Lay Juries, Professional Arbitrators, and the Arbitrator Selection Hypothesis

Donald Wittman, *University of California, Santa Cruz*

Do civil juries follow the broad dictates of the law? For example, do those plaintiffs who suffer greater damages receive greater awards? Are juries consistent? Do juries empty deep pockets? In many states automobile accidents are first tried by a professional arbitrator and then by a jury if one of the litigants is dissatisfied with the outcome. How do the decisions made by professional arbitrators compare to the decisions made by juries? This article seeks to answer these questions by first developing a model of arbitrator selection and then undertaking an empirical study of 380 automobile accident cases that went through both an arbitration and a jury trial.

Do civil juries follow the broad dictates of the law? For example, in automobile accident cases do plaintiffs with greater medical bills receive higher awards from juries, and do more reckless defendants pay a greater share of the damages, as is appropriate under comparative negligence? Do juries really go after deep pockets? And if so, how deep do they reach? Are jury verdicts predictable, or do awards depend on the vagaries of the particular jury?

Approximately half the state courts, as well as ten federal district courts, have implemented court-annexed arbitration, which substitutes a legal professional for a judge and jury, and proceeds under less formal rules, but is otherwise similar to an ordinary civil trial. Is there a different

More than perfunctory thanks go to the referee and the editor for their ideas regarding the econometrics and the theory. I would also like to thank Gary Charness, John Morgan, and Mitchell Polinsky for helpful input.

Send correspondence to: Donald Wittman, Economics Department, University of California, Santa Cruz, CA 95064; Fax: (831) 459-5900; E-mail: wittman@cats.ucsc.edu.

type of justice being served to those who resolve the dispute differently? That is, do the alternative methods produce similar or different outcomes? For example, do plaintiffs tend to receive less when they face a professional arbitrator instead of a jury?

This article provides answers to these questions by considering 380 automobile accident cases that were tried in California by both professional arbitrator and jury. This data set allows us to discover the relationship between case characteristics (such as damages to the plaintiff) and the amount awarded by the jury or arbitrator. We are also able to investigate the importance of deep pockets by discovering the impact on the jury (or arbitrator) award when the defendant is a business or government agency.

1. The Arbitrator Selection Hypothesis

In California, as in many other states, most automobile accident litigation must first be tried by an arbitrator rather than by a jury. However, either side may request a jury trial if the litigant is not satisfied with the arbitrator decision. The jury is not informed that arbitration has taken place, let alone the outcome. However, if the judgment is not more favorable to the requesting party than the arbitration award, then the party requesting the trial must pay the arbitrator's fee and costs incurred after the request for a trial de novo is filed. Any member of the state bar and any active or retired judge may serve as an arbitrator. The arbitration administrator selects a short list of three arbitrators. The litigants must agree on one.[1]

I argue that arbitrators who are not good predictors of jury outcomes will not be chosen to arbitrate. As a consequence, arbitration verdicts will look very similar to jury verdicts. I label this the "arbitrator selection hypothesis."

1. Arbitration takes place on a less formal basis than an ordinary court trial. The rules of evidence are relaxed, and the case is usually tried within three months. Arbitration is generally required when the amount in controversy does not exceed $50,000. For larger amounts, there is arbitration when both parties stipulate to a hearing. There is no limit on the amount that the arbitrator can award. There is some variation from jurisdiction to jurisdiction on the particular details of the program. The judicial arbitrator is not at all similar to a mediator, whose job is to get both sides to compromise and come to an agreement in a dispute.

The basic intuition behind the arbitrator selection hypothesis is as follows: The objective of court-imposed arbitration is to reduce the number of cases going to jury trial.[2] If two arbitrators are unbiased predictors of the expected jury outcome (that is, their decisions on average are the same as jury decisions), then arbitration administrators and litigants will prefer the lower-variance arbitrator. The lower the variance, the more weight that each litigant will put on the arbitrator's decision relative to his or her own independent estimate of the jury outcome, and the more likely that the differential in the litigants' expectations will be less than the cost of going to a jury trial. That is, the lower the arbitrator variance, the more informative the arbitrator's decision, and the more likely that the litigants will either accept the arbitrator verdict or settle afterwards, thereby avoiding a costly jury trial.[3] The arbitration administrator and the litigants would prefer to save on needless court costs, so they will choose the lower-variance arbitrator. On the other hand, given two arbitrators with identical variances, but differing biases vis-à-vis jury decisions, the litigants may accept the arbitrator decision and eliminate the postarbitration negotiation costs if the arbitrator decision is not too far away from the expected jury decision. So it may benefit one or both parties when a less biased arbitrator is chosen.

In short, arbitrators are chosen in the shadow of the potential jury outcome. Those who are hired will have reputations for making awards close to the jury award. Successful arbitrators will either knowingly try to emulate juries (and because of their experience as lawyers and ex-judges will be able to do so) or naturally have the same attitudes as the average jury does. Thus, the competitive market for arbitrators implies that their decisions should be much more consistent with jury verdicts than would be the case if arbitrators were randomly chosen.[4]

---

2. The formal argument is found in Appendix A.

3. Note that the litigants gain information about the jury outcome from the arbitrator's decision and not from knowledge of the arbitrator. This can readily be seen when the litigants have diffuse priors over the jury award. Then knowing that an arbitrator is an unbiased predictor of the jury outcome also yields a diffuse prior over the arbitrator decision. Once the arbitrator has made her decision, the litigants update their beliefs.

4. Labor economists (see Ashenfelter, 1998, and Bloom and Cavanagh, 1986, for examples) have a model of arbitrator selection, but it is significantly different from the one presented here. Here arbitration takes place in the shadow of a jury trial; in the labor economist model the arbitrator is the last in line. Here the litigants desire arbitrators that are low-variance, unbiased estimators of the jury trial outcome; in contrast, unions and

Even if the arbitrator selection hypothesis did not hold, we would expect to see some similarities between jury and arbitrator decisions. After all, they are members of the same culture. Thus, we would expect that both juries and arbitrators would follow the broad dictates of the law. For example, other things being equal, larger plaintiff medical bills would lead to greater jury and arbitrator awards.

2. The Data

The data are drawn from *Jury Verdicts Weekly* over a nine-year period. My focus is on traffic accidents involving automobiles, trucks, or buses (accidents involving motorcycles, pedestrians, or deaths are not included in the sample). Focusing on one type of case, rather than pooling across issues such as malpractice and breach of contract, reduces the "noise" from extraneous differences in the cases. Traffic accidents were chosen because they are the most common civil jury case. *Jury Verdicts Weekly* asks lawyers from both sides to report various details of the case. The reports are then edited and published weekly. They cover such details as the jury award (JAWARD), the plaintiff's medical bills (MEDICAL.BILLS), and the arbitrator award (ARBAWARD) when the case was previously tried by an arbitrator (variables used in estimation will be denoted by capitals).[5] *Jury Verdicts Weekly*'s coverage of civil juries is very good. For example, it covers more than 85% of all jury verdicts in San Francisco (see Gross and Syverud, 1991; Shanley and Peterson, 1983).

California employs comparative negligence (this is the case for most other states as well). The defendant pays the plaintiff an amount equal to the defendant's relative negligence (RELNEG) times the amount of damage to the plaintiff. For example, if the plaintiff has suffered $100 in damages and the defendant was very negligent and the plaintiff was not at all negligent, then the plaintiff should be awarded $100; if the plaintiff and the defendant were equally negligent, then the plaintiff should be awarded

corporations choose unpredictable arbitrators who have not made choices significantly different from other arbitrators (the exchangeability hypothesis). If the labor arbitrator were predictable, choosing an arbitrator would be tantamount to settling.

5. Summary statistics are found in Table 7.

\$50.00. The measure for relative negligence of the defendant is based on the following formula:

$$\text{RELNEG} = (\text{DCULPABILITY})/(\text{DCULPABILITY} + \text{PCULPABILITY}) \quad (1)$$

Jury Verdicts Weekly describes the facts of the case. A value of 3 was ascribed to DCULPABILITY if the defendant had been driving recklessly; a value of 1 if the defendant's driving was normal; and a value of 2 for something in between. A similar scheme was devised for PCULPABILITY. Thus if both litigants had a score of 2, then the defendant's RELNEG would equal 1/2. If both litigants had a score of 1, then RELNEG would again be 1/2.[6]

A jury's or arbitrator's subjective estimate of the total damage should presumably be related to some objective measure of damage. The objective measure of damage here is MEDICAL.BILLS. Multiplying RELNEG times MEDICAL.BILLS, we get RELNEG.MED. Hence, RELNEG.MED is the empirical counterpart to the rule of comparative negligence, which apportions the liability for the damage according to relative negligence.

Since MEDICAL.BILLS includes estimated future medical costs, disputes about the size of the MEDICAL.BILLS may arise. Letting DISPUTE = 1 if a dispute arises, and letting dispute = 0 if there is no dispute, we create a new variable RNG.MED.DISP = RELNEG.MED∗ DISPUTE.[7] Other things being equal, both juries and arbitrators should award less than otherwise when there is a dispute about the size of the medical bills.

6. The case description and assignment of numbers was done without reference to the jury or arbitrator decision. I use the word *culpability* instead of negligence to emphasize that the DCULPABILITY numbers come from the description of the case rather than from the jury's determination of negligence. The jury's determination of the percent negligence was rarely reported and data for it was not collected. In Appendix B, I consider some other ways of defining relative negligence. These alternative measures yield very similar results. Of course, one could quibble with any definition of RELNEG. So it is important to note that, with one or two exceptions, other empirical studies have no measure of RELNEG whatsoever, which means that these studies are implicitly assuming that degree of defendant negligence is identical across cases.

7. The MEDICAL.BILLS variable is reported by the plaintiff. We expect that the larger the MEDICAL.BILLS, the larger the amount of the disagreement, given that a dispute takes place.

Previous research (e.g., Wittman, 1986) has suggested that juries award plaintiffs more when the defendant is a business or government. In order to test the deep pockets hypothesis, the variable DCORP.RNG.MED was created. Letting DCORP = 1 if the defendant is a business or government, and letting DCORP = 0, otherwise, DCORP was multiplied times RELNEG.MED, creating the new variable DCORP.RNG.MED.

Other things being equal, one would expect that juries (and arbitrators) would extract less from a defendant with deep pockets when there was a DISPUTE over MEDICAL.BILLS. Therefore, the variable DC.RNG.MED.DISP = DCORP.RNG.MED ∗ DISPUTE was also created.

3. The Econometric Models and Their Parameter Estimates

We are now ready to specify the underlying econometric model to be estimated. As is often the case, there is more than one possible way of specifying the model. I consider three versions here and several more in Appendix B. Fortunately, the results are very similar. The main reason for considering the various permutations is to show the robustness of the results and dispel the notion that still some other specification would make an important difference.

3.1. The Basic Jury Award and Arbitration Award Equations

The basic econometric model is closely related to the presentation of the data in section 2.

$$\begin{aligned}\text{JAWARD} = {} & \alpha_0 + \alpha_1\text{RELNEG.MED} + \alpha_2\text{RNG.MED.DISP} \\ & + \alpha_3\text{DCORP.RNG.MED} + \alpha_4\text{DC.RNG.MED.DISP} \\ & + \upsilon_\alpha + \varepsilon_\alpha \qquad (2)\end{aligned}$$

ε_α is the error term arising from the particular jury's idiosyncratic behavior. ε_α has a mean equal to 0. υ_α is the error term arising from factors observed by the jury but not by me. Examples of such factors include documentation of the plaintiff's pain and suffering, the particular nature of the damages, and the plaintiff's and defendant's demeanors when testifying. The reason for dividing the error term into two component parts will be made clear shortly.

Recall that RELNEG.MED is relative negligence times medical bills; it is our characterization of comparative negligence. Our a priori expectation is that juries will broadly follow the dictates of the law and award more to a plaintiff, the greater the defendant's relative negligence and the greater the damage to the plaintiff (where medical bills is the measure of damage). That is, our a priori expectation is that $\alpha_1 > 0$.[8] Since RNG.MED.DISP = RELNEG.MED $\times$ DISPUTE, we expect that $\alpha_2 < 0$ and that $|\alpha_2| < \alpha_1$. That is, the award will be less than otherwise if there is a DISPUTE over the medical bills, but the effect of a medical dispute will not be so large that greater medical bills lead to a lower award.

Essentially, DCORP.RNG.MED measures the change in the AWARD when the defendant is a business firm or government (since we already have RELNEG.MED). My a priori expectation (based both on previous empirical work and casual reading of the newspapers) is that juries tend to award more to the plaintiff when the defendant has deep pockets ($\alpha_3 > 0$). DC.RNG.MED.DISP = DCORP.RNG.MED $*$ DISPUTE. The jury's inclination to empty deep pockets should be tempered when there is a DISPUTE. That is, I expect that $\alpha_4 < 0$ and $|\alpha_4| < \alpha_3$.

I next consider the arbitration award equation:

$$\begin{aligned}\text{ARBAWARD} = {} & \beta_0 + \beta_1\text{RELNEG.MED} + \beta_2\text{RNG.MED.DISP} \\ & + \beta_3\text{DCORP.RNG.MED} + \beta_4\text{DC.RNG.MED.DISP} \\ & + \upsilon_\beta + \theta_\beta \qquad (3)\end{aligned}$$

I expect that the coefficients of the variables have the same sign as the corresponding coefficients in the jury award equation. In accordance with the arbitrator selection theory, I entertain the stronger hypothesis that $\alpha_0 = \beta_0$, $\alpha_1 = \beta_1$, and $\alpha_2 = \beta_2$. Now the theory of arbitrator selection also suggests that $\alpha_3 = \beta_3$ and $\alpha_4 = \beta_4$, but experience tells me otherwise. It is not uncommon for juries to really sock it to corporations in product liability, so I would expect juries to be more aggressive than arbitrators in emptying deep pockets, although arbitrators would still not be slouches in this regard. That is, my a priori expectation is that, $\alpha_3 > \beta_3 > 0$ and

8. The coefficients need not equal 1 since juries may account, on the one hand, for pain and suffering, which tends to make the coefficients greater than 1, and, on the other hand, for out-of-court compensation to the plaintiff (such as medical insurance), which tends to make the coefficients less than 1.

$\alpha_3 + \alpha_4 > \beta_3 + \beta_4 > 0$. Of course, in testing these latter hypotheses, I will be testing the arbitrator selection hypothesis at the same time.

Both the jury and the arbitrator have case information that is inaccessible to me. I have only a short summary of the case. The arbitrator and jury have access to much more information that becomes available through the trial itself (for example, the previously mentioned pain and suffering of the plaintiff may be presented as evidence in the trials). The arbitrator selection hypothesis suggests that the information would affect the arbitrator and the jury in the same way. That is, $\upsilon_\alpha = \upsilon_\beta$.

Therefore, one might want to consider using seemingly unrelated regressions in simultaneously estimating both the jury and arbitrator award equations. Unfortunately, the set of independent variables is the same in both cases; so, seemingly unrelated regression techniques are of no use. Nevertheless, it would be nice to get a *feel* for the joint variability of the jury and arbitrator awards not accounted for by our present set of independent variables. Therefore, I also undertake the following two-stage procedure: I use the error term (JURY.RESID) from the estimated first-stage jury award equation as an additional independent variable in the second-stage arbitration award equation and the error term (ARB.RESID) from the estimated first-stage arbitration award equation as an additional independent variable in the second-stage jury award equation. Obviously, adding ARB.RESID to the basic jury award equation will not change the coefficients of the original independent variables—the statement about seemingly unrelated regressions still holds.

The underlying rationale can best be understood by reference to an example. Suppose that, in a particular case, the plaintiff's body language indicates that she is lying. Both the jury and the arbitrator can observe this, but my data do not account for such a possibility. My first-stage results when I regress ARBAWARD against RELNEG.MED, RNG.MED.DISP, DCORP.RNG.MED, and DC.RNG.MED.DISP would overestimate the arbitration award. The actual minus the predicted arbitration award would be negative. For similar reasons, the actual minus the predicted jury award would also be negative (and by the same amount if the arbitrator selection hypothesis were true). I use the information from my first-stage arbitration regression in estimating my second-stage jury award equation. The residual terms from my first-stage arbitration regression are used instead of υ_α in the second-stage award regression.

Similarly, the residual terms from my first-stage jury regression are used instead of υ_β in the second-stage arbitration regression. If the arbitration selection model is true, then the coefficients of these variables are positive and equal to 1 in the limit as the other sources of error go to 0.

3.2. Viewing Expected Award as the Product of P and A

Recall that the expected jury award is $R = P * A$, where P is the defendant's relative negligence and A is the amount that the jury would award if the defendant's relative negligence were 100%. Instead of seeing R as the basic econometric equation, we can view R as the product of two underlying econometric equations—$A(w)$ and $P(z)$. Our econometric model of $A(w)$ is

$$A(w) = \alpha'_0 + \alpha'_1\text{MEDICAL.BILL} + \alpha'_2\text{MED.DISP} + \alpha'_3\text{DCORP.MED} + \alpha'_4\text{DC.MED.DISP} + \upsilon'_\alpha + \varepsilon'_\alpha. \quad (4)$$

And our econometric model of $P(z)$ is

$$P(z) = \text{RELNEG} + \nu_\alpha, \quad (5)$$

where ν_α, the error term, has a 0 mean and is independent of ε_α and the other independent variables in the award equation.

We do not observe $A(w)$ or $P(z)$ separately but only their product, the jury award:

$$\text{JAWARD}' = A(w)P(z) = \alpha'_0\text{RELNEG} + \alpha'_1\text{RELNEG.MED} + \alpha'_2\text{RNG.MED.DISP} + \alpha'_3\text{DCORP.RNG.MED} + \alpha'_4\text{DC.RNG.MED.DISP} + \text{RELNEG}\ \upsilon'_\alpha + \theta'_\alpha, \quad (6)$$

where θ is a marker for all the terms involving ε_α and ν_α.

By a similar process we can derive the arbitration award equation:

$$\text{ARBAWARD}' = \beta'_0\text{RELNEG} + \beta'_1\text{RELNEG.MED} + \beta'_2\text{RNG.MED.DISP} + \beta'_3\text{DCORP.RNG.MED} + \beta'_4\text{DC.RNG.MED.DISP} + \text{RELNEG}\upsilon'_\beta + \theta'_\beta. \quad (7)$$

The main difference between equations (2) and (3) and equations (6) and (7) is that RELNEG in equations (6) and (7) is substituted for the intercept terms in equations (2) and (3). I had two qualms about estimating such a model: (1) there is no intercept term, making the meaning and interpretation of the various statistics problematic; and (2) RELNEG appears one more time, increasing the possibility of multicollinearity. It happens that my qualms about multicollinearity were unfounded. Indeed, the coefficients for RELNEG.MED, RNG.MED.DISP, DCOPRP.RNG.MED, and DC.RNG.MED.DISP were very similar in the two sets of equations.

Once again a two-stage estimating procedure can be employed. In this case, the residual terms from the JAWARD equation are used instead of RELNEG υ'_β in the ARBAWARD equation, and the residual terms from the ARBAWARD equation are used instead of RELNEG υ'_α in the JAWARD equation.

3.3. Accounting for a Plaintiff Verdict

In the main body of the study, I regress award against the independent variables; I do not break the estimation down into two parts—the probability of a plaintiff verdict and the award given a plaintiff verdict. The rationale for this procedure is as follows: Under comparative negligence, greater relative negligence by the plaintiff reduces the award toward 0, so there is no need to distinguish between a verdict in favor of a defendant and a verdict in favor of the plaintiff. A change in the verdict from the former to the latter would not change the award (it would be 0 in either case).

This is in contrast to a system of negligence with contributory negligence, which in principle awards either full compensation for damages or nothing at all to the plaintiff. Under negligence with contributory negligence the plaintiff might have received a large award if the verdict had been in her favor, but did not because the verdict was in favor of the defendant. Such an outcome would have to be distinguished from the situation in which the verdict was in favor of the plaintiff but the award was nevertheless small because the harm was small. Under a regime of negligence with contributory negligence, a simple award regression (that including both defendant and plaintiff verdicts) would be inappropriate and other estimation techniques, which account for the probability of a plaintiff verdict, should be employed.

However, under comparative negligence, the award is not all or nothing, because greater negligence by the plaintiff reduces the award toward 0, so that standard regression is appropriate. In fact, in the data set there are numerous cases in which the award is very small (less than $200), and in a few cases the plaintiff received nothing even though the verdict was in favor of the plaintiff.

Despite the these arguments, in Appendix B I consider other estimating techniques that account for the probability of a plaintiff verdict. These include estimating the award conditional on a plaintiff verdict, the simultaneous estimation of the probability of a plaintiff verdict and the award given a plaintiff verdict, the sample selection of trials into a plaintiff verdict, and a Tobit equation. In Appendix Table 5 I report the estimated coefficients based on these other methods. What is most striking about these results is how similar they are to each other and the standard regression results reported in Tables 1–4.

3.4. Other Variations

In Appendix B I discuss other possible equations and estimating techniques, and then present the empirical estimates. These methods include using different measures of relative negligence and characterizing the model as a log equation. These variations yield very similar results.

4. Regression Results

We are now ready to look at our regression results. Recall that we use two different methods of estimating the jury and arbitration award equations (actually, we have two variations on two different equations). Looking at Tables 1–4, it is immediately apparent that the different methods of estimation produce very similar results—with one exception, the comparable coefficients are identical to one decimal point. Therefore, to avoid unnecessary detail when discussing particular coefficients, I will refer only to the first jury award and arbitration award regressions reported in Tables 1 and 2.

4.1. Parameter Estimates

Looking at Tables 1–4, we see in all of the equations the coefficient of RELNEG.MED is positive as predicted and statistically significant (the tables report the two-tailed test probabilities; the hypothesis concerning the coefficient of RELNEG.MED and most of the other hypotheses are one-tailed). Suppose that the defendant is not a corporation and there is no dispute about the medical bills. Looking at the first jury award

Table 1. JAWARD Equation

Variable	**Parameter**	**OLS T**	***p* > |T|**	**2SP T**	***p* > |T|**
Intercept	7,867.957599	2.274	.0235	3.364	.0009
RELNEG.MED	0.878445	2.211	.6407	3.271	.0012
RNG.MED.DISP	−0.333303	−0.467	.0276	−0.691	.4901
DCORP.RNG.MED	8.149804	7.974	.0001	11.793	.0001
DC.RNG.MED.DISP	−7.049640	−5.603	.0001	−8.287	.0001
ARB.ERROR	0.790006			20.899	.0001

Notes: OLS: F value: 25.957; prob > F: .0001; root MSE: 57,589; r^2; .2205. 2SP: F value: 132.769; prob. > F: .0001; root MSE: 38,938; r^2: .6446.

Table 2. ARBAWARD Equation

Variable	**Parameter**	**OLS T**	***p* > |T|**	**2SP T**	***p* > |T|**
Intercept	12,315.000000	3.812	.0002	5.638	.0001
RELENG.MED	0.826748	2.229	.0264	3.297	.0011
RNG.MED.DISP	−0.105780	−0.159	.8739	−0.235	.8145
DCORP.RNG.MED	6.358402	6.663	.0001	9.854	.0001
DC.RNG.MED.DISP	−5.409051	−4.604	.0001	−6.810	.0001
JURY.ERROR	0.688710			20.899	.0001

Notes: OLS: F value: 21.364; prob. > F: .0001; root MSE: 53,770; r^2: .1889. 2SP: F value: 124.742; prob. > F: .0001; root MSE: 3,6356; r^2: .6302.

Table 3. JAWARD Equation, No Intercept, with r^2 Redefined

Variable	**Parameter**	**OLS T**	***p* > |T|**	**2SP T**	***p* > |T|**
RELNEG	13,856.000000	2.571	.0105	3.793	.0002
RELNEG.MED	0.817022	2.048	.0413	3.020	.0027
RNG.MED.DISP	−0.319889	−0.449	.6536	−0.663	.5081
DCORP.RNG.MED	8.110550	7.949	.0001	11.725	.0001
DC.RNG.MED.DISP	−6.981654	−5.560	.0001	−8.201	.0001
ARB.ERROR	0.786203			20.797	.0001

Notes: OLS: F value: 29.847; prob. > F: .0001; root MSE: 57,477; r^2: .2891 2SP: F value: 126.204; prob. > F: .0001; root MSE: 38,966; r^2: .6742.

Table 4. ARBAWARD, No Intercept, with r^2 Redefined

Variable	**Parameter**	**OLS T**	***p* > |T|**	**2SP T**	***p* > |T|**
RELNEG	18,898.000000	3.746	.0002	5.526	.0001
RELNEG.MED	0.813026	2.177	.0301	3.211	.0014
RNG.MED.DISP	−0.083987	−0.126	.8527	−0.186	.8527
DCORP.RNG.MED	6.359100	6.658	.0001	9.821	.0001
DC.RNG.MED.DISP	−5.423037	−4.613	.0001	−6.805	.0001
JURY.ERROR	0.688946			20.797	.0001

Notes: OLS: F value: 30.279; prob. > F: .0001; root MSE: 53,805; r^2: .2891. 2SP: F value: 126.987; prob. > F: .0001, root MSE: 36,477; r^2: .6755.

equation, we see that if the plaintiff and defendant have been equally careless, then RELNEG is 1/2 and the plaintiff can expect to receive $0.44 for each additional dollar in medical bills. If the defendant's relative negligence is 100%, then the plaintiff can expect $0.88 for each extra $1.00 of medical bills. When the defendant does not have deep pockets, juries and arbitrators do not appear to compensate for pain and suffering.

In comparison to a situation in which there is no dispute, a dispute over the size of the medical bills should reduce the award. In all of the equations the coefficient of RNG.MED.DISP is negative as predicted, but in none of the equations is it significantly less than 0 (here again "$p > |T|$" should be divided by 2 so that the results are not so dismal as they first appear). Also in all of the equations, the absolute value of the coefficient of RNG.MED.DISP is less than the coefficient of RELNEG.MED, as predicted.

In all equations the coefficient of DCORP.RNG.MED is positive as predicted and statistically significant, meaning that juries and arbitrators empty deep pockets. The coefficients are economically significant, as well. Consider the case where RELNEG = 1 and there is no dispute over medical bills. Looking at the regression results in Table 1, we see that when the defendant is not a business or government entity the jury awards $0.88 for every $1.00 in medical bills; when the defendant is a business or government agency, then the jury awards $9.03 for every $1.00 of medical bills. Arbitrators also empty deep pockets, but by not as much. When the defendant is a business of government agency, then the arbitrator awards $7.18 for each $1.00 of medical bills.

In all of the equations the coefficient of DC.RNG.MED.DISP is negative as predicted and statistically highly significant. And in all of the equations, as predicted, the absolute value of the coefficient of DC.RNG.MED.DISP is less than the coefficient of RNG.MED.DISP and highly significant (all at the .0001 level). Thus, a dispute over the medical bills again lowers the amount awarded but not enough to cancel out the positive affect of medical bills on the amount awarded.

4.2. Consistency

Do different juries treat like cases alike? That is, are jury verdicts predictable, or do awards depend on the vagaries of the particular jury? A similar set of questions can be asked about arbitrators.

A serious problem is that the data are inherently biased against a finding of consistent behavior by juries (or arbitrators). Obviously, the data do not account for all of the legally relevant variables. Therefore, juries and arbitrators are likely to appear more arbitrary than they in fact are and would appear if we had access to the data available to them.[9]

Furthermore, it is the least predictable cases that are the most likely to go to trial. For an "open and shut" case, there will be little disagreement between the litigants about the trial outcome. Therefore, the case is likely to be settled out of court in order to avoid the extra cost of a trial. Since juries and arbitrators tend to try the unpredictable cases, they will appear to be less consistent than if all disputes ended in trial. This is an example of "case selectivity bias" because jury (arbitration) trials do not sample randomly from all disputes; rather they tend to select from those disputes in which the trial outcome is the least predictable.[10]

Finally, the most inconsistent arbitrators (those with the highest variance) are the most likely to have their cases retried by a jury. Thus, the set of cases that are retried by a jury tends to select from high-variance arbitrators.[11]

For all these reasons, juries and arbitrators will look much less consistent than they would if we had access to all the information they have and if all disputes, not just the subset of unpredictable cases, were tried.

A good measure of consistency is r^2. Sixty-four percent of the variation in the jury award and 63% of the variation in the arbitration award is explained by the independent variables (including the residual from the other equation). Whether these numbers are signs of consistency or inconsistency is open to subjective judgment. However, in light of the fact that the least predictable cases go to trial, these numbers appear to indicate considerable consistency.

4.3. Comparing Juries and Arbitrators

Do arbitrators and juries dispense the same type of justice? The vague concept of justice can be operationalized by breaking it down into

9. To some degree I have taken account of this problem by including the residual from one equation when estimating the other equation.

10. This analysis undermines Huber (1988) and others who criticize the tort system for being unpredictable. See Osborne (1999) for a counterargument.

11. Judges and lawyers with the highest variance will not be hired to arbitrate.

parts: the difference in the average award, the difference in the equation parameters, the difference in the percentage of plaintiff verdicts, and the difference in the unexplained variance.

Perhaps the easiest way to see the difference between jury and arbitration verdicts is to create a new variable, jury award minus arbitration award, and regress it against our set of independent variables (REL-NEG.MED, RNG.MED.DISP, DCORP.RNG.MED, and DC.RNG.MED.DISP). The results are reported in Tables 5 and 6.

If the arbitrator selection hypothesis worked to the fullest extent possible, then the hypothesis would predict that all of the coefficients (including the intercept) are 0, the average jury award minus arbitration award is equal to 0, and the correlation between the jury award and arbitration award is 100%.[12]

Unfortunately, all of this makes the usual interpretation of the tests of significance misleading because *both* the average and standard deviation equal 0 when the arbitrator selection hypothesis is true. Typically, one tests the hypothesis that a variable (or coefficient) is equal to 0 under the assumption that it has a strictly positive variance. The arbitrator selection hypothesis says that the variable JAWARD.MINUS.ARBAWARD should equal 0 and its standard deviation should equal 0; the hypothesis says nothing about the ratio of the average JAWARD.MINUS.ARBAWARD to its standard deviation. In particular, this ratio need not converge to 0 when both elements of the ratio converge to 0. To illustrate, the average jury award minus arbitration award is $-\$3,295$. This is 1.6 standard deviations from 0 (with an 11% level of significance). Suppose that each observation were 1/100000 as large so that the average differential would be $0.03. This would still be 1.6 standard deviations from 0 as the standard error would also be 1/100000 as large. The more effective the arbitrator

12. This is a high standard that most, if not all, economic theories would fail. For example, empirical studies do not uphold the law of one price. Classical hypothesis testing with the null hypothesis equal to 0 is not a very good way to test a theory, because a rejection leaves one with nothing and there is no metric on how different the coefficients are in comparison to how different they might be. See Wittman, forthcoming, for further elaboration.

Table 5. JAWARD.MINUS.ARBAWARD Equation

Variable	Parameter Estimate	Standard Error	T for $H0$: Parameter = 0	Prob. > |T|
Intercept	−4,446.610193	2,432.4936206	−1.828	.0684
RELNEG.MED	0.051697	0.27929615	0.185	.8533
RNG.MED.DISP	−0.227523	0.50175725	−0.453	.6505
DCORP.RNG.MED	1.791402	0.71862725	2.493	.0131
DC.RNG.MED.DISP	−1.640589	0.88462131	−1.855	.0645

Notes: F value: >1.864; prob. > F: .1161; root MSE: 40,490; r^2: 0.0199. JAWARD.MINUS.ARBAWARD Mean: −3,074.44. Test that all five coefficients = 0: F value: 1.92; prob. > F: .09. Test that first three coefficients = 0: F value: 1.60; prob. > F: .16.

Table 6. JAWARD.MINUS.ARBAWARD Equation, No Intercept, with r^2 Redefined

Variable df	Parameter Estimate	Standard Error	T for $H0$: Parameter = 0	Prob. > |T|
RELNEG	−5,041.792324	3,804.3711539	−1.325	.1859
RELNEG.MED	0.003996	0.28169574	0.014	.9887
RNG.MED.DISP	−0.235902	0.50280696	−0.469	.6392
DCORP.RNG.MED	1.751450	0.72031910	2.431	.0155
DC.RNG.MED.DISP	−1.558617	0.88651487	−1.758	.0796

Notes: F value: 1.598; prob; > F: .1598; root MSE: 40,577; r^2: 0.0213. Test that all five coefficients = 0: F value : 1.48; prob. > F: .22. Test that first three coefficients = 0: F value : 0.95; prob > F: .42.

selection hypothesis, the greater the simultaneous reduction in both the average JAWARD.MINUS.ARBAWARD and its standard deviation is likely to be. The variance of JAWARD.MINUS.ARBAWARD equals the variance of JAWARD plus the variance of ARBAWARD minus twice the covariance of JAWARD and ARBAWARD. The problem arises because the covariance increases as the impact of arbitrator selection increases.

To overcome the problem posed in the preceding paragraph, I also treat JAWARD and ARBAWARD as independent variables and find the difference between their means. In this case the overall variance is just the sum of the variances, which does not go to 0 when the arbitrator selection hypothesis holds perfectly. This test is a legitimate test of the arbitrator selection hypothesis. With this method, the difference between the JAWARD and ARBAWARD means is 0.74 standard deviations, which produces a significance level of 47%. So this evidence does not reject the arbitrator selection hypothesis.[13]

A similar problem to the one identified above affects our interpretation of the significance levels of the coefficients of the regression equation. When the arbitrator selection hypothesis holds, both the coefficients in the equation and their standard errors should be 0.

Again a possible solution is to treat the ARBAWARD and JAWARD equations as being independent and then test whether their coefficients are identical, rather than looking only at the jury award minus arbitration award results in Tables 5 and 6. Nevertheless, I consider just the results as reported in Tables 5 and 6, realizing that the t-statistics may overstate the differences.

Looking first at the individual coefficients for both equations, we see that the coefficients for RELNEG.MED and RNG.MED.DISP are not at all statistically significantly different from 0, that the coefficients for DCORP.RNG.MED and DC.RNG.MED.DISP are the most significantly different from 0, and that the intercept term is somewhere in between these two extremes of significance. The joint test that all five coefficients are equal to 0 can be rejected at the 9% level of significance for the regression reported in Table 5 but only at the 22% level for the regression reported in Table 6.[14] Clearly, in both regressions the significance levels are due to the different levels of response by juries and arbitrators to corporate defendants (as I had anticipated). The joint test that the first three coefficients (the intercept, RELNEG.MED and RNG.MED.DISP) are equal to 0 can be rejected only at the 16% and 42% level of significance for the regressions reported in Tables 5 and 6, respectively.

The one area where there is a clear difference between juries and arbitrators is the percentage of plaintiff verdicts. As can be seen in Table 7,

---

13. Unfortunately, this method ignores that the individual jury and arbitration verdicts are related. Therefore, I also undertook a nonparametric test that counted the number of times JAWARD.MINUS.ARBAWARD was positive and negative. Under the null hypothesis these numbers should be the same. But they were not. JAWARD.MINUS.ARBAWARD was much more often negative than positive (we will come back to this point shortly). The hypothesis that they were the same could be rejected at .001 significance level.

14. Because of missing values for some of the independent variables, the jury-award-minus-arbitration-award equation has slightly fewer observations than are recorded in Table 7. As a result, the reported levels of significance are somewhat different from those reported in Table 7.

Table 7. Case Characteristics

Variable	n	Mean	Std. Dev.
MEDICAL.BILLS	373	9,049.530	17,623.38
LIABILITY.ADMIT	383	0.3315927	0.4714013
RELNEG	381	0.6358268	0.1163167
MED.DISPUTE	379	0.2401055	0.4277120
DCORP	383	0.2245431	0.4178269
JAWARD	383	19,226.96	63,993.79
ARBAWARD	383	22,521.50	58,662.13
JAWARD.MINUS.ARBAWARD	383	−3,294.54	40,314.96
J.PLAINTIFF.VERDICT	383	0.7284595	0.4453359
ARB.PLAINTIFF.VERDICT	383	0.9843342	0.1243413
RELNEG.MED	373	5,768.26	11,793.02
RNG.MED.DISP	372	2,067.39	9,597.88
RNGCMED	373	1,738.81	8,903.77
DCORP.RNG.MED	373	1,738.81	8,903.77
DC.RNG.MED.DISP	372	962.87	8,463.25

Notes: Reported standard deviations are standard deviations of the variable itself and not of the average. J.PLAINTIFF.VERDICT, ARB.PLAINTIFF.VERDICT = 0 when the verdict is in favor of the defendant; J.PLAINTIFF.VERDICT, ARB.PLAINTIFF.VERDICT = 1, otherwise. LIABILITY.ADMIT = 1 when liability admitted by the defendant; LIABILITY.ADMIT = 0, otherwise.

arbitrators rule in favor of the plaintiff 98% of the time, whereas juries rule in favor of the plaintiff 73% of the time. The arbitrator selection theory predicts that arbitrators are more likely than juries to "split the difference" and rule in favor of the plaintiff. For example, if a jury has a 50% chance of awarding $1,000 and a 50% chance of ruling a defendant verdict, then the arbitrator will choose to award $500.[15] Splitting the difference reduces variance, which increases the probability of the arbitrator being chosen.

As already noted, the average JAWARD.MINUS.ARBAWARD equals $-\$3,295$. With an average jury verdict of $19,227 and an average arbitration verdict of $22,521, arbitration verdicts are on average 17% higher than jury verdicts. This is not an insignificant economic difference, but it is probably not much larger, if at all, than the transaction costs of going to a jury trial.

15. Bloom and Cavanagh (1986) have argued that labor arbitrators tend "to split the difference" in order to look fair and continue being hired. Splitting the difference also accounts for the results in our nonparametric test. Arbitration verdicts are more often greater than their related jury verdict than the reverse because there are numerous cases in which the jury awards nothing and the arbitrator awards something.

The correlation between jury awards and arbitration awards is 74%. While clearly not 100%, this correlation is nonetheless very high.[16] The similarities between arbitrators and juries appear to be stronger than their differences.

It should be realized that the nature of the data tends to exaggerate the differences between jury and arbitrator verdicts. The data set consists of those cases that were tried by both an arbitrator and a jury. Those arbitration cases for which the verdicts are the farthest away from the expected jury verdict are the most likely to be retried. If all arbitration cases were retried we would likely see an even greater similarity.

We next consider the issue of consistency. Suppose that the coefficients of the estimated jury award and arbitration award equations are identical; it is still possible that the error variance of one of the two equations is much larger than the error variance of the other. Under reasonable assumptions, the higher variance would indicate less consistency.[17] In determining whether juries or arbitrators are more consistent, we will consider two choices for a metric: comparing the root mean-square errors of the jury and arbitrator award equations and comparing their r^2. Looking back at Tables 1 and 2, we see little difference between arbitrators and juries in this regard—the jury equation has an r^2 of .6446 and a root mean-squared error of 38,937, and the arbitration equation has an r^2 of .6302, with a root mean-square error of 36,355.[18]

These data contradict Bernstein (1996), who argues that juries are “a disaster for the civil justice system” because they “undermine certainty.” Here, arbitrators are no more predictable than juries, even though the arbitrator selection theory suggests that arbitrators are chosen with the trait of consistency in mind (though the least predictable of these arbitrators tend to be in the data set). If there were only arbitrators and no juries, the impetus for appointing consistent arbitrators would likely be lessened.

---

16. The high correlation makes the standard error of the JAWARD.MINUS. ARBAWARD variable much smaller than it would be otherwise and is the reason for my alternative methods of testing.

17. One can make the opposite argument, however. Suppose that the unmeasured variables should affect the award, but that the arbitrator does not pay attention to them but the jury does. Then the jury will look less consistent.

18. When the arbitration residual was not included in the jury award equation and the jury residual was not included in the arbitration award equation, the r^2 and root mean-square errors were again very similar.

In sum, with the exception of the deep pockets issue, the jury and arbitrator equations are very similar regarding both their coefficients and their sum of squared errors. And even on the deep pockets issue the coefficients are reasonably close in terms of size (if not statistically so).

5. Conclusion

There has been a long and contentious debate about the relative merits of juries versus judges and legal professionals, and there have been numerous empirical studies. The work here differs in two respects from previous work.

First, I have data on cases that were actually tried by both a jury and an arbitrator.[19] Other studies compare one set of cases tried by judges and another set of cases tried by juries. The set of cases chosen are often quite dissimilar in terms of case category (e.g., there is a different mix of product liability and malpractice cases facing judges and juries). As Helland and Tabarrok (2000) have argued, a major explanation for the differences between juries and judges is that they try different categories of cases.[20] Of course, in these studies judges and juries may try cases with greatly different characteristics even within categories.

Second, in my data, set arbitrators are chosen within the shadow of the jury. To the degree that the arbitrator selection hypothesis is true, arbitrators and juries will have on average similar awards. This equivalence need not hold true if arbitration verdicts could not be retried by juries.[21]

Econometrics is the art of the possible. One can always raise statistical questions about any econometric study. However, the relative richness of

19. The classic work by Kalven and Zeisel (1966) asked judges what their decisions would have been if they, instead of the juries, had decided the cases. See Hans and Vidmar (1991) for a review of the subsequent literature.

20. See also Claremont and Eisenberg (1992), who show that differences between judge and jury vary by category.

21. Other studies have found similarites between judges and jurors. Eisenberg et al. (2000) show that there is no substantive difference between judges and juries regarding the rate at which they award punitive damages or the ratio of punitive damages to compensatory damages. A similar conclusion is reached by Robbennolt (2001). Wissler, Hart, and Saks (1999) surveyed judges, lawyers, and ordinary citizens on how they would have decided very short summaries of actual cases. They find considerable similarities between judges and "juries," but substantial differences between these two and defense lawyers.

the data set and the many results reported here should serve to increase our confidence in the data and the econometric models employed.[22] The text reports two equations each with four coefficients (plus the intercept) and numerous other statistics. Nearly all of these coefficients are consistent with our prior expectations regarding sign and magnitude. Across equation results are also consistent with our prior expectations (the tests are not all statistically independent, however). These empirical results are robust to alternative estimating techniques and specifications of the model (see Appendix B and accompanying tables).

To summarize the main results: Both arbitrators and civil juries follow the broad outline of the law. Both award more to plaintiffs with greater medical damages and both award less where these damages are in dispute. The one instance in which both bend the intent of the law is when the defendant is a business or government entity. "Deep pockets" are emptied by juries and somewhat less so by arbitrators. Consistent with the arbitrator selection hypothesis, arbitrators tend to split the difference and consequently are much more likely to find a verdict in favor of the plaintiff. Arbitrators tend to award more for small cases, whereas juries tend to award more for big cases in which the defendant is a corporation or government entity. Despite these differences (which are exaggerated because arbitration verdicts close to the expected jury verdict are less likely to be retried by a jury and be in this data set), the overall impression is one of great similarity between arbitrator and jury decisions. This similarity between jury and arbitration awards corroborates the arbitrator selection hypothesis.

Appendix A

This Appendix presents a formal model of arbitrator selection. Let R be the expected jury award. The litigants make informed predictions about the outcome of a trial. The generation of the litigants' expectations can be characterized by the following two equations:

$$R_p = R + \omega_p; \tag{A1}$$

$$R_d = R + \omega_d. \tag{A2}$$

22. Many other data sets have no information on case characteristics such as damages suffered by the plaintiff or plaintiff recklessness, and some studies have data only on who won, not on the size of the award.

ω_p is the plaintiff's error in estimating the expected jury award, R. We will assume that ω_p has a normal distribution with expected value, $E[\omega_p]$, equal to 0, and variance, σ^2.[23] Thus, R_p is the plaintiff's unbiased estimate of the jury award (that is, $E[R_p] = R$).

ω_d is the defendant's error in estimating R. We will assume that φ_d also has a normal distribution with expected value, $E[\omega_d]$, equal to 0, and variance, σ^2. Thus, R_d is the defendant's unbiased estimate of the jury award. ω_p and ω_d need not be independent; their covariance may be positive.

The litigants may have inconsistent estimates of R, either because their observations cannot be credibly conveyed to the other party, or because they have inconsistent priors.

Let $C = C_p + C_d$ be the cost to the litigants of going to trial.

Suppose that the litigants are risk neutral and that the litigants will always settle if the differential in their expectations is less than or equal to the cost of going to court. Then the following conditions would be necessary and sufficient for a case going to trial:

$$C < R_p - R_d = \omega_p - \omega_d.$$

Because of strategic bargaining and risk aversion (each working in the opposite direction), these conditions are, in general, neither necessary nor sufficient. Therefore, we will assume the more general formulation that the probability of a trial (weakly) increases as the expectation differential, $R_p - R_d = \omega_p - \omega_d$, increases and that the probability of a trial is 0 when $R_p \leq R_d$.[24] Since we are looking at cases that have gone through an arbitration trial, $R_p > R_d$.

B^i is arbitrator i's percent bias in favor of the plaintiff, which is known by all of the litigants and the court administrator.[25] If arbitrator i is not biased, then $B^i = 0$; if arbitrator i is biased against the plaintiff, then $B^i < 0$. Let ARBAWARD^i be the actual award by arbitrator i.

$$\text{ARBAWARD}^i = (1 + B^i)(R + \zeta_i). \tag{A3}$$

23. A normal distribution is not required. It just makes the formula for Bayesian updating simpler. A negative R_p would mean that the plaintiff expects to pay the defendant in a cross-complaint.

24. This latter assumption is not necessary, but it does simplify the proof.

25. A more complicated model could assume that the litigants are uncertain about the degree of bias. Such a consideration would add little to our understanding and make the analysis unnecessarily complicated.

Equivalently,

$$\frac{\text{ARBAWARD}^i}{1 + B^i} = R + \zeta_i, \tag{A4}$$

where ζ_i has a normal distribution with mean 0 and variance σ_i^2. That is, arbitrator i's actual choice involves a random idiosyncratic component that is independent of the facts of the case and unobservable by the other parties (to make the analysis simpler, we will assume that ζ_i is independent of ω_p and ω_d). The second formula will be referred to as the arbitrator's implied estimate of the jury award.

The litigants update their estimates of the expected jury award after they learn ARBAWARD^i.[26] That is, the litigants make use of the arbitration verdict in reestimating the outcome of the jury trial.

The formula for updating is

$$R_p^{\text{new}} = \frac{\sigma^2 \frac{\text{ARBAWARD}^i}{1+B^i} + \sigma_i^2 R_p}{\sigma^2 + \sigma_i^2} \tag{A5}$$

and

$$R_d^{\text{new}} = \frac{\sigma^2 \frac{\text{ARBAWARD}^i}{1+B^i} + \sigma_i^2 R_d}{\sigma^2 + \sigma_i^2} \tag{A6}$$

Proposition 1. Suppose that $B^i = 0$ for all i, but the variances differ across arbitrators and this is common knowledge. Then the court administrators and litigants will want to choose those arbitrators with the smallest variances.

Proof of Proposition 1: Recall that for cases going to arbitration, $R_p > R_d$. For any given B, the lower the variance of the arbitrator, the greater the weight that the litigants will place on the arbitrator's implied estimate of the jury award and the closer their *postarbitration* estimates of the jury trial will be. Therefore, the probability of a settlement will be higher. In particular, the lower the variance of the arbitrator's implied estimate, the greater the likelihood that the additional costs of a jury trial will outweigh the litigants' expectation differential. Therefore the litigants will be more

26. Evidence of this is found in the data. For example, the postarbitration, prejury trial plaintiffs' demands and defendants' offers are more highly correlated with the arbitration verdicts than their prearbitration demands and offers are correlated with the arbitration verdicts.

likely to avoid a jury trial by either accepting the arbitration verdict or settling afterwards.

The litigants would prefer to save on the cost of a trial and therefore, other things being equal, would choose to have a lower-variance arbitrator. The court administrator wants to minimize the number of jury trials; this is the rationale for setting up the arbitration system in the first place. Therefore, other things being equal, the court administrator will choose low-variance arbitrators. Q.E.D.

We next turn our attention toward the issue of bias. We will make the following additional assumptions.

If one side or the other does not accept the arbitration verdict, then the cost to each litigant of postarbitration negotiation is $0.5C^N$

If there is a settlement, S, we assume that a priori one side has no bargaining advantage, so that on average $S = (R_p + R_d)/2$ prearbitration and $S = (R_p^{\text{new}} + R_d^{\text{new}})/2$ postarbitration (given an average arbitrator bias of 0).

If the postarbitration negotiation is not successful, then the cost to each litigant of a jury trial is $0.5C^T$ (that is, each side bears half the cost of a jury trial) unless the jury award is worse than the arbitration award for the side requesting a jury trial. In which case, the full cost is shifted to the requester: $0 < C^N < C^T$.

Proposition 2. Given arbitrators with equal variance, a litigant will choose an arbitrator that is not biased over an arbitrator that is biased against the litigant.

Proof of Proposition 2: First, if the case goes to jury trial, then the expected trial cost to the litigant will be less when the arbitrator is not biased against her. Sometimes, the litigant will pay for the other litigant's trial cost (when the litigant rejected the arbitration verdict but did worse with the jury trial), sometimes the other litigant will pay for his or her cost, whereas at other times each will share the burden, but on average, the cost of trial to each litigant will be $0.5C^T$ when the arbitrator is not biased. In contrast, when the arbitrator is biased against one of the litigants, call her A, then A's expected trial cost is greater than $0.5C^T$ because A is more likely to reject the arbitration verdict and be unlucky enough to have the jury award be worse than the arbitrator award than for the other litigant, call him B, to reject the arbitration verdict and be unlucky enough to have the jury award be worse than the arbitrator award from B's perspective. Even if the case does not get to jury trial, A will be at a disadvantage in the negotiations and thus be willing to settle for less than otherwise.

The second reason for preferring an unbiased arbitrator over a biased one is that postarbitration trial costs will tend to be less. Despite the difference in bias, the arbitrators' decisions are equally as informative because their variances are the same. But the biased arbitrator's decision is more likely to be rejected because the biased arbitrator's verdict is less likely to fall within the parameters for acceptance by both parties. Costly postarbitration negotiations will then have to be undertaken even if none would have been needed if the arbitrator's verdict were unbiased.

Hence each litigant will want to eliminate the arbitrator biased against the litigant, in order to reduce the likelihood of unnecessary, but costly, postarbitration negotiations, as well as to reduce the likelihood of paying more than her share of court costs if the case goes to jury trial. Q.E.D.

Although the relative impact of these two propositions depends on the cost of a jury trial relative to the cost of negotiating a settlement after the arbitration verdict, and on the variation in the size of the bias relative to the variation in arbitrator variance, I suspect that Proposition 1 is more powerful in practice. That is, high variance arbitrators will be eliminated more than biased arbitrators, because high-variance arbitrators do not aid the settlement process.

Appendix B

Econometrics involves reasoned choices. However, not everyone will agree with my reasoning. Hence, in this Appendix I describe other possible specifications, argue for and against their use, and summarize the regression results. I report results when other samples, equation forms, and estimation techniques are used.

Different Definitions of Relative Negligence

In the main part of the article, relative negligence (RELNEG) was defined as defendant culpability/(defendant culpability + plaintiff culpability), where a value of 3 was ascribed to CULPABILITY for reckless driving; a value of 1, for appropriate driving; and a value of 2, for something in between. Now under comparative negligence, when the plaintiff is negligent and the defendant is not, the defendant is 0% liable, and when the defendant is negligent and the plaintiff is not, the defendant is 100% liable. One interpretation of this rule is that the defendant's relative negligence is 0 when DCULP = 1 and PCULP = 2 or 3, while RELNEG = 1 when DCULP = 2 or 3 and PCULP = 1 (otherwise, RELNEG conforms to the ratio defined earlier). I will call this the comparative negligence version of relative negligence. In comparison to the definition of RELNEG used in the main part of the article, this measure loses information and therefore I expected the results to be less precise.

A simpler, but still less precise approach, is to set RELNEG = 1 when the defendant admits liability, and RELNEG = 0.5 when the defendant does not admit liability.

Regressions using these alternative methods of measuring relative negligence are reported in Appendix Tables 1–4. As can be seen, with one statistically insignificant exception (the coefficient of RNG.MED.DISP for one of the arbitrator award equations), the results are similar in scale and sign to the results reported in Tables 1–4 in the main body of the article.

Alternative Procedures

In the main body of the article, I estimated the expected award equation directly and not as a product of its component parts (the probability of a plaintiff verdict and the award given a plaintiff verdict). I argued that under comparative negligence, greater relative negligence by the plaintiff reduces the award toward 0 so there is no need to distinguish between a plaintiff verdict and a defendant verdict. A change from a defendant verdict to a plaintiff verdict would not change the award (it would be 0 in either case).

Award, given a plaintiff win. The estimated equations, given a plaintiff win (Appendix Table 5, columns 1 and 5), look very similar to earlier estimates not conditioned on the plaintiff's winning (Tables 1–4). Nevertheless, in this Appendix, I will consider other statistical procedures for estimating the award, which account for the probability of a plaintiff verdict.

Appendix Table 1. JAWARD—Comparative Negligence Definition of RELNEG

Variable	Parameter	OLS T	*p* > |T|	2SP T	*p* > |T|
Intercept	8,625.503176	2.492	.0131	3.740	.0002
RELNEG.MED	0.701464	2.316	.0211	3.476	.0006
RNG.MED.DISP	−0.177686	−0.291	.7711	−0.437	.6624
DCORP.RNG.MED	5.845296	7.745	.0001	11.624	.0001
DC.RNG.MED.DISP	−5.275555	9.360	.0001	−8.045	.0001
ARB.ERROR	0.797393			21.464	.0001

Notes: OLS: *F* value: 24.315; prob. > *F*: .0001; root MSE: 57,993; r^2: .2009. 2SP: *F* value: 135.956; prob. > *F*: .0001; root MSE: 38,639; r^2: .6500.

Appendix Table 2. ARBAWARD—Comparative Negligence Definition of RELNEG

Variable	Parameter	OLS T	*p* > |T|	2SP T	*p* > |T|
Intercept	13,682	4.222	.0001	6.337	.0001
RELNEG.MED.BILLS	0.505242	1.782	.0756	2.674	.0078
RNG.MED.DISP	0.100248	0.175	.8608	0.263	.7925
DCORP.RNG.ME	4.344338	6.148	.0001	9.228	.0001
DC.RNG.MED.DISP	−3.598798	−3.906	.0001	−5.862	.0001
JURY.ERROR	0.698870			21.464	.0001

Notes: OLS: *F* value: 19.197; prob. > *F*: .0001; root MSE: 54,292; r^2: .1730. 2SP: *F* value: 126.734; prob. > *F*: .0001; root MSE: 36,174; r^2: .6339.

Appendix Table 3. JAWARD—RELNEG Based on Admitted Liability

Variable	Parameter	OLS T	*p* > |T|	2SP T	*p* > |T|
Intercept	10,353	2.953	.0033	4.433	.0001
RELNEG.MED	0.588946	1.491	.1369	2.237	.0259
RNG.MED.DISP	−0.325751	−0.564	.5731	−0.847	.3977
DCORP.RNG.MED	5.645002	6.953	.0001	10.437	.0001
DC.RNG.MED.DISP	−4.828384	−4.964	.0001	−7.451	.0001
ARB.ERROR	0.818859			21.470	.0001

Notes: OLS: *F* value: 20.473; prob. > *F*: .0001; root MSE: 58,977; r^2: .1824. 2SP: *F* value: 129.097; prob. > *F*: .0001; root MSE: 39,289; r^2: .6382.

Appendix Table 4. ARBAWARD—RELNEG Based on Admitted Liability

Variable	Parameter	OLS T	*p* > |T|	2SP T	*p* > |T|
Intercept	12,773	3.996	.0001	5.999	.0001
RELNEG.MED	0.927829	2.575	.0104	3.866	.0001
RNG.MED.DISP	−0.480695	−0.913	.3619	−1.370	.1714
DCORP.RNG.MED	4.261736	5.757	.0001	8.642	.0001
DC.RNG.MED.DISP	−3.345694	−3.772	.0002	−5.662	.0001
JURY.ERROR	0.680718			21.470	.0001

Notes: OLS: *F* value: 21.350; prob. > *F*: .0001; root MSE: 53,773; r^2: .1888. 2SP: *F* value: 130.678; prob. > *F*: .0001; root MSE: 35,822; r^2: .6410.

Unfortunately, there is little consensus on the appropriate estimating technique, and all contenders have serious shortcomings.

Insight into the issues surrounding estimation is best gained by first considering negligence with contributory negligence. Under a system of negligence with contributory negligence (NCN), the plaintiff receives all or nothing at all. If the plaintiff is not negligent and the defendant is, then the verdict is in favor of the plaintiff and the award fully compensates the plaintiff for her damages; otherwise, the verdict is in favor of the defendant. One could legitimately restrict the sample to plaintiff wins and regress the award against the independent variables. If cases won by the defendant were reversed, this would not alter the estimated coefficients, because there is no censoring of the dependent variable. Since the probability of a plaintiff verdict depends on only the litigants' levels of care and not on any of the variables effecting award (under NCN), the probability of a plaintiff verdict could be estimated independently. The estimated probability of winning would then be multiplied times the estimated award, given a plaintiff win, to generate the overall expected award.

Comparative negligence has a different configuration. Under comparative negligence, less care by the defendant increases the award given a plaintiff win. Thus, under comparative negligence, probability of a plaintiff win and the size of the award, given a plaintiff win, are not independent.

Jointly estimating probability of winning and the award. In order to account for this joint dependence, I reformulated the equations as follows. Let P be the probability of a plaintiff win.

$$p = e^{\gamma_0} + \gamma_1 Z + \varepsilon_\gamma \tag{A7}$$

$$\mathrm{LOG}(p) = \gamma_0 + \gamma_1 Z + \varepsilon_\gamma \tag{A8}$$

where Z = LIABAD, and the observed $p = .001$ or 1.001. LIABAD = 1 when the defendant admits liability, and LIABAD = 0, otherwise. Because LIABAD takes on only two values, probability of winning takes on only two values. Therefore one can use ordinary regression as a perfect substitute for probit.

The jury award equation is then

$$e^{\mathrm{JAWARD}} = p_0 e^{\alpha_0 + \alpha_1 \mathrm{RELNEG.MED} + \alpha_2 \mathrm{RNG.MED.DISP} + \alpha_3 \mathrm{DCORP.RNG.MED} + \alpha_4 \mathrm{DC.RNG.MED.DISP} + \varepsilon\alpha}, \tag{A9}$$

Appendix Table 5. Accounting for the Probability of a Plaintiff Verdict

	JAWARD				AAWARD			
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
Intercept	13,087	13,570	14,431	7,731	12, 508	12,401	39,772	21,426
	(2.852)	(5.244)	(1.425)	(2.468)	(3.828)	(5.736)	(0.375)	(9.576)
RELNEG.MED	1.051	0.885	1.048	0.848	0.821	0.824	0.765	0.685
	(2.051)	(3.399)	(2.062)	(2.509)	(2.201)	(3.338)	(.463)	(2.728)
RNG.MED.DISP	−0.591	−0.425	−0.598	−.208	−0.109	−0.106	−0.158	0.053
	(0.676)	(0.907)	(0.689)	(0.347)	(0.163)	(0.240)	(0.053)	(0.118)
DCORP.RNG.MED	8.538	8.013	8.515	7.969	6.351	6.356	6.212	5.731
	(6.934)	(11.951)	(6.923)	(9.326)	(6.620)	(10.012)	(1.453)	(8.860)
DC.RNG.MED.DISP	−7.420	−6.906	−7.391	−6.935	−5.395	−5.404	−5.196	−4.859
	(4.933)	(8.251)	(4.916)	(6.599)	(4.568)	(6.914)	(0.986)	(4.191)
Residual		0.789		0.803		0.709		0.687
		(21.539)		(17.30)		(21.54)		(20.837)
Other		2,936	−3,104	47,572		955.32	−935,729	36,402
		(4.577)	(0.148)	(22.471)		(0.367)	(0.257)	(27.114)

Notes: Columns 1 and 5: Award, given plaintiff verdict. Columns 2 and 6: Award and probability jointly estimated. “Other” is the predicted log (plaintiff verdict + .001) for juries and for arbitrators. “Residual” is the residual from the arbitration award equation for the jury award equation and the residual from the jury award equation for the arbitration award equation. Columns 3 and 7: The Heckman sample selection model (selection into plaintiff verdict): “Other” is the inverse Mills ratio. Columns 4 and 8: Estimation using Tobit. “Other” is sigma.

or

$$\begin{aligned}\text{JAWARD} = {} & \lambda\,\text{LOG}(p) + \alpha_0 + \alpha_1\text{RELNEG.MED} \\ & + \alpha_2\text{RNG.MED.DISP} + \alpha_3\text{DCORP.RNG.MED} \\ & + \alpha_4\text{DC.RNG.MED.DISP} + \varepsilon_\alpha. \qquad \text{(A10)}\end{aligned}$$

The equations were estimated using seemingly unrelated regressions. The results are reported in Appendix Table 5, columns 2 and 5. The coefficients are very similar to those obtained by considering only plaintiff wins (columns 1 and 4) and again are similar to the original estimates reported in Tables 1–4.

Sample selection into plaintiff wins. The award equation could be conceived as a sample selection into the set of cases where the plaintiff wins. There are two problems with this approach. One, this is not really a problem of censoring (in the conventional econometric sense); and, two, the selection variable, probability of winning, is not really independent. Nevertheless, I have estimated the JAWARD and ARBAWARD equations (see Appendix Table 5, columns 3 and 7), using the Heckit procedure. Letting J.P.VERDICT = 1 when the jury verdict is in favor of the plaintiff and J.P.VERDICT = 0, otherwise, I first ran a probit with J.P.VERDICT (or ARB.P.VERDICT) as the dependent variable and LIABAD as the independent variable. With the exception of the intercepts, the coefficients were quite similar to the coefficients in the other equations using different estimating techniques. In accordance with our argument, the coefficients of the inverse Mills ratio were insignificantly different from 0.

Tobit. Alternatively, one could employ a Tobit. There are problems with this approach, as well. The formulation assumes that the underlying variable can be negative, which does not seem sensible in the context of litigation.[27] The estimated coefficients are reported in Appendix Table 5, columns 4 and 8. With the exception of the intercept, the coefficients of the equations are again similar to the previous results.

27. Kessler (1995) views each case as being two cases—the plaintiff in one case is the defendant in the other, with both litigants having identical medical bills. Then the Tobit model would be appropriate. My data set is not composed of such pairs.

Appendix Table 6. Estimated Log Equation, LN.JAWARD

Variable	Parameter	OLS T	$p > \|T\|$	2SP T	$p > \|T\|$
Intercept	8.174823	8.657	.0001	8.744	.0001
L.RELNEG	7.049045	7.254	.0001	7.328	.0001
L.MED.DISP	0.697238	0.574	.5667	0.579	.5627
L.DCORP	0.713944	0.980	.3276	0.990	.3227
L.MEDICAL.BILLS	0.188763	1.851	.0649	1.870	.0623
ARB.ERROR	0.455705			2.908	.0039

Notes: OLS: F value: 14.200; prob. $> F$: .0001; root MSE: 4.0059; r^2: 0.1340. 2SP: F value: 13.282; prob. $> F$: .0001; root MSE: 4.0185; r^2: 0.1536.

Appendix Table 7. Estimated Log Equation, LN.ARB

Variable	Parameter	OLS T	$p > \|T\|$	2SP T	$p > \|T\|$
Intercept	7.166633	23.015	.0001	23.248	.0001
L.RELNEG	0.326346	1.019	.3091	1.029	.3042
L.MED.DISP	−0.735246	−1.834	.0675	−1.853	.0648
L.DCORP	0.169836	0.707	.4799	0.714	.4755
L.MEDICAL.BILLS	0.288003	8.566	.0001	8.652	.0001
JURY.ERROR	0.049551			2.908	.0039

Notes: OLS: F value: 19.085; prob. $> F$: .0001; root MSE: 1.33849; r^2: .1722. 2SP: F value: 17.269; prob. $> F$: .0001; root MSE: 1.32510; r^2: .1909.

Different Variables

Perhaps I have misspecified the true underlying equation. If we redefine MED.DISP = MED.DISP + 0.5 and DCORP = DCORP + 0.5, the following is a reasonable alternative specification:

$$\text{JAWARD} = \alpha_0[\text{RELNEG}]^{\alpha_1} * [\text{MEDICAL.BILLS}]^{\alpha_2} * [\text{MED.DISP}]^{\alpha_3} * [\text{DCORP}]^{\alpha_4} * e^{\varepsilon\alpha} \qquad \text{(A11)}$$

Essentially, this says that the variables are multiplied times each other. RELNEG is multiplied times the MEDICAL.BILLS. This amount is then increased by some percentage if the defendant is a corporation and reduced by some percentage if there is a dispute over the medical bills. Logs of both sides are taken and the resulting equation is estimated.[28]

The results are reported in Appendix Tables 6 and 7. As can be seen, the basic hypotheses are robust to this alternative specification. All of the

28. This equation is not a simple log transformation of our original equation in the main body of the article (such a transformation would be inappropriate). Therefore, one cannot use a Box and Cox test to ascertain whether the log version is a better form.

coefficients have the correct sign except MED.DISP in the LN.JAWARD equation (but it is not at all significant). Again, juries empty deep pockets more than arbitrators do.

These results are reasonably consistent with those of Viscusi (1986), who also used a log equation in his study of products liability. Looking only at products liability verdicts in favor of the plaintiff, he found a coefficient of 0.83 for MEDICAL.BILLS. Considering that defendants in products liability cases have deep pockets, this is within a reasonable range of 0.18, the coefficient of LOG (MEDICAL.BILLS) in Appendix Tables 6 and 7.

References

Ashenfelter, Orley. 1998. "Arbitration," in Peter Newman, ed., *The New Palgrave Dictionary of Economics and the Law*. London: Macmillan.

Bernstein, David. 1996. "Procedural Tort Reform: Lessons from other Nations," 19 *Regulation* 67–79.

Bloom, David, and Christopher Cavenagh. 1986. "An Analysis of the Selection of Arbitrators," 76 *American Economic Review* 408–22.

Clermont, Kevin, and Theodore Eisenberg. 1992. "Trial by Jury or Judge: Transcending Empiricism," 77 *Cornell Law Review* 1124–78.

Eisenberg, Theodore, N. LaFountain, B. Ostrom, D. Rottman, and M. Wells. 2000. "Juries, Judges, and Punitive Damages: An Empirical Study." Cornell Law School Working Paper.

Gross, Samuel, and Kent Syverud. 1991. "Getting to No: A Study of Settlement Negotiations and the Selection of Cases for Trial," 90 *Michigan Law Review* 319–93.

Hans, Valerie P., and Neil Vidmar. 1991. "The *American Jury* at Twenty Five Years," 16 *Law and Social Inquiry* 323–51.

Helland, Eric, and Alexander Tabarrok. 2000. "Runaway Judges? Selection Effects and the Jury," 16 *Journal of Law, Economics, & Organization* 306–33.

Huber, Peter W. 1988. *Liability: The Legal Revolution and Its Consequences*. New York: Basic Books.

Jury Verdicts Weekly. California.

Kalven, Harry, and Hans Zeisel. 1966. *The American Jury*. Boston: Little, Brown.

Kessler, Daniel. 1995. "Fault, Settlement and Negligence Law," 26 *RAND Journal of Economics* 296–313.

Osborne, Evan. 1999. "Courts as Casino? An Empirical Investigation of Randomness and Efficiency in Civil Litigation," 28 *Journal of Legal Studies* 187–204.

Robbennolt, Jennifer. 2001. “Punitive Damage Decision Making: The Decisions of Citizens and Trial Court Judges.” University of Missouri Law School Working Paper.

Shanley, M. G., and M. A. Peterson. 1983. *Comparative Justice: Civil Jury Verdicts in San Francisco and Cook Counties, 1959–1980*. Santa Monica: RAND.

Viscusi, W. Kip. 1986. “The Determinants of the Disposition of Product Liability Claims and Compensation for Bodily Injury,” 15 *Journal of Legal Studies* 321–46.

Wissler, Roselle, Allen Hart, and Michael Saks. 1999. “Decisionmaking about General Damages: A Comparison of Jurors, Judges and Lawyers,” 98 *Michigan Law Review* 751–826.

Wittman, Donald. 1986. “The Price of Negligence under Differing Liability Rules,” 29 *Journal of Law & Economics* 151–63.

———. 2002. “Normative Public Finance without Guilt: Why Normative Public Finance is Positive Public Finance,” in Stanley Winer and Hirofumi Shibata, eds., *Political Economy and Public Finance: The Role of Political Economy in the Theory and Practice of Public Finance*. Cheltenham: Elgar.