Bargaining with Optimism: Identification and Estimation of a Model of Medical Malpractice Litigation

Antonio Merlo and Xun Tang [1]
Department of Economics
Rice University
September 19, 2018

Abstract

We study a model of bargaining with optimism where players have heterogeneous beliefs about the final resolution. Beliefs and bargaining surplus are identified from the settlement probability and the distribution of accepted transfers. Using data from medical malpractice lawsuits in Florida, we estimate doctor and patient beliefs and the distribution of potential compensation. We find that patients are more optimistic and doctors more pessimistic when the severity of injury is higher, and the joint optimism diminishes as severity increases. We quantify the increase in settlement probability and the reduction in accepted settlement offers under counterfactual caps on the total compensation.

Key words: Bargaining, optimism, litigation costs, nonparametric identification, medical malpractice litigation

JEL codes: C14, C78

[1] We thank seminar participants at Brown, CEMFI (Madrid), CREST (Paris), Emory, HKUST, LSE, Ohio State, Princeton, SUFE (Shanghai), Texas A&M, TSE (Toulouse), Tsinghua, UCL, U Wisconsin (Madison) and attendants of 2014 Econometric Society North America Summer Meeting, 2015 Cowles Foundation Conference, and the 11th World Congress of Econometric Society (Montreal 2015) for feedbacks. This research is funded by NSF Grant # SES-1448257. We thank Devin Reily, Michael Shashoua and Michelle Tyler for capable research assistance. Special thanks to Ms. Ying Xu, a practicing attorney from the Law Offices of Eric K. Chen, for sharing her knowledge about the institutional details related to medical malpractice litigation.

1 Introduction

Optimism is often invoked as a possible explanation for why parties involved in a negotiation sometimes fail to reach an agreement even though a compromise could be mutually beneficial. For example, consider a medical malpractice dispute where a patient (the plaintiff) suffered a damage allegedly caused by a doctor's (the defendant's) negligence or wrongdoing. If the plaintiff and the defendant are both overly optimistic about their chances of getting a favorable jury verdict, there may not be any settlement that can satisfy both parties' exaggerated expectations. The general argument dates back to Hicks (1932), and was later developed by Shavell (1982), among others, in the context of legal disputes. A recent theoretical literature originated by the work of Yildiz (2003, 2004), extends this insight and studies a general class of bargaining models with optimism (see Yildiz (2011) for a survey). These models have also been used in a variety of empirical applications that range from pretrial negotiations in medical malpractice lawsuits (Watanabe (2006)), to negotiations about market conditions (Thanassoulis (2010)), and cross-license agreements (Galasso (2012)).

Despite the recent surge of interest in the theory and application of bargaining with optimism, none of the existing contributions formally addresses the issue of identification in this class of models. That is, under what conditions can the structural elements of the model (beliefs and bargaining surplus) be unambiguously recovered from the history of bargaining outcomes reported in the data? In a typical context of negotiations, the beliefs of both parties interact with the bargaining surplus to determine the outcome. As a result, the distribution of bargaining outcomes reported in a typical data environment is conditional on complex events involving these model elements. Such a selection issue causes major challenges in the identification of this class of models. One of the contributions of this paper is to deal with these challenges in identification using a minimum set of nonparametric assumptions.

We consider a bilateral bargaining environment where players hold optimistic beliefs about whether a stochastic outcome favors them if they fail to reach an agreement. The players have a one-time opportunity for reaching an agreement at an exogenously scheduled date during the bargaining process, and make decisions about whether or not to settle and, if so, the amount of the settlement based on their beliefs, the bargaining surplus and the time discount factor. We show that all structural elements in the model are identified nonparametrically from the probability of reaching an agreement and the distribution of transfers in the final resolution of the dispute. The identification strategy is robust in the sense that it does not rely on any parametrization of the distribution of beliefs or bargaining surplus.

We apply this model to analyze medical malpractice disputes in the State of Florida between 1984 and 1999. In his seminal work, Sloan (1993) provided a thorough empirical analysis of medical malpractice litigation using this source of data. Sieg (2000) was the first structural paper that estimated a bargaining model to understand the incentives of patients and doctors during medical malpractice disputes. Sieg (2000) was also the first paper that structurally estimated a bargaining model with asymmetric information. In comparison, we use a qualitatively different model to study how optimism in beliefs impact the final outcome. In addition to establishing the nonparametric identification of this model, we are interested in the following questions: How do the characteristics of lawsuits affect the litigation outcome through their impact on the parties' beliefs (and optimism)? To what extent are these beliefs consistent with the actual pattern of jury decisions observed in court? How does the total potential compensation for the alleged malpractice depend on these characteristics? What are the consequences of a tort reform that restricts the maximum compensation possible? To answer these questions, we propose a Maximum Simulated Likelihood (MSL) estimator based on flexible parametrization of the beliefs of both sides in the litigation. We then use the structural estimates for the belief and compensation distribution to quantify their relation to the case characteristics, and to evaluate the impact of the proposed tort reform.

The bargaining environment we consider is simpler than the one studied by Yildiz (2004). Rather than allowing for multiple rounds of offers and counteroffers, in our model there is a single settlement opportunity for the players to reach an agreement. Hence, in our case there are no dynamic "learning" considerations in the players' decisions, and the dates of the final resolution of the bargaining episodes are solely determined by the players' optimism, their patience, and their perception of the surplus available for sharing. Following Sieg (2000), we model the medical malpractice lawsuits as bargaining episodes with one-time opportunities for the players to settle out of court.[2]

Our specification of the bargaining environment is motivated by both theoretical and empirical concerns. First, data limitation would prevent us from deriving robust (parametrization-free) arguments for the identification of structural elements in general models of bargaining with optimism that admit multiple rounds of offers and counteroffers. For instance, none of the data sets which are used in empirical applications of bargaining models with optimism contains information on the sequence of proposers in a negotiation or the timing and size of rejected offers. By abstracting from the dynamic learning aspects (which would be introduced into the theoretical analysis if we were to consider a more general bargaining environment with multiple rounds of unobserved offers and counteroffers), we take a pragmatic approach and specify a model that is identifiable under realistic data requirements and mild econometric assumptions. Second, despite this simplification, our model captures the key insight of bargaining with optimism in that the incidence of agreement is determined by the players' optimism and their patience. Thus, our work represents a first important

[2] Da Silveira (2017) also used a one-shot bargaining model with asymmetric information to study the outcome of plea bargaining between a prosecutor and a defendant in criminal cases. His analysis provides conditions for the non-parametric identification of that model.

step toward addressing the issue of nonparametric identification in more general models of bargaining with optimism.[3] Third, our modeling choice is motivated by the specific empirical context of medical malpractice disputes in Florida. The law of the State of Florida (Florida Statues, Title XLV, Chapter 766, Section 108), requires that a one-time, mandatory settlement conference between the plaintiff and the defendant be held "at least three weeks before the date set for trial". The settlement conference is scheduled by the county court, is held before the court, and is mediated by court-designated legal professionals.

Our identification method consists of two steps. First, we recover the settlement probability and the distribution of transfers conditional on the unreported wait-time between the scheduled settlement conference and the court trial. In order to do so, we tap into a recent literature that uses eigenvalue decomposition to identify finite mixture models or structural models with unobserved heterogeneity (see, for example, Hu (2008), Hu and Schennach (2008), Kasahara and Shimotsu (2009), An, Hu and Shum (2010) and Hu, McAdams and Shum (2013)). In particular, we exploit the institutional details in our environment to group lawsuits into clusters defined by the county and the month in which the lawsuit is filed. We argue that the lawsuits within each cluster can be plausibly assumed to share the same, albeit unobserved, wait-time. We then use the cases in the same cluster as instruments for each other and apply eigenvalue decomposition to the joint distribution of settlement decisions and accepted offers within the cluster. This identifies the probability for settlement and the distribution of accepted settlement offers conditional on unreported wait-time. A novel feature in our first step of identification is that we show how the major identifying assumptions for models with unobserved heterogeneity (e.g., rank conditions in Hu (2008) and invertibility conditions in Hu and Schennach (2008)) are implied by intuitive restrictions on structural primitives in our model.

Second, we identify all structural elements of the model by exploiting the interaction between the length of wait-time, the beliefs and the potential compensation in the outcome of settlement decisions and accepted offers. To do so, we use the conditional distribution of outcomes recovered from the first step, and take full advantage of two implications of the model: (1) With orthogonality between beliefs and potential compensation, the distribution of transfers to the plaintiff as ruled by the court is directly related to the marginal distribution of potential compensation and the settlement probability; (2) The distribution of accepted offers under settlement is based on an additive transformation of the beliefs and potential compensation distribution.

Our structural estimates show that on average the potential compensation decreases with the patient's age, but increases with the severity of injury due to the alleged malpractice and the median household income in the county where the lawsuit is filed. As for the beliefs

[3] Watanabe (2006) studied medical malpractice disputes in the context of a dynamic model of bargaining with optimism and learning. His analysis is fully parametric and does not address the issue of identification.

about jury verdict, we find that patients tend to be relatively more optimistic and doctors relatively more pessimistic for the cases with higher severity. This is consistent with the effect of severity on court decisions observed in the data. Our estimates also show that the patient's and the doctor's beliefs are negatively correlated, and that the optimism diminishes as severity increases. In addition, we find that doctor qualifications (i.e., the doctor's board certification status and educational background) affect the beliefs of both parties in ways that are inconsistent with their actual marginal effects on jury verdicts observed in the data.

We use our structural estimates to predict the probability for settlement and the distribution of accepted offers under hypothetical caps on potential compensation. For each level of severity, we impose a cap equal to the 75th empirical percentile of compensations paid by defendants following jury verdicts in the data. While these caps only increase the probability for settlement by small margins, they lead to sizeable reductions in the accepted settlement offers on average. For example, among the lawsuits against board certified doctors, the rates of reduction in the mean of accepted offers under the caps vary between 15% and 22%, depending on the severity level. For the other cases involving doctors with no board certification, this range is between 16% and 20%.

The rest of the paper is organized as follows. Section 2 introduces the model of bilateral bargaining with optimism. Section 3 establishes the identification of structural elements in the model. Section 4 presents the Maximum Simulated Likelihood (MSL) estimator. Section 5 describes the data and the institutional details regarding medical malpractice lawsuits in Florida. Section 6 presents our estimation results. Section 7 evaluates the impact of the proposed tort reform. Section 8 concludes. Proofs are contained in the appendix.

2 The Model

Consider a lawsuit following an alleged instance of medical malpractice between a plaintiff (patient) and a defendant (doctor). The potential compensation, or the "bargaining surplus," is denoted by $C \in \mathbb{R}_{++}$. It is a monetary measure of the damage suffered by the patient as a consequence of the alleged malpractice, and is common knowledge between both parties. It is the total potential compensation to be paid to the plaintiff (if the court finds the doctor guilty) before subtracting any litigation costs.

After the filing of the lawsuit, both parties are notified of a date for a court trial and a date for a settlement conference, which is mandatory by law. (See, for example, Title XLV, Chapter 766, Section 108 of the Florida Statutes.[4]) The settlement conference requires attendance by both parties and their attorneys as well as legal professionals appointed by the county court where the lawsuit is filed. It must be scheduled at least three weeks before

[4] The current Florida Statutes pertaining to medical malpractice and related matters are available online at http:// www. flsenate. gov/ Laws/ Statutes/ 2014/ Chapter766

the date for the trial. During the settlement conference, the defendant makes a settlement offer S to the plaintiff. If the plaintiff accepts the offer, then the case is settled outside the court with no trial, and the plaintiff receives S from the defendant. Otherwise, the case needs to be resolved by a trial in the court after a hearing and jury deliberation. Let $A = 1$ denote the event that a settlement is reached at the conference, and $A = 0$ that the case goes to trial.

The date of the trial is determined by the court schedule and the backlog of lawsuits filed at the county court. Cases are assigned randomly among judges within a county court, based on their availability. If the court rules in favor of the plaintiff, then the defendant pays C to the plaintiff. Otherwise, the charges against the defendant are dropped and the plaintiff receives no compensation. Let $D = 1$ denote the event that the jury rules in favor of the plaintiff, and $D = 0$ otherwise.

Litigation costs for defendants and plaintiffs play an important role in determining the outcome from settlement conferences. It is worth emphasizing that defendants and plaintiffs typically face different cost schemes in practice. Attorneys representing defendants are paid based on the total number of hours they spend on the case. We use K to denote the per period legal fees paid by the defendant to her lawyer throughout the litigation process. In contrast, lawyers hired by plaintiffs are paid through contingency contracts. Such contracts entitle plaintiffs' lawyers to a fixed proportion γ of the total potential compensation C if the court rules in favor of their clients. On the other hand, if the court rules in favor of the defendant, the lawyers receive no additional compensation.[5]

The plaintiff and the defendant have heterogeneous beliefs about a jury verdict in their favor if the case goes to trial. Let $\mu_p \in (0, 1)$ denote the plaintiff's subjective probability for the event "$D = 1$" and $\mu_d \in (0, 1)$ denote the defendant's subjective probability for "$D = 0$". Optimism arises from the assumption that the joint support of (μ_p, μ_d) is $\mathcal{M} \equiv \{(\mu, \mu') \in (0, 1)^2 : 1 < \mu + \mu' < 2\}$. Here we use the term "optimism" to refer to the situation where parties' beliefs add up to be greater than one. The realized value of (μ_p, μ_d) is common knowledge between both parties at the time of the settlement conference. Let T denote the wait-time, or the length of the interval between the settlement conference and the trial; let δ denote the time discount factor that is exogenously fixed and known. At the settlement conference, the plaintiff's expected compensation from a court trial is $\delta^T \mu_p C$; and the defendant's expected loss from a court trial is $\delta^T (1 - \mu_d) C + \sum_{r=1}^{T} \delta^r K$.

We now characterize the Nash equilibrium at the settlement conference. The plaintiff accepts an offer S if and only if it exceeds his expected compensation from a trial. That is, $A = 1$ if and only if $S \geq \delta^T \mu_p C$. The defendant offers the plaintiff $S = \delta^T \mu_p C$ if and only if

[5] In practice, plaintiffs may pay a fixed amount of initiation fees to their lawyers upfront. However, such initiation fees are sunk costs for plaintiffs at the time of settlement conferences. Hence, we ignore such initiation fees without loss of generality.

this is lower than her expected loss if the case goes to trial. That is,

$$\delta^T \mu_p C \leq \delta^T (1-\mu_d) C + \textstyle\sum_{r=1}^{T} \delta^r K,$$

where the second term on the right-hand side is the incremental litigation costs for the defendant if no agreement is reached at the settlement conference. Therefore

$$A = 1 \text{ if and only if } YC \leq \phi(T)K, \tag{1}$$

where $\phi(t) \equiv \sum_{r=1}^{t} \delta^{r-t} = 1 + \frac{1}{\delta} + \frac{1}{\delta^2} + ... + \frac{1}{\delta^{t-1}}$ is increasing in t, and $Y \equiv \mu_p + \mu_d - 1$ denotes the joint optimism of the plaintiff and the defendant. Note that the plaintiff lawyer's compensation ratio γ does not affect the bargaining outcome.[6]

It is straightforward to incorporate into the model heterogeneity across lawsuits, such as the severity of injury due to the alleged malpractice, demographics of the plaintiff or professional qualifications of the defendant. The characterization of equilibrium remains valid after conditioning the distribution of (μ_p, μ_d), C, D, K and T on such heterogeneity.

3 Identification

This section shows the identification of the distribution of potential compensation and the beliefs of patients and doctors, assuming the bargaining outcomes reported in the data are rationalized by Nash equilibria. The main challenge for identification arises from a classical selection issue. That is, the distribution of bargaining outcome reported in the data is conditional on complex events involving beliefs and bargaining surplus.

We consider an environment where for each lawsuit the data reports whether a settlement occurs during the mandatory conference (A). For each case settled at the conference, the data reports the amount paid by the defendant to the plaintiff (S). For each of the other cases that were resolved through a court trial, the data reports the jury decision (D) and, if the court rules in favor of the plaintiff, the compensation paid by the defendant (C). The data also provide information about defendants' litigation costs per period (K). However, the scheduled dates of the settlement conference and the court trial are never reported in the data.[7] As a result the wait-time between the settlement conferences and the scheduled

[6] Also, any legal fee paid by the defendant prior to the settlement conference is a sunk cost and would therefore not affect the bargining outcome.

[7] The data we use in Section 5 reports the "date of final disposition" for each case. However, for a case settled outside the court, this date is defined not as the exact date of the settlement conference, but as the day when official administrative paperwork is finished and the claim is declared closed by the insurer. There is a substantial length of time between the two. For instance, for a large proportion of cases that are categorized as "settled within 90 days of the filing of lawsuits", the reported "dates of final disposition" are actually more than 150 days after the initial filing. Similar issues also exist for cases that were taken to the court in that the reported "dates of final disposition" do not correspond to the actual dates of court hearings.

court trial, which is known to both parties at the time of the conference, is not reported in the data. The following conditions are instrumental for our identification method.

Assumption 1 *(i) The vector (μ_p, μ_d, C, D, K) is independent of the wait-time T. (ii) K, D and (μ_p, μ_d, C) are mutually independent.*

Assumption 1 allows the beliefs of plaintiffs and defendants to be correlated and asymmetric with different marginal distributions. This generality allows for informational asymmetry between the two parties (e.g., doctors may be better informed about the cause and severity of the damage) and unobserved individual heterogeneities. It also accommodates the correlation between the plaintiff and defendant beliefs through unobserved heterogeneity on the case level. For example, both parties may be aware of certain aspects that are related to the cause and severity of the damage but not recorded in data. Assumption 1 also allows for correlation between beliefs and potential compensation.

Independence of wait-time from the beliefs, the potential compensation, the jury decision and the litigation costs per period is plausible, especially in the current context where the wait-time is determined by the availability of judges and juries in the county court where the lawsuit is filed. Such availability in turn depends on the court schedule and the backlog of cases for the judges, which are idiosyncratic and orthogonal to the beliefs of patients and doctors.

The orthogonality of D from C is also tenable. The potential compensation C is a monetary measure of the damage inflicted upon the plaintiff, regardless of its cause. In contrast, D depends on the jury's perception about whether the damage is due to the doctor's negligence, based on evidence presented at the trial. Once we condition on relevant characteristics (such as severity of damage and the plaintiff's costs of living), the jury decision D is likely to be orthogonal to the measure of damage C.

The assumption that the jury decision D is independent from beliefs μ_p, μ_d (conditional on the case characteristics in data) is somewhat restrictive. Suppose certain information/signal related to the doctor's negligence is known to the jury as well as both sides of litigation, but not reported in the data. Such information adds a new dimension of unobserved heterogeneity of the lawsuit that could lead to correlation between beliefs and jury decisions. To account for such correlation we would need a model that includes the unobserved signal distribution and additional parameters summarizing its impact on beliefs and jury decisions. The dimension of parameters in such a generalized model would be higher, and the method of robust identification in the current paper could not be applied.

Under Assumption 1, the distribution of accepted offers conditional on the wait-time $T = t$ and litigation costs per period $K = k$ is:

$$\Pr\left(S \leq s \mid A = 1, T = t, K = k\right) = \Pr\left(\mu_p C \leq s\delta^{-t} \mid YC \leq \phi(t)k\right), \tag{2}$$

and the distribution of potential compensation conditional on no settlement, a jury verdict in favor of the plaintiff and $T = t$, $K = k$ is:

$$\Pr(C \leq c \mid A = 0, D = 1, T = t, K = k) = \Pr(C \leq c \mid YC > \phi(t)k). \tag{3}$$

In practice, data often reports heterogeneity among the lawsuits. For example, the data we use in this paper reports defendants' professional qualifications (board certification and educational background), plaintiffs' demographics (age and gender) as well as severity of the injury due to alleged malpractice. Such a vector of observed case heterogeneity, denoted X, is correlated with the potential compensation C and the beliefs (μ_p, μ_d) in general.

To identify the model, we maintain that Assumption 1 holds conditional on X. The structural link between the data and the model elements are characterized as above, except that each distribution involved need to be conditional on X. In order to simplify notation, we suppress the dependence of structural elements on X in our presentation below. We only make such dependence explicit in the estimation section when it is needed to avoid ambiguity.

3.1 Conditional distribution of outcome

The first step in identifying the model is to recover the distribution of the litigation outcome conditional on the unreported wait-time T. To do so, we tap into the institutional background to construct clusters of independent litigation cases which share the same wait-time T. We then recover the conditional outcome distribution by applying an eigenvalue decomposition to the joint distribution of outcomes in the cases from the same cluster. Hu, McAdams and Shum (2013) had adopted a similar approach to identify first-price private-value auctions with discrete unobserved heterogeneity (UH) known to the bidders.

We exploit an implicit panel structure in the data. In particular, we note that lawsuits filed with the same county court during the same period practically share the same wait-time T. The reason for such a pattern is as follows: First, the dates for settlement conferences are determined by availability of court officials, and are assigned on a "first-come, first-served" basis. Thus, the settlement conferences for the cases filed with the same court at the same time are practically scheduled in the same period. Besides, the dates for court trials are determined by the schedule and the backlog of cases for judges at the court. Hence, the cases filed with the same county court simultaneously can be expected to be scheduled for a court trial at the same time in the future. This allows us to group the lawsuits into clusters with the same wait-time, despite unobservability of T in data. We formalize such an implicit panel structure as follows.

Assumption 2 *The data is partitioned into known clusters, each of which consists of multiple litigation cases that share the same wait-time T. The variables (μ_p, μ_d, C, K) and D (if necessary) are drawn independently across the cases within a cluster.*

This panel structure allows us to use accepted settlement offers within the same cluster as instruments for each other and apply an argument based on eigenvalue decomposition proposed by Hu (2008) and Hu and Schennach (2008) to recover the distribution of bargaining outcomes conditional on T. In the next subsection we will use these quantities to back out the joint distribution of beliefs using variations in T. Let the wait-time T be discrete with a finite support (i.e. $|\mathcal{T}| < \infty$). Denote the support of wait-time by $\mathcal{T} \equiv \{1, 2, ..., |\mathcal{T}|\}$.

For any discrete random vector R_2 and continuous random vector R_1, let $f_{R_1}(r_1, R_2 = r_2|.)$ be shorthand for $\frac{\partial}{\partial \tilde{r}} \Pr\{R_1 \leq \tilde{r},\, R_2 = r_2 \mid .\}|_{\tilde{r}=r_1}$. For any three lawsuits i, j, l that share the same wait-time T, let $\mathbf{k} \equiv (k_i, k_j, k_l)$ denote the vector of defendants' litigation costs in each post-conference period, and let $\mathcal{E}_{i,l}$ denote the event that "the lawsuit l was resolved through a settlement and lawsuit i through a court trial which ruled against the defendant ($A_i = 0, D_i = 1$ and $A_l = 1$)".

The following lemma links the joint distribution of the bargaining outcome reported in the data to their distribution conditional on the wait-time. It follows from the independence between lawsuits in the same cluster and an application of the law of total probability.

Lemma 1 *Suppose Assumption 1 and 2 hold. Then*

$$\begin{aligned} & f_{C_i,S_l}(c, s, A_j = 1|\mathcal{E}_{i,l}, \mathbf{k}) \\ = \ & \textstyle\sum_{t \in \mathcal{T}} f_{C_i}(c|A_i = 0, D_i = 1, T = t, k_i) \mathbb{E}(A_j|T = t, k_j) f_{S_l}\left(s, T = t|\mathcal{E}_{i,l}, k_i, k_l\right) \end{aligned} \tag{4}$$

and

$$f_{C_i,S_l}(c, s|\mathcal{E}_{i,l}, \mathbf{k}) = \textstyle\sum_{t \in \mathcal{T}} f_{C_i}(c|A_i = 0, D_i = 1, T = t, k_i) f_{S_l}\left(s, T = t|\mathcal{E}_{i,l}, k_i, k_l\right). \tag{5}$$

To better illustrate the method, we need to introduce matrix notation. For a generic integer M, let $\mathcal{D}_M$ denote a partition of the *unconditional* support of potential compensation $\mathcal{C}$ into M intervals, each denoted by d_m. Likewise, let $\mathcal{B}_M$ denote a partition of the *unconditional* support of the accepted settlement offers $\mathcal{S}$ into M intervals, each denoted by b_m. Fix a vector of per period costs $\mathbf{k}$. For a given pair of partition $\mathcal{D}_M$ and $\mathcal{B}_M$, let L_{C_i,S_l} be a M-by-M matrix whose (m, m')-th entry is the probability that "$C_i \in d_m$ and $S_l \in b_{m'}$" conditional on $\mathbf{k}$ and $\mathcal{E}_{i,l}$; and let Λ_{C_i,S_l} be a M-by-M matrix with its (m, m')-th entry being the probability that "$C_i \in d_m$, $A_j = 1$ and $S_l \in d_{m'}$" conditional on $\mathbf{k}$ and $\mathcal{E}_{i,l}$. Note that the definition of Λ_{C_i,S_l} and L_{C_i,S_l} is specific to the vector $\mathbf{k}$ as well as the partitions involved. For simplicity, we suppress their dependence on $\mathbf{k}$ in notation.

By definition, both Λ_{C_i,S_l} and L_{C_i,S_l} are identifiable directly from the data. Under Assumption 1 and 2, a discretized version of (4) is:

$$\Lambda_{C_i,S_l} = L_{C_i|T} \Delta_j L_{T,S_l} \tag{6}$$

where $L_{C_i|T}$ is a M-by-$|\mathcal{T}|$ matrix with (m, t)-th entry being $\Pr(C_i \in d_m|A_i = 0, D_i = 1, T = t, k_i)$; Δ_j be a $|\mathcal{T}|$-by-$|\mathcal{T}|$ diagonal matrix with the t-th diagonal entries being $\mathbb{E}(A_j|T = t, k_j)$;

and L_{T,S_l} be a $|\mathcal{T}|$-by-M matrices with its (t,m)-th entry being $\Pr(T=t, S_l \in b_m|\mathcal{E}_{i,l}, k_i, k_l)$. Likewise a discretized version of (5) is

$$L_{C_i,S_l} = L_{C_i|T} L_{T,S_l} \tag{7}$$

under Assumption 1 and 2. Note that $L_{C_i|T}$ depends on k_i, Δ_j depends on k_j, and L_{T,S_l} depends on (k_i, k_l) respectively. Again we suppress such dependence to simplify notation.

Assumption 3 *The potential compensation C is continuously distributed over its support $\mathcal{C} \equiv (0, \bar{c}) \subset \mathbb{R}_{++}$, which does not vary with beliefs (μ_p, μ_d). The defendant's litigation cost per period K is continuously distributed over $\mathcal{K} \equiv (0, \bar{k}) \subset \mathbb{R}_{++}$.*

This assumption implies that the support of potential compensation does not vary with the beliefs of both parties. For each $k \in \mathcal{K}$, define $\tau(k) \equiv \max\{t \in \mathcal{T} : \phi(t)k < \bar{c}\}$. By definition, $\tau(k) = |\mathcal{T}|$ if $\phi(|\mathcal{T}|)k < \bar{c}$, and $\tau(k) \leq |\mathcal{T}| - 1$ otherwise. The next lemma states that there exists sufficient dependence between doctor-patient transfers across the lawsuits within the same cluster. Such dependence between transfers arises from their respective correlation with the wait-time. This sufficient dependency is key to identification, because it implies the following rank condition which is necessary for applying the eigenvalue decomposition approach to identify finite mixture models.

Lemma 2 *Suppose Assumption 1, 2 and 3 hold. Then for any $k_i, k_l \in \mathcal{K}$ there exists a partition $\mathcal{D}_{\tau(k_i)}$ on $\mathcal{C}$ and a partition $\mathcal{B}_{\tau(k_i)}$ on $\mathcal{S}$ such that L_{C_i,S_l} defined over $\mathcal{D}_{\tau(k_i)}$ and $\mathcal{B}_{\tau(k_i)}$ has full rank $\tau(k_i)$.*

If $\phi(|\mathcal{T}|)k_i < \bar{c}$ (or equivalently $\tau(k_i) = |\mathcal{T}|$), Lemma 2 simply states that for any $k_l \in \mathcal{K}$ there exists a partition $\mathcal{D}_{|\mathcal{T}|}$ on $\mathcal{C}$ and a partition $\mathcal{B}_{|\mathcal{T}|}$ on $\mathcal{S}$ such that L_{C_i,S_l} has full rank $|\mathcal{T}|$. The intuition for this result as follows. Under the support condition in Assumption 3, there is sufficient variation in the conditional distribution of C and that of S as the wait-time changes. Under the panel structure and orthogonality conditions in Assumption 1, these two sources of variation interact with each other and induce substantial dependence between observed transfers C and S even after the unobserved wait-time is integrated out.

The next proposition states that the conditional distribution of outcomes given the wait-time is identified by exploiting the joint distribution of transfers C_i and S_l.

Proposition 1 *Suppose Assumptions 1, 2 and 3 hold. Then $\Pr(A_j = 1|T = t, K_j = k)$ and $f_{S_l}(.|A_l = 1, T = t, K_l = k)$ are identified for all $k \in \mathcal{K}$ and $t \in \mathcal{T}$; and $f_{C_i}(.|(1-A_i)D_i = 1, T = t, K_i = k)$ is identified for all $k \in \mathcal{K}$ and $t \in \mathcal{T}$ such that $\phi(t)k < \bar{c}$.*

This proposition states that all conditional distributions of bargaining outcome are identified. Note that the density $f_{C_i}(.|A_i = 0, D_i = 1, T = t, K_i = k)$ is not well-defined for any

(t, k) with $\phi(t)k \geq \bar{c}$, because $\Pr(A_i = 0|T = t, K_i = k) = 0$ for such (t, k). We now sketch the main idea behind this proposition and explain why the correlation between the transfers C_i and S_l established in Lemma 2 is instrumental for this identification result.

Consider $k_i, k_j \in \mathcal{K}$ such that $\tau(k_i) = \tau(k_j) = |\mathcal{T}|$. By Lemma 2 there exists partitions $\mathcal{D}_{|\mathcal{T}|}$ and $\mathcal{B}_{|\mathcal{T}|}$ such that L_{C_i,S_l} and $L_{C_i|T}$ are non-singular. It then follows from (6) and (7) that

$$\Lambda_{C_i,S_l}(L_{C_i,S_l})^{-1} = L_{C_i|T}\Delta_j \left(L_{C_i|T}\right)^{-1}, \tag{8}$$

where the left-hand side consists of directly identifiable quantities only. The right-hand side of (8) takes the form of an eigenvalue-decomposition, which is unique up to a scale normalization and unknown matching between the columns in $L_{C_i|T}$ (and diagonal entries in Δ_j) and the wait-time T.

To pin down the scale of $L_{C_i|T}$ and match its columns with specific values of $t \in \mathcal{T}$, we exploit implications of the model structure. First, the scale in the eigenvalue-decomposition is identified because each column in $L_{C_i|T}$ corresponds to a conditional probability mass function and therefore adds up to one. Second, the question of unknown indexing is also resolved, because under Assumption 1 and 3, we know that for the k_j considered, $\Pr(A_j = 1|T = t, k_j) = \Pr(Y_j C_j \leq \phi(t)k_j)$ in equilibrium and is monotonically increasing in t over the support of wait-time $\mathcal{T}$. This rules out duplicate eigenvalues in the decomposition, and helps to match eigenvalues and eigenvectors with specific elements in $\mathcal{T}$. Once Δ_j and $L_{C_i|T}$ are recovered through the decomposition, we can recover L_{T,S_l} as $\left(L_{C_i|T}\right)^{-1} L_{C_i,S_l}$.

A symmetric argument identifies a square matrix $L_{S_l|T}$ whose (m, t)-th entry is defined as $\Pr(S_l \in b_m|A_l = 1, T = t, k_l)$ with $b_m \in \mathcal{B}_{|\mathcal{T}|}$ and $k_l \in \mathcal{K}$. With the discretized distribution $L_{C_i|T}$, L_{T,S_l} and $L_{S_l|T}$ identified, we recover the conditional density functions $f_{S_l}(s_l|A_l = 1, T = t, K_l = k_l)$ and $f_{C_i}(c_i|(1 - A_i)D_i = 1, T = t, K_i = k_i)$ for each s_l and c_i on their respective domains. Identification for the other cases where either $\tau(k_i) < |\mathcal{T}|$ or $\tau(k_j) < |\mathcal{T}|$ follows from similar arguments. See the proof of Proposition 1 in Appendix B for details.

3.2 Joint distribution of beliefs

The intermediate step in Section 3.1 recovers the marginal (but not the joint) distribution of C or S conditional on the jury verdict. In this section, we identify the joint distribution of beliefs and potential compensation (μ_p, μ_d, C) from the conditional distribution of outcomes recovered in Proposition 1.

Assumption 4 *The joint distribution of (μ_p, μ_d) is independent from C.*

This condition requires the magnitude of potential compensation to be independent from plaintiffs' and defendants' beliefs. This condition is plausible because C is meant to capture an objective monetary measure of the severity of damage inflicted upon the patient. On the other hand, the beliefs (μ_p, μ_d) should depend on the evidence available as to whether the defendant's neglect is the main cause of such damage.

For all t, k such that $\phi(t)k < \bar{c}$, define

$$\psi(k,t,c) \equiv \Pr(C_i \leq c | A_i = 0, D_i = 1, T = t, K_i = k) \Pr(A_i = 0 | T = t, K_i = k).$$

Recall from Section 3.1 that under Assumption 1, 2 and 3, $\Pr(A_i = 0|T = t, K_i = k)$ is identified, and so is $f_C(.|A_i = 0, D_i = 1, T = t, K_i = k)$ over the conditional support of C for all t, k such that $\phi(t)k < \bar{c}$. Thus $\psi(k,t,c)$ is directly identifiable for all $c \in \mathcal{C}$ and $t \in \mathcal{T}$, $k \in \mathcal{K}$ with $\phi(t)k < c$. By construction,

$$\begin{aligned}
\psi(k,t,c) &= \frac{\Pr(C_i \leq c, Y_i C_i > \phi(t)k | D_i = 1, t, k)}{\Pr(Y_i C_i > \phi(t)k | D_i = 1, t, k)} \Pr(A_i = 0 | t, k) \\
&= \Pr(Y_i C_i > \phi(t)k, C_i \leq c),
\end{aligned}$$

where the second equality holds under Assumption 1. Then

$$\begin{aligned}
\tfrac{\partial}{\partial c}\psi(k,t,c) &= \tfrac{\partial}{\partial \hat{c}} \left[\int^{\hat{c}} \Pr(Y_i C_i > \phi(t)k | C_i = \tilde{c}) f_C(\tilde{c}) d\tilde{c} \right]_{\hat{c}=c} \\
&= \tfrac{\partial}{\partial \hat{c}} \left[\int^{\hat{c}} \Pr(Y_i > \phi(t)k/\tilde{c}) f_C(\tilde{c}) d\tilde{c} \right]_{\hat{c}=c} \\
&= \Pr(Y_i > \phi(t)k/c) f_C(c),
\end{aligned}$$

where the first equality follows from the law of total probability and the second from Assumption 4.

Then we identify the marginal density of potential compensation f_C as follows. Fix any $c_0 \in \mathcal{C}$. For any $c \in \mathcal{C}$, we can find a triple (t, k_0, k) with $k_0/k = c_0/c$ and $\phi(t)k < c$ thanks to Assumption 3. Thus by construction, for any $t \in \mathcal{T}$, we have

$$\frac{\partial\psi(k,t,c)/\partial c}{\partial\psi(k_0,t,c_0)/\partial c} = \frac{f_C(c)}{f_C(c_0)} \equiv \varphi_0(c). \tag{9}$$

Because $\int_{\mathcal{C}} f_C(c) dc = 1$ by construction, (9) implies that the marginal density of potential compensation is identified as

$$f_C(c) = \frac{\varphi_0(c)}{\int_{\mathcal{C}} \varphi_0(\tilde{c}) d\tilde{c}}$$

for all $c \in \mathcal{C}$.[8] In fact the marginal density f_C is over-identified, because the method above can be applied using any quadruple (c_0, t, k_0, k) that satisfies $k_0/c_0 = k/c$ and $\phi(t)k < c$.

[8] An alternative argument for identification is as follows. With Y_i continuously distributed over $(0,1)$, we have $\lim_{k' \to 0} \partial\psi(k',t,c)/\partial c = \Pr(Y_i > 0) f_C(c)$ for all $t \in \mathcal{T}$ and $c \in \mathcal{C}$. With $\Pr(Y_i > 0) = 1$, the density $f_C(.)$ is over-identified.

Next, we show how to recover the marginal distribution of Y. Under Assumption 3, for any $\alpha \in (0,1)$ there exist $t_\alpha \in \mathcal{T}$ and $k_\alpha \in \mathcal{K}$, $c_\alpha \in \mathcal{C}$ in the interior of support such that $\phi(t_\alpha)k_\alpha/c_\alpha = \alpha$. Then

$$\frac{\partial\psi(k_\alpha, t_\alpha, c_\alpha)}{\partial c} = \Pr(Y_i > \phi(t_\alpha)k_\alpha/c_\alpha)f_C(c_\alpha) \Rightarrow \Pr(Y_i > \alpha) = \frac{\partial\psi(k_\alpha, t_\alpha, c_\alpha)/\partial c}{f_C(c_\alpha)}.$$

Because ψ is identified for such t_α, k_α and all c in an open neighborhood around c_α, so is the partial derivative $\partial\psi(k_\alpha, t_\alpha, c_\alpha)/\partial c$. With the density f_C identified above, this means the marginal distribution of Y is identified over its full support $(0,1)$.

To identify the distribution of (μ_p, μ_d), we exploit the variation in

$$\Psi(k,t,s) \equiv \Pr(S \leq s, A = 1|K = k, T = t) = \Pr(\mu_p C \leq s/\delta^t, YC \leq \phi(t)k),$$

which is identified for all t, k and $s \in \mathcal{S}$ under the following support condition.

Assumption 5 $\phi(|\mathcal{T}|)\bar{k} \geq \bar{c}$.

Under this condition there is positive probability for the extreme case where the total litigation costs for the defendant exceeds the potential compensation. In such a case, the probability for settlement is one, because $\Pr(YC \leq \phi(T)K \mid T = |\mathcal{T}|, K = \bar{k}) = \Pr(YC \leq \phi(|\mathcal{T}|)\bar{k}) = 1$. Without this assumption, we can only partially identify this model.

Applying a logarithm transform, we can write

$$\Psi(k,t,s) = \Pr(V_1 + W \leq \log s - t\log\delta, V_2 + W \leq \log\phi(t) + \log k),$$

where $V_1 \equiv \log\mu_p$, $V_2 \equiv \log Y$ and $W \equiv \log C$. Thus the joint distribution of (V_1+W, V_2+W) is identified over its full support under Assumption 5 and Proposition 1.

Define $\mathbf{V} \equiv (V_1, V_2)$, $\mathbf{W} \equiv (W, W)$ and let $\varphi_{\mathbf{V}+\mathbf{W}}, \varphi_{\mathbf{V}}, \varphi_{\mathbf{W}}$ denote the characteristic function of $\mathbf{V}+\mathbf{W}$, $\mathbf{V}$ and $\mathbf{W}$ respectively. Then by the independence of (μ_p, μ_d) and C,

$$\varphi_{\mathbf{V}+\mathbf{W}}(\mathbf{u}) = \varphi_{\mathbf{V}}(\mathbf{u})\varphi_{\mathbf{W}}(\mathbf{u}) \text{ for all } \mathbf{u} \equiv (u_1, u_2) \in \mathbb{R}^2.$$

Let φ_W denote the characteristic function of the scalar variable W. By the argument above, the density of C is identified. By construction, $\varphi_{\mathbf{W}}(\mathbf{u}) = \varphi_W(u_1 + u_2)$ for all $\mathbf{u} \in \mathbb{R}^2$. Hence we treat the characteristic function of $\mathbf{W}$ as known in subsequent identification exercises. Furthermore, the characteristic function $\varphi_{\mathbf{V}+\mathbf{W}}$ is also known because the joint distribution of $(V_1 + W, V_2 + W)$ is identified above.

Assumption 6 *$\varphi_W(.)$, is non-vanishing over $\mathbb{R}$.*

This condition holds if $\varphi_{\mathbf{V}+\mathbf{W}}(\cdot)$ is non-vanishing, which is directly testable using the data available. Under this assumption, the characteristic function $\varphi_{\mathbf{W}}$ is also non-vanishing over

$\mathbb{R}$. This implies that $\varphi_{\mathbf{V}}$ is identified as the ratio between known complex numbers $\varphi_{\mathbf{V}+\mathbf{W}}(\mathbf{u})$ and $\varphi_{\mathbf{W}}(\mathbf{u})$ for all $\mathbf{u} \in \mathbb{R}^2$. Therefore the joint distribution of $\mathbf{V} \equiv (\log \mu_p, \log(\mu_p + \mu_d - 1))$ is known, which in turn implies the joint distribution of (μ_p, μ_d) is identified via Jacobian transformation.

We summarize the identification results discussed throughout this subsection in the following proposition.

Proposition 2 *Under Assumption 1, 2, 3, 4, 5 and 6, the joint distribution of (μ_p, μ_d) and the distribution of C are identified.*

4 Maximum Simulated Likelihood Estimation

A constructive nonparametric estimator based on the identification result in Section 3 would require a large data set that allows us to estimate the sample analogs of the distributions in Section 3. If the data report case-level characteristics that affect both parties' beliefs (such as severity of damage due to alleged malpractice and defendant qualification), then they should be conditioned on in estimation. This aggravates the "curse of dimensionality". To deal with heterogeneous cases in moderate-sized data, we propose a maximum simulated likelihood (MSL) estimator based on flexible parametrization of beliefs.

Suppose the data consists of N clusters. Each cluster is indexed by n and consists of $m_n \geq 1$ cases, each of which is indexed by $i = 1, ..., m_n$. For each case i in cluster n, let $A_{n,i} = 1$ when there is an agreement for settlement outside the court and $A_{n,i} = 0$ otherwise. Let $K_{n,i}$ denote the per-period litigation cost for the defendant in that lawsuit; denote the observed transfer in the data by $Z_{n,i}$ so that $Z_{n,i} \equiv S_{n,i}$ if $A_{n,i} = 1$; $Z_{n,i} \equiv C_{n,i}$ if $A_{n,i} = 0$ and $D_{n,i} = 1$; and $Z_{n,i} \equiv 0$ otherwise. Let T_n denote the wait-time between the settlement conference and the scheduled date for court decision, which is shared by all cases in cluster n. We propose an MSL estimator for the joint distribution of beliefs that also uses variation in the heterogeneity of lawsuits reported in the data. Throughout this section, we assume the identifying conditions in Assumption 1-6 hold after conditioning on such observed heterogeneity of the lawsuits.

Let $x_{n,i}$ denote a vector of case-level variables reported in the data that affect the distribution of C. (We allow $x_{n,i}$ to contain a constant component in the estimation below.) The potential compensation C in a lawsuit with observed features $x_{n,i}$ is drawn from an exponential distribution truncated at \$2.5 million with its latent rate given by

$$\lambda_{n,i}(\beta) \equiv \exp\{x_{n,i}\beta\},$$

where β is a vector of unknown parameters.

Let $w_{n,i}$ denote the vector of case-level variables in the data that affects the joint belief distribution. (In general, vectors $x_{n,i}$ and $w_{n,i}$ may have overlapping components.) We suppress the subscripts n, i for simplicity when there is no ambiguity. We estimate the belief distribution conditional on such a vector of case-level variables using maximum simulated likelihood. The estimate $\hat{\beta}$ above is used as an input to construct the simulated likelihood.

We adopt a flexible parametrization of the joint distribution of (μ_p, μ_d) conditional on W as follows. For each realized w, let $(\tilde{Y}, Y, 1-\tilde{Y}-Y)$ be drawn from a Dirichlet distribution with concentration parameters $\alpha_j(w) \equiv \exp\{w\rho_j\}$ for $j = 1, 2, 3$, where $\rho \equiv (\rho_1, \rho_2, \rho_3)$ are constant coefficient vectors. We suppress w in the notation $\alpha_j(w)$ for simplicity. Let $\mu_p = 1 - \tilde{Y}$ and $\mu_d = \tilde{Y} + Y$. The support of (μ_p, μ_d) is $\{(\mu, \mu') \in (0,1)^2 : 1 < \mu + \mu' < 2\}$, which is consistent with our model with optimism. (Table C1 in Appendix C shows how flexible such a specification of the joint distribution of (μ_p, μ_d) is in terms of the range of moments and the location of the model it allows.) Also, note by construction $Y = \mu_p + \mu_d - 1$ is a measure of optimism.

Let $q_{n,i} \equiv \Pr(D_{n,i} = 1 | A_{n,i} = 0, w_{n,i}, x_{n,i})$. Under the orthogonality conditions in Assumption 1 conditional on $w_{n,i}, x_{n,i}$, the probability $q_{n,i}$ does not depend on $c_{n,i}$, and is directly identifiable from the data. Let $h_n(.;\theta)$ denote density of the wait-time T_n in cluster n. This density in general may depend on cluster-level variables reported in the data, and is specified up to an unknown vector of parameters θ.

The log-likelihood of our model is:

$$L_N(\rho, \theta, \beta) \equiv \sum_{n=1}^{N} \ln \left[\sum_{t \in \mathcal{T}} h_n(t;\theta) \prod_{i=1}^{m_n} f_{n,i}(t;\rho,\beta) \right]$$

where $f_{n,i}(t;\rho,\beta)$ is shorthand for the conditional density of $Z_{n,i}, A_{n,i}, D_{n,i}$ given $T_n = t$, $W_{n,i} = w_{n,i}$, $K_{n,i} = k_{n,i}$ and with parameter ρ, evaluated at $(z_{n,i}, a_{n,i}, d_{n,i})$. Specifically,

$$\begin{aligned} f_{n,i}(t;\rho,\beta) \quad \equiv \quad & [g_{1,n,i}(t;\rho,\beta)]^{a_{n,i}} \times [g_{0,n,i}(t;\rho,\beta)]^{(1-a_{n,i})d_{n,i}} \times \\ & \{[1 - p_{n,i}(t;\rho,\beta)]\,(1 - q_{n,i})\}^{(1-a_{n,i})(1-d_{n,i})} \end{aligned}$$

where

$$p_{n,i}(t;\rho,\beta) = \int_0^{\bar{c}} \Pr(Y \le k_{n,i}\phi(t)/c | w_{n,i};\rho) f_C(c|x_{n,i};\beta) dc \tag{10}$$

and

$$g_{0,n,i}(t;\rho,\beta) = q_{n,i} \Pr(Y > k_{n,i}\phi(t)/z_{n,i} | w_{n,i};\rho) f_C(z_{n,i}|x_{n,i};\beta) \tag{11}$$

with $f_C(.|x_{n,i};\beta)$ being conditional density of C at β. Furthermore,

$$g_{1,n,i}(t;\rho,\beta) = \int_0^{1-z_{n,i}/(\delta^t \bar{c})} B_{n,i}(\rho) f_{\tilde{Y}}(\tau|w_{n,i};\rho) \frac{f_C\left(\delta^{-t} z_{n,i}/(1-\tau) \middle| x_{n,i};\beta\right)}{\delta^t(1-\tau)} d\tau \tag{12}$$

where $B_{n,i}(\rho;t)$ denotes the Beta distribution evaluated at $\delta^t \phi(t) k_{n,i}/z_{n,i}$ and parameters $(\exp(w_{n,i}\rho_2), \exp(w_{n,i}\rho_3))$ and $f_{\tilde{Y}}(.|w_{n,i};\rho)$ is a Beta p.d.f. evaluated at parameters $(\exp(w_{n,i}\rho_1)$, $\exp(w_{n,i}\rho_2) + \exp(w_{n,i}\rho_3))$. Details for deriving (10), (11) and (12) are presented in Appendix C2.

For each (n, i, t) and parameter (ρ, β), let $\hat{p}_n(t; \rho, \beta)$ and $\hat{g}_{1,n,i}(t; \rho, \beta)$ be estimators for $p_n(t; \rho, \beta)$ and $g_{1,n,i}(t; \rho, \beta)$ using $S^* > N$ simulated draws of τ, which are generated from a uniform distribution over the interval defined by integral limits. It follows from the Law of Large Numbers that $\hat{g}_{1,n,i}(t; \lambda, \rho)$ is an unbiased estimator for each n, i and (λ, ρ). Our maximum simulated likelihood estimator is

$$(\hat{\rho}, \hat{\theta}, \hat{\beta}) \equiv \arg \max_{(\rho,\theta,\beta)} \hat{L}_N(\rho, \theta, \beta). \tag{13}$$

where $\hat{L}_N(\rho, \theta, \beta)$ is an estimator for $L_N(\rho, \theta, \beta)$ by replacing $g_{1,n,i}(t; \rho, \beta)$ with $\hat{g}_{1,n,i}(t; \rho, \beta)$ and replacing $q_{n,i}$ with a parametric (logit) estimate $\hat{q}_{n,i}$ using the regressors in $w_{n,i}, x_{n,i}$.

Under appropriate regularity conditions, $\sqrt{N}[(\hat{\rho}, \hat{\theta}, \hat{\beta}) - (\rho, \theta, \beta)]$ converges in distribution to a zero-mean multivariate normal distribution with some finite covariance as long as $N \to \infty$, $S^* \to \infty$ and $\sqrt{N}/S^* \to 0$. The covariance matrix can be consistently estimated using the analog principle, which involves the use of simulated observations. (See Equation (12.21) in Cameron and Trivedi (2005) for a detailed formula.)

5 Data Description

Since 1975 the State of Florida has required all insurers that cover medical malpractice lawsuits to file reports on their resolved claims to the Florida Department of Financial Services. Using this source of data, we construct a sample that consists of 6,405 medical malpractice cases filed in Florida between 1984 and 1999.[9] Our sample includes cases that are either resolved through the mandatory settlement conference or a court trial. For each lawsuit, the data reports the date when it is officially filed with a court (*Suit_Date*), the county in which it is filed (*County*), and the date of the final disposition (*Date_of_Disp*). This latter date is defined as the date when the claim is closed with the insurer, and is typically later than the actual date for settlement conferences or court trials due to unrecorded administrative delays. The data does not report any dates of settlement conferences or scheduled court trials. Thus the wait-time between these dates can not be inferred from the data.

For each lawsuit, the data reports whether it is resolved through an agreement at the settlement conference or a court trial ($A = 1$ or $A = 0$). The data also reports the size of the transfer from the defendant to the plaintiff upon the resolution of the lawsuit. This transfer is equal to the settlement offer accepted by the plaintiff (S) if the case is settled out of court. Otherwise, this transfer is equal to the total compensation paid to the plaintiff if the court rules in his favor (C). In addition, we observe case-level variables that are relevant to beliefs

[9] Sieg (2000) and Watanabe (2009) also use the same source of data for their empirical analyses of medical malpractice lawsuits. In our analysis we exclude all cases that resulted in the death of the patient.

and/or the potential compensation. These variables include the severity of injury due to the alleged malpractice (*Severity*) which is discretized to be “low”, “medium” or “high”, the age (*Age*) and the gender (*Gender*) of the patient, whether the doctor sued is board certified (*Board*), and whether the doctor holds a degree from a medical school outside the United States (*Graduate*). Specifically, *Gender* $= 1$ if the patient is male and *Gender* $= 0$ otherwise; *Board* $= 1$ means the doctor is reported to be certified by at least one professional board and *Board* $= 0$ otherwise; *Graduate = 1* means the doctor holds a degree from a non-U.S. medical school, and *Graduate = 0* otherwise.

[Insert Figure 1]

We measure the length of the legal process (*Duration*) by the number of quarters between the date of the initial filing (*Suit_Date)* and the date of final disposition (*Date_of_Disp*). As explained above, this provides a noisy measure of the duration of each litigation process. Figure 1 reports the histograms of such a noisy measure of the duration conditional on the method of disposition (via settlement or jury decisions in court). By construction, the support of *Duration* in the cases with settlement is substantially larger than those resolved through court hearing and jury decisions. Moreover, Figure 1 illustrates that conditional on the means of disposition (via settlement or trial), there is substantial variation in this noisy measure of the duration of litigation.

[Insert Table 1(a)]

The data also reports the total litigation costs paid by each defendant to the attorneys throughout the complete legal process. Table 1(a) provides a descriptive regression analysis of the determinants and the structure of litigation costs for defendants. The dependent variable is the total legal costs for defendants throughout the litigation process. The explanatory variables are case characteristics, which include dummy variables for the severity of injury (*Low, High*), the defendant’s qualification (*Board*, *Graduate*) as well as the length of the litigation process (*Duration*). We also condition on certain measure of the income level in the geographic area where the lawsuit was filed (*Income*), which is defined in Section 5.2 below. On the one hand there is no statistical evidence for any impact of the severity of injury on the total litigation costs; on the other hand it is significantly more costly to defend a doctor who is educated outside the United States. The regression coefficients also confirm that the defendant’s litigation costs is highly positively related to the income level of the geographic area in which the lawsuit was filed.

More importantly, the duration of the litigation process is highly significant at 1% level in each specification considered. Besides, there is strong evidence that these litigation costs increase linearly with the length of litigation process, because the coefficients for the quadratic terms (*Duration*^{2}) are insignificant in both specifications. These stylized facts provide some justification for our maintained assumption that the litigation costs for defendants are linear in the duration of the legal process. We recover the defendant's litigation cost per period ($Costs$) by dividing the total litigation costs by the duration of the legal process in quarters. (Note that the legal process is by definition longer than the unreported wait-time between settlement conferences and court trials, and that parts of total litigation costs reported are sank costs for defendants at the settlement conferences.) The sample mean of $Costs$ is $\$3,569$, the sample standard deviation is $\$3,432$, and the 25th, 50th and 75th sample percentiles are $\$1,457$, $\$2,547$ and $\$4,411$ respectively.

Notwithstanding the reduced-form evidence above, we are aware that the assumption of linear litigation costs is a simplification and abstracts away from other cost factors not measured in the data available (e.g., the complexity of the case and the competence of the lawyer). As such the assumption can be viewed as a first-order approximation of more complex cost structures, and is justifiable due to data limitations. It is also worth mentioning that the identification and estimation method we present could also be applied to a more general framework, provided the relation between the litigation costs and the duration of the legal process is monotonic and known.

5.1 Summary of litigation outcomes

Table 1(b) and 1(c) summarize the relation between the litigation outcomes and the characteristics of lawsuits in the data. Table 1(b) reports the sample proportion of cases settled outside the court and the sample mean of accepted settlement offers after controlling for the doctor's board certification and the severity of the injury due to the alleged malpractice. Using estimates from Table 1(b), we conduct one-sided z-tests for the null hypothesis that the settlement probability is higher when the doctor is board certified. The p-value for such a null hypothesis is less than 0.001 when conditioning on low severity, and is 0.007 (and 0.354) when conditioning on medium (and high) severity. Table 1(c) reports the p-values in two-sided pair-wise z-tests for the equality of settlement probabilities across different groups.

[**Insert Table 1(b) and (c)**]

These results from Table 1(b) and 1(c) demonstrate mixed patterns about how the characteristics of the lawsuit affect settlement. First, when the doctor is board certified, the probability for settlement increases significantly with severity. In contrast, when the doctor is not board certified, the settlement probability does not vary significantly with severity. Second, the doctor's board certification has a significant impact on settlement probability only when the severity is low or medium. Third, for any given board certification status, the mean of accepted settlement offers is significantly higher for cases with higher severity. Last, board certification has a significant positive effect on expected settlement offers only for the cases with medium severity.

To interpret these mixed patterns, recall that a settlement takes place if and only if optimism-weighted compensation (YC) is small relative to the potential savings in litigation costs due to settlement. The severity of injury affects settlement through its impacts on the size of potential compensation *and* on the joint optimism Y. The signs of these impacts are opposite: While potential compensation increases with severity, the patient and doctor's joint optimism may well diminish with severity. (This is confirmed by our structural estimates reported in Figure 3 and Table 8 in Section 6 below.) The qualification of the doctor also has an impact on the plaintiff and defendant beliefs about the outcome of court trials. Furthermore, the mean of accepted settlement offers is the expectation of discounted expected transfers ($\delta^t \mu_p C$) conditional on settlement ($YC \leq \phi(t)k$). Thus, it is the interaction of *Severity* and *Board* that drives the likelihood for settlement and the mean of accepted offers. This explains the lack of monotone patterns in the settlement probability and the mean of accepted offers in Table 1(b) and 1(c).

[**Insert Table 2**]

In total there are 1,247 lawsuits that were not settled at scheduled conferences and had to be resolved through court trials. In 204 of these lawsuits the court ruled in favor of the plaintiff. Table 2 reports the sample mean of the compensation (C) paid to plaintiffs among these cases, conditioning on *Severity* and *Age*. Across all age groups, there are substantial differences (statistically significant at 5% level in one-sided tests) in the average compensation ruled by the court for the cases with low and medium severity. However, the difference in the average compensation between the cases with medium and high severity is only significant for the youngest group. Besides, Table 2 also shows that the average compensation does not vary significantly with the patient's age, for any level of severity.

To interpret these patterns, note these average compensations are all conditional on the absence of settlement ($YC > \phi(t)k$). The unconditional distribution of C depends on the interaction of the severity of injury and the age of the patient. Besides, as mentioned above, the severity level has a mixed effect on the settlement probability. Therefore, the lack of a monotonic pattern in Table 2 is due to the interaction of these multiple factors.

5.2 Descriptive analyses

We now report results from descriptive analyses about how the characteristics of lawsuits are related to the outcomes from settlement conferences and court trials. We control for the income level in the county where the lawsuit is filed. This is meant to capture the impact of county-level income on the compensation for patients. We collect data on the median household income in all counties in Florida in the years of 1989, '93, '95, '97, '98 and '99 from the Small Area Income and Poverty Estimates (*SAIPE*) produced by the U.S. Census

Bureau.[10] We also collect a time series of state-wide median household income in Florida each year between 1984 and 1999 from the U.S. Census Bureau's Current Population Survey. We combine this latter state-wide information with the county-level information from *SAIPE* to extrapolate the median household income in each Florida county in the years 1984-89, '92, '94 and '96.[11] We then incorporate this yearly data on household income in each county while estimating the distribution of total compensation next year. All income values are normalized in terms of U.S. dollars in 1990 using historical data on U.S. inflation/CPI.

[**Insert Tables 3 and 5**]

Table 3 reports three logit regressions of the settlement dummy (A) on the case characteristics, using all 6,405 observations. *Low* and *High* are dummy variables for low and high severity of injury. Across these nested specifications, *Board* and *Graduate* are statistically significant at 5% level with negative and positive effects respectively on the probability for settlement: A doctor's board certification would reduce the probability for settlement while a doctor's non-U.S. background of medical education would increase this probability. Both characteristics are unlikely to have any impact on the total compensation, but may well affect the settlement probability through their marginal effect on patients' and doctors' beliefs.

Other patient and case characteristics (*Age*, *Income*, *Costs* and dummies for *Severity*) enter the logit model in multiple regressors. To quantify the impact of these variables on settlement, we report their average marginal effects (A.M.E.) and the p-values of likelihood ratio tests for their significance in Table 5. In each specification, the likelihood-ratio test fails to reject the null hypothesis that the vector of coefficients for all *Age*-related regressors is jointly zero at the 10% level. On the other hand, the likelihood ratio test shows that *Income* is significant at 3% and 1% level in the two richer specifications respectively. The A.M.E. of *Income* is negative and small in both of these specifications. This could be explained by a positive association between potential compensation and the income level of the county in which the lawsuit is filed (which is confirmed by our structural estimates in Section 6). This is because higher compensation makes it less likely for optimism-weighted compensation YC to be small relative to savings in litigation costs due to settlement. Across all specifications, the severity dummies are almost always significant at 1% level. The signs

[10] See http://www.census.gov/did/www/saipe/data/statecounty/data/index.html

[11] The extrapolation is done based on a mild assumption that a county's growth rate relative to the state-wide growth rate remains steady in adjacent years. For example, if the ratio between the growth rate in County A between 1993 and 1995 and the contemporary state-wide growth rate is α, then we maintain the yearly growth rates in County A in 1993-94 (and 1994-95) are both equal to $\sqrt{\alpha}$ times the state-wide growth rates in 1993-94 (and 1994-95 respectively). With the yearly growth rate in County A beteen 1993-1995 calculated, we then extrapolate the median household income in County A in 1994 using the data from the *SAIPE* source.

of the average marginal effects of these severity dummies suggest that higher severity leads to greater probability for settlement *paribus ceteris*.[12]

[**Insert Tables 4 and 6**]

In Table 4, we report estimates in three logit regressions of the binary outcome of court trials (D) on the characteristics of doctors, patients and lawsuits, using 1,247 observations that did not reach any settlement and had to be resolved through court trials. Across all specifications in Table 4, the court is significantly less likely to rule in favor of the plaintiff in the lawsuits that involve a board certified doctor. In contrast, the educational background (*Graduate*) does not have any significant bearing on the outcome of a court trial. Table 6 shows the severity dummies are again almost always significant at 1% level. The signs of the average marginal effects of these dummies suggest that, holding other factors fixed, the court is more likely to rule in favor of the patients as the severity level increases. Table 6 also shows that the courts are statistically more likely to rule against the doctors when the patient is older or the lawsuit is filed in a county with greater household income.

6 Estimation Results

In this section, we discuss how to implement the Maximum Simulated Likelihood estimator, and report our structural estimates related to the distribution of potential compensation C and beliefs (μ_p, μ_d). Using these estimates, we examine how the characteristics of the lawsuit affect the beliefs and the potential compensation.

• **Definition of Clusters**

We resort to institutional details to construct clusters used in identification and estimation. To reiterate, a "cluster" is a group of independent cases that are reasonably assumed to have the same length of the unreported wait-time. Chapter 766 ("Medical Malpractice and Related Matters") of Title XLV ("Torts") in The Florida Statute delineates the legal and logistic procedures in a typical medical litigation process.[13] In that chapter, Section 108 ("Mandatory Settlement Conference in Medical Negligence Actions") states that the timing of settlement conferences is subject to binding schedule constraints of the judges or other legal professionals appointed by the court who have the authority to formalize settlement.

In addition, we consulted with practicing attorneys and learned that the dates for settlement conferences as well as jury hearings in a county court are often determined by the backlog of cases filed with the court and the availability of its judges and authorized legal professionals. Based on such information, we conclude that it is plausible that the lawsuits

[12] This is consistent with the patterns in Table 1(b), which does not condition on case characteristics and aggregates over lawsuits within subgroups defined by *Severity* and *Board*.

[13] Web link to The Florida Statute: http://www.leg.state.fl.us/Statutes/

filed with the same county court in the same month would be scheduled for settlement conferences and court proceedings in the same period. Hence we maintain that all lawsuits filed with the same county court in the same month share the same length of wait-time between the settlement conferences and court hearings.

• **Econometric Specification**

The reduced-form analyses in Section 5.2 shows that the severity of injury interacts with other patient and doctor characteristics to determine the litigation outcomes. We partition the data into three subsets with different levels of *Severity*, and apply the MSL estimator defined in Section 4 conditioning on *Severity.* Our specification for $x_{n,i}$ in the distribution of potential compensation consists of a constant, *Age, Gender, Board, Graduate, Income, Costs* and Age^2; our choice of $w_{n,i}$ in the belief distribution consists of a constant, *Age, Gender, Board, Graduate, Income, Costs* and *Age*$\times$*Income.* We maintain that the distribution of the wait-time in the data is binomial with a maximum value of 25 quarters.

The data consists of 2,464 clusters defined by county-month pairs. In total, there are 771 clusters which report at least three medical malpractice lawsuits. About 93.9% of these clusters (724 clusters) include at least two lawsuits that were settled during the settlement conference. These features of the data confirm that we can apply our identification strategy from Section 3 to recover the joint distribution of patient and doctor beliefs. It is worth emphasizing that the likelihood in our MSL estimation is calculated using all 6,405 observations from the 2,464 clusters in order to improve the efficiency of the estimator, even though identification only requires the joint distribution of settlement decisions and accepted offers from a subset of clusters with at least three cases. In our estimation, we use a quarterly discount factor of 99% (which is consistent with a 4% annual inflation rate).

• **Estimates of Potential Compensation**

For each lawsuit in the data, we estimate the rate parameter in the distribution of potential compensation $\hat{\lambda}_{n,i} \equiv \exp\{x_{n,i}\hat{\beta}\}$, where $\hat{\beta}$ is the MSL estimate. We then calculate the expected compensation in each lawsuit, which is a closed-form function of $\hat{\lambda}_{n,i}$. Figure 2 reports the histograms of estimated mean compensation for all 6,405 cases, conditioning on *Age* and *Severity.* It is worth emphasizing that, unlike Table 2, the histograms in Figure 2 pertains to the complete, untruncated population of lawsuits. In comparison, Table 2 only reports the sample means of compensation *conditional on absence of any settlement.*

[**Insert Figure 2**]

Figure 2 reveals two patterns. First, within each age group, the expected compensation tends to increase with the level of severity. This is not surprising because by definition the total compensation is meant to be positively associated with the patient's loss of welfare due to the damage. Second, for a fixed level of severity, the expected compensation tends to be lower for groups with elder patients. This latter pattern is more pronounced for cases with medium and high severity. This pattern could be explained by the fact that a patient's life span following the alleged malpractice is shorter for groups with more senior patients.

[Insert Table 7]

We regress the estimates of expected compensation on the characteristics of lawsuits in Table 7, including dummy variables for each level of severity. We calculate the standard errors using 200 bootstrap samples, each drawn with replacement from the original estimation sample of 6,405 observations. Table 7 also reports the average Adjusted R^2 across these 200 bootstrap samples. The coefficients for the severity dummies are all significant at 1% level, and their point estimates increase with the level of severity. Pairwise t-tests show that the difference in these estimates are all statistically significant at 1% level. This confirms the intuition that, holding other factors fixed, the expected compensation increases with the level of severity.

F-tests for significance of *Age* and *Income* both yield p-values less than 0.01. The average marginal effect of *Age* is estimated to be -3.426. This means on average a one-year increase in patient age reduces the expected compensation by over $\$3,426$. Again this is consistent with the intuition that compensation is negatively related to the life span after the alleged malpractice. In addition, our estimate shows *Income* is significant with a positive average marginal effect. Specifically, an increase of $\$1,000$ in the county's median household income on average leads to an increase of \$885 in the expectation of potential compensation. This is evidence that the jury verdicts in the court may have at least partially taken into account a patient's prior living standards.

• **Estimates of Patient and Doctor Beliefs**

Next, we report our estimates for the parameters in the joint distribution of beliefs. For each one of the 6,405 lawsuits in the data, we estimate the concentration parameters by plugging in the estimates for the belief parameters $\hat{\alpha}_{j,n,i} \equiv \exp\{w_{n,i}\hat{\rho}_j\}$ for $j = 1, 2, 3$, where $\hat{\rho}_j$'s are MSL estimates. We then estimate the mean, standard deviation, skewness and correlation of patient and doctor beliefs $(\mu_{p,n,i}, \mu_{d,n,i})$ in each lawsuit by plugging $\hat{\alpha}_{n,i} \equiv (\hat{\alpha}_{j,n,i})_{j=1,2,3}$ into the analytical expression for these parameters. (See Table C1 in Appendix C for closed-form expressions.) Table 8 reports the sample average of estimates for these distributional parameters conditioning on *Severity*. To visualize estimates of the joint belief distribution, we plot in Figure 3 the contour graphs of the joint density of beliefs based on the sample mean of $\hat{\alpha}_{n,i}$ conditional on *Severity*. The figure also reports histograms of the estimates for correlation between $\mu_{p,n,i}$ and $\mu_{d,n,i}$ in each lawsuit.

[Insert Table 8]

Table 8 and Figure 3 reveal some interesting patterns regarding the beliefs of both parties. First, Table 8 shows that on average patients are more optimistic and doctors are more pessimistic in cases with higher severity. Both monotonicity patterns are statistically significant at 1% level. Furthermore, Table 8 suggests the distribution of the patient's belief is increasingly skewed to the left as *Severity* increases while that of the doctor's belief is increasingly skewed to the right. This is consistent with the pattern in the means of beliefs. Recall that the descriptive analysis of court trial outcomes (Table 6) shows the level of severity has a positive marginal effect on the probability that the court rules against the defendant ($D = 1$). Thus, these monotonicity patterns suggest that the beliefs held by both parties are consistent with the impact of severity on the outcome of court trials in data.

[**Insert Figure 3**]

Second, the histograms in Figure 3 and the last row in Table 8 show that, conditional on *Severity*, the correlation between the patient and doctor beliefs is significantly more negative for the cases with higher severity. The contour graphs in Figure 3 suggest that these patterns may arise because the joint density of beliefs tends to shift the mode (and possibly the probability mass) to the lower-right corner of the support (with greater values for μ_p and smaller values for μ_d) as severity increases. Third, the joint optimism ($Y \equiv \mu_p + \mu_d - 1$) diminishes as severity increases. Pairwise t-tests conclude that the mean of Y decreases significantly as the level of severity increases.

Put together, these patterns in the estimates for the joint density provide evidence that the doctors and patients may have some partial consensus in terms of how they account for the impact of severity on court decisions in their beliefs. For cases with higher severity the doctor and patient beliefs tend to be relatively more consistent with each other, thus reducing the discrepancies leading to the joint optimism.

• **Impact of Doctor Qualification on Beliefs**

We now analyze how the doctor's qualifications (*Board* and *Graduate*) affect the beliefs of both parties. For this purpose, we report in Table 9 the sample average of the estimated mean of $\mu_{p,n,i}$ and $\mu_{d,n,i}$ after controlling for these qualifications.

[**Insert Table 9**]

The first panel in Table 9 shows that, on average, patients are more optimistic when suing board-certified doctors, regardless of the level of severity. In addition, conditioning on *Severity*, board-certified doctors are on average more pessimistic than those who report no certification from any professional board. The second panel in Table 9 shows that the mean of patient beliefs is significantly higher in cases with low and medium severity when the doctor holds a non-U.S. medical degree. On the other hand, for cases with high severity, patients become slightly less optimistic when doctors have a non-U.S. education background. The doctor's belief demonstrates a similar pattern with reversed signs of changes. Doctors holding non-US medical degrees are significantly more pessimistic except for cases with high severity, where the difference in the mean belief between U.S.- and Non-U.S.-educated doctors is small.

These patterns from Table 9 are not conformable to the marginal effect of doctor qualifications on jury verdicts revealed in our descriptive analysis. Recall the logit regression of outcomes from court trials in Table 4. Across all specifications in Table 4, the doctor's board certification status is shown to have significantly negative effects on the probability that the court rules in favor of the plaintiff. In addition, the doctor's educational background (*Graduate*) is shown to have no significant impact on this probability. Thus we conclude from the patterns in Table 9 that both parties share some misperception about how doctor qualifications affect the outcome from court trials. Specifically, both sides are inclined to believe that a non-U.S. educational background would reduce the doctor's chance to win the lawsuit while the data does not suggest any significant role of the education background. They also tend to misinterpret a doctor's board certification as a factor that decreases the chance for a jury verdict in favor of the defendant, while it in fact increases such chance in data.

Finally, Table 9 suggests that the plaintiff's and the defendant's misperception about doctor qualifications is much less pronounced in the cases with high severity. To see this, note the discrepancies in the mean beliefs under different doctor qualifications are much smaller when *Severity* is high. A possible explanation is that with high severity the expected compensation at stake is larger, and thus both parties may decide to make greater effort to improve the accuracy of their perceptions.

7 Impact of Compensation Caps

Using our structural estimates, we evaluate the consequences of a hypothetical tort reform which imposes caps on the maximum compensation that may be awarded to the patient if the court rules against the defendant. The goal of such a tort reform is to reduce the social and administrative costs that arise in court trials by lowering the number of filed cases that need to be resolved through court procedures. The rationale for imposing caps on the potential compensation is that such caps would limit the liability of defendants and thus dampen plaintiffs' incentives to file lawsuit or reject settlement offers.

We study the impact of *Severity*-specific caps on the potential compensation C. For each level of severity, we set these caps at the 75-th empirical percentile of the compensations reported in the data. Using the MSL estimates from Section 4, we calculate the counterfactual probability for settlement outside the court as well as the mean and quartiles of accepted settlement offers. (See Appendix C3 for calculation details.)

[**Insert Table 10**]

Table 10 compares the empirical settlement probability in data with the counterfactual settlement probability predicted using MSL estimates. The three columns in the table report respectively the sample proportion of cases that were settled outside the court ("Data"), the average of predicted settlement probabilities without caps ("Est.") and the average of counterfactual settlement probabilities under proposed caps ("C.f."). The sample proportions in the first column and the estimated mean of settlement probabilities in the second are reasonably close to each other, with discrepancies being small relative to their standard errors. This shows that our estimates for the likelihood of settlement in the model fit quite well with the empirical settlement probability in data.

For lawsuits against board-certified doctors, the caps increase the probability for settlement by 1.39%, 1.19% and 3.40% for the cases with low, medium and high severity respectively. For lawsuits against the doctors with no reported board certification, increases in such probabilities are 0.48%, 4.66% and 1.24% respectively. These increases in settlement probabilities are small but statistically significant. For example, let $\bar{p}_{low}$ denote the mean settlement probability under the proposed caps for low severity cases and board-certified doctors. Then a t-statistic for the null hypothesis $H_0 : \bar{p}_{low} \leq 71.23\%$ is $(72.62\% - 71.23\%)/0.003 = 4.633$, which yields a p-value less than 1%. Likewise pairwise t-tests for other severity levels and doctor certification status also reject the one-sided nulls at the 1% significance level.

That the increase in settlement probability is small can be explained by the interaction of the distribution of joint optimism $\mu_p + \mu_d - 1$ and the unobserved wait-time between settlement conferences and court trials. The proposed tort reform affects the settlement probability by censoring the distribution of potential compensation (C) at the caps. The impact of such censoring is smaller if the optimism is heavily skewed to the right with larger probability mass close to zero and if the potential savings in defense litigation costs $\phi(t)k$ is large relative to the potential compensation. As the contour graphs in Figure 3 and Table 8 show, the distribution of optimism is indeed positively skewed with mean close to zero.

Next, we report estimates of the distribution of accepted offers under the proposed tort reform in Table 11. For each lawsuit reported in the data, we predict the counterfactual expectation and quartiles of accepted settlement offers under the compensation caps. Table 11 reports the sample averages of these estimates across lawsuits.

[**Insert Table 11**]

Table 11 shows the counterfactual means and quartiles of accepted settlement offers are all increasing in the level of severity, and such monotonic patterns are mostly significant. As a measure of model fit, we also report the estimated means of settlement offers with no binding caps in the column labeled "Est. means" in Table 11. Comparison between these estimated means and the conditional mean in data reported in Table 1(b) suggests our model provides a good fit to the data.

We compare the estimates for counterfactual means in Table 11 with the sample means of accepted offers in Table 1(b), and conclude that the reduction in mean settlement offer is quite substantial. For cases against board-certified doctors, the reduction rate is 15.48%, 20.88% and 21.95% for low, medium and high severity respectively. For cases against doctors who report no board certification, these reduction rates are 16.85%, 13.65% and 19.80%. The magnitude of the reduction is significantly increasing in severity as well. This pattern can be attributed to thicker tails in the distribution of potential compensation under higher severity levels.

8 Concluding Remarks

A fundamental step in the empirical analysis of bargaining outcomes is to show how assumptions of a model warrant that its structural elements can be unambiguously recovered from the data. Merlo and Tang (2012) discuss the identification of stochastic sequential bargaining models under various scenarios of data availability. In the current paper, we have addressed the same identification question in a prototypical model of bargaining with optimism. We have shown that all structural elements of the model are identified nonparametrically under realistic data requirements.[14]

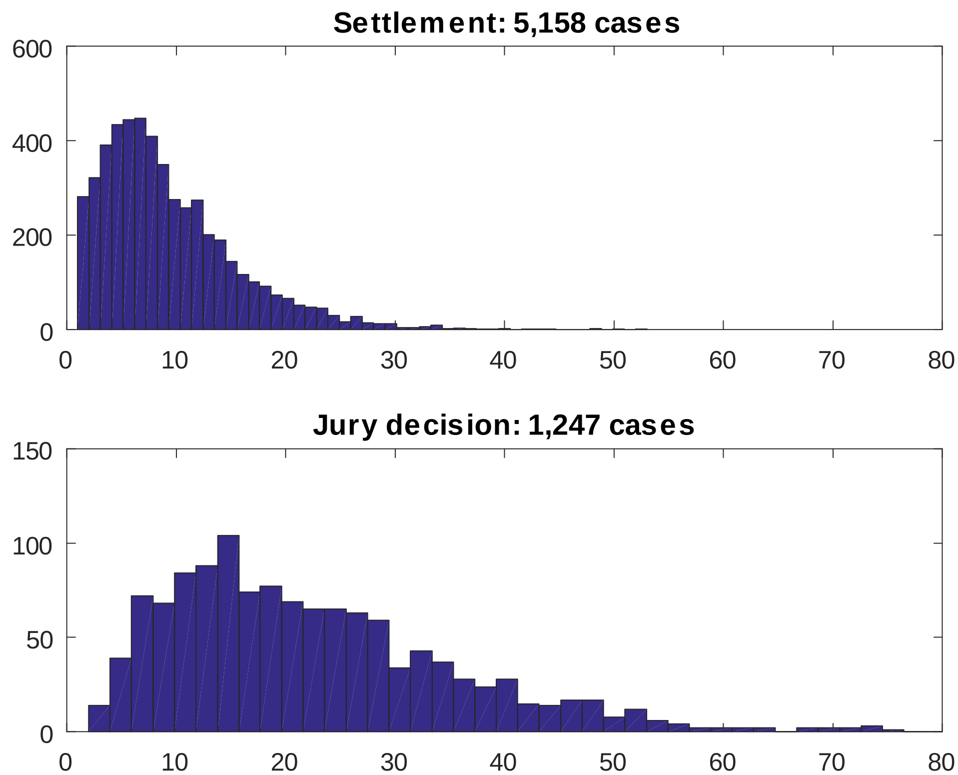
Based on our identification result, we have proposed a feasible estimation procedure using maximum simulated likelihood, and applied it to a data set on medical malpractice lawsuits in Florida in the 1980s and 1990s. We have found that patients tend to be more optimistic and doctors more pessimistic for the cases with relatively higher severity. This is consistent with the effect of severity on court decisions observed in the data. Our estimates have also shown that joint optimism exists between patients and doctors and diminishes as severity increases. This is evidence that both parties seem to correctly assess the impact of severity in their beliefs. On the other hand, there is evidence that both parties are subject to some shared misperception about the impact of doctors' qualifications (board certification and educational background) on jury verdicts. Finally, we quantify an increase in the settlement probability and a substantial reduction in the accepted settlement offers under a hypothetical tort reform that restricts the maximum compensation possible for plaintiffs.

[14]Whether our method can be extended and applied in another environment of bilateral bargaining with unobserved heterogeneity remains an open question. The answer would depend on the specifics of bargaining protocol and source of unobserved heterogeneity in the context considered.

Appendix A: Tables and Figures

Figure 1. Proxy Duration of Litigation Process

(Unit: one quarter/92 days)



Note: The histograms summarize the proxy measure of total duration of litigation, which is defined as the duration between the reported dates of the initial filing of the lawsuit and the dates of final disposition. Upper panel: cases settled outside the court; lower panel: cases resolved by jury decisions in court.

Table 1(a). Total Litigation Costs

(6,405 observations. Unit: \$1k)

	(1)	(2)	(3)
Constant	-3.414 (2.503)	-3.617 (2.613)	-3.366 (2.756)
Low	-1.784* (1.035)	-1.791* (1.036)	-1.490 (1.933)
High	-0.377 (0.940)	-0.381 (0.940)	-1.212 (1.744)
Board	2.974*** (0.871)	2.964*** (0.872)	2.962*** (0.872)
Graduate	3.876*** (0.949)	3.873*** (0.949)	3.871*** (0.949)
Income	0.342*** (0.082)	0.342*** (0.082)	0.343*** (0.082)
Duration	2.335*** (0.065)	2.380**8 (0.178)	2.349*** (0.204)
$Duration^2$		-0.002 (0.006)	-0.001 (0.006)
$Low \times Duration$			-0.028 (0.166)
$High \times Duration$			0.085 (0.152)
R-square	0.1756	0.1756	0.1757
F-test	<0.01	<0.01	<0.01

Note: This table reports the coefficient estimates in a regression of defendants' total litigation costs on the case characteristics. Standard errors are reported reported in the parentheses. 'F-test' reports the p-value in a F-test for the joint significance of all coefficients.

***: signif. at 1% level; **: signif. at 5% level; *: signif. at 10% level.

Table 1(b). Settlement probability and accepted offers

(6,405 observations. Unit: \$1k)

Board Certified	Severity	# obs	$\hat{p}_{A=1}$		$\hat{E}_{S\|A=1}$	
Yes	low	987	0.712	(0.014)	41.948	(2.159)
	medium	1,572	0.792	(0.010)	104.948	(3.814)
	high	1,867	0.835	(0.009)	271.382	(7.677)
No	low	711	0.812	(0.015)	36.681	(2.705)
	medium	679	0.851	(0.014)	82.669	(4.288)
	high	589	0.844	(0.015)	253.841	(13.800)

Notes: The table reports sample proportions of lawsuits settled outside the court ($\hat{p}_{A=1}$), and sample means of accepted settlement offers ($\hat{E}_{S|A=1}$) given the severity of injury and the doctor's board certification status. Standard errors are reported in the parentheses.

Table 1(c). Tests for Equal Settlement Probability

(6,405 observations)

	(u,l)	(u,m)	(u,h)		(c,l)	(c,m)	(c,h)
(u,l)	–	0.161	0.277	(c,l)	–	0.001	<0.001
(u,m)		–	0.809	(c,m)		–	0.023
(u,h)			–	(c,h)			–

Notes: This table reports the p-values of two-sided tests for equal settlement probability between groups defined by severity and board certification status. The letters {u,c} are short-hand for {uncertified, certified}; and {l,h,m} for {low, medium, high} respectively.

Table 2. Average Compensation to Plaintiffs

(204 observations. Unit: \$1k)

	Low		*Medium*		*High*	
$Age \leq 32$	65.587	(21.703)	178.863	(37.352)	485.388	(84.741)
$32 < Age \leq 51$	128.061	(27.898)	277.994	(57.824)	593.831	(149.117)
$Age > 51$	77.613	(18.674)	257.179	(86.618)	530.802	(126.947)

Notes: This table reports sample means of potential compensation (C) among lawsuits that are resolved in favor of the patient through court trials, conditioning on patient age and severity of injury. The cutoffs defining age groups are the 33rd and 66th percentiles in sample. Standard errors for sample means are reported in parentheses.

Table 3. Logit Regression of Settlement

(6,405 observations)

	(1)	(2)	(3)
Constant	1.519*** (0.233)	1.859*** (0.516)	0.025 (1.363)
Low	-0.340*** (0.115)	-0.981** (0.482)	-0.006 (1.981)
High	0.323*** (0.118)	0.264 (0.486)	0.003 (1.954)
Age	0.003 (0.002)	-7.271e-6 (0.011)	-1.178e-5 (0.010)
Gender	-0.308*** (0.064)	-0.308*** (0.065)	-0.322*** (0.065)
Board	-0.345*** (0.073)	-0.348*** (0.073)	-0.334*** (0.073)
Graduate	0.171** (0.079)	0.177** (0.080)	0.174** (0.080)
Income	0.007 (0.007)	-4.415e-4 (0.018)	0.153 (0.107)
Costs	-0.015 (0.032)	-0.018 (0.032)	-0.081* (0.048)
$Age \times Costs$	-5.303e-5 (3.809e-4)	-2.411e-5 (4.026e-4)	-6.657e-5 (4.057e-4)
$Low \times Costs$	-0.020 (0.028)	-0.021 (0.028)	0.048 (0.063)
$High \times Costs$	-0.013 (0.023)	-0.015 (0.023)	0.081 (0.057)
$Costs^2$	3.112e-5 (0.001)	1.112e-4 (0.001)	0.005 (0.003)
$Age \times Income$		-1.964e-4 (3.428e-4)	-2.481e-4 (3.373e-4)
$Low \times Income$		0.023 (0.017)	-0.060 (0.159)
$High \times Income$		0.002 (0.017)	0.010 (0.156)
Age^2		1.206e-4* (6.901e-5)	1.501e-4* (6.964e-5)
$Low \times Age^2$		-2.125e-5 (4.870e-5)	-5.443e-5 (4.897e-5)
$High \times Age^2$		-1.114e-5 (4.777e-5)	-2.038e-5 (4.809e-5)
$Income^2$			-0.003 (0.002)
$Low \times Income^2$			0.002 (0.003)
$High \times Income^2$			-1.497e-4 (0.003)
$Low \times Costs^2$			-0.005 (0.004)
$High \times Costs^2$			-0.007 (0.004)
Log likelihood	-3102.108	-3098.658	-3096.810
p-value for L.R.T.	<0.001	<0.001	<0.001

Note: This table reports the coefficient estimates in logit regressions of the settlement dummy (A) on the case characteristics. Standard errors are reported in the parentheses. The last row reports the p-values for LR tests of the joint significance of all regressors. ***: significant at 1% level; **: significant at 5% level; *: significant at 10% level.

Table 4. Logit Regression of Court Decisions

(1,247 observations)

	(1)	(2)	(3)
Constant	2.464 (2.057)	-1.338* (0.715)	0.0128 (2.230)
Low	-2.187 (1.312)	-0.683 (0.778)	-0.054 (1.407)
High	0.482 (1.072)	-0.592 (0.603)	0.035 (1.296)
Age	0.015** (0.007)	-0.009 (0.023)	0.008 (0.025)
Gender	-0.469*** (0.160)	-0.490*** (0.160)	-0.517*** (0.161)
Board	-0.387** (0.172)	-0.406** (0.172)	-0.330* (0.173)
Graduate	-0.291 (0.207)	-0.275 (0.206)	-0.193 (0.204)
Income	-0.316** (0.161)	0.008 (0.017)	-0.151 (0.167)
$Low \times Age$	-0.015 (0.011)	0.032 (0.038)	-0.005 (0.037)
$High \times Age$	-0.006 (0.009)	0.069** (0.031)	0.056* (0.033)
$Low \times Income$	0.079* (0.043)		0.007 (0.042)
$High \times Income$	-0.006 (0.038)		-0.012 (0.040)
$Income^2$	0.006* (0.003)		0.003 (0.003)
Age^2		2.83e-4 (2.71e-4)	1.17e-4 (2.88e-4)
$Low \times Age^2$		-5.89e-4 (4.45e-4)	-2.18e-4 (4.34e-4)
$High \times Age^2$		-0.001*** (3.84e-4)	-8.85e-4** (3.99e-4)
Log likelihood	-534.371	-535.248	-533.364
p-value for L.R.T.	<0.001	<0.001	<0.001

Note: This table reports coefficient estimates in logit regressions of jury decisions (D) on the case characteristics. Standard errors are reported in the parentheses. The last row reports the p-values for LR tests of the joint significance of all regressors. ***: significant at 1% level; **: significant at 5% level; *: significant at 10% level.

Table 5. Marginal Effects of Regressors on Settlement

(6,405 observations)

	(1)		(2)		(3)	
	A.M.E.	p-value	A.M.E.	p-value	A.M.E.	p-value
Age	0.0005	0.1464	5.087e-4	0.2092	4.602e-4	0.1533
Income	0.0011	0.3020	-2.190e-5	0.0287	-0.0012	<0.0001
Costs	-0.0041	0.1350	-0.0045	<0.0001	-0.0054	<0.0001
Low	-0.0634	<0.0001	-0.0701	<0.0001	-0.0648	<0.0001
High	0.0426	0.0021	0.0374	<0.0001	0.0360	0.0307

Note: The table reports average marginal effects of regressors on settlement probability in the three logit regressions in Table 3. "A.M.E." stands for the "average marginal effect"; p-values are reported for LR tests for significance of regressors.

Table 6. Marginal Effects on Outcomes from Court Trials

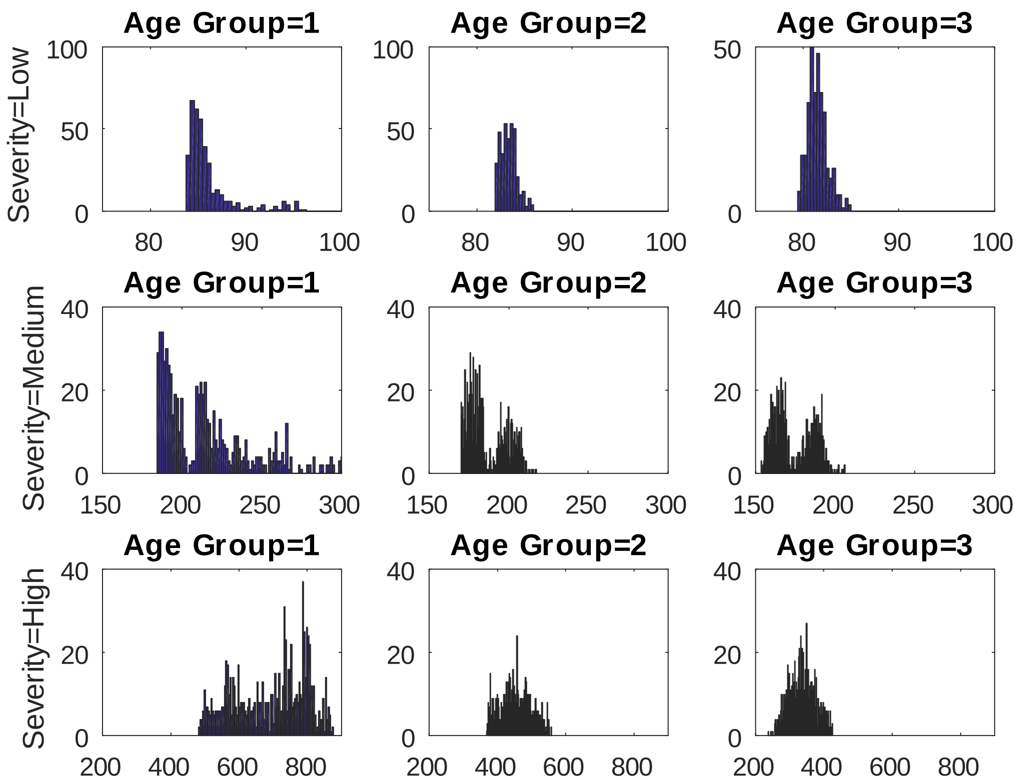
(1,247 observations)

	(1)		(2)		(3)	
	A.M.E.	p-value	A.M.E.	p-value	A.M.E.	p-value
Age	0.0012	0.0229	0.0008	0.0169	0.0008	0.0726
Income	0.0044	0.0210	0.0010	0.0119	0.0036	0.0390
Low	-0.0795	0.0018	-0.0765	<0.0001	-0.0673	<0.0001
High	0.0076	0.5865	0.0169	<0.0001	0.0204	<0.0001

Note: The table reports average marginal effects of regressors on jury decisions in the three logit regressions in Table 4. "A.M.E." stands for average marginal effect; p-values are reported for LR tests for the significance of regressors.

Figure 2. Histogram of Estimated Mean Compensation

(6,405 observations. Unit: $1k)



Note: This figure plots the histograms of the estimated expected compensation (i.e., $E(C)$) in each lawsuit. The definition of age groups are the same as in Table 2. Dotted lines plot the means of expected compensation within each group defined by *Severity* and *Age*.

Table 7. Regression: Estimated Mean Compensation (Units: \$1K)
(6,405 observations)

	Estimates	Std. Err	t-statistic
Low	227.612***	22.098	10.300
Medium	327.251***	24.309	13.462
High	614.842***	29.390	20.920
Age	-5.953***	0.487	-12.223
Gender	-5.236***	1.896	-2.761
Income	2.326	2.138	1.088
Age^2	0.051***	0.002	24.337
$Age \times Income$	-0.060***	0.013	-4.733
$Income^2$	0.019	0.033	0.569
Adjusted R^2		0.872	

Note: This table reports coefficients in a regression of estimated mean compensation (C) on case characteristics. The expected compensation for each observation (lawsuit) is calculated using the MSL estimates $\hat{\beta}$ in (13). Standard errors and Adjusted R^2 are calculated using 200 bootstrap samples. ***: significant at 1% level.

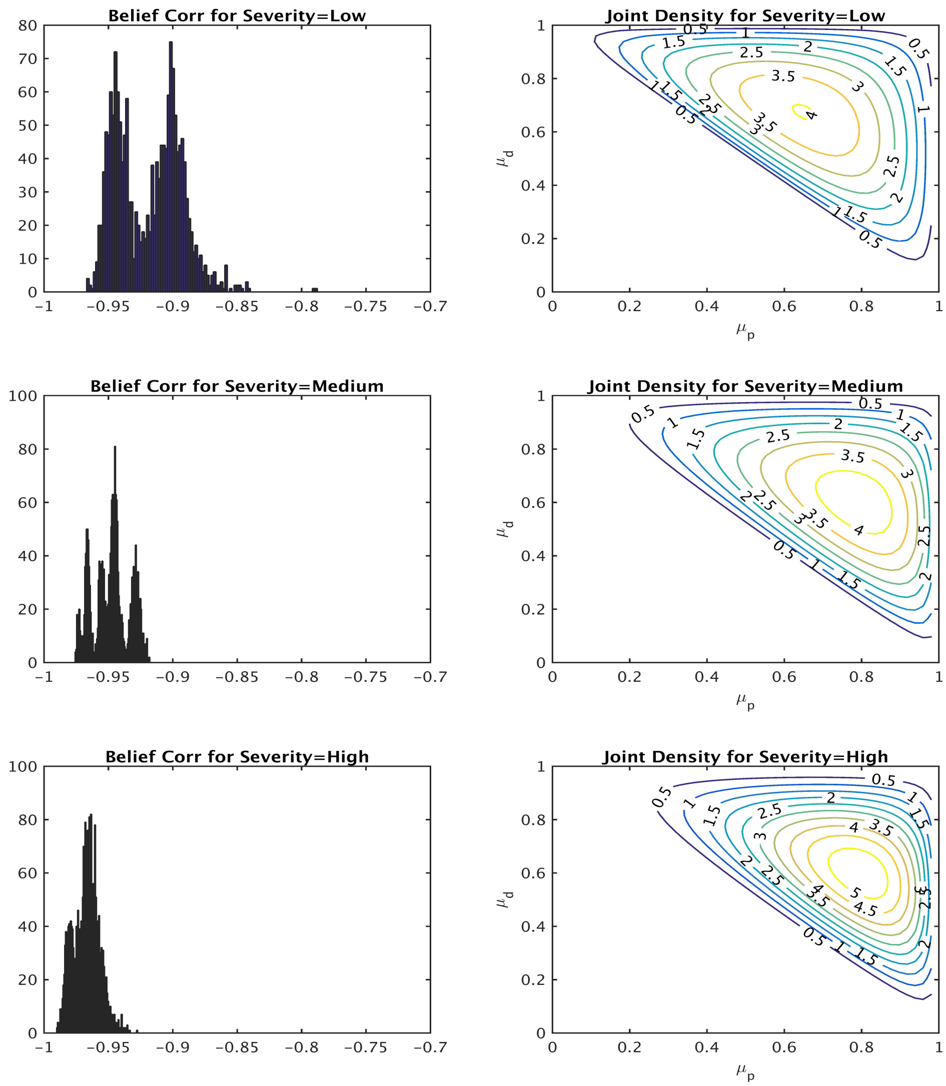
Table 8. Estimates of Belief Distribution
(6,405 observations)

	Low		Medium		High	
	μ_p	μ_d	μ_p	μ_d	μ_p	μ_d
Mean	0.5123	0.5292	0.5794	0.4467	0.5937	0.4224
	(0.0020)	(0.0019)	(0.0011)	(0.0010)	(0.0009)	(0.0009)
Std. dev.	0.0502	0.0504	0.0512	0.0524	0.0457	0.0463
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Skewness	-0.0479	-0.0835	-0.2470	0.1635	-0.2774	0.2266
	(0.0062)	(0.0056)	(0.0033)	(0.0032)	(0.0027)	(0.0025)
Correlation	-0.9187		-0.9479		-0.9670	
	(0.0061)		(0.0029)		(0.0020)	

Note: This table reports the sample averages of estimated mean, standard deviation, skewness and correlation of beliefs (μ_p, μ_d) in all observed lawsuits. Standard errors of the sample means of these parameter estimates are reported in the parentheses.

Figure 3. Estimates for the Distribution of Beliefs

(6,405 observations)



Note: This figure plots the contour graphs of the joint density of beliefs conditional on each level of severity based on the sample averages of $\hat{\alpha}_{n,i}$. It also reports the histograms of estimated correlation between μ_p and μ_d in each lawsuit conditional on the severity level.

Table 9. Mean beliefs Conditional on Doctor Qualification

(6,405 observations)

Board Certification Status

Severity	Uncertified		Certified	
	$E(\mu_p)$	$E(\mu_d)$	$E(\mu_p)$	$E(\mu_d)$
Low	0.4530	0.5752	0.5551	0.4960
	(0.0025)	(0.0024)	(0.0021)	(0.0021)
Medium	0.5127	0.5056	0.6081	0.4213
	(0.0011)	(0.0012)	(0.0006)	(0.0008)
High	0.5778	0.4396	0.5988	0.4170
	(0.0019)	(0.0018)	(0.0010)	(0.0010)

Educational Background

Severity	U.S.-based		International	
	$E(\mu_p)$	$E(\mu_d)$	$E(\mu_p)$	$E(\mu_d)$
Low	0.5056	0.5375	0.5393	0.4963
	(0.0022)	(0.0020)	(0.0044)	(0.0041)
Medium	0.5700	0.4574	0.6109	0.4103
	(0.0012)	(0.0011)	(0.0021)	(0.0020)
High	0.5978	0.4203	0.5817	0.4285
	(0.0010)	(0.0010)	(0.0018)	(0.0018)

Note: The table reports sample means of estimated expectation of beliefs across lawsuits, conditional on the severity of injury and doctor qualifications (*Board* and *Graduate*). Standard errors of sample means are reported in the parentheses.

Table 10. Settlement Probability under Caps

(6,405 observations)

Severity	Board Certified			Not Board Certified		
	Data	Est.	C.f.	Data	Est.	C.f.
Low	0.7123	0.7180	0.7262	0.8115	0.8092	0.8163
	(0.0144)	(0.0030)	(0.0030)	(0.0147)	(0.0027)	(0.0027)
Medium	0.8014	0.8065	0.8133	0.7736	0.8133	0.8202
	(0.0127)	(0.0015)	(0.0015)	(0.0157)	(0.0018)	(0.0018)
High	0.8055	0.8329	0.8395	0.8298	0.8361	0.8422
	(0.0126)	(0.0012)	(0.0012)	(0.0141)	(0.0022)	(0.0021)

Note: This table compares the empirical settlement probability in the data with counterfactual settlement probability predicted using MSL estimates. Column "Data" reports the empirical settlement probability (sample proportion for $A = 1$); "Est." reports the mean of estimated settlement probability calculated from the MSL estimates $(\hat{\rho}, \hat{\beta}, \hat{\theta})$; "C.f." reports the mean of counterfactual settlement probability under proposed caps predicted from the MSL estimates. Standard errors are reported in the parentheses.

Board Certified	Severity	# obs	$\hat{p}_{A=1}$		$\hat{E}_{S\|A=1}$	
Yes	low	987	0.712	(0.014)	41.948	(2.159)
	medium	1,572	0.792	(0.010)	104.948	(3.814)
	high	1,867	0.835	(0.009)	271.382	(7.677)
No	low	711	0.812	(0.015)	36.681	(2.705)
	medium	679	0.851	(0.014)	82.669	(4.288)
	high	589	0.844	(0.015)	253.841	(13.800)

Notes: The table reports sample proportions of lawsuits settled outside the court ($\hat{p}_{A=1}$), and sample means of accepted settlement offers ($\hat{E}_{S|A=1}$) given the severity of injury and the doctor's board certification status. Standard errors are reported in the parentheses.

Table 11. Distribution of Settlement Offers under Caps

(6,405 observations)

Sev.	Board Certified					Not Board Certified				
	est. mean	c.f. mean	25%	50%	75%	est. mean	c.f. mean	25%	50%	75%
Low	42.142	35.451	7.015	23.267	54.483	37.967	30.500	6.035	20.281	45.888
	(1.684)	(3.047)	(2.327)	(3.009)	(5.726)	(1.372)	(3.161)	(1.632)	(2.858)	(5.364)
Med.	107.704	83.034	20.577	60.208	131.380	85.211	71.381	17.112	50.953	110.304
	(2.471)	(4.484)	(2.441)	(5.791)	(8.366)	(2.638)	(4.277)	(2.577)	(5.539)	(7.778)
High	276.007	211.802	61.151	169.466	332.351	260.678	203.573	54.938	156.216	318.399
	(4.022)	(39.476)	(22.531)	(51.873)	(64.563)	(10.697)	(36.783)	(19.279)	(45.662)	(62.565)

Note: This table reports sample averages of the estimated mean settlement offers (with no caps) and the predicted mean and quartiles of settlement offers under the hypothesized caps. The means and quartiles for each lawsuit observed in the data are calculated using the MSL estimates $(\hat{\rho}, \hat{\beta}, \hat{\theta})$. Standard errors are reported in the parentheses.

Appendix B: Proofs

Proof of Lemma 1. By the law of total probability,

$$f_{C_i,S_l}(c,s,A_j=1|\mathcal{E}_{i,l},\mathbf{k})=\sum_{t\in\mathcal{T}}\left\{\begin{array}{c} f_{C_i}(c|S_l=s,A_j=1,\mathcal{E}_{i,l},\mathbf{k},T=t)\times \\ \mathbb{E}[A_j|S_l=s,\mathcal{E}_{i,l},\mathbf{k},T=t]\times f_{S_l}(s,T=t|\mathcal{E}_{i,l},\mathbf{k}) \end{array}\right\}$$

In equilibrium, "$S=\delta^T\mu_pC$ when $A=1$" and "$A=1$ if and only if $YC\leq\phi(T)K$". Hence the first term in the product of the summand is

$$\begin{aligned} & f_{C_i}(c|Y_iC_i>\phi(t)k_i,D_i=1,Y_lC_l\leq\phi(t)k_l,\mu_{l,p}C_l=s/\delta^t,Y_jC_j\leq\phi(t)k_j,\mathbf{k},T=t) \\ = \;& f_{C_i}(c|A_i=0,D_i=1,k_i,T=t). \end{aligned}$$

The second equality holds under Assumption 1-(i) and Assumption 2, which state (μ_p,μ_d,C,D,K) are independent draws across observations within the same cluster and are independent from T. By a similar argument, the second term in the product is

$$\begin{aligned} & \Pr(Y_jC_j\leq\phi(t)k_j|Y_lC_l\leq\phi(t)k_l,\mu_{l,p}C_l=s/\delta^t,Y_iC_i>\phi(t)k_i,D_i=1,\mathbf{k},T=t) \\ = \;& \mathbb{E}(A_j|T=t,k_j), \end{aligned}$$

and the last term in the product is

$$f_{S_l}\left(s,T=t|Y_iC_i>\phi(T)k_i,D_i=1,Y_lC_l\leq\phi(T)k_l,\mathbf{k}\right)=f_{S_l}\left(s,T=t|\mathcal{E}_{i,l},k_i,k_l\right),$$

where the equality holds because the distribution of T and the outcomes in cases i and l does not depend on k_j. This proves (4). Equation (5) follows from the law of total probability and similar arguments. □

Proof of Lemma 2. For a generic partition of $\mathcal{S}$ into M intervals (each denoted b_m), let $L_{S_l|T}$ be a M-by-$|\mathcal{T}|$ matrix with its (m,t)-th entry being $\Pr(S_l\in b_m|A_l=1,T=t,k_l)$. Under Assumption 1 and 2, $L_{C_i,S_l}=L_{C_i|T}\Sigma_{i,l}\left(L_{S_l|T}\right)'$ where $\Sigma_{i,l}$ is a diagonal matrix with its t-th diagonal entry being $\Pr(T=t|\mathcal{E}_{i,l},k_i,k_l)$ and $\left(L_{S_l|T}\right)'$ denotes the transpose of $L_{S_l|T}$.

Case 1: $k_i\in\mathcal{K}$ with $\phi(|\mathcal{T}|)k_i<\bar{c}$. In this case, $\phi(t)k_i\in(0,\bar{c})$ for all $t\in\mathcal{T}$. Hence $\tau(k_i)=|\mathcal{T}|$. Under Assumption 3, all diagonal entries in the diagonal matrix $\Sigma_{i,l}$ are non-zero. Hence $\Sigma_{i,l}$ has full rank regardless of the value of k_l. We first show that there exists a partition $\mathcal{B}_{|\mathcal{T}|}$ on $\mathcal{S}$ such that the matrix $L_{S_l|T}$ defined using $\mathcal{B}_{|\mathcal{T}|}$ has full-rank $|\mathcal{T}|$. Recall the joint support of $(\mu_{l,p},\mu_{l,d})$ is $\{(\mu,\mu')\in(0,1)^2:1<\mu+\mu'<2\}$. Hence by construction, the joint support $(\mu_{l,p},Y_l)\equiv(\mu_{l,p},\mu_{l,p}-(1-\mu_{l,d}))$ is $\{(r',r)\in(0,1)^2:0<r'-r<1\}$. Under Assumption 1 and 3, this implies for any $t\in\mathcal{T}$ and *any* $k_l\in\mathcal{K}$, the support of $\mu_{l,p}C_l$ conditional on "$Y_lC_l\leq\phi(t)k_l$, $T=t$ and $K_l=k_l$" is $(0,\bar{c})$. Thus conditional on "$A=1$, $T=t$ and $K=k_l$", the accepted offer $S_l=\delta^t\mu_{l,p}C_l$ is continuously distributed over $(0,\delta^t\bar{c})$.

Let $\mathcal{B}_{|\mathcal{T}|}$ be a partition of the unconditional support of settlement offers $\mathcal{S}$ into $|\mathcal{T}|$ intervals, which are characterized by the sequence of endpoints $\bar{s}(1) > \bar{s}(2) > \bar{s}(3) > .. > \bar{s}(|\mathcal{T}|) > \bar{s}(|\mathcal{T}|+1)$, where $\bar{s}(t) \equiv \delta^t \bar{c}$ for $t = 1, 2, ..., |\mathcal{T}|$ and $\bar{s}(|\mathcal{T}|+1) \equiv 0$. (That is, the t-th interval in $\mathcal{B}_{|\mathcal{T}|}$ is $[\bar{s}(|\mathcal{T}|-t+2), \bar{s}(|\mathcal{T}|-t+1)]$ for $t = 1, 2, ..., |\mathcal{T}|$.) Because conditional on $A = 1$ and t, k, the settlement offer S is continuously distributed over $(0, \delta^t \bar{c})$, the square matrix $L_{S_i|T}$ defined using $\mathcal{B}_{|\mathcal{T}|}$ is triangular with strictly positive diagonal entries, and therefore is non-singular with full rank $|\mathcal{T}|$.

Next, we show that there exists a partition $\mathcal{D}_{|\mathcal{T}|}$ on $\mathcal{C}$ such that the matrix $L_{C_i|T}$ defined using $\mathcal{D}_{|\mathcal{T}|}$ has full rank $|\mathcal{T}|$. Under Assumption 1 and 3, the support of C_i conditional on "$Y_i C_i > \phi(t) k_i$, $D_i = 1$, $T = t$ and $K_i = k_i$" is $(\phi(t) k_i, \bar{c})$ because the support of Y is $(0, 1)$. For any $t \in \mathcal{T} \equiv \{1, 2, ., |\mathcal{T}|\}$, let $\underline{c}(t) \equiv \phi(t) k_i$ denote the inf of the support of C conditional on "$A_i = 0$, $D_i = 1$, $T = t$, $K_i = k_i$". By construction, $\underline{c}(1) < \underline{c}(2) < \underline{c}(3) < ... < \underline{c}(|\mathcal{T}|) < \bar{c} \equiv \underline{c}(|\mathcal{T}|+1)$ under Assumption 3 for k_i s.t. $\phi(|\mathcal{T}|) k_i < \bar{c}$. Let $\mathcal{D}_{|\mathcal{T}|}$ be a partition of the unconditional support of $\mathcal{C}$ into $|\mathcal{T}|$ intervals, which are characterized by the sequence of endpoints $\underline{c}(t), t = 1, 2, .., |\mathcal{T}|+1$. (That is, the t-th interval in $\mathcal{D}_{|\mathcal{T}|}$ is $(\underline{c}(t), \underline{c}(t+1))$ for $t = 1, 2, ..., |\mathcal{T}|$.) The square matrix $L_{C_i|T}$ defined using $\mathcal{D}_{|\mathcal{T}|}$ is triangular with strictly positive diagonal entries, and therefore is non-singular with full rank $|\mathcal{T}|$.

To sum up, under the stated conditions, for any $k_i \in \mathcal{K}$ with $\phi(|\mathcal{T}|) k_i < \bar{c}$ and *any* $k_l \in \mathcal{K}$, the matrices $L_{C_i|T}$, $\Sigma_{i,l}$ and $L_{S_l|T}$ defined using $\mathcal{B}_{|\mathcal{T}|}$ and $\mathcal{D}_{|\mathcal{T}|}$ have full rank. Hence $L_{C_i,S_l} = L_{C_i|T} \Sigma_{i,l} \left(L_{S_l|T}\right)'$ has full rank.

Case 2: $k_i \in \mathcal{K}$ with $\phi(|\mathcal{T}|) k_i \geq \bar{c}$. In this case $\tau(k_i) \leq |\mathcal{T}| - 1$. Then for any such k_i and for *all* $k_l \in \mathcal{K}$, $\Pr(T = t | \mathcal{E}_{i,l}, k_i, k_l) > 0$ for $t \leq \tau(k_i)$ and $\Pr(T = t | \mathcal{E}_{i,l}, k_i, k_l) = 0$ for $t > \tau(k_i)$. This is because $\Pr(YC > \phi(t) k_i) = 0$ for all $t > \tau(k_i)$ under the support conditions in Assumption 3. Thus for a generic partition of $\mathcal{C}$ into M intervals and partition of $\mathcal{S}$ into M intervals, we can write $L_{C_i,S_l} = \tilde{L}_{C_i|T} \tilde{\Sigma}_{i,l} (\tilde{L}_{S_l|T})'$ where $\tilde{\Sigma}_{i,l}$ is a $\tau(k_i)$-by-$\tau(k_i)$ diagonal matrix with its t-th diagonal entry $\Pr(T = t | \mathcal{E}_{i,l}, k_i, k_l)$ and $\tilde{L}_{C_i|T}$ is a M-by-$\tau(k_i)$ matrix with its (m, t)-th entry being $\Pr(C_i \in d_m | A_i = 0, D_i = 1, T = t, k_i)$, and $\tilde{L}_{S_l|T}$ is a M-by-$\tau(k_i)$ matrix with its (m, t)-th entry being $\Pr(S_l \in b_m | A_l = 1, T = t, k_l)$.

For any such k_i and any k_l we can construct a partition $\mathcal{D}_{\tau(k_i)}$ on $\mathcal{C}$ (in a similar fashion to Case 1 using the infima of supports of C conditional on "$(1 - A_i) D_i = 1, t, k_i$" as the grids) and a partition $\mathcal{B}_{\tau(k_i)}$ on $\mathcal{S}$ (by combining some of the intervals in $\mathcal{B}_{|\mathcal{T}|}$ in Case 1) so that $\tilde{L}_{C_i|T}$ and $\tilde{L}_{S_l|T}$ are triangular with positive diagonal entries. Besides, the diagonal matrix $\tilde{\Sigma}_{i,l}$ is non-singular by construction. It then follows that L_{C_i,S_l} defined by $\mathcal{D}_{\tau(k_i)}$ and $\mathcal{B}_{\tau(k_i)}$ has rank $\tau(k_i)$ for all k_l. $\square$

Proof of Proposition 1. **Step 1.** Consider any pair of (k_i, k_j) such that $\phi(|\mathcal{T}|) k_i < \bar{c}$ and $\phi(|\mathcal{T}|) k_j < \bar{c}$ (that is, $\tau(k_i) = \tau(k_j) = |\mathcal{T}|$). By Lemma 2, for *all* k_l there exists a partition $\mathcal{D}_{|\mathcal{T}|}$ on $\mathcal{C}$ and a partition $\mathcal{B}_{|\mathcal{T}|}$ on $\mathcal{S}$ such that the square matrix L_{C_i,S_l} defined using $\mathcal{D}_{|\mathcal{T}|}$ and $\mathcal{B}_{|\mathcal{T}|}$ has full rank $|\mathcal{T}|$. Let d_m and b_m denote the m-th interval in $\mathcal{D}_{|\mathcal{T}|}$ and $\mathcal{B}_{|\mathcal{T}|}$ respectively. The argument presented in the text of Section 3.1 shows that $L_{C_i|T}$, Δ_j and L_{T,S_l} are identified for all k_l and any such pair (k_i, k_j). For simplicity, we suppress dependence of $L_{C_i|T}$, Δ_j and L_{T,S_l} on k_i, k_j and (k_i, k_l) in notation respectively.

It remains to show that the conditional density $f_{C_i}(.|A_i = 0, D_i = 1, T = t, k_i)$ is identified over its full domain for all $t \in \mathcal{T}$. For any $c \in \mathcal{C}$ let l_c denote a $|\mathcal{T}|$-vector where the m-th coordinate is $f_{C_i}(c, S_l \in b_m|\mathcal{E}_{i,l}, k_i, k_l)$. By construction,

$$l_c = (L_{T,S_l})' \lambda_c \tag{14}$$

where λ_c is a $|\mathcal{T}|$-vector with the t-th component being $f_{C_i}(c|A_i = 0, D_i = 1, T = t, k_i)$.[15] The coefficient matrix L_{T,S_l} does not depend on the value of c while both vectors λ_c and l_c do. By an argument in the proof of Lemma 2, L_{T,S_l} is invertible and identified. Thus, with l_c directly identifiable and with L_{T,S_l} identified and non-singular, λ_c is recovered as the unique solution of the linear system in (14) for any $c \in \mathcal{C}$.

Next we show that the conditional density $f_{S_l}(.|A_l = 1, T = t, k_l)$ is identified over its full domain for all $t \in \mathcal{T}$ and $k_l \in \mathcal{K}$. As before, let $L_{S_l|T}$ denote a $|\mathcal{T}|$-by-$|\mathcal{T}|$ matrix defined using $\mathcal{B}_{|\mathcal{T}|}$, with its (m,t)-th entry being $\Pr(S_l \in b_m|A_l = 1, T = t, k_l)$. Let L_{T,C_i} denote a $|\mathcal{T}|$-by-$|\mathcal{T}|$ matrix defined using $\mathcal{D}_{|\mathcal{T}|}$, with its (t,m)-th entry being $\Pr(T = t, C_i \in d_m|\mathcal{E}_{i,l}, k_i, k_l)$. Then a symmetric argument using transposes of Λ_{C_i,S_l} and L_{C_i,S_l} identifies the non-singular matrices $L_{S_l|T}$ and L_{T,C_i}. Specifically, let Λ_{S_l,C_i} and L_{S_l,C_i} denote the transpose of Λ_{C_i,S_l} and L_{C_i,S_l} respectively. By construction, $\Lambda_{S_l,C_i} = L_{S_l|T}\Delta_j L_{T,C_i}$ and $L_{S_l,C_i} = L_{S_l|T}L_{T,C_i}$. By a similar argument, $L_{S_l|T}$ is non-singular and $\Lambda_{S_l,C_i}(L_{S_l,C_i})^{-1} = L_{S_l|T}\Delta_j \left(L_{S_l|T}\right)^{-1}$. Thus $L_{S_l|T}$ is identified as the matrix of eigenvectors in the eigenvalue-decomposition of the left-hand side. It then follows that $L_{T,C_i} = (L_{S_l|T})^{-1}L_{S_l,C_i}$ is identified. For any $s \in \mathcal{S}$ let l_s denote a $|\mathcal{T}|$-vector l_s whose m-th coordinate is $f_{S_l}(s, C_i \in d_m|\mathcal{E}_{i,l}, k_i, k_l)$. Then

$$l_s = (L_{T,C_i})'\lambda_s \tag{15}$$

where λ_s is a $|\mathcal{T}|$-vector with the t-th component being $f_{S_l}(s|A_l = 1, T = t, k_l)$. Thus, with l_s directly identifiable and with L_{T,C_i} identified and non-singular, λ_s is recovered as the unique solution of the linear system in (15) for any $s \in \mathcal{S}$ and for all t and k_l.

Step 2. Consider any (k_i, k_j) such that $\phi(|\mathcal{T}|)k_i \geq \bar{c}$ and $\phi(|\mathcal{T}|)k_j < \bar{c}$ (i.e., $\tau(k_i) \leq |\mathcal{T}|-1$ and $\tau(k_j) = |\mathcal{T}|$). By construction, for all $t > \tau(k_i)$, $\Pr(A_i = 0|T = t, k_i) = 0$ and therefore $f_C(.|A_i = 0, D_i = 1, T = t, k_i)$ is not defined for such pairs of (t, k_i). Hence we only need to identify $f_C(.|A_i = 0, D_i = 1, T = t, k_i)$ for $t \leq \tau(k_i)$.

[15] To see this, note for any c and b_m, the law of total probability implies $f_{C_i}(c, S_l \in b_m|\mathcal{E}_{i,l}, k_i, k_l)$ equals

$$\textstyle\sum_{t\in\mathcal{T}} f_{C_i}(c|T = t, S_l \in b_m, \mathcal{E}_{i,l}, k_i, k_l) \Pr(T = t, S_l \in b_m|\mathcal{E}_{i,l}, k_i, k_l)$$

where the density in the summand equals $f_{C_i}(c|\ A_i = 0,\ D_i = 1,\ T = t,\ k_i)$ due to Assumption 1 and 2.

By Lemma 2, for all k_l there exists a partition $\mathcal{D}_{\tau(k_i)}$ on $\mathcal{C}$ and a partition $\mathcal{B}_{\tau(k_i)}$ on $\mathcal{S}$ such that the square matrix L_{C_i,S_l} defined using $\mathcal{D}_{\tau(k_i)}$ and $\mathcal{B}_{\tau(k_i)}$ has full rank $\tau(k_i)$. (Recall that L_{C_i,S_l} is a $\tau(k_i)$-by-$\tau(k_i)$ matrix defined using $\mathcal{D}_{\tau(k_i)}$ and $\mathcal{B}_{\tau(k_i)}$ with (m,m')-th entry being $\Pr(C_i \in d_m, S_l \in d_{m'}|\mathcal{E}_{i,l}, k_i, k_l)$, where d_m and b_m denote the m-th interval in $\mathcal{D}_{\tau(k_i)}$ and $\mathcal{B}_{\tau(k_i)}$ respectively). For any k_l, let Λ_{C_i,S_l} denote the $\tau(k_i)$-by-$\tau(k_i)$ matrix defined using $\mathcal{D}_{\tau(k_i)}$ and $\mathcal{B}_{\tau(k_i)}$ whose (m,m')-th entry is $\Pr(C_i \in d_m, A_j = 1, S_l \in d_{m'}|\mathcal{E}_{i,l}, k_i, k_j, k_l)$; $\tilde{L}_{C_i|T}$ denote the $\tau(k_i)$-by-$\tau(k_i)$ matrix defined using $\mathcal{D}_{\tau(k_i)}$, whose (m,t)-th entry is $\Pr(C_i \in d_m|A_i = 0, D_i = 1, T = t, k_i)$; $\tilde{L}_{T,S_l}$ the $\tau(k_i)$-by-$\tau(k_i)$ matrix defined using $\mathcal{B}_{\tau(k_i)}$, whose (t,m)-th entry is $\Pr(T = t, S_l \in b_m|\mathcal{E}_{i,l}, k_i, k_l)$; and $\tilde{\Delta}_j$ denote a $\tau(k_i)$-by-$\tau(k_i)$ diagonal matrix with the t-th diagonal entry being $\mathbb{E}(A_j|T = t, k_j)$.

By construction, $\Lambda_{C_i,S_l} = \tilde{L}_{C_i|T}\tilde{\Delta}_j\tilde{L}_{T,S_l}$, and $L_{C_i,S_l} = \tilde{L}_{C_i|T}\tilde{L}_{T,S_l}$ is non-singular. Note in the application of the law of total probability here we have used the fact that when $\phi(|\mathcal{T}|)k_i \geq \bar{c}$ (i.e., $\tau(k_i) \leq |\mathcal{T}| - 1$), it must be the case that $\Pr(T = t|\mathcal{E}_{i,l}, k_i, k_l) > 0$ for $t \leq \tau(k_i)$ and $\Pr(T = t|\mathcal{E}_{i,l}, k_i, k_l) = 0$ for $t > \tau(k_i)$ for *all* $k_l \in \mathcal{K}$. The same argument as in the text shows that $\tilde{L}_{C_i|T}$, $\tilde{\Delta}_j$ and $\tilde{L}_{T,S_l}$ are identified using the same argument as in the text of Section 3.1. Again, under conditions of the lemma, $\mathbb{E}(A_j|T = t, k_j)$ is monotone in $t \in \{1, 2, ..., |\mathcal{T}|\}$ for k_j with $\phi(|\mathcal{T}|)k_j < \bar{c}$. This monotonicity allows us to correctly label the eigenvalues in $\tilde{\Delta}_j$ with $t \in \{1, 2, ..., \tau(k_i)\}$.

We now show that $f_{C_i}(.|A_i = 0, D_i = 1, T = t, k_i)$ is identified over its full domain for all $t \leq \tau(k_i)$. For any $c \in \mathcal{C}$ let $\tilde{l}_c$ denote a $\tau(k_i)$-vector where the m-th coordinate is $f_{C_i}(c, S_l \in b_m|\mathcal{E}_{i,l}, k_i, k_l)$. By construction,

$$\tilde{l}_c = \left(\tilde{L}_{T,S_l}\right)' \tilde{\lambda}_c \tag{16}$$

where $\tilde{\lambda}_c$ is a $\tau(k_i)$-vector with the t-th component being $f_{C_i}(c|A_i = 0, D_i = 1, T = t, k_i)$. The coefficient matrix $\tilde{L}_{T,S_l}$ does not depend on the realization of $C_i = c$ while both vectors $\tilde{\lambda}_c$ and $\tilde{l}_c$ do. The matrix $\tilde{L}_{T,S_l}$ is invertible and identified as above. Thus, with $\tilde{l}_c$ directly identifiable and with $\tilde{L}_{T,S_l}$ identified and non-singular, $\tilde{\lambda}_c$ is recovered as the unique solution of the linear system in (14) for any $c \in \mathcal{C}$.

Step 3. The last step is to identify $\mathbb{E}(A_j|T = t, k_j)$ for all $t \in \mathcal{T}$ and k_j such that $\phi(|\mathcal{T}|)k_j \geq \bar{c}$ (i.e., $\tau(k_j) \leq |\mathcal{T}| - 1$). Under Assumption 1, for all such k_j, $\Pr(A_j|T = t, k_j) = \Pr(Y_j C_j \leq \phi(t)k_j) = 1$ for all $t > \tau(k_j)$. Hence it only remains to identify $\mathbb{E}(A_j|T = t, k_j)$ for $t \leq \tau(k_j)$ and such k_j. This is done by repeating the eigenvalue decomposition in Step 2 for a pair (k_i, k_j) with $\phi(|\mathcal{T}|)k_i < \bar{c}$ and $\phi(|\mathcal{T}|)k_j \geq \bar{c}$, and use the monotonicity of $\mathbb{E}(A_j|T = t, k_j)$ over $t = 1, 2, ..., \tau(k_j)$ to label the eigenvalues. $\square$

Appendix C: Computational Details

C1. Parametrization of joint belief distribution

For simplicity, consider a simple design where cases are homogenous (no $x_{n,i}$'s). Let the data-generating process be defined as follows. Let $(y, \tilde{y}) \in \{(y, \tilde{y}) \in [0,1]^2 : 0 \leq y + \tilde{y} \leq 1\}$ and let $(\tilde{y}, y, 1 - y - \tilde{y})$ follow a Dirichlet distribution with concentration parameters $(\alpha_1, \alpha_2, \alpha_3)$. Let $\mu_p = 1 - \tilde{Y}$ and $\mu_d = Y - (1 - \tilde{Y}) + 1 = Y + \tilde{Y}$. By construction, the support of (μ_p, μ_d) is $\{(\mu, \mu') \in [0,1]^2 : 1 \leq \mu + \mu' \leq 2\}$, which is consistent with our model of bargaining with optimism. The marginal distribution of μ_p is $Beta(\alpha_2 + \alpha_3, \alpha_1)$ (because the marginal distribution of $\tilde{Y}$ is $Beta(\alpha_1, \alpha_2 + \alpha_3)$); and the marginal distribution of μ_d is $Beta(\alpha_1 + \alpha_2, \alpha_3)$.[16] Let $\alpha_0 \equiv \alpha_1 + \alpha_2 + \alpha_3$. Table C1 below summarizes the relation how the concentration parameters determine the key features of the distribution of (μ_p, μ_d):

Table C1: Summary of the Joint Distribution of (μ_p, μ_d)

	μ_p	μ_d
Marginal distr'n	$Beta(\alpha_2 + \alpha_3, \alpha_1)$	$Beta(\alpha_1 + \alpha_2, \alpha_3)$
Mean	$\frac{\alpha_2+\alpha_3}{\alpha_0}$	$\frac{\alpha_1+\alpha_2}{\alpha_0}$
Variance	$\frac{\alpha_1(\alpha_2+\alpha_3)}{\alpha_0^2(\alpha_0+1)}$	$\frac{\alpha_3(\alpha_1+\alpha_2)}{\alpha_0^2(\alpha_0+1)}$
Skewness	$\frac{2(\alpha_1-\alpha_2-\alpha_3)\sqrt{\alpha_0+1}}{(\alpha_0+2)\sqrt{\alpha_1(\alpha_2+\alpha_3)}}$	$\frac{2(\alpha_3-\alpha_1-\alpha_2)\sqrt{\alpha_0+1}}{(\alpha_0+2)\sqrt{\alpha_3(\alpha_1+\alpha_2)}}$
Mode (marginal)	$\frac{\alpha_2+\alpha_3-1}{\alpha_0-2}$	$\frac{\alpha_1+\alpha_2-1}{\alpha_0-2}$
Correlation	$-\frac{\sqrt{\alpha_1\alpha_3}}{\sqrt{(\alpha_2+\alpha_3)(\alpha_1+\alpha_2)}}$	

To sum up, Table C1 shows that the parametrization allows for flexible patterns in the joint distribution of beliefs (μ_p, μ_d).

C2. Details in estimation

While estimating the conditional distribution of compensation given $x_{n,i}$, we let $f(c_{n,i}|x_{n,i}; \beta)$ be specified as the density of a truncated exponential density:

$$\frac{\lambda_{n,i}(\beta) \exp\{-\lambda_{n,i}(\beta) c_{n,i}\}}{1 - \exp\{-\lambda_{n,i}(\beta)\bar{c}\}}$$

where $\bar{c}$ is the sup of the support (i.e., the truncation point).

Next, we give details in the derivation of the likelihood in Section 4. Under the specification of belief distribution in Section 4, the marginal distribution of $\tilde{Y}$ conditional on $W = w$

[16] The covariance between μ_p and μ_d is $Cov(\mu_p, \mu_d) = Cov(1 - \tilde{Y}, \tilde{Y} + Y) = -Var(\tilde{Y}) - Cov(\tilde{Y}, Y)$, which is used to calculate the expression reported in Table C1.

is $Beta(\alpha_1, \alpha_2 + \alpha_3)$, where α_j is shorthand for $\exp\{w\rho_j\}$ for $j = 1, 2, 3$. The conditional distribution of Y given $\tilde{Y} = \tau, W = w$ is $(1-\tau)Beta(\alpha_2, \alpha_3)$. For any y and $\tau \in (0, 1)$, we can write

$$\Pr\{Y \leq y \mid \tilde{Y} = \tau, W = w\} = \Pr\left\{ \tfrac{Y}{1-\tau} \leq \tfrac{y}{1-\tau} \middle| \tilde{Y} = \tau, W = w \right\}$$

where the conditional distribution of $Y/(1-\tau)$ given $\tilde{Y} = \tau$ is $Beta(\alpha_2, \alpha_3)$.

By definition of the log-likelihood in the text,

$$\begin{aligned} p_{n,i}(t; \rho, \beta) &\equiv \Pr(A_{n,i} = 1 | T_n = t, w_{n,i}, x_{n,i}, k_{n,i}; \rho, \beta) = \Pr(YC \leq \phi(t) k_{n,i} | w_{n,i}, x_{n,i}; \rho) \\ &= \int_0^{\bar{c}} \Pr(Y \leq k_{n,i}\phi(t)/c | w_{n,i}; \rho) f_C(c | x_{n,i}; \beta) dc, \end{aligned}$$

and

$$\begin{aligned} g_{0,n,i}(t; \rho, \beta) &\equiv \left. \frac{\partial \Pr(C_{n,i} \leq c, A_{n,i} = 0, D_{n,i} = 1 | T_n = t, x_{n,i}, w_{n,i}, k_{n,i}; \rho, \beta)}{\partial c} \right|_{c = z_{n,i}} \\ &= \left. \tfrac{\partial}{\partial c} \left[\int^{c} q_{n,i} \Pr(Y > \phi(t) k_{n,i}/\tilde{c} | w_{n,i}; \rho) f_C(\tilde{c} | x_{n,i}; \beta) d\tilde{c} \right] \right|_{c = z_{n,i}} \\ &= q_{n,i} \Pr(Y > \phi(t) k_{n,i}/z_{n,i} | w_{n,i}; \rho) f_C(z_{n,i} | x_{n,i}; \beta) \end{aligned}$$

with $f_C(.|x_{n,i}; \beta)$ being the density of C conditional on $x_{n,i}$. Furthermore,

$$\begin{aligned} g_{1,n,i}(t; \rho, \beta) &\equiv \left. \frac{\partial \Pr(S_{n,i} \leq s, A_{n,i} = 1 \mid T_n = t, w_{n,i}, x_{n,i}, k_{n,i}; \rho, \beta)}{\partial s} \right|_{s = z_{n,i}} \\ &= \left. \tfrac{\partial}{\partial s} \left[\int_0^{\bar{c}} \Pr\left(\tilde{Y} \geq 1 - s/(c\delta^t),\ Y \leq \phi(t) k_{n,i}/c \middle| w_{n,i}; \rho \right) f_C(c | x_{n,i}; \beta) dc \right] \right|_{s = z_{n,i}}. \end{aligned}$$

In the derivation above, we have used the conditional independence between $C_{n,i}$ and $D_{n,i}$, T_n, $(\mu_{p,n,i}, \mu_{d,n,i})$ conditional on $w_{n,i}, x_{n,i}$. Under regularity conditions that allow for the change of the order of integration and differentiation, $g_{1,n,i}(t; \rho, \beta)$ is equal to

$$\int_{z_{n,i}/\delta^t}^{\bar{c}} \Pr\left\{ Y \leq \phi(t) k_{n,i}/c \middle| \tilde{Y} = 1 - \tfrac{z_{n,i}}{c\delta^t}, w_{n,i}; \rho \right\} f_{\tilde{Y}} \left(1 - \tfrac{z_{n,i}}{c\delta^t} \middle| w_{n,i}; \rho \right) \tfrac{f_C(c | x_{n,i}; \beta)}{c\delta^t} dc,$$

where the lower limit is $z_{n,i}\delta^{-t}$ because the integrand is non-zero if and only if $1 - z_{n,i}\delta^{-t}/c \in (0, 1)$, or $c > z_{n,i}\delta^{-t}$. Changing variables between c and $\tau \equiv 1 - z_{n,i}\delta^{-t}/c$ for any i, n and t, we can write $g_{1,n,i}(t; \rho, \beta)$ as:

$$\int_0^{1 - \frac{z_{n,i}}{\delta^t \bar{c}}} \Pr\left\{ \frac{Y}{1-\tau} \leq \frac{\delta^t \phi(t) k_{n,i}}{z_{n,i}} \middle| \tilde{Y} = \tau, w_{n,i}; \rho \right\} f_{\tilde{Y}}(\tau | w_{n,i}; \rho) \frac{f_C\left(\frac{z_{n,i}}{\delta^t(1-\tau)} \middle| x_{n,i}; \beta \right)}{\delta^t (1-\tau)} d\tau$$

where $Y/(1-\tau)$ given $\tilde{Y} = \tau$ follows a Beta distribution with parameters (α_2, α_3), and $f_{\tilde{Y}}(.|w_{n,i}; \rho)$ denotes a Beta p.d.f. with parameters $(\alpha_1, \alpha_2 + \alpha_3)$. (Recall α_j are shorthands for $\exp(w_{n,i}\rho_j)$ for $j = 1, 2, 3$.)

C3. Counterfactual prediction

We now provide details about how we calculate the settlement probability and the distribution of accepted settlement offer under counterfactual compensation caps in Section 7. Let x denote state variables that affect the distribution of C (e.g., age, income and severity); let $f(. \mid x)$ denote the density of C conditional on x. Let w denote state variables that affect the distribution of (μ_p, μ_d); let $h(.)$ denote the probability mass function for the wait-time T (which is orthogonal to μ, C).

First off, we calculate the conditional probability for $A = 1$ under a compensation cap $\hat{c} < \bar{c}$. By construction,

$$\Pr\{A = 1 \mid w, x, k\} = \textstyle\sum_t \Pr\{YC \leq \phi(t)k \mid w, x\} h(t) \tag{17}$$

due to independence between K, T and (μ, C). Under a binding cap $\hat{c} < \bar{c}$,

$$\Pr\{YC \leq \phi(t)k \mid w, x\} = \int_0^{\hat{c}} \Pr\left\{Y \leq \tfrac{\phi(t)k}{c} \middle| w\right\} f(c|x)dc + \Pr\{C \geq \hat{c}|x\} \Pr\left\{Y \leq \tfrac{\phi(t)k}{\hat{c}} \middle| w\right\}.$$

Next, we calculate the distribution of $S \mid A = 1$ under a cap $\hat{c} < \bar{c}$. By construction,

$$\Pr\{S \leq s \mid A = 1, w, x, k\} = \tfrac{\Pr\{S \leq s, A=1 \mid w,x,k\}}{\Pr\{A=1 \mid w,x,k\}}, \tag{18}$$

where the denominator on the right-hand side is as described above. The numerator on the right-hand side of (18) is:

$$\textstyle\sum_t \left[\int_0^{\bar{c}} \Pr\left\{\tilde{Y} \geq 1 - \tfrac{s}{c\delta^t}, Y \leq \tfrac{\phi(t)k}{c} \middle| w\right\} f(c \mid x)dc\right] h(t) \tag{19}$$

where $\tilde{Y} \equiv 1 - \mu_p$ and $Y \equiv \mu_p + \mu_d - 1$. With a binding cap $\hat{c} < \bar{c}$, the term in the square brackets in (19) becomes

$$\int_0^{\hat{c}} \Pr\left\{\tilde{Y} \geq 1 - \tfrac{s}{c\delta^t}, Y \leq \tfrac{\phi(t)k}{c} \middle| w\right\} f(c|x)dc + \Pr\left\{\tilde{Y} \geq 1 - \tfrac{s}{\hat{c}\delta^t}, Y \leq \tfrac{\phi(t)k}{\hat{c}} \middle| w\right\} \Pr\{C \geq \hat{c}| x\}.$$

If $\hat{c} \leq \frac{s}{\delta^t}$, then the expression above is simplified to

$$\int_0^{\hat{c}} \Pr\left\{Y \leq \tfrac{\phi(t)k}{c} \middle| w\right\} f(c \mid x)dc + \Pr\left\{Y \leq \tfrac{\phi(t)k}{\hat{c}} \middle| w\right\} \times \Pr\{C \geq \hat{c}| x\}.$$

Given our MSL estimates for the distribution of C and $(\tilde{Y}, Y)$, we calculate the conditional distribution of settlement offers under a cap $\hat{c}$ using these formulas and the simulation-based integration. Inverting this estimated distribution gives our estimates of the quantiles of settlement offers under caps.

REFERENCES

An, Y., Hu, Y. and M. Shum, "Estimating First-Price Auctions with an Unknown Number of Bidders: A Misclassification Approach," *Journal of Econometrics*, vol 157 (2010), 328-341

Cho, I., "Delay and Uncertainty in Bargaining," *Review of Economic Studies*, vol 57 (1990), 575-595.

Da Silveira, B., "Bargaining with Asymmetric Information: An Empirical Study of Plea Negotiations," *Econometrica*, Vol. 85 (2017), 419-452

Galasso, A., "Broad Cross-License Negotiations," *Journal of Economics and Management Strategy*, vol 21 (2012), 873–911

Hall, P. and X. Zhou, "Nonparametric estimation of component distributions in a multivariate mixture," *Annals of Statistics*, vol 31 (2003), 201-224

Hicks, J., *The Theory of Wages*, (London, Macmillan, 1932)

Hu, Y., "Identification and estimation of nonlinear models with misclassification error using instrumental variables: A general solution," *Journal of Econometrics*, vol. 144 (2008), 27-61

Hu, Y., McAdams, D. and M. Shum, "Nonparametric identification of first-price auctions with non-separable unobserved heterogeneity," *Journal of Econometrics*, vol 174 (2013), 186-193.

Hu, Y. and S. Schennach, "Instrumental variable treatment of nonclassical measurement error models," *Econometrica*, vol 76 (2008), 195-216

Kasahara, H. and K. Shimotsu, "Nonparametric identification of finite mixture models of dynamic discrete choices," *Econometrica*, vol 77 (2009), 135-175

Merlo, A. and X. Tang, "Identification and Estimation of Stochastic Sequential Bargaining Models," *Econometrica*, Vol. 80 (2012), 1563-1604

Merlo, A. and C. Wilson, "A Stochastic Model of Sequential Bargaining with Complete Information," *Econometrica*, Vol 63 (1995), 371-399

Shavell, S., "Suit, Settlement, and Trial: A Theoretical Analysis Under Alternative Methods for the Allocation of Legal Costs," *Journal of Legal Studies*, Vol. 11 (1982), 55-81

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

Sloan, F., Githens, P., Clayton, E., Hickson, G., Gentile, D. and D. Partlett., *Suing for Medical Malpractice*, (The University of Chicago Press, 1993)

Thanassoulis J., "Optimal stalling when bargaining," *Journal of Economic Dynamics and Control*, vol 34 (2010), 101–120

Watanabe, Y., "Learning and bargaining in dispute resolution: theory and evidence from medical malpractice litigation," Working Paper, Northwestern University, 2006

Yildiz, M., "Bargaining without a common prior: An immediate agreement theorem," *Econometrica*, vol 71 (2003): 793-811.

Yildiz, M., "Waiting to Persuade," *Quarterly Journal of Economics*, vol 119 (2004): 223-248.

Yildiz, M., "Bargaining with Optimism," *Annual Review of Economics*, Vol 3 (2011), Article 3.