A Structural Misclassifcation Model to Estimate the Impact of Physician Incentives on Healthcare Utilization

Arrieta, Alejandro

November 2007

A Structural Misclassification Model to Estimate the Impact of Physician Incentives on Healthcare Utilization

Alejandro Arrieta
Rutgers University

(Job Market Paper)

November 3, 2007

Abstract

The issue of over-utilization of medical procedures has generated strong debate in the United States. It is well acknowledged that, in the agency relationship between physicians and patients, the informational advantage gives doctors an incentive to deviate from the appropriate treatment as defined for a patient's health status, thus incurring over- or under-utilization. However, the empirical consequence of this problem has not been adequately considered. In particular, physician agency breaks the correspondence between appropriate treatment and observed treatment, generating a problem whose characteristics and effects on estimation are analogous to a classification error. However, the error is non-random. Empirical literature that does not consider the misclassification problem understates the impact of clinical and non-clinical factors on healthcare utilization.

This paper proposes a structural misclassification model in which the physician behavior is modeled to characterize the structure of the measurement error. The model captures the interaction between a physician's incentives and a patient's health status, and returns consistent estimators. It also lets us identify the degree of deviation from appropriate treatment (misclassification probability) due to physician incentives, and to compute risk-adjusted utilization rates based on clinical factors only. The model is applied to the cesarean section deliveries performed in the state of New Jersey during the 1999-2002 period. Our results show a moderate but growing rate of non-clinically required c-sections of around 3.2%. We conclude that the growth of the c-section rates in New Jersey over these years is explained mainly by non-clinical factors.

1 Introduction

Ever since Arrow's paper on uncertainty in the healthcare market (Arrow, 1963), the informational inequality in the doctor-patient relationship has become an issue that is now well identified (McGuire, 2000). The informational advantage of physicians regarding patients' health status creates incentives to overuse or underuse medical procedures according to specific physician's objectives. Literature has analyzed this problem from different perspectives. In the health economic field, the focus has been put on physician-induced demand (Fuchs, 1978; Dranove, 1988), where doctors may exert influence over patients and intentionally shift the patient demand curve, thus increasing health care services against patients' best interest. Inducement, however, has a cost for doctors in terms of professional ethics. Therefore, over-utilization of medical procedures results when the revenues of inducement overcome the intrinsic cost of acting against professional ethics.

In health service research literature, growing attention has been placed on explaining the large variation of utilization rates across geographic areas (Wennberg, 2002; Fuchs et al., 2001). There is strong evidence that these regional disparities are not related to clinical factors. Fear of litigation, racial biases, socioeconomic differences and institutional differences are among the non-clinical factors that have an influence on clinical outcomes. However, when it comes to utilization of medical procedures, this literature is unable to identify under- and over- healthcare utilization clearly. Because the observed rates are currently affected by non-clinical factors, it is difficult to construct a health-related benchmark to define appropriate utilization rates.

The literature that has focused on health care quality has also found a similar problem. In this case, there are at least two goals: First, the definition of medical standards or guidelines for specific treatments. Second, the elaboration of quality indexes to report hospital or physician quality level ratings using standard methodologies to compute case-mix risk adjusted rates (Iezzoni, 2003). In both cases, it is important to find the clinical factors that define an appropriate treatment. However, studies that are based on observed treatments may be "contaminated" with the effect of non-clinical factors. As the patient-physician relationship involves a particular interaction between the patient's health status and the physician's incentives, controlling for non-clinical factors by itself is not enough to correct the problem.

Although the fact that physician incentives affect health care utilization rates is known, it has not been adequately considered in empirical literature. The usual approach has been to estimate binary dependent models (usually logit or probit) and control for clinical or non-clinical factors depending on the variables of interest. However, this approach does not consider that physician incentives break the correspondence between appropriate treatment and observed treatment. The definition of appropriate treatment is based on patient health status only, the latter being only observed by the physician. When incentives are strong enough, the physician deviates from appropriate treatment and, therefore, the observed treatment will not reflect the clinical characteristics of the patient. In that regard, this problem can be seen as a misclassification problem where the measurement error in the binary dependent variable is proportional to the strength of physician incentives to deviate from appropriate treatment.

Two examples are illustrative. The first is related to racial differences in healthcare access. One of the most studied cases is lower access to cardiovascular procedures in African Americans (Kressin and Peterson, 2001; Ford and Cooper, 1995; Van Ryn and Burke, 2000). In this case, an African American patient with a poor health condition requires a cardiovascular surgery. Based on health status, the appropriate treatment -observed only by the doctor- should be the utilization of the procedure. However, if the doctor has a racial bias, he may influence the treatment choice by omitting to suggest the surgery. In that case, the observed outcome will be "misclassified" resulting in under-utilized cardiovascular procedures for African Americans.

However, some authors suggest that, in the case of cardiovascular surgeries, the problem is not under-use for African Americans, but over-use among white patients (Schneider et al., 2001). This means that, based on additional non-clinical factors, the doctor may also influence treatment choice for white patients by suggesting the surgery when it is not required. In this case, both over- and under-healthcare utilization coexist, and the observed outcome will thus be "misclassified" in both directions: observation of surgery when it was not required, and no observation of surgery when it was required.

The second example is related to the demand inducement theory, where fee-for-service pricing creates financial incentives for the physician to recommend unnecessary medical procedures. A well known case is cesarean section delivery (Gruber and Owings, 1996; Das, 2002; Tussing and Wojtowycz, 1993). While the doctor sees which delivery method is warranted by the woman's health status, this "true" response is not seen by the econometrician. If financial incentives are strong enough to overcome professional ethics, the doctor will influence a woman to have a c-section even though it is not clinically necessary. In this case, the appropriate choice is affected by the physician's decision, resulting in a "misclassified" outcome that is identified by the econometrician. Note that in this example, "misclassification" runs in one direction only: observation of c-section when it is not required. This happens because there is no monetary incentive to perform a vaginal delivery, and, to top it off, there is fear of litigation. These factors result in a strong disincentive to avoid a vaginal delivery when a c-section is required.

Since this problem is analogous to the misclassification problem, it also shares its consequences. In general, measurement error on limited dependent variables (misclassification) leads to biased and inconsistent estimators. When misclassification is not adequately corrected, it understates parameter estimates, and overstates standard errors. Literature has focused mainly on the case where misclassification is originated randomly by errors in the report or record of a categorical variable (Hausman et al., 1998; Magder and Hughes, 1997; see Kenkel et al., 2004 for an application). Additionally, Abrevaya and Hausman (1999) and Lewbel (2000) have considered the case in which misclassification depends on some covariates, imposing strong conditions for identification. However, there are cases where a decision maker is able to alter the true outcome. When this happens, the classification error will not be random, but a behavioral model can be used to incorporate the structure of the measurement error into the estimation process in order to find consistent estimators.

The main contribution of this paper is a methodology based on a structural misclassification model to estimate the impact of physician incentives on healthcare utilization. This brings three improvements compared to previous literature: First, we obtain consistent estimates for the patients' health risk factors and the physicians' incentives. Second, we model a physician's behavior and its interaction with a patient's health status. Third, we are able to estimate the rate of inappropriate treatments defined as those related to non-clinical factors (misclassification probability) and the risk-adjusted utilization rate based only on health characteristics (removing non-clinical factors).

The second section of this paper describes the structural misclassification model and establishes a parametric solution. In the third section, we use a Monte Carlo study to compare the effect on estimators' consistency of four different approaches to estimate risk-adjusted utilization rates, ranging from not considering the misclassification problem to considering it adequately. The fourth section provides an application to the case of cesarean section deliveries in New Jersey. The last section provides conclusions to this paper, discusses related research in progress and directions for future research.

2 A parametric estimation of structural misclassification

In the examples presented in section I, the physician observes the true health condition of his patient. Conditioned on the patient's health status, the doctor may deviate from appropriate treatment. This will happen if his personal goals overcome his professional ethics within his utility function. Therefore, the structure of this model is a simplified version of game theory models of inducement (De Jaeguer and Jegers, 2001; Xie et al. 2006). It is a simple version because we do not explicitly consider the patient's decisions. The physician's decision tree is shown in Figure 1. In the first stage, nature determines a patient's state (healthy $h < 0$, or sickly $h \geq 0$), and this can only be observed by the physician. Two treatments are considered: A and B. However, only one treatment is appropriate for each patient's health state.

In the second stage, the physician must choose the treatment based on his incentives (i). Based on the patient's health state (h), he may decide to perform the treatment that is appropriate for the patient ($i < 0$) or an inappropriate one ($i \geq 0$). However, the physician will choose the treatment that, after discounting the intrinsic cost of acting against professional ethics, gives him more utility in terms of monetary and non-monetary factors. Therefore, the physician will choose treatment A or B.For instance, a doctor may inappropriately recommend treatment A (cardiovascular surgery) for a healthy White patient and treatment B (no surgery) for a sickly African American patient. On the other hand, an obstetrician may inappropriately recommend treatment A (c-section) for both, risk and non-at-risk pregnant women.

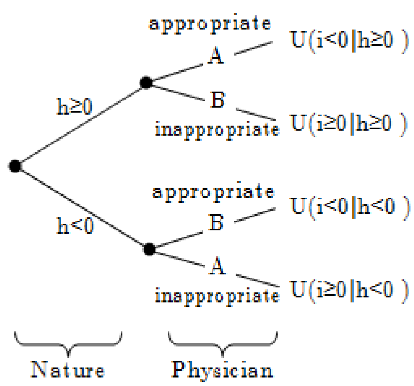


Figure 1: Physician's decision tree

Notice that a more complete model may explicitly include a third stage where the patient accepts or rejects the treatment based on obtained medical information (Xie et al. 2006). However, Figure 1 may be seen as a one-step backward induction in which physicians take patients' actions into consideration. In that regard, index i in the utility function represents incentives that are net of the cost implied by acting against the doctor's ethical standards, but also the expected loss that would occur when a patient that is well informed decides to leave the doctor. The expected loss depends on how well the physician knows his patient (that is, if he knows the patient could collect medical information on his own, as well as their degree of resistance to physician's influence).

The econometric model is described following figure 1. In the first stage, a patient's health status (h) is determined by a set of observable clinical characteristics or risk factors (x) and unobserved risk factors (ϵ_h) .The physician can observe a patient's health condition:

$$h = x\beta + \epsilon_h \tag{1}$$

This is the health status equation. There are two possible treatments, $\tilde{y} = \{0, 1\}$. The patient will require treatment $\tilde{y} = 1$ if health status equals or exceeds zero

$$\tilde{y} = \begin{cases} 1, \text{ if } & h = x\beta + \epsilon_h \geq 0 \\ 0, \text{ otherwise} \end{cases}$$

The Econometrician observes the doctor's treatment choice y but not $\tilde{y}$. Without physician incentives to alter the required treatment $y = \tilde{y}$, and in that case any binary model estimation (logit or probit) will be consistent because the probability of observing the treatment choice and the probability of not observing it are respectively

$$\begin{aligned} \Pr(y = 1) &= \Pr(\tilde{y} = 1) = \Pr(h \geq 0) \\ \Pr(y = 0) &= \Pr(\tilde{y} = 0) = \Pr(h < 0) \end{aligned}$$

However, if the physician decides to do the surgery when it is not needed -like in the cesarean section case- or not to do the procedure when it is required -like in the case of cardiovascular surgery- then the binary model estimation will be inconsistent because $\Pr(y = 1) \neq \Pr(\tilde{y} = 1)$. In those cases the econometrician observes a "misclassified" treatment.

In the second stage, the physician decides to alter the required treatment based on his incentives (i). Incentives depend on doctor's characteristics and patient's characteristics that are observed (z) and unobserved (ϵ_i) by the doctor.

$$i = z\gamma + \epsilon_i \tag{2}$$

This is the physician incentives equation. When incentives equal or exceed a threshold 0, the doctor may proceed, depending on health status, with the inappropriate treatment, thus altering the appropriate choice with probabilities (misclassification probabilities)

$$\alpha_0 \equiv \Pr(y = 1 \mid \tilde{y} = 0) = \Pr(i \geq 0 \mid h < 0) \tag{3a}$$
$$\alpha_1 \equiv \Pr(y = 0 \mid \tilde{y} = 1) = \Pr(i \geq 0 \mid h \geq 0) \tag{3b}$$

α_0 is the probability of over-utilization (doctor performs a surgery when it is not required). α_1 is the probability of under-utilization (doctor does not perform a surgery when it is required). These two probabilities define the probability of observing the surgery:

$$\begin{aligned}
\Pr(y=1) &= \Pr(i<0 \mid h \geq 0)\Pr(h \geq 0) + \Pr(i \geq 0 \mid h<0)\Pr(h<0) \\
&= \alpha_0 + (1-\alpha_0-\alpha_1)\Pr(h \geq 0) \qquad (4)
\end{aligned}$$

Note that the second equality in equation (4) corresponds to equation (5) in Hausman et al. (1998). Clearly, without physician incentives, $\alpha_0 = \alpha_1 = 0$ and therefore $\Pr(y=1)$ collapses to $\Pr(h \geq 0) = \Pr(\tilde{y}=1)$, returning the consistent estimation of binary models. For the cardiovascular example, it is expected that $\alpha_0 > 0$ when the patient is white and $\alpha_1 > 0$ when the patient is an African American. This is the case when misclassification runs in both directions. In this case the probability of observing a patient with $(y=1)$ or without $(y=0)$ cardiovascular surgery is respectively:

$$\begin{aligned}
\Pr(y=1) &= \Pr(i<0 \mid h \geq 0)\Pr(h \geq 0) + \Pr(i \geq 0 \mid h<0)\Pr(h<0) \\
\Pr(y=0) &= 1 - \Pr(i<0 \mid h \geq 0)\Pr(h \geq 0) - \Pr(i \geq 0 \mid h<0)\Pr(h<0)
\end{aligned}$$

For the cesarean section example, we expect over-utilization $(\alpha_0 > 0)$ because of the disincentives to proceed with a vaginal delivery when a c-section is needed. It means that misclassification runs in only one direction given that $\alpha_1 = 0$. In this case the probability of observing a c-section $(y=1)$ and the probability of a vaginal delivery $(y=0)$ are respectively:

$$\begin{aligned}
\Pr(y=1) &= \Pr(h \geq 0) + \Pr(i \geq 0 \mid h<0)\Pr(h<0) \\
\Pr(y=0) &= \Pr(i<0 \mid h<0)\Pr(h<0)
\end{aligned}$$

Haussman et al. (1998) and Lewbel (2000) discuss the conditions for identification of misclassification models. In particular, the monotonicity condition (MC) is required to identify the parameters. For the general misclassification model, the MC is $\alpha_0 + \alpha_1 < 0$. In terms of our problem, this condition is generally satisfied because the degree of physician incentives is relatively small given the punishment in terms of reputation and lawsuits that result if inappropriate treatments are very high. Note that in the case of misclassification in one direction (either over- or under- utilization), the MC is automatically satisfied since there is only one probability, which we safely assume is below 1.

The parameters β and γ in equations (1) and (2) can be estimated with MLE, maximizing the likelihood function

$$\mathcal{L}(\beta,\gamma) = \prod \Pr(y=1)^{y} \Pr(y=0)^{1-y} \tag{5}$$

Notice that if the errors ϵ_h, ϵ_i are not independently distributed (with correlation ρ), the problem becomes a bivariate model. Additionally, if it is assumed that the error terms are normally distributed, the problem becomes a bivariate probit (Amemiya, 1985). For the general case described by equation (4), the likelihood function is

$$\mathcal{L}(\beta,\gamma,\rho) = \prod \left[2\Phi_2(x\beta,-z\gamma;\rho)\right]^{y} \left[1-2\Phi_2(x\beta,-z\gamma;\rho)\right]^{1-y} \tag{6}$$

Where Φ_2 is standard bivariate normal CDF. Notice that this model is a variety of Poirier's partial observability model (Poirier, 1980). However, there are important differences. The partial observability model considers two agents making decisions based on a common set of information. In this structural misclassification model, there is only one decision maker: the physician. The patient's health status is not a decision maker and consequently equations (1) and (2) are generally functions of two separate sets of variables, in contrast to the partial observability model. A variable may be in both equations if it carries information about the patient's health status and the physician's incentives (some examples are age, sex, weight, etc. depending on the analyzed treatment). Partial observability models have been used to address misspecification in simple probit or logit models (see for example Abowd and Farber, 1982), while in this paper partial observability is obtained after adding structure to the misclassification problem. Finally, the likelihood function of the structural misclassification model is different to the partial observability model, but both models converges when misclassification run in only one direction (under- or over- utilization). Note that the loss of information due to limited observability reduces efficiency of the maximum likelihood estimator as in the partial observability models (Poirier, 1980; Meng and Schmidt, 1985).

The structural misclassification error model presented in this paper rests on strong parametric assumptions. We have based our estimation on a bivariate probit, but the model can be easily extended to a bivariate logit. However, a natural extension is to get rid of the parametric assumptions and estimate this model semi-parametrically based on a multiple index model as in Ichimura and Lee (1991).

3 Comparing methodologies: Monte Carlo simulation

In this section, we study the consequences of omitting or mis-specifying the impact of physician incentives on observed health care utilization. In particular, we study how consistency estimation is affected when we use the current methodology used in literature to estimate the influence of clinical factors on health outcome. The variable of interest is a dichotomous variable indicating whether the patient received the treatment under study or the alternative treatment (or no treatment). We consider the case in which physicians have incentives to alter the required treatment given the patient's health status. For simplicity and with the purpose to connect this study with the empirical application described in section IV, we focus on the case of physician incentives to over-utilize medical procedures.

In estimating the effects of clinical risk factors associated with specific treatments, the literature has followed two common approaches[1]: (i) Estimation of a simple binary (SB) model (logit or probit) with clinical factors as the only regressors, and (ii) Estimation of a simple binary model with controls (SBC), where non-clinical factors are added to the SB model as control variables. Because the agency problem in the physician-patient relationship creates important interactions between a patient's health status and non-medical factors, the omission of non-clinical characteristics creates a serious omitted variable bias. When non-clinical factors are added as controls as in the SBC model, the bias is reduced but interactions are not appropriately captured, and the misclassification bias described in previous sections is not adequately corrected.

In that regard, we also study the following two approaches based on the structural misclassification model developed in section II: (iii) Structural misclassification model with independent errors (SMCI), and (iv) Structural misclassification model where errors are allowed to be dependent (SMC). From an empirical perspective, the restriction of independent errors may be strong. Because there are many health characteristics that are non-observable to the econometrician but observable to the physician, the latter could make decisions based on information conveyed in ϵ_h. Given that some variables related to physician incentives are also unobserved by the econometrician, ϵ_i and ϵ_h may be correlated. In that regard, the error correlation also measures physician incentives that are difficult to observe because either the clinical factors do not define clearly what an appropriate treatment is or the physician's incentives go beyond measurable characteristics. Moreover, it is expected that the error correlation is negative for over-utilized procedures and positive for under-utilized procedures. Consider the case of the racial bias example described in sections 1 and 2. If the doctor observes health characteristics that cannot be easily captured by diagnosis codes or medical guidelines, but that imply a need to perform surgery on an African American patient, then it will be easier for a biased doctor to deviate from the appropriate treatment, and regarding any

[1] To get rid of the potential "misclassification" problem, other approaches have considered the reviewing of medical records in light of professional guidelines based on clinical trials or expert opinions. Good examples of these approaches are the RAND guidelines and the American College of Cardiology/American Heart Association guidelines. For a discussion of the own problems of these methodologies see Leape et al. (2003).

observable characteristic, the physician's incentive growths implying a positive error correlation. A similar argument can be used to expect a negative correlation for over-utilized procedures. For that reason, it is important to evaluate the consequence of imposing the independence restriction.

In order to assess the impact of these four approaches on estimator bias and consistency, I examine the results of Monte Carlo simulation. The true model representing equations (1) and (2) is

$$\begin{aligned} h &= -1.5 + 0.5x_1 - x_2 + 2x_3 + \epsilon_h \\ i &= -2.5 - 1.5z_1 + z_2 + 0.5z_3 + 2z_4 + \epsilon_i \end{aligned}$$

Covariates x and z include dummy variables and continuous variables that were drawn from uniform distributions and trimmed chi-squared distributions to avoid outliers. The error disturbances ϵ_h, ϵ_i, are drawn jointly from a bivariate standard normal distribution with correlation $\rho = 0.25$. For the design of the Monte Carlo study we consider 1000 independent random draws of a sample size of 5000. Table 1 reports the sample mean and standard error of parameters estimated over the 1000 draws.

The results of the Monte Carlo simulation are consistent with the misclassification problem described by Hausman et al. (1998). In particular, the simple probit (SB and SBC models) understates the coefficients. The probit model with only patient's health related variables (SB) produces estimates that are biased downward by 35-50% (column 1). When physician incentives are added as control variables (SBC model), bias is still substantial in the case of health risk factors. Coefficient estimates of control variables (non-clinical characteristics) have a more substantial downward bias of around 45-70% (column 2).

Bias size in the health status and the non-clinical coefficients depends, among others, on two parameters: the error correlation and the degree of physician incentives (misclassification). Different Monte Carlo designs (not shown) were used to see the impact of both parameters on estimator biasedness. First, the bias increases in both sets of estimators when the error correlation gets closer to 1. Given the previous discussion related to the sign of the error correlation, this result implies that bias will be larger in the case of under-utilized procedures than in the over-utilized procedures. It is also important to note that even in the case of small correlation, the bias does not vanish. Second, the bias in the health status estimators decreases and in the doctor's incentives estimators increases when the degree of incentives falls. A rough exercise shows that the bias in the health status estimators decreases almost proportionally with the reduction in the degree of incentives. On the other hand, a reduction in the degree of incentives increases the bias in the incentive equation more than proportionally when incentive is high, and less than proportionally when it is low.

Table 1. Monte Carlo simulations

Parameter	True parameter value	Simple Binary Model (1)	Simple Binary Model with Controls (2)	Structural Misclassif. w/indp errors (3)	Structural Misclassif. Model (4)
			Health related variables		
β_0	-1.50	-0.752	-1.34	-1.508	-1.502
		(0.044)	(0.074)	(0.092)	(0.089)
β_1	0.50	0.311	0.349	0.495	0.502
		(0.129)	(0.138)	(0.178)	(0.180)
β_2	-1.00	-0.544	-0.626	-0.985	-1.005
		(0.045)	(0.047)	(0.085)	(0.087)
β_3	2.00	1.249	1.408	1.972	2.003
		(0.068)	(0.072)	(0.114)	(0.117)
			Patient and Physician related variables		
γ_0	-2.50	-	-	-2.548	-2.505
				(0.215)	(0.211)
γ_1	-1.50	-	-0.51	-1.488	-1.503
			(0.143)	(0.321)	(0.321)
γ_2	1.00	-	0.301	0.989	0.995
			(0.046)	(0.127)	(0.126)
γ_3	0.50	-	0.176	0.493	0.498
			(0.066)	(0.144)	(0.145)
γ_4	2.00	-	1.069	1.978	2.005
			(0.048)	(0.127)	(0.126)
ρ	0.25	-	-	-	0.239
					(0.193)
Prob. approp. treatment †	0.249	-	-	0.244	0.250
				(0.015)	(0.015)
Prob. physic incentives ‡	0.112	-	-	0.100	0.110
				(0.018)	(0.020)

n=5000, 1000 simulations. Standard deviations in parentheses.

† Calculated using the marginal probability $Pr(h \geq 0)$

‡ Calculated using the marginal probability $Pr(i \geq 0)$

Estimators of the structural misclassification models (SMCI and SMC) are consistent but present larger standard errors (columns 3 and 4). Compared with these models, the simple probit (SB and SBC models) overstates the precision of its estimates. As an implication for health care quality studies, the impact of health risk factors on utilization rates will appear to be less important than they really are when probit or logit is used. Confidence intervals will be narrower too.

There is a small discrepancy between the structural misclassification model estimators when error independence is erroneously imposed (column 3). In general, estimated coefficients of the SMCI model are biased by 1-2% in the case of risk factors and physician and patient characteristics (column 3). For the structural misclassification model that allows correlated errors (SMC model), the estimates are unbiased (less than 0.5%) and show slightly higher standard errors than the SMCI model (column 4).

An important feature of the structural misclassification model is that it allows us to neatly separate the estimated physician incentives probability and the utilization rate due to health status only. Table 1 reports the degree of doctor's incentives or percentage of misclassification. It was calculated as the marginal probability that physician incentives exceed threshold zero: $\Pr(i \geq 0) = \int \Pr(i \geq 0, h)\, dh$. The design of the Monte Carlo study implies a true degree of incentives of 0.112. The degree of incentives estimated by the SMCI model is downward biased by 11% (column 3), while the SMC model presents a smaller bias (column 4).

If the goal is to estimate risk-adjusted utilization rates, the appropriate measure that discards the effect of physician incentives on the "misclassification" bias will be the marginal probability of appropriate treatment. For this particular design where misclassification is related to over-utilization, the appropriate treatment is obtained only when health status exceeds zero, and therefore $\Pr(h \geq 0) = \int \Pr(i, h \geq 0)\, di$ will be the estimated utilization rate based on health status only, under the counterfactual that there are no physician incentives to influence the appropriate health outcome. For this Monte Carlo study, true probability of appropriate treatment is 24.9%. The SMCI model understates the probability by 1.7% (column 3). The bias is almost zero (less than 0.5%) when the probability of appropriate treatment is estimated using the SMC model (column 4).

4 An application to cesarean section deliveries in New Jersey

4.1 Data

This section applies the structural misclassification model to births in the state of New Jersey in the period 1999-2002. We use Hospital Patient Discharge Data collected by the New Jersey Department of Health and Senior Services. This data contains detailed information on each discharge from an acute care hospital including identification of the hospital, patient demographics and zip code of residence, diagnosis and surgical procedures classified by ninth revision of the International Statistical Classification of Diseases and Related Health Problems(ICD-9) codes and Diagnosis Related Group numbers (DRG), source of admission, and identification of payers. Additional socioeconomic information was collected from the US Census 2000, using the patient's zip code as the key variable for matching. Births were identified by DRG codes 370-375. Cesarean sections were identified by DRG 370-371 or ICD-9 code 74xx excluding 7491. The selected sample includes women aged 15 to 49. We excluded deliveries performed in hospitals that in a particular year had less than 100 births (0.03% of the sample). Finally, we also exclude patients with wrong zip codes (0.5%) and patients with missing or wrong reported information in the variables of analysis (1.7%). The final sample used for the estimation considers a total of 403,660 women.

This estimation uses two sets of variables. A first set comprises women's health characteristics identified according to diagnosis codes. We follow previous health service research to select the clinical variables that were seen as more relevant to explain c-sections (Keeler et al., 1997; Aron et al., 1998; DiGiuseppe et al., 2001; and Rahnama et al., 2006). The second set of variables comprises patient and physician characteristics that may drive doctor's incentives. The complete list of variables and their mean values for vaginal and c-section deliveries are reported in Table A1 in the Appendix.

4.2 A Model of physician incentives

We use the structural misclassification model described in section 2 to measure physician incentives and to estimate the probability of appropriate cesarean section rates by removing non-clinical factors. It is important to highlight that in a different context, the physician incentive equation may capture patient's choice rather than physician's influence. However, because c-sections by women's choice are not allowed in the USA and because professional guidelines consider c-sections for non-medical reasons "*to fall outside the bounds of best professional practice*"[2], c-sections related to non-clinical factors are associated to physician incentives. Even though this model can be easily extended to test for physician-induced demand (PID), the data collected for this application does not allow us to identify PID because of lack of an exogenous shock on medical income on that period.

Again, in the case of c-sections the misclassification runs in only one direction: over-utilization. Assuming standard normal distributions for the error terms, the model is estimated by MLE using the likelihood function described in equation (6) that in this case becomes:

$$\mathcal{L}(\beta, \gamma, \rho) = \prod \left[1 - \Phi_2(-x\beta, -z\gamma; \rho)\right]^y \left[\Phi_2(-x\beta, -z\gamma; \rho)\right]^{1-y}$$

where the probability of observing a c-section is:

$$\Pr(y+1) = 1 - \Phi_2(-x\beta, -z\gamma; \rho)$$

[2]FIGO Statement on Cesarean Section. January 2007. http://www.figo.org/Cesarean.asp

In general, variables in the physician incentive equation may be classified in two types: First, variables observed by the doctor that signal the degree of patient-obtained medical information. These are mainly socioeconomic characteristics that can be observed directly (employment, marital status, etc.) or that can be inferred from patient population (patient ethnicity, patient geographic area residence, etc.) Usually, patients with lower socioeconomic status have less access to information, and also have less capability to use such information in the medical visit (Xie et al. 2006). To capture this effect we include four variables: (i) Social support, measured as the presence of a partner (married or life partner), may be perceived to improve the degree of information because decisions maybe taken on a couple basis. It is expected that it reduces physician incentives and therefore the probability of c-section. (ii) Woman's employment status may signal a potential compliance to c-sections. It is expected that fully employed women may prefer a c-section delivery because of the convenience in terms of scheduling and the lower pre-partum work. Consequently we expect higher physician incentives for full employed women. (iii) Ethnicity (a White, Black or Hispanic woman) may be perceived as a signal of access to medical information. It is expected that minorities have less access and worse use of medical information, which makes them more vulnerable to physician influence. (iv) Patient socioeconomic status is not easily observed by the doctor. Instead, the physician can infer the socioeconomic status from the zip code of the patient's residence. We include the average household income at zip code level. It is expected that a zip code with low income is perceived as lower socioeconomic status, thus increasing vulnerability to physician incentives (Pauly, 1980).

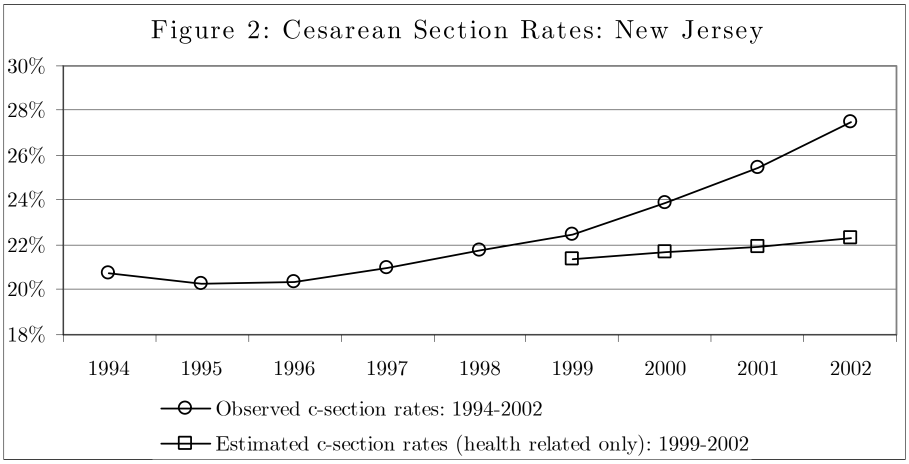
The second type of variables comprises factors that affects directly physician incentives. These are mainly institutional and contractual variables related to the physician itself or the health facility. We consider three variables: (i) The method of payment or patient's insurance condition (uninsured, Medicaid, HMO, other insurance). The type of insurance is important for doctor's incentives because it sets the method of payment. It is expected that the uninsured have the lowest rate of c-sections compared to non-HMO private insurance, since they must pay for the procedure. The capitation payment of an HMO reduces physician incentives, and so does prospective payment system under Medicaid. An additional factor that reduces incentives in the case of Medicaid patients is the low fees for obstetric procedures observed in New Jersey. (ii) The size of the hospital measured as the average yearly number of births observed in each hospital. It has been shown that hospital size has an impact of over-utilization due to the supply-sensitive service phenomenon (Wennberg, 2002). Larger hospitals usually have larger fixed costs increasing the incentive to use more expensive treatments to keep returns. Consequently, a higher probability of c-sections in larger hospitals is expected. (iii) The physician specialty (Ob/Gyn specialty) captures the tendency of over-utilization in more specialized doctors. It is expected to observe more c-sections when the delivery is attended by a specialist. Finally, year dummy variables are also included in the estimation of the physician incentive equation to capture the trend related to non-observed factors.

The set of variables related to clinical factors is not discussed here and we refer to the specialized health service literature for details (see section 4.1). However, it is important to highlight that all these variables are related to risk of pregnancy and delivery, and therefore they are expected to increase the probability of a c-section.

4.3 Results

Table A2 in the Appendix reports the estimation using three econometric methods: The simple binary model (SBM), the structural misclassification model with independent errors (SMCI), and the structural misclassification model (SMC). Compared to the SMC model, the simple probit (SBM) understates both the impact of woman's health characteristics and physician incentives (column 1). However, the bias is greater for incentive related variables, and that is explained by the small probability of non-clinically required c-sections (misclassification probability) observed in the data (estimated at 3.2%). The difference between the SMCI model and the SMC model is small in spite of the high and statistically significant negative error correlation (column 3). As it was discussed in section 3, a negative error correlation is expected for over-utilized treatments.

For the whole period, the estimated marginal probability of physician incentives was 3.2%. This means that 3.2% of healthy, non-risk women had a non-clinically required cesarean section in the period between 1999-2002, meaning that each year around 2,500 women have unnecessary c-sections in New Jersey. Even though the percentage of unnecessary c-sections is relatively small, a more detailed inspection of the results shows a positive trend in the doctor's incentive equation given by the year dummy variables. As a consequence, it is expected that most of the growth in the observed c-section rates in recent years can be explained by physician incentives rather than changes in health conditions in the population. To test this hypothesis, we compute the probability of appropriate c-sections measured as the marginal probability of c-sections due to health conditions only (see section 3 for further discussion). This estimated c-section rate is the rate without any non-clinical influence, and therefore without misclassification. Figure 2 compares both, the observed and the only-health-related estimated c-section rate. According to this figure, c-section rates in New Jersey grew from 22.5% in 1999 to 27.5% in 2002. However, the rapid growth started in 1997 after a long period in which cesarean rates were about 20%. The only-health-related estimated c-section rate is based on 1999-2000 data. For that period, it is shown that the rapid growth of c-sections was explained mainly by non-clinically required c-sections. Without physician incentives, the c-section rate in New Jersey would have remained almost constant at around 22%, i.e. just above the levels observed before 1997 when c-sections soared, and more in line with the recommended rate of 15% of Healthy People and the World Health Organization (1985).



What are the determinants of physician incentives? The estimated incentive equation is shown in Table A2. We structure the discussion of these results according to the two types of variables described before, and using marginal calculations for each estimated parameter. We found the expected direction for all variables related to signaling of the degree of patient-obtained medical information. Income and ethnicity are observed by physicians as indicators of patient observed medical information. For the average household income, an increment of five thousand dollars reduces the probability of c-section by 0.10%. This small but significant impact was also observed by Pauly (1980) for ambulatory care. Ethnicity had an important impact on the probability of c-sections. Black and Hispanic women have respectively 2.20% and 2.30% higher probability of having a c-section compared with other non-white ethnicities. White women have a 2.10% lower probability of c-section delivery. These results are consistent with previous literature (Aron et al. 2000).

Social support measured as married (or joint in life) women reduces physician incentives, implying a 1.90% lower probability of a c-section. Compliance to physician's influence was captured with women's employment status. As we expected, full-time employed women have a 7% higher probability of c-section. Li et al. (2003) show also higher c-section rates for employed women.

The second type of variables related to factors that directly affect physician incentives also had the expected effect on c-sections. The most important and studied variable is payment source. With respect to non-HMO private insured patients, uninsured women are the least affected by doctor's incentives, with reduction in probability of c-section of around 9.20%. Medicaid beneficiaries are the next in lower influence with reduction in the probability of c-section of 3.40%. Finally, the capitated payment system of HMO reduces the probability of physician incentives, reducing probability of c-section in 1.30%. With respect to hospital size, our results confirm the supply-sensitive service hypothesis. In general, births in larger hospitals have higher probability of c-section. For a mid-size hospital, increasing births in 500 hundred per year raises the probability of a c-section in 0.10%. A similar argument is validated when we observe that women attended by more specialized physicians (Ob/Gym) have higher probabilities of c-section (2.60% more).

5 Conclusions

This paper develops an econometric method to estimate over- and under- utilization of medical procedures. When a physician has incentives that keep him from choosing the appropriate treatment for a patient, the patient's health status loses correspondence with the observed treatment. This generates a problem whose characteristics and effects on estimation are analogous to a classification error. However, this particular measurement error is not random. This paper proposes a structural model where the classification error is characterized by a physician behavior structure. That allows us to consistently estimate risk-adjusted utilization rates based on clinical factors only, and the probability of inappropriate treatments based on non-clinical factors (misclassification probability). Both measures can be neatly separated to test over- or under- healthcare utilization.

The results of the Monte Carlo study suggest that methodologies based on bivariate models (e.g. logit or probit) report biased estimates even when clinical or non-clinical factors are added as control variables. There are important interactions in the physician-patient relationship that can be captured by the structural misclassification model developed in this paper. We apply the model to cesarean section deliveries performed in New Jersey from 1999 to 2002. The results show that around 3.2% of healthy, non-risky women had c-sections not due to clinical factors but physician incentives. This rate implies that each year nearly 2,500 women have c-sections for non-medical reasons implying an excess cost of around $17.5 millions per year. Finally, it is estimated that physician incentives explain the rapid growth of c-section rates observed in New Jersey over these years.

The results of the applied section give direction for further research. A deeper analysis will be done using non-public data related to physician's and hospital's characteristics to understand the main drivers of physician incentives. Additionally, more complete clinical data will be incorporated to measure risk-adjusted utilization rates at the level of hospitals and physicians. A logit version of the model will be developed to be comparable with results from the health service research literature.

A second extension of the model is the reduction in the parametric assumptions. Because this rests heavily on the structure of physician behavior and distributional assumptions, there is a potential misspecification error that may set a bias on the estimation. A natural extension is to estimate the structural misclassification model semi-parametrically. In particular, because the model results in a partially observable bivariate model, a double-index semi-parametric model will be explored.

References

[1] Abowd, John and H. Farber, 1982. Job Queues and the Union Status of Workers. Industrial and Labor Relations Review 35, 354-367.

[2] Abrevaya, Jason, and J. Hausman, 1999. Semiparametric Estimation with Mismeasured Dependent Variables: An Application to Duration Models for Unemployment Spells. Mimeo.

[3] Amemiya, Takeshi, 1985. Advanced Econometrics. Harvard University Press.

[4] Aron, D., D. Harper, L. Shepardson and G. Rosenthal. 1998. Impact of risk- adjusting cesarean delivery rates when reporting hospital performance. Journal of American Medical Association 279, 1968-1972.

[5] Aron, D., H. Gordon, D. DiGiuseppe, D. Harper and G. Rosenthal. 2000. Variations in risk adjusted cesarean delivery rates according to race and health insurance. Medical Care 38, 35-44.

[6] Arrow, Kenneth, 1963. Uncertainty and the Welfare Economics of Medical Care. American Econonomic Review 53, 941-73.

[7] Das, Mitali. 2002. Is there evidence against the induced demand hypothesis? Explaining the large reduction in cesarean rates. Discussion Paper #:0102-40, Columbia University.

[8] De Jaegher, Kris and M. Jegers, 2001. The Physician–Patient Relationship as a Game of Strategic Information Transmission. Health Economics 10: 651–668.

[9] DiGiuseppe, D., D. Aron, S. Payne, R. Snow, L. Dierker and G. Rosenthal. 2001. Risk Adjusting Cesarean Delivery Rates: A comparison of hospital profiles based and medical record and birth certificate data. Health Services Research 36, 959-977.

[10] Dranove, David, 1988. Demand Inducement and the Physician/Patient Relationship Economic Inquiry 26, 281-298.

[11] Ford, Earl and R. Cooper, 1995. Racial/ethnic differences in health care utilization of cardiovascular procedures: a review of the evidence. Health Service Research 30, 237–252.

[12] Fuchs, Victor R., 1978. The Supply of Surgeons and the Demand for Operations. NBER Working Paper No. 0236.

[13] Fuchs V, M. McClellan M and J. Skinner, 2001. Area differences in utilization of medical care and mortality among U.S. elderly. NBER Working Paper No. 8628.

[14] Hausman, J., J. Abrevaya, and F.M. Scott-Morton, 1998. MisclassiÞcation of the dependent variable in a discrete-response setting. Journal of Econometrics 87, 239-269.

[15] Gruber, J. and M. Owings. 1996. Physician financial incentives and cesarean section delivery. RAND Journal of economics 27: 99 – 123.

[16] Ichimura, Hidehiko and L. Lee, 1991. Semiparametric least squares estimation of multiple index models: Single equation estimation. In Nonparametric and semiparametric methods in econometrics and statistics. Barnett, Powell and Tauchen (Eds.). Cambridge University Press.

[17] Iezzoni, Lisa (Ed.), 2003. Risk Adjustment for Measuring Healthcare Outcomes. Third edition. AcademyHealth/HAP.

[18] Keeler, D., R. Park, R. Bell, G. Spelliscy and J. Keesey. 1997. Adjusting Cesarean Delivery Rates for Case mix. HSR: Health Services Research 32, 511-528.

[19] Kenkel, Donald, D. Lillard and A. Mathios, 2004. Accounting for misclassification error in retrospective smoking data. Health Economics 13, 1031-1044.

[20] Kressin, Nancy and L. Petersen, 2001. Racial Differences in the Use of Invasive Cardiovascular Procedures: Review of the Literature and Prescription for Future Research. Annals of Internal Medicine 135, 352-366.

[21] Leape, Lucian, J. Weissman, E. Schneider, R. Piana, C. Gatsonis, and A. Epstein, 2003. Adherence to Practice Guidelines: The Role of Specialty Society Guidelines. American Heart Journal 145, 19-26.

[22] Lewbel, Arthur, 2000. Identification of the Binary Choice Model with Misclassification. Econometric Theory 16, 603-609.

[23] Li, T., G. Roads, J. Smulian, K. Demissie, D. Wartenberg and L. Kruse, 2003. Physician Cesarean Delivery rates and risk adjusted perinatal outcomes. Obstetrics & Gynecology 101, 1204-1212.

[24] Magder, Laurence and J. Hughes, 1997. Regression when the Outcome is Measured with Uncertainty. American Journal of Epidemiology 146, 195-203.

[25] McGuire, Thomas G. 2000. Physician agency. In Handbook of Health Economics, A.J. Culyer and J.P. Newhouse (Eds.), Volume 1, Part 1, 461-536.

[26] Meng, Chun-Lo and P. Schmidt, 1985. On the Cost of Partial Observability in the Bivariate Probit Model. International Economic Review 26, 71-85.

[27] Pauly, Mark, 1980. Doctors and Their Workshops: Economic Models of Physician Behaviour. NBER monograph.

[28] Poirier, Dale, 1980. Partial Observability in Bivariate Probit Models. Journal of Econometrics 12, 209-217.

[29] Rahnama, P., S. Ziaei, and S. Faghihzadeh. 2006. Impact of early admission in labor on method of delivery. International Journal of Gynecology and Obstetrics 92, 217-220.

[30] Schneider, Eric, L. Leape, J. Weissman, R. Piana, C. Gatsonis, and A. Epstein, 2001. Racial Differences in Cardiac Revascularization Rates: Does "Overuse" Explain Higher Rates among White Patients?. Annals of Internal Medicine 135, 328-337.

[31] Tussing, D. y M. Wojtowycz. 1993. The effect of physician characteristics on clinical behavior: Cesarean section in New York State. Social Science and Medicine 37, 1251-1260.

[32] Van Ryn, M. and J. Burke, 2000. The effect of patient race and socio-economic status on physicians' perceptions of patients. Social Science and Medicine 50, 813-828.

[33] Wennberg, John E., 2002. Unwarranted variations in healthcare delivery: Implications for academic medical centres. BMJ 325, 961-964.

[34] World Health Organization, 1985. Appropriate technology for birth. Lancet 2, 436-437.

[35] Xie, Bin, D. Dilts and M. Shor, 2006. The physician-patient relationship: The impact of patient-obtained medical information. Health Economics 15, 813–833.

6 Appendix

Table A1. Sample mean of health and non-health related variables: New Jersey 1999-2002 (Percentages unless noted)

Variable	Vaginal Delivery	Cesarean Section	Full Sample
Cesarean delivery	0.00	100.00	24.83
Clinical Variables			
Age (years)	28.62	30.47	29.08
Previous cesarean delivery	4.01	42.00	13.44
Multiple gestation	0.71	2.84	1.24
Admission by emergency	5.58	3.30	5.02
Long labor	0.76	0.78	0.77
Elderly primigravida $\geq$ 35 y.o.	0.69	1.39	0.86
Breech or transverse lie presentation	2.40	22.98	7.51
Diabetes	3.45	5.53	3.97
Hypertension	3.24	3.57	3.33
Pre-eclampsia	1.62	1.42	1.57
Oligohydramnios	0.21	0.21	0.21
Polyhydramnios	0.30	0.94	0.46
Abruptio placenta	0.47	0.60	0.50
Full or partial placenta previa	0.10	0.87	0.29
Patient and Physician related variables			
Woman is married	65.90	71.48	67.28
Woman is full time employed	34.85	40.46	36.24
White (non-Hispanic)	43.43	44.16	43.61
Black (non-Hispanic)	13.19	12.14	12.93
Hispanic	17.40	17.46	17.42
Zip code mean household income (thousands, $)	56.00	57.15	56.29
Patient payment (uninsured)	8.22	6.41	7.77
Medicaid payment	10.62	8.60	10.11
HMO payment	57.04	58.01	57.28
Yearly average of births in Hospital (thousands)	2.48	2.61	2.51
Obs/Gyn Physician	89.67	91.09	90.02
Number of observations			
Total	303,434	100,226	403,660
Year 1999	76,610	22,193	98,803
Year 2000	77,571	24,371	101,942
Year 2001	76,273	26,027	102,300
Year 2002	72,980	27,635	100,615

**Table A2: Model estimation of cesarean section deliveries.
New Jersey 1999-2002**

Variables	Simple Binary Model with Controls (1)	Structural Misclassification w/indp. errors (2)	Structural Misclassification Model (3)
Clinical variables			
Age	0.009 *	0.010 *	0.010 *
	(0.001)	(0.001)	(0.001)
Previous cesarean delivery	1.856 *	1.982 *	1.950 *
	(0.009)	(0.011)	(0.010)
Multiple gestation	0.487 *	0.517 *	0.507 *
	(0.027)	(0.023)	(0.022)
Admission by emergency	-0.207 *	-0.290 *	-0.277 *
	(0.015)	(0.018)	(0.017)
Long labor	0.244 *	0.251 *	0.248 *
	(0.034)	(0.030)	(0.030)
Elderly primigravida (35+ years old)	0.518 *	0.605 *	0.590 *
	(0.031)	(0.026)	(0.025)
Breech or transverse lie presentation	1.702 *	1.830 *	1.799 *
	(0.012)	(0.012)	(0.011)
Diabetes	0.209 *	0.241 *	0.235 *
	(0.015)	(0.014)	(0.014)
Hypertension	0.107 *	0.121 *	0.118 *
	(0.017)	(0.016)	(0.015)
Pre-eclampsia	0.063 *	0.067 *	0.065 *
	(0.025)	(0.024)	(0.023)
Oligohydramnios	0.085 ***	0.055	0.057
	(0.067)	(0.062)	(0.061)
Polyhydramnios	0.692 *	0.764 *	0.748 *
	(0.044)	(0.035)	(0.035)
Abruptio placenta	0.115 *	0.082 **	0.083 **
	(0.042)	(0.039)	(0.038)
Full or partial placenta previa	1.357 *	1.476 *	1.448 *
	(0.060)	(0.046)	(0.046)
Intercept	-1.510 *	-1.669 *	-1.659 *
	(0.023)	(0.023)	(0.023)

Dependent variable is mode of delivery. 1 if it was a cesarean section, 0 if it was a vaginal delivery.
Estimation was done in GAUSS. Program code is available under request.
Standard errors in parenthesis.
* Significant at 1%. ** Significant at 5%. *** Significant at 10%

Table A2: Model estimation of cesarean section deliveries. New Jersey 1999-2002 (Continued...)

Variables	Simple Binary Model with Controls (1)	Structural Misclassification w/indp. errors (2)	Structural Misclassification Model (3)
Patient and Physician related variables			
Woman is married	-0.042 *	-0.123 *	-0.111 *
	(0.008)	(0.020)	(0.019)
Woman is full time employed	0.153 *	0.419 *	0.417 *
	(0.007)	(0.021)	(0.021)
White (non-Hispanic)	-0.035 *	-0.121 *	-0.115 *
	(0.008)	(0.022)	(0.021)
Black (non-Hispanic)	0.017 ***	0.150 *	0.131 *
	(0.011)	(0.026)	(0.025)
Hispanic	0.063 *	0.129 *	0.131 *
	(0.010)	(0.025)	(0.024)
Zip code mean household income	-0.002 *	-0.004 *	-0.004 *
	0.000	0.000	0.000
Patient payment (uninsured)	-0.139 *	-0.449 *	-0.459 *
	(0.013)	(0.056)	(0.057)
Medicaid payment	-0.068 *	-0.169 *	-0.175 *
	(0.012)	(0.030)	(0.031)
HMO payment	-0.021 *	-0.073 *	-0.070 *
	(0.007)	(0.017)	(0.017)
Yearly average of births in Hospital	0.010 *	0.026 *	0.027 *
	(0.002)	(0.005)	(0.005)
Ob/Gyn Physician	0.031 *	0.157 *	0.163 *
	(0.011)	(0.037)	(0.037)
Year 2000	0.052 *	0.135 *	0.144 *
	(0.009)	(0.027)	(0.027)
Year 2001	0.106 *	0.197 *	0.228 *
	(0.009)	(0.027)	(0.027)
Year 2002	0.177 *	0.366 *	0.397 *
	(0.009)	(0.027)	(0.027)
Intercept	-	-2.058 *	-2.125 *
	-	(0.067)	(0.068)
Correlation	-	-	-0.422 *
	-	-	(0.018)
Degree of physician incentives	-	0.034	0.032
(Mean of marginal probability)	-	(0.012)	(0.012)
Log-Likelihood function	-159,942.87	-160,363.97	-160,307.28
Number of Observations	403,660	403,660	403,660

Dependent variable is mode of delivery. 1 if it was a cesarean section, 0 if it was a vaginal delivery.
Estimation was done in GAUSS. Program code is available under request.
Standard errors in parenthesis.
* Significant at 1%. ** Significant at 5%. *** Significant at 10%