A Structural Model of Obstetrician Location and Treatment Decisions

Adam S. Hyde

November 2009

Department of Economics, University of Virginia. Mail: PO BOX 400182, Charlottesville VA 22904. Email: adamhyde@virginia.edu. Phone: (708) 574-5735. Fax: (434) 982-2904.

JEL Classification: H0, I1, J1, J6, K1 , L0, L5
Keywords: Healthcare, Defensive Medicine, Supplier Induced Demand, Malpractice, Cesarean Section, Obstetrician, Location Decision

Acknowledgements: I am grateful to Steven Stern, Amalia Miller, and John Pepper for all of their invaluable comments, advice and support over the years. I would also like to thank The Bankard Fund for Politcal Economy and The John M. Olin Program in Law and Economics for financial support and Thomson Reuters (formerly Medstat) and Robert Oshel of the National Practitioner Data Bank for their assistance obtaining data.

Abstract:
This paper develops and estimates a joint model of an obstetrician's choices of practice location, and treatments by cesarean section or vaginal birth. The model specifically incorporates the impacts of medical malpractice risk, expected health outcomes of patients, and earned income from procedures on obstetricians' decisions. Liability risk may affect an obstetrician's location choice and may cause unnecessary medical services to be supplied in an effort to reduce that liability risk, an action referred to as "defensive medicine." Outside of the obstetrician's incentive to insulate against this risk, there may be a profit motive for the obstetrician to induce demand for a particular method of delivery. The information asymmetry between obstetrician and patient can be exploited to provide the obstetrician with higher income or to protect against a potential lawsuit. Nonstructural results suggest that this is in fact a serious and potentially harmful concern. Specifically, I predict that, on average, the marginal effect of performing a cesarean decreases the probability of facing a lawsuit by 35% but increases the probability of a poor health outcome by 23%. Malpractice liability risk is measured by explicitly modeling the market for liability insurance using data on damages from the National Practitioner Data Bank (NPDB), insurance premiums from the Medical Liability Monitor (MLM), and legal expenses from the Physician Insurers Association of America (PIAA). Combining these data with the National Center for Healthcare Statistics' Birth Cohort Linked Birth/Infant Death Data from 1999 and obstetrician counts from the American Medical Association, a model of obstetrician choice is estimated and structural parameters of the obstetrician's utility function recovered. Estimation of this model illuminates how obstetricians choose locations based on the values of these parameters and allows comparisons of the endogenous distributions of these parameters across areas. Differences in the underlying parameters of obstetricians may help explain why treatment styles vary dramatically across areas. The recovery of these preference parameters enables me to simulate the effects of potential policy changes, such as changes to financial reimbursement schemes or certain tort reforms, on obstetrician behavior.

1 Introduction

A recent Department of Health and Human Services report declared the US medical liability system to be in crisis, claiming that the ever increasing costs of liability insurance are limiting patients' access to care, jeopardizing the quality of care, and contributing to healthcare costs that continue to grow faster than GDP.[1] Physicians can respond to increases in liability risk by relocating or retiring, both of which can cause potential shortages and limit access to care, but they can also respond by supplying medically unnecessary services in an effort to appear less negligent should the patient experience a negative health outcome. This action is commonly referred to as defensive medicine.[2] Defensive medicine can be considered one form of Supplier Induced Demand (SID). Information asymmetries between physicians and patients allow the physician to act as an imperfect agent by inducing demand for services for either defensive purposes or for financial gain.[3]

While liability fears and profit motive can cause these problems in all fields of medicine, my analysis will focus on obstetrics, an important specialty that is commonly thought to be among the hardest hit by the malpractice crisis.[4] A recent survey by the American College of Obstetricians and Gynecologists (ACOG) found that 65% of obstetricians responded that they had made changes to their practices because of the risk or fear of liability claims or litigation and that 37% had increased the number of their deliveries by cesarean section "because of risk of

[1]U.S. Department of Health and Human Services (HHS), "Addressing the New Health Care Crisis: Reforming the Medical Litigation System to Improve the Quality of Health Care." (2003) http://aspe.hhs.gov/daltcp/reports/medliab.pdf

[2] Defensive medicine is any treatment for which costs outweigh benefits and is prescribed by the physicians primarily to shield against liability risk. For more, see Kessler and McClellan (1996).

[3] I interpret the term "induced demand" as indicative of the patient utilizing a medical service that she would not otherwise consume if she had the same information and training as the physician.

[4] On April 6, 2004, Senate Majority Leader Bill Frist, M.D. stated on the Senate floor: "While the crisis affects all people seeking access to quality care, it affects those who are seeking help from high risk specialist physicians the most.…Our litigation system is increasingly forcing needed medical specialty doctors like neurosurgeons and obstetricians to…move to states not in crisis or to simply retire early from the practice of medicine."

liability claims or fear of being sued." [5] Cesarean sections are a more intensive form of treatment than a vaginal delivery, and therefore the obstetrician may appear to be less negligent in the event of an adverse health outcome. In addition to the potential insulation from liability risk, cesarean sections are, on average, reimbursed at higher rates than vaginal deliveries.[6] When these two facts combine with information problems, the physicians are left to weigh their own financial and legal benefits as well as the interests of the patient. But the question remains how heavily do each of these three factors weigh into treatment decisions?

Survey data and anecdotal claims, such as the above, recently have sparked intense debate among lawmakers and researchers about malpractice liability laws and the need for tort reform. Despite these beliefs, evidence on the magnitude and importance of both defensive medicine and labor supply responses in obstetrics remains mixed. The majority of the research to this point has focused on one particular source of variation and how that affects one specific outcome (e.g., do tort reforms lower the cesarean rate?), but, given that there are multiple behaviors through which an obstetrician may insulate against liability risk, it is not surprising that looking at any one of them alone does not provide clear evidence either way.

To solve this problem, I develop and estimate a structural model of obstetricians choosing a state in which to practice and subsequently making decisions of whether to treat patients by cesarean section or vaginal delivery. By endogenizing the location decision, my model allows for potential differences in the composition of obstetrician labor supply across states. Most importantly, I allow obstetricians to have heterogeneous preferences for both earned income and avoidance of liability risk. To my knowledge, this is the first research to model the mechanism

[5] http://www.acog.org/from_home/publications/press_releases/nr11-03-06.cfm
[6] I estimate that average reimbursement for Cesarean sections was $1,112 higher than a vaginal delivery using data from Thomson Medstat MarketScan database (2000), and the Nationwide Inpatient Sample (2000).

behind location choice and also the first to examine the joint decision of location and subsequent choices of treatment.

My analysis builds upon studies that have examined the effect of malpractice liability on physician labor supply. Using different measures of liability risk, Kessler, Sage, and Becker (2005), Baicker and Chandra (2004), Klick and Stratmann (2003), and Matsa (2007) all find that liability risk has a modest negative impact on the number of physicians practicing in a state.[7] Mello et al. (2007) finds this effect to be more pronounced when looking at obstetrics in particular, but none of these studies examine variation in the composition of labor supply. Physician labor supply is important to the extent that it affects not only cost and access to care, but also the quality of that care. Given the endogeneity of location choice, simply looking at the levels of physician labor supply may cause us to miss important variation in how physicians sort themselves based upon their preferences for avoiding liability risk. These differences may have important implications for many dimensions of quality but especially for physicians' propensities to practice medicine defensively.

Previous research examining cesarean sections also has largely ignored variation in the composition of labor supply. Using data from the National Center for Healthcare Statistics (NCHS) birth certificate data, Dubay, Kaestner, and Waidmann (1999) estimate that a $10,000 drop in malpractice premium translates to between a 0.0-0.4 percentage point drop in the primary cesarean rate,[8] but, without knowing the underlying preferences of obstetricians, it is unclear whether such a change would have different effects across locations, or even itself cause some relocation. Focusing on financial incentives, Gruber and Owings (1996) find some evidence that obstetricians smooth their incomes by performing more cesareans in the face of declining

[7] Kessler et al. (2005) estimate an increase of approximately 3% in physician labor supply in response to tort reforms such as damage caps.
[8] This is equivalent to a 23% drop in the premium translating to a 0.0-2.3% drop in the cesarean rate.

fertility rates but do not look at health outcomes as a result of the treatments or account for measures of liability risk. My research will extend both works by examining the mechanism affecting physician behavior as well as utilizing different measures of maternal and infant health outcomes.[9]

In a detailed micro data study, Grant and McInnes (2004) obtain similar (when aggregated) estimates to those of Dubay, Kaestner and Waidmann (1999). They find that personally experiencing a large claim significantly affects the future practice behavior of obstetricians. Their estimates indicate that obstetricians in Florida who experience an indemnity payout of average magnitude increase their risk adjusted cesarean rates by 1 percentage point. For small claims, however, they find that cesarean rates may actually drop.

More recently, Kim (2007) and Currie and MacLeod (2007) examine the effect of liability on obstetrics using data from birth certificates, and the National Practitioner Data Bank (NPDB), which collects data on amounts and circumstances surrounding medical malpractice payments. Kim (2007) does not find strong evidence that obstetricians respond to increased liability risk. However, Currie and MacLeod (2007) conclude that joint and several liability[10] reforms reduce cesareans as well as some complications during labor and delivery, but that caps on damages actually increase cesareans and complications. They conclude that, by lowering potential liability, caps on damages actually reduce the obstetrician's incentive to treat the patient carefully, leaving the obstetrician free to choose the more lucrative cesarean with less fear of facing a lawsuit. By allowing for differential impacts of different types of tort reform, their results provide a potential explanation for the mixed findings in the previous literature.

[9] Dubay et al. does examine infants' APGAR scores, but does not look at infant mortality, injury from birth, or any maternal health outcomes.
[10] Joint and several liability allows a patient to be compensated fully by any one of several defendants, and thus reforms can limit one's liability for the actions of others.

Another potential explanation is that obstetricians are choosing locations based upon their varied preferences for avoiding liability risk. If obstetricians are moving their practices to avoid potential liability, it is unclear whether one would expect an increase in liability risk to increase or decrease defensive medicine because of the relevant selection issues.

Consider a simple example where physicians in state X experience an increase in liability risk from L_{X0} to L_{X1}, but there is no similar increase liability risk in state Y, L_{Y0}. Physicians in X may react to this by treating their patients more defensively and staying in X, or they may respond by moving to Y. Defensive medicine in X may increase or decrease, depending on the magnitude of the difference between L_{X0} and L_{X1} relative to the change in the composition of preferences of physicians that choose to stay in X. Further, by simply looking for defensive medicine in X, researchers would miss a potential increase in defensive medicine in Y. If it is the case that $L_{X1} > L_{Y0} > L_{X0}$ then the physicians that relocate from X to Y may be more likely to engage in defensive medicine in Y than they initially were in X. In this simple example, researchers might find evidence of a labor supply response to liability risk but find no evidence of defensive medicine in state X, particularly when using behavior in Y as a control group for behavior in X. Selection has changed the composition of the labor supply in both states, a clear violation of the assumptions upholding difference-in-differences and similar models.

My estimation explicitly allows for this potential heterogeneity in obstetrician preferences. Using a nested logit framework, I estimate a structural model of obstetricians choosing locations and subsequent treatments that allows for heterogeneous preferences over avoiding liability risk, achieving good health outcomes for patients, and collecting revenue from procedures. I quantify malpractice risk by explicitly modeling the malpractice insurance industry as perfectly competitive and backing out the probability of facing a lawsuit for performing a particular treatment on a specific patient. Estimation of these patient specific probabilities takes place in stages. I estimate an instrumental variables model predicting treatment choices, health outcomes, and then the filing of lawsuits.

While estimation of these probabilities is a necessary step as an input for the utility maximization problem, the estimation by itself provides an interesting first look at how treatments are decided and how they impact health outcomes and lawsuit filings. I describe this simplified model and estimation in section 4. I compute the marginal effect of prescribing delivery by cesarean on the probability of experiencing a bad health outcome and the probability of filing a lawsuit should a bad health outcome occur. I estimate that on average prescribing a cesarean would increase the probability of a bad health outcome by 23% yet lower the probability of facing a lawsuit by 35%. While this effect is quite heterogeneous across patients, the difference indicates the there may exist mismatches in the incentives facing patients and obstetricians. These mismatches may cause the obstetrician to act as an imperfect agent and are consistent with a story of defensive medicine. To get a definitive picture of the mechanism driving these results I must examine the composition of obstetrician preferences in different areas, and estimate a full structural model of location and treatment choice together.

Using simulation methods, I estimate the distribution of obstetrician preference parameters based upon the outcomes from location and treatment decisions from the National Center for Healthcare Statistics' Birth Cohort Linked Birth/Infant Death Data from 1999 and state obstetrician counts from the American Medical Association. Estimation of this model illuminates how obstetricians choose locations based on the values of these parameters and allows comparisons of the endogenous distributions of these parameters across states. Differences in the underlying parameters of obstetricians across states may help explain why treatment styles can vary dramatically across areas.[11] The recovery of these preference parameters enables me to simulate the effects of potential policy changes, such as changes to the financial reimbursement system or certain types of tort reform, on obstetrician behavior.

2 Model

I model an obstetrician as a utility maximizing agent who derives utility from the expected health outcome of his patients and his income and experiences some disutility from having a patient file a lawsuit against him.[12] The obstetrician must first choose a state in which to practice and then decide how best to treat the patients he sees in that location. I model this as a nested discrete choice of treatments and location in a static framework. The obstetrician calculates the sum over patients of the utility he expects to receive from the optimal treatment of patients in each state and then chooses the state that maximizes his own utility. First, I will describe an obstetrician's choice of treatment for an individual patient, *i*, conditional on practicing in a particular state *k*, and then second, I will consider the choice of practice state.

2.1 Obstetrician Treatment Decisions

I will refer to the expected value of obstetrician *j* treating patients all of his patients in *k*, $I_k(\omega_j)$, as the inclusive value to obstetrician *j* from choosing to practice in *k*. The inclusive value is defined in terms of individual treatment decisions, where the utility to obstetrician *j* from treating patient *i* in state *k* by treatment *t* is written as

(1) $$u(E(H \mid X_i, t), Y_k(X_i, t), \omega_j, t, \varepsilon_{ijk}^t) = \omega_{1j} E[H(X_i, t)] + \omega_{2j} Y_k(t) + \omega_{3j} \alpha_k(X_i, t) + \varepsilon_{ijk}^t$$

where X_i represents a vector of patient demographic characteristics and medical risk factors, $Y_k(X_i, t)$ denotes the income that the obstetrician receives from treating patient *i* with treatment *t*

[11] In my sample, county (state) cesarean rates vary between 12% (14.7%) and 33% (27%)
[12] As it is the case that all obstetrics patients are female, I will refer to the patient in the feminine and obstetrician in the masculine to avoid confusion.

in state k, $E[H(X_i, t)]$ is the expected health outcome of patient i, and $\alpha_k(X_i,t)$ represents the obstetrician's belief about the probability of incurring a lawsuit, and is an extreme value error. For the sake of simplifying notation, I denote the three obstetrician specific coefficients, ω_{1j}, ω_{2j}, and ω_{3j} by the vector ω_j and write the utility function as the deterministic component and the error term,

$$u(X_i, \omega_j, t, \varepsilon_{ijk}^t) = \tilde{u}(X_i, \omega_j, t) + \varepsilon_{ijk}^t .$$

Because the only two options are treating with a cesarean section or a vaginal birth, the obstetrician compares the two utilities and chooses the treatment that maximizes his own utility. Conditional on the demographic characteristics and medical risk factors present at the time of birth, X_i, I model the probability of obstetrician j treating patient i by cesarean section as,

(2) $$\Pr(t = c \mid X_i, \omega_j) = \Pr[\tilde{u}(X_i, \omega_j, c) - \tilde{u}(X_i, \omega_j, v) > \varepsilon_{ijk}^v - \varepsilon_{ijk}^c].$$

With $\varepsilon_{ijk}^v - \varepsilon_{ijk}^c$ being the difference of two extreme value errors, the resulting probability is,

$$\Pr(t = c \mid X_i, \omega_j) = \frac{\exp\{\tilde{u}(X_i, \omega_j, c) - \tilde{u}(X_i, \omega_j, v)\}}{1 + \exp\{\tilde{u}(X_i, \omega_j, c) - \tilde{u}(X_i, \omega_j, v)\}} .$$

Because the individual obstetrician is not observed, I evaluate the mathematical expectation over the distribution of obstetrician characteristics in area k, $G_k(\omega)$. This probability, denoted by C, is a function of parameters of the $G_k(\omega)$ distribution, Ω and θ,

(3) $$C(\Omega, \theta \mid X_i) = \Pr(t = c \mid X_i) = \int \frac{\exp\{\tilde{u}(X_i, \omega_j, c) - \tilde{u}(X_i, \omega_j, v)\}}{1 + \exp\{\tilde{u}(X_i, \omega_j, c) - \tilde{u}(X_i, \omega_j, v)\}} dG_k(\omega_j) .$$

To obtain the inclusive value, I aggregate the utility from individual treatments to include all treatment decisions that the obstetrician expects to make in state k. The obstetrician does not know ahead of time exactly how many or which patients he will treat when choosing a state. Consequently, in addition to aggregating the treatment decisions, he must also evaluate the expectation over the distribution of patient characteristics in state k. Define W_k as a measure of potential market size in k and n_k as the number of obstetricians practicing in k. I assume that patients and obstetricians are randomly matched, that patients follow the treatment prescribed, and that the marginal valuation of market size is constant rather than diminishing.[13] Letting $F_k(X)$ be the distribution function of patient characteristics in area k, the inclusive value is

$$I_k(\omega_j) = \int E[\underset{t}{\text{Max}}\, u(X_i, \omega_j, t, \varepsilon_{ijk}^t)] dF_k(X_i)(\frac{W_k}{n_k}).$$

With ε_{ijk}^t as an extreme value error, the inclusive value has the functional form

$$I_k(\omega_j) = \int \ln[\sum_t \exp\{\tilde{u}(X_i, \omega_j, t)\}] dF_k(X_i)(\frac{W_k}{n_k}).$$

2.2 Obstetrician's Location Decision

The obstetrician must first choose a state in which to practice and then decide how best to treat the patients he sees in that location. Let U_{jk} represent total utility obstetrician j receives from practicing in area k,

(4) $$U_{jk} = I_k(\omega_j) + \theta Z_k + \eta_{jk}.$$

This is a function of I_k, the expected utility from making future utility maximizing treatment decisions for patients in k, some community characteristics, Z_k, and an extreme value error, η_{jk}. Obstetricians are allowed to vary in their preferences according to the three element parameter vector, ω_j: how much they value a patient's expected health outcome, ω_1; how much they value income from a procedure, ω_{2j}; and their disutility of facing a lawsuit, ω_{3j}.

Now I can define the probability that an obstetrician with a specific ω_j chooses to locate in area k as

[13] I examine the assumption of random matching of patients and obstetricians in more detail in Appendix B and conclude that search and matching do not have a large effect on the treatment that is chosen.

(5) $$\Pr(U_{jk} > U_{jk'} \forall k' \neq k \mid \omega_j) = \frac{\exp\{I_k(\omega_j) + \theta Z_k\}}{\sum_{k'} \exp\{I_{k'}(\omega_j) + \theta Z_{k'}\}}.$$

This has the familiar logit form for a discrete choice because of the assumption that the errors, η_{jk}, are i.i.d. extreme value. Because the number of obstetricians practicing in a particular location is observed but not the individual decisions, I evaluate the expectation of the probability of a specific obstetrician's choice over the distribution of all obstetricians' characteristics, $G(\omega)$. Defining Ω as the vector of parameters associated with $G(\omega)$, the theoretical share of obstetricians choosing k, $S_k(\Omega, \theta)$, is the unconditional probability that location k is chosen and is written as

(6) $$S_k(\Omega, \theta) = \Pr(U_{jk} > U_{jk'} \forall k' \neq k) = \int \frac{\exp\{I_k(\omega_j) + \theta Z_k\}}{\sum_{k'} \exp\{I_{k'}(\omega_j) + \theta Z_{k'}\}} dG(\omega_j).$$

A novel and crucial feature of the model is that the distribution of ω within state k, which I will denote as $G_k(\omega)$, is generated endogenously by the location decisions. The $G_k(\omega)$ distribution function depends on the exogenous national distribution of ω as well as the probability of location k being chosen conditional on a particular value of ω. The distribution function can be written as

$$G_k(\omega) = \frac{\int_{-\infty}^{\omega} \Pr(U_{jk} > U_{jk'} \forall k' \neq k \mid s) g(s) ds}{\Pr(U_{jk} > U_{jk'} \forall k' \neq k)} = \frac{\int_{-\infty}^{\omega} \frac{\exp\{I_k(s) + \theta Z_k\}}{\sum_{k'} \exp\{I_{k'}(s) + \theta Z_{k'}\}} g(s) ds}{\int \frac{\exp\{I_k(s) + \theta Z_k\}}{\sum_{k'} \exp\{I_{k'}(s) + \theta Z_{k'}\}} g(s) ds}.$$

2.3 Medical Liability Insurance Premium

Another contribution I make to this literature is to explicitly model the premium for medical liability insurance. I model the insurer's decision of the optimal premium taking as given the treatment decisions of obstetricians and the litigation decisions of patients. If the market for liability insurance is perfectly competitive and premiums are not experience rated, then each insurer would set the premium equal to the expected loss plus the difference between expenses and the expected income from investing the premium.[14] I model the malpractice premium offered by company in area k, M_k, as a function of the expected loss that the insurer faces for each of four different lawsuit dispositions: 1) the case is dropped or dismissed, 2) the case goes to trial and a verdict is returned for the defendant, 3) the case goes to trial and a verdict is returned for the plaintiff, and 4) the case is settled out of court.[15] Measures of the legal expenses by disposition have not been included in the past when modeling the premium.

Previous analyses of medical malpractice premiums, like Baicker and Chandra (2004), have ignored the costs of all lawsuits whether a payment is made or not. By omitting these costs, they are capturing only a portion of the expenses against which obstetricians are insured.[16] I include measures of average legal expenses by case disposition, which allows me to more accurately estimate the patient suit rates that the insurer expects the physician to face for different treatments. This provides a more precise measure of a physician's belief about exposure to risk than just the premium alone or indicators for reforms.

I allow the probability of each of these four dispositions to differ when the obstetrician chooses treatment by cesarean or vaginal birth. The resulting premium is then a function of the probability, π^t_{sk}, of each of the four suit dispositions, s, in state k when treatment t is chosen, and the average loss associated with each event. This loss includes the average legal expense, E_{sk}, and the average damage paid, D^t_{sk}, in k for a suit of disposition s when treatment t is chosen.

[14] There is evidence to suggest that there is only very limited, if any, experience rating. See Danzon (2000) and Sloan (1990) for further discussion.
[15] Outcomes 1 and 4 do not necessarily imply that an actual case has been filed. It is possible for a patient and a physician's insurer to begin an inquiry into an incident or to arrange a settlement without a case ever being filed with the legal system.
[16] In 2003, the average cost of defending a lawsuit that was ultimately dismissed or dropped was $17,408**,** but that number goes up to $123,000 for cases that the defendant ultimately lost. Data from the 2003 Physician Insurer's Association (PIAA) Data Sharing Report.

The damages are zero when the case is dropped, s = 1, or the verdict is for the defendant, s = 2. Defining λ_k as the fraction of births in area k delivered by cesarean section, μ_i as a geographically invariant but company-specific constant, and e_{ik} as an idiosyncratic error, I can write the premium charged by company i in state k as

(7) $$M_{ik} = \mu_i + \lambda_k \sum_{s=1}^{4} \pi_{sk}^{c} (D_{sk}^{c} + E_{sk}) + (1-\lambda_k) \sum_{s=1}^{4} \pi_{sk}^{v} (D_{sk}^{v} + E_{sk}) + e_{ik} .$$

The parameters of interest in equation (7) are the eight π_{sk}^{t} outcome probabilities. This implies that the obstetrician's probability of facing any lawsuit for a particular treatment can be represented as $\sum_{s=1}^{4} \pi_{sk}^{t}$. Given that the obstetrician does not explicitly have to pay D_{sk}^{t} or E_{sk} when defending a case, it could be the case that the obstetrician experiences the same disutility from a case regardless of disposition.[17]

3 Data

3.1 Data Sources[18]

The data on treatments and outcomes are from the National Center for Healthcare Statistics' Birth Cohort Linked Birth/Infant Death Data for 1999. These files are the universe of birth certificates in the United States and are matched with any subsequent infant deaths. These data have a great deal of information about each birth. In addition to some parental demographic characteristics, they contain information on maternal health and medical risk factors such as cephalopelvic disproportion, breech presentation, hypertension, diabetes, excessive bleeding, and

---

[17] Alternatively, if the disutility associated with defending a case is a time cost, then this should vary across the four dispositions. If this is the case, then equation (2) would need to be rewritten to allow each π_{sk}^{t} to enter the physician's utility function separately instead of through α_k.

[18] See Table 4 for overview of data sources.

if the mother has previously given birth by cesarean.[19] These data also contain records of events that occur during the delivery such as prolonged labor, tocolysis, induction of labor, method of delivery, and subsequent complications like fetal distress, birth injury, assisted ventilation, neonatal seizures, and infant mortality. Certain states[20] are omitted from the analysis because of data reporting problems, leaving 2,675,779 births across 44 states delivered by physicians in my sample.[21] Many of these observations exhibit identical risk factors, outcomes, and occur in the same locations, leaving 551,939 unique types of observations. Summary statistics of these data are presented in Table 1.

The American Medical Association (AMA) conducts a census of physicians by specialty that is available through the Area Resources File and includes the number of obstetricians in every county in 1999. These data show the number of obstetricians choosing to practice in location *k* in 1999 and also provide these numbers by age groups.

On the insurance side, the medical malpractice premiums for obstetricians, general surgeons, and internal medicine are available through the Medical Liability Monitor's Annual Rate Survey. They collect average premiums for at least three insurers in those three specialties by state, or, in some states, by county or grouping of counties. Data from malpractice liability insurers and their expenses associated with claims of different dispositions are taken from the Data Sharing Report from the Physician Insurers Association of America (PIAA). They collect data from member organizations detailing all claims but unfortunately will share only the

[19] For definitions of medical terms, see Appendix A

[20] Texas ,New York, and Nebraska are omitted because they lack some health outcomes and risk factor variables reported on birth certificates. Washington DC, Hawaii, Oklahoma, and South Carolina are omitted also because they lack adequate data for the model of liability insurance discussed in section 3.2.2

[21] This number does not include the roughly 7.7% of births delivered by midwives or certified nurse midwives. There may be some reliability issues with these data as only one attendant at birth is recorded. For example, if a birth was intended to be delivered by a midwife, but was subsequently transferred because of complications, that birth may not have a record of the midwife's involvement. Miller (2006) does not find that evidence that these midwives provide very different treatment from physicians, but does find that patients experience lower death rates.

nationally aggregated data. I estimate the legal expenses incurred by the insurer when defending each type of case in each state by adjusting the national expense estimates with state specific price indices.

The National Practitioner Data Bank (NPDB) has collected information on all payments made to patients on behalf of physicians for all states since 1990.[22] These data allow me to estimate the average damages paid for suits resulting from different treatments by either settlement or judgment in different states. When combined with the insurance market data these data allow me to estimate the probability of facing a lawsuit.[23]

Data on physician fees and reimbursement amounts is taken from Medstat MarketScan Data collected by Thomson-Medstat. These data are for private insurance companies and contain claims in 2000 for 27,027 births. For each hospital stay, these data include procedures, diagnoses, payment amounts, and payment sources. For each birth, the total payments are reported, and some observations have a breakdown detailing how much of the payment went to the obstetrician. Some estimation is required to construct measures of the income that obstetricians stand to receive for a cesarean and a vaginal delivery for all patients.

In addition to the Medstat data on privately insured patients, I use the Nationwide Inpatient Sample (NIS) from the Healthcare Cost and Utilization Project (HCUP) which collects data from patients regardless of insurance status. I use the NIS to estimate the relationship between charges to private insurance, Medicaid, and self pay patients, as well as the inflation of

---

[22] The NPDB has been criticized recently for the presence of a "corporate shield," which refers to the lack of a record of payments made on behalf of hospitals or other organizations instead of on behalf of individual physicians. The extent of this problem is not known, but is not problematic unless the awards "shielded" are systematically different than those brought against individuals.
[23] See section 4.1.2 for estimation methodology and results.

medical costs from 1999 to 2000.[24] In the 28 states represented in the NIS, there are 834,233 births in 2000 and 784,966 births in 1999.

To measure the attractiveness of different states I use data from the US Statistical Abstract of 2001 for two of the Z_k's in equation (5). I use the ten year average for annual precipitation in a state and total crime per 100,000 people in 1999. I also use the state's top tax rate on wages in 1999 as one of the Z_k's, which I take from the National Bureau of Economic Research website.[25]

3.2 Data Construction

The main empirical goal of this analysis is to estimate the parameters of the distribution of obstetrician preferences both nationally and the parameters of their endogenously determined distributions in each state. However, before estimating the structural model I must do some preliminary estimation to obtain the arguments in the obstetricians' utility from prescribing treatment t to patient i. Specifically, I must estimate the probability of facing a lawsuit and the financial reimbursement for each treatment.

3.2.1 Liability Risk

To measure the liability risk an obstetrician perceives from a particular patient, I first estimate the aggregate probability by modeling the insurance market. I then disaggregate those probabilities based upon the distribution of patient characteristics, treatments, and health outcomes in each state. In the process of estimating these patient specific probabilities, I conduct a nonstructural analysis of the dependence between these patient characteristics, treatments, health outcomes, and lawsuits. This analysis provides an interesting first look at this dependence

[24] The Medstat MarketScan data are not available to me in 1999. To estimate what the equivalent payments would be in 1999, I estimate a deflator of 0.901 from the NIS data.
[25] NBER Tax Sim data : https://nber15.nber.org/taxsim-temp/state-rates/maxrate.html

and produces some interesting results that highlight the possible mismatch in incentives facing obstetricians and patients. On average, an obstetrician can reduce the probability of facing a lawsuit by delivering by cesarean section, but on average, that would be harmful to the health of the patients.

3.2.2 State Aggregate Probabilities

The probability of facing a lawsuit of disposition *s*, resulting from treatment *t*, and in state *k*, is estimated from the premium data, legal expenses, and average damage awards. In order to estimate the lawsuit probabilities in the premium model in equation (7), I must estimate the expected loss to the insurer which in this case is the average damage payment.[26] I then use the estimated damages for each treatment in each state to estimate the probability of facing a lawsuit.

Even with M_{ik}, λ_k, D^t_{sk}, and E_{sk} all coming from data and prior estimation, the π^t_{sk} terms in equation (7) are identified only with the addition of two assumptions. I assume that cases of each disposition appear in the same relative proportions in all states for a particular treatment,

A1: $\pi^t_{sk} = a_s \pi^t_{1k}$, for $s = \{1,2,3,4\}$ and $t = \{c, v\}$,

where the a_s terms are nationally aggregated data from the PIAA. I assume that that these relative proportions are the same in all states. This allows all the π^t_{sk} for one treatment to be expressed in terms of π^t_{1k}. I also assume that cases resulting from each treatment appear in a state specific proportion across dispositions,

A2 : $\pi^v_{sk} = b_k \pi^c_{sk}$, for $s = \{1,2,3,4\}$.

[26] Estimation methodology and results are reported in Appendix C

Unfortunately the b_k terms are not as easily accessible from data as the a_s terms. The NPDB can be utilized to gain some insight, but the NPDB records only cases in which a payment was exchanged, which means cases that are dropped, or returned verdicts for the defendant, are not represented. Assuming that the b_k term is the same for all dispositions assumes that the proportion of cases in the NPDB resulting from cesareans vs. vaginal deliveries is the same for the observed outcomes as it is for the unobserved outcomes.

Combining A1 and A2 allows all the π^t_{sk}'s to be expressed in terms of π^c_{1k}. Equation (7) can now be rewritten as,

$$(8) \qquad M_{ik} = \mu_i + \lambda_k \sum_{s=1}^{4} a_s \pi^c_{1k}(D^c_{sk} + E_{sk}) + (1-\lambda_k)\sum_{s=1}^{4} b_k a_s \pi^c_{1k}(D^v_{sk} + E_{sk}) + e_{ik},$$

where $a_1 = 1$ and π^c_{1k} is the only probability parameter to be estimated. For some states the malpractice premium is reported at the state level, while for others it is reported for a county, or a group of counties. In these states, the malpractice premium offered by company *i* in state *k*, is a population weighted average of the premiums in all the counties or metropolitan areas in the state. The E_{sk} terms are the legal expenses associated with defending a particular type of case in each state and are constructed by weighting the PIAA national data on legal expenses by a state level price index.[27] Some states are omitted from the analysis because of insufficient data on one or more variables.[28]

Equation (8) is estimated using OLS and the estimated π^t_{sk}'s are reported in Table 4. With the π^t_{sk} terms estimated, I can now analyze them to see how they depend on the distribution of patient characteristics and risk factors. This simplified model of cesareans, health outcomes,

[27] State level price indices are constructed from housing price and non-housing price survey data compiled by Edgar Olsen. MSA level price indices are population weighted to form the state level indices.
[28] States where the MLM reports only one insurance company's price, and that company doesn't serve any other states cannot be used in this analysis. These states are: Washington DC, Hawaii, Oklahoma, South Carolina.

lawsuits and patient medical risk factors and demographic characteristics and estimation are described in Section 4.

3.2.3 Financial Incentives

In addition to the possibility of being driven by liability fears, differences in financial reimbursement may also play a role in obstetricians' treatment choices. Physicians are generally aware of a patient's insurance status and have a reasonable estimate of the fees that they stand to receive from a particular patient. If there is any difference in income from providing one treatment over another, then an obstetrician would clearly have an incentive to prescribe the treatment that maximizes his hourly wage, or a similar measure of dollars per unit of effort. To account for this, I estimate a parameter that allows the income from a cesarean to impact utility differentially than the income from a vaginal delivery.

Few studies have actually used true *prices* when examining physician incentives. Most use the amount a physician charges, but often that amount is different from what the insurance company or patient actually pays. The Medstat MarketScan data provide true prices for the privately insured, but the prices for patients with other payer sources are not observed. To account for alternate payer sources, I use the NIS data to estimate the ratios of private insurance charges to both Medicaid and self-pay patients. The charges reported in the NIS are not the total charges of the case, but instead have had professional fees removed prior to data collection. This means that the obstetrician's fees have been explicitly removed. However, this is not a problem for my purpose, as the charges are only used to estimate ratios. Under the assumption that obstetrician fees represent similar portions of the total charges across payer sources, absolute levels of these data are not important and these ratios are unbiased. The ratios of charges by payer source relative to private insurance are reported in Appendix C. At first glance, it seems surprising that charges to Medicaid patients are 16-17% higher than those to privately insured patients. This could be consistent with a selection story, in which Medicaid patients require more care at the time of delivery because they are on average less healthy or have received lower levels of prenatal care. A similar selection story might also explain why self-pay patients are charged 90-93% of the private patients' charges.[29]

Price endogeneity is a common problem in economics, frequently necessitating special estimation techniques to avoid bias. This will be a problem if there is something affecting both the treatment choice and either the level of prices for obstetrical services or the difference between the prices for cesarean and vaginal deliveries that is not captured in the model. One common candidate for such problems in the health literature is the omission of liability costs or fears, but I explicitly model their effect on both treatment and location decisions. Thinking about the supply curve for obstetrical services is of some use here as well. It's unclear that the supply curve necessarily slopes upward when thinking about increasing the number of cesareans versus vaginal deliveries. It seems more appropriate to think about the supply curve sloping upward when the relevant quantity is the number of patients a particular obstetrician is seeing. Increasing marginal cost is commonly one factor causing supply to slope upward, but, given that an obstetrician has agreed to take some number of patients, the treatment choice and its price for those patients might be completely unaffected by the treatments his other patients receive.

While the above stories have been about supply, thinking about demand is also useful here. Given the power of third party payers in healthcare, it is harder than in other markets to construct a story that fits the typical pattern of endogeneity. The insurance company would have to agree to reimburse more highly (relative to cost) for one treatment over another, which could

[29] This could also be consistent with other stories such as self-pay patients receiving discounts from hospitals or providers.

be possible if they thought it would improve health outcomes of their clients. However, this will not be a problem in my model because the health outcomes are allowed to affect the treatment choice explicitly.

4 A Simplified Model of Treatments, Outcomes, and Lawsuits

Before estimating the full structural model, it is useful to consider a simplified model that abstracts away from the location decision. I estimated the average probability of an obstetrician facing a lawsuit for a particular treatment in state *k*, in section 3.2.2, but, when an obstetrician estimates a particular patient's propensity to file a lawsuit, he can observe that patient's characteristics and risk factors. This allows him to refine his estimate further than simply an average probability. The utility an obstetrician derives from treating a particular patient, as specified in equation (1), depends upon the probability of getting sued conditional on a particular treatment and all of the patient's relevant characteristics.

To construct these patient-specific probabilities, I model the relationship between, treatment choice, health outcomes, lawsuits, and patient characteristics by combining the state average probabilities with individual data on treatments, patients, and health outcomes from birth certificates. This estimation is necessary as an input in the utility function, but this simplified model is useful and interesting in its own right. From this estimation, I can calculate the marginal effects of cesarean sections on the health outcomes and lawsuit probabilities as well as examine which subgroups are most likely to receive cesareans, experience bad health outcomes, and file lawsuits.

I estimate a three stage model of treatment choices, health outcomes, and lawsuit filings. First, the obstetrician makes a decision between cesarean section and vaginal delivery. Next, the patient experiences a health outcome either good or bad, conditional on the treatment choice.

Finally, the patient decides to file a lawsuit or not which is conditional on the occurrence of a bad health outcome and the treatment choice. The first two stages are relatively straightforward to estimate, but, because I do not have individual level data for lawsuits, the third stage requires more complex methods to estimate variables' effects on an individual patient's probability of filing a lawsuit.

4.1 Nonstructural Model

Let C_{ik}*, H_{ik}*, and S_{ik}* be latent variables indexing the propensity for patient i in state k to receive a cesarean, experience a good health outcome, and to file a lawsuit should a bad health outcome occur. C_i is a dummy variable equal to 1 when $C_{ik}^* > 0$, while $H_{ik}^* > 0$ (H=1) indicates a good health outcome, and $S_{ik}^* > 0$ (S_i=1) indicates that a lawsuit was filed as the result of a bad health outcome. Let their reduced form dependency be expressed as a system of three equations,

(10) $C_{ik}* = X_i\theta_{1c} + \theta_{kc} + Z_i\theta_z + \varepsilon_{ic}$

(11) $H_{ik}* = X_i\theta_{1h} + \theta_{kh} + C_i\theta_{2h} + (C_iX_i)\theta_{3h} + \varepsilon_{ih}$

(12) $S_{ik}* = X_i\theta_{1s} + \theta_{ks} + C_i\theta_{2s} + (C_iX_i)\theta_{3s} + \varepsilon_{is}$, where,

$$\begin{pmatrix} \varepsilon_{ic} \\ \varepsilon_{ih} \\ \varepsilon_{is} \end{pmatrix} \sim N(0,\Sigma), \quad \text{and} \quad \Sigma = \begin{bmatrix} 1 & \sigma_{ch} & \sigma_{cs} \\ \sigma_{ch} & 1 & \sigma_{hs} \\ \sigma_{cs} & \sigma_{hs} & 1 \end{bmatrix}.$$

X_i is a vector of demographic characteristics and medical risk factors, Z_i is the instrument in the equation (10), and all the θ's are parameters, with θ_k's representing fixed effects for each state k. The error terms are assumed to be draws from a mean zero Trivariate Normal distribution with covariance terms to be estimated.

In order for any type of lawsuit to be filed, an adverse health outcome, injury, or harm must occur. The birth certificates have several variables that capture bad health outcomes for mother and child. For the purposes of this analysis, I include anesthetic complications, birth injury, hyaline membrane syndrome, meconium aspiration, assisted ventilation for less than 30 minutes, assisted ventilation for greater than 30 minutes, neonatal seizures, and infant death.[30] In the following analysis, I define a bad health outcome as any one of these events occurring. While this is not an exhaustive list of bad health outcomes associated with labor and delivery, it is fairly comprehensive for infant related outcomes.

4.2 Nonstructural Estimation

One of the goals of this research is to examine the effect of cesareans on health and liability risk, but this treatment choice is almost certainly endogenous in one or both equations. For this reason, an instrumental variables approach is taken during estimation. I estimate the parameters of equations (10), (11), and (12) using an approach adapted from Heckman (1978) that accounts for the endogeneity of the treatment decision. Using Heckman's terminology, I estimate a multivariate probit model with structural shift.

The fraction of all other patients receiving a cesarean in patient *i*'s county is used as an instrument for whether a cesarean was performed for patient *i*.[31] This should be uncorrelated with unobservables specific to patient *i* that may affect the likelihood of a cesarean and also influence the health outcome, but correlated with whether or not patient *i* had a cesarean if, for example, obstetricians in particular areas are simply more likely to perform cesareans for any individual patient than in other areas.

I estimate the model in three stages. The first stage is a univariate probit model estimating the parameters of equation (10), the results of which are reported in Table 5. The second stage is a bivariate probit model that conditions on whether or not the patient received a cesarean in the first stage. The parameters of the equation (11) are estimated at this stage, as

[30] See Appendix A for definitions of medical terms.
[31] Sometimes the reference group is literally a county, sometimes it is a grouping of counties, and sometimes it is a state.

well as σ_{ch}, the covariance between the errors in equations (10) and (11). Results of the second stage are reported in Table 6.

Unfortunately, individual level data on lawsuits is not available as it is for treatments and health outcomes. The aggregate probabilities however, are available from the previous estimation of the premium equation. This makes the third stage less straightforward than the first two, and it is at this point that my analysis deviates from the traditional model of this type. I will define *S* as equal to 1 if any type of lawsuit occurs, implying,

(13) $$P(H=0,S=1\mid C=1,k) = \sum_{s=1}^{4}\pi_{sk}^{c}$$, and

(14) $$P(H=0,S=1\mid C=1,k) = \sum_{i}P(H_i=0,S_i=1\mid C=1,k,X_i)P(X_i\mid C=1,k).$$

Because I have specified the error structure as Trivariate Normal, I can rewrite equation (14) as,

(15) $$P(H=0,S=1\mid C=1,k)=\sum_{i}\frac{T(\hat{C}_{ik}*,-\hat{H}_{ik}*,\tilde{S}_{ik}(\theta_s);\Sigma)P(X_i\mid t,k)}{\Phi(\hat{C}_{ik}*)}$$ where,

$\hat{C}_{ik}*=X_i\hat{\theta}_{1c}+\hat{\theta}_{kc}+Z_i\hat{\theta}_z$,

$\hat{H}_{ik}*=X_i\hat{\theta}_{1h}+\hat{\theta}_{kh}+C_i\hat{\theta}_{2h}+(C_iX_i)\hat{\theta}_{3h}$, and

$\tilde{S}_{ik}(\theta_s)=X_i\theta_{1s}+\theta_{ks}+C_i\theta_{2s}+(C_iX_i)\theta_{3s}$.

With the parameters from equations (10) and (11) already estimated from the first two stages, the only parameters left to estimate are the σ_{cs} and σ_{hs} terms in Σ and the θ_s vector. In practice however, there are not enough observations to separately identify all of these parameters.

There are only two treatments and 44 states yielding 88 aggregate observations, which is too few to estimate all of the 115 θ_s parameters. A principal components approach however, can reduce the number of parameters to something manageable. I group all of X_i, and their C interaction terms together with the state fixed effects and determine their principal components. I define $\tilde{X}_i$ to be the vector containing the largest principal component of those variables and leave C outside the principal components to estimate a separate coefficient. In the second stage I estimated that $\hat{\sigma}_{ch}$ = 0.9996, which in the interest of cutting the number of parameters to estimate even further, leads me to the assumption that $\sigma_{ch} = 1$ which implies that $\sigma_{hs} = \sigma_{cs}$. These assumptions leave the optimization problem only three parameters to estimate. To estimate these I use non linear least squares and match the theoretical moments,

$$P(H = 0, S = 1 \mid C = 1, k) = \sum_i \frac{T(\hat{C}_{ik}*, -\hat{H}_{ik}*, \tilde{X}_i \tilde{\theta}_{1s} + C_i \tilde{\theta}_{2s}; \Sigma) P(X_i \mid t, k)}{\Phi(\hat{C}_{ik}*)},$$

to the empirical moments from equation (15). This yields the following optimization problem:

$$\min_{\tilde{\theta}, \Sigma} \{ \sum_{s=1}^{4} \pi_{sk}^{c} - \sum_i \frac{T(\hat{C}_{ik}*, -\hat{H}_{ik}*, \tilde{X}_i \tilde{\theta}_{1s} + C_i \tilde{\theta}_{2s}; \Sigma) P(X_i \mid t, k)}{\Phi(\hat{C}_{ik}*)} \}^2 \quad (16)$$

Results of this estimation are reported in Table 7.

4.3 Results and Counterfactuals from the Simplified Model

This estimation allows me to construct a fitted value for the obstetrician's estimated probability of a particular patient having a bad health outcome and subsequently filing a lawsuit conditional on patient risk factors, the treatment, and the state in which the birth occurs. Marginal effects for selected variables are presented in Table 8. These marginal effects are calculated by creating fitted probabilities for both outcomes for each patient, changing the variable of interest, differencing the probabilities, and dividing by the average to give a percentage change.

Of particular interest is the marginal effect of delivery by cesarean section on health outcomes and lawsuit probabilities. While these effects vary quite a bit across patients, on average, a cesarean raises the probability of a bad health outcome by 23%, yet lowers the probability of a lawsuit being filed by 35%. This is a somewhat striking result that highlights a mismatch in incentives that is potentially harmful to patients. If obstetricians' treatment decisions are influenced by the probability of facing a lawsuit then they may be harming patients physically, financially, or both.[32]

The simplified model alone allows the simulation of some interesting counterfactuals that begin to explore these issues. I examine what the model predicts would happen to cesarean rates, health outcomes, and lawsuits under two different decision rules for prescribing treatment. The results of these two counterfactuals appear in table 9.

<u>Decision Rule 1:</u> Always choose the treatment that maximizes expected health. If we forced obstetricians to choose treatments this way, cesarean rates in 1999 would not have been 23.3%, but instead would have fallen to 18.5%, expected health would increase from 0.85 to 0.89, and litigation rates would fall slightly as well.

<u>Decision Rule 2:</u> Always choose the treatment that minimizes the probability of facing a lawsuit. If obstetricians were deciding treatments in this way, cesarean rates and expected health outcomes would have remained virtually unchanged, but the probability of facing a lawsuit would be cut almost in half.

From these two simple experiments it is clear that obstetricians are allowing their decisions to be influenced by something other than the health outcomes of their patients. Observed cesarean rates and health outcomes are more closely approximated by the obstetrician

[32] I estimate that the payment an obstetrician receives for a cesarean section on average is nearly twice that of a vaginal delivery.

choosing solely based upon law suit probabilities, but because the probability of a lawsuit falls under the second decision rule, it must be the case that there is some substitution in who is being treated with a cesarean. One possible explanation is that their decisions are also being influenced by the financial reimbursement component that is missing from the simplified model.

While this exposes a potential for the practice of defensive medicine, it is not clear evidence that it actually occurs. Obstetricians' preferences over how much to weigh the potential liability risk against other factors is an important next step in understanding whether or not defensive medicine occurs and which areas may be particularly affected.

5 Structural Model Estimation and Identification

For each obstetrician, there are two types of decisions to be made, one of location and many of treatment. For each of the K states in the sample, the number of obstetricians is converted into the share of obstetricians practicing in state k, yielding K-1 = 43 moment conditions matching the predicted share of obstetricians choosing location k and the observed share of obstetricians,

$$E[\frac{n_k}{\sum_{k'} n_{k'}} - S_k(\Omega,\theta)] = 0 .$$

For each of the 551,939 types of births, there is a moment condition matching the predicted probability that patient i receives a cesarean and an indicator variable equal to one when the patient does receive a cesarean,

$$E[1(t_i = c) - C(\Omega,\theta \mid X_i)] = 0 .$$

These moment conditions are then weighted to reflect the original sample frequencies of the 2,675,779 births. Together the moments from location decisions and the treatment decisions can be written as a quadratic form to be minimized over the parameter vectors, Ω, θ. To construct these moments, conditional on a guess of the parameters, I simulate draws from the population distribution of preference parameters, ω. Using these draws, I then compute the theoretical shares in equation (6) and generate the distribution of ω in each state. Using this distribution, I simulate the ω's for obstetricians in that state and compute the treatment decision moments in equation (3).[33]

Identification comes from two helpful sources of covariation. The first is covariation in observed location decisions and the expected health, income and probability of facing a lawsuit associated with practicing in the area. The second is covariation in the indicator for a patient receiving a cesarean and differences in the expected health, income and lawsuit variables associated with a cesarean over a vaginal birth. The distribution of all the elements of ω will not be separately identified. I fix the mean of ω_1 to 1 and measure the others relatively.

6 Results

7 Policy Experiments

7.1 Financial Reimbursement

There are many useful policy experiments that can be conducted with the estimates from my model. With much current debate about the current financial structure of our healthcare system, one useful policy experiment is to alter the financial incentives for choosing treatments. For example, how would obstetricians' choices of locations and treatments differ if they faced a system which paid them on a per patient basis instead of by treatment? Using the recovered preference parameters, I can estimate changes to the geographic distribution of obstetricians, and subsequent cesarean rates in response to an altered reimbursement scheme. If supplier induced

[33] For a thorough description of the simulation algorithm, see Appendix C3

demand is driving some obstetricians to prescribe cesareans then the rate should fall in response to the removal of differential reimbursement.

Another interesting policy experiment might be to examine the effects of a pay-for-performance reimbursement scheme where physicians receive payment only if the patient experiences a good health outcome, or based on a risk adjusted performance scale.

7.2 Tort Reform

Another interesting policy experiment is to estimate the impact of different tort reforms. In my model, the probability that a patient files suit is being treated as the result of the patient's optimal behavior without that behavior being explicitly modeled. I can simulate the effect of changing lawsuit probabilities on obstetrician behavior in a partial equilibrium framework, but without modeling the litigation decision I cannot predict the general equilibrium impact of particular tort reforms.[34] Even focusing on just the partial equilibrium effects, there are many useful experiments that can be conducted. One such experiment is the implementation of no-fault compensation funds. This eliminates all legal expenses from the premium equations and might lower a physician's psychic cost of facing a lawsuit. These changes may alter the distribution of obstetricians across states and cesarean rates in those states. An interesting variant to this experiment is to compare the results when the fund is implemented at the state vs. the national level.

8 Conclusion

[34] At this point in time I am not aware of any data that would allow me to estimate such a model. Currently the best data available for analyzing malpractice cases is from the Florida Department of Financial Services, but unfortunately Florida's malpractice system for obstetrics is not representative of the rest of the United States. Since 1988, Florida has had a no fault compensation fund for infants that are neurologically injured.[34] Obstetricians can choose to participate in this fund and in the event of an injury a patient can receive compensation from the fund to cover related expenses to the injury. In order to receive compensation, the patient must forfeit the right to file a malpractice case. Virginia is the only other state to have implemented a similar program, and as such, behavior in either state likely differs substantially from that of the population in other states.

This paper has taken an important first step in understanding the complex problem of how Obstetricians make their decisions about where to locate their practices and how to treat their patients in those locations. While there are limitations, such as the use of a static model to approximate dynamic behavior, this study is the first to combine these choices of location and treatment, and provides a great deal of insight into a problem that has garnered much attention in recent years. Estimation of the simplified model by itself exposes what appears to be a striking contrast in the incentives of patients and their obstetricians: obstetricians do not always choose the treatment with the highest expected health outcome for their patients. When estimation of the full structural model is complete, I will be able to simulate counterfactuals and inform policy making decisions in a new and more robust way.

REFERENCES

American Medical Association, 2005. "America's Medical Liability Crisis," http://www.ama-assn.org/ama1/pub/upload/mm/399/mlr_tp.pdf.

Area Resources File (2001)

Avraham, Ronen. "Dataset of Tort Law Reforms 1980-2005," <http://www.law.northwestern.edu/faculty/profiles/RonenAvraham/>

Baicker, K., Chandra, A., 2004. "The Effect of Malpractice Liability on the Delivery of Healthcare." NBER Working Paper 10709.

Baldwin, L. et al. 1995. "Defensive Medicine and Obstetrics." *Journal of the American Medical Association* vol. 274, no. 20: 1606-1610.

Currie, J., and MacLeod, B., 2007. "First Do No Harm Tort Reform and Birth Outcomes? *Quarterly Journal of Economics*, 123(2): 795-830.

Danzon, Pauly, Kington 1990. "The Effects of Malpractice Litigation on Physicians' Fees and Incomes," *American Economic Review*, 80(2): 122-127.

Danzon, P.M., 2000, "Liability for Medical Malpractice," *Handbook of Health Economics* 1B: 1339-1404.

Dor, A., Koroukian, S., Grossman , M., 2004. "Managed Care Discounting: Evidence from the MarketScan Database" NBER Working Paper No. 10437.

Dubay, L., Kaestner, R.,Waidmann, T., 1999. "The Impact of Malpractice Fears on Cesarean Section Rates," *Journal of Health Economics* 18: 491–522.

Epstein, A., Ketcham, J., Nicholson, S., 2008, NBER Working Paper , w14070.

Grant,D. and McInnes, M. M., 2004. "Malpractice Experience and the Incidence of Cesarean Delivery: a Physician-level Longitudinal Analysis," *Inquiry*, 41, 2: 170-188.

Gruber, J., Owings, M., 1996. "Financial Incentives in Cesarean Delivery." *The RAND Journal of Economics*, 27(1): 99-123.

Heckman, J., 1978. "Dummy Endogenous Variables," *Econometrica*, 46, (4): 931-959.

Kessler, D.P., McClellan, M.B., 1996. "Do doctors practice defensive medicine?" *Quarterly Journal of Economics,* 111: 353–390.

-----------------, 1997. "The Effects of Malpractice Pressure and Liability Reforms on Physicians' Perceptions of Medical Care," *Law and Contemporary Problems* 60: 81–106.

----------------, 2002. “How Liability Law Affects Medical Productivity,” *Journal of Health Economics*, vol. 21, no. 6: 931-55.

----------------, 2002. “Malpractice Law and Health Care Reform: Optimal Liability Policy in an Era of Managed Care,” *Journal of Public Economics*, 84(2): 175-97

Kessler, D.P., Sage, W., Becker, D., 2005. “Impact of Malpractice Reforms on the Supply of Physician Services,” *Journal of the American Medical Association*, 293(21): 2618-2625.

Kim, Beomsoo, 2007. “The Impact of Malpractice Risk on the Use of Obstetrics Procedures.” *Journal of Legal Studies*, 37(S2).

Klick, J., Stratmann, T., 2003. “Medical Malpractice Reform and Physicians in High-Risk Specialties?” *Journal of Legal Studies*, 37(S2).

Localio A.R., et al, 1993. “Relationship Between Malpractice Claims and Cesarean Delivery,” *Journal of the American Medical Association*, vol. 269: 366-373.

Matsa, David A., 2007. “ Does Malpractice Libility Keep the Doctor Away? Evidence from Tort Reform Damage Caps.” *Journal of Legal Studies*, 37(S2)

Medical Liability Monitor, Annual Rate Survey 1999.

McGuire 2000. “Physician Agency,” *The Handbook of Health Economics* 1A: 462-536.

Miller, Amalia, 2006. “The Impact of Midwifery-Promoting Public Policies on Medical Interventions and Health Outcomes,” *B. E. Journal of Economic Analysis & Policy*: Vol. 6: No. 1, Article 6.

National Practitioner Data Bank. http://www.npdb-hipdb.hrsa.gov .

Physician Insurers’ Association of America (PIAA), 2003. Data Sharing Report.

Sloan, F.A. 1990. “Experience Rating: Does It Make Sense for Medical Malpractice Insurance,” *American Economic Review* 80(2): 128-33.

Studdert, D., et al. 2005. “Defensive Medicine Among High-Risk Specialist Physicians in a Volatile Malpractice Environment,” *Journal of the American Medical Association* vol. 293, no. 21: 2609-2617.

United States. General Accounting Office. “Medical Malpractice Insurance: Multiple factors have contributed to increased premium rates.” Washington, D.C.: U.S. General Accounting Office; 2003.

Table 1: NCHS Birth Certificate Summary Statistics

Variable	Mean	Std Dev	Min	Max
C-Section	0.222137	0.415683	0	1
Previous C-Section	0.110894	0.314001	0	1
Non-white	0.199141	0.399354	0	1
Age<19	0.119528	0.324409	0	1
Age>35	0.097851	0.297113	0	1
Intermediate prenatal care	0.173968	0.379082	0	1
Inadequate Prenatal care	0.046985	0.211608	0	1
Previous Child	0.588097	0.492178	0	1
Anemia	0.018599	0.135105	0	1
Cardiac Disease	0.004317	0.065563	0	1
Lung Disease	0.009013	0.094508	0	1
Diabetes	0.026365	0.16022	0	1
Herpes	0.00895	0.094181	0	1
Hydramnios	0.012543	0.111289	0	1
Hemoglobinopathy	0.000785	0.028004	0	1
Chronic Hypertension	0.007041	0.083614	0	1
Pregnancy Related Hypertension	0.037597	0.190219	0	1
Eclampsia	0.003174	0.056247	0	1
Incompetent Cervix	0.002566	0.050593	0	1
Previous 4000g+ infant	0.010966	0.104142	0	1
Preterm Labor	0.01233	0.110352	0	1
Renal Disease	0.002346	0.048377	0	1
Uterine Bleeding	0.005894	0.076547	0	1
Febrile	0.015097	0.12194	0	1
Meconium Staining	0.053834	0.225691	0	1
Ruptured membrane	0.023719	0.152171	0	1
Abruptio Placenta	0.00558	0.074488	0	1
Placenta Previa	0.003176	0.056264	0	1
Excessive Bleeding	0.004985	0.070429	0	1
Maternal Seizures	0.000339	0.018397	0	1
Breech Presentation	0.035128	0.184103	0	1
Cephalopelvic Disproportion	0.02129	0.14435	0	1
Cord Prolapse	0.001798	0.042367	0	1
Fetal Distress	0.040575	0.197304	0	1
Anesthetic Complications	0.000494	0.022214	0	1
Birth Injury	0.002803	0.052871	0	1
Hyaline Membrane Syndrome	0.005338	0.072867	0	1
Meconium Aspiration Syndrome	0.00196	0.044227	0	1
Assisted Ventilation < 30 min	0.01803	0.133059	0	1
Assisted Ventilation > 30 min	0.008493	0.091765	0	1
Neonatal Seizures	0.000597	0.024423	0	1
Infant Mortality	0.005487	0.073869	0	1
County C-section Rate	0.204915	0.031258	0.117	0.327

N=2,675,779

Table 2: Average Damage Awards and Settlements by State and Treatment

State	D^v_{4k}	D^v_{3k}	D^c_{4k}	D^c_{3k}
AL	60,650.85	85,232.11	148,735.86	83,677.50
AK	30,905.71	362,429.20	74,101.23	184,002.35
AZ	30,975.24	319,965.68	193,736.04	587,790.45
AR	46,384.17	611,438.47	118,564.90	290,570.00
CA	51,012.43	214,660.67	102,939.43	391,540.34
CO	43,008.15	237,820.85	312,520.66	365,924.40
CT	448,514.03	2,586,333.96	555,381.98	1,714,672.87
DE	911,115.81	58,859.48	44,708.86	192,031.15
DC	2,752,202.64	15,026,074.96	221,269.26	1,205,964.44
FL	73,408.14	256,151.26	184,763.66	1,652,877.18
GA	167,761.11	1,933,166.88	293,966.11	208,996.42
HI	53,397.27	48,199.09	247,693.91	47,342.22
ID	31,734.91	344,786.98	487,779.07	1,122,148.34
IL	85,192.71	950,347.22	263,504.31	505,674.97
IN	23,096.87	270,855.52	107,139.47	266,040.33
IA	249,047.57	951,987.42	168,262.36	856,472.09
KS	183,493.62	2,326,231.45	124,141.75	503,769.60
KY	362,336.15	1,638,929.77	298,247.34	422,867.14
LA	39,780.94	157,813.62	125,827.62	153,652.76
ME	57,940.78	201,828.19	548,117.06	404,281.94
MD	138,500.96	802,924.27	227,710.74	482,452.24
MA	373,737.74	9,637,062.53	418,263.75	871,414.35
MI	55,258.86	41,892.54	101,613.83	119,262.11
MN	153,964.31	3,313,200.63	427,098.68	1,440,083.20
MS	89,798.25	1,702,067.71	298,875.05	1,032,816.20
MO	83,330.22	645,780.58	233,745.11	2,509,782.95
MT	102,096.43	8,161,717.72	87,842.98	1,486,933.59
NE	41,423.50	87,501.17	235,121.77	185,234.86
NV	233,374.78	5,564,745.70	119,019.18	1,081,001.86
NH	71,446.82	1,729,404.76	519,894.97	5,412,937.61
NJ	107,152.30	141,966.01	287,474.14	769,069.16
NM	243,784.72	565,415.87	157,371.32	77,285.96
NY	130,118.46	525,602.29	246,592.79	196,213.15
NC	71,211.75	3,015,590.85	263,124.49	2,242,711.64
ND	234,587.81	124,804.36	1,147,176.47	407,180.01
OH	77,788.39	1,013,117.56	209,213.07	445,070.23
OK	10,703.01	61,073.40	160,700.94	194,168.19
OR	59,685.18	780,517.64	141,684.22	185,818.99
PA	15,729,003.00	478,001.84	223,486.42	183,985.46
RI	67,932.65	4,291.97	154,340.92	2,064,774.78
SC	103,527.72	535,644.33	203,167.06	103,681.29
SD	31,152.07	863,895.85	230,213.57	1,351,819.86
TN	466,591.00	516,344.26	281,326.90	66,294.12
UT	82,799.45	352,653.81	136,734.93	185,485.83
VT	49,305.30	104,089.25	223,402.35	99,865.05
VA	106,018.33	1,010,448.68	216,371.01	1,502,141.53
WA	611,168.48	2,875,390.74	244,007.31	333,167.61
WV	26,631.61	274,095.30	474,877.40	340,711.27
WI	87,015.81	190,572.11	260,824.59	4,208,727.04
WY	38,517.77	451,695.33	625,299.79	1,552,695.38

S=3 is a judgment — All numbers in 1999 Dollars.

S=4 is a settlement — N=1,840

Source: Author's calculations based upon National Practitioner Data Bank entries occurring in 1999

Table 3: Overview of Data Sources

Data Source	Unit of Observation	Enters Model Through	Notes
NCHS Birth Certificates	Individual Patient	$X_i, C_i, H_i(t)$	Medical Risk Factors, Demographic Characteristics, Treatments, and Health Outcomes.
Area Resource File (AMA Physician Masterfile)	County	S_k	Obstetrician counts by county, which I aggregate and convert to State shares.
Medical Liability Monitor	County, County Groupings, or States	M_{ik} in Premium Equation (8)	OB Premiums that I aggregate and average at the state level.
Physician Insurers Association of America	National Averages	E_{sk} in Premium Equation (8)	Average Legal Expenses by Disposition
National Practitioner Data Bank	Individual Claims against physicians	D^t_{sk} in Premium Equation (8)	Used to calculate Average Damage Payments in Malpractice for OBs by state, treatment, and disposition.
Medstat MarketScan Database	Individual Patient	$Y_k(t)$	Used to calculate average reimbursement by state and treatment
HCUP's Nationwide Inpatient Sample	Individual Patient	$Y_k(t)$	Used to calculate average reimbursement by state and treatment
US Statistical Abstract	State	Z_k	Total Crime, Average annual precipitation.
NBER TaxSim	State	Z_k	Top State Tax Rate on Wages

Table 4: Estimated Probability of facing a lawsuit by State and Treatment

State	$\sum_s \pi_{sk}^c$	$\sum_s \pi_{sk}^v$		State	$\sum_s \pi_{sk}^c$	$\sum_s \pi_{sk}^v$
AL	0.00154395	0.00315461		**MS**	0.00021367	0.00042646
AK	0.00192582	0.00176388		**MO**	0.00027459	0.00057412
AZ	0.00109892	0.00114803		**MT**	0.00011748	0.00022380
AR	0.00029344	0.00033481		**NV**	0.00031358	0.00075770
CA	0.00140053	0.00136044		**NH**	0.00009146	0.00013356
CO	0.00066969	0.00115393		**NJ**	0.00054740	0.00116132
CT	0.00033717	0.00036201		**NM**	0.00135917	0.00120714
DE	0.00174849	0.00170613		**NC**	0.00002566	0.00005050
FL	0.00059517	0.00093192		**ND**	0.00041012	0.00059225
GA	0.00037322	0.00087176		**OH**	0.00104543	0.00121784
ID	0.00030942	0.00024087		**OR**	0.00066273	0.00062866
IL	0.00131655	0.00150060		**PA**	0.00062606	0.00115053
IN	0.00024837	0.00035185		**RI**	0.00682863	0.00480529
IA	0.00041255	0.00040002		**SD**	0.00005937	0.00009532
KS	0.00055404	0.00106052		**TN**	0.00099637	0.00105852
KY	0.00068574	0.00081920		**UT**	0.00201692	0.00158634
LA	0.00137612	0.00380211		**VT**	0.00241451	0.00161731
ME	0.00102734	0.00112191		**VA**	0.00024807	0.00033554
MD	0.00106434	0.00151709		**WA**	0.00047958	0.00087726
MA	0.00029144	0.00014353		**WV**	0.00309315	0.00323994
MI	0.00349431	0.00509308		**WI**	0.00006802	0.00008905
MN	0.00008744	0.00013125		**WY**	0.00078047	0.00041243

Source: Author's calculations based upon estimation of Equation (8) using National Practitioner Data Bank entries occurring in 1999, Medical Liability Monitor Annual Rate Survey, and PIAA Data Sharing Report.

Table 5: First Stage Cesarean Section Outcome IV Estimation Results
Dependent Variable : C (C=1 denotes delivery by Cesarean Section)

Variable	Estimate		Std error
Non-White	0.0313	**	0.002678
Age<18	-0.0164	**	0.003694
Age>35	0.1477	**	0.003288
No Diploma	-0.0452	**	0.005584
Some College	0.051	**	0.002714
College Degree	0.0501	**	0.002769
Married	-0.0471	**	0.002602
Previous C-section	1.9911	**	0.002835
Intermediate Prenatal Care	-0.099	**	0.002903
Poor Prenatal Care	-0.205	**	0.005296
Anemia	-0.0315	**	0.007296
Cardiac Disease	-0.0184		0.014307
Lung Disease	-0.0025		0.010123
Diabetes	0.2979	**	0.005565
Herpes	0.5397	**	0.008356
Hydramnios	0.3534	**	0.007774
Chronic Hypertension	0.3996	**	0.010342
Pregnancy Hypertension	0.4683	**	0.004434
Eclampsia	0.7003	**	0.013391
Incompetent Cervix	0.0503	**	0.015568
Previous 4000g+ Infant	-0.2425	**	0.010328
Previous preterm or small for age	-0.3948	**	0.009413
Renal Disease	0.0472	**	0.019479
Uterine	0.0215	*	0.012075
Febrile	0.2314	**	0.006522
Meconium Staining	0.0147	**	0.004182
Uterine Rupture	0.068	**	0.005697
Abruptio Placenta	1.0325	**	0.009731
Placenta Previa	1.9024	**	0.014991
Excessive Bleeding	-0.0582	**	0.012225
Breech Presentation	2.1118	**	0.004528
Cephalopelvic Disproportion	2.976	**	0.009268
Cord Prolapse	1.3074	**	0.01551
Fetal Distress	1.2386	**	0.003848
Plurality	0.8995	**	0.004474
County C-Section Rate	3.3287	**	0.052106

Log Likelihood = -923,233.05
N = 2,675,779
*,** indicates significance at the 5%, and 1% levels
See Appendix A for definitions of medical terms.

Table 6: Second Stage Health Outcome IV Estimation Results

Dependent Variable : H (H=0 denotes a "bad" health outcome)

Variable	Estimate		Std error	Variable	Estimate		Std error
Non-White	-0.0289	**	0.006012	**Non-White x C**	0.0863	**	0.01143
Age<18	0.0048		0.006857	**Age<18 x C**	-0.036	**	0.016995
Age>35	-0.1326	**	0.008506	**Age>35 x C**	0.2373	**	0.013535
No Diploma	0.0376	**	0.011001	**No Diploma x C**	-0.0301		0.024478
Some College	-0.0442	**	0.005895	**Some College x C**	0.0691	**	0.011311
College Degree	-0.0393	**	0.006297	**College Degree x C**	0.0785	**	0.011987
Married	0.0439	**	0.005524	**Married x C**	-0.065	**	0.011028
Previous C-section	-1.5023	**	0.098179	**Previous C-section x C**	2.1136	**	0.098601
Intermediate Prenatal Care	0.0796	**	0.005672	**Intermediate Prenatal Care x C**	-0.1738	**	0.011952
Poor Prenatal Care	0.1462	**	0.008739	**Poor Prenatal Care x C**	-0.3536	**	0.020453
Anemia	0.0177		0.01358	**Anemia x C**	-0.1201	**	0.024464
Cardiac Disease	-0.0174		0.02629	**Cardiac Disease x C**	-0.1373	**	0.047599
Lung Disease	-0.0231		0.0177	**Lung Disease x C**	-0.1499	**	0.032018
Diabetes	-0.2786	**	0.014294	**Diabetes x C**	0.3808	**	0.020656
Herpes	-0.483	**	0.02621	**Herpes x C**	0.6774	**	0.037582
Hydramnios	-0.3508	**	0.017377	**Hydramnios x C**	0.3048	**	0.024291
Chronic Hypertension	-0.3676	**	0.028279	**Chronic Hypertension x C**	0.4363	**	0.036726
Pregnancy Hypertension	-0.4334	**	0.012038	**Pregnancy Hypertension x C**	0.5259	**	0.017064
Eclampsia	-0.6626	**	0.039372	**Eclampsia x C**	0.555	**	0.048149
Incompetent Cervix	-0.2751	**	0.022498	**Incompetent Cervix x C**	0.0408		0.040676
Previous 4000g+ Infant	0.159	**	0.017038	**Previous 4000g+ Infant x C**	-0.377	**	0.034492
Previous	0.1596	**	0.013064	**Previous preterm x C**	-0.5594	**	0.025621
Renal Disease	-0.0844	**	0.033619	**Renal Disease x C**	-0.0072		0.061345
Uterine	-0.1008	**	0.01982	**Uterine x C**	-0.1053	**	0.031812
Febrile	-0.2477	**	0.014064	**Febrile x C**	0.2019	**	0.025754
Meconium Staining	-0.1011	**	0.006784	**Meconium Staining x C**	-0.0458	**	0.016159

Table 6: Second Stage Health Outcome IV Estimation Results (Continued)

Dependent Variable : H (H=0 denotes a "bad" health outcome)

Variable	Estimate		Std error	Variable	Estimate		Std error
Uterine Rupture	-0.182	**	0.008668	**Uterine Rupture x C**	-0.0684	**	0.017208
Abruptio Placenta	-0.9851	**	0.029468	**Abruptio Placenta x C**	0.8381	**	0.034395
Placenta Previa	-1.3385	**	0.352266	**Placenta Previa x C**	1.6741	**	0.352884
Excessive Bleeding	-0.0044		0.021604	**Excessive Bleeding x C**	-0.2269	**	0.038093
Breech Presentation	-1.4812	**	0.149951	**Breech Presentation x C**	2.0455	**	0.150202
Cephalopelvic Disproportion	-0.2039	**	0.094613	**Cephalopelvic Disproportion x C**	1.4642	**	0.095356
Cord Prolapse	-1.2107	**	0.083914	**Cord Prolapse x C**	1.2768	**	0.089063
Fetal Distress	-1.1461	**	0.014411	**Fetal Distress x C**	1.1372	**	0.017955
Plurality	-0.879	**	0.013708	**Plurality x C**	0.9571	**	0.017316
				C	-1.1118	**	0.015118
				Sigma c,h	0.9993	**	0.008297

ln likelihood= -187,600.69
n= 2,675,779
*,** indicates significance at the 5%, and 1% levels
See Appendix A for definitions of medical terms.

Table 7: Third Stage Lawsuit Probability Estimation Results

Dependent Variable: $P(H=0, S=1 | C=1, k)$

Variable	Estimate		std err
Factor 1	-0.1586		-0.0553
C	-0.2201		-0.2245
Sigma c,s (h,s)	-0.9781	**	-3.2522

n=88
*,** indicates significance at the 5%, and 1% levels

Table 8: Selected Marginal Effects

Comparison	Pr(C=1) Mean % Difference	Pr(H=1) Mean % Difference	Pr(S=1|H=0) Mean % Difference
HS vs. No Diploma	-6.63	0.64	-16.08
HS vs. Some College	6.22	-0.3	3.63
HS vs. College	4.26	-0.02	28.52
Unmarried vs. Married	-2.37	0.19	11.42
White vs. Non-White	3.8	0.13	-21.06
18-35 vs. <18	-2.17	-0.21	-19.11
18-35 vs. >35	19.98	-1.01	19.61
Vaginal vs. C-Section	-	-23.08	-35.02

$$\text{A vs. B} = \frac{1}{n}\sum_{i=1}^{n} \Pr(:|\text{B})_i - \Pr(:|\text{A})_i$$

Table 9: Effects on C,H,S of Counterfactual Decision Rules for Treatment

Decision Rule	Pr(C=1)	Pr(H=1)	Pr(S=1)
Baseline	0.2339	0.8566	0.000597
Max{Pr(H=1}	0.1852	0.891	0.000500
Min{Pr(S=1)}	0.2386	0.8559	0.000335

Appendix A: Definitions of medical terms

Maternal Risk Factors
Anemia--Hemoglobin level of less than 10.0 g/dL during pregnancy or a hematocrit of less than 30 percent during pregnancy.
Cardiac disease--Disease of the heart.
Acute or chronic lung disease--Disease of the lungs during pregnancy.
Diabetes--Metabolic disorder characterized by excessive discharge of urine and persistent thirst; includes juvenile onset, adult onset, and gestational diabetes during pregnancy.
Genital herpes--Infection of the skin of the genital area by herpes simplex virus.
Hydramnios/oligohydramnios--Any noticeable excess (hydramnios) or lack (oligohydramnios) of amniotic fluid.
Hemoglobinopathy--A blood disorder caused by alteration in the genetically determined molecular structure of hemoglobin (for example, sickle cell anemia).
Hypertension, chronic--Blood pressure persistently greater than 140/90, diagnosed prior to onset of pregnancy or before the 20th week of gestation.
Hypertension, pregnancy-associated--An increase in blood pressure of at least 30 mm Hg systolic or 15 mm Hg diastolic on two measurements taken 6 hours apart after the 20th week of gestation.
Eclampsia--The occurrence of convulsions and/or coma unrelated to other cerebral conditions in women with signs and symptoms of pre-eclampsia.
Incompetent cervix--Characterized by painless dilation of the cervix in the second trimester or early in the third trimester of pregnancy, with prolapse of membranes through the cervix and ballooning of the membranes into the vagina, followed by rupture of membranes and subsequent expulsion of the fetus.
Previous infant 4,000+ grams--The birthweight of a previous live-born child was over 4,000 grams (8 lbs 13 oz). *Previous preterm or small-for-gestational-age infant*--Previous birth of an infant prior to term (before 37 completed weeks of gestation) or of an infant weighing less than the 10th percentile for gestational age using a standard weight-for-age chart.
Renal disease--Kidney disease.
Rh sensitization--The process or state of becoming sensitized to the Rh factor as when an Rh-negative woman is pregnant with an Rh-positive fetus.
Uterine bleeding--Any clinically significant bleeding during the pregnancy, taking into consideration the stage of pregnancy; any second or third trimester bleeding of the uterus prior to the onset of labor.

Complications
Febrile--A fever greater than 100 degrees F. or 38 C. occurring during labor and/or delivery.
Meconium, moderate/heavy--Meconium consists of undigested debris from swallowed amniotic fluid, various products of secretion, excretion, and shedding by the gastrointestinal tract; moderate to heavy amounts of meconium in the amniotic fluid noted during labor and/or delivery.
Premature rupture of membranes (more than 12 hours)--Rupture of the membranes at any time during pregnancy and more than 12 hours before the onset of labor.
Abruptio placenta--Premature separation of a normally implanted placenta from the uterus.
Placenta previa--Implantation of the placenta over or near the internal opening of the cervix.
Other excessive bleeding--The loss of a significant amount of blood from conditions other than abruptio placenta or placenta previa.
Seizures during labor--Maternal seizures occurring during labor from any cause.
Precipitous labor (less than 3 hours)--Extremely rapid labor and delivery lasting less than 3 hours.
Prolonged labor (more than 20 hours)--Abnormally slow progress of labor lasting more than 20 hours.
Dysfunctional labor--Failure to progress in a normal pattern of labor.
Breech/malpresentation--At birth, the presentation of the fetal buttocks rather than the head, or other malpresentation.
Cephalopelvic disproportion--The relationship of the size, presentation, and position of the fetal head to the maternal pelvis prevents dilation of the cervix and/or descent of the fetal head.
Cord prolapse--Premature expulsion of the umbilical cord in labor before the fetus is delivered.
Anesthetic complications--Any complication during labor and/or delivery brought on by an anesthetic agent or agents.
Fetal distress--Signs indicating fetal hypoxia (deficiency in amount of oxygen reaching fetal tissues).

Appendix A: Definitions of medical terms (continued)

Infant Health Problems

Anemia--Hemoglobin level of less than 13.0 g/dL or a hematocrit of less than 39 percent.

Birth injury--Impairment of the infant's body function or structure due to adverse influences that occurred at birth.

Fetal alcohol syndrome--A syndrome of altered prenatal growth and development occurring in infants born of women who consumed excessive amounts of alcohol during pregnancy.

Hyaline membrane disease/RDS--A disorder primarily of prematurity, manifested clinically by respiratory distress and pathologically by pulmonary hyaline membranes and incomplete expansion of the lungs at birth.

Meconium aspiration syndrome--Aspiration of meconium by the fetus or newborn, affecting the lower respiratory system.

Assisted ventilation (less than 30 minutes)--A mechanical method of assisting respiration for newborns with respiratory failure.

Assisted ventilation (30 minutes or more)--Newborn placed on assisted ventilation for 30 minutes or longer.

Seizures--A seizure of any etiology.

Appendix B: Random Assignment of Patients and Obstetricians

In the model, an obstetrician chooses a location based on the expected utility from treating the patients in that location. Because the identity of the obstetrician treating a particular patient is unknown, the model assumes that patients are randomly assigned to an obstetrician. In reality, this is certainly not the case, but what is of importance is only whether the process by which patients and obstetricians match with one another affects the treatment decision in any systematic way. Therefore, the legitimacy of this assumption must be explored at greater depth.

Epstein, Ketcham, and Nicholson (2007) investigate many aspects of sorting and matching in obstetrics, including whether patients with preferences for cesareans are cared for by physicians with a higher propensity to perform cesareans for all patients. They examine the patients and physicians of group practices in New York and Florida from 1999-2004. They do not find evidence that this is the case, although they do find evidence that suggests patients who appear to be at high risk for receiving a cesarean may be seen by the members of the practice who are more skilled in that delivery method.

The key source of variation in their analysis comes from the assumption that patients who labor during a weekend are seen by the physician who is randomly assigned to be on call that weekend, whereas the patients delivering during the week are more likely to be seen by their regular obstetrician, with whom they may have matched. They estimate a physician specific fixed effect on the propensity to perform a cesarean during the week, and separately estimate an analogous effect for births occurring on a weekend. They compare the variance of these two terms, with the hypothesis that the weekday effect will have a greater variance if matching based upon unobserved patient preferences for treatment style occurs, but ultimately reject this hypothesis. The following analysis is similar in spirit to theirs, but is econometrically different.

Consider a model in which patients have a desire or belief about their own likelihood of receiving a cesarean, measured by a linear index:

$$C_i^* = X_i\beta + \xi_i + \varepsilon_i, \text{ where } \begin{matrix} \xi_i \sim N(0,\sigma_\xi^2) \\ \varepsilon_i \sim N(0,\sigma_\varepsilon^2) \end{matrix},$$

X_i is a vector of patient i's clinical indications or risk factors for receiving a cesarean, ξ_i is a patient specific taste for receiving a cesarean that is known to the patient ahead of time, and ε_i is a shock that occurs at the time of the birth. Let an obstetrician have a similar index measuring the likelihood to treat the patient by cesarean,

$$C_{ij}^* = X_i\beta + \mu_j + \varepsilon_i, \text{ where } \mu_j \sim N(0,\sigma_\mu^2),$$

and μ_j is a obstetrician specific taste for performing a cesarean that is observable to the patient.

Assume that at least some patients match with obstetricians in an effort to minimize the absolute value of the difference between μ and ξ, and that there is a continuum of obstetricians who do not experience capacity constraints so that patient i can find an obstetrician for whom $\mu_j = \xi_i$.[35] Assume further that, regardless of whether they matched or not, births occurring on a weekend are delivered by a randomly assigned on call obstetrician and that a patient does not have any control over treatment once they pick or are assigned to an obstetrician.

Let the probability that patient i receives a cesarean from obstetrician j be written as,

(1) $$\Pr(C_{ij}^* > 0 \mid X_i) = \Pr(X_i\beta + \mu_j + \varepsilon_i > 0),$$

and let $\mu_j = \xi_i$ for patients who have chosen an obstetrician based upon treatment preference have an obstetrician, implying that

(2) $$\Pr(C_{ij}^* > 0 \mid X_i) = \Pr(X_i\beta + \xi_i + \varepsilon_i > 0),$$

[35] This assumption is not explicitly necessary but will simplify notation to come and provide better intuition. It must be the case however that patients search and match as to $\underset{\mu_j}{Min} \mid \mu_j - \xi_i \mid$.

if treated by the obstetrician with whom they matched. If this is true, then the probability of receiving a cesarean for the matched patients delivering during the week can be written as (2), but the probability of cesarean for any patient delivering on the weekend cannot and must be written as (1), regardless of their prior decision to match or not.

Given that obstetricians have years of medical training and experience that most patients do not, it seems likely that there would be more consensus among obstetricians about the propriety of a cesarean, conditional on risk factors, than in exists among patients.[36] This implies that $\sigma_\mu^2 < \sigma_\xi^2$ and trivially that,

(3) $$Var(\mu_j + \varepsilon_i) < Var(\xi_i + \varepsilon_i)\ .$$

Because I do not observe the identity of the delivering obstetrician, I cannot estimate equations (1) and (2) to test (3), but instead I can estimate and test a related set of equations. I estimate a heteroskedastic Probit model with latent index $C*$,

$$C_i^* = X_i\beta + \gamma W_i + \tilde{\varepsilon}_i\ , \text{ and } \quad \tilde{\varepsilon}_i \sim N(0, \sqrt{1+\delta})\ , W_i\text{=1, if the birth is on a weekend,}$$

$$\text{but} \quad \tilde{\varepsilon}_i \sim N(0,1)\ , W_i\text{=0, if the birth is during the week.}$$

The implication analogous to equation (3) in this case is that $\delta < 0$. This implies that births occurring on the weekend are unobservably different from those occurring during the week, and that those differences systematically alter the probability that the patient receives a cesarean. One potential problem with this is that the particular day a patient delivers might be endogenously determined (e.g. a scheduled cesarean). For this reason I take an instrumental variables approach in estimation.

The first stage regression is a Probit estimation with W as the dependent variable and,

[36] A model with some positive search cost could generate a selection effect which would yield essentially the same implication that follows (i.e. equation 3) without imposing the assumption that $\sigma_\mu^2 < \sigma_\xi^2$.

$$\Pr(W_i = 1 \mid X_i, \overline{W_k}, \nu_i) = X_i\Pi + \theta\overline{W_k} + \nu_i ,$$

where $\overline{W_k}$ is the instrument, the fraction of births in area k occurring on the weekend. The results of this regression are reported in Table B1. The first stage estimates are then used to construct a predicted value for W_i, $\hat{W}_i$. Following Mallar (1977) the heteroskedastic Probit equation can then be written as

(4) $$\Pr(C_i = 1 \mid X_i, \hat{W}_i, \tilde{\varepsilon}_i) = \Pr(C_i^* > 0) = \Pr(X_i\beta + \gamma\hat{W}_i + \tilde{\varepsilon}_i) ,$$

where $\hat{W}_i = X_i\hat{\Pi} + \hat{\theta}\overline{W}$.

The results of this estimation are presented in Table B2. The parameter of interest, δ, is estimated to be -0.0232, and is statistically significant. This implies that the standard deviation of the error term for weekend births is roughly 2% smaller than that of a birth during the week. The small magnitude of the estimate of delta combined with the results from Epstein, Ketcham, and Nicholson (2007) lead me to believe my structural results will not be sensitive to the assumption that patients are randomly assigned to obstetricians.

Table B1: First Stage Results for IV Probit Modeling Pr(W=1)

Variable	Value		Std error
Constant	-1.5091	**	0.012
Prev C-sec	-0.3231	**	0.003
Anemia	0.0241	**	0.005
Cardiac disease	-0.0355	**	0.011
Lung disease	-0.0038		0.007
Diabetes	-0.0873	**	0.005
Hydramnios	-0.1132	**	0.007
Hemoglobnopathy	-0.0588	**	0.027
Chronic hypertension	-0.0928	**	0.009
Pregnancy hypertension	-0.0899	**	0.004
Eclampsia	0.0384	**	0.014
Incompetent cervix	-0.0518	**	0.015
Previous 4000g+ infant	-0.1141	**	0.008
Prev preterm or small for age	0.0582	**	0.007
Renal disease	-0.0132		0.015
Febrile	0.0721	**	0.006
Meconium staining	0.0888	**	0.003
Uterine rupture	0.167	**	0.005
Abruptio placenta	0.1569	**	0.01
Placenta previa	-0.1127	**	0.014
Excessive bleeding	0.0432	**	0.011
Maternal Seizures	0.0874	**	0.04
Dysfunctional labor	0.0484	**	0.005
Breech presentation	-0.142	**	0.004
Cephalopelvic disproportion	-0.0324	**	0.006
Cord prolapse	-0.0036		0.017
Weekend Birth Fraction	3.4566	**	0.056

log likelihood = -1185183.4

N = 3,436,749

Table B2: Second Stage Results for IV Heteroskedastic Probit Modeling Pr (C=1)

Variable	Value		Std error
Constant	-1.1443	**	0.018
Prev C-sec	2.0415	**	0.007
Anemia	-0.0618	**	0.006
Cardiac disease	0.0029		0.012
Lung disease	-0.0232	**	0.008
Diabetes	0.3218	**	0.005
Hydramnios	0.4245	**	0.007
Hemoglobnopathy	0.0427		0.029
Chronic hypertension	0.4749	**	0.009
Pregnancy hypertension	0.4985	**	0.004
Eclampsia	0.7103	**	0.013
Incompetent cervix	0.0755	**	0.015
Previous 4000g+ infant	-0.2415	**	0.009
Prev preterm or small for age	-0.4167	**	0.009
Renal disease	-0.0192		0.016
Febrile	0.2087	**	0.006
Meconium staining	0.0687	**	0.004
Uterine rupture	-0.0098	*	0.006
Abruptio placenta	1.1342	**	0.009
Placenta previa	1.9595	**	0.013
Excessive bleeding	-0.0722	**	0.011
Maternal Seizures	0.7961	**	0.037
Dysfunctional labor	1.5522	**	0.004
Breech presentation	2.2453	**	0.005
Cephalopelvic disproportion	2.9094	**	0.009
Cord prolapse	1.3956	**	0.014
Fitted Weekend Dummy	0.1572	**	0.019
delta	-0.0232	**	0.002

log likelihood = -1185183.4
N = 3,436,749

Appendix C: Empirical Appendix

1 Average Damage Payments

The first step is using the NPDB data to estimate the average damage awards for all outcomes. The NPDB records the amount of a given payment along with some data about the nature of the malpractice. There are a number of data fields that provide increasingly detailed information about the case, but unfortunately these are also reported with irregularity increasing with that level of detail. I look only at cases brought against physicians where the nature of the claim is listed as "Obstetrics Related." There are two additional variables further classifying the nature of the allegation. These can take on one of ninety one different values, two of which particularly interest me: "improperly performed C-Section" and "improperly performed vaginal delivery." While some are coded with specific allegations like the previous two or "Failure to Treat Fetal Distress," many are simply coded as "Cannot be Determined from Available Records," or as "Allegation – Not Otherwise Classified."

There are four outcomes of interest in the data. These outcomes are: cesarean settlement, cesarean judgment, vaginal delivery settlement, and vaginal delivery judgment. It is clear in which outcome the cases belong when they are classified as "improperly performed C-Section" or "improperly performed vaginal delivery." Presumably, all or nearly all cases in the data could be classified as one of those two, but those identifiers only appear in 12% of entries. In a simplified coding system it would be easier to identify which cases belong to which categories, but the level of detail goes beyond the interest of this investigation. A consequence of this detail is that some states do not have claims clearly identified as being in one of the four outcome categories. If these cases were identified as such, it would be a simple matter of taking an average of the payments made for each of the four outcomes, but because this is not the case I use the data to estimate a hedonic regression model for payment amounts.

Payments are modeled so as to allow for estimation of payments for all outcomes even when such an outcome is not observed:

$$\ln(Payment_i) = St_i + Disp_i + Allegation_i + (St_i \text{ x } Disp_i \text{ x } Allegation_i) + \varepsilon_i,$$

where *St* is a vector of state fixed effects, *Disp* is an indicator for whether the payment was a settlement or a judgment, and *Allegation* is a vector of indicators for the nature of the allegation, and x is used to denote the full set of interactions between *St*, *Disp* and *Allegation*. By including a full set of secondary and tertiary interactions, this model perfectly predicts the state level means for all four outcomes when they are observed, and yields an estimated average for those unobserved outcomes. The two elements of *Allegation* in which I am specifically interested are improperly performed cesareans and improperly performed vaginal deliveries. This model predicts fitted values for the D_{sk}^t terms that are to be used in estimating equation (7). For example, D_{3k}^c will be equal to the exponent of the sum of all the fixed effects and interactions for a judgment awarded in state *k* where a cesarean was improperly performed. Payment estimates for each state by treatment, and disposition are reported in Table C1.

Table C1: Charges for Births in 2000 from Nationwide Inpatient Sample

	Average Total Charges C-sections and Vaginal Deliveries						Ratio Relative to Private			
	Medicaid		Private		Self-Pay		Medicaid		Self-Pay	
State	c	v	c	v	c	v	c	v	c	v
AZ	11,013.37	5,952.95	10,607.28	6,046.69	9,647.39	5,123.29	1.04	0.98	0.91	0.85
CA	15,096.35	6,588.17	14,755.43	7,397.70	12,764.26	6,503.68	1.02	0.89	0.87	0.88
CO	10,869.91	4,913.13	11,145.88	4,970.03	9,413.54	4,460.21	0.98	0.99	0.84	0.90
CT	6,660.94	3,290.78	7,070.96	3,681.57	6,139.61	3,556.21	0.94	0.89	0.87	0.97
FL	10,903.15	5,703.02	10,618.48	5,963.28	9,968.59	5,129.67	1.03	0.96	0.94	0.86
GA	8,242.82	4,280.73	8,452.74	4,705.77	7,793.83	4,204.00	0.98	0.91	0.92	0.89
HI	7,650.14	4,999.52	7,119.56	4,988.72	7,045.26	3,737.98	1.07	1.00	0.99	0.75
IL	11,150.49	5,361.41	10,345.85	5,446.55	11,053.29	4,996.62	1.08	0.98	1.07	0.92
IA	6,717.12	3,230.46	6,594.50	3,206.70	6,077.38	3,136.43	1.02	1.01	0.92	0.98
KS	7,633.77	3,998.68	7,541.57	4,275.15	6,434.31	3,215.77	1.01	0.94	0.85	0.75
KY	7,263.51	4,069.61	6,940.89	4,377.36	5,504.31	3,818.06	1.05	0.93	0.79	0.87
ME	6,932.49	3,092.71	6,873.92	3,136.32	6,846.80	2,563.12	1.01	0.99	1.00	0.82
MD	5,485.38	3,443.65	4,969.87	3,208.45	4,187.63	2,890.84	1.10	1.07	0.84	0.90
MA	10,340.02	6,009.57	8,967.32	5,685.21	9,119.18	5,453.81	1.15	1.06	1.02	0.96
MO	8,568.09	4,490.78	7,613.76	4,372.94	7,768.53	3,829.32	1.13	1.03	1.02	0.88
NJ	13,157.67	8,025.40	13,813.63	7,946.20	13,410.71	8,581.71	0.95	1.01	0.97	1.08
NY	9,002.42	5,069.62	8,315.42	4,514.97	9,011.52	4,512.11	1.08	1.12	1.08	1.00
NC	7,625.38	3,944.40	6,406.16	3,613.58	7,042.61	3,428.10	1.19	1.09	1.10	0.95
OR	7,232.61	3,272.14	5,781.45	2,965.01	7,088.38	3,122.61	1.25	1.10	1.23	1.05
PA	12,069.82	7,039.31	9,417.66	5,566.36	6,740.89	5,733.48	1.28	1.26	0.72	1.03
SC	7,922.32	4,457.69	7,035.27	4,332.98	6,482.47	3,434.88	1.13	1.03	0.92	0.79
TN	7,019.90	3,806.04	6,608.28	3,653.11	9,240.31	4,089.78	1.06	1.04	1.40	1.12
TX	7,842.06	4,288.38	8,549.66	4,880.08	6,578.35	3,323.75	0.92	0.88	0.77	0.68
UT	5,819.13	2,816.78	6,295.33	3,216.06	5,554.46	2,697.70	0.92	0.88	0.88	0.84
VA	7,738.96	4,016.48	7,453.07	4,179.32	6,493.26	3,644.02	1.04	0.96	0.87	0.87
WA	8,323.28	3,842.01	7,869.38	4,145.14	6,478.97	3,779.35	1.06	0.93	0.82	0.91
WV	6,656.40	2,848.33	6,157.44	2,913.59	6,180.67	3,979.01	1.08	0.98	1.00	1.37
WI	8,870.00	4,510.51	7,604.77	3,555.77	7,782.98	3,325.15	1.17	1.27	1.02	0.94
Mean	8,707.41	4,548.65	8,247.34	4,533.74	7,780.34	4,152.52	1.06	1.01	0.95	0.92
Median	7,882.19	4,284.56	7,573.17	4,352.96	7,043.93	3,798.71	1.05	0.99	0.92	0.90
Min	5,485.38	2,816.78	4,969.87	2,913.59	4,187.63	2,563.12	0.92	0.88	0.72	0.68
Max	15,096.35	8,025.40	14,755.43	7,946.20	13,410.71	8,581.71	1.28	1.27	1.40	1.37
Std Dev	2,301.02	1,301.31	2,301.17	1,276.33	2,165.68	1,290.52	0.09	0.10	0.14	0.13
N = 834,233						**Ratios estimated by OLS Regression:** (e.g. $\text{Medicaid}_{c,st} = \beta\text{Private}_{c,st}+\varepsilon_{c,st}$)	1.17	1.16	0.93	0.90

2 Malpractice Insurance Premia

The probability of facing a lawsuit of disposition *s*, resulting from treatment *t*, and in state *k*, is estimated from the premium data, legal expenses, and average damage awards. In order to estimate the lawsuit probabilities in the premium model in equation (7), I must estimate the expected loss to the insurer which in this case is the average damage payment.[37] I then use the estimated damages for each treatment in each state to estimate the probability of facing a lawsuit.

Even with M_{ik}, λ_k, D^t_{sk}, and E_{sk} all coming from data and prior estimation, the π^t_{sk} terms in equation (7) are identified only with the addition of two assumptions. I assume that cases of each disposition appear in the same relative proportions in all states for a particular treatment,

A1: $\pi^t_{sk} = a_s \pi^t_{1k}$, for $s = \{1,2,3,4\}$ and $t = \{c, v\}$,

where the a_s terms are nationally aggregated data from the PIAA. I assume that that these relative proportions are the same in all states. This allows all the π^t_{sk} for one treatment to be expressed in terms of π^t_{1k}. I also assume that cases resulting from each treatment appear in a state specific proportion across dispositions,

A2 : $\pi^v_{sk} = b_k \pi^c_{sk}$, for $s = \{1,2,3,4\}$.

Unfortunately the b_k terms are not as easily accessible from data as the a_s terms. The NPDB can be utilized to gain some insight, but the NPDB only records cases in which a payment was exchanged, which means cases that are dropped, or returned verdicts for the defendant, are not represented. Assuming that the b_k term is the same for all dispositions assumes that the proportion of cases in the NPDB resulting from cesareans vs. vaginal deliveries is the same for

[37] Estimation methodology and results are reported in Appendix C

the observed outcomes as it is for the unobserved outcomes. Because b_k is assumed to be the same across dispositions, b_k can be rewritten as,

$$b_k = \frac{P(s=3,4 \mid c,k)}{P(s=3,4 \mid v,k)} \text{, and by Bayes' Theorem}$$

$$= \frac{\dfrac{P(s=3,4 \cap c \cap k)}{P(c \mid k)P(k)}}{\dfrac{P(s=3,4 \cap v \cap k)}{P(v \mid k)P(k)}}$$

$$b_k = \frac{P(s=3,4 \cap c \cap k)}{P(s=3,4 \cap v \cap k)} \bullet \frac{P(v \mid k)}{P(c \mid k)}.$$

Entries in the NPDB file that are clearly identified as resulting from a cesarean or vaginal delivery can be used to estimate the relative proportion of these cases, but as was the case with estimating the average payments, there are some states which do not have entries for all of the possible outcomes, and some for which the numbers are very small. In the NPDB, the number of cases of where a judgment was returned for the plaintiff (s=3) or a settlement was made (s=4) resulting from cesareans in a particular state is often small or zero. It is not surprising that the analogous entries for cases resulting from vaginal deliveries occur more frequently, given that there are fewer cesareans performed.[38] Because these data are small non-negative integers, a Poisson regression framework is the most appropriate choice for modeling.

Defining the number of cases of dispositions three or four resulting from a cesarean in state k as S_k^c, I will assume that,

$$S_k^c \sim Poisson(\Lambda_k^c) \text{, with}$$

$$\Lambda_k^c = \exp\{\delta_0 \ln(S_k^v) + \delta' X_k\},$$

[38] Overall rate of cesarean section in 1999 was 22.2%

where S_k^v is the number of cases of dispositions three or four resulting from vaginal deliveries in state *k*, X_k is a vector of explanatory variables including the number of cesareans performed in state *k* and indicators capturing the characteristics of the liability laws in state *k*, and the δ's are the parameters to be estimated.[39] Results of this estimation are reported in Table C2. This method allows the calculation of a fitted value for Λ_k^c relative to S_k^v. By rewriting equation (8), this fitted value can then be used to calculate b_k,

$$b_k = E(\frac{S_k^c}{S_k^v}) \bullet \frac{P(v \mid k)}{P(c \mid k)}.$$

The results of this estimation are reported in Table C3.

Combining A1 and A2 allows all the π_{sk}^t's to be expressed in terms of π_{1k}^c. Equation (7) can now be rewritten as,

$$(8) \qquad M_{ik} = \mu_i + \lambda_k \sum_{s=1}^{4} a_s \pi_{1k}^c (D_{sk}^c + E_{sk}) + (1 - \lambda_k) \sum_{s=1}^{4} b_k a_s \pi_{1k}^c (D_{sk}^v + E_{sk}) + e_{ik},$$

where a_1 =1 and π_{1k}^c is the only probability parameter to be estimated. For some states the malpractice premium is reported at the state level, while for others it is reported for a county, or a group of counties. In these states, the malpractice premium offered by company *i* in state *k*, is a population weighted average of the premiums in all the counties or metropolitan areas in the state. The E_{sk} terms are the legal expenses associated with defending a particular type of case in each state and are constructed by weighting the PIAA national data on legal expenses by a state level price index.[40] Some states are omitted from the analysis because of insufficient data on one

[39] Data on liability laws are taken from Ronen Avraham's Dataset of Tort Law Reforms 1980-2005, available at http://www.law.northwestern.edu/faculty/profiles/RonenAvraham/.
[40] State level price indices are constructed from housing price and non-housing price survey data compiled by Edgar Olsen. MSA level price indices are population weighted to form the state level indices.

or more variables.[41] Equation (8) is estimated using OLS and the estimated π_{sk}^{t}'s are reported in Table 4. With the π_{sk}^{t} terms estimated, I can now analyze them to see how they depend on the distribution of patient characteristics and risk factors. This model and estimation are described in Section 4.

Table C2: Results of Poisson Regression Predicting Lawsuits Resulting from Cesareans Relative to Vaginal Deliveries

Variable	Estimate		t-statistic	std err
Constant	-1.0022	**	-3.05459872	0.328085
$\ln(S_k^v)$	0.9023	**	9.22837946	0.097775
State level # of C-sections	0.0051		0.90087475	0.005646
Cap on punitive damages	-0.4231	**	-2.01730437	0.209759
Cap on total damages	0.0788		0.26511614	0.29727
Cap on non economic damages	0.1604		1.03081649	0.155641
Joint & several liability reform	-0.187		-1.0386976	0.180005
Punitive damages evidentiary	-0.3149		-1.4350086	0.219433
Split Recovery (punitive damages)	0.3367		0.83512485	0.403126
Collateral source rules	0.174		0.64292541	0.270638
Contingency fees	0.2781		0.82164996	0.338411
Patient Compensation fund	0.1633		0.71513307	0.228385
Periodic payments rules used with court discretion	-0.1157		-0.64200022	0.18027
Periodic payments rules required to be enforced	-0.4273	**	-2.30442356	0.18542

log likelihood = -101.779

[41] States where the MLM reports only one insurance company's price, and that company doesn't serve any other states cannot be used in this analysis. These states are: DC, HI, OK, SC.

N = 88 ** indicates significance at the 1% level
Source: National Practitioner Data Bank and Ronen Avraham's Dataset of Tort Law Reforms

Table C3: Estimated Ratios of Cases resulting from Vaginal Deliveries to Cesareans

State	$b_k = E(\frac{S_k^c}{S_k^v}) \bullet \frac{P(v \mid k)}{P(c \mid k)} = \frac{\pi_{sk}^v}{\pi_{sk}^c}$		State	$b_k = E(\frac{S_k^c}{S_k^v}) \bullet \frac{P(v \mid k)}{P(c \mid k)} = \frac{\pi_{sk}^v}{\pi_{sk}^c}$
AL	0.489427		**MO**	0.478283
AK	1.091806		**MT**	0.524956
AZ	0.957219		**NE**	0.852892
AR	0.876445		**NV**	0.413858
CA	1.029465		**NH**	0.684806
CO	0.580358		**NJ**	0.471365
CT	0.931396		**NM**	1.12594
DE	1.024829		**NC**	0.508232
FL	0.638647		**ND**	0.692481
GA	0.428128		**OH**	0.858429
ID	1.284587		**OR**	1.054204
IL	0.877348		**PA**	0.544151
IN	0.705906		**RI**	1.421065
IA	1.031301		**SD**	0.622847
KS	0.522427		**TN**	0.941287
KY	0.837089		**UT**	1.271426
LA	0.361937		**VT**	1.49292
ME	0.915706		**VA**	0.73932
MD	0.701569		**WA**	0.546681
MA	2.030434		**WV**	0.954693
MI	0.68609		**WI**	0.763833
MN	0.666233		**WY**	1.892346
MS	0.501031			

These ratios are not adjusted for prevalence of medical risk factors.
See Section 3.2.2 for full description of methodology

3 Simulation Algorithm ***UNFINISHED***

There are a number of integrals to evaluate for computation of the theoretical moments, all of which depend on the joint density $g(\omega_1, \omega_2, \omega_3)$ in some form. The distribution is assumed to be Trivariate Normal with parameters, $\Omega = [\mu, \tilde{\Omega}]$. Because I do not observe individual obstetricians, I take an initial guess of Ω, and then in order to simulate the integrals I draw values for the three *ω*'s for each of the moment conditions. The following algorithm describes how I draw the *ω*'s:

1. Draw and store (551939 + 44)x(3) times from a uniform distribution , u^r.
2. Compute $g(\omega_1, \omega_2, \omega_3)$ at all deciles [0,1] (11x11x11)
3. Compute marginal distribution $G_1\left(\omega_1^i\right) = \sum_{i=1}^{i} g_1\left(\omega_1^i\right)\left(\omega_1^{i+1} - \omega_1^i\right)$
    a. $g_1\left(\omega_1^i\right) = \sum_{j=1}^{10}[\sum_{k=1}^{10} g(\omega_1^i, \omega_2^j, \omega_3^k)(\omega_3^{k+1} - \omega_3^k)](\omega_2^{j+1} - \omega_2^j)$
4. Interpolate to make $G_1\left(\omega_1^i\right)$ a continuous function.
    a. If $G_1\left(\omega_1^i\right) < u^r < G_1\left(\omega_1^{i+1}\right)$ then
    b. $u^r(\omega_1^r) = \left[\frac{G_1\left(\omega_1^{i+1}\right) - G_1\left(\omega_1^i\right)}{\omega_1^{i+1} - \omega_1^i}\right]\omega_1^r + G_1\left(\omega_1^i\right)$
5. $\omega_1^r = [u^r - G_1\left(\omega_1^i\right)]\left[\frac{\omega_1^{i+1} - \omega_1^i}{G_1\left(\omega_1^{i+1}\right) - G_1\left(\omega_1^i\right)}\right]$
6. Compute conditional marginal distribution $G_2\left(\omega_2^m | \omega_1^r\right)$ where $\omega_1^i < \omega_1^r < \omega_1^{i+1}$
    a. Compute $G_2\left(\omega_2^m | \omega_1^i\right)$ and $G_2\left(\omega_2^m | \omega_1^{i+1}\right)$
    b. $G_2\left(\omega_2^m | \omega_1^i\right) = \sum_{i=1}^{m} g_2\left(\omega_2^i | \omega_1^i\right)\left(\omega_2^{i+1} - \omega_2^i\right)$
    c. $g_2\left(\omega_2^i | \omega_1^i\right) = \sum_{j=1}^{10} g(\omega_1^i, \omega_2^i, \omega_3^j)\left(\omega_3^{j+1} - \omega_3^j\right)$

d. $G_2\ (\omega_2^m|\omega_1^r) = \left[\frac{G_2\ (\omega_2^m|\omega_1^{l+1}) - G_2\ (\omega_2^m|\omega_1^l)}{\omega_1^{l+1} - \omega_1^l}\right] \omega_1^r + G_2\ (\omega_2^m|\omega_1^l)$

7. Interpolate to make $G_2\ (\omega_2^m|\omega_1^r)$ a continuous function

    a. If $G_2\ (\omega_2^m|\omega_1^r) < u^r < G_2\ (\omega_2^{m+1}|\omega_1^r)$ then

    $$u^r\,(\omega_2^r) = \left[\frac{G_2\ (\omega_2^{m+1}|\omega_1^r) - G_2\ (\omega_2^m|\omega_1^r)}{\omega_2^{m+1} - \omega_2^m}\right] \omega_2^r + G_2\ (\omega_2^m|\omega_1^r)$$

8. $\omega_2^r = [\,u^r - G_2\ (\omega_2^m|\omega_1^r)]\left[\frac{\omega_2^{m+1} - \omega_2^m}{G_2\ (\omega_2^{m+1}|\omega_1^r) - G_2\ (\omega_2^m|\omega_2^r)}\right]$

9. Compute conditional marginal distribution $G_3\ (\omega_3^n|\omega_1^r, \omega_2^r)$ where $\omega_1^l < \omega_1^r < \omega_1^{l+1}$ and $\omega_2^m < \omega_2^r < \omega_2^{m+1}$

    a. $G_3\ (\omega_3^n|\omega_1^l, \omega_2^m) = \sum_{i=1}^{n}\ g_3\ (\omega_3^i\,|\omega_1^l, \omega_2^m)\,(\omega_3^{i+1} - \omega_3^i)$

    b. $G_3\ (\omega_3^n|\omega_1^r, \omega_2^m) = \left[\frac{G_3\ (\omega_3^n|\omega_1^{l+1}, \omega_2^m) - G_3\ (\omega_3^n|\omega_1^l, \omega_2^m)}{\omega_1^{l+1} - \omega_1^l}\right] \omega_1^r + G_3\ (\omega_3^n|\omega_1^l, \omega_2^m)$

    c. $G_3\ (\omega_3^n|\omega_1^r, \omega_2^r) = \left[\frac{G_3\ (\omega_3^n|\omega_1^r, \omega_2^{m+1}) - G_3\ (\omega_3^n|\omega_1^r, \omega_2^m)}{\omega_2^{m+1} - \omega_2^m}\right] \omega_2^r + G_3\ (\omega_3^n|\omega_1^r, \omega_2^m)$

10. Interpolate to make $G_3\ (\omega_3^n|\omega_1^r, \omega_2^r)$ a continuous function

    a. If $G_3\ (\omega_3^n|\omega_1^r, \omega_2^r) < u^r < G_3\ (\omega_3^{n+1}|\omega_1^r, \omega_2^r)$ then

    $$u^r\,(\omega_3^r) = \left[\frac{G_3\ (\omega_3^{n+1}|\omega_1^r, \omega_2^r) - G_3\ (\omega_3^n|\,\omega_1^r, \omega_2^r)}{\omega_3^{n+1} - \omega_3^n}\right] \omega_3^r + G_3\ (\omega_3^n|\omega_1^r, \omega_2^r)$$

11. $\omega_3^r = [\,u^r - G_3\ (\omega_3^n|\omega_1^r, \omega_2^r)]\left[\frac{\omega_3^{n+1} - \omega_3^n}{G_3\ (\omega_3^{n+1}|\omega_1^r, \omega_2^r) - G_3\ (\omega_3^n|\omega_1^r, \omega_2^r)}\right]$

12. Use antithetic acceleration by repeating steps 5, 8, and 11 using $u^{rc} = 1 - u^r$ in place of u^r.