Bargaining with Heterogeneous Beliefs:
A Structural Analysis of Florida Medical Malpractice Lawsuits

Antonio Merlo and Xun Tang [1]
Department of Economics
Rice University
May 16, 2015

Abstract

We propose a structural bargaining model where players hold heterogeneous beliefs about the final resolution if no settlement is reached outside the court. We show the distribution of their beliefs and the stochastic surplus are nonparametrically identified from the probability for reaching an settlement and the distribution of final transfers between players. We then use a Simulated Maximum Likelihood (SML) approach to estimate the beliefs of doctors and patients in medical malpractice lawsuits in Florida in the 1980s and 1990s. We find strong evidence that the beliefs for both parties vary with the severity of the injury and the qualification of the doctors in the lawsuits, even though these characteristics are statistically insignificant in explaining whether the court rules in favor of the plaintiff or the defendant.

Key words: Bargaining with heterogeneous beliefs, nonparametric identification, medical malpractice lawsuits

[1] We thank participants at Brown, OSU and Econometric Society North America Summer Meeting 2014 for useful feedbacks. This research is funded by NSF Grant # xxxxxx. We thank Devin Reily and Michelle Tyler for capable research assistance. Errors (if any) are our own.

1 Introduction

A major theme in recent development of bargaining theory is to rationalize the delay in reaching an agreement that are prevalent in real-world bargaining episodes. (See, for example, Cho (1990), Merlo and Wilson (1995) and Yildiz (2004).) One explanation for the delay is that the parties involved in bargaining are "too optimistic" about their respective bargaining power in the absence of a common prior belief. Specifically, consider a bilateral bargaining episode where players could learn about each other's bargaining power through history of negotiation. In the presence of optimism, a player i will decide to wait in hopes that the other player j will learn about his (i's) self-perceived strong position and agree to his (i's) terms. As time passes, the learning slows down, and it becomes no longer worthwhile to wait for the other parties' learning. That is when they reach an agreement. Yildiz (2004) introduced this model and showed that there is a deterministic settlement date, which is predetermined by the prior and the discount factor of the players, such that players will wait until this date to reach an agreement. Since its introduction, the model of bargaining with optimism has been applied in a wide range of empirical context, such as pretrial negotiations in medical malpractice lawsuits (Watanabe (2006)), negotiation about market conditions (Thanassoulis 2010), and cross-license agreements (Galasso (2006)).

Despite the recent surge in the theory and the application of bargaining with optimism, we are not aware of any existing work which addresses the identification question in such a model formally. That is, under what conditions can the structural elements of the model be unambiguously recovered from the history of bargaining reported in the data. One of the objectives of our paper is to fill in this gap between theory and empirical work by introducing a framework for the structural estimation of bargaining without a common prior. In particular, we propose a model for bilateral bargaining where players have optimism about the stochastic final outcome in case no agreement is reached. The players have a one-time opportunity for reaching an agreement at an exogenously scheduled date during the bargaining process, and make decisions about the settlement based on their beliefs and time discount factors. We show that all structural elements in this model are identified nonparametrically from the probability for reaching an agreement and the distribution of transfer in the final resolution. The identification strategy does not rely on any parametrization of the structural primitives such as players beliefs or the surplus distribution. We then propose a Simulated Maximum Likelihood (SML) estimator based on flexible parametrization of the joint beliefs.

The model we introduce is a simplification of the model of bargaining without a common prior in Yildiz (2004). We model a one-time settlement opportunity for the players to reach an agreement. Consequently there is no dynamic "learning" consideration in players' decisions, and the dates of the final resolution of the bargaining episodes are determined by players' optimism, their patience and their perception of the surplus to be shared.

There are both theoretical and empirical motivation for such simplification and modification. First off, in real-world bargaining episodes, data limitation prevents researchers from deriving robust (parametrization-free) arguments for the identification of structural elements in a full-fledged model of bargaining with uncommon prior. For instance, the identity of proposers and the timing or the size of rejected offers in negotiations are seldom reported in the data available to researchers. By abstracting away from the dynamic learning aspects in Yildiz (2004), we adopt a pragmatic approach to build a model that is identifiable under less stringent data requirements and mild econometric assumptions. Despite this simplification, our model captures a key aspect of models without a common prior in that the timing of the agreement is determined by the players' optimism and their patience. Thus our work provides a benchmark for understanding what additional data or econometric assumptions are needed to recover the primitives in more elaborate models of bargaining without common priors.

Our modeling choices are also motivated by the empirical question addressed in this paper: In medical malpractice lawsuits, what do the settlement decisions by patients and doctors tell us about their respective perception of how likely the court would rule in their favor in case a court hearing is necessary ? The law of the State of Florida requires that there should be a one-time mandatory settlement conference between the plaintiff and the defendant, which is scheduled by the county court sometime prior to the hearing and is mediated by court-designated legal professionals. Besides, unlike Yildiz (2004) where a player's bargaining power is modeled as his chance for being the proposer, we model bargaining power as a player's perception of the probability that he will receive the favored outcome in the final resolution in case no agreement is reached. This modification is meant to be a better approximation of the actual decision environment in the legal context.

Our strategy for identifying this model builds on a couple of insights: First off, if the length of time between the scheduled settlement conference and the court hearing (a.k.a. the "wait-time") were reported in the data, we would be able to recover the distribution of optimism by observing how the conditional settlement probability varies with the length of wait-time. Second, the distribution of the potential surplus to be divided between players can be recovered from the distribution of total compensation awarded to the plaintiff by the court decision, provided the surplus distribution is orthogonal to the beliefs. Third, because the accepted settlement offers reflects a plaintiff's time-discounted expectation of his share of the total surplus, we can identify the distribution of the plaintiff's belief conditional on settlement using the distribution of accepted settlement offers given the length of the wait-time. This is done through a deconvolution argument using the distribution of surplus recovered above. Last, since optimism is defined as the sum of both parties' beliefs minus one, the objects identified from the preceding steps can be used to back out the joint distribution of the beliefs through a standard Jacobian transformation.

A key challenge for implementing the identification strategy in the environment of malpractice lawsuits is that the length of wait-time is not directly reported in the data. In order to solve this issue of unobserved wait-time, we tap into a branch of recent literature that uses an approach based on eigenvalue decomposition to identify finite mixture models or structural models with unobserved heterogeneity. (See for example Hall and Zhou (2003), Hu and Schennach (2008), Kasahara and Shimotsu (2009), An, Hu and Shum (2010) and Hu, McAdams and Shum (2013).) To do so, we first exploit the institutional details in our environment to group lawsuits into smaller clusters (defined by the county and the month in which a lawsuit is officially filed) that can be plausibly assumed to share the same (albeit unobserved) length of wait-time between settlement conference and court hearings. We then use the cases in the same cluster as instruments for each other and apply an eigenvalue decomposition to the joint distribution of settlement decisions and accepted offers within the cluster. This allows us to recover the settlement probability and the distribution of accepted offers conditional on the unobserved wait-time. Then the arguments from the preceding paragraph applies to identify the joint distribution of beliefs.

The inference of doctors' and patients' beliefs in medical malpractice lawsuits is an interesting empirical question in its own right. In particular, a central issue in the reform of U.S. health care system is how to minimize the litigation costs in medical malpractice lawsuits, which are known to constitute a large portion of the soaring insurance expenses. Knowing how settlement outside the court depends on patients' and doctors' optimism in the bargaining process could shed lights on policy design.

Using data from medical malpractice lawsuits in Florida in the 1980s and 1990s, we find clear evidence in our estimates that the beliefs of the doctors and patients vary with observed characteristics of the lawsuits such as the severity of the injury and the qualification of the doctors. This contrasts with the reality that the court and jury decisions depend mostly on the nature and the cause of the malpractice and not so much on these observed case characteristics (which is another fact revealed in our estimates in the application). Our estimates can be used for answering future policy design questions such as how the settlement probability would change if the distribution of the wait-time is changed or some caps on put on the potential compensation possible.

The rest of the paper is organized as follows: Section 2 introduces the model of bilateral bargaining with uncommon beliefs. Section 3 establishes the identification of structural elements in the model. Section 4 defines the Simulated Maximum Likelihood (SML) estimator. Section 5 describes the data and the institutional details in the application of medical malpractice lawsuits in Florida. Section 6 presents and discusses the estimation results. Proofs and a monte carlo study are presented in the appendices.

2 The Model

Consider a lawsuit following an incidence of medical malpractice involving a plaintiff (or patient) and a defendant (or doctor). The total amount of potential compensation C is common knowledge among the plaintiff and the defendant. (It should be interpreted as a sunk cost for the defendant, analogous to the money paid by the defendant for bailout.) After the filing of a lawsuit, the plaintiff and the defendant are notified of a date for a *one-time* settlement conference, which is mandatory by the State Statutes in Florida. The conference requires attendance by both parties (and their attorneys), as well as legal professionals designated by the county court where the lawsuit is filed. Such settlement conferences take place within 120 days after the filing of the lawsuit.[2]

During the conference, the defendant has the opportunity to make an settlement offer of $S \leq C$ to the plaintiff. If the plaintiff accepts it, then the legal process ends with plaintiff receiving S and the defendant reclaiming $C - S$. Otherwise the case needs to go through a court hearing process that culminates in jury decisions. Both the defendant and the plaintiff are aware that the court hearing needs to be at least three weeks later than the settlement conference; and the exact date is determined by the schedule and the backlogs of all judges available at the count court. Let T denote the length of time between the settlement conference and the scheduled court hearing date.

Let $A \equiv 1$ if a settlement is reached at the conference; and $A \equiv 0$ otherwise. In the latter case, at the end of the court hearing process, the jury makes a binary decision D as to whether the plaintiff gets compensated with the full amount C (i.e. $D = 1$) or the defendant is acquitted with no compensations to the plaintiff required (i.e. $D = 0$). The plaintiff and the defendant believe their chances of winning are μ_p and $\mu_d \in [0, 1]$ respectively. These beliefs are common knowledge between the parties, but are not reported in data. The joint support of beliefs is $\Omega_\mu \equiv \{(\mu_p, \mu_d) \in (0, 1]^2 : 1 < \mu_p + \mu_d \leq 2\}$. This means excessive optimism always occurs (i.e. $\mu_p + \mu_d > 1$ with probability 1). We maintain the following assumption throughout the paper.

Assumption 1 *(i) (μ_p, μ_d) and C are independent from the wait time T; and the distributions of (μ_p, μ_d) is continuous with positive densities over Ω_μ. (ii) Conditional on $A = 0$, the jury decision D is orthogonal to C and T.*

Assumption 1 allows plaintiffs' and defendants' beliefs to be correlated with each other and asymmetric with different marginal distributions. This is empirically relevant because the marginal distribution of beliefs may well differ between patients and doctors due to

[2]See Section 108 in Chapter 776 of Florida Statutes 2012. Web link: http:// www. flsenate. gov/ Laws/ Statutes/ 2012/ 766.108

factors such as informational asymmetries (e.g. doctors are better informed about the cause and severity of the malpractice) or unobserved individual heterogeneities.

Beliefs of plaintiffs and defendants are also likely to be correlated through unobserved heterogeneity of the case of malpractice. For example, they may both observe aspects related to severity or cause of the malpractice that are not recorded in data. Such aspects lead to correlations between patients' and plaintiffs' beliefs from an outsider's perspective. Assumption 1 also accommodates correlation between (μ_p, μ_d) and C.

The independence between the wait time T and the beliefs is a plausible condition, because the wait time T is mostly determined by availability of judges and juries in the county court during the lawsuits. This depends on the schedule and backlogs of judges, which are idiosyncratic and orthogonal to parties' beliefs (μ_p, μ_d).

The orthogonality of C from D given T and $A = 0$ in condition (ii) is also justified. On the one hand, C is a monetary measure of the magnitude of the damage inflicted on the plaintiff regardless of its cause; on the other hand, D captures the jury's judgement about the cause of damage based on court hearings. It is likely that the jury decision is correlated with specific features of the lawsuit that are reported in data and that may also affect the beliefs of both parties. Nevertheless, once conditional on such observable features, jury decisions are most likely to be orthogonal to measure of damage captured by C. At the end of this section, we discuss how to extend our model to account for heterogeneities across lawsuits reported in data.

We now summarize how the distributions that are directly identifiable from data are linked to model primitives under the assumption that both parties follow rational strategies. At the settlement conference, the plaintiff accepts an offer if and only if $S \geq \delta^T \mu_p C$, where δ is a constant time discount factor fixed throughout the data-generating process and available in data. The defendant offers the plaintiff $S = \delta^T \mu_p C$ if the remainder of the potential compensation $C - S$ exceeds $\delta^T \mu_d C$. Hence a settlement occurs during the conference if and only if:

$$C - \delta^T \mu_p C \geq \delta^T \mu_d C \;\Leftrightarrow \mu_d + \mu_p \leq \delta^{-T}.$$

The resulted distribution of settlements, conditional on the wait time between the settlement conference and scheduled court hearing being $T = t$, is:

$$\Pr\left(S \leq s \mid A = 1, T = t\right) = \Pr\left(\mu_p C \leq \delta^{-t} s \mid \mu_d + \mu_p \leq \delta^{-t}\right). \tag{1}$$

where lower cases denote realized values for random variables; and the equality follows from part (i) in Assumption 1.

Besides, the distribution of potential compensation, conditional on the absence of settlement in the conference $T = t$ periods ahead of the court hearing and conditional on the jury ruling in favor of the plaintiff, is:

$$\Pr(C \leq c \mid A = 0, D = 1, T = t) = \Pr(C \leq c \mid \mu_d + \mu_p > \delta^{-t}) \tag{2}$$

where the equality follows from both conditions in Assumption 1.

In practice, the data reports differences in the characteristics of plaintiffs and defendants, such as the professional qualification of the defendant or the demographics of the plaintiff. Besides, the data also reports features related to the cause and the severity of malpractice in question. Such information available in data (denoted by a vector X) are correlated with total compensation C and beliefs (μ_p, μ_d).

The simplistic model above can incorporate such observed case heterogeneities by letting the primitives (i.e. distributions of beliefs (μ_p, μ_d), compensations C, jury decisions D and the wait-time T) depend on X. If both restrictions in Assumption 1 hold conditional on X, then rational strategies are characterized in the same way as (1) and (2) except that all distributions needs to be conditioned on X.

More importantly, the identification strategy proposed in Section 3 below are applicable when data reports heterogeneities across lawsuits. Formally, the results in Section 3 (Lemma 1 and Proposition 1) hold after conditioning on X, provided the identifying conditions (Assumptions 2, 3, 4 and 5) are formulated as conditional on X. Nonetheless, in order to simplify exposition of the main idea for identification, we choose to suppress dependence on observable case characteristics in Section 3, and only incorporate them explicitly later in the estimation section.

3 Identification

This section shows how to recover the distribution of both parties' beliefs from the probability for reaching settlements and the distribution of accepted settlement offers. We consider an empirical environment where for each lawsuit the data reports whether a settlement occurs during the mandatory conference (A). For cases settled at the conference, the data reports the amount paid by the defendant to the plaintiff (S). For the other cases that underwent court hearings, the data reports jury decisions (D) and, if the court rules in favor of the plaintiff, the amount of total compensations paid by the defendant (C). However, exact dates of settlement conferences and scheduled dates for court hearings (if necessary) are never reported in data.[3] Thus the wait-time T between settlement conference and scheduled court hearings, which is known to both parties at the conference, is not available in data.

[3] For example, the data we use in Section 5 reports "Dates of Final Disposition" for each case. However, for cases settled outside the court, these dates are defined not as the exact date of the settlement conference, but as the day when all official administrative paperwork are concluded. 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 underwent court hearings in that the reported "dates of final disposition" are not identical to the actual date of court hearings.

To address this issue with unreported wait-time, we propose sequential arguments that exploit an implicit panel structure of the data in the current context. In particular, we note that lawsuits filed with the same county court during the same period (week) practically share the same wait-time T. The reason for such a pattern is as follows: First, the dates for settlement conferences are mostly determined by availability of authorized legal professionals affiliated with the count court, and are assigned on a "first-come, first served" basis. Thus settlement conferences for cases filed with the same county court at the same time are practically scheduled for the period. Besides, the dates for potential court hearing are determined by the schedule and backlog of judges at the county court. Hence cases filed with the same county court simultaneously can be expected to be handled in court in the same period in the future. This allows us to effectively group lawsuits into clusters with the same T, despite unobservability of T in data. We formalize this implicit panel structure as follows.

Assumption 2 *Researchers have sufficient information to divide the data into clusters, each of which consists of at least three lawsuits sharing the same wait-time T. Across the cases within the same cluster, the beliefs (μ_p, μ_d), the total compensation C and the potential jury decision D (if necessary) are independent draws from the same distribution.*

This implicit panel structure in our data allows us to use accepted settlement offers in the lawsuits within the same cluster as instruments for each other, and apply eigen-decomposition-based arguments in Hu and Schennach (2008) to recover the joint settlement probability and distributions of accepted settlement offers conditional on the unobserved T. We then use these quantities to back out the joint distribution of beliefs using exogenous variations in T.

For the rest of this section, we first present arguments for the case where T is discrete (i.e. $|\mathcal{T}| < \infty$). At the end of this section, we explain how to generalize them for identification when T is continuously distributed.

3.1 Conditional distribution of settlement offers

An intermediate step for identifying the joint distribution of beliefs is to recover the conditional settlement probability and the distribution of settlement offers given the wait-time before court hearings T. Let $\mathcal{S}, \mathcal{T}$ denote the unconditional supports of S, T respectively.

Assumption 3 *(i) The support of T is finite ($|\mathcal{T}| < \infty$) with a known cardinality and* $\inf\{\delta^t : t \in \mathcal{T}\} \geq 1/2$*. (ii) Given any (μ_p, μ_d), the potential compensation C is continuously distributed with positive density over a connected support* $[0, \overline{c}]$*.*

We focus on the model with finite $\mathcal{T}$ with known cardinality because of its empirical relevance. Without loss of generality, denote elements in $\mathcal{T}$ by $\{1, 2, ., |\mathcal{T}|\}$. Condition (i) also rules out unlikely cases where a court hearing is scheduled so far in the future or the one-period discount factor is so low that the compounded discount factor is less than one half. Condition (i), together with the non-increasingness of $\mathbb{E}[A_i \mid T = t]$ over $t \in \mathcal{T}$ due to Assumption 1, pin down the index for eigenvalues and eigenvectors in the aforementioned decomposition.

Part (ii) in Assumption 3 is a mild condition on the conditional support of potential compensation. A sufficient condition for this is that C is orthogonal from (μ_p, μ_d) with a bounded support.[4] The role of part (ii) will become clear as we discuss the identification result below.

Lemma 1 *Under Assumptions 1, 2 and 3,* $\mathbb{E}(A \mid T = t)$ *and* $f_S(s \mid A = 1, T = t)$ *are jointly identified for all* t *and* s*.*

This intermediate result uses arguments similar to that in Hu, McAdams and Shum (2013) for identifying first-price sealed-bid auctions with non-separable auction heterogeneities. It exploits the panel structure of the data and the conditional independence of beliefs across lawsuits in Assumption 2. These conditions allow us to break down the joint distribution of the incidence of settlement and the size of accepted settlement offers across multiple lawsuits within one cluster into the composition of three linear operators.

More specifically, let $f_{R_1}(r_1, R_2 = r_2 \mid .)$ be a shorthand for $\frac{\partial}{\partial \tilde{r}} \Pr\{R_1 \leq \tilde{r}, R_2 = r_2 \mid .\}|_{\tilde{r}=r_1}$ for any discrete random vector R_2 and continuous random vector R_1. For any three lawsuits i, j, k sharing the same wait-time T, let $A_{i,k} = 1$ be a shorthand for "$A_i = A_k = 1$". By construction,

$$\begin{aligned}
& f_{S_i,S_k}(s, s', A_j = 1 \mid A_{i,k} = 1) \\
= & \sum_{t \in \mathcal{T}} f_{S_i}(s \mid S_k = s', A_j = 1, T = t, A_{i,k} = 1)\mathbb{E}[A_j \mid S_k = s', T = t, A_{i,k} = 1] f_{T,S_k}(t, s' \mid A_{i,k} = 1) \\
= & \sum_{t \in \mathcal{T}} f_{S_i}(s \mid A_i = 1, T = t)\mathbb{E}[A_j \mid T = t] f_{T,S_k}(t, s' \mid A_{i,k} = 1) \,. \qquad (3)
\end{aligned}$$

The second equalities follow from Assumption 1; from "$S = \delta^T \mu_p C$ whenever $A = 1$" and "$A = 1$ if and only if $\mu_p + \mu_d \leq \delta^{-T}$"; and from the fact that beliefs (μ_p, μ_d) and potential compensation C are independent draws across the lawsuits i, j, k according to Assumption 2.

To illustrate the identification argument, it is useful to adopt matrix notations. Let $\mathcal{D}_M$ denote a partition of the *unconditional* support of accepted settlement offers $\mathcal{S}$ into

[4] It is worth noting that our identification argument remains valid even with $\bar{c}$ being unbounded, as long as the full-rank condition in Lemma A1 holds for some partitions of $\mathcal{S}$.

M intervals. Each of the intervals has a non-degenerate interior and is denoted by d_m.[5] For a given partition $\mathcal{D}_M$, let L_{S_i,S_k} be a M-by-M matrix whose (m, m')-th entry is the probability that $S_i \in d_m$ and $S_k \in d_{m'}$ conditional on $A_{i,k} = 1$ (settlements are reached in cases i and k); and let Λ_{S_i,S_k} be a M-by-M matrix with its (m, m')-th entry being $f(S_i \in d_m, A_j = 1, S_k \in d_{m'} \mid A_{i,k} = 1)$. Note that both Λ_{S_i,S_k} and L_{S_i,S_k} are directly identifiable from data. Thus a discretized version of (3) is :

$$\Lambda_{S_i,S_k} = L_{S_i|T}\ \Delta_j\ L_{T,S_k} \tag{4}$$

where $L_{S_i|T}$ be a M-by-$|\mathcal{T}|$ matrix with (m, t)-th entry being $\Pr(S_i \in d_m \mid A_i = 1, T = t)$; Δ_j be a $|\mathcal{T}|$-by-$|\mathcal{T}|$ diagonal matrix with diagonal entries being $[\mathbb{E}(A_j \mid T = t)]_{t=1,.,|\mathcal{T}|}$; and L_{T,S_k} be a $|\mathcal{T}|$-by-M matrices with its (t, m)-th entry being $\Pr(T = t, S_k \in d_m \mid A_{i,k} = 1)$. Besides,

$$L_{S_i,S_k} = L_{S_i|T}\ L_{T,S_k} \tag{5}$$

due to conditional independence in Assumption 2.

Part (ii) in Assumption 3 implies the supreme of the conditional support of accepted offers given $T = t$ is $\delta^t \bar{c}$ and hence decreases in t. This, in turn, guarantees there exists a partition $\mathcal{D}_{|\mathcal{T}|}$ such that $L_{S_i|T}$ as well as L_{S_i,S_k} are non-singular (as proved in Lemma A1 in Appendix B). Then (4) and (5) imply

$$\Lambda_{S_i,S_k}\ (L_{S_i,S_k})^{-1} = L_{S_i|T}\ \Delta_j\ \left(L_{S_i|T}\right)^{-1} \tag{6}$$

where the L.H.S. consists of directly identifiable quantities. The R.H.S. of (6) takes the form of an eigen-decomposition of a square matrix, which is unique up to a scale normalization and unknown indexing of the columns in $L_{S_i|T}$ and diagonal entries in Δ_j (i.e. it remains to pin down a specific value of $t \in \mathcal{T}$ for each diagonal entry in Δ_j).

The scale in the eigen-decomposition is implicitly fixed because the eigenvectors in $L_{S_i|T}$ are conditional distributions and needs to sum up to one. The question of unknown indices is solved because in our model $\mathbb{E}[A_j \mid T = t]$ is monotonically decreasing in t over $\mathcal{T}$ provided the parties follow rational strategies described in Section 2. This is again due to the independence between timing and the beliefs in Assumption 1 and the moderate compounded discounting in Assumption 3. This establishes the identification of Δ_j and $L_{S_i|T}$, which are used for recovering L_{T,S_k} and then the conditional density of accepted settlement offers over its full support (see proof of Lemma 1 in Appendix B).

3.2 The joint belief distribution

We now explain how to identify the joint distribution of beliefs (μ_p, μ_d) from the quantities recovered from Lemma 1 under the following orthogonality condition.

[5] That is, $d_m \equiv [s_m, s_{m+1}]$ for $1 \leq m \leq M$, with $(s_m : 2 \leq m \leq M)$ being a vector of ordered endpoints on $\mathcal{S}$ such that $s_1 < s_2 < .. < s_M < s_{M+1}$ and $s_1 \equiv \inf \mathcal{S}$, $s_{M+1} \equiv \sup \mathcal{S}$.

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

This condition requires the magnitude of potential compensation to be independent from plaintiff 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. It then follows from (2) that the distribution of C is directly identified as:

$$\Pr(C \leq c) = \Pr(C \leq c \mid A = 0, D = 1). \tag{7}$$

Let $\mathcal{S}_t \equiv [0, \bar{c}\delta^t]$ denote the conditional support of accepted settlement offers $S = \delta^T \mu_p C$ given $A = 1$ and $T = t$;[6] and let $\varphi_t(s)$ denote the probability that a settlement is reached when the length of wait-time between the settlement conference and the date for court hearing is t and that the accepted settlement offer is no greater than s. That is, for all (s, t),

$$\varphi_t(s) \equiv \Pr\left(S \leq s, A = 1 \mid T = t\right) = \Pr\left(\mu_p C \leq s/\delta^t, \mu_d + \mu_p \leq 1/\delta^t\right) \tag{8}$$

where the equality is due to Assumption 1. The non-negativity of C and (μ_p, μ_d) and an application of the law of total probability on the right-hand side of (8) implies:

$$\varphi_t(s) = \int_0^{\bar{c}} \Pr\left(\tfrac{1}{\mu_p} \geq \tfrac{c}{s}\delta^t,\ \tfrac{1}{\mu_d+\mu_p} \geq \delta^t\right) f(c)dc = \int_0^{\bar{c}} h_t\left(c/s\right) f_C(c)dc \tag{9}$$

where $f_C(c)$ is the density of C and $h_t(v) \equiv \Pr\{\mu_p^{-1} \geq v\delta^t,\ (\mu_d + \mu_p)^{-1} \geq \delta^t\}$; and the first equality is due to orthogonality between C and (μ_p, μ_d).

Changing variables between C and $V \equiv C/S$ for any fixed t and s, we can write (9) as:

$$\varphi_t(s) = \int_0^{\infty} h_t(v)\kappa(v, s)dv \tag{10}$$

where $\kappa(v, s) \equiv s f_C(vs) 1\{v \leq \bar{c}/s\}$. With the distribution (and hence density) of C recovered from (7), the kernel function $\kappa(v, s)$ is considered known for all (v, s) hereinafter for identification purposes. Also note for any $s > 0$, $\kappa(., s)$ is a well-defined conditional density with support $[0, \bar{c}/s]$.[7] Let $F_{V|A=1,T=t}$ denote the distribution of V given $T = t$ and $A = 1$ (or equivalently $\mu_d + \mu_p \leq \delta^{-T}$), whose support is denoted as $\mathcal{V}_t$.

Assumption 5 *For any t and $g(.)$ such that $\mathbb{E}[g(V) \mid A = 1, T = t] < \infty$, the statement "$\int_0^\infty g(v)\kappa(v, s) = 0$ for all $s \in \mathcal{S}_t$" implies the statement "$g(v) = 0$ a.e. $F_{V|A=1,T=t}$".*

[6] In general, we could also allow supports $\otimes_{S,T}$ and $\mathcal{S}_t$ to depend on observed heterogeneities of lawsuits as well. Nonetheless, throughout this section, we refrain from such generalization in order to simplify exposition.

[7] This is because $\kappa(v, s) > 0$ for any $v \geq 0$, $s > 0$. Besides $\int_0^\infty \kappa(v, s)dv = \int_0^{\bar{c}/s} s f_C(vs)dv = 1$ for any s.

This condition, known as the "completeness" of kernels in integral operators, was introduced in Lehmann (1986) and used in Newey and Powell (2003) for identification of nonparametric regressions with instrumental variables. Andrews (2011) and Hu and Shiu (2012) derived sufficient conditions for various versions of such completeness conditions when $g(.)$ is restricted to belong to difference classes. This condition is analogous to a "full-rank" condition on κ if the conditional supports of S and V were finite.[8]

Proposition 1 *Under Assumptions 1-5,* $\Pr(\mu_p \leq \mu \ , \ \mu_p + \mu_d \leq \delta^{-t})$ *is identified for all* $\mu \in (0, 1]$ *and* $t \in \mathcal{T}$*.*

For the rest of this section, we discuss how to generalize results above when $\mathcal{T}$ is infinite (T is continuously distributed over a known interval). First off, the key idea of using eigen-decompositions in Section 3.1 remains applicable, except that $L_{S_i|T}$ and L_{T,S_k} become linear integral operators, and their invertibility needs to be stated as an assumption as opposed to being derived from restrictions on model primitives and implications of rational strategies (as is the case when T is discrete).

Under the support condition that $\inf\{\delta^t : t \in \mathcal{T}\} \geq 1/2$, the eigenvalues in the decomposition $\mathbb{E}[A_j \mid T = t]$ remains strictly monotonic over the interval support $\mathcal{T}$ when T is continuously distributed. On the other hand, the argument that uses monotonicity of the eigenvalues over a finite support $\mathcal{T}$ to index them is no longer applicable when T is continuously distributed. However, rational strategies in our model imply the supremum of the support of accepted settlement offers given $T = t$ must be $\delta^t \bar{c}$. With the supremum of the support of compensations $\bar{c}$ identified and known, this means t can be expressed through a known functional of the eigenvectors $f_{S_i}(. \mid A_i = 1, T = t)$ in the eigen-decomposition identified in the first step. Thus the issue with indexing eigenvalues is also solved.

The remaining step of identifying the joint distribution of $(\mu_p, \mu_p + \mu_d)$ from $f_S(. \mid A = 1, T = t)$ and $\mathbb{E}[A \mid T = t]$ follow from the same argument above. It is worthy of note that an additional step based on Jacobian transformation leads to identification of the joint distribution of (μ_p, μ_d) when T is continuously distributed.

[8] If the support of potential compensation is unbounded, there are plenty of examples of parametric families of densities that satisfy the completeness conditions. For example, suppose potential compensations follow a Gamma distribution with parameters $\alpha, \beta > 0$. That is, $f_C(t) = \frac{\beta^\alpha}{\Gamma(\alpha)} t^{\alpha-1} \exp\{-t\beta\}$. Then, with $s > 0$, the kernel $\kappa(v, s) \equiv s f_C(vs) = \frac{[s\beta]^\alpha}{\Gamma(\alpha)} v^{\alpha-1} \exp\{-v\,(s\beta)\}$ is a density of a Gamma distribution with a shape parameter $\alpha > 0$ and a scale parameter $s\beta > 0$. That is, $\kappa(v, s)$ remains a conditional density within the exponential family, and satisfies the sufficient conditions for the completeness condition in Theorem 2.2 in Newey and Powell (2003).

4 Simulated Maximum Likelihood Estimation

Our identification results in Section 3 lay the foundation for nonparametric estimation of the belief distribution. However, a nonparametric estimator based on those arguments would require a large data set, and the "curse of dimensionality" aggravates if the data also report case-level variables that may affect both parties' beliefs (such as the severity of injury inflicted upon the plaintiff and the qualification of the defendant) and therefore should be conditioned on in estimation. To deal with case-heterogeneities in moderate-sized data, we propose in this section a Maximum Simulated Likelihood estimator based on a flexible parametrization of the joint belief distribution.

Consider a panel-structure data contain 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. Define $Z_{n,i} \equiv S_{n,i}$ 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 decisions. We propose a Maximum Simulated Likelihood estimator for the joint beliefs (μ_p, μ_d) that also exploits variation in the heterogeneity of lawsuits reported in the data. Throughout this section, we assume the identifying conditions also hold once conditional on such observed heterogeneity of the lawsuits.

Let $x_{n,i}$ denote the vector of case-level variables reported in the data that affects the distribution of C. (We allow $x_{n,i}$ to contain a constant in the estimation below.) These include the age, severity, and county average/median income (as well as their interaction terms). The total potential compensation C in a lawsuit with observed features $x_{n,i}$ is drawn from an exponential distribution with the rate parameter given by:

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

for some unknown constant vector of parameters β. In the first step, we pool all observations where the jury is observed rule in favor of the plaintiff to estimate β:

$$\hat{\beta} \equiv \arg\max_{\beta} \textstyle\sum_{n,i} d_{n,i}(1 - a_{n,i}) \left[x_{n,i}\beta - \exp\{x_{n,i}\beta\}c_{n,i}\right].$$

Next, let $w_{n,i}$ denote the vector of case-level variables in the data that affects the joint belief distribution. (The two vectors $x_{n,i}$ and $w_{n,i}$ are allowed to have overlapping elements.) We suppress the subscripts n, i for simplicity when there is no confusion. In the second step, we estimate the belief distribution conditional on such a vector of case-level variables W using $\hat{\beta}$ above as an input in the likelihood. To do so, we adopt a flexible parametrization of the joint distribution of (μ_p, μ_d) conditional on W as follows. For each realized w, let $(Y_1, Y_2, 1 - Y_1 - Y_2)$ be drawn from a Dirichlet distribution with concentration parameters $\alpha_j \equiv \exp\{w\rho_j\}$ for $j = 1, 2, 3$ for some constant vector $\rho \equiv (\rho_1, \rho_2, \rho_3)$. In what follows, we suppress the dependence of α_j on w to simplify the notation.

Let $\mu_p = 1 - Y_1$ and $\mu_d = Y_1 + Y_2$. 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 with optimism. (Table C1 and Figure C1 in Appendix C show 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 $Y_2 = \mu_p + \mu_d - 1$ by construction, so it can be interpreted as a measure of optimism. Under this specification, the marginal distribution of Y_1 conditional on $W = w$ is $Beta(\alpha_1, \alpha_2 + \alpha_3)$, where of course α_j's are functions of w. The conditional distribution $Y_2 \mid Y_1 = \tau, W = w$ is the same as the distribution of $(1-\tau)Beta(\alpha_2, \alpha_3)$. For any y and $\tau \in (0,1)$, we can write:

$$\Pr\{Y_2 \leq y \mid Y_1 = \tau, W = w\} = \Pr\left\{ \frac{Y_2}{1-\tau} \leq \frac{y}{1-\tau} \middle| Y_1 = \tau, W = w \right\}$$

where the right-hand side is the c.d.f. of a $Beta(\alpha_2, \alpha_3)$ evaluated at $y/(1-\tau)$.

Let $q_{n,i} \equiv \Pr(D_{n,i} = 1 \mid A_{n,i} = 0, W_{n,i} = w_{n,i})$. Recall that we maintain D is orthogonal to (T, C) once conditional on $A = 0$ and W. Hence $q_{n,i}$ does not depend on $c_{n,i}$. This conditional probability is directly identifiable from the data. Let $h_n(t;\theta)$ denote density of the wait-time T_n at $T_n = t$ in cluster n. This density 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:

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

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

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

where

$$\begin{aligned} p_{n,i}(t;\rho) &\equiv \Pr(A_{n,i} = 1 \mid T_n = t, W_{n,i} = w_{n,i};\rho) = \Pr(\mu_{p,n,i} + \mu_{d,n,i} \leq \delta^{-t} \mid w_{n,i};\rho) \\ &= \Pr(Y_2 \leq \left(1 - \delta^t\right)/\delta^t \mid w_{n,i};\rho); \end{aligned}$$

$$g_{0,n,i}(t;\beta) \equiv g_0(z_{n,i}, x_{n,i}, t;\beta) \equiv \left.\tfrac{\partial \Pr(C_{n,i} \leq Z \mid A_{n,i}=0, T_n=t, X_{n,i}=x_{n,i};\beta)}{\partial Z}\right|_{Z=z_{n,i}} = f_C(z_{n,i} \mid x_{n,i};\beta), \quad (11)$$

with $f_C(. \mid x_{n,i};\beta)$ being the conditional density of the potential compensation given $X_{n,i} = x_{n,i}$; and

$$\begin{aligned} g_{1,n,i}(t;\beta,\rho) &\equiv g_1(z_{n,i}, w_{n,i}, x_{n,i}, t;\beta,\rho) \equiv \left.\tfrac{\partial \Pr(S_{n,i} \leq Z, A_{n,i}=1 \mid T_n=t, w_{n,i}, x_{n,i};\beta,\rho)}{\partial Z}\right|_{Z=z_{n,i}} \\ &= \left.\frac{\partial}{\partial Z} \int_0^\infty \Pr\left(Y_1 \geq 1 - Z/(c\delta^t),\ Y_2 \leq \tfrac{1-\delta^t}{\delta^t} \middle| w_{n,i};\rho \right) f_C(c \mid x_{n,i};\beta) dc \right|_{Z=z_{n,i}} \quad (12) \end{aligned}$$

In the derivations 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 mild regularity conditions, the order of integration and differentiation in (12) can be exchanged. That is, $g_{1,n,i}(t;\beta,\rho)$ equals:

$$\int_{z_{n,i}\delta^{-t}}^{\infty} \Pr\left\{ Y_2 \leq \tfrac{1-\delta^t}{\delta^t} \middle| Y_1 = 1 - z_{n,i}\delta^{-t}/c, w_{n,i}; \rho \right\} f_{Y_1}(1 - z_{n,i}\delta^{-t}/c \mid w_{n,i}; \rho) \frac{f_C(c \mid x_{n,i}; \beta)}{c\delta^t} dc$$

where the lower limit is $z_{n,i}\delta^{-t}$ because the integrand is nonzero only when $1 - z_{n,i}\delta^{-t}/c \in (0,1) \Leftrightarrow c \in (\frac{z_{n,i}}{\delta^t}, +\infty)$.Changing variables between c and $\tau \equiv 1 - z_{n,i}\delta^{-t}/c$ for any i,n and fixed t, we can write $g_{1,n,i}(t;\beta,\rho)$ as:

$$\int_0^1 \Pr\left\{ \frac{Y_2}{1-\tau} \leq \tfrac{1-\delta^t}{\delta^t(1-\tau)} \middle| Y_1 = \tau, w_{n,i}; \rho \right\} \frac{f_{Y_1}(\tau \mid w_{n,i}; \rho) f_C\left(\frac{z_{n,i}}{\delta^t(1-\tau)} \mid x_{n,i}; \beta \right)}{\delta^t(1-\tau)} d\tau$$

where the first conditional probability in the integrand is a Beta c.d.f. evaluated at $\frac{1-\delta^t}{\delta^t(1-\tau)}$ and parameters $(\alpha_2(w_{n,i};\rho_2), \alpha_3(w_{n,i};\rho_3))$ and the second term $f_{Y_1}(\tau \mid w_{n,i};\rho)$ is the Beta p.d.f. with parameters $(\alpha_1(w_{n,i}), \alpha_2(w_{n,i}) + \alpha_3(w_{n,i}))$. For each n, i, t and a fixed vector of parameters (λ, ρ), let $\hat{g}_{1,n,i}(t;\lambda,\rho)$ be an estimator for $g_{1,n,i}(t;\lambda,\rho)$ using $S > N$ simulated draws of τ. (We experiment with various forms of density for simulated draws.) 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 for the belief parameters ρ in the second step is

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

where $\hat{L}_N(\rho,\theta,\beta)$ is an estimator for $L_N(\rho,\theta,\beta)$ by replacing $g_{1,n,i}(t;\beta,\rho)$ with $\hat{g}_{1,n,i}(t;\beta,\rho)$ and replacing $q_{n,i}$ with a parametric (logit or probit) estimate $\hat{q}_{n,i}$; and $\hat{\beta}$ is the estimates for the parameters in the distribution of potential compensation in the first step.

Under regularity conditions, $(\hat{\rho}, \hat{\theta})$ converge at a $\sqrt{N}$-rate to a zero-mean multivariate normal distribution with some finite covariance as long as $N \to \infty$, $S \to \infty$ and $\sqrt{N}/S \to \infty$. 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 medical malpractice insurers to file reports on their resolved claims to the Florida Department of Financial Services. Using this source, we construct a sample that consists of 13,351 lawsuits filed in Florida between 1984 and 1999. Sieg (2000) and Watanabe (2009) also used the same source of data. Our sample includes the cases that are either resolved through the mandatory settlement conference or by a jury decision that followed the court proceedings. For each lawsuit, the data reports the date when it is filed (*Suit_Date*) and the county court with which it is filed (*County_Code*), the date of the final disposition (*Year_of_Disp*) (when the claim was closed with the insurer), and whether the case is resolved through a settlement conference or by a jury decision in court ($A=1$ if settled outside the court). The data also reports the size of the transfer from the defendant to the plaintiff upon the resolution of the lawsuit. This equals the amount of accepted offer to the plaintiff (S) if a settlement is reached outside the court, or the total compensation awarded to the plaintiff according to the court decision (C) otherwise. In addition we also observe case-level variables that may be relevant to the distribution of the joint belief or that of the potential compensation. These include the severity of the injury due to negligence (*Severity*), the age (*Age*) and gender of patients and whether the doctors responsible are board-certified (*Board_Code*). (That is, Board_Code = 1 if the doctor is certified by at least one professional board and = 0 otherwise.)

For the lawsuits settled outside the court, the dates for settlement conferences are not reported in the data. The scheduled dates for court hearings are not reported for the cases resolved by court decisions either. Furthermore, the recorded dates for the final disposition only reveal when the claim is closed with the insurer, which are typically later than the actual dates when an agreement is reached in a settlement conference or when a decision is made by the judge in the court. Therefore, the lengths of the time between scheduled court hearings and the settlement conferences are not directly measured in the data.

Despite these data limitations, we define clusters within which the cases could be reasonably assumed to share the same length of wait-time. It is plausible that the lawsuits filed with the same county court in the same month would be scheduled for court proceedings in the same month. This is of course because the schedule for hearings in a county court is mostly determined by the backlog of unresolved cases filed with that court, and by the availability of judges and other legal professional from the court. By the same token, the schedule for settlement conferences, which require the presence of court officials who have authority to coordinate a settlement, are also mostly determined by the backlog cases as well as the availability of attorney representing both parties. Due to these empirical considerations, we maintain that the wait-time between settlement conferences and court hearings are identical for the cases filed with the same county in the same months. As explained in Section 3, the distribution of settlement decisions and accepted offers in lawsuits from these clusters are sufficient for recovering the joint beliefs of plaintiffs and defendants.

The data consists of 3,545 clusters defined by month-county pairs. In total there are 1,344 clusters which report at least three medical malpractice lawsuits. About half of these clusters (661 clusters) contain at least six cases. Besides, among these 1,344 clusters, 1,294 have at least two lawsuits that were settled outside the court due to the mandatory conference.

These numbers confirm that we can apply our identification strategy from Section 3 to recover the joint distribution of patients' and doctors' beliefs. It is worth noting that in our SML estimation the likelihood includes all 3,545 clusters to improve the efficiency of the estimator, even though in theory identification only requires the joint distribution of settlement decisions and accepted offers from the subset of clusters that have at least two settlements out of three or more cases.

Table 1(a): Settlement probability and accepted offers

Board Cert'n	Severity	# obs	$\hat{p}_{settle}$	$s.e.(\hat{p}_{settle})$	$\hat{\mu}_{S\|A=1}$ (\$1k)	$s.e.(\hat{\mu}_{S\|A=1})$ (\$1k)
certified	low	1,129	0.717	0.013	42.553	2.313
	medium	2,996	0.805	0.007	158.521	4.234
	high	2,623	0.831	0.007	263.089	7.213
uncertified	low	1,783	0.812	0.009	41.300	2.256
	medium	2,192	0.863	0.007	128.796	4.748
	high	2,628	0.880	0.006	333.161	12.357

Next, we report some evidence from the data that the belief of the plaintiffs and the defendants are affected by certain observed characteristics in the lawsuits. Table 1(a) summarizes the settlement probability and the average size of accepted offers in the sample after controlling for the doctors' qualification and the level of severity. There is evidence that both the settlement probability and the size of accepted settlement offers differ systematically across the sub-groups. Table 1(b) reports the p-values of two-sided t-tests (using the unequal variance formula) for the equality of settlement probabilities in sub-groups. We let (u,c) and (l,m,h) be shorthand for the realized values of (*uncertified, certified*) in *Board_ code* and (*low, medium, high*) in *Severity* respectively. With the exception of three pair-wise tests, the nulls in the other tests are all rejected at the 1% significance level. Among the three exceptions, the null for equal settlement probability between (u,l) and (u,h) is also rejected at the 10% level. The only two cases where the null can not be rejected even at the 10% significance level are "(u,l) versus (c,m)" and "(u,l) versus (c,h)". This is somewhat consistent with the intuition that a plaintiff may tend to be more optimistic that the jury would rule in his favor when the injury inflicted is more severe, or when the doctor's qualification is not supported by board certification. Our estimates in the next section are also consistent with this intuition.

The failure to reject the null of equal settlement probability between the two subgroups (u,l) and (c,h) for example may be due to the fact that the impacts of severity and of board certification on the plaintiff's belief offset each other. Pairwise t-tests for the equality of average accepted settlement offers between the sub-groups defined severity and doctor qualification also demonstrate similar patterns. Specifically, the null of equal average settlement offers is almost always rejected at the 1% significance level for all pair-wise t-tests using unequal variances, with the only exception being the tests comparing (u,l) versus (c,l).

Table 1(b): p-values for t-tests: settlement probability

	u,l	u,m	u,h	c,l	c,m	c,h
u,l	–	< 0.001	< 0.001	< 0.001	0.601	0.105
u,m		–	0.072	< 0.001	< 0.001	0.002
u,h			–	< 0.001	< 0.001	< 0.001
c,l				–	< 0.001	< 0.001
c,m					–	0.014
c,h						–

The data also contain some evidence that the distribution of total compensation may be partly determined by the age of the plaintiff and the severity of the injury. Out of the total 2,298 lawsuits which were not resolved through settlement, 359 were ruled in favor of the plaintiff by the court. The observations of the realized total compensation in these cases are useful for inference of the distribution of C. Figure 1(a) and 1(b) in Appendix A report histograms of the accepted offers (S) from the cases settled outside the court and the total compensation (C) from the cases where the court ruled in favor of the plaintiff, conditioning on the information about the plaintiffs. The variable *Age* is discretized into three categories: young ($Age < 33$), older ($Age > 54$) and middle, with the cutoffs being the 33rd and the 66th percentiles in the data. Figure 1(a) suggests the younger plaintiffs tend to receive higher transfers either through accepted offers in settlement or through the total compensation paid by the defendant when the court rules in favor of the plaintiff. Figure 1(b) shows the cases with more severe injuries in general are associated with higher transfers. Both patterns are intuitive, and consistent with our estimates in the next section.

To further compare the distribution of the accepted offers with that of the total compensation ruled by the court, we compare the percentiles of both variables conditional on *Age* and *Severity*. We find that the 10th, 25th, 50th, 75th and 90th conditional percentiles of the accepted offers are consistently lower than those of the total compensation ruled by the court. This is consistent with the notion that the accepted settlement offers are the discounted expectation of the total compensation to be ruled by the court.

The qualification of the doctors does not seem to have any noticeable effect on the distribution of the total compensation. Figure 1(c) reports the histogram of the total compensation for the cases where the court ruled in favor of the plaintiff, conditioning on the board certification of the doctors. A t-test for the equality of the average compensation for the two subgroups with and without board certification reports an asymptotic p-value of 0.5036 (assuming unequal population variance). Besides, a one-sided Komolgorov-Simirnov test against the alternative that the distribution of C is stochastically lower when the defendant is board-certified yields a test statistic of 0.0705 and an asymptotic p-value of 0.4078. Thus in either test the null can not be rejected even at the 15% significance level.

On the other hand, it is reasonable to postulate that the total potential compensation in a malpractice lawsuit is positively correlated with the contemporary income level in the county where the lawsuit is filed. In order to control for such an income effect, we collect data on household income in all counties in Florida between 1981 and 1999. We first collect the data on the median household income in each Florida county in 1989, '93, '95, '97, '98 and '99 from the Small Area Income and Poverty Estimates (*SAIPE*) produced by the U.S. Census Bureau.[9] We also collect a time series of state-wide median household income in Florida each year between 1984 and 1999 from U.S. Census Bureau's the Current Population Survey. We then 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. [10] We then incorporate this yearly data on household income in each county while estimating the distribution of total compensation next year.

6 Estimation Results

As the first step in estimation, we use a logit regression to fit the court decisions in those lawsuits that are resolved through court hearings. The goal is to provide some evidence about whether the jury decisions were affected by case characteristics reported in the data. Besides, the predicted probability for $D = 1$ (the jury ruled in favor of the plaintiff) from the logit regression will be used in the SML estimation of the joint beliefs of doctors and patients.

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

[10] 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.

Table 2. Logit Estimates for Court Decisions[11]

(Response Variable: D. Total # of observations: 2,289 cases.)

	(1)	(2)	(3)
Board_ Code	-0.0701 (0.120)	-0.207 (0.282)	-0.233 (0.288)
Severity.	0.045** (0.023)	0.032 (0.033)	0.083 (0.056)
Age	0.003 (0.003)	0.003 (0.003)	0.021* (0.012)
Severity×*Board_ Code*		0.025 (0.046)	0.029 (0.046)
Age^2			-0.014 (0.011)
Severity×*Age*			-0.012 (0.011)
Constant	-2.042*** (0.189)	-1.954*** (0.228)	-2.393*** (0.391)
Log likelihood	-953.5592	-953.4136	-952.2051
Pseudo-R^2	0.0026	0.0028	0.0041
p-value for L.R.T.	0.1676	0.2533	0.2557

Notes: Standard errors are reported in parentheses. (*** significant at 1%; ** sig. at 5%; * sig. at 10%).Age^2 is reported in units of “100 yr^2”.

Table 2 reports the logit regression estimates under different specifications, using 2,289 lawsuits from the data that were not settled outside the court and thus had to be resolved through scheduled court hearings. The case heterogeneity used in the logit regressions include *Board_ Code*, *Severity* and the age of the patients *Age*. In all three logit regressions, the constant term is highly statistically significant at the 1% level. The severity is statistically insignificant in the latter two specifications. Besides, the age of the patient is only significant at 10% level in the third specification. The board certification of doctors and the interaction terms in the logit regressions are all insignificant.

The pseudo R-squares are low for all three specifications. This suggests that the patient and case characteristics considered are rather insignificant in explaining the court decisions. Furthermore the p-values for the likelihood ratio tests of the joint significance of all slope coefficients are 0.1676, 0.2533 and 0.2557 in the three specifications respectively. Therefore we conclude from Table 2 that the doctor's board certification, the severity of the malpractice and the age of the plaintiff do not have significant impact on jury decisions in the court.

Next, we estimate the distribution of total potential compensation using a subset of the observations of lawsuits above where the court ruled in favor of the plaintiffs ($A = 0$ and $D = 1$). The descriptive statistics in Section 5 show that the severity of the injury and the age of the plaintiffs have a noticeable impact on the size of the total potential compensation, while the doctors' board certification does not. In one of the specifications, we include the

[11] Standard errors are reported in the parenthesis. “*B.C.*” is shorthand for “*Board_ Code*”, and “*Sev.*” for “*Severity*”. The variable Age^2 is reported in units of “100 yr^2”.

county-level income in the same year when the lawsuit is filed in order to allow for the effect of income level. We adopt an exponential specification where the density of the compensation given case characteristics $x_{n,i}$ is $\lambda_{n,i}\exp\{-\lambda_{n,i}c\}$, where $\lambda_{n,i} \equiv \exp\{x_{n,i}\beta\}$.

Table 3(a). Coefficient Estimates in the Distribution of Total Compensation

(Total # of observations: 351 cases.)

	(1)	(2)
M.Sev.	-1.198*** (0.157)	-1.014*** (0.392)
H.Sev.	-1.855*** (0.159)	-0.707** (0.348)
M.Sev.×Age		-0.003 (0.009)
H.Sev.×Age		-0.023*** (0.008)
Age	0.004 (0.003)	0.016** (0.007)
Income (in $1k)	-0.026*(0.014)	-0.031** (0.014)
Constant	-3.810*** (0.506)	-4.288*** (0.554)
Log likelihood	-1179.27	-1167.48
Pseudo-R^2	0.513	0.541
p-value for LRT	<0.001	<0.001

Notes: *** significant at 1%; ** significant at 5%; * significant at 10%.

Table 3(a) presents the maximum likelihood estimates for β. We calculate the standard errors using a standard robust form that consists of estimates for the Hessian and cross-products of the Jacobian of the likelihood. In both specifications, likelihood ratio tests for the joint significance of all slope coefficients yield p-values that are lower than 0.001, thus rejecting the null of joint insignificance even at the 1% level. In both specifications, signs of the estimates show that more severe injuries lead to higher total potential compensation on average; and the expectation of such compensation is higher when the local median household income in the county is higher. (Note that under our specification the conditional mean of the compensation is $1/\lambda_{n,i} = \exp\{-x_{n,i}\beta\}$, so a negative coefficient implies a positive effect on the conditional mean.)

The patient's age also has a significant impact on the total compensation awarded in the second (richer) specification. A likelihood ratio test for the joint significance of all three age-related coefficients yields a p-value of 0.002, thus offering strong evidence that age matters in determining the size of compensation. Furthermore, it is worth mentioning that the sign of the estimated age effects vary across malpractices with different severity. For cases with low and medium severity, a patient's age is estimated to have a negative impact on the conditional mean of compensation (because $0.016 > 0$ and $0.016 - 0.003 > 0$); for case with high severity, the estimated effect is positive ($0.016 - 0.023 < 0$). This pattern may be due to the nature of the interaction between the patient's age and the damage inflicted by malpractice. When the damage is moderate (severity="low" or "medium"), the compensation that is required to keep the patients' life-quality is likely to be proportional to the patient's remaining life span. On the other hand, the seniority in age may aggravate severe damage (severity="high") much more than it does moderate damage. Thus costs for maintaining the life-quality (or even sustaining the life) of the patient when the severity is high could increase drastically with the patient's age.

To better understand the magnitude of age, severity and income effects in terms of monetary units, we report estimates for the average marginal effects (AME) of severity, income and age on the total potential compensation in Table 3(b). The standard errors are calculated using a bootstrap resampling method. While reporting the AME of *Income* and *Age*, we condition on the severity of the malpractice because the latter is shown in Table 3(a) to be statistically significant in determining compensation.

Table 3(b). Average Marginal Effects on Potential Compensation
(units: one thousand US $)

	Estimates	90% C.I.
Severity		
low to medium	237.108	$[39.149, 342.710]$
medium to high	228.637	$[177.517, 626.792]$
Income		
low sev.	3.046	$[1.203, 14.433]$
medium sev.	10.780	$[3.640, 57.358]$
high sev.	20.240	$[7.331, 254.106]$

The average marginal effect on the total potential compensation is estimated to be $\$237,108$ when the severity increases from low to medium, and $\$228,637$ when from medium to high. In both cases, the 90% confidence interval does not include zero, suggests such increases in mean compensation are statistically significant. A $\$1k$-increment in the local income raises the average compensation by $\$3,046$ when the severity of malpractice is low; and raises it by $\$10,780$ (and $\$20,240$) if the severity is medium (and high). That such an increase is greater for more severe damage may be ascribed to the fact that more severe damage requires more resources for care, which cost more in counties with household income.

Our last step is to estimate the joint distribution of plaintiff and defendant beliefs. To do so, we plug in the logit estimates from specification (3) in Table 2 and the MLE estimates from the second specification in Table 3 (a) in our Simulated Maximum Likelihood estimator defined in (13). As explained in Section 4, we maintain the specification that $\mu_p = 1 - Y_1$

and $\mu_d = Y_1 + Y_2$, where Y_1, Y_2 are the first two components in the draw from a Dirichlet distribution with concentration parameters $(\alpha_1, \alpha_2, \alpha_3)$. We allow α_j's to vary across the classification of cases based on the severity and the board certification of the doctors. We define a period in the model as a quarter in the calendar year, and adopt a binomial specification for the distribution of the wait-time T. Specifically, we let $T\tilde{}binomial(\overline{T}, p)$ where $\bar{T} \equiv 4$. We use a quarterly discount factor of 0.99 (which is consistent with a 4% annual inflation rate). While implementing SML, we use $S = 2,000$ simulated draws from the standard uniform distribution to evaluate the integral in the likelihood for each observation.

Table 4. Estimates for Parameters in the Belief Distribution

		Board Certified: Yes		Board Certified: No	
		μ_p	μ_d	μ_p	μ_d
Sev. = low	mean	0.402	0.692	0.407	0.691
		[0.366, 0.480]	[0.607, 0.785]	[0.396, 0.507]	[0.666, 0.703]
	std.dev.	0.173	0.163	0.168	0.158
		[0.167, 0.249]	[0.144, 0.244]	[0.167, 0.181]	[0.157, 0.166]
	skewness	0.251	-0.523	0.233	-0.509
		[0.059, 0.345]	[-0.867, -0.321]	[-0.017, 0.262]	[-0.564, -0.439]
	correlation	-0.613		-0.607	
		[-0.642, -0.411]		[-0.646,-0.447]	
Sev. = med.	mean	0.431	0.671	0.542	0.549
		[0.427, 0.464]	[0.625, 0.677]	[0.502, 0.602]	[0.543, 0.600]
	std.dev.	0.172	0.163	0.183	0.183
		[0.169, 0.172]	[0.162, 0.165]	[0.179, 0.183]	[0.175, 0.185]
	skewness	0.173	-0.452	-0.108	-0.128
		[0.088, 0.183]	[-0.468, -0.314]	[-0.274, -0.006]	[-0.266, -0.112]
	correlation	-0.804		-0.833	
		[-0.846, -0.701]		[-0.845, -0.671]	
Sev.= high	mean	0.510	0.585	0.733	0.338
		[0.499, 0.522]	[0.571, 0.596]	[0.728, 0.746]	[0.334, 0.366]
	std.dev.	0.176	0.173	0.163	0.175
		[0.173, 0.177]	[0.171, 0.175]	[0.162, 0.166]	[0.173, 0.181]
	skewness	-0.026	-0.216	-0.686	0.445
		[-0.056, 0.002]	[-0.246, -0.176]	[-0.746, -0.669]	[0.361, 0.457]
	correlation	-0.825		-0.844	
		[-0.845, -0.714]		[-0.864, -0.783]	

For each severity level and the certification of doctors, Table 4 reports the point estimates and 90% confidence intervals for the mean, standard deviation, skewness and correlation of the beliefs of the doctor and the patients, which are calculated using the SML estimates for $\hat{\alpha}_j$. (The closed form for mappings from the concentration parameters α_j to the mean, standard deviation, skewness and the correlation of (μ_p, μ_d) is presented in the appendix.) The confidence intervals reported in Table 4 are constructed using the empirical distribution of the respective estimates from $B = 200$ bootstrap samples. Figure 2 and Figure 3 in the appendix plots the estimated marginal and joint distribution of (μ_p, μ_d).

The estimates in Table 4 demonstrate several informative patterns about how the joint distribution of plaintiff and defendant beliefs vary with case characteristics. First, the plaintiffs tend to be more optimistic about their chance of winning the lawsuit when the injuries are more severe, regardless of the doctor's certification status. This is evident from the pattern that the estimated mean of μ_p is greater for high severity cases, and that the estimated skewness of μ_p decreases (shifting more probability masses toward 1) as severity increases. Comparing confidence intervals for the mean and skewness of μ_p across severity levels, we can see that such differences in the patient's belief are statistically significant at 10% or 15% level for most cases. The only exception is that this difference may be statistically insignificant as severity increase from low to medium when the doctor is board certified.

Second, defendants are less optimistic about the court decision as severity increases. Our estimates show the defendant belief μ_d becomes more positively skewed (shifting probability mass toward 0) with a lower mean when the injuries are more severe. Again, the differences in the mean and skewness of μ_d across classifications are mostly statistically significant, except for when the severity increases from low to medium and the doctor is board certified.

One explanation for the two patterns above could be the "sympathy factor" in the jury decision perceived by plaintiffs and defendants. That is, both parties might believe the jury is inclined to rule in favor of the patient out of sympathy if the severity of damage is high. Such an explanation is consistent with the earlier observation that the differences in beliefs are more pronounced when the severity reaches a high level as opposed to a medium level.

Third, for cases with medium or high severity, the plaintiffs tend to be significantly more optimistic (with beliefs μ_p more skewed to the left with a greater mean) when the doctors are not board certified. However, for cases with low severity, the doctor's certification does not seem to have any significant impact on the skewness the patient's belief. In comparison, the distribution of defendant beliefs move in an opposite direction, with the doctors with board certification being more optimistic. This is probably because board certification serves as a professional endorsement of a doctor's qualification and capabilities. Thus, both parties of the lawsuit may perceive board certification as a potential influence on the jury's decision that is favorable to the defendant.

Last but not the least, the joint distribution of (μ_p, μ_d) is estimated to be significantly negatively correlated, regardless of the severity and the doctor's qualification. The point estimates for the correlation is smaller for cases with lower severity, which conforms with the patterns mentioned above.

7 Policy Experiment

We now use our estimates to predict the impact of a hypothetical tort reform that limits the liability of defendants. In particular, we consider a policy which imposes a binding cap on the maximum compensation payable by the defendant if the court rules in favor of the plaintiff. For practical concerns, we allow these caps to vary with the reported severity of malpractice, and are set to 75% of the maximum compensation reported in the estimation data for a given severity level. Using our estimates from Table 4, we calculate the counterfactual mean of accepted settlement offers (that is, the mean of $S \mid A = 1$) under caps, and compare them with the reported empirical mean in the data (where there were no cap on total compensation to the best of our knowledge). Table 5 reports the point estimates as well as the 90% confidence intervals calculated using bootstrap resampling.

Table 5: Impact of compensation caps on mean accepted settlement offers
(Units: $\$1k$)

		Board Certified: Yes		Board Certified: No	
Low sev.	emp.	42.159	[38.166, 46.794]	41.203	[37.239, 45.749]
	c.f.	28.250	[21.118, 41.724]	27.522	[19.901, 34.190]
	diff.	13.909	[-1.280, 21.841]	13.680	[5.618, 22.195]
Med sev.	emp.	159.001	[151.180, 165.098]	129.145	[120.558, 137.454]
	c.f.	88.326	[73.371, 99.440]	117.121	[92.572, 132.746]
	diff.	70.675	[56.109, 88.235]	12.024	[-4.538, 36.286]
High sev.	emp.	263.059	[251.389, 277.277]	332.938	[312.392, 356.387]
	c.f.	199.944	[174.914, 226.242]	314.611	[260.384, 350.431]
	diff.	63.115	[32.691, 91.445]	18.327	[-28.745, 73.341]

Notes: "emp.": observed mean in the data; "c.f.": counterfactual mean under caps; "diff.": equals "emp."−"c.f."

Table 5 shows that on average the binding caps could induce sizable reductions in the accepted settlement offers. For example, the point estimates suggest that when the severity level is low, imposing a cap on total potential compensation that equals the 75% empirical quantile of court-ruled compensation would lead to over thirty-percent reduction in the mean of accepted settlement offer (32.99% when the doctor is reported to be board-certified; and 33.20% when there is no reported board certification). On the other hand, there is also evidence that the impact of compensation caps (conditional on medium and high severity) interacts with the reported qualification of the doctors. For cases with medium severity, the reduction in mean accepted offer is 44.45% when the doctors is board certified, compared with 9.31% when there is not certification reported in the data. Likewise, when severity is high, the reduction is as high as 23.99% with board certification but only 5.5% when no certification is reported in data. The bootstrap confidence intervals suggest that the impact of the means are statistically insignificant for cases with medium and high severity without board certification information.

The heterogeneity in the estimated impact of compensation caps across severity and doctor qualification can be explained by the difference in the joint beliefs and the distribution of compensation across these categories. Recall that the caps we consider condition on the severity levels but not the doctor's qualification. If we ignore the difference in the joint belief of plaintiffs and defendants across categories, the impact of caps on the mean of accepted settlement should be greater when the distribution of total compensation has a thinner (far-stretched) tail beyond the binding cap (imposed at 75% empirical quantile). For instance, when there is no certification information, the reduction in the counterfactual mean is greater with medium severity than with low severity. This is partly attributed to the fact that the range of realized compensations in the data that are censored by the caps (which are imposed at the 75% quantile) is larger in the former category, as seen the right panels in Figure 1(b).

Moreover the difference in the joint beliefs also interacts with that in the compensation distribution to affect the proportion of reduction. To see this, note that as severity becomes high, the range of censored compensation is even greater than the medium severity category. Nevertheless, as our estimates in Table 4 suggest, the plaintiffs subject to high-severity damage are significantly more optimistic. This could in part explain why the reduction in the mean accepted settlement is not as high as in the case with medium severity.

To further investigate the impact of caps on the distribution of $S \mid A = 1$, we report in Table 6 the estimated quantiles of accepted settlement offers between two scenarios: one with a binding cap on total compensation; and the other with no binding cap (that is, as if the cap were set to maximum reported in the data).

Table 6: Impact of compensation caps on quantiles of settlement offers
(Units: $\$1k$)

		Board Certified: Yes		Board Certified: No	
Low sev.	25%.	0.138	[0.039, 0.456]	0.038	[-0.013, 0.330]
	50%	2.154	[1.072, 4.957]	1.650	[0.718, 4.010]
	75%	9.066	[-18.000, 13.867]	8.299	[-12.149, 11.319]
Med sev.	25%.	0.729	[0.353, 1.263]	0.863	[0.533, 1.562]
	50%	9.938	[7.750, 15.158]	13.565	[9.841, 20.102]
	75%	92.947	[-97.099, 44.544]	60.188	[48.003, 79.916]
High sev.	25%.	1.410	[0.525, 2.364]	2.135	[0.518, 3.318]
	50%	15.676	[11.699, 26.511]	22.250	[14.389, 33.895]
	75%	76.946	[54.004, 98.928]	108.094	[75.332, 154.836]

The estimates suggest the conditional quantiles of accepted settlement offers are lower when a binding cap is imposed. Thus these estimates conforms to the theoretical implication that under a binding cap the distribution of accepted settlement offers is stochastically lower once a binding cap is imposed. Moreover, the difference in quantiles is statistically more significant for cases with greater severity. This is consistent with the pattern shown in Figure 1(b) that the range of compensation values censored by the caps is greater when severity is higher.

8 Concluding Remarks

There are several directions for further investigation. First, with the structural estimates of beliefs and surplus process, we can address policy questions such as the counterfactual probability of settlement or distribution of compensation under alternative schemes for scheduling the mandatory conference, or putting caps on the total non-economic losses possible. Second, while our identification strategy is nonparametric, the implementation in estimation in the current article is parametric. It would be interesting to construct fully nonparametric estimators based on the identification argument, say those using sieve approximation method. Third, in a more complicated contexts where parties make several rounds of offers and counter offers, identification of model primitives may still be achievable with additional observation of rejected offers. It will be useful to formalize the conditions that are needed for recovery of belief distribution and the surplus process in this case.

Appendix A: Figures

Figure 1(a): Distribution of transfers by plaintiff age

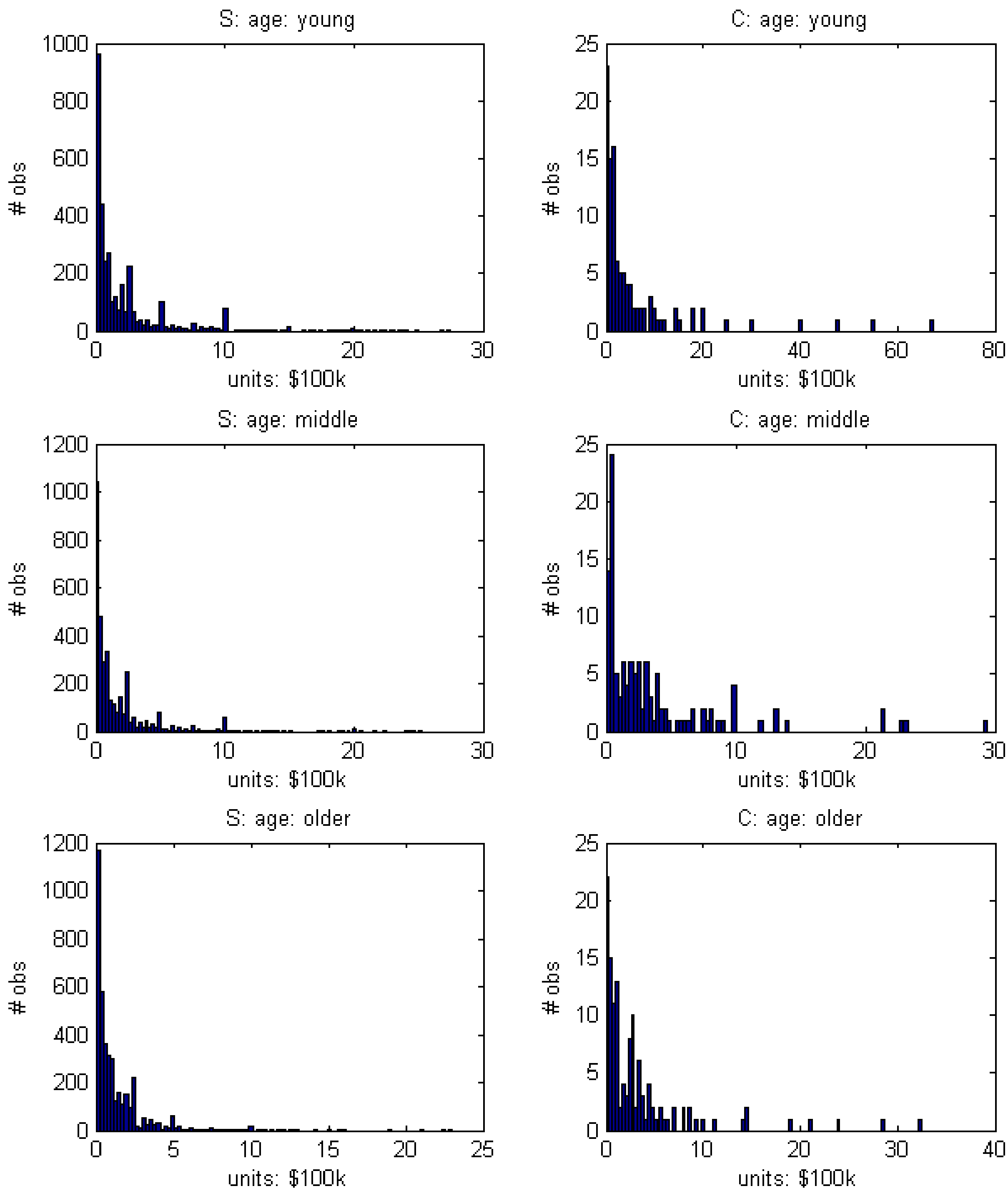


Figure 1(b): Distribution of transfers by severity

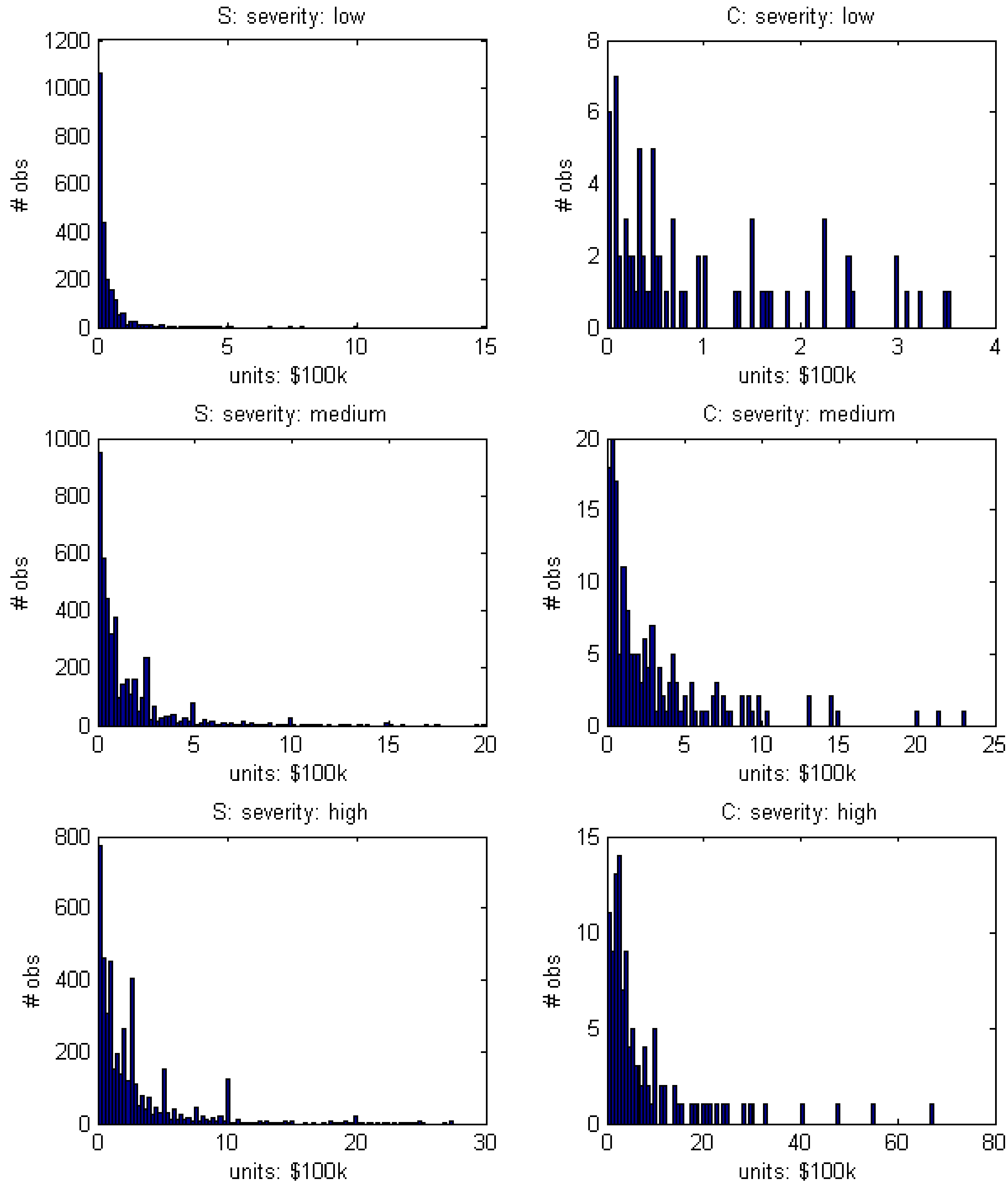
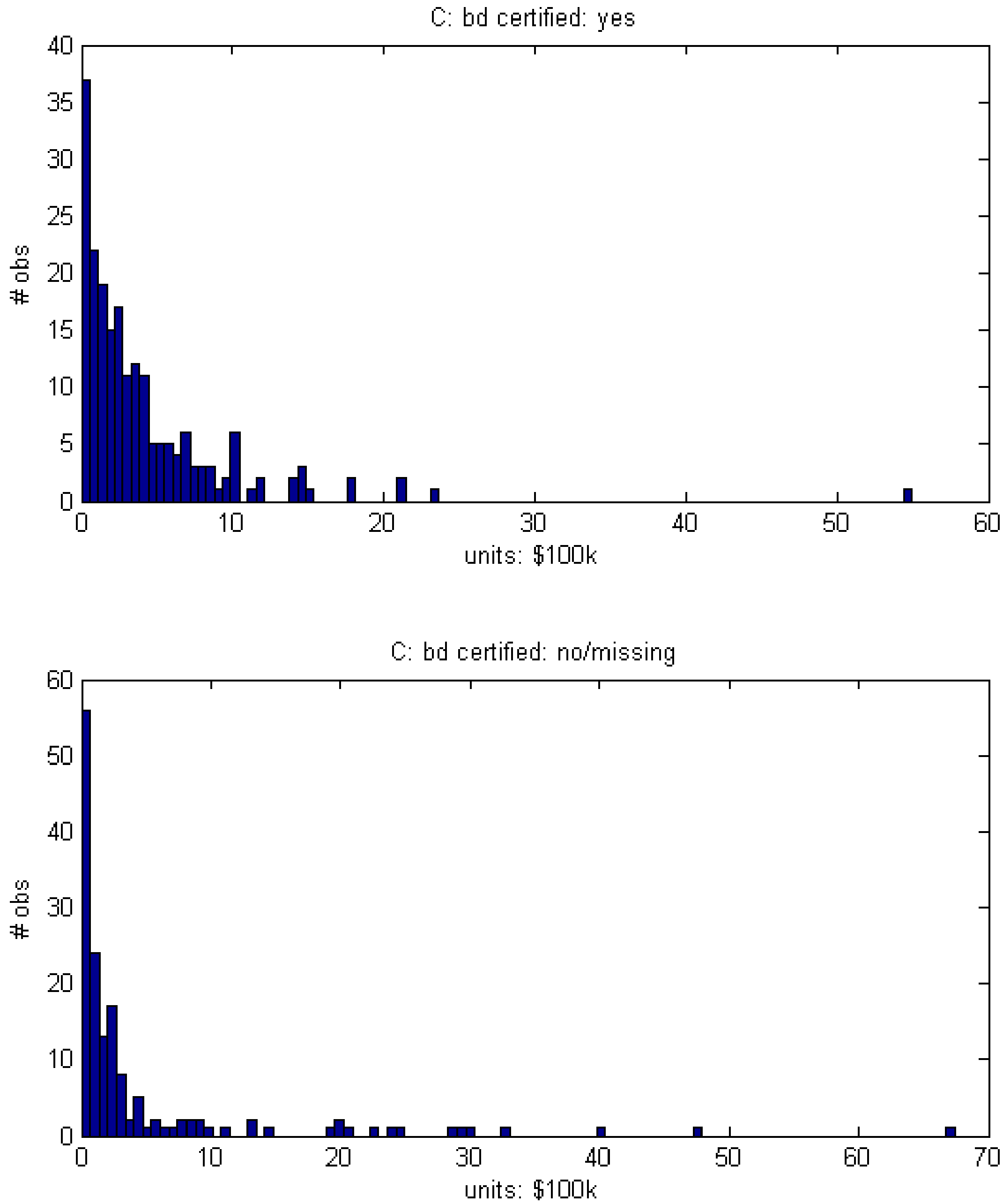
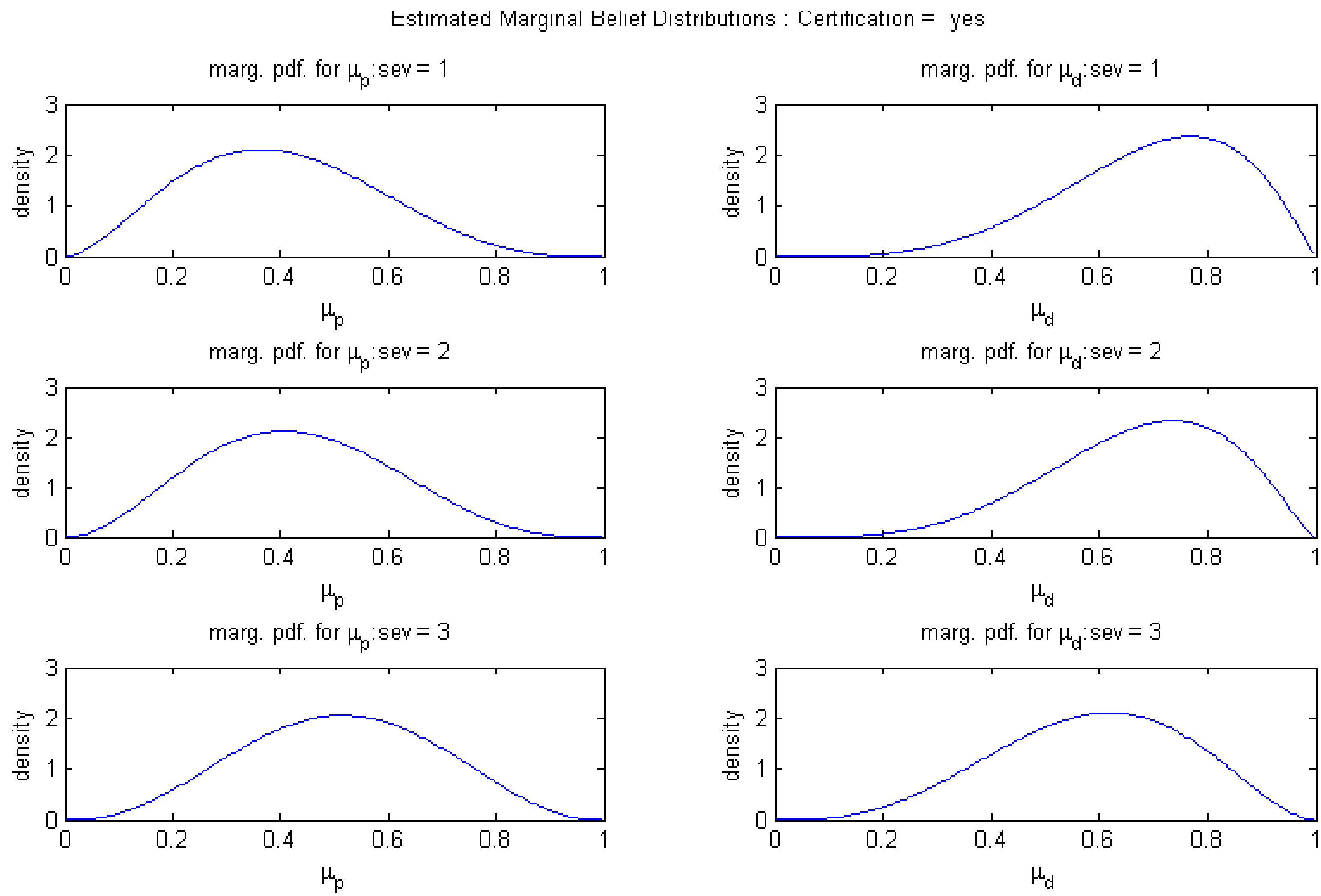
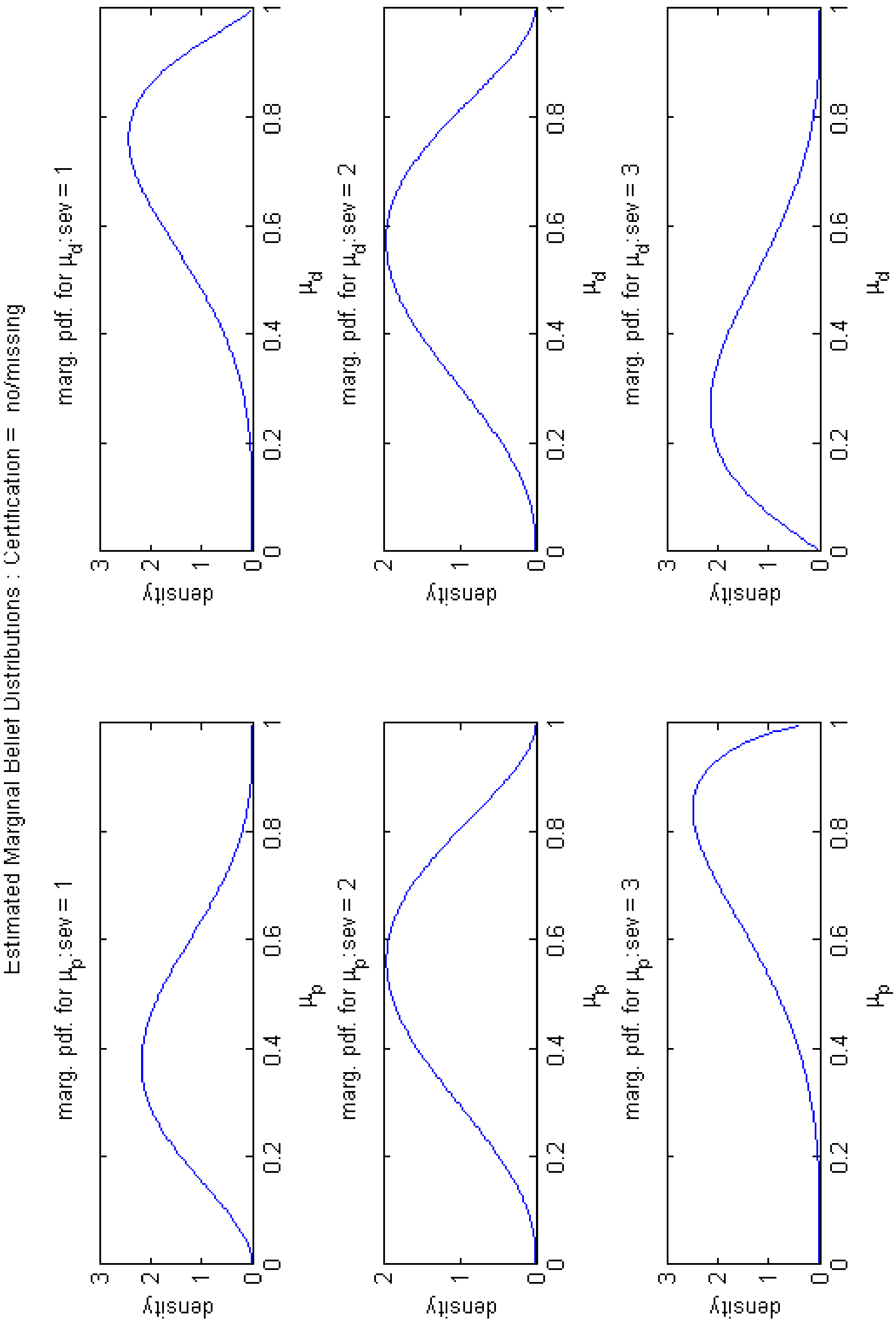


Figure 1(c): Distribution of transfers by certification




Estimated Marginal Belief Distributions : Certification = yes
marg. pdf. for μp:sev = 1
marg. pdf. for μd:sev = 1
marg. pdf. for μp:sev = 2
marg. pdf. for μd:sev = 2
marg. pdf. for μp:sev = 3
marg. pdf. for μd:sev = 3
density
μp
μd
0
0.2
0.4
0.6
0.8
1
2
3


Estimated Marginal Belief Distributions : Certification = no/missing
marg. pdf. for μ_p:sev = 1
marg. pdf. for μ_d:sev = 1
marg. pdf. for μ_p:sev = 2
marg. pdf. for μ_d:sev = 2
marg. pdf. for μ_p:sev = 3
marg. pdf. for μ_d:sev = 3
density
μ_p
μ_d

Appendix B: Proofs

Proof of Lemma 1. Lemma A1 below shows that, under Assumptions 1, 2 and 3, there exists a partition $\mathcal{D}_{|\mathcal{T}|}$ such that L_{S_i,S_k} has full rank $|\mathcal{T}|$. This necessarily means $L_{S_i|T}$ must also be invertible (see proof of Lemma A1). Hence

$$L_{T,S_k} = \left(L_{S_i|T}\right)^{-1} L_{S_i,S_k}. \tag{14}$$

Substituting (14) into (4) leads to

$$\begin{aligned} & \Lambda_{S_i,S_k} = L_{S_i|T}\ \Delta_j\ \left(L_{S_i|T}\right)^{-1}\ L_{S_i,S_k} \\ \Leftrightarrow\ & \Lambda_{S_i,S_k}\ (L_{S_i,S_k})^{-1} = L_{S_i|T}\ \Delta_j\ \left(L_{S_i|T}\right)^{-1}. \end{aligned} \tag{15}$$

Note the two matrices on the left-hand side of (15), L_{S_i,S_k} and Λ_{S_i,S_k}, are directly identifiable from data. The equation (15) suggests $\Lambda_{S_i,S_k}(L_{S_i,S_k})^{-1}$ admits an eigenvalue-eigenvector decomposition, with each eigenvalue being $\mathbb{E}\left(A_j \mid T = t\right)$ and corresponding eigenvector being $[\Pr(S_i \in d_m \mid A_i = 1, T = t)]_{m=1,..,|\mathcal{T}|}$ for any $t \in \mathcal{T}$. Because each eigenvector is a conditional probability mass function with entries summing up to 1, the scale in the decomposition is fixed implicitly. Furthermore, Assumption 1-(i) and Assumption 3-(i) imply that $\mathbb{E}[A_j \mid T = t]$ must be strictly decreasing in t over the support of wait-time between settlement conferences and court hearings $\mathcal{T}$. This rules out the possibility of duplicate eigenvalues in the decomposition, and also uniquely links the eigenvalues and eigenvectors to each specific elements in $\mathcal{T}$. Thus $L_{S_i|T}$ and Δ_j are identified using the partition $D_{|\mathcal{T}|}$. It then follows that L_{T,S_k} is also identified from (14) once $L_{S_i|T}$ is identified.

It remains to show that $f_{S_i}(. \mid A_i = 1, T = t)$ is identified over its full support. (Note $L_{S_i|T}$ identified above is only a discretized version of this conditional density.) For any $s \in \mathcal{S}$, define a $|\mathcal{T}|$-vector l_{s,S_k} whose m-th coordinate is given by $f_{S_i}(s, S_k \in d_m \mid A_{i,k} = 1)$, where d_m is the m-th interval in the partition $D_{|\mathcal{T}|}$ used for identification of $L_{S_i|T}$ and Δ_j above. By construction,

$$l_{s,S_k} = (L_{T,S_k})'\,\lambda_s \tag{16}$$

where $(L_{T,S_k})'$ is the transpose of L_{T,S_k}; and λ_s is a $|\mathcal{T}|$-vector with the t-th coordinate being $f_{S_i}(s \mid A_i = 1, T = t)$.[12] The coefficient matrix $(L_{T,S_k})'$ does not depend on the realization of $S_i = s$ while vectors λ_s and l_{s,S_k} both do. With L_{T,S_k} invertible and identified above and

[12] To see this, note for any s and d_m, the Law of Total Probability implies $f_{S_i}(s, S_k \in d_m \mid A_{i,k} = 1)$ can be written as:

$$\begin{aligned} & \textstyle\sum_{t\in\mathcal{T}} f_{S_i}(s \mid T = t, S_k \in d_m, A_{i,k} = 1) \Pr(T = t, S_k \in d_m \mid A_{i,k} = 1) \\ = \ & \textstyle\sum_{t\in\mathcal{T}} f_{S_i}(s \mid A_i = 1, T = t) \Pr(T = t, S_k \in d_m \mid A_{i,k} = 1) \end{aligned}$$

where the equality follows from Assumption 1.

with l_{s,S_k} directly identifiable, λ_s is recovered as the unique solution of the linear system in (16) for any $s \in \mathcal{S}$. $\square$

Lemma A1 *Under Assumption 1, 2 and 3, there exists a partition $\mathcal{D}_{|\mathcal{T}|}$ such that L_{S_i,S_k} has full-rank $|\mathcal{T}|$.*

Proof of Lemma A1. By construction, the supremum and the infimum of the support of μ_p given $A = 1$ ($\mu_p + \mu_d \leq \delta^{-t}$) and $T = t$ are 1 and 0 respectively. It then follows from (ii) in Assumption 3 implies the supremum and the infimum of the support of $S = \delta^T \mu_p C$ conditional on "$A = 1$ and $T = t$" are $\delta^t \times 1 \times \overline{c} = \delta^t \overline{c}$ and $\delta^t \times 0 \times 0 = 0$ respectively. For any $t \in \mathcal{T} \equiv \{1, 2, ., |\mathcal{T}|\}$, let $\bar{s}(t) \equiv \overline{c}\delta^t$ denote the supreme of the support of S conditional on "$A = 1$ and $T = t$". Let $\mathcal{D}_{|\mathcal{T}|}$ be a partition of the unconditional support of settlement offers $\mathcal{S}$ into $|\mathcal{T}|$ intervals, which are characterized by the sequence of endpoints $\bar{s}(1) > \bar{s}(2) > \bar{s}(3) > .. > \bar{s}(|\mathcal{T}|) > \bar{s}(|\mathcal{T}| + 1) \equiv 0$. (That is, the t-th smallest interval in $\mathcal{D}_{|\mathcal{T}|}$ 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 $T = t$ the settlement offer S is continuously distributed over $[0, \delta^t \overline{c}]$ with positive densities, the square matrix $L_{S_i|T}$ based on the partition $\mathcal{D}_{|\mathcal{T}|}$ must be triangular with full rank $|\mathcal{T}|$. Next, note that $L_{S_i,S_k} = L_{S_i|T}\ D_T\ \left(L_{S_k|T}\right)'$ where $\left(L_{S_k|T}\right)'$ is the transpose of $L_{S_k|T}$ and D_T is a diagonal matrix with diagonal entries being $[\Pr(T = t)]_{t=1,.,|\mathcal{T}|}$. Since $L_{S_k|T}$ has full rank by symmetric arguments and D_T is non-singular by construction, it then follows that L_{S_i,S_k} has full rank. $\square$

Proof of Proposition 1.. By definition, the function $\varphi_t(s)$ on the L.H.S. of (10) is directly identifiable for all s, t. For any given $t \in \mathcal{T}$, suppose there exists $\tilde{h}_t \neq h_t$ such that $\varphi_t(s) = \int_0^\infty \tilde{h}_t(v)\kappa(v, s)dv$ for all $s \in \mathcal{S}_t$. Then (10) implies:

$$\int_0^\infty \left[\tilde{h}_t(v) - h_t(v)\right] \kappa(v, s)dv = 0$$

for all $s \in \mathcal{S}_t$. It then follows from Assumption 5 that $\tilde{h}_t(v) = h_t(v)$ almost everywhere $F_{V|t}$ for such t. This establishes the identification of $h_t(v)$ for any $t \in \mathcal{T}$ and $v \in \mathcal{V}_t$. Thus $\Pr(\frac{1}{\mu_p} \geq b, \frac{1}{\mu_p+\mu_d} \geq \delta^t)$ is identified for all $t \in \mathcal{T}$ and $b \in [1, +\infty)$, i.e. the support of μ_p^{-1} given $\mu_p + \mu_d \leq \delta^{-t}$. (To see this, note that for any t, the support of $V\delta^t$ given $\mu_p + \mu_d \leq \delta^{-t}$ is by construction identical to that of μ_p^{-1} given $(\mu_p + \mu_d)^{-1} \geq \delta^t$.) It then follows that $\Pr(\mu_p \leq \mu\ ,\ \mu_p + \mu_d \leq \delta^{-t})$ is identified for all $t \in \mathcal{T}$ and $\mu \in (0, 1]$. $\square$

Appendix C: Monte Carlo Study

In this section we present some evidence of the finite-sample performance of the Maximum Simulated Likelihood estimator proposed in Section 4. For the sake of simplicity, we focus on a simple design where cases are homogenous. Let the data-generating process be defined as follows. Let $R \in \{r \in [0,1]^3 : r_1 + r_2 + r_3 = 1\}$ follow a Dirichlet distribution with concentration parameters $(\alpha_1, \alpha_2, \alpha_3)$. Let $\mu_p = 1 - R_1$ and $\mu_d = R_2 - (1 - R_1) + 1 = R_1 + R_2$. 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 R_1 is $Beta(\alpha_1, \alpha_2 + \alpha_3)$); and the marginal distribution of μ_d is $Beta(\alpha_1 + \alpha_2, \alpha_3)$.[13] 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
Marg. distr.	$Beta(\alpha_2 + \alpha_3, \alpha_1)$	$Beta(\alpha_1 + \alpha_2, \alpha_3)$
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	$-\frac{\sqrt{\alpha_1\alpha_3}}{\sqrt{(\alpha_2+\alpha_3)(\alpha_1+\alpha_2)}}$	

We also use the simple design where the distribution of the cake: $gamma(1,1)$. $\Pr(D = 1) = 0.05$ and that the distribution of wait-time T is Binomial with parameters 5 and 0.4. The results are reported in the following tables:

[13] The covariance between μ_p and μ_d is given by: $Cov(\mu_p, \mu_d) = Cov(1 - R_1, R_1 + R_2) = Var(R_1) - Cov(R_1, R_2)$, which is used to calculate the expression reported in Table C1.

Table C2: Results for DGP # 1: $\alpha \equiv$[1.25,1.50,2.70]

DGP1		mean	std. dev	l.q.	med.	h.q.	r.m.s.e.	m.a.e.
$N = 200$	α_1	1.580	0.864	0.945	1.456	1.996	0.923	0.467
	α_2	1.635	0.860	1.155	1.401	1.801	0.868	0.328
	α_3	3.090	1.334	2.157	2.841	3.734	1.387	0.723
	λ	0.385	0.213	0.208	0.343	0.539	0.213	0.172
$N = 400$	α_1	1.424	0.599	0.998	1.350	1.705	0.623	0.350
	α_2	1.482	0.489	1.131	1.403	1.711	0.489	0.312
	α_3	2.757	0.960	2.149	2.645	3.224	0.960	0.541
	λ	0.378	0.175	0.232	0.362	0.517	0.176	0.149
$N = 800$	α_1	1.291	0.431	0.999	1.210	1.495	0.431	0.246
	α_2	1.330	0.318	1.126	1.275	1.494	0.359	0.266
	α_3	2.489	0.574	2.127	2.439	2.920	0.609	0.436
	λ	0.355	0.135	0.256	0.324	0.453	0.142	0.111

Table C3: Results for DGP # 2: $\alpha =$[3.60, 2.00, 1.40]

DGP1		mean	std. dev	l.q.	med.	h.q.	r.m.s.e.	m.a.e.
$N = 200$	α_1	5.014	2.406	3.503	4.465	5.798	2.785	1.187
	α_2	1.958	0.655	1.515	1.832	2.274	0.655	0.443
	α_3	2.056	0.916	1.460	1.828	2.290	1.124	0.432
	λ	0.263	0.161	0.142	0.220	0.351	0.211	0.199
$N = 400$	α_1	4.562	1.496	3.428	4.248	5.400	1.775	0.788
	α_2	1.727	0.363	1.481	1.672	1.921	0.454	0.374
	α_3	1.871	0.563	1.474	1.730	2.154	0.733	0.342
	λ	0.230	0.105	0.165	0.204	0.277	0.200	0.197
$N = 800$	α_1	4.383	0.964	3.644	4.273	4.964	1.239	0.743
	α_2	1.695	0.259	1.533	1.696	1.798	0.400	0.333
	α_3	1.828	0.378	1.559	1.759	2.087	0.570	0.359
	λ	0.231	0.089	0.168	0.216	0.273	0.191	0.184

Appendix D: Calculating Counterfactuals

In this appendix we provide further details for calculating the counterfactual mean settlement offer under compensation caps considered in Section 7. First, we calculate the counterfactual density of $S \mid A = 1$ under the proposed caps on C. Let x denote the variables that affect the distribution of C (that is, age, income and severity, which is suppressed in the notation below). Let $f(. \mid x)$ denote the conditional density of C; h denote the probability mass for wait-time T (which is orthogonal to μ, C). By construction,

$$\tfrac{\partial}{\partial s} \Pr\{S \leq s \mid A = 1, X = x\} = \frac{\frac{\partial}{\partial s}\Pr\{S \leq s, A=1 \mid X=x\}}{\Pr\{A=1 \mid X=x\}} \tag{17}$$

where the denominator is

$$\textstyle\sum_t \Pr\left\{\mu_p + \mu_d - 1 \leq \frac{1-\delta^t}{\delta^t}\right\} h(t)$$

due to independence between T and (μ, C); and the numerator is:

$$\begin{aligned} & \tfrac{\partial}{\partial s} \textstyle\sum_t \left[\displaystyle\int_0^{\bar{c}} \Pr\left(\mu_p \leq \tfrac{s}{c\delta^t}, \mu_p + \mu_d \leq \tfrac{1}{\delta^t}\right) f(c \mid x) dc\right] h(t) \\ = \; & \textstyle\sum_t \left[\tfrac{\partial}{\partial s} \displaystyle\int_0^{\bar{c}} \Pr\left(Y_1 \geq 1 - \tfrac{s}{c\delta^t}, Y_2 \leq \tfrac{1-\delta^t}{\delta^t}\right) f(c \mid x) dc\right] h(t) \end{aligned}$$

where $Y_1 \equiv 1 - \mu_p$ and $Y_2 = \mu_p + \mu_d - 1$. Assuming the order of integration and differentiation can be changed and using the fact that "$1 - s\delta^{-t}/c \in (0, 1) \Leftrightarrow c \in (\frac{s}{\delta^t}, +\infty)$", we can write the term in the square brackets on the right-hand side for each t as

$$\begin{aligned} & \int_{s\delta^{-t}}^{\infty} \Pr\left\{Y_2 \leq \tfrac{1-\delta^t}{\delta^t} \,\middle|\, Y_1 = 1 - s\delta^{-t}/c\right\} f_{Y_1}(1 - s\delta^{-t}/c) \frac{f_C(c \mid x)}{c\delta^t} dc \\ = \; & \int_0^1 \Pr\left\{\frac{Y_2}{1-\tau} \leq \tfrac{1-\delta^t}{\delta^t(1-\tau)} \,\middle|\, Y_1 = \tau\right\} \frac{f_{Y_1}(\tau) f_C\left(\frac{s}{\delta^t(1-\tau)} \mid x\right)}{\delta^t(1-\tau)} d\tau \end{aligned}$$

where the equality follows from changing variables between c and $\tau \equiv 1 - \frac{s}{c\delta^t}$.

Suppose we put a cap $\hat{c}$ on total compensation, then the expression in the numerator on the right-hand side of (17) becomes

$$\tfrac{\partial}{\partial s} \textstyle\sum_t \left[\begin{array}{l} \int_0^{\hat{c}} \Pr\left(\mu_p \leq \frac{s}{c\delta^t}, \mu_p + \mu_d \leq \frac{1}{\delta^t}\right) f(c \mid x) dc + \\ \Pr\{C \geq \hat{c} \mid x\} \Pr\left(\mu_p \leq \frac{s}{\hat{c}\delta^t}, \mu_p + \mu_d \leq \frac{1}{\delta^t}\right) \end{array}\right] h(t).$$

By construction, $\frac{\partial}{\partial s} \Pr\left(\mu_p \leq \frac{s}{\hat{c}\delta^t}, \mu_p + \mu_d \leq \frac{1}{\delta^t}\right)$ is

$$\tfrac{\partial}{\partial s} \Pr\left(Y_1 \geq 1 - \tfrac{s}{\hat{c}\delta^t}, Y_2 \leq \tfrac{1-\delta^t}{\delta^t}\right) = \Pr\left(Y_2 \leq \tfrac{1-\delta^t}{\delta^t} \,\middle|\, Y_1 = 1 - \tfrac{s}{\hat{c}\delta^t}\right) \frac{f_{Y_1}(1 - s\delta^{-t}/\hat{c})}{\hat{c}\delta^t}.$$

A similiar expression exists with $\hat{c}$ replaced by c. Given our estimates for the distribution of C and (Y_1, Y_2) from Section 4, we calculate the counterfactual mean of settlement offers using these formulas and the simulation-based integration. The estimated counterfactual means and their differences with empirical means in the data are reported in Table 5.

Next, we explain how to calculate the distribution of $S \mid A = 1$ under counterfactual caps on C. Recall by construction

$$\Pr\{S \leq s \mid A = 1, X = x\} = \frac{\Pr\{S \leq s, A=1 \mid X=x\}}{\Pr\{A=1 \mid X=x\}}$$

where the denominator is calculated as before. The numerator under a cap $\hat{c}$ is

$$\sum_t \left[\int_0^{\hat{c}} \Pr\left(\mu_p \le \tfrac{s}{c\delta^t}, \mu_p+\mu_d \le \tfrac{1}{\delta^t}\right) f(c\mid x)dc + \Pr(C \ge \hat{c}\mid x)\Pr\left(\mu_p \le \tfrac{s}{\hat{c}\delta^t}, \mu_p+\mu_d \le \tfrac{1}{\delta^t}\right)\right] h(t). \qquad (18)$$

To calculate the term in the square brackets, we need to consider two cases.

Case 1: $\hat{c} \le s\delta^{-t}$. Then $\frac{s}{\hat{c}\delta^t} \ge 1$ and

$$\Pr\left(\mu_p \le \tfrac{s}{\hat{c}\delta^t}, \mu_p+\mu_d \le \tfrac{1}{\delta^t}\right) = \Pr\left(Y_2 \le \tfrac{1-\delta^t}{\delta^t}\right).$$

Besides,

$$\begin{aligned}
&\int_0^{\hat{c}} \Pr\left(\mu_p \le \tfrac{s}{c\delta^t}, \mu_p+\mu_d \le \tfrac{1}{\delta^t}\right) f(c\mid x)dc = \int_0^{\hat{c}} \Pr\left(Y_1 \ge 1-\tfrac{s}{c\delta^t}, Y_2 \le \tfrac{1-\delta^t}{\delta^t}\right) f(c\mid x)dc \\
= \;& \Pr\left(Y_2 \le \tfrac{1-\delta^t}{\delta^t}\right)\int_0^{\hat{c}} f(c\mid x)dc = \Pr\left(Y_2 \le \tfrac{1-\delta^t}{\delta^t}\right)\Pr\left(C \le \hat{c}\mid x\right)
\end{aligned}$$

where the second equality holds because $\hat{c} \le s\delta^{-t}$ implies $1-\frac{s}{c\delta^t} < 0$ for all $c \le \hat{c}$. Therefore the square bracket in (18) is

$$\begin{aligned}
&\Pr\left(Y_2 \le \tfrac{1-\delta^t}{\delta^t}\right)\Pr\left(C \le \hat{c}\mid x\right) + \Pr(C \ge \hat{c}\mid x)\Pr\left(Y_2 \le \tfrac{1-\delta^t}{\delta^t}\right) \\
= \;& \Pr\left(Y_2 \le \tfrac{1-\delta^t}{\delta^t}\right)
\end{aligned}$$

Case 2: $\hat{c} > s\delta^{-t}$. Then $\frac{s}{\hat{c}\delta^t} < 1$ and

$$\begin{aligned}
&\int_0^{\hat{c}} \Pr\left(\mu_p \le \tfrac{s}{c\delta^t}, \mu_p+\mu_d \le \tfrac{1}{\delta^t}\right) f_C(c\mid x)dc \\
= \;& \int_{s\delta^{-t}}^{\hat{c}} \Pr\left(Y_1 \ge 1-\tfrac{s}{c\delta^t}, Y_2 \le \tfrac{1-\delta^t}{\delta^t}\right) f_C(c\mid x)dc + \Pr\left(Y_2 \le \tfrac{1-\delta^t}{\delta^t}\right)\Pr(C \le s\delta^{-t}\mid x) \\
= \;& \int_0^{1-\frac{s}{\hat{c}\delta^t}} \Pr\left(Y_1 \ge \tau, Y_2 \le \tfrac{1-\delta^t}{\delta^t}\right) f_C\left(\frac{s}{\delta^t(1-\tau)}\middle| x\right)\frac{s}{(1-\tau)^2\delta^t}d\tau + \Pr\left(Y_2 \le \tfrac{1-\delta^t}{\delta^t}\right)\Pr(C \le s\delta^{-t}\mid x)
\end{aligned}$$

where the first equality uses the fact that $1-\frac{s}{c\delta^t} < 0$ if $c > s\delta^{-t}$ and the second uses the change of variables between c and $\tau \equiv 1-\frac{s}{c\delta^t}$. Furthermore

$$\Pr\left(Y_1 \ge \tau, Y_2 \le \tfrac{1-\delta^t}{\delta^t}\right) = \int_\tau^1 \Pr\left(Y_2 \le \tfrac{1-\delta^t}{\delta^t}\middle| Y_1 = \zeta\right) f_{Y_1}(\zeta)d\zeta.$$

Hence the square bracket in (18) is

$$\begin{aligned}
&\int_0^{1-\frac{s}{\hat{c}\delta^t}} \Pr\left(Y_1 \ge \tau, Y_2 \le \tfrac{1-\delta^t}{\delta^t}\right) f_C\left(\frac{s}{\delta^t(1-\tau)}\middle| x\right)\frac{s}{(1-\tau)^2\delta^t}d\tau \\
+ \;& \Pr\left(Y_2 \le \tfrac{1-\delta^t}{\delta^t}\right)\Pr(C \le s\delta^{-t}\mid x) + \Pr(C \ge \hat{c}\mid x)\Pr\left(Y_1 \ge 1-\tfrac{s}{\hat{c}\delta^t}, Y_2 \le \tfrac{1-\delta^t}{\delta^t}\right).
\end{aligned}$$

Again given our estimates for the distribution of C and (Y_1, Y_2) from Section 4, we can calculate the counterfactual distribution of settlement offers using these formulas and simulation-based integration. These estimated counterfactual distributions of $S \mid A = 1$ are then inverted to estimate counterfactual quantiles, whose differences with the empirical quantiles are reported in Table 6.

REFERENCES

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

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

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

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

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

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

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

Kasahara, H. and K. Shimotsu (2009): "Nonparametric identi cation of nite mixture models of dynamic discrete choices," Econometrica, vol 77: 135-175

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

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

Thanassoulis J. (2010): "Optimal stalling when bargaining," Journal of Econ. Dyn. Control, vol 34: 101–20

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

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