

integral tells us that all the high type plaintiffs with $\pi \in \left[\pi^{NC}, \pi^{HTC}\right]$ who used to go to trial under Regime NC in the second period, under Regime HTC settle in the first period with certainty.[28] That first effect reduces the length of resolution of disputes under Regime HTC. Second, the difference in the argument of the integral tells us that the low type plaintiffs with $\pi \geq \pi^{HTC}$ may take longer than one period to solve their dispute while high type plaintiffs with $\pi \geq \pi^{HTC}$ resolve their disputes in the third or fourth and not in the second period as was the case in Regime NC and is seen in $(4')$. That second effect increases the length dispute resolution under Regime HTC. In Appendix B we show that the second effect dominates the first and thus the resolution length in Regime HTC is longer than that in Regime NC under certain conditions: (1) The discount factor, δ, and the expectation of a strike down, α, are not too low, and (2) the cost of negotiation is not too high. The result is valid regardless of the values of the recoveries, the cap, and the distribution of beliefs about plaintiff's type.

The reason the results are conditional on the values of c, δ and α is as follows: If the costs of negotiation, c, are not very high the defendant is more inclined to make an offer that will induce the plaintiff to wait for a second round, because that extra round of negotiations is not too expensive. On the other side, if the discount factor and the expectation of a strike down, δ and α respectively, are large, the defendant is more inclined to make an offer that will induce the plaintiff to wait for a second round, because the offer that the plaintiff demands for *not waiting*, $\delta^3 \bar{x}$, becomes larger.

The following proposition summarizes the analysis above.

Proposition 1: (Time of dispute resolution)

a) Expected dispute resolution is shorter in Regime LTC than in Regime NC.

b) There exists $(\bar{\bar{c}}, \bar{\bar{\alpha}}, \bar{\delta})$ such that for all (c, α, δ) satisfying $c < \bar{\bar{c}}$, $\alpha > \bar{\bar{\alpha}}$ and $\delta > \bar{\delta}$ the expected time of dispute resolution is longer in Regime HTC than in Regime NC.

Proof: See Appendix B.

[28] There are no differences for the low type plaintiffs.

REMARK As was mentioned in footnote 10 above Avraham and Bustos (2008) show some preliminary empirical evidence that supports the veracity of Proposition 1.

4.2 Welfare Implications

In this Section we consider the welfare implications of our model. We ask questions such as whether caps increase total legal costs, or cause plaintiffs to recover more or less. In order to answer these questions, we start by noticing that caps (regardless of the size of the expected trim) tend to increase the percentage of disputes that are settled rather than litigated. Later, we show that, relative to Regime NC, Regime LTC tends to reduce litigation expenses while Regime HTC tends to *increase* them. Finally, we show that in both Regime LTC and HTC, plaintiffs with high value claims typically end worse off while plaintiffs with low value claims may end better or worse off compared to similar plaintiffs in Regime NC.

4.2.1 Proportion of Disputes Settled

As shown above, it is commonly thought that caps will increase the fraction of disputes that are settled instead of litigated. Because the expected recovery in a trial is smaller, the parties would prefer to settle and save the cost of trials more frequently. Our model offers a slightly different reason why caps increase the fraction of disputes settled. In our model, caps drive defendants to make *high offers* (offers that induce both types of plaintiffs to settle in the first period) more frequently than in the no-caps case because those "high offers" required to induce pooling equilibria are lower when caps are in place.[29]

More formally, under Regime NC, trials take place only for $\pi > \pi^{NC}$ and a high type plaintiff. Under Regime LTC, trials take place only for $\pi > \pi^{LTC}$ and a high type plaintiff. Because $\pi^{LTC} > \pi^{NC}$, it is clear that the set of pairs of defendants and plaintiffs

[29] In addition, in Regime HTC the high type plaintiff may decide to settle in the third period whereas in Regime NC, if the high type does not settle in the first period she would go to trial in the second period.

that resolve their disputes through settlement is higher in the regime with caps and low trim than in the no-caps regime. Analogously, under Regime HTC trials take place with probability p^D only for $\pi > \pi^{HTC}$ and a high type plaintiff. Because $\pi^{HTC} > \pi^{NC}$, it is clear that the set of pairs of defendants and plaintiffs that resolve their disputes through settlement is higher in the regime with caps and high trim than in the no-caps regime.

4.2.2 Do Caps Increase or Decrease Litigation Expenses?

Given that disputes are solved more quickly and are more likely to be settled than litigated under Regime LTC than under Regime NC, it is not surprising that litigation expenses (costs of negotiation and trial) are lower under Regime LTC. However, the same does not hold for the comparison between regimes NC and HTC. Proposition 1 proved that regime HTC may have longer times of resolution of disputes than regime NC, ergo we may expect that litigation expenses will increase as well. Although we will find that indeed that is the case, we show that a longer time of resolution of disputes is not enough to conclude that legal costs are larger. For example, if the reduction in the proportion of trials that take place in the second round is large enough to dominate the increment in the negotiation costs (both characteristics of the HTC solution) then, the total legal costs will be smaller.

To proceed we first write the expressions for the expected costs of litigation in regimes NC and HTC respectively

$$E^{NC} = c + \int_{\pi^{NC}}^{\pi^{HTC}} (1-x)\delta k f(x)dx + \int_{\pi^{HTC}}^{1} (1-x)\delta k f(x)dx \qquad (7)$$

$$E^{HTC} = c + \int_{\pi^{HTC}}^{1} \{ xp^{LP}\delta^2 c + (1-x)\left(\delta^2 c + p^D \delta^3 k\right)\} f(x)dx \qquad (8)$$

There are three main dissimilarities between (7) and (8). First, disputes are settled instead of litigated in a higher proportion in Regime HTC than in Regime NC. This "settlement effect" is captured by the expression $\int_{\pi^{NC}}^{\pi^{HTC}} (1-x)\delta k f(x)dx$, which appears in (7) but not in (8) and represents the extra costs of litigation for Regime NC. Second, trials

under Regime HTC take place at the fourth instead of the second period. This implies that the trial costs are lower under Regime HTC. This "trial effect" is captured by the difference between $(1-x)p^D\delta^3 k$ in (8) and $(1-x)\delta k$ in (7) ($(1-x)p^D\delta^3 k < (1-x)\delta k$). Third, disputes may be resolved over longer periods of time, which implies additional costs of negotiation under Regime HTC. This "length effect" is captured by the expression $xp^{LP}\delta^2 c + (1-x)\delta^2 c$, which is in (8) but not in (7).

While the first two effects (the "settlement effect" and the "trial effect") tend to decrease litigation expenses under Regime HTC relative to Regime NC, the third effect (the "length effect") tends to increase them. As a result, caps may actually increase rather than reduce litigation expenses.

Proposition 1 offers the starting point for determining when the "length effect" will dominate the "settlement" and "trial" effects. If c = 0, δ =1 and α =1, then total legal expenses are equal under both regimes. The reason is that the "length effect" disappears when c = 0, the "trial effect" disappears when δ = 1 and the "settlement effect" disappears when in addition to c = 0 and δ =1 it is also true that α =1.[30] Once one realizes that E^{HTC} is a concave function on the value of c[31], and is a strictly increasing function of δ and α, but that those same parameters do not affect E^{NC}, one can conclude that Regime HTC generates more litigation expenses than Regime NC if the negotiation costs are small enough and/or the discount factor together with the expectations of a strike down are large enough.

The following proposition summarizes the analysis above.

Proposition 2: (Litigation expenses)

a) Expected litigation expenses are smaller in Regime LTC than in Regime NC.

[30] The recovery of the high type plaintiff is x_H regardless of the value of the cap because the cap will be struck down for sure. As we mentioned before, when the parameters take these particular values we have that $\pi^{NC} = \pi^{HTC} = k/(x_H - x_L + k)$.

[31] It increases for small values of c but decreases for large values of c

b) There exists $(c^*, \alpha^*, \delta^*)$ such that for all (c, α, δ) satisfying $c < c^*$, $\alpha > \alpha^*$ and $\delta > \delta^*$ the expected litigation expenses are larger in Regime HTC than in Regime NC.

Proof: See Appendix B.

A comparison of $(c^*, \alpha^*, \delta^*)$ from Proposition 2 with $(\bar{c}, \bar{\alpha}, \bar{\delta})$ from Proposition 1 shows that because $L^{HTC}(c = 0, \delta = 1, \alpha = 1) > L^{NC}$ but $E^{HTC}(c = 0, \delta = 1, \alpha = 1) = E^{NC}$ then $\bar{c} > c^*$, $\bar{\alpha} < \alpha^*$ and $\bar{\delta} < \delta^*$. That means that for all cases in which $c \in [c^*, \bar{c}]$ the expected time of resolution of disputes (all else being equal) is longer in Regime HTC than in Regime NC but the litigation expenses are *not* increased. The same holds for all cases in which $\alpha \in [\bar{\alpha}, \alpha^*]$ and/or $\delta \in [\bar{\delta}, \delta^*]$. This shows that the longer time of resolution of disputes in Regime HTC than in Regime NC does not necessarily imply that the legal costs are larger.

We acknowledge that Proposition 2 only takes into account short-term effects. That is, the additional litigation costs incur within the time period in which the uncertainty surrounding the cap is unresolved.[32] This means that even if HTC induces short term increments in litigation expenses, those increments may be offset by long term reductions in expenses (in the future after the uncertainty is resolved and the HTC regime becomes either LTC or NC *permanently*). Accordingly, society might not lose by enacting caps. While this observation is correct, one has to remember that as a matter of fact, the average time to resolving the uncertainty surrounding the constitutionality of caps in the U.S. is about 10 years (See Avraham [2006]). Moreover, as it is stated in the next lemma and proved in appendix B, even when we take into account long-term effects, part (b) of Proposition 2 still holds.

Lemma 3: If there is an infinite sequence of pairs of plaintiffs and defendants such that the uncertainty about the cap is resolved in the game played by the first pair then there

[32] We thank Tom Kelly for pointing this out to us.

exists $(c^*, \alpha^*, \delta^*)$ such that for all (c, α, δ) satisfying $c < c^*$, $\alpha > \alpha^*$ and $\delta > \delta^*$ the expected time of dispute resolution is longer in a Regime HTC than in a Regime NC.
Proof: See Appendix B.

Evidently, the parameters in Lemma 3 relative to Proposition 2 are more stringent. That is, (α^*, δ^*) which are defined by Lemma 3 are larger than the equivalent values which were defined by Proposition 2, and (c^*) which was defined by Lemma 3 is smaller than the equivalent value which was defined by Proposition 2. This is because after a period in which society experiences a welfare loss due to uncertainty in the constitutionality of the caps (increase in litigation expenses), there is a period in which it experiences a gain due to the elimination of that uncertainty (decrease in litigation expenses). Lemma 3 shows that there always exist cases in which the aggregate effect is negative.

4.2.3 Do Caps Increase or Decrease Plaintiffs' Recoveries?

At this point we wonder whether caps have any systematic effects on the recoveries obtained by plaintiffs. One may expect that plaintiffs should be able to recover less in states with caps. Indeed, that is the overall effect. Nevertheless, when we distinguish by the type of the plaintiff, we uncover an unexpected result. Unlike high type plaintiffs, who always recover less when caps are in place, low type plaintiffs may recover the same, more, or less when caps are in place, depending on various factors.

Some low type plaintiffs may recover less because high type plaintiffs also recover less. Recall that in a pooling equilibrium low type plaintiffs recover the same dollar amount as high type plaintiffs. Hence, low type plaintiffs who would have been pooled with high type plaintiffs in any case will recover in any of the caps regimes less than they would in Regime NC.

Some low type plaintiffs may recover more because some low type plaintiffs are included in pooling equilibria under the caps regimes (low types recover δx^c in LTC and $\delta^3 \bar{x}$ in HTC in the case of pooling equilibria), rather than falling under the separating

equilibria as they would under the no caps regime (where low types recover only δx_L). We know this effect exists because the cut-off that divides pooling from separating equilibrium is larger under the cap regimes than the no cap regime. More formally

Proposition 3: (Recoveries)

a) For all values of π, high type victims receive lower recoveries in regimes with caps (whether LTC or HTC) than in a regime without caps.

b) For all values of $\pi < \pi^{NC}$, low type victims receive smaller recoveries in regimes with caps (whether LTC or HTC) than in a regime without caps. For all values of $\pi \in \left[\pi^{NC}, \max(\pi^{LTC}, \pi^{HTC})\right]$ low type victims receive larger or equal (expected) recoveries in regimes with caps (whether LTC or HTC) than in a regime without caps. For all values of $\pi > \max(\pi^{LTC}, \pi^{HTC})$ low type victims receive equal (expected) recoveries in regimes with caps (whether LTC or HTC) and without caps.

Proof: See Appendix B.

REMARK. That high-type plaintiffs would be under-compensated if subjected to caps was observed by many. For example Viscusi (1991) pp 97 argued that the effect of caps is that "victims with major injuries would be limited in making their claims while those with minor injuries would be unaffected." As a result victims of brain damage, para- or quadriplegia, and cancer will be the most disadvantaged. For these reasons, among others, a study by the American Legal Institute (ALI) rejected caps (See ALI [1991] pp 219-220). Our study is the first to formally show (in addition to the under-compensation of high-type plaintiffs) the possibility that caps will cause victims with minor injuries to be over-compensated, and not simply "unaffected".

4.3 Discussion: Which states should enact caps?

Our analysis suggests that states' legislators should be cautious when enacting caps because they might decrease total welfare by increasing litigation costs. To see that in more detail we go back to figure 2 and discuss the optimal decision of four hypothetical states denoted in the figure with the letters A, B, C and D.

To simplify the discussion we can assume that the discount factor δ is larger than δ^* and all the states face the same costs of negotiation c which are smaller than c^*.

A state such as the one denoted A (which, for example, may correspond to Montana in Avraham and Bustos (2008)) which enacted a relatively small cap and faces a low probability that its Supreme Court would strike down the cap, will probably benefit from the enactment. The reason is that although the cap defines a high trim equilibrium the low value of α not only implies large "settlement" and "trial" effects but in addition a small "length" effect.[33] In other words, the reduction in trial costs generated by the cap is larger than the increase in negotiation costs also generated by the cap.[34]

A state such as B (which, for example, may correspond to Maine) which faces the same probability of a strike down as state A but enacted a middle level cap, benefits with certainty from the cap. The obvious reason is that parameters (α, x^c) define a low trim equilibrium.

Interestingly, a state such as D (which, for example, may correspond to North Dakota) also benefits with certainty from the cap as the equilibrium is low trim. Nevertheless, because in the case of North Dakota there would be more separating equilibria than in the case of Maine,[35] more cases will be decided by a judge instead of settled and so litigation costs in North Dakota may be higher than in Maine. This result is completely driven by the difference in the size of the cap; expectations regarding the strike down do not play any role.

Finally, a state such as C (which, for example, may correspond to Illinois) is the only one in our hypothetical analysis which does not benefit from enacting a cap. C has the same α as D and the same x^c as B, two states with low trim equilibrium, nevertheless, both parameters taken together, define a high trim equilibrium. One may wonder why D and A are so different: Whereas in A, even when the equilibrium was also high trim we

---

[33] From Lemma 1 we know that π^{HTC} is decreasing in α which guarantees a high settlement effect. It is easy to show that p^D is increasing with α which guarantees a high trial effect. Finally, it is also straight forward to see that p^{LP} is decreasing in α which implies a low length effect.

[34] Strictly speaking, the result holds if and only if $\alpha < \alpha^*$.

[35] Because π^{LTC} decreases with x^c.

expected a social gain, in D we do not. The answer is that in D, α is large enough to guarantee that the increase in negotiation costs dominates the reduction in trial costs.[36]

5 Conclusions

That parties delay settlements in the shadow of caps may seem counterintuitive. After all, caps reduce the uncertainty associated with jury awards, and are therefore expected to ease settlements. In contrast, we showed that if the parties expect that caps will be struck down in the near future, they might delay settlement. In Avraham and Bustos (2008) we test empirically some of our predictions and find supporting evidence. There we show that a) when the caps are relatively small and the probability of their strike down is large, parties delay settlements, until the fate of the caps becomes clear, and b) that when the caps are high and the probability of their strike down is small, parties will expedite their settlements, relative to states with no caps.

From a welfare perspective, Proposition 2 is probably the most important theoretical finding of this paper: Low Expected Trim Caps decrease legal expenses whereas High Expected Trim Caps *may* increase them. While our model deals with the short run, the insight that there exists an ex-ante cutoff probability of a strike-down due to unconstitutionality of the caps, α, above which enacting caps will be welfare decreasing remains true even when the long run is considered. But that cutoff will naturally be higher the longer the time period considered. An intuitive policy implication emerges from this analysis: States legislatures that believe the high court of their state is highly likely to strike down the reform, and care much about the short term effects of settlement delays, may be better off enacting relatively high caps or not enacting them at all.

Our model suggests that caps hit plaintiffs with large claims the hardest because they either receive lower recoveries if the caps are struck down, or delayed settlements if the caps are not. Interestingly, the model predicts that, in some circumstances, High Expected Trim caps may make plaintiffs with small claims (perhaps plaintiffs with frivolous lawsuits) better off as it enables them to sometimes receive the same settlement offer that plaintiffs with high claims receive, which is higher than what they would get in regime

[36] As in footnote 35, strictly speaking, the result holds if and only if $\alpha > \alpha^*$.

with no caps. Since there are no nation-wide datasets which contain plaintiffs' original claims, nor do we have data on plaintiffs' characteristics, it seems difficult to empirically test these predictions.[37]

References

ALI (1991) American Legal Institute, Enterprise Responsibility for Personal Injury: Approaches to Legal and Institutional Change, Volume II.

Arlen Jennifer, *Tort Damages: A Survey*, in ENCYCLOPEDIA OF LAW & ECONOMICS (Boudewijn Bouckaert & Gerrit De Geest, eds., 2000).

Atiyah, P.S. (1980). *Accidents, Compensation and the Law*. (3rd Ed).

Avraham, R. (2006a) Datbase on State Tort Law Reforms, available at ssrn.com

Avraham, R. (2006b). "Putting a Price on Pain-and-Suffering Damages: A Critique of the Current Approaches and a Preliminary Proposal for Change", *Northwestern Univ. Law Rev.*, 100, pp 87.-120.

Avraham, R. (2007). "The Impact of Tort Reforms on Medical Malpractice Settlement Payment" *Journal of Legal Studies* 36 (2), pp 183-229.

Avraham, R. and A. Bustos (February 2008) "The Unexpected Effect of Tort Reform: Do Caps Accelerate Settlements?" (on file with authors).

Babcock L and Pogarsky G. (1999) "Damage Caps and Settlement: A Behavioral Approach", *Journal of Legal Studies*, 28 341–70.

Bebchuk, L. A. (1984) "Litigation and settlement under imperfect information", *RAND Journal of Economics*, 15, pp 404-415.

Bebchuk, L. A. (1988) "Suing solely to extract a settlement offer", *Journal of Legal Studies*, 17, pp 437-449.

Currie, J., and MacLeod W. Bentley (2008), "First Do No Harm? Tort Reform and Birth Outcomes", *The Quarterly Journal of Economics*, May 2008, pp 795-830.

[37] In a previous study, Avraham (2007) showed that cases subject to caps on non-economic damages that were not struck down have lower average recoveries by 65 to 72 percent. (Yet, Avraham (2007) did not distinguish between high claims victims and low claims victims. In an empirical analysis (not reported here) we defined claims as being high or low based on the fields of the physicians. We found weak support for the model's predictions. While we find that Low Expected Trim caps significantly reduce high-claim plaintiffs' recoveries, we find that High Expected Trim Caps increased the recovery of victims with low claims by 10%, yet this increase was not found to be significant.

Danzon, P.M., (2000). "Liability for Malpractice," In Anthony J. Culyer and Joseph P. Newhouse, eds., Handbook of Health Economics. New York: Elsevier.

Daughety, A.F. (2000) 'Settlement', In *Encyclopedia of Law and Economics, Volume 5*. B. Bouckaert and G. De Geest, eds., Edward Elgar Publishing Co.

Daughety, A. F. and Reinganum, J.F. (1994) "Settlement negotiations with two-sided asymmetric information: Model duality, information distribution and efficiency", *International Review of Law and Economics*, 14, pp 283-298.

Kessler, D.P. (1996). "Institutional Causes of Delay in the Settlement of Legal Disputes", *Journal of Law, Economics, and Organizations*, 12(2), pp 432-460.

Kessler, D.P., and Daniel L. Rubinfeld. (2004). ''Empirical Study of the Civil Justice System.'' NBER Working Paper 10825.

Landes, W.M. (1971) "An economic analysis of the courts", *Journal of Law and Economics*, 14, pp 61-107.

Nalebuff, B. (1987) "Credible pretrial negotiation", *RAND Journal of Economics*, 18, pp 198-210.

Parsons C. (2005) "Trial lawyers target cap on malpractice", Chicago Tribune (Aug 25, 2005).

Posner, R.A. (1973) "An economic approach to legal procedure and judicial administration", *Journal of Legal Studies*, 2, pp 399-458.

Priest, G. and Klein, B. (1984), "The selection of disputes for litigation", *Journal of Legal Studies*, 13, pp 1-55.

Reinganum, J. and Wilde, L. (1986) "Settlement, litigation, and the allocation of litigation costs", *RAND Journal of Economics*, 17, pp 557-68.

Rubin, P. (1993). *Tort Reform by Contract*. Published by the American Enterprise Institute for Public Policy Research (Aei Pr), Washington DC.

Schweizer, U. (1989) "Litigation and settlement under two sided incomplete information", *Review of Economic Studies*, 56, pp 163-177.

Sieg, H. (2000). "Estimating a Bargaining Model with Asymmetric Information: Evidence from Medical Malpractice Disputes", *Journal of Political Economy*, 108(5), pp. 1006-1021.

Spier, K. (1992). "The Dynamics of Pretrial Negotiations", *Review of Economic Studies*, 59 (1), pp 93-108.

Spier, K. (2005). Litigation. *The Handbook of Law and Economics*. Chapter 4, Vol 1.

Zukerman S., Koller C. and Bovbjerg R. (1986). "Information on Malpractice: A Review of Empirical Research on Major Policy Issues", 49, *Law and Contemporary Problems*, 85, pp101-03.

U.S. Congress, Office of the Technology Assessment. 1993. Impact of Legal Reforms on Medical Malpractice Costs, at 57-75, 105-111.

U.S. Congress, Congressional Budget Office. 2004. The Effects of Torts Reform: Evidence from the States.

U.S Department of Health and Human Services. 2003. Addressing the New Health Care Crisis: Reforming the Medical Litigation System to improve the Quality of Health Care

Viscusi, K. (1991). *Reforming products liability*. Harvard University Press.

Watanabe, Y. (2006). "Learning and Bargaining in Dispute Resolution: Theory and Evidence from Medical Malpractice Litigation", working paper, Kellogg School of Management.

Tables, Diagrams and Figures

Table 1- Caps on Non-Economic Damages Enacted or Struck-Down Between 1991-2005			
State	Cap Size	Enacted	Struck-Down
AL	400		1991
IL	500	1995	1997
MT	250	1995	
ND	500	1995	
SD	500	1996	
OH	500	1997	1999
OR	500		1999
ME	400	2000	
MS	500	2003	
OH	500	2003	
OK	300	2004	
TX	250	2004	
FL	450	2004	
TN	250	2005	
NV	350	2005	
WI	350		2005

Diagram 1- Resolution of uncertainty in states with caps

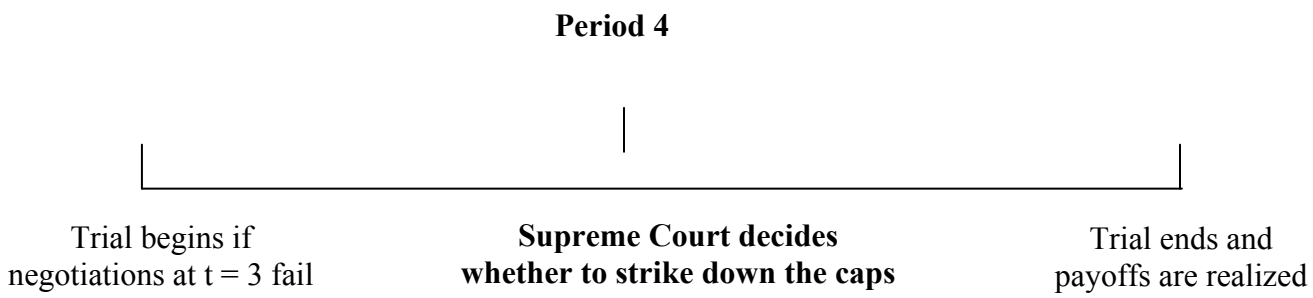


Figure 1: Equilibria for states with and without caps

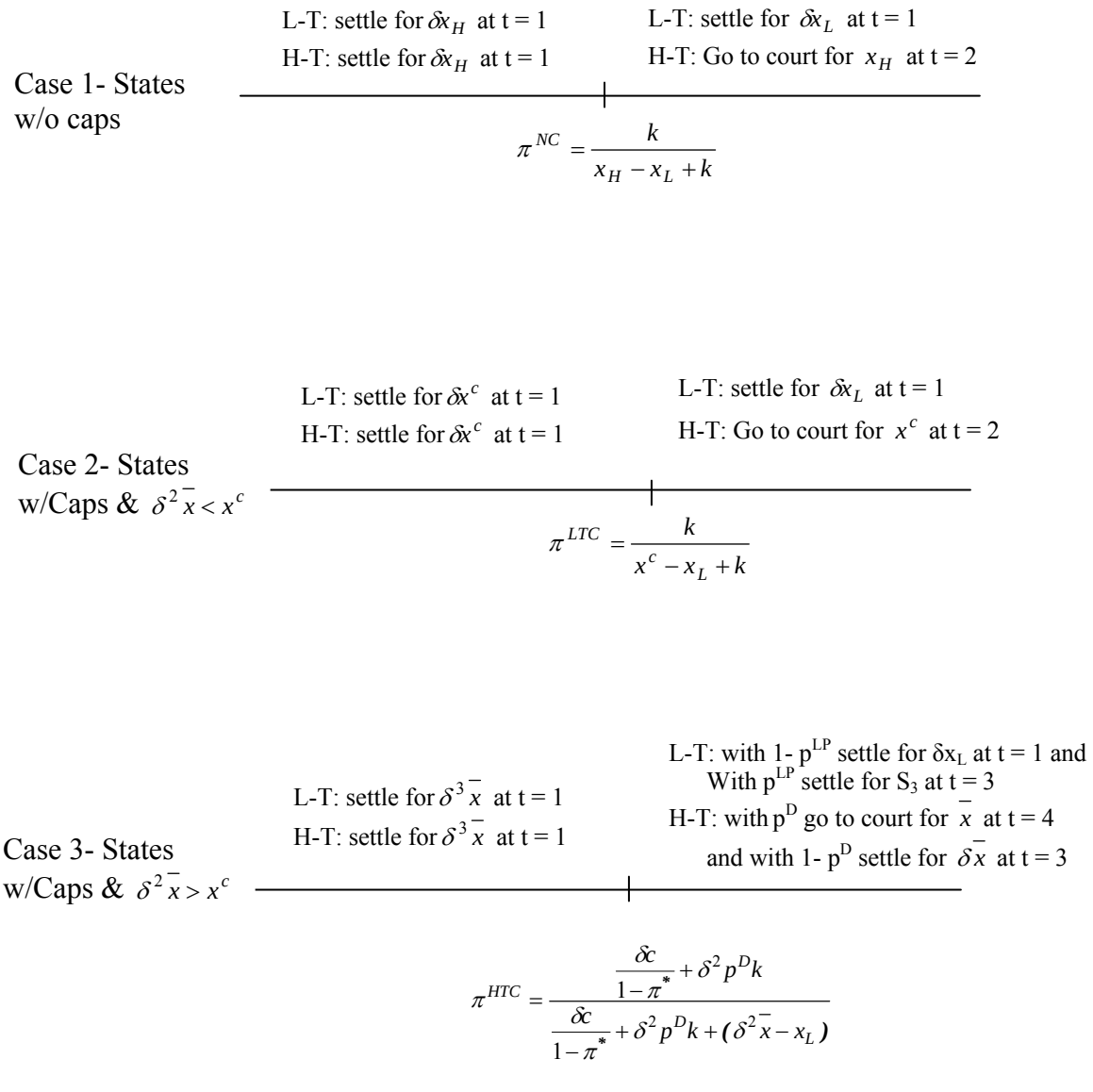
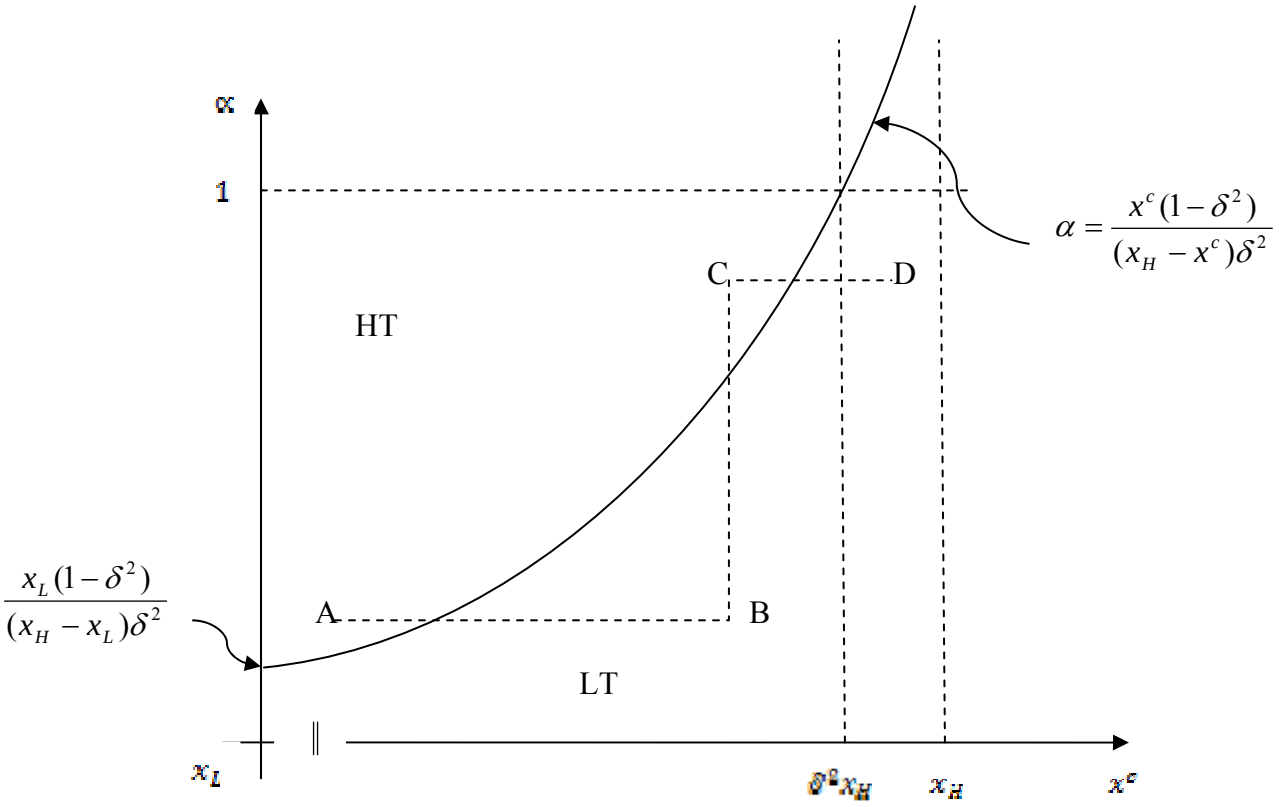


Figure 2: HTC and LTC equilibria in the space (α, x^c)



Appendix A: Characterization of the Equilibria

Solution for states without caps

At t = 4. High type plaintiff recovers x_H while low type plaintiff recovers x_L. By its side, the defendant pays recovery plus negotiation and litigation costs c + k.

At t = 3. We denote π_3 the Bayesian update of the probability that the plaintiff is the low type at the beginning of the third period. The cost faced by the defendant C conditional on the third period settlement offer S_3 is given by

If $S_3 < \delta x_L$ then $C = \delta[\pi_3 x_L + (1-\pi_3) x_H + k] + c$ (O1)
If $S_3 \in [\delta x_L, \delta x_H)$ then $C = \pi_3 S_3 + \delta[(1-\pi_3)(x_H + k)] + c$ (O2)
If $S_3 \geq \delta x_H$ then $C = S_3 + c$ (O3)

In the calculation of C, we used the fact that no plaintiff accepts a settlement offer lower than the discounted value of his harm (what she obtains at period 4). That structure of settlement offers implies that $S_3 = \delta x_L$ if and only if $\pi_3 > k/(x_H - x_L + k)$, otherwise $S_3 = \delta x_H$ (to see that note that $S_3 = \delta x_L$ in (O2) dominates any offer in (O1) or (O2). Hence, in order to determine when he should offer δx_L or δx_H the defendant only needs to determine when $\pi_3 \delta x_L + \delta[(1-\pi_3)(x_H + k)]$ is smaller or larger than δx_H). In the case, the defendant offers $S_3 = \delta x_L$, the low type plaintiff settles while the high type goes to trial, in the case that he offers $S_3 = \delta x_H$, both settle. The negotiation costs are irrelevant.

At t = 2. High type plaintiff recovers x_H while low type plaintiff recovers x_L. High type plaintiff never waits for t = 3 because she knows that at that time she can only recover δx_H. The dominant strategy of the high type immediately implies that $\pi_3 = 1$, hence the low type plaintiff doesn't have incentives to wait for t = 3 as she gets $S_3 = \delta x_L$ with certainty. By its side, at t = 2, the defendant pays recovery plus litigation cost k.

At t = 1. The high (low) type plaintiff accepts all settlement offers higher or equal than δx_H (δx_L) while he rejects all inferior offers. As the defendant knows that the plaintiff can get these same amounts in the second period but in that case he pays the litigation cost, he induces either both types or the low type to settle immediately. In other words, he makes settlement offer δx_L when $\pi \delta x_L + (1-\pi)\delta(x_H + k) + c < \delta x_H + c$ and settlement offer δx_H when $\pi \delta x_L + (1-\pi)\delta(x_H + k) + c > \delta x_H + c$. Or;

If $\pi > k/(x_H - x_L + k) = \pi^{NC}$ then $S_1 = \delta x_L$. In that case, while the low type plaintiff settles in the first period, the high type plaintiff litigates in the second and gets δx_H. Neither of them wants to wait for a third period because in that case they get the same amounts.[38]

[38] Notice that plaintiffs' strategy of randomizing between first period settlement and second period trial doesn't support a mixed-strategies equilibrium (after all they get the same payoff) because in that case the

If $\pi < k/(x_H - x_L + k) = \pi^{NC}$ then $S_1 = \delta x_H$. Both types settle in the first period.

Solution for states with caps

Unlike in the solution for states without caps, here we distinguish two cases

First Case: If $\delta^2 \bar{x} < x^c$ (caps and low expected trim) then

At t = 4. The solution is the same that in the case without caps but instead of recovering x_H the high type expects to recover $\bar{x}$.

At t = 3. The solution is the same that in the case without caps but instead of offering δx_L and δx_H the defendant offers δx_L and $\delta \bar{x}$.

At t = 2. The solution is the same that in the case without caps but instead of recovering x_H the high type expects to recover $\bar{x}$.

At t = 1. The strategies of the defendant and the plaintiffs are the same that in the case without caps with the following exceptions. First, instead of offering δx_L and δx_H the defendant offers δx_L or δx^c and second, the high type plaintiff accepts all settlement offers higher or equal than δx^c (not δx_H) while litigate all inferior offers. The settlement offer in the first period depends on the probability that the plaintiff is a low type.

If $\pi > k/(x^c - x_L + k) = \pi^{LTC}$ then $S_1 = \delta x_L$. While the low type plaintiff settles, the high type gets more going to trial. Both actions take place in the first period.

If $\pi < k/(x^c - x_L + k) = \pi^{LTC}$ then $S_1 = \delta x^c$. Both types settle in the first period.

Second Case: If $\delta^2 \bar{x} > x^c$ (caps and high expected trim) then

At t = 4. High type plaintiff expects to recover $\bar{x}$ while low type plaintiff recovers x_L. By its side, the defendant pays recovery plus negotiation and litigation costs c + k.

---

defendant can always increase the settlement offer by ε, induce the plaintiffs to accept immediately and save the extra negotiation and litigation costs.

At t = 3. As in the case without caps, the plaintiffs settle only if they get at least the discounted value of their expected recovery at trial in t = 4. That is δx_L and $\delta \bar{x}$ respectively. The strategy followed by the defendant is the same as in the case without caps with two differences: first, instead of offering δx_L and δx_H he offers δx_L and $\delta \bar{x}$; second, it may be the case that he randomizes between these two offers. More specifically

$$S_3 = \begin{cases} \delta x_L \text{ if } \pi_3 > \pi^* \\ \delta \bar{x} \text{ if } \pi_3 < \pi^* \\ \text{randomizes between } \delta x_L \text{ and } \delta \bar{x} \text{ if } \pi_3 = \pi^* = \dfrac{k}{\bar{x} - x_L + k} \end{cases}$$

The defendant could randomize at t = 3 because at time t = 2 both plaintiffs may decide to wait for the second settlement offer. The high type has incentives to wait for a third period if she gets offer δx_L at t = 1 as she cannot recover more than δx^c by going to trial at t = 2 and the low type has incentives to wait for a third period because she can get more by mimicking the high type.

We define p^D as the probability that the defendant offers δx_L in the third period, p^{LP} as the probability that the low type plaintiff waits for the third period if she receives offer δx_L and p^{HP} as the probability that the high type plaintiff waits for the third period if she receives offer δx_L. Although we need to wait for the considerations made by the defendant at period 1 to determine the exact value of these probabilities, at this point we notice that due to Bayes rule

$$\pi_3 = \frac{p^{LP}\pi}{p^{LP}\pi + p^{HP}(1-\pi)} = \pi^*$$

Hence it is true that

$$p^{LP} = \frac{\pi^*}{1-\pi^*}\frac{1-\pi}{\pi}\, p^{HP}$$

At t = 2. High type plaintiffs never go to trial because by waiting for a second round of settlement/trial they guarantee a recovery of $\delta^3 \bar{x}$ which is larger than δx^c. Low type plaintiffs wait for period 3 if and only if $\pi_3 \le \pi^*$ (we will see in the analysis at t =1 that it is always true that $\pi_3 = \pi^*$, hence the low type plaintiff never goes to trial. You may think that that cannot be true because in the case of a mix strategies solution, the low type gets δx_L both in the second and third periods, consequently the defendant can always save future litigation costs by offering the plaintiff ε more either at t =1 or t = 2 and inducing her to accept immediately. But, recall that in the second period the plaintiff cannot get more than her true harm, and notice that there is no equilibrium if the defendant always accepts the first period offer.

At t = 1. The defendant chooses between offering δx_L and $\delta^3 \bar{x}$. If the probability that the plaintiff is low type is small enough, the defendant offers $\delta^3 \bar{x}$, because in that case both plaintiffs accept it right away (as they don't expect to get more if they wait) and the defendant saves in future litigation expenses. Instead, if the probability that the plaintiff is low type is high enough the defendant offers δx_L because in that case some plaintiffs accept and the defendant saves in settlement payments. Hence, there exists π^{HTC} such that

$$S_1 = \begin{cases} \delta x_L \text{ if } \pi > \pi^{HTC} \\ \delta^3 \bar{x} \text{ if } \pi < \pi^{HTC} \end{cases}$$

Obviously π^{HTC} is the belief that makes the defendant indifferent between the two offers. More specifically, the belief that satisfies

$$\begin{aligned} &\pi\left[\left(1-p^{LP}\right)\delta x_L + p^{LP}\delta^2\left(p^D\delta x_L + (1-p^D)\delta\bar{x}+c\right)\right] \\ &+(1-\pi)\left[(1-p^{HP})\delta(x^c+k)+p^{HP}\delta^2(p^D\delta(\bar{x}+k)+(1-p^D)\delta\bar{x}+c)\right] + c = \delta^3\bar{x}+c \end{aligned} \quad \text{(A1)}$$

The left hand side expression corresponds to the defendant's expected cost if the offer is δx_L. The right hand side expression corresponds to the defendant's expected cost if the offer is $\delta^3 \bar{x}$. The right hand side expression is completely determined because in the case that the offer is $\delta^3 \bar{x}$ both types of plaintiff accept it immediately. The left hand side expression is not completely determined because we need to calculate the values of p^D, p^{LP} and p^{HP}. First, it is easy to see that $p^{HP} = 1$ because the high type plaintiff never accepts offer δx_L as in the third period she can make $\delta^3 \bar{x}$ (remember that the high type never goes to trial in the second period as $\delta^3 \bar{x} > \delta x^c$.) Second, notice that there is no equilibrium that supports a pure strategy for the low type plaintiffs. For if the low type always accept δx_L, the defendant knows at t = 3 that he is dealing with high types and offer $\delta\bar{x}$. As $\delta^3 \bar{x} > \delta x_L$ the low type has incentives not to accept at t = 1 and then that strategy cannot be an equilibrium. On the other side, if the low type always rejects δx_L and gets $S_3 \geq \delta x_L$ as expected recovery in the third period the defendant can offer $S_3 + \varepsilon$ and induce her to settle in the first period, saving the extra litigation expenses, hence $S_1 = \delta x_L$ cannot be an equilibrium.

Now we calculate the values of p^D and p^{LP}. Given that the low type follows a mixed strategy, she must be indifferent between waiting and settling when he is offered δx_L and that happens if and only if $\delta x_L = \delta^3(p^D x_L + (1-p^D)\bar{x})$. That identity allows us to determine that

$$p^D = \frac{\delta^2\bar{x} - x_L}{\delta^2(\bar{x} - x_L)}$$

In addition, because $p^{HP} = 1$ we know that

$$p^{LP} = \frac{\pi^*}{1-\pi^*}\frac{1-\pi}{\pi}$$

Replacing those expressions in (A1) we get that

$$\pi^{HTC}\left[x_L + \frac{\pi^*}{1-\pi^*}\frac{1-\pi^{HTC}}{\pi^{HTC}}\delta c\right] + (1-\pi^{HTC})\left[\delta^2\bar{x} + p^D\delta^2 k + \delta c\right] = \delta^2\bar{x}$$

which leads us to conclude that:

If $\pi < \dfrac{\frac{\delta c}{1-\pi^*} + \delta^2 p^D k}{\frac{\delta c}{1-\pi^*} + \delta^2 p^D k + (\delta^2\bar{x} - x_L)} = \pi^{HTC}$ then $S_1 = \delta^3\bar{x}$. Both types settle in the first period.

If $\pi > \dfrac{\frac{\delta c}{1-\pi^*} + \delta^2 p^D k}{\frac{\delta c}{1-\pi^*} + \delta^2 p^D k + (\delta^2\bar{x} - x_L)} = \pi^{HTC}$ then $S_1 = \delta x_L$ and $S_3 = \begin{cases} \delta x_L \text{ with } p^D \\ \delta\bar{x} \text{ with } 1-p^D \end{cases}$.

The high type plaintiff always wait for the third period in which case she settles when she receives offer $\delta\bar{x}$ but goes to trial when receives offer δx_L. The low type settles in the first period with probability $p^{LP} = \dfrac{\pi^*}{1-\pi^*}\dfrac{1-\pi}{\pi}$ otherwise she settles in the third period, that is, with probability $1 - p^{LP}$.

Appendix B: Proofs

Proof of Lemma 1:

If we differentiate (3) with respect to c we have that

$$\frac{\partial\pi^{HTC}}{\partial c}(x_L - \delta^2\bar{x}) + (1-\pi^{HTC})\frac{\delta}{1-\pi^*} - \frac{\partial\pi^{HTC}}{\partial\delta}\left[\frac{c\delta}{1-\pi^*} + p^D\delta^2 k\right] = 0$$

which implies that

$$sign\left(\frac{\partial\pi^{HTC}}{\partial c}\right) = sign\left(\frac{1-\pi^{HTC}}{1-\pi^*}\right)\delta = +$$

If we differentiate (3) with respect to δ we have that

$$-\pi^{HTC}2\delta\bar{x}+\frac{\partial\pi^{HTC}}{\partial\delta}(x_L-\delta^2\bar{x})+\frac{(1-\pi^{HTC})}{(\bar{x}-x_L)}\left[c(k+\bar{x}-x_L)+2\delta\bar{x}k\right]$$
$$-\frac{\partial\pi^{HTC}}{\partial\delta}\left[\frac{c\delta(k+\bar{x}-x_L)}{\bar{x}-x_L}+\frac{\delta^2\bar{x}-x_L}{\bar{x}-x_L}k\right]=0$$

which implies that

$$sign\left(\frac{\partial\pi^{HTC}}{\partial\delta}\right)=sign\left(\frac{1-\pi^{HTC}}{\bar{x}-x_L}\left[c(k+\bar{x}-x_L)+2\delta\bar{x}k\right]-2\pi^{HTC}\delta\bar{x}\right)$$
$$=sign\left(\left(\frac{\delta^2\bar{x}-x_L}{\bar{x}-x_L}\right)\left[c(k+\bar{x}-x_L)+2\delta\bar{x}k\right]-2\delta\bar{x}\left[\frac{c\delta(k+\bar{x}-x_L)}{\bar{x}-x_L}+\frac{\delta^2\bar{x}-x_L}{\bar{x}-x_L}k\right]\right)$$
$$=sign\left(\left(\frac{-\delta^2\bar{x}-x_L}{\bar{x}-x_L}\right)\left(k+\bar{x}-x_L\right)c\right)=-$$

If we differentiate (3) with respect to α we have that

$$-\pi^{HTC}\delta^2(x_H-x^c)+\frac{\partial\pi^{HTC}}{\partial\alpha}(x_L-\delta^2\bar{x})+\frac{(1-\pi^{HTC})}{(\bar{x}-x_L)}\left[(x_H-x^c)(x_L(1-\delta^2)-\delta c)k\right]$$
$$-\frac{\partial\pi^{HTC}}{\partial\alpha}\left[\frac{c\delta(k+\bar{x}-x_L)}{\bar{x}-x_L}+\frac{\delta^2\bar{x}-x_L}{\bar{x}-x_L}k\right]=0$$

which implies that

$$sign\left(\frac{\partial\pi^{HTC}}{\partial\alpha}\right)=sign\left(\frac{1-\pi^{HTC}}{\bar{x}-x_L}\left[(x_H-x^c)(x_L(1-\delta^2)-\delta c)k\right]-\pi^{HTC}\delta^2\left(x_H-x^c\right)\right)$$
$$=sign\left(\left(\frac{\delta^2\bar{x}-x_L}{\bar{x}-x_L}\right)\left[(x_H-x^c)(x_L(1-\delta^2)-\delta c)k\right]-\delta^2(x_H-x^c)\left[\frac{c\delta(k+\bar{x}-x_L)}{\bar{x}-x_L}+\frac{\delta^2\bar{x}-x_L}{\bar{x}-x_L}k\right]\right)$$
$$=sign\left(\left(\frac{x_H-x^c}{\bar{x}-x_L}\right)\left(\left(\delta^2\bar{x}-x_L\right)(1-\delta^2)(x_L-k\delta^2)-\delta c\left(k+2\left(\delta^2\bar{x}-x_L\right)\right)\right)\right)=-$$

End Proof.

Proof Proposition 1:

Part (a) The proof is direct by inspection of (4) and (5).

Part (b) Notice that $\pi^{HTC}(c=0,\delta=1,\alpha=1)=\pi^{NC}$. Additionally, as the argument of the integral in (6) is always larger than the argument of the integral in (4) we have that $L^{HTC}(c=0,\delta=1,\alpha=1)>L^{NC}$. Consequently, it is enough to show that L^{HTC} is decreasing in c, it is increasing in δ and it is increasing in α when c = 0 and δ = 1. The conclusion follows from an argument of continuity.

First, notice that

$$\frac{\partial L^{HTC}}{\partial c}=-\left(\frac{2(\bar{x}-x_L+k)}{\bar{x}-x_L}+p^D\right)(1-\pi^{HTC})f(\pi^{HTC})\frac{\partial \pi^{HTC}}{\partial c}<0$$

because $\frac{\partial \pi^{HTC}}{\partial c}>0$ as was shown in Lemma 1. Next, notice that

$$\frac{\partial L^{HTC}}{\partial \delta}=\int_{\pi^{HTC}}^{1}\frac{2(\bar{x}-x_L)x_L}{\delta^3(\delta^2\bar{x}-x_L)}(1-x)f(x)dx-\left(\frac{2(\bar{x}-x_L+k)}{\bar{x}-x_L}+p^D\right)(1-\pi^{HTC})f(\pi^{HTC})\frac{\partial \pi^{HTC}}{\partial \delta}>0$$

because $\frac{\partial \pi^{HTC}}{\partial \delta}<0$ as was shown in Lemma 1 and finally, notice that

$$\left.\frac{\partial L^{HTC}}{\partial \alpha}\right|_{c=0,\delta=1}=\int_{\pi^{HTC}}^{1}\frac{2(x_H-x_L)k}{(\bar{x}-x_L)^2}(1-x)f(x)dx-\left(\frac{2(\bar{x}-x_L+k)}{\bar{x}-x_L}+1\right)(1-\pi^{HTC})f(\pi^{HTC})\frac{\partial \pi^{HTC}}{\partial \alpha}>0$$

because $\left.\frac{\partial \pi^{HTC}}{\partial \alpha}\right|_{c=0,\delta=1}=-\frac{(x_H-x^c)k}{(\bar{x}-x_L)^2}<0$.

End Proof.

Proof Proposition 2:

Part (a) Proposition 2 tells us that the fraction of disputes resolved at trial instead of settlement is larger under Regime NC than under Regime LTC. As trials are more expensive than settlements, it is direct that litigation expenses are higher under Regime NC than under Regime LTC.

Part (b) Expected litigation expenses under Regime NC are

$$E^{NC}=c+\int_{\pi^{NC}}^{1}(1-x)\delta k f(x)dx$$

While expected litigation expenses under Regime HTC are

$$c+\int_{\pi^{HTC}}^{1}\{xp^{LP}\delta^2 c+(1-x)\left(\delta^2 c+p^D\delta^3 k\right)\}f(x)dx$$

$$E^{HTC}=c+\int_{\pi^{HTC}}^{1}(1-x)\left(\frac{\delta^2 c}{1-\pi^*}+\delta^3 p^D k\right)f(x)dx$$

First notice that E^{HTC} coincide with E^{NC} when $c=0;\delta=1;\alpha=1$. Next we show the behavior of $E^{HTC}-c$ with respect to c, δ and α when we start at point $c=0;\delta=1;\alpha=1$.

$$\left.\frac{\partial(E^{HTC}-c)}{\partial c}\right|_{\delta=1,\alpha=1} = \int_{\pi^{HTC}}^{1} \frac{(x_H - x_L + k)}{(x_H - x_L)}(1-x)f(x)dx - \frac{(1-\pi^{NC})(x_H - x_L + k)}{(x_H - x_L)} cf(\pi^{HTC})$$

As the first expression is decreasing in c while the second one increasing in c, we conclude that $E^{HTC} - c$ first increases and later decreases with that variable.

$$\left.\frac{\partial(E^{HTC}-c)}{\partial \delta}\right|_{c=0,\alpha=1} = \int_{\pi^{HTC}}^{1} \frac{2\delta^2 x_H}{\delta^2(x_H - x_L)}(1-x)f(x)dx$$

which is positive for all values of δ and

$$\left.\frac{\partial(E^{HTC}-c)}{\partial \alpha}\right|_{c=0,\delta=1} = (1-\pi^{NC})k^2 \frac{(x_H - x^c)(x_H - x_L)}{(\bar{x} - x_L)^2} f(\pi^{NC})$$

which is positive for all values of α. Then, by a continuity argument, there must exist $(c^*, \delta^*, \alpha^*)$ such that for all $c < c^*$, $\alpha > \alpha^*$ and $\delta > \delta^*$ it is true that LE^{HTC} is larger than E^{NC}.

End Proof.

Proof Proposition 3:

Part (a): In Regime NC high-type plaintiffs recover δx_H for all values of π. In Regime LTC high-type plaintiffs recover δx^c for all values of π. In Regime HTC high-type plaintiffs recover $\delta^3 \bar{x}$ for all values of π. As $x_H > \delta^2 \bar{x} > x^c$ then the result follows.

Part (b): For all values of $\pi < \pi^{NC}$ the low-type plaintiffs recover δx_H in Regime NC which is larger than the maximum recovered in Regimes LTC and HTC given by $\max\{\delta^3 \bar{x}, \delta x^c\}$. For all values of $\pi \in [\pi^{NC}, \max\{\pi^{HTC}, \pi^{LTC}\}]$ the low-type plaintiffs recover δx_L in Regime NC which is smaller than δx^c which is what they recover in Regime LTC and smaller than $\delta^3 \bar{x}$ which is what they recover in Regime HTC. Finally, for all values of $\pi > \max\{\pi^{HTC}, \pi^{LTC}\}$ the low-type plaintiffs recover an expected value of δx_L in all regimens.

End Proof.

Proof of Lemma 3:

We provide details for the derivation of α^*, the analysis is analogous for c^* and δ^*. We show that for all values of c and δ there exists $\alpha^* \in [0,1]$ such that for all $\alpha > \alpha^*$ the expected litigation expenses are larger in a Regime HTC than in Regime NC. The rest of the proof follows as in Proposition 2.

We call $E^{NC}(x)$ the litigation expenses in a cycle of 4 periods when there are no-caps and the recoveries of the high type is x; E^{LTC} the litigation expenses in a cycle of four periods when the caps involve low expected trim and $E^{HTC}(\alpha)$ the litigation

expenses in a cycle of four periods when the caps involve high expected trim and the caps are stroke down with probability α.

Then, in an infinite horizon litigation expenses in HTC are higher than in NC if and only if:

$$E^{HTC}(\alpha)+\alpha\sum_{i=1}^{\infty}(\delta^4)^i LE^{NC}+(1-\alpha)\sum_{i=1}^{\infty}(\delta^4)^i LE^{NC}(x^c)>\sum_{i=0}^{\infty}(\delta^4)^i E^{NC}(x_H)$$

which is equivalent to

$$E^{HTC}(\alpha)-E^{NC}(x_H)>(1-\alpha)\left[E^{NC}(x_H)-E^{NC}(x^c)\right]\sum_{i=1}^{\infty}(\delta^4)^i$$

$$E^{HTC}(\alpha)-E^{NC}(x_H)>(1-\alpha)\left[E^{NC}(x_H)-E^{NC}(x^c)\right]\frac{\delta^4}{1-\delta^4}$$

$$\alpha>\frac{\left[E^{NC}(x_H)-E^{NC}(x^c)\right]\frac{\delta^4}{1-\delta^4}-\left[E^{HTC}(\alpha)-E^{NC}(x_H)\right]}{\left[E^{NC}(x_H)-E^{NC}(x^c)\right]\frac{\delta^4}{1-\delta^4}}$$

The last inequality is satisfied for all the values of $\alpha>\alpha^*$ where α^* is implicitly defined by

$$\alpha^*=\frac{\left[E^{NC}(x_H)-E^{NC}(x^c)\right]\frac{\delta^4}{1-\delta^4}-\left[E^{HTC}(\alpha^*)-E^{NC}(x_H)\right]}{\left[E^{NC}(x_H)-E^{NC}(x^c)\right]\frac{\delta^4}{1-\delta^4}}$$

Notice that $\alpha^*\in[0,1]$ because the right hand side expression is decreasing in α. It takes a value larger than 1 when $\alpha=0$ and a value smaller than 1 when $\alpha=1$ (because $E^{HTC}(0)-E^{NC}(x_H)<0$ as indeed we are in a case of LTC and $E^{HTC}(1)-E^{NC}(x_H)>0$).

End Proof.