Econometrica, Vol. 71, No. 4 (July, 2003), 1215–1238

BAYESIAN INFERENCE FOR HOSPITAL QUALITY IN A SELECTION MODEL

BY JOHN GEWEKE, GAUTAM GOWRISANKARAN, AND ROBERT J. TOWN[1]

This paper develops new econometric methods to infer hospital quality in a model with discrete dependent variables and nonrandom selection. Mortality rates in patient discharge records are widely used to infer hospital quality. However, hospital admission is not random and some hospitals may attract patients with greater unobserved severity of illness than others. In this situation the assumption of random admission leads to spurious inference about hospital quality. This study controls for hospital selection using a model in which distance between the patient's residence and alternative hospitals are key exogenous variables. Bayesian inference in this model is feasible using a Markov chain Monte Carlo posterior simulator, and attaches posterior probabilities to quality comparisons between individual hospitals and groups of hospitals. The study uses data on 74,848 Medicare patients admitted to 114 hospitals in Los Angeles County from 1989 through 1992 with a diagnosis of pneumonia. It finds the smallest and largest hospitals to be of the highest quality. There is strong evidence of dependence between the unobserved severity of illness and the assignment of patients to hospitals, whereby patients with a high unobserved severity of illness are disproportionately admitted to high quality hospitals. Consequently a conventional probit model leads to inferences about quality that are markedly different from those in this study's selection model.

KEYWORDS: Bayesian inference, hospital quality, simultaneous equations, MCMC, Medicare, pneumonia, mortality.

1. INTRODUCTION

THIS PAPER DEVELOPS new econometric methods to estimate hospital quality and other models with discrete dependent variables and nonrandom selection. Assessing the quality of care in hospitals is an important problem for public policy and a challenge for applied econometrics.[2] Policy changes in Medicare reimbursement rates and the rise of managed care as well as technological innovations

[1] We thank Pat Bajari, Lanier Benkard, Gary Chamberlain, Mike Chernew, Tom Holmes, Steven Stern, a co-editor, anonymous referees, and seminar participants at Duke, Federal Reserve Bank of Chicago, Georgetown, Harvard/MIT, Iowa, Michigan, Princeton, Stanford, UC Davis, Irvine, and Riverside, Virginia, Yale, the Econometric Society Seventh World Congress, and the Society of Economic Dynamics 1999 Annual Meeting for helpful comments, and Anita Todd for editorial assistance. Any remaining errors are the sole responsibility of the authors. The first author acknowledges support from NSF Grant SBR-9819444, and the second author acknowledges support from the University of Minnesota Supercomputer Institute and from the Federal Reserve Bank of San Francisco. The views expressed herein are solely those of the authors and do not represent those of the Federal Reserve System, the Federal Reserve Bank of San Francisco, or any other institution.

[2] As described by a leading study, "Quality of care is the degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge. . . ," Lohr (1990, p. 4).

have affected hospital incentives and, through that, hospital quality.[3] These quality changes have large welfare effects and hence the potential for large deadweight losses.[4]

Hospital patient discharge databases provide several indicators plausibly associated with hospital quality. Since they cover large numbers of patients and hospitals and are much less expensive to obtain and access than other sources of information, they have been widely used in comparisons of hospital quality. Mortality has been the most popular indicator of hospital quality in the literature: it is unambiguously defined and its link with quality of care is so strong as to be tautological.[5]

In this widely used framework, the conceptual experiment, hospital-specific mortality rates following random assignment of a population of patients to hospitals, reveals hospital quality. Patients, however, are not randomly assigned to hospitals. Patients or their physicians are likely to choose hospitals based on factors such as location, convenience, and severity of illness. If assignment were nonrandom, although still random conditional on observed characteristics, then conventional dichotomous outcome models could be used to infer the outcome of the conceptual experiment from the available data. However, discharge data contain only crude summaries of medically pertinent information and hence many aspects of the severity of illness are unobserved. Thus, the assumption of random conditional assignment is not tenable and patients with the same observed characteristics are not equally likely to be admitted to all hospitals. For instance, if patients with high unobserved severity of illness select high quality hospitals, then observed mortality rates for high quality hospitals will be inconsistent and upwardly biased measures of mortality from the conceptual experiment. This will be true even after controlling for observed measures of severity of illness. Conventional statistical methods that ignore unobserved severity will produce misleading inferences about hospital quality. This has led prominent medical experts to make a pessimistic assessment of the usefulness of discharge data in assessing hospital quality.[6]

Recent work by Gowrisankaran and Town (1999) developed a framework to control for the nonrandom assignment of patients. This work modeled mortality as a function of indicator variables for each hospital and patient discharge information. The authors treat mortality as continuous and directly apply linear instrumental variables methods. The identifying assumption is that a patient's

[3] See Cutler (1995), Kessler and McClellan (2000), and McClellan and Noguchi (1998) for studies of the effects on medical outcomes of Medicare policy, of managed care, and of technological change, respectively.

[4] For instance, if changes in Medicare policies caused hospitals to reduce their pneumonia mortality rates by one percentage point, this would translate to over 6,000 lives saved annually in the U.S.

[5] Strictly speaking mortality is an indicator of hospital mediocrity; mortality is an inverse indicator of quality. Subsequently we provide a precise definition of hospital quality in the context of the model developed in this study.

[6] Leading medical researchers, including Iezzoni et al. (1996), and government studies (U.S. GAO (1994)) have both argued that discharge databases are problematic, for this reason.

mortality is not affected by how far that patient's residence is from alternative hospitals. Combined with the demonstrable fact that patients are more likely to choose hospitals that are closer to home, other things equal, the conventional conditions for consistency of instrumental variables estimation in a linear model are satisfied. Conceptually, the estimator would predict hospital A to be of higher quality than hospital B if patients residing near hospital A have lower mortality than patients residing near hospital B, after controlling for their medical and demographic characteristics.

The difficulty with this approach is that because the outcome variable, mortality, is dichotomous, any internally consistent model of hospital quality and choice must be nonlinear. This paper develops a logically coherent model designed to infer the outcome of the conceptual experiment that randomly assigns patients to hospitals, given data that has nonrandom patient assignment.[7] Inference with this model is challenging because the amount of information per observation is small.[8] This paper develops an approach to inference in this model that is practical with the large data sets required to extract signal from noise in hospital patient discharge databases. This approach is potentially applicable to a wide range of policy evaluations of economic interest where the outcome variable is dichotomous.[9]

The model developed here incorporates hospital choice and mortality as endogenous variables and fixed hospital and patient characteristics as exogenous variables. Hospital choice is described by a multinomial probit model and mortality by a binary probit model. The mortality model includes indicator variables for each hospital to accommodate hospital-specific differences in quality as well as demographic variables and observed disease characteristics. The mortality model is structural in the sense that it predicts outcomes given alternative assignments of patients to hospitals including random assignment. The multinomial probit model is a reduced form relationship that provides probabilities of hospital choice conditional on observed covariates that are a function of demographic characteristics and distance of the hospital from the patient's home. The random component in the binary probit model includes unobserved severity of illness and is permitted to be correlated with the random component in the multinomial choice model. If, after controlling for the observed covariates in the hospital choice model, patients with high unobserved severity of illness are more likely to be admitted to hospital A than patients with low unobserved severity, this will imply a positive correlation between the shock in the mortality equation and the shock in the hospital A choice equation.

[7] Though the methods of Gowrisankaran and Town (1999) are much simpler than the ones developed in this paper, there is no formal statistical model that rationalizes their approach.

[8] Simple measures of fit always indicate that most variation in mortality cannot be ascribed to covariates. Even if all the difference in mortality rates were attributable to quality, the variation in these rates is small.

[9] Examples include the effect of school performance based on graduation rates, of prison rehabilitation programs based on recidivism rates, and of job training programs based on the incidence of harassment complaints, and many medical outcome evaluations.

We estimate this selection model using Bayesian inference from data on 74,848 Medicare patients admitted to 114 hospitals in Los Angeles County during the period from 1989 to 1992 with a diagnosis of pneumonia. By transforming the integration problem posed by the latent variables into a simulation problem, our approach to inference computes the parameter estimates orders of magnitude faster than the method of maximum likelihood. This makes inference feasible for this type of simultaneous equations model.[10] The basis for the simulation procedure is the fact that the model is similar to the conventional linear simultaneous equations model conditional on latent variables. Using Markov chain Monte Carlo techniques, we iteratively simulate latent variable values conditional on data and parameters, and parameters conditional on data and latent variables. The second step is computationally similar to classical instrumental variables, differing principally in the appearance of the discrete hospital choice in the mortality probit equation, which does not pose a problem. The simulation methods simultaneously recover the joint posterior distribution of parameters and latent variables.[11] Albert and Chib (1993) used this approach in the binary probit model and Geweke, Keane, and Runkle (1997) extended it to the multinomial probit model. The methods developed here extend this approach to a new class of models.

We use these methods to address the motivating policy questions directly. First, to what extent is hospital quality associated with observed characteristics of hospitals, such as size and ownership status? Second, with what degree of confidence can it be said that one hospital is of higher quality than another? We model hospital quality using hierarchical priors. This approach, which combines some characteristics of classical fixed- and random-effects models, specifies the quality of each hospital as a separate parameter, but assigns a more important role to the data in determining whether these parameters are similar for hospitals with similar observable characteristics, relative to a normal prior. Our approach provides an efficient method for extracting the signal from the noise, which is particularly important given this type of data.

The remainder of the paper is organized as follows. Section 2 provides the specification of the model and methods for inference. The database is described in Section 3. Section 4 presents findings on hospital quality and the role of nonrandom admission to hospitals. Section 5 concludes. Five appendices are available in the working paper version of this paper.[12] Appendix A1 details the construction of the prior, Appendix A2 details the likelihood function and

[10] Maximum likelihood evaluation for one parameter vector for one individual would require evaluating the joint density of the mortality and hospital choice outcomes for that individual. Given that we have 114 endogenous variables and that the mortality error and hospital choice error are correlated, this would take several minutes on a fast supercomputer. Multiplied by a data set of roughly 75,000 patients (necessary because of the small signal to noise ratio), it would take months to evaluate the likelihood for a single parameter vector.

[11] Surveys that discuss convergence to the posterior include Chib and Greenberg (1996) and Geweke (1997, 1999).

[12] See the NBER working paper Geweke, Gowrisankaran, and Town (2001).

computation, Appendix A3 gives evidence on the numerical accuracy of our Markov chain Monte Carlo algorithm, Appendix A4 provides posterior rankings for the hospitals in our data set, and Appendix A5 provides robustness results with alternative priors.

2. THE MODEL

The central component of the model is a structural probit equation, in which the probability of mortality is a function of the hospital to which a patient is admitted, the observed severity of the patient's illness, and the observed demographic characteristics of the patient. The objective is to learn about the way the hospital to which the patient is admitted influences the probability of mortality in this equation. A multinomial probit model of hospital admission supplements the mortality model, to permit nonrandom assignment of patients to hospitals. This section describes, in turn, the specification of the model, the prior distribution of the model parameters, and methods to recover the posterior distribution of these parameters.

2.1. *Model Specification*

Let $i = 1, \ldots, n$ index the patients in the sample, and let $j = 1, \ldots, J$ index hospitals in the sample. There are two groups of exogenous variables in the model. The $k \times 1$ vector x_i consists of individual characteristics of patient i that may affect mortality, including indicators for age, race, sex, and disease stage, and measures of income. The $q \times 1$ vector z_{ij}, which consists of characteristics specific to the combination of individual i and hospital j, includes distance between the home of patient i and hospital j and interactions of distance with observable patient characteristics. The specifics of these variables are given in Section 3.

There are two sets of endogenous variables in the model. The mortality indicator m_i is 1 if the patient dies in the hospital within ten days of admission and is 0 otherwise. The $J \times 1$ indicator vector c_i has jth entry 1 if patient i is admitted to hospital j, and 0 otherwise.

To present the structural mortality equation, let ε_i $(i = 1, \ldots, n)$ be independent $N(0, \sigma^2)$ random variables conditional on the exogenous variables. The mortality probit m_i^* is a latent random variable,

$$m_i^* = c_i'\beta + x_i'\gamma + \varepsilon_i. \tag{1}$$

The mortality indicator m_i equals 1 if $m_i^* > 0$ and 0 if $m_i^* \leq 0$. The structural interpretation of (1) is that if patient i were randomly assigned to hospital j, then $m_i^* = \beta_j + x_i'\gamma + \varepsilon_i$ and consequently $P(m_i = 1) = \Phi((\beta_j + x_i'\gamma)/\sigma)$. Note that the parameters β and σ are jointly unidentified in (1) because they can be scaled by the same arbitrary positive constant without changing the behavior of m_i. In the conventional probit model this problem is avoided by setting $\sigma = 1$. We return to this matter in the context of the complete model below.

If c_i were in fact independent of ε_i—as it would be if patients were randomly assigned to hospitals, for example—then c_i would be exogenous in (1). After resolution of the above identification issue this model would conform with the conventional textbook specification of the binary probit model. However, it is likely that in observed data, c_i depends in part on ε_i: the admission of patient i to hospital j takes into account information that is correlated with ε_i. The conventional probit model is then misspecified.

To develop a more plausible model of hospital choice, we assume that patients become infected with one of the many bacterial or viral agents that can cause pneumonia and it has been determined that they are sufficiently ill to benefit from inpatient treatment. At that point the patient (or the patient's agent) selects from the set of J hospitals the hospital to which the patient will be admitted. The actual choice decision will be a complex function of many factors, such as severity of illness, characteristics of the hospital, the patient's primary care physician, etc. One important observable influence on choice is distance: previous research has shown that the farther a patient is from a hospital, the less likely the patient is to be admitted to that hospital, other observables constant.[13]

To present the reduced form model of hospital choice, define the $J \times q$ matrix $\widetilde{Z}_i$, $\widetilde{Z}_i = [z_{i1}, z_{i2}, \ldots, z_{iJ}]'$. Let the $J \times 1$ vectors $\tilde{\eta}_i \sim N(0, \widetilde{\Sigma})$ $(i = 1, \ldots, n)$ be mutually independent conditional on the exogenous variables, and let $\tilde{\rho}_j$, $j = 1, \ldots J$, denote the correlation between ε_i and $\tilde{\eta}_{ij}$. Define the $J \times 1$ hospital choice latent vector multinomial probit $\tilde{c}_i^* = (\tilde{c}_{i1}^*, \ldots, \tilde{c}_{iJ}^*)'$ as

$$\tilde{c}_i^* = \widetilde{Z}_i \alpha + \tilde{\eta}_i. \tag{2}$$

The choice indicator vector $c_i = (c_{i1}, \ldots, c_{iJ})'$ has entry $c_{ij} = 1$ if $\tilde{c}_{ij}^* \geq \tilde{c}_{ik}^*$ $(k = 1, \ldots, J)$ and $c_{ij} = 0$ otherwise. As above with (1), the parameters α and $\widetilde{\Sigma}$ are jointly unidentified since scaling α by any positive constant and $\widetilde{\Sigma}$ by the square of that constant leaves the distribution of c_i conditional on Z_i unaffected. We return to this matter in the context of the prior distribution in Section 2.2.

As is customary in models with J choices, it is easier to work with $J-1$ latent utilities, and normalize the Jth utility to 0. Accordingly, we define the $(J-1) \times q$ matrix $Z_i = [\tilde{z}_{i1} - \tilde{z}_{iJ}, \tilde{z}_{i2} - \tilde{z}_{iJ}, \ldots, \tilde{z}_{i,J-1} - \tilde{z}_{iJ}]'$, the $(J-1) \times 1$ vectors $\eta_i = [\tilde{\eta}_{i1} - \tilde{\eta}_{iJ}, \ldots, \tilde{\eta}_{i,J-1} - \tilde{\eta}_{iJ}]'$ and $c_i^* = [\tilde{c}_{i1}^* - \tilde{c}_{iJ}^*, \ldots, \tilde{c}_{i,J-1}^* - \tilde{c}_{iJ}^*]'$, and the $(J-1) \times (J-1)$ matrix $\Sigma = \operatorname{var}(\eta_i)$. Note that

$$c_i^* = Z_i \alpha + \eta_i. \tag{3}$$

If the unobserved severity of illness affects hospital choice, the mortality and choice error terms will be correlated. Let ρ_j denote the correlation between ε_i and η_{ij} $(j = 1, \ldots, J-1)$. The larger is ρ_j, the more likely is a patient with a high unobserved severity of illness (ε_i) to be admitted to hospital j. Thus we shall refer to ρ_j as the *hospital j severity correlation*. The hospital severity correlations

[13] See Luft et al. (1990) and Burns and Wholey (1992).

are a useful way to characterize severity of illness by hospital since they are independent of the scale of ε_i, which we know from (1) is unidentified.

Now, we can write the variance of the joint error terms as

$$\text{(4)} \qquad \text{var}(\varepsilon_i, \eta_i') = \begin{bmatrix} \sigma^2 & \pi' \\ \pi & \Sigma \end{bmatrix}$$

where π is a $(J-1)\times 1$ vector with $\pi_j = \rho_j \sigma \Sigma_{jj}^{1/2}$.

To permit unobserved severity of illness to affect hospital choice in any way consistent with the model, the only restriction we place on π is that var (ε_i, η_i') be positive definite. Since this implies complicated restrictions on π, a more graceful treatment is to work with the population regression of the shock ε_i in (1) on the shock vector η_i in (3):

$$\text{(5)} \qquad \varepsilon_i = \eta_i'\delta + \zeta_i; \quad \text{cov}(\eta_i, \zeta_i) = 0.$$

In this regression δ is a $(J-1)\times 1$ parameter vector and the scale of ε_i is normalized by var $(\zeta_i) = 1$. This specification simultaneously resolves the identification problem due to the scaling in (1) and incorporates all permissible values of $\pi = \Sigma\delta$ in (4).

With this reparameterization, the variance of the shock in the mortality probit equation is $\sigma^2 = \delta'\Sigma\delta + 1$, and the correlation between ε_i and η_{ij} is

$$\text{(6)} \qquad \rho_j = \left(\sum_{k=1}^{J-1} \delta_k \Sigma_{kj}\right) / [\Sigma_{jj}(\delta'\Sigma\delta + 1)]^{1/2}.$$

In the hypothetical experiment in which patient i is admitted to hospital j by means of a random assignment c_i, $P(m_i = 1|x_i)$ is equal to $\Phi[(c_i'\beta + x_i'\gamma)/(\gamma'\Sigma\delta + 1)^{1/2}]$. We shall refer to

$$\text{(7)} \qquad q_j = -\beta_j/(\delta'\Sigma\delta + 1)^{1/2},$$

as the *hospital j quality probit*. Differences in these probits across hospitals may be used to address quality comparisons for individual hospitals. In the conventional probit model with normalization $\sigma = 1$, the hospital j quality probit is $q_j^* = -\beta_j$. To compare groups of hospitals, define $q_G = \sum_{j\in G} \omega_j q_j$, where G is the group of interest and the weight ω_j is proportional to the number of patients admitted to hospital j; define ρ_G and q_G^* analogously.

2.2. *Prior Distributions*

The number of free parameters in Σ is $J(J-1)/2 - 1$, that is, 6,441 in our sample with $J = 114$ hospitals. We make one major simplification, that $\widetilde{\Sigma} = I_J$, so that after differencing, $\Sigma = I_{J-1} + e_{J-1}e_{J-1}'$, where e_n denotes an $n\times 1$ vector of units. We introduce some evidence on the plausibility of this assumption in

Section 4.4. Estimating these parameters would increase the computation time by orders of magnitude and also complicate our MCMC simulation algorithm.[14]

We choose independent prior distributions for the parameter vectors, α, δ, γ, and β so as to include all reasonable parameter values well within their support. We discuss specific aspects of these priors here.[15]

First, we utilize a variance component structure and a hierarchical prior to specify that hospital qualities are similar ex ante while allowing the data to determine the degree of similarity ex post. Each hospital, j, is in one of four ownership categories, k, and one of four size categories, l, detailed in Section 3.2. If hospital j is of ownership category k and size category l, then decompose $\beta_j = \beta_1 + p_k + s_l + u_j$. The prior distributions of the components $\beta_1, p_1, \ldots, p_4, s_1, \ldots, s_4$, and $u_1, \ldots, u_{114}$ are jointly Gaussian, mean zero, and mutually independent. The common term β_1 has standard deviation 3 (essentially a flat prior). The other components have variances τ_p^2, τ_s^2, and τ_u^2, respectively, grouped together in the vector $\tau' = (\tau_p^2, \tau_s^2, \tau_u^2)$. Given τ, the prior specifies that hospital quality is more strongly correlated between hospitals that share the same size or ownership specification. However, we employ a hierarchical prior distribution with the variance terms having independent prior distributions $1.25/\tau_j^2 \sim \chi^2(5)$ $(j = p, s, u)$ in the standard probit model.[16]

Second, since $\eta_{ij} = \tilde{\eta}_{ij} - \tilde{\eta}_{iJ}$, an iid prior on δ implies a prior on $\tilde{\rho}$ that is not exchangeable with respect to the Jth hospital, which is undesirable since the numbering of hospitals is arbitrary. We use the prior $\delta \sim N(0, \sigma_\delta^2 \Sigma^{-1})$ with $\sigma_\delta = 0.196$, which implies an exchangeable and diffuse prior for $\tilde{\rho}$.[17]

Third, the priors for the selection model need to be carefully scaled relative to the conventional probit model to account for the different values of σ across the models. From (5), $\sigma^2 = \delta' \Sigma \delta + 1$ in the selection model, but $\sigma = 1$ in the probit model. Since $\delta \sim N(0, \sigma_\delta^2 \Sigma^{-1})$ it follows that $\delta' \Sigma \delta + 1 \sim \sigma_\delta^2 \chi^2(J-1) + 1$ and $E(\delta' \Sigma \delta + 1) = \sigma_\delta^2 (J-1) + 1$. Thus, in the hierarchical hospital quality prior in the selection model, $1.25[\sigma_\delta^2 (J-1) + 1]/\tau_j^2 \sim \chi^2(5)$. Similarly, we scale the selection model prior standard deviations for β_1 and γ by $[\sigma_\delta^2 (J-1) + 1]^{1/2}$ relative to the probit model.

The choice of the prior distributions of α and γ is relatively straightforward. As with β and δ the governing principle is that reasonable values must be well within the support of the prior distribution, and care must be taken to maintain the same scale in the probit and selection models. With respect to the last consideration, note in particular that the impact of covariates in the selection model,

[14] Keane (1992) shows that Σ is the source of irregularity in the multinomial probit likelihood function.

[15] Appendix A1 of the working paper version of this paper (Geweke, Gowrisankaran, and Town (2001)) contains detailed descriptions of all the priors.

[16] The centered 99% prior credible interval for each τ_j^2 is (.22, 1.7). Robustness of our results with respect to variation in these and other priors is summarized in Section 4.4 and is detailed in Appendix A5 of the working paper.

[17] Appendix A1 of the working paper documents further details of this prior distribution including the reasoning leading to the choice $\sigma_\delta = 0.196$.

corresponding to γ in the probit model, is $\gamma/(\delta'\Sigma\delta+1)^{1/2}$ by means of the same reasoning leading to (7).

2.3. *Inference*

The observed data are $(x_i, Z_i, c_i, m_i, i = 1, \ldots, n)$, which can be abbreviated as y. The model contains latent variables $(m_i^*, c_i^*, i = 1, \ldots, n)$, which can be abbreviated as y^*. The parameter vectors are α, β, γ, and δ, which can be collected in the vector θ. The model specified in Section 2.1 provides the density $p(y, y^*|\theta)$ and the prior distributions in Section 2.2 provide $p(\theta)$. Explicit expressions for these densities are given in Appendix A2 of the working paper. From Bayes rule, the distribution of the unobservables y^* and θ conditional on the data and model specification is

$$p(y^*, \theta|y) = p(\theta)p(y, y^*|\theta)/p(y) \propto p(\theta)p(y, y^*|\theta). \tag{8}$$

The objective is to obtain the posterior distribution of functions such as the hospital quality probits q_j, and $\Phi(-q_j + x_i'\gamma)$, the probability of mortality under random hospital admission of a patient with observed characteristics x_i to hospital j. This objective requires integrating a highly nonlinear function over millions of dimensions, most of which correspond to latent variables. This cannot be accomplished analytically.

The parameter vector and latent variables can be partitioned into groups, such that the posterior distribution of any one group conditional on all the others is of a single, easily recognized form that is easy to simulate. Details of the partition are given in Appendix A2 of the working paper. The problem is then well suited to attack by execution of a Gibbs sampling algorithm (Gelfand and Smith (1990), Geweke (1999)). In this approach, each group of parameters and latent variables is simulated conditional on all the others. Following each pass through the entire vector of latent variables and parameters, all parameter values are recorded in a file.

As detailed in Appendix A2 of the working paper, the Gibbs sampling algorithm is ergodic and its unique limiting distribution is the posterior distribution. Therefore, dependent draws from the posterior distribution of any function of the parameters $g(\theta)$ can be made by computing the value of g corresponding to the recorded parameter values, after discarding initial iterations of the Gibbs sampling algorithm to allow for convergence. We use parallel computing methods and a supercomputer, exploiting the fact that in each iteration of the Gibbs sampling algorithm the latent variables $(m_i^*, c_i^*, i = 1, \ldots, n)$ are conditionally independent across individuals. The iterations themselves are executed serially. The results reported in Section 4 are based on every tenth draw from 19,000 successive iterations (a total of 1,900 draws), after discarding 1,000 burn-in iterations based on convergence diagnostics. For comparison purposes, we apply the same procedures to a conventional probit model for mortality, using the Gibbs sampling algorithm described in Albert and Chib (1993). Appendix A3 of the working paper provides details on the numerical accuracy of our Gibbs sampling algorithm.

3. THE DATA

The primary source of data for this study is the Version B Discharge Data from the State of California Office of Statewide Health Planning and Development. These data provide records for all patients discharged from any California acute-care hospital during the years 1989 through 1992. We confine our attention to patients who were over 65 at the time of admission. During this time period, the vast majority of patients over 65 were covered by traditional Medicare fee-for-service insurance, which has standardized hospitalization benefits. We confine our attention to Los Angeles County. A large metropolitan area is best suited to our purposes because it has a large base of patients and contains multiple hospitals in every size and ownership class. We limit our study to a single disease, because there is evidence that the relation between mortality and covariates is disease specific.[18] We choose pneumonia in particular for three reasons. First, it is a common disease[19] that provides the large sample needed to draw inferences about hospital quality. Second, in-hospital death is a relatively frequent outcome for pneumonia patients, which makes it a relevant disease to examine through the medium of hospital discharge records. Third, there is independent evidence that an appropriately adjusted in-hospital mortality rate for pneumonia is correlated with the quality of in-hospital care.[20]

The secondary source of data is the Annual Survey of Hospitals Database published by the American Hospital Association (AHA). Among other information, the AHA data contain the addresses, ownership status, and size of each hospital in our sample.

3.1. *Sample Construction*

The sample was selected through a process of eliminating patients from the 1989–1992 Version B Discharge Data. The first qualification for selection is that the patient live in a Los Angeles County zip code, be admitted to a Los Angeles County hospital, and be over 65 at the time of admission.

The second qualification is that one of the five ICD-9-CM disease codes specified in the discharge data be 48.1, 48.2, 48.5, 48.6, or 48.38, as suggested by Iezzoni et al. (1996) to define pneumonia.

The third qualification is that the source of admission must be either routine or from the emergency room. This eliminates patients transferred into the hospital from another medical facility or admitted from an intermediate care or skilled nursing facility. To the extent that placement in these facilities is correlated with unobserved disease severity, and to the extent that such facilities may be systematically located near higher quality hospitals, the key assumption that

[18] See Wray et al. (1997).

[19] Pneumonia and influenza combined constitute the sixth leading cause of death in the U.S., and the fourth leading cause of death for those over 65 (Pickle et al. (1996)). Pneumonia is also the leading cause of death among patients with nosocomial (hospital-acquired) infections (Pennington (1994)).

[20] See Keeler et al. (1990) and McGarvey and Harper (1993).

distance from the hospital is exogenous in our model would be violated. This step eliminates approximately 23 percent of the patients from the sample.

The fourth qualification is that the patient be admitted to a hospital with at least 80 admissions for pneumonia in our data set. This screen reduces J and thereby computation time. Its potential to introduce sample selection bias is limited by the fact that it eliminates fewer than one percent of the patients.

3.2. *Variable Construction*

The covariate vector x_i in the mortality probit equations contains an indicator for each year, demographic variables, and indicators of disease severity. Most of the demographic variables are constructed from the discharge records. We include age indicators (65–69, 70–74, 75–79, 80–84, and 85 or older), an indicator for gender, and indicators for race (white, black, Hispanic, Native American, and Asian). The discharge records contain no information on socioeconomic status. As a proxy for the patient's household income, we use the mean 1990 census household income for households with the same zip code, race, and age class as the patient.[21]

Indicators of disease severity in x_i are constructed from the admission disease staging information contained in the discharge records. Disease staging has been shown to be as good as some risk adjustment data based on chart review of medical records.[22] Since some of the 13 stages have very few patients, we aggregated stages into five groups: stage 1.1, stages 1.3 through 2.3, stages 3.1 through 3.6, stage 3.7, and stage 3.8. Indicator variables for all but stage 1.1 are included in x_i.

The indicator for mortality, m_i, is set to 1 if the patient died in the hospital within ten days of admission; otherwise $m_i = 0$. The horizon for mortality is limited to ten days, because beyond this point hospitals sometimes transfer terminally ill patients to other facilities, and this decision appears to vary considerably by hospital. To control for differential patient transfer, Gowrisankaran and Town (1999) used a hazard model as an alternative to the 10-day inpatient mortality, but found little difference between the two specifications. In two separate studies of heart disease patients, McClellan, McNeil, and Newhouse (1994) and McClellan and Staiger (1999b), find that there is a very strong correlation between 7-day mortality and 30-day mortality rates across hospitals.[23]

[21] The census provides only two relevant age categories, 65–74 and 75 and older, instead of four. Thus, we aggregated the discharge data age categories to this level. Additionally, the census provides income only within cells. To find the mean income, we took the mean value for each cell as the income for each household in that cell. For the highest cell, $100,000 or more, we assumed a mean income of $140,000. Income is measured in units of $100,000 and income squared in units of billions of dollars squared.

[22] See Thomas and Ashcroft (1991). Iezzoni et al. (1996) showed excellent agreement of disease stage with the ratings of other systems.

[23] As caveats, note that heart disease is very different from pneumonia and that these studies examine mortality, not inpatient mortality.

TABLE I

FREQUENCY AND MORTALITY RATES BY AGE, DISEASE STAGE, RACE, AND SEX CATEGORIES[a]

Severity and Demographic Categories		Age Categories					Row Totals
		65–69	70–74	75–79	80–84	85+	
Disease Stage	1.1	8,407	10,254	11,524	11,168	14,864	56,217
		5.01	5.09	5.83	5.82	10.18	6.94
	1.3–2.3	846	1,021	1,017	912	1,069	4,865
		5.91	5.97	6.88	10.09	10.20	7.85
	3.1–3.6	670	769	1,018	973	1,478	4,908
		12.69	12.87	14.83	16.07	21.99	16.70
	3.7	1,350	1,598	1,707	1,381	1,664	7,700
		15.33	14.77	16.81	22.13	28.18	19.56
	3.8	156	228	218	239	317	1,158
		45.51	42.10	44.03	56.49	53.94	49.14
Race	White	7,100	9,301	10,796	10,542	14,256	51,995
		7.20	7.68	8.75	10.44	13.89	10.10
	Black	1,498	1,405	1,376	1,207	1,433	6,919
		9.74	8.61	7.92	10.60	13.32	10.04
	Hispanic	2,013	2,032	2,098	1,978	2,709	10,830
		6.31	5.41	6.86	7.79	11.04	7.70
	Asian	794	1,106	1,189	930	971	4,990
		6.17	6.06	6.39	8.27	11.33	7.59
	Native American	24	26	25	16	23	114
		4.17	7.69	8.00	37.50	26.09	14.91
Sex	Female	5,726	7,010	8,116	7,955	12,092	40,899
		6.61	6.22	7.34	9.25	13.24	9.14
	Male	5,703	6,860	7,368	6,718	7,300	33,949
		8.12	8.42	9.23	10.87	13.51	10.12
Column Totals		11,429	13,870	15,484	14,673	19,392	74,848
		7.30	7.31	8.24	9.99	13.34	9.59

[a]The first number in each cell is the cell frequency, and the second number is the mortality rate in that cell.

Table I provides a summary of the distribution of demographic characteristics and disease severity in the sample, together with mortality rates. Within each age group the composition of the sample by race and sex closely reflects the demographics of Los Angeles County. Older individuals enter the sample in greater proportion to their numbers in the population than do younger ones. Within each age group three-quarters of the sample is classified in the least severe disease stage. Mortality rates increase gradually with age, increase sharply with disease stage, are a little higher for men than for women, and are lower for Asians and Hispanics than for whites or blacks.

The covariate matrix Z_i contains variables specific to the combination of patient i and each hospital. The additional information in Z_i not contained in x_i is the distance of the patient's home from each hospital. The discharge data include patient zip codes and the AHA data include hospital zip codes. The Census Bureau's TIGER database provides the latitude and longitude of the centroid

TABLE II
HOSPITAL FREQUENCY, PATIENTS TREATED, AND MORTALITY RATE BY HOSPITAL CLASSIFICATION[a]

	150 Beds or Less	151–200 Beds	201–300 Beds	Over 300 Beds	Row Totals
Private,	9	4	18	19	50
Not-for-Profit	4,741	2,369	15,526	21,545	44,181
	9.17	11.11	9.42	9.71	9.62
Private,	32	15	7	1	55
For-profit	9,792	6,627	4,412	973	21,804
	9.24	9.57	10.54	10.48	9.66
Private				5	5
Teaching	0	0	0	6,802	6,802
				9.17	9.17
Public			1	3	4
	0	0	232	1,829	2,061
			8.62	9.57	9.46
Column	41	19	26	28	114
Totals	14,533	8,996	20,170	31,149	74,848
	9.22	9.97	9.65	9.61	9.59

[a]The first number in each cell is the number of hospitals in that category, the second number is the total number of pneumonia patients discharged from hospitals in that cell, and the third number is the mortality rate (patient-weighted) for patients who were discharged from hospitals in that cell.

of each zip code. Given these, standard great circle trigonometric formulas provide the distance between each patient home and hospital.[24] The five variables in Z_i are distance (in hundreds of kilometers); distance-squared; the product of distance and an age indicator (1 for 65–69, 2 for 70–74, 3 for 75–79, 4 for 80–84, 5 for 85+); the product of distance and disease stage $(1.1, \ldots, 3.8)$; and the product of distance and income (in units of $100,000).

The prior distribution and subsequent analyses require the size and ownership status of each hospital. This information was obtained from the AHA survey and is summarized in Table II. We specify private teaching, public (operated by Los Angeles County), other not-for-profit, and for-profit hospitals as four mutually exclusive ownership categories, and 150 or less, 151–200, 201–300, and 301 or more beds as four mutually exclusive size categories.

While mortality rates differ slightly by ownership category, none of the differences are significant at conventional levels. The same is true by size category. Contrasts in mortality rates are stronger between cross-classified cells in Table II. For example, the mean of the cells private, not-for-profit with 151–200 beds (11.11%) and private, for-profit with 201–300 beds (10.54%) are significantly greater than the overall mean at the 5 percent level.

[24] For zip codes that contain more than one hospital, we use address-level latitude and longitude data from the Census Bureau's TIGER database, which stores the geographic location of every block corner and will interpolate from that to find the latitude and longitude of any address.

4. FINDINGS

The model set forth in Section 2 applied to the data described in Section 3 yields evidence on systematic differences in quality across hospitals, provides insight into the interaction between hospital choice and hospital quality, and suggests quality orderings among hospitals. This section summarizes these findings.

4.1. *Patient Mortality and Hospital Choice*

Table III presents the posterior means and standard deviations of some parameters and functions of parameters in the selection and standard probit models. Table III details q_G, ρ_G, $\gamma/(\delta'\Sigma\delta+1)^{1/2}$, and $\tau^2/\Sigma\delta+1)$ for the selection model and q_G^*, γ, and τ^2 for the probit model.[25]

The mortality equation has three groups of covariates: demographics, disease severity, and hospital indicators. In the case of the demographic and disease severity covariates, coefficient posterior means in the selection and probit models are similar to each other and closely reflect the mortality rates presented in Table I. Posterior standard deviations indicate substantial information about differences in mortality probabilities across demographic groups.

In the case of the hospital quality probits, there are greater and more interesting differences between the selection model, the probit model, and the raw data. Both the probit model and the raw data (Table II) do not draw any sharp distinctions in hospital quality by size or ownership class. However, the selection model finds sharp distinctions by size. This suggests that controls for both observed and unobserved severity of illness are important.

The posterior means of the hyperparameters τ_j^2 carry forward the substantial uncertainty—about hospital qualities in the prior distribution, combined with the information in the data. The prior mean of each τ_j^2 is 0.41. In the case of the four ownership components p_k and size components s_l, the data combine with the prior to lower the posterior mean to 0.21. In the case of the 114 individual hospital components u_j the data provide more information about the common variance and lower the posterior mean to 0.037.[26]

Posterior means and standard deviations of the choice covariate coefficient vector α show that, as expected, distance is an important factor in describing the hospital of admission. The posterior mean of -13.65 implies that a hospital that is 20 kilometers farther from a patient's home than another has a normalized probit that is $13.65 \times 0.2/\sqrt{2} \cong 2$ units lower. The quadratic term in the equation is highly significant, but since distances are at most 100 kilometers within Los Angeles County, its substantive effect is not great. Interactions of distance

[25] The normalization of γ and τ^2 facilitates comparison between the two models.

[26] The mean of an inverted gamma distribution for τ^2 of the form $s^2/\tau^2 \sim \chi^2(v)$ is $E(\tau^2) = s^2/(v-2)$. If the prior were conjugate, then the posterior mean of each τ_j^2 would be $(1.25+d^2)/(n+3)$, where d^2 is the sum of squares due to p_k, s_l, or u_j and $n=4$ in the first two cases and $n=114$ in the last. The lower bound on the posterior mean would then be $1.25/(n+3)$, or 0.18 in the first two cases and 0.011 in the last case.

TABLE III

POSTERIOR MEANS AND STANDARD DEVIATIONS[a]

Coefficient	Selection model		Probit model	
	$\gamma/(\delta'\Sigma\delta+1)^{1/2}$		γ	
Demographic covariates				
Age 70–74	−0.009	(0.024)	−0.008	(0.025)
Age 75–79	0.065	(0.023)	0.068	(0.025)
Age 80–84	0.184	(0.023)	0.187	(0.024)
Age 85+	0.369	(0.022)	0.374	(0.023)
Female	−0.087	(0.013)	−0.087	(0.013)
Black	−0.020	(0.028)	−0.025	(0.028)
Hispanic	−0.120	(0.022)	−0.126	(0.023)
Native	0.152	(0.130)	0.168	(0.134)
Asian	−0.091	(0.030)	−0.091	(0.031)
Income	0.223	(0.207)	0.253	(0.201)
Income^2	−0.028	(0.024)	−0.030	(0.024)
	$\gamma/(\delta'\Sigma\delta+1)^{1/2}$		γ	
Disease severity covariates				
Emergency admit	0.180	(0.015)	0.181	(0.016)
Disease stages 1.3–2.3	0.089	(0.028)	0.089	(0.028)
Disease stages 3.1–3.6	0.493	(0.023)	0.496	(0.023)
Disease stage 3.7	0.635	(0.019)	0.640	(0.018)
Disease stage 3.8	1.396	(0.038)	1.412	(0.037)
	q_G		ρ_G		q_G^*	
Hospital group quality probits and severity correlations						
150 beds or less	0.018	(0.021)	0.001	(0.022)	0.007	(0.012)
151 to 200 beds	−0.069	(0.032)	−0.017	(0.024)	−0.032	(0.018)
201 to 300 beds	−0.023	(0.027)	−0.010	(0.032)	−0.003	(0.013)
Over 300 beds	0.039	(0.019)	0.022	(0.023)	0.004	(0.012)
Private, not-for-profit	0.0055	(0.018)	0.003	(0.026)	−0.001	(0.009)
Private, for-profit	0.0074	(0.015)	0.008	(0.024)	−0.008	(0.009)
Private teaching	0.019	(0.041)	0.006	(0.023)	0.021	(0.023)
Public	−0.072	(0.089)	−0.017	(0.038)	−0.017	(0.041)
	$\gamma/(\delta'\Sigma\delta+1)$		τ^2	
Variance of quality				
Size	0.20	(0.14)	0.21	(0.15)
Ownership	0.20	(0.14)	0.21	(0.15)
Individual hospital	0.037	(0.0062)	0.030	(0.0048)
	α			
Hospital choice covariates				
Distance	−13.65	(0.147)	—	
Distance^2	12.43	(0.080)	—	
Distance × Age	−0.45	(0.025)	—	
Distance × Severity	−0.31	(0.034)	—	
$10^{-5}\times$ Distance × Income	−0.974	(0.258)	—	

[a]Specifications also include indicators for each year.

with age and severity both have negative coefficients with posterior standard deviations small relative to their posterior means. Given that age class varies between 1 and 5 and observed severity varies between 1.1 and 3.8, the posterior mean of the distance coefficient varies between -14.44 and -17.08, with distance decreasing in age and observed severity of illness. The reason for this is likely due to the increased cost and difficulty of transport for severely ill patients. Patients in zip codes with higher average income are more likely to be admitted to nearby hospitals.

Table IV provides explicit posterior probabilities for hospital group quality comparisons using the selection model and also lists the mean and standard deviation of the posterior probability of mortality at each type of hospital given a 10 percent mortality (roughly the sample mean) at other types. There are sharp differences based on hospital size (panel A). The posterior probability that the group hospital quality probit for the largest-hospital group exceeds that of the smallest-hospital group is 0.71, and the posterior probability that it is greater than that of the other two size groups exceeds 0.95. The posterior probability that the

TABLE IV

POSTERIOR PROBABILITY COMPARISONS OF GROUP HOSPITAL QUALITY PROBITS, SELECTION MODEL[a]

A. Hospitals Grouped by Size				
	≤150 beds	151–200 beds	201–300 beds	>300 beds
≤ 150 beds	—	1%	16%	71%
	0.10 (—)	0.086 (0.007)	0.089 (0.007)	0.104 (0.006)
151–200 beds	99%	—	82%	100%
	0.117 (0.009)	0.10 (—)	0.109 (0.009)	0.121 (0.007)
201–300 beds	84%	18%	—	98%
	0.108 (0.007)	0.093 (0.008)	0.10 (—)	0.112 (0.006)
>300 beds	29%	0%	2%	—
	0.097 (0.006)	0.083 (0.006)	0.090 (0.005)	0.10 (—)
B. Hospitals Grouped by Ownership Classification				
	Private not-for-profit	Private for-profit	Private teaching	Public
Private not-for-profit	—	54%	60%	23%
	0.10 (—)	0.101 (0.005)	0.103 (0.008)	0.088 (0.015)
Private for-profit	46%	—	56%	20%
	0.100 (0.005)	0.10 (—)	0.103 (0.009)	0.088 (0.014)
Private teaching	40%	44%	—	22%
	0.098 (0.008)	0.099 (0.009)	0.10 (—)	0.087 (0.017)
Public	77%	80%	78%	—
	0.116 (0.019)	0.116 (0.018)	0.118 (0.022)	0.10 (—)

[a]The first number in each cell is the posterior probability that the group quality probit q_G in the column category exceeds q_G in the row category. The second number is the posterior mean probability of mortality in the row category given a 10 percent probability of mortality in the column category, with the posterior standard deviation of this statistic in parentheses.

smallest-hospital group quality probit exceeds that of the second-smallest group similarly exceeds 0.95. This is reflected in a mortality rate of 11.7 percent for the 151–200 bed category given a mortality rate of 10 percent for the smallest size of hospital.

These findings are in rough agreement with the literature. A study by Keeler et al. (1992), which examined the relationship between hospital quality and size using a very detailed and expensive data set that included pneumonia patients along with patients with other, more complex diagnoses, found that hospital quality increases with bed size. However, in their study they did not allow for a nonlinear relationship between hospital size and mortality rates; thus they could not uncover the U-shaped relationship between hospital quality and size that we do. Successful pneumonia treatments are linked to identifying the pathogen responsible for the infection and administering the appropriate antibacterial agent early in the progression of the disease, and subsequently monitoring and adjusting the dosage of the drug (Rello and Valles (1998), Pennington (1994), McGarvey and Harper (1993)). There is evidence that smaller hospitals may be better at the timely administration of antibiotics (Fine et al. (1998)), which may explain why we observe that they have better outcomes. Furthermore, since small hospitals are likely to treat a disproportionate number of pneumonia patients relative to more technically challenging illness,[27] they may also develop expertise in this disease. That, in turn, may overcome advantages that medium-sized hospitals may have in other dimensions, such as laboratory facilities.

There are less-sharp differences in the selection model based on ownership (panel B). Overall, private teaching hospitals have the highest quality, public hospitals have the lowest quality, and other hospitals are in the middle. However, from the posterior standard deviations of the mortality rates it is evident that there are no definitive comparisons among ownership categories.

There is debate in health policy circles regarding the role that for-profit hospitals should play in the U.S. health system (Gray (1991), Sloan (2000)). Some have argued that private, not-for-profit hospitals may better serve the public interest because they are more likely to provide better care. Our results indicate that for the treatment of pneumonia in older patients and the hospitals in our sample, there is no evidence of this. Keeler et al. (1992) also found public hospitals in large cities to be of lower quality, while the difference in quality between for-profit and not-for-profit hospitals is less pronounced. McClellan and Staiger (1999a) conclude that the quality difference in for-profit and not-for-profit hospitals is small, and if anything for-profits likely provide better care in the treatment of heart attacks. Private teaching hospitals, which are generally viewed as providing superior care (Keeler et al. (1992)), do appear to offer higher quality according to the selection model.

[27] Performing a simple multinomial logit regression of Southern California patients, we found that pneumonia patients were more likely to be admitted to smaller hospitals than were hospital patients generally. In contrast, acute myocardial infarction (heart attack) patients were more likely to be admitted to larger hospitals than the average hospital patient. Unlike pneumonia treatments, acute myocardial infarction treatments often include high-technology surgery such as cardiac catheterization, angioplasty, or bypass.

TABLE V

RELATIONS BETWEEN HOSPITAL QUALITY PROBITS AND SEVERITY CORRELATIONS IN THE SAMPLE[a]

A. Variances and correlations of posterior means of q_j, q_j^*, and ρ_j			
q_j	.0148	.766	.324
q_j^*	.0105	.0128	−.325
ρ_j	.0018	−.0017	.0022
B. OLS regression of ρ_j (posterior means) on q_j (posterior means)			
$\rho_j = .124 q_j$; $R^2 = .105$, $s = .044$. (.034)			
C. OLS regression of q_j^* (posterior means) on q_j and ρ_j (posterior means)			
$q_j^* = .905 q_j - 1.553 \rho_j$; $R^2 = .954$, $s = .022$. (.020) (.052)			
D. OLS regression of q_j^* (posterior means) on q_j (posterior means)			
$q_j^* = .712 q_j$; $R^2 = .587$, $s = .073$. (.056)			

[a]Panel A shows variances on the diagonal, covariances below the main diagonal, and correlations above the main diagonal. Panels B–D show standard errors in parentheses.

4.2. *Selection and Selection Bias*

We present some statistics on the relationship between the posterior means of q_j, q_j^*, and ρ_j across the 114 hospitals in Table V. These statistics allow us to uncover the importance of selection and the relationship between selection and quality.

We start by analyzing the quantitative importance of selection in influencing patient mortality. In the simple probit model, the variance in unobserved disease severity ε_i is normalized to be 1. From the posterior means of the coefficients on observed disease severity in the model (Table III) and the distribution of observed patient characteristics in the population (Table I), one may approximate the variance in the contribution of observed demographics and disease severity to the mortality probit: it is about 0.45. The variance in the mortality probit due to variation in hospital quality is about 0.013 (Table V, panel A), much smaller than the variance due to unobserved severity of illness, which is normalized to 1. This decomposition of variance is about the same in the selection model—variation in hospital quality is slightly higher (Table V, Panel A)—but it is still quite small relative to disease severity.

In the selection model the variation in unobserved disease severity is decomposed into a component that is independent of the hospital assignment process (ζ_i from (5)) with variance 1, and a component that is a function of the hospital assignment probits, $\eta_i'\delta$ (also from (5)). The variance of the latter term, $\delta'\Sigma\delta$, has a posterior mean of 8.7, which is much larger than the independent component. This constitutes strong evidence against random assignment of patients, and suggests that the simple probit model provides misleading inferences about hospital quality.

Since patient selection is important, we are interested in understanding the relationship between selection and quality. Table V, panel A reveals a positive relationship between the posterior means of q_j and ρ_j: the correlation between posterior means is 0.517 (panel A) and a simple least squares regression of the posterior means of the ρ_j on the posterior means of the q_j shows a slope coefficient of 0.183 that is significantly positive (t of over 6).[28] Thus, hospitals with higher quality (higher q_j) have a greater propensity to be selected by patients with greater unobserved disease severity (higher ε_j). This is also reflected in Table III, which shows similar patterns of q_G and ρ_G across types of hospitals.

In any selection model, conditional on observed characteristics (including observed severity), the observed mortality rate for each hospital will be decomposed into a hospital quality component and an unobserved severity component. Panel C of Table V shows that in this relationship hospital quality q_j^* in the probit model is well described as a linear function of hospital quality q_j, and severity correlation ρ_j, in the selection model. From the regression relation reported in panel C of Table V, it is clear that variation in hospital severity correlation substantially drives variation in inferred hospital quality q_j^* in the probit model. From the regressions in panels B and C, one can infer the slope coefficient of .712 ($=.905 - 1.553 \times .124$) in panel D. Thus, variation in hospital severity correlation accounts for a substantial portion of the variation in hospital mortality rates in the selection model, whereas in the simple probit model this variation must be attributed to quality differences.

4.3. *Ordering by Quality*

The model and approach to inference described in Section 2 provide the complete posterior distribution of all the parameters in the model, and any functions of these parameters. In particular, corresponding to the parameter values in any iteration of the Gibbs sampling algorithm, it is a simple matter to compute the corresponding hospital quality probits q_j. The 1,900 draws used to obtain the posterior moments reported in this section therefore also provide 1,900 draws from the joint distribution of the hospital quality probits q_j. Pairwise comparisons between hospitals are then straightforward. For example, for two hospitals j and k, the numerical approximation to the posterior probability that $q_j > q_k$ is the fraction of iterations in which $q_j > q_k$, and the joint distribution of q_j and q_k can easily be plotted.

Comparing all 114 hospitals simultaneously is more challenging. A formal approach to ordering hospitals by quality would begin with a loss function for orderings. Suppose the 114-element vector of quality ranks is r, and the estimated

[28] Since results in Table V are based on posterior means, they do not take into account dispersion in the posterior. To account for this dispersion, one can examine the sample relation between q_j, q_j^*, and ρ_j as a function of the parameters, and consider the posterior uncertainty associated with this relationship. This would yield values of Table V for each draw from the posterior simulator. One can then compute the mean value across the draws. This method yields similar results.

quality rank vector is $\hat{r}$. If the loss function is $(\hat{r}-r)'A(r-\hat{r})$, where A is a positive definite matrix, then $\hat{r}$ should be the posterior mean of r.[29] This estimate may, in turn, be approximated numerically by sorting hospital qualities q_j in each iteration of the Gibbs sampler, finding the corresponding rank for each hospital, and then averaging the ranks across all iterations. The resulting estimated ranks $\hat{r}_j$ are generally not integers. If the loss function were $\Sigma_{j=1}^{114} a_j|\hat{r}_j - r_j|$, where all $a_j > 0$, then $\hat{r}_j$ should be the median of the posterior distribution of r_j, which in turn is an integer (with probability one).

Appendix A4 in Geweke, Gowrisankaran, and Town (2001) provides rankings based on both loss functions. The choice of loss function turns out not to have a large effect on the orderings of relative quality. The rankings produced by these alternative loss functions are similar. The posterior distributions of r and of the hospital qualities convey the uncertainty associated with the rankings. For most pairwise combinations of hospitals in the top and bottom quartiles, the posterior probability that the quality of the former exceeds the latter is rarely less than 0.8 and exceeds 0.9 more often than not. An approximate rule of thumb for the accuracy of rankings is that if a hospital is ranked at quantile x, then the posterior probability that its true rank is above the median is also x. Appendix A4 provides all the rankings and several aspects of their joint posterior distribution.

4.4. *Specification and Robustness*

A key assumption in the selection model is that the distances between the patients' homes and the 114 hospitals in the sample constitute variables that may be used to control for the nonrandom assignment of patients to hospitals. Because of the nonlinear relationship between the endogenous variables (hospital choice) in the mortality equation and the instruments, this relationship was modeled explicitly. Table III reveals an indisputably strong link between the measures in Z and the choice of hospital. For instance, distance and its square explain about 30 percent of the variance of the probits. The findings are in accord with the literature.[30]

The further assumption that distances from hospitals to patients are uncorrelated with unobserved disease severity cannot be examined so directly. One plausible alternative is that there remain geographic variations in unobserved disease severity after accounting for the observed covariates listed in the first two panels of Table III. We examined this possibility from three angles. First, in a conventional probit model for mortality using the observed covariates, hospital choice dummies, and patient zip code dummies, the zip code dummies are insignificant. Second, the same is true if dummies for nearest hospital replace zip code dummies. In both equations, the coefficients on the hospital choice dummies are jointly significant in the presence of the zip code dummies. Finally,

[29] See, for example, Bernardo and Smith (1994, Section 5.1.5) for this standard result, as well as the one on medians used in the next paragraph.

[30] See Luft et al. (1990) and Gowrisankaran and Town (1999).

we conducted a more direct examination by retrieving the unobserved disease severity component from the mortality probit equation in each iteration of the MCMC algorithm. In the regression of this component on zip code dummies and the other regressors, the dummies were jointly insignificant in every iteration. All these findings are consistent with the absence of any unobserved geographic component of disease severity.

Given the large number of endogenous variables in the selection model, quite a few assumptions about functional form were required. The dimensionality of the problem is perhaps most evident in the 6,440 potentially independent free parameters in Σ, the prior variance matrix in the multinomial hospital assignment model. The selection model takes the extreme step of assuming that shocks to the probits in this model are iid normal before differencing (Section 2.2). If this assumption is reasonable, then the 113×1 vectors of posterior shocks η_i $(i = 1, \ldots, n)$, which may be retrieved in each iteration of the MCMC algorithm, should be consistent with the specification $\Sigma = I_{j-1} + e_{j-1}e'_{j-1}$. If it is not—for example, if patients with certain characteristics all choose from one small group of hospitals—then this will be evidenced by a constructed covariance matrix $S = (n-1)^{-1}\Sigma_{i=1}^{n}(\eta_i - \bar{\eta})(\eta_i - \bar{\eta})'$ being substantially different from Σ. A conventional goodness of fit test, carried out at the 5 percent level, rejects the null hypothesis in slightly over half the iterations of the MCMC algorithm. We conclude that there may well be misspecification of the covariance structure in the multinomial hospital assignment covariance matrix, but it is probably not severe. Due to the large number of parameters in Σ, information about the covariance structure beyond the data would be required to deal constructively with this potential misspecification.

The sensitivity of findings to the specification of the prior distribution can be examined in a number of ways. To convey the nature of the sensitivity we set up three further variants of the selection model. Variant A effectively eliminates the instruments from the entire model by scaling the prior standard deviations of the coefficient vector α in the multinomial hospital assignment model by the factor 10^{-6}. This variant leaves only the functional form to identify the hospital-specific parameters in the mortality equation. Variant B scales the prior standard deviations of α in the original selection model downward by a factor of 5 and τ^2 downward by a factor of 25. Variant C is like variant B except that prior standard deviations are increased by a factor of 5 relative to the base model. Thus, variants B and C provide alternative priors that are plausible from the perspective of the subjective prior in the base selection model.

Appendix A5 of the working paper provides a detailed set of results for each of these prior distributions. As one might expect, coefficients on covariates in the mortality probit equation show very little sensitivity to the choice from among the four prior distributions. The same is true in the hospital choice multinomial probit model, with the obvious exception of prior variant A. The findings about hospital mortality (Section 4.1) are the same in variants B and C as in the base selection model: quality is a U-shaped function of size; private teaching hospitals have the highest and public hospitals the lowest quality with differences in this

dimension remaining small. By contrast variant A shows little effect of size, or ownership, and the point estimates display neither the U shape for size nor the ownership ranking of the base model. The correlations between hospital quality posterior means in the base selection model and variants B and C are both 0.80. By contrast, the correlation between hospital quality posterior means in the base selection model and variant A is only 0.34. We conclude that reasonable variants on the prior produce distinct but insubstantial differences, whereas elimination of the instruments from the model has strong and substantial effects.

5. CONCLUSION

This study has extended existing econometric methods in order to measure hospital quality using the experience of patients admitted to hospitals in nonrandom fashion. Using discharge records for almost 75,000 older pneumonia patients from 114 hospitals in Los Angeles County, we find evidence of differences in quality between hospitals of different size and ownership classifications. The smallest and largest hospitals exhibit higher quality than other hospitals. We also detect substantial differences in quality for a sizable minority of individual hospitals.

As an important by-product, our methods produce information about the hospital admissions process. Patients with greater unobserved severity of illness tend, overall, to be admitted to hospitals of higher quality. Consequently more conventional methods that ignore nonrandom admission, when applied to this data set, tend to lower the inferred quality of good hospitals and raise that of poor ones, relative to our findings. We find that variation across individual hospitals in the unobserved severity of illness is at least as great as variation in quality, and that this variation accounts for most of the large discrepancy between inference about hospital quality in our model and in more conventional methods.

The procedures used here are at the current frontier of intensive computational methods in econometrics. A supercomputer and several days of computing were required to obtain the results reported here. Recent and imminent innovations in numerical methods and computing technology should sharply reduce the real costs of these procedures in the near term. Given the policy importance of assessing quality of care in hospitals, we believe there is a significant return to further investment in these methods and their application to similar questions in health policy and related fields.

Departments of Economics and Statistics, University of Iowa, W382 Pappajohn Business Building, Iowa City, IA 52242-1000, U.S.A.; john-geweke@uiowa.edu;

Department of Economics, Harvard University, Littauer Center, Cambridge, MA 02138-3001, U.S.A., and National Bureau of Economic Research; gautam_gowrisankaran@nber.org;

and

Health Services Research and Policy, School of Public Health, University of Minnesota, Mayo Mail Code 729, 420 Delaware St. S.E., Minneapolis, MN 55455-0392, U.S.A.; rjtown@umn.edu

Manuscript received February, 2000; final revision received July, 2002.

REFERENCES

ALBERT, J., AND S. CHIB (1993): "Bayesian Analysis of Binary and Polychotomous Response Data," *Journal of the American Statistical Association*, 88, 669–679.

BERNARDO, J., AND A. F. M. SMITH (1994): *Bayesian Theory*. Chichester U.K.: John Wiley and Sons.

BURNS, L., AND D. WHOLEY (1992): "The Impact of Physician Characteristics in Conditional Choice Models for Hospital Care," *Journal of Health Economics*, 11, 43–62.

CHIB, S., AND E. GREENBERG (1996): "Markov Chain Monte Carlo Simulation Methods in Econometrics," *Econometric Theory*, 12, 409–431.

CUTLER, D. M. (1995): "The Incidence of Adverse Medical Outcomes Under Prospective Payment," *Econometrica*, 63, 29–50.

FINE, J. M., J. D. SCINTO, D. H. GALUSHA, M. K. PETRILLO, AND T. P. MEEHAN (1998): "Patient and Hospital Characteristics Associated with Timely Care of Elderly Patients Hospitalized with Pneumonia: Results from the Medicare Quality Indication System Pneumonia Module," *Abstract Book*, Association of Health Services Research, 15.

GELFAND, A. E., AND A. F. M. SMITH (1990): "Sampling Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, 85, 398–409.

GEWEKE, J. (1997): "Posterior Simulators in Econometrics," in *Advances in Economics and Econometrics: Theory and Applications*, Vol. III, ed. by D. Kreps and K. F. Wallis. Cambridge U.K.: Cambridge University Press, 128–165.

——— (1999): "Using Simulation Methods for Bayesian Econometric Models: Inference, Development and Communication" (with discussion and rejoinder), *Econometric Reviews*, 18, 1–126.

GEWEKE, J., G. GOWRISANKARAN, AND R. J. TOWN (2001): "Bayesian Inference for Hospital Quality in a Selection Model," Working Paper 8497, National Bureau of Economic Research, Cambridge MA.

GEWEKE, J., M. KEANE, AND D. RUNKLE (1997): "Statistical Inference in the Multinomial, Multiperiod Probit Model," *Journal of Econometrics*, 80, 125–165.

GOWRISANKARAN, G., AND R. J. TOWN (1999): "Estimating the Quality of Care in Hospitals Using Instrumental Variables," *Journal of Health Economics*, 18, 747–767.

GRAY, B. (1991): *The Profit Motive and Patient Care*. Cambridge MA: Harvard University Press.

IEZZONI, L. I., M. SHWARTZ, A. S. ASH, J. S. HUGHES, J. DALEY, AND Y. D. MACKIERNAN (1996): "Severity Measurement Methods and Judging Hospital Death Rates for Pneumonia," *Medical Care*, 34, 11–28.

KEANE, M. P. (1992): "A Note on Identification in the Multinomial Probit Model," *Journal of Business and Economic Statistics*, 10, 193–200.

KEELER, E. B., K. L. KAHN, D. DRAPER, M. J. SHERWOOD, L. V. RUBENSTEIN, E. J. REINISCH, J. KOSECOFF, AND R. H. BROOK (1990): "Changes in Sickness at Admission Following the Introduction of the Prospective Payment System," *Journal of the American Medical Association*, 264, 1962–1968.

KEELER, E. B., L. V. RUBENSTEIN, K. L. KAHN, D. DRAPER, E. R. HARRISON, M. J. MCGINTY, W. H. ROGERS, AND R. H. BROOK (1992): "Hospital Characteristics and Quality of Care," *Journal of the American Medical Association*, 268, 1709–1714.

KESSLER, D., AND M. MCCLELLAN (2000): "Is Hospital Competition Socially Wasteful?" *Quarterly Journal of Economics*, 115, 577–615.

LOHR, K. N., ED. (1990): *Medicare: A Strategy for Quality Assurance*, Volume I. Washington DC: National Academy Press.

LUFT, H. S., D. W. GARNICK, D. H. MARK, D. J. PELTZMAN, C. S. PHIBBS, E. LICHTENBERG, AND S. J. MCPHEE (1990): "Does Quality Influence Choice of Hospital?" *Journal of the American Medical Association*, 263, 2899–2906.

MCCLELLAN, M., B. MCNEIL, AND J. NEWHOUSE (1994): "Does More Intensive Treatment of Acute Myocardial Infarction in the Elderly Reduce Mortality?" *Journal of the American Medical Association*, 272, 859–866.

MCCLELLAN, M., AND H. NOGUCHI (1998): "Technological Change in Heart-Disease Treatment: Does High Tech Mean Low Value," *American Economic Review Papers and Proceedings*, 88, 90–96.

MCCLELLAN, M., AND D. STAIGER (1999a): "Comparing Hospital Quality at For-Profit and Not-For-Profit Hospitals," Working Paper 7324, National Bureau of Economic Research, Cambridge MA.

——— (1999b): "The Quality of Health Care Providers," Working Paper 7327, National Bureau of Economic Research, Cambridge MA.

MCGARVEY, R., AND J. HARPER (1993): "Pneumonia Mortality Reduction and Quality Improvement in a Community Hospital," *Quality Review Bulletin*, 19, 124–130.

PENNINGTON, J. (1994): *Respiratory Infections: Diagnosis and Management*, 3rd Edition. New York: Raven Press.

PICKLE, L., M. MUNGIOLE, G. JONES, AND A. WHITE (1996): *Atlas of United States Mortality*, PHS 97-1015. Hyattsville MD: National Center for Health Statistics.

RELLO, J., AND J. VALLES (1998): "Mortality as an Outcome in Hospital-Acquired Pneumonia," *Infection Control and Hospital Epidemiology*, 19, 795–797.

SLOAN, F. (2000): "Not-for-profit Ownership and Hospital Behavior," in *The Handbook of Health Economics*, Vol. 1, ed. by A. J. Culyer and J. P. Newhouse. Amsterdam: Elsevier Science.

THOMAS, J. W., AND M. L. F. ASHCROFT (1991): "Measuring Severity of Illness: Six Severity Systems and Their Ability to Explain Cost Variations," *Inquiry*, 28, 39–55.

UNITED STATES GENERAL ACCOUNTING OFFICE (1994): *"Report Cards" Are Useful But Significant Issues Need to be Addressed* (GAO/HEHS-94-219). Washington, DC: United States General Accounting Office.

WRAY, N., J. HOLLINGSWORTH, N. PETTERSEN, AND C. ASHTON (1997): "Case-Mix Adjustment Using Administrative Databases: A Paradigm to Guide Future Research," *Medical Care Research and Review*, 54, 326–356.