Bayesian efficiency analysis through individual effects: Hospital cost frontiers

Gary Koop a, Jacek Osiewalski b, Mark F.J. Steel* : c

a Department of Economics, University of Toronto, Toronto, Ont. M5S 1A1, Canada
b Department of Econometrics, Academy of Economics, 31-510 Kraków, Poland
c CentER and Department of Econometrics, Tilburg University, 5000 LE Tilburg, The Netherlands

(Received May 1994; final version received June 1995)

Abstract

This paper develops Bayesian tools for making inferences about firm-specific inefficiencies in panel data models. We begin by establishing a Bayesian setting in which fixed and random effects models are defined. What distinguishes these classes of models is the marginal prior independence of the effects. We show how such models can be analyzed using Monte Carlo integration or Gibbs sampling. These techniques are applied to a panel of U.S. hospitals. Our empirical findings illustrate the different characteristics of both types of models, as well as the influence of the particular priors used on the firm effects.

Key words: Stochastic frontier models; Panel data; Fixed effects; Random effects; Gibbs sampler; Hospital cost function

JEL classification: C11; C23; 110

1. Introduction

This paper develops Bayesian tools for making inferences about firm-specific inefficiencies in panel data models. We use the stochastic cost frontier methodology for panel data described in, for instance, Schmidt and Sickles (1984). We

* Corresponding author.

We would like to thank the VA Management Science Group, Bedford, for providing the data used in this study. The first author acknowledges the hospitality of the Center for Economic Research, Tilburg University. The second author acknowledges the hospitality of the Department of Econometrics, Tilburg University, during early work on this paper, and was visiting CORE, Université Catholique de Louvain, partly supported by the Belgian Program on Interuniversity Poles of Attraction during the preparation of the present version. An associate editor and two anonymous referees made many helpful suggestions.
introduce firm-specific or individual effects, which are assumed to be constant over time. No time effects will be considered in this paper. This methodology is typically implemented using one of two approaches which are often called the fixed and random effects models. The differences between these two types of models can be viewed, from a Bayesian perspective, as a difference in the structure of the prior information.

We develop a Bayesian framework in which we define fixed effects and random effects models. Our Bayesian fixed effects models are characterized by marginal prior independence between the individual effects, which are thus not linked across firms, but are only assumed constant over time. We distinguish the standard individual effects (SIE) model, where an improper prior on individual intercepts is used. We call our second fixed effects model the marginally independent efficiency distribution (MIED) model. In the latter model we use a proper prior on the firm-specific effects, which are still independent. In this context, the stochastic frontier interpretation implies that the inefficiency error should be one-sided, and the Bayesian approach allows us to easily incorporate this prior information.

In the so-called Bayesian random effects models we assume prior links between the individual effects: their means can be functionally related to certain firm characteristics, which defines the varying efficiency distribution (VED) model, or they can all be drawn from a common distribution, leading to the common efficiency distribution (CED) model.

With respect to the existing Bayesian stochastic frontier literature, i.e., van den Broeck, Koop, Osiewalski, and Steel (1994) and Koop, Osiewalski, and Steel (1994a), we have incorporated the following methodological advances: i) treatment of panel data as opposed to a cross-section analysis, ii) explicitly allowing efficiencies to depend on firm characteristics, iii) providing Bayesian counterparts to the classical fixed and random effects stochastic frontier models.

Our work relates closely to previous Bayesian work on random coefficient models, particularly McCulloch and Rossi (1994). These authors develop computational tools for working with panel data models with individual effects. Recent extensions and empirical applications can be found in Allenby, McCulloch, and Rossi (1995). Our work differs from previous Bayesian work in that the one-sided nature of the inefficiency distribution implies that the individual effect takes a different form from traditional specifications.

This paper uses data from a large panel of U.S. hospitals from 1987–1991 to investigate hospital efficiencies. Our application indicates that the techniques we propose are computationally feasible. We explicitly state the consequences of certain prior assumptions, and illustrate the differences between Bayesian fixed

---

1 The interested reader can find early work on the closely related topic of Bayesian hierarchical models in Lindley and Smith (1972). The use of the Gibbs sampler with such models was pioneered by Gelfand and Smith (1990).
and random effects models. The estimated frontier is largely consistent with economic regularity conditions, and prior sensitivity of inference on efficiencies is examined. In addition to measuring hospital-specific efficiencies, we also seek to explain why some hospitals are more efficient than others by incorporating exogenous variables which may be expected to affect efficiency (i.e., nonprofit, for-profit, or government-run dummies and a measure of workers per patient). Our random effects models clearly indicate that for-profit hospitals are less efficient than nonprofit or government-run hospitals. The VED model indicates that hospitals with more clinical workers per patient are less efficient, which perhaps indicates that there are important quality dimensions that our output measures do not capture.

2. Efficiency analysis with panel data

In order to measure hospital efficiencies, we adopt the stochastic frontier framework first developed by Meeusen and van den Broeck (1977) and Aigner, Lovell, and Schmidt (1977). This methodology postulates a cost frontier, reflecting technology common to all firms, which represents the minimum attainable cost of producing a given level of output(s). Deviations from this frontier reflect either measurement error or inefficiency. Measurement error is assumed to be symmetric and Normally distributed, while inefficiency has some one-sided distribution. For instance, if \( y_i \) is the log of costs of firm \( i \) and \( x_i \) is a vector of \( k \) appropriate explanatory variables, then a typical stochastic frontier model may be specified as

\[
y_i = x_i'\beta + v_i + u_i, \tag{1}
\]

where the \( k \)-dimensional vector \( \beta \) describes the frontier, \( u_i \) and \( v_i \) are independent of each other, \( v_i \) is i.i.d. Normal and \( u_i \) is i.i.d. with some one-sided distribution, \( i = 1, \ldots, n \). Given these assumptions, the likelihood function can be derived and inferences made about the firm-specific inefficiencies. We analyze overall productive efficiency in the sense of Farrell (1957) (see Kopp and Diewert, 1982). We interpret (1) as defining a conditional model for \( y_i \) given \( x_i \), and thus we assume that sufficient conditions for the weak exogeneity of \( x_i \) are fulfilled (see Engle, Hendry, and Richard, 1983). On account of the independence across firms, we can then validly use (1) for predictions corresponding to, as yet, unobserved firms, as explained in Osiewalski and Steel (1996).

In previous work (van den Broeck, Koop, Osiewalski, and Steel, 1994), we have argued for the adoption of a Bayesian perspective for making inferences from such models, since such an approach yields exact finite sample results, allows us to mix over models, to conduct inference on the actual efficiencies, and surmounts some difficult statistical issues which arise in classical analyses. In the present application, we have panel data and new issues arise which necessitate
an extension of our previous work. The classical econometric analysis of firm efficiency with panel data is described in Schmidt and Sickles (1984). If we extend (1) to allow for a time component \( t = 1, \ldots, T \) and assume that efficiency is constant over time for a given firm, we obtain

\[
y_i = (x_0 + u_i)T + X_i\beta + v_i,
\]

where \( y_i \) is now a \( T \times 1 \) vector containing observations for firm \( i \), \( X_i \) is \( T \times k \), \( 1_T \) is a \( T \times 1 \) vector of ones, and \( v_i \) is i.i.d. \( N(0, \sigma^2I_T) \), \( i = 1, \ldots, n \). There are two ways of proceeding with this model, which correspond to the distinction between the fixed effects and the random effects model used in panel data analysis.

In the classical fixed effects model, we define \( x_i = x_0 + u_i \), where the \( u_i \)'s are the individual effects. The inefficiency is thus associated with the firm-specific intercept. Schmidt and Sickles define \( h_i = \hat{x}_i - \hat{x} \), where \( \hat{x}_i \) are OLS estimates of the intercepts and \( \hat{x} = \min_j(\hat{x}_j) \). The \( h_i \)'s (or \( \exp(-h_i) \)'s) are used as measures of inefficiency (efficiency). Note that this approach assumes that the firm with the smallest \( \hat{x}_i \) is fully efficient and measures inefficiencies as deviations from this firm, i.e., it leads to the analysis of relative efficiencies. Furthermore, the use of the min operator above makes the classical distribution theory for the \( h_i \)'s difficult and, hence, it is hard to calculate standard errors for the efficiencies. The Bayesian approach provides the tools to surmount such problems, as will be stressed below.

In the classical random effects model, it is assumed that \( u_i \) has some one-sided distribution and maximum likelihood estimation can be carried out as described in Pitt and Lee (1981) and Schmidt and Sickles (1984). The disadvantage of this approach over the fixed effects approach is that a distributional assumption must be made for the inefficiencies. The advantage is that the problems with the distributional theory of the \( h_i \)'s described above do not occur. Furthermore, Simar (1992) argues that, in practice, the fixed effects model may produce poor estimates of the parameters and the efficiencies in the (usual) case where \( T \ll n \) and regressors do not vary much over time.

McCulloch and Rossi (1994) rightly remark that 'in the Bayesian point of view, there is no distinction between fixed and random effects models, only between hierarchical and nonhierarchical models'. However, we feel it is useful to construct a Bayesian framework in which the widely used panel data terminology is preserved, and fixed and random effects models are clearly distinguished. The distinguishing characteristic of both groups of models will, of course, not be the deterministic or random nature of the effects, but rather the marginal prior links between the effects. Typically, these \( n \) effects will be assumed to be \textit{a priori} independent, either marginally, which gives us the Bayesian fixed effects structure, or conditionally upon a small number \( m \ll n \) of additional parameters, which link the effects and introduce marginal dependence of these effects across time or individuals. In the latter case, we talk of Bayesian random effects models, which, by construction, require nontrivial hierarchical prior structures.
2.1. Bayesian fixed effects models

**Standard individual effects (SIE) model**

We start from the basic individual effects model in (2). As in the classical analysis, we define $\alpha_i = \alpha_0 + u_i$ and let $\alpha = (\alpha_1, \ldots, \alpha_n)'$. Under the standard noninformative prior, $p(\alpha, \beta, \sigma^{-2}) \propto \sigma^2$, the full Bayesian model, therefore, is given by:

$$p(y, \alpha, \beta, \sigma^{-2} | X) = c\sigma^2 \prod_{i=1}^{n} f_N^T(y_i | X_i \beta + \alpha_i \theta, \sigma^2 \theta),$$

(3)

where $c > 0$, and $f_N^T(\cdot | a, B)$ denotes the probability density function of a $T$-variate Normal distribution with mean vector $a$ and covariance matrix $B$. We immediately infer from (3) that the marginal posterior distribution of $(\alpha, \beta)$ is the $(n + k)$-variate Student-$t$ distribution with $n(T - 1) - k$ degrees of freedom. In our application $n = 382$ and $T = 5$, so degrees of freedom are around 1,500 (the exact value depends on $k$). In the light of this, the Student-$t$ posterior will be virtually identical to a Normal distribution and, for the rest of the SIE case, we present results in terms of this Normal approximation. Details of the posterior distribution are given in the Appendix.

In the present paper, interest centres on firm-specific and predictive efficiency. In the classical analysis of Schmidt and Sickles (1984), the authors assumed one firm was fully efficient and measured inefficiencies relative to this firm. In our Bayesian analysis of this model with an improper uniform prior on $\alpha$, we, too, need to measure inefficiency relative to the most efficient firm, but we do not assign this status to one particular firm. In fact, our approach allows us to formally treat the uncertainty implicit in deciding which is the most efficient firm. That is, due to parameter uncertainty, it is not necessarily the case that the firm with the smallest $\hat{\alpha}_i$ is the most efficient, where $\hat{\alpha}_i$ is the posterior mean of $\alpha_i$ (see Appendix). Formally, we define relative firm-specific efficiency as $r_i^{rel} = \exp(-u_i^{rel})$, where $u_i^{rel} = u_i - \min_j(u_j) = \alpha_i - \min_j(\alpha_j)$. The nonnegativity restriction, following directly from the interpretation of the inefficiency term, is not imposed on $u_i$ here, but rather on $u_i^{rel}$. As the definition of $u_i^{rel}$ depends on the number of firms in the sample under consideration, this makes the implicit prior on $r_i^{rel}$ a function of $n$. The implied prior distributions for the efficiencies, $r_i^{rel}$, are characterized by a point mass of $1/n$ at full efficiency ($r_i^{rel} = 1$) and $p(r_i^{rel}) \propto 1/r_i^{rel}$ for $r_i^{rel} \in (0, 1)$. The latter is an L-shaped improper density, which for an arbitrarily small $a \in (0, 1)$ puts an infinite mass in $(0, a)$, but only a finite mass in $[a, 1)$. Thus the implied prior strongly favors low efficiency.

If we knew which firm was most efficient, it would be straightforward to calculate inefficiencies relative to this firm. However, we do not know this, and must formally incorporate this uncertainty into the analysis. The marginal posterior distributions of $u_i^{rel}$ and of the relative efficiency of the $i$th firm, $r_i^{rel}$, have a
point mass at full efficiency given by

\[ P(u_i^{rel} = 0 \mid y, X) = P(\alpha_i = \min_j \alpha_j \mid y, X) \equiv P_i, \]  

and we have the following density function for \( u_i^{rel} > 0 \):

\[ p(u_i^{rel} \mid y, X) = \sum_{j=1, j \neq i}^n p(u_i^{rel} \mid y, X, u_j^{rel} = 0)P(u_j^{rel} = 0 \mid y, X) \]

\[ = \sum_{j=1, j \neq i}^n P_j p(u_i^{rel} \mid y, X, u_j^{rel} = 0). \]  

In other words, we calculate the distribution of the efficiency of firm \( i \) relative to firm \( j \) for all \( j \), and then weight by the probability that the \( j \)th firm is the most efficient. This allows us to calculate exact, finite sample results for the relative efficiencies, and deals with the difficult distributional issues which arise in the classical analysis of Schmidt and Sickles as a result of the min operator being present.

It remains to describe how to calculate \( P_j \) and \( p(u_i^{rel} \mid y, X, u_j^{rel} = 0) \). \( P_j \) is the probability that firm \( j \) is most efficient and can be expressed as

\[ P_j = P\left( \bigwedge_{i} \alpha_i - \alpha_j \geq 0 \mid y, X \right) = P(\eta^{(j)} \geq 0 \mid y, X), \]

where \( \eta^{(j)} \) is the \((n-1)\times1\) vector consisting of \( \alpha_i - \alpha_j \) for \( i = 1, \ldots, n, i \neq j \). Since the \( \alpha_i \)'s are Normally distributed, it follows that the marginal posterior of \( \eta^{(j)} \) is the \((n-1)\)-variate Normal distribution with mean and covariance matrix given in the Appendix. Thus, \( P_j \) is the posterior probability mass located in the positive orthant of the \((n-1)\)-dimensional space of \( \eta^{(j)} \). It is possible to obtain analytical approximations to \( P_j \), but we choose to perform Monte Carlo integration since the Monte Carlo draws used for calculating \( P_j \) can also be used for calculating \( p(u_i^{rel} \mid y, X, u_j^{rel} = 0) \). That is, for each \( j \), we draw random vectors from the appropriate \((n-1)\)-dimensional Normal distribution and count the proportion of draws which have all elements positive; this proportion is an estimate for \( P_j \). Note that this procedure theoretically is very computationally intensive since it must be performed for all \( n \) firms. In practice, however, it will usually be the case that only a few firms have nonnegligible \( P_j \)'s. For this reason, we pursue the following strategy: the firms are ordered from the smallest \( \tilde{\alpha}_j \) to the largest. We start by computing \( P_1 \) (which is typically the largest), followed by \( P_2 \), etc. We stop computation when \( \sum P_j > 0.999 \). All subsequent \( P_j \)'s are set to zero. Hence, the computational demands of this approach are much reduced.

\( p(u_i^{rel} \mid y, X, u_j^{rel} = 0) \) can be calculated as a by-product of the Monte Carlo integration procedure described in the previous paragraph. That is, \( p(u_i^{rel} \mid y, X, u_j^{rel} = 0) \) is equivalent to \( p(\eta^{(j)} \mid y, X, \eta^{(j)} \geq 0) \), which is the appropriate marginal from the joint truncated Normal distribution. Hence, the accepted
Monte Carlo draws used in calculating the $P_j$'s automatically provide draws from $p(u^{rel}_j | y, X, u^{rel}_j = 0)$. As log costs is the dependent variable, and interest centres on $\exp(-u^{rel}_j)$, the Monte Carlo draws are transformed and used to plot the efficiency measure.\footnote{One more point should be noted before proceeding to the one-sided individual effects case. It is often of interest to allow for firm-specific inefficiency to depend upon some other variables. For example, for hospitals it is of interest to see if different organizational structures (e.g., nonprofit vs. for-profit) tend to imply different efficiency levels. However, variables such as for-profit status do not vary over time. If we were to introduce such time-invariant variables, then the $k \times k$ matrix $S$ defined in the Appendix would be singular and the posterior would not be defined. Hence, for the standard individual effects case we do not allow for firm-specific efficiencies to depend on organizational structure or other firm characteristics.}

**Marginally independent efficiency distribution (MIED) model**

Our second Bayesian fixed effects model is still characterized by the absence of links between the individual effects, which translates into marginal prior independence between the $u_i$'s. In the preceding discussion of the standard model, a noninformative prior was used for $\alpha$. We now use an informative prior. When dealing with stochastic frontiers, the individual effects, $u_i$, are measures of inefficiency and thus, by definition, are nonnegative. This fact has motivated various classical maximum likelihood studies (e.g., Pitt and Lee, 1981; Schmidt and Sickles, 1984) and will be at the basis of our models with proper distributions on the individual effects.

As before, Eq. (2) provides the basic model, and let $u = (u_1, \ldots, u_n)'$ be the vector of firm-specific inefficiencies. The nonnegativity of the $u_i$'s can be thought of as prior information that the researcher should impose. We assume a particular one-sided prior distribution for $u$ that is similar to that used in van den Broeck, Koop, Osiewalski, and Steel (1994). The reader is referred to this paper for more details regarding this prior. We let $u_i$ be independent of $v_i$ and i.i.d. exponential with firm-specific mean $\lambda_i$. Note that the specification of a distribution for the $u_i$'s allows us to talk of absolute inefficiencies, unlike the standard individual effects case. The parameters $\lambda_i^{-1}$ are assumed to have independent exponential priors with means all equal to $-1/\ln(r^*)$ ($i = 1, \ldots, n$). The marginal prior of the efficiency of firm $i$, $r_i = \exp(-u_i)$, is given by $p(r_i) = r_i^{-1} \int_0^\infty (-\ln(r) | 1, 1, -\ln(r^*))$, where

$$f_{\lambda}(z | a, b, c) = \frac{\Gamma(a + b)}{c^a \Gamma(a) \Gamma(b)} \left[ \frac{1 + \frac{z}{c}}{z} \right]^{b-1} \left[ \frac{1 + \frac{z}{c}}{z} \right]^{-(a+b)}$$

denotes the density function of the three-parameter inverted Beta or Beta prime distribution (see Zellner, 1971, pp. 375–376). Since $r^*$ can be shown to be the prior median efficiency, prior elicitation can be performed based on an easily understood quantity (see van den Broeck, Koop, Osiewalski, and Steel, 1994, Sec. 6). As a result of the prior independence of the $\lambda_i$'s, efficiencies are marginally...
prior independent. We stress that an explicit parameterization in terms of $\lambda_i$'s is not formally required, as we could immediately start from the marginal Inverted Beta prior distributions on the $u_i$'s. The corresponding Bayesian model then becomes

$$p(y, u, \alpha_0, \beta, \sigma^{-2} | X) = p(y | X, u, \alpha_0, \beta, \sigma^{-2}) p(u) p(\alpha_0, \beta, \sigma^{-2})$$

$$= c \sigma^2 \prod_{i=1}^{n} f_{N}(y_i | X_i \beta + (\alpha_0 + u_i) \sigma^2 \sigma^{-2})$$

$$\times f_{ID}(u_i | 1, 1, -\ln(r^*)) \text{ (6)}.$$  

However, introducing the incidental parameters $\lambda_i$ and thus using a trivial hierarchical prior structure considerably facilitates the numerical analysis of this model through Gibbs sampling. In addition, it allows for a more direct comparison with the random effects models, to be introduced later.

In the absence of data corresponding to a particular firm $f$, the individual effect $u_f$ will not be updated by the observations in the sample. Note that the latter result requires both prior independence between $u_f$ and $(u, \alpha_0, \beta, \sigma^{-2})$ as well as sampling independence over firms, which is assumed throughout. Therefore, in Bayesian fixed effects models the sample cannot help us in predicting individual effects (efficiencies) of unobserved firms.

We wish to make inferences about $\alpha_0$, $\beta$, $\sigma^2$, and $u$. It turns out that the joint posterior distribution is very difficult to work with. However, conditional posterior distributions have relatively simple forms. This suggests that a Gibbs sampler can be set up for this model. Details of the Gibbs sampler are given in the Appendix.

2.2. Bayesian random effects models

In this Subsection, we focus on models where the individual effects are in some way related across firms. Here we distinguish between two models depending on the way in which these links are implemented. In the usual panel context, one often considers a hierarchical prior for $\alpha$ where $\alpha$ is Normal with mean $\mu_\alpha$ and variance $\sigma^2_\alpha$. The parameters $\mu_\alpha$ and $\sigma^2_\alpha$ in turn have a Normal-Gamma prior distribution (see, e.g., Box and Tiao, 1973). It can easily be seen that this model has the same type of structure as a classical Normal random effects model, which is commonly assumed in the panel data literature, as, e.g., in Simar (1992). The resulting variance component model with an intra-class covariance structure is typically analyzed using generalized least squares in a classical framework.

\footnote{An introduction to the Gibbs sampler is given in Gelfand and Smith (1990) or Koop (1994). A discussion of the use of Gibbs sampling techniques in stochastic frontier models with cross-sectional data is given in Koop, Steel, and Osiewalski (1995). A discussion of Gibbs sampling techniques for random coefficients models with panel data is given in McCulloch and Rossi (1994).}
(see, e.g., Schmidt and Sickles, 1984; Simar, 1992), and can readily be treated using Gibbs sampling in a Bayesian context (see Gelfand and Smith, 1990; McCulloch and Rossi, 1994). However, we are here dealing with individual effects, \( u_i \), that are nonnegative by definition, and thus we will use hierarchical structures that reflect this property. For classical maximum likelihood estimation of random one-sided individual effects models, see Pitt and Lee (1981) and Schmidt and Sickles (1984).

**Varying efficiency distribution (VED) models**

One might reasonably assume that the efficiency of hospitals with similar characteristics could be related. One way of implementing such links is to parameterize the mean inefficiencies \( \lambda_i \), which were all independent in the MIED model, as \( \exp(-w_i'\gamma) \), where \( w_i \) is an \( m \times 1 \) vector of exogenous dummy variables, which do not have a time component, to be used in explaining firm-specific inefficiency (e.g., nonprofit status) and \( \gamma = (\gamma_1 \ldots \gamma_m)' \) is an \( m \times 1 \) parameter vector, linking the firm effects across the sample. The parameter vector \( \gamma \) is assumed to have a proper prior, independent of the other parameters. All other assumptions are identical to the MIED model. The Bayesian model is now:

\[
p(y, u, \gamma, \alpha_0, \beta, \sigma^{-2} | X, W)
= p(y | X, u, \gamma, \alpha_0, \beta, \sigma^{-2}) p(u, \gamma | W) p(\alpha_0, \beta, \sigma^{-2})
= c \sigma^2 p(\gamma) \prod_{i=1}^{n} f_N(y_i | X_i \beta + (\alpha_0 + u_i) \iota, \sigma^2 I_T) f_G(u_i | 1, \exp(w_i'\gamma)),
\]

where \( W = (w_1 \ldots w_n)' \). We assume \( w_{i1} = 1 \) for all \( i \). It is worth noting that our framework explicitly allows for functional links between \( X_i \) and \( w_i \), given that we condition on \( X \) (and \( W \)) throughout. That is, we could allow for the individual effects to be correlated with the variables describing the frontier.

The conditional posterior for \( \gamma \) depends on the form of the prior for \( \gamma \). In order to maintain comparability with the common efficiency distribution case, to be introduced later, we reparameterize as

\[
\exp(w_i'\gamma) = \prod_{j=1}^{m} \phi_j^{w_{ij}},
\]

where \( \phi_j = \exp(\gamma_j) \).\(^4\) We express our prior information in terms of the \( \phi_j \)'s. In particular, we assume that they are, a priori, i.i.d. Gamma with parameters

---

\(^4\)The assumption the \( w_{ij} \)'s are 0–1 dummy variables is of importance for technical reasons in that the conditional distributions of the \( \phi_i \)'s have a simple Gamma form only with dummy variables. An early version of this paper (Koop, Osiewalski, and Steel, 1994b) allowed for the \( w \)'s to be other than dummies. In this case, the conditionals for the \( \phi_i \)'s do not have a convenient form. To surmount this problem we used an independence Metropolis algorithm (see Tierney, 1991). The reader is referred to our earlier working paper for more details.
and $g_j$, $j = 1, \ldots, m$. Despite the conditional prior independence of the $u_i$'s, inefficiencies are marginally linked through the common parameter $\gamma$ in the conditional mean. The properties of this prior will be discussed shortly, and details regarding the Gibbs sampler for this model are given in the Appendix.

**Common efficiency distribution (CED) model**

In the previous VED model only the efficiencies of the firms with the same characteristics (as measured by $w_i$) are drawn from a common distribution. Here we will develop an important special case of the VED model, where $m = 1$, and, since $w_{i1} = 1$, this amounts to assuming that all individual effects are independent drawings from the same distribution. Thus, the links between firm effects will be even stronger than in the previous model.

We now assume that $u_i$ is still independent of $v_i$ and i.i.d. exponentially distributed with a common mean $\mu$. Thus, the CED is a special case of the VED model with $m = 1$ and $\mu = \phi_1^{-1}$. The Bayesian model is given by (7) with $W = n\mu$, and $\gamma = -\ln(\mu)$. The prior on $\mu^{-1}$ is now taken to be exponential with mean $-1/\ln(r^*)$, i.e., the same as the marginal prior for each $\lambda_j^{-1}$ used to parameterize the MIED model. Therefore, the marginal distribution of $r_i$ will be exactly the same as in the latter model, but the $r_i$'s will no longer be *a priori* independent. As $\mu$ now has the interpretation of a common mean inefficiency, we will also be interested in inference on $\mu$. For $T = 1$ our CED model reduces exactly to the exponential model used for analyzing cross-sectional data in van den Broeck, Koop, Osiewalski, and Steel (1994). Again, the Gibbs sampler is a natural method to treat the numerical integration required for posterior and predictive inference and details are given in the Appendix.

We remind the reader that the CED model is a special case of the VED model, and we would like to specify a prior on the latter that is consistent with the prior assumed for the CED model. In the VED model each $\phi_j$ had a Gamma prior distribution with parameters $a_j$ and $g_j$, $j = 1, \ldots, m$. To ensure that the prior for the VED model is consistent with that of the CED model, we set $g_1 = -\ln(r^*)$, $g_j = 1$ for $j = 2, \ldots, m$, and $a_j = 1$ for $j = 1, \ldots, m$. In other words, we centre the prior over the CED model with the same hyperparameter, $r^*$, which is no longer the prior median if $m > 1$. In the VED case, $P(r_i < r^*) = \mathbb{E}[1 + \prod_{j=2}^m \phi_j]^{-1}$, where the expectation is with respect to the prior of $\phi_2, \ldots, \phi_m$. Applying Jensen's inequality, we conclude that $P(r_i < r^*) > 0.5$, i.e., the prior median efficiency is less than $r^*$ whenever $m > 1$.

Tables 1 and 2 and Fig.1 illustrate some properties of the prior for the varying and common efficiency distribution cases. Table 1 contains prior efficiency means and standard deviations for various values of $r^*$ for $a_j = g_j = 1$ ($j = 2, \ldots, m$). Although changing $r^*$ changes the prior moments, a comparison of $m = 1$ with

---

5 The derivations in this paragraph are straightforward extensions of derivations given in van den Broeck, Koop, Osiewalski, and Steel (1994, Sec. 6).
Table 1
Prior means of efficiency for different values of \( r^* \) (prior standard deviations in parentheses)

<table>
<thead>
<tr>
<th>( r^* )</th>
<th>0.01</th>
<th>0.20</th>
<th>0.50</th>
<th>0.80</th>
<th>0.99</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIED</td>
<td>0.16</td>
<td>0.32</td>
<td>0.48</td>
<td>0.69</td>
<td>0.96</td>
</tr>
<tr>
<td>CED</td>
<td>(0.26)</td>
<td>(0.32)</td>
<td>(0.34)</td>
<td>(0.30)</td>
<td>(0.11)</td>
</tr>
<tr>
<td>VED</td>
<td>0.10</td>
<td>0.19</td>
<td>0.28</td>
<td>0.42</td>
<td>0.78</td>
</tr>
<tr>
<td>( (m = 4) )</td>
<td>(0.23)</td>
<td>(0.31)</td>
<td>(0.35)</td>
<td>(0.39)</td>
<td>(0.32)</td>
</tr>
</tbody>
</table>

Table 2
Prior means of efficiency for different values of \( a_j \) and \( g_j \), \( j > 1 \), \( r^* = 0.8 \) (prior standard deviations in parentheses)

<table>
<thead>
<tr>
<th>( a_j = g_j = )</th>
<th>0.6</th>
<th>1</th>
<th>5</th>
<th>10</th>
<th>25</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>VED</td>
<td>0.31</td>
<td>0.42</td>
<td>0.62</td>
<td>0.65</td>
<td>0.67</td>
<td>0.68</td>
</tr>
<tr>
<td>( (m = 4) )</td>
<td>(0.39)</td>
<td>(0.39)</td>
<td>(0.34)</td>
<td>(0.32)</td>
<td>(0.31)</td>
<td>(0.31)</td>
</tr>
</tbody>
</table>

Fig. 1. Marginal prior of efficiencies, one-sided individual effects \( (r^* = 0.8) \), SIE model.

\( m = 4 \) for \( r^* = 0.8 \) in Fig. 1 indicates that the priors have roughly similar properties with respect to efficiency.

Table 2 investigates the effect of changing \( a_j \) and \( g_j \) \( (j = 2, \ldots, m) \) for the \( m = 4 \) case. We keep \( a_j = g_j \), but let them take on a common value different from 1. This implies that the \( \phi_j \)'s still have mean 1 (and hence are centred over the \( m = 1 \) model), but allows their prior variances to take on values different from 1. It can be seen that changing these prior variances has a reasonably
small effect on the prior means of efficiency relative to prior standard deviations. The latter change very little across priors. Hence, in our empirical work, we set $a_j = g_j = 1$ ($j = 2, \ldots, m$), which is judged a reasonable value, and do not investigate other values for these prior hyperparameters. Clearly, as $a_j$ and $g_j$ grow for $j = 2, \ldots, m$, the prior for $\phi_j$ ($j > 1$) will become tighter around one and prior efficiency for the case $m = 4$ will tend to that with $m = 1$ (which can be derived from the model with $m > 1$ by restricting all $\phi_j$ ($j > 1$) to be one).

Fig. 1 plots the marginal prior density of efficiency for the $m = 1$ ($r^* = 0.8$) and $m = 4$ ($r^* = 0.8$, $a_j = g_j = 1$ for $j = 2, \ldots, m$) cases. For $m = 4$ we take the average values of the $w_{ij}$. In our prior sensitivity analysis, to be discussed in Section 4, we find that posterior results for $m = 1$ are extremely robust for the random effect models, even to enormous changes in $r^*$. In view of this, the small differences in prior between the $m = 1$ and $m = 4$ cases will undoubtedly have negligible empirical consequences. In contrast, the choice of $r^*$ is found to be much more important in the MIED model, in line with the fixed effects structure of that model. In addition, Fig. 1 contains the continuous part of the implied prior on the relative efficiency $r_i^{rel}$ for the SIE model (with arbitrary scaling), from which it is obvious that this model corresponds to a very strong prior belief in low efficiency. The proper priors, on the other hand, convey a genuine sense of lack of strong prior information. The particular U-shaped form of their densities reflects the thick tail of the marginal prior on $u_i$, which is evident from (6) in the case of the MIED model.

3. Hospital cost function estimation

The theory of the multiple-product firm implies that a firm’s costs should depend upon the quantity of each output produced as well as the input prices faced by the firm. Given data on outputs and input prices, the researcher can select a functional form for the cost frontier and estimate its parameters. Hospital cost function estimation poses problems which render it difficult to apply such a strategy in a straightforward manner. A modern hospital produces a myriad of outputs that are hard to quantify. For most standard functional forms (e.g., the translog), the large number of outputs causes the number of parameters to be

---

6 Cowing, Holtmann, and Powers (1983) provide a discussion of some of the difficulties inherent in hospital cost function estimation.

7 Most theories of the firm imply that cost minimization should be a reasonable objective function even for nonprofit or government-run institutions. Empirical evidence for or against the assumption of cost minimization is scant. One exception is Eakin and Kniesner (1988), which estimates a long-run cost function (using 1975-76 data) that allows for systematic allocative inefficiency and rejects the assumption of cost minimization. However, since estimated differences between shadow and observed marginal costs are small, the authors conclude that use of traditional minimum cost functions may yield fairly accurate estimates of output concepts, such as economies of scale and scope.
estimated to be very large. As a result, researchers have worked with highly aggregated data. Early work in this area avoided such flexible functional forms and worked with ad hoc (usually linear) specifications, where the class of explanatory variables was expanded beyond that implied by economic theory (see Breyer, 1987, for a discussion). Much of the recent work (see Vita, 1990; Granneman, Brown, and Pauly, 1986) has criticized these ad hoc specifications and worked with flexible functional forms such as the translog. In this paper, we intend to follow the path of these latter authors and work with highly aggregated data and a translog functional form.

In addition to the general area of hospital cost analysis, there has also been a great deal of interest in the specific area of hospital efficiency analysis. A recent special issue of the *Journal of Health Economics* (Vol. 13, No. 3, 1994) is devoted to hospital efficiency analysis using mathematical programming or sampling-theory statistical methods. The interested reader is referred to this issue for more details. Suffice it to note here that none of the empirical work in the special issue uses panel data (see Zuckerman, Hadley, and Iezzoni, 1994; Vitaliano and Toren, 1994a), but that the potential usefulness of panel data is stressed in the discussion (see Skinner, 1994; Dor, 1994; Vitaliano and Toren, 1994b). Hence, our paper can be thought of as a logical extension of the existing empirical hospital efficiency literature.

Breyer (1987), in a survey of the hospital cost function literature, argues that the true `output' of a hospital is improvement in patient health. Defined in this way, `output' is impossible to measure, so Breyer recommends using observable intermediate products as proxies for output. In particular, he suggests three important hospital output dimensions that can easily be measured: i) number of cases (as a proxy for medical services), ii) number of inpatient days (as a proxy for nursing, accommodation, and other `hotel' services), and iii) number of beds (to satisfy an operating demand for hospital services). In this present paper, we use these three variables as measures of output. In addition, we include the number of outpatient visits and a case mix index as other aspects of outputs that are included in our cost frontier. We use data for \( n = 382 \) nonteaching U.S. hospitals (over the years 1987–1991) which were selected so as to constitute a relatively homogeneous sample. The Data Appendix to Koop, Osiewalski, and Steel (1994b) provides more detail.

In our cost frontier, we also include a measure of capital stock (total fixed assets). The fact that such a variable is included means that all our results should be interpreted as applying in the short run.

In terms of inputs, labour is predominant. Our data source (described in Koop, Osiewalski, and Steel, 1994b) contains only one aggregate wage index. The other major input in hospital technology is general materials and supplies.

\[^8\text{It has been argued by some researchers that the number of beds is better considered as a particular type of capital, this should be kept in mind when considering our empirical results.}\]
Unfortunately, we have no measurements on this. However, given the wholesale buying power of most hospitals, it is reasonable to assume that the price of materials is fairly constant across hospitals. Hence, we treat the price of materials as a constant. Such a treatment is undoubtedly reasonable cross-sectionally, but is not reasonable over time. Thus, we add a time trend and a time trend squared as explanatory variables in our cost frontier to try to capture the missing time dimension of the price of materials or other dynamics that are not modelled explicitly.

With this data, we can specify a standard cost function where costs depend on five output categories, one capital stock, two input prices, a time trend, and a time trend squared. We choose a translog specification and impose linear homogeneity in prices. The imposition of linear homogeneity allows us to normalize with respect to the price of materials (which is a constant), yielding the resulting cost frontier:

\[
\ln C = \alpha_0 + \sum_{i=1}^{5} \beta_i \ln Y_i + \beta_6 \ln P + \sum_{i=1}^{5} \sum_{j=i}^{5} \psi_{ij} \ln Y_i \ln Y_j + \beta_7 (\ln P)^2 \\
+ \sum_{i=1}^{5} \beta_7+1 \ln Y_i \ln P + \beta_{13} \ln K + \sum_{i=1}^{5} \beta_{13+i} \ln Y_i \ln K \\
+ \beta_{19} \ln P \ln K + \beta_{20} (\ln K)^2 + \beta_{21} t + \beta_{22} t^2,
\]

where \(\psi_{ij}\) are the remaining 15 elements of \(\beta(\psi_{ij} = \psi_{ji})\). A brief description of the variables is given in Table 3, and more detail is provided in the Data Appendix to Koop, Osiewalski, and Steel (1994b). The explanatory variables (\(w_{ij}\)) for the varying efficiency distribution model are taken to be the ownership status (i.e., nonprofit or for-profit, the omitted category being government-run) and a dummy variable which equals one if the ratio of clinical personnel to average daily census is above average.

4. Empirical results

In this section we discuss our empirical findings for the models considered in Section 2: the standard individual effects (SIE) model, the marginally independent efficiency distribution (MIED) model, the varying efficiency distribution (VED) model, and the common efficiency distribution (CED) model. We remind the reader that the first two of these models are Bayesian fixed effects models, the second pair constitute Bayesian random effects models. These models are arranged in order of increasing prior links between the firm-specific efficiencies. The findings from these models are discussed and compared in different subsections.
Table 3

Description of variables

| \( C \)  | costs (facility operating expenditure) |
| \( Y_1 \) | number of discharges |
| \( Y_2 \) | number of inpatient days |
| \( Y_3 \) | number of beds |
| \( Y_4 \) | number of outpatient visits |
| \( Y_5 \) | case mix index |
| \( P \)  | wage index |
| \( K \)  | capital stock |
| \( t \)  | time trend |
| \( w_1 \) | intercept |
| \( w_2 \) | dummy variable for nonprofit hospitals |
| \( w_3 \) | dummy variable for for-profit hospitals (dummy for government-run hospitals dropped) |
| \( w_4 \) | dummy which equals 1 if ratio (averaged over years) of clinical workers to average daily census is above average (= 0 otherwise) |

Properties of the efficiency measures

The Bayesian techniques used in this paper yield exact posterior distributions for the efficiency measures of each of our \( n \) hospitals. These are probably of greatest interest for policy purposes. However, space precludes a detailed presentation of efficiencies for each hospital. Instead, we present a detailed analysis of the three firms which have minimum, median, and maximum values for \( \hat{\alpha}_i \) in (A.3), thus representing low, medium, and high efficiency levels. The corresponding efficiencies will be denoted by \( r_{\text{min}} \), \( r_{\text{med}} \), and \( r_{\text{max}} \), respectively. For the random effects models we consider an ‘average’ hospital. The notion of the efficiency of an ‘average’ or ‘typical’ out-of-sample firm, \( r_f \), is discussed in van

\[ \text{Since the emphasis of the paper is on efficiency analysis we do not report the properties of the frontier itself. The interested reader is referred to Koop, Osiewalski, and Steel (1994b) for results in this area.} \]
den Broeck, Koop, Osiewalski, and Steel (1994). Essentially, $r_f$ is the efficiency of a hypothetical unobserved hospital. The posterior distribution of $r_f$ is given by

$$p(r_f | y, X) = r_f^{-1} \int_0^\infty f_G(-\ln(r_f) | 1, \mu^{-1}) p(\mu^{-1} | y, X) \, d\mu^{-1}$$

and by

$$p(r_f | y, X) = r_f^{-1} \int f_G \left( -\ln(r_f) | 1, \prod_{j=1}^m \phi_j^{w_j} \right)$$

$$\times p(\phi_1, \ldots, \phi_m | y, X) \, d\phi_1 \cdots d\phi_m,$$

for the CED and VED models, respectively. For the latter model, we have characterized the 'typical' firm as having average values for the $w_{ij}$'s i.e., $\bar{w}_j$. Note that this implies a somewhat artificial 'average' hospital since the $w_{ij}$'s are dummy variables. In other words, $p(r_f | y, X)$ is posterior to the data on all observed firms, but prior to the unobserved output for hypothetical firm $f$. The updating of $r_f$ is only possible due to prior links between efficiencies, and hence considering $r_f$ does not make sense for the fixed effects models.

Fig. 2 plots the marginal posterior density of $r_f$ for the CED and VED cases. It can be seen that the plots are almost identical. If we compare Fig. 2 with the prior for the one-sided efficiency distribution models in Fig. 1, it can clearly be seen that the data rule out very low efficiencies. The posterior means (standard deviations) of $r_f$ for the CED and VED cases are 0.856 (0.125) and 0.862 (0.120), respectively. As inference on $r_f$ in the random effects models is conducted without knowledge of $y_f$ and $X_f$, its distribution is much more spread out than the posterior distributions of firm-specific efficiencies. The distribution of $r_f$ depicted in Fig. 2 also captures the uncertainty about the type of hospital. For example, firm $f$ could resemble a very inefficient firm, but could also be like a very efficient one.

Figs. 3 through 6 plot the posterior distributions of $r_{rel}^{min}$, $r_{rel}^{med}$, and $r_{rel}^{max}$ for the SIE, and of $r_{min}$, $r_{med}$, and $r_{max}$ for the MIED, VED, and CED models, respectively. Since Figs. 4, 5, and 6 are almost identical, we will focus on comparing CED with SIE (i.e., comparing Fig. 6 with Fig. 3). Remember that, for the SIE model, efficiency is measured relative to the most efficient firm. It turns out that one firm is almost certainly the most efficient firm (i.e., the posterior probability that it is most efficient is 0.9997). For this reason the distribution of $r_{rel}^{max}$ is a point mass at one. However, even for the CED model (which allows for the calculation of absolute efficiency), $r_{max}$ has all of its probability mass very close to one. It is with regard to $r_{med}$ and $r_{min}$ that major differences occur. In particular, posterior means are much smaller for the SIE model than for the other models (0.23 vs. around 0.70 for $r_{rel}^{min}$ and $r_{min}$, respectively). This behaviour is consistent
with the important difference between the SIE and one-sided distribution models. The implied prior on $r_{i}^{\text{rel}}$ (see Fig. 1) simply does not rule out very low relative efficiencies for many firms. From the fact that the posterior results for the MIED model (with $r^{*}=0.8$) are quite close to those of the random effects models, we infer that it is not the fixed effects nature that induces the SIE model to behave so differently from the rest. Rather, it is the improper prior structure on the intercepts.
If we replace the one-sided prior distribution (with $r^* = 0.8$) on the inefficiencies by the improper prior structure of the SIE model, we tend to substantially decrease the hospital efficiencies. Overall, the average posterior mean efficiency is 0.47 for the SIE model and around 0.85 for the one-sided efficiency distribu-
tion models. However, the differences go beyond merely decreasing efficiency of each hospital; in many cases the ranking of efficiencies changes. The Spearman rank correlation between the n-vector of posterior means of hospital-specific efficiencies for the SIE and CED cases is only 0.43.

To illustrate the fact that the standard individual effects model considers relative, rather than absolute, measures of efficiency, we have eliminated the firm with the highest value of $\hat{a}_i$ (which was most efficient with probability 0.9997). Then we need to consider six candidates for most efficient firm in order to capture a probability mass of at least 0.999, and the average posterior mean efficiency jumps to 0.55. Using the CED model on this reduced sample leads to an average posterior mean efficiency of 0.85, which hardly differs from that with the full sample. The results from the one-sided models are, in our subjective opinion, much more reasonable than those of the SIE model, both in terms of the efficiency measures (is it reasonable that, as the SIE model would have, there are many hospitals with efficiencies less than 30% of the most efficient firm?) and the frontier itself (fewer suggestions that regularity conditions are violated for the one-sided cases) Thus, we would advise against use of the SIE model. Therefore, in the rest of the discussion of our empirical results, we will primarily concentrate on the one-sided efficiency distribution models.

Even for the one-sided distribution models, the inference on firm-specific efficiencies suggests that cost-minimization is not achieved by many of the hospitals in our sample. Our analysis cannot tell us, however, whether this fact is due to managerial error, to some missing output (e.g., quality of care) or to a different objective function.
Prior sensitivity analysis

The one-sided individual effects models are based upon proper priors for efficiencies. We believe that we have elicited quite reasonable priors, but it is always important to carry out a sensitivity analysis to see if the choice of prior hyperparameters (in our case $r^*$) has an important effect on our results. This prior hyperparameter $r^*$ has the interpretation of the prior median efficiency for both the MIED and CED models. Let us compare these two models, as this enables us to isolate the effect of prior independence. For our previous discussions, we have set $r^* = 0.8$, implying that we assign a prior probability of 0.5 to any hospital being less than 80 percent efficient.

Fig. 7 displays the average of posterior mean efficiencies over the 382 hospitals within the sample, say $\bar{\epsilon}$, as a function of $r^*$. The striking feature of these graphs is that $\bar{\epsilon}$ is virtually constant over the whole range of $r^*$ from 0.01 until 0.99 for the CED model. It seems the random effect nature of this model links the efficiencies sufficiently to ensure that the data dominate the prior information. In sharp contrast, the fixed effects MIED model does not lead to such robustness, as $\bar{\epsilon}$ varies substantially with $r^*$. The independence assumption inherent to this model, combined with the small value of $T$, implies that the data information on each individual efficiency is much weaker than in the CED case. If we take a value of $r^*$ in line with the data, say $r^* = 0.8$, then both models lead to virtually the same inference on firm efficiencies, but if we deviate from such a prior median efficiency, differences between the models grow. As $r^*$ becomes very small, the MIED model will have marginal priors on the efficiencies that are still proper, but will start to look like the marginal priors on relative efficiencies for the SIE model. As both models are fixed effects models and their only difference lies in the form of the marginal priors, we find, indeed, that results for the MIED model with $r^* = 0.01$ are relatively close to the SIE model.

To further illustrate the robustness of our results for the CED model to extreme changes in our prior, Fig. 8 plots the posterior distribution of $r_f$ for $r^* = 0.01, 0.80$ and 0.99. It can be seen that the plots are virtually indistinguishable. Posterior means of $r_f$ for these three values of $r^*$ are 0.843, 0.856, and 0.857 and posterior standard deviations 0.133, 0.125, and 0.124. This is strong evidence that results reported for the CED case are not dependent on the specific values of the prior hyperparameter.

Explaining efficiencies

The VED model explicitly allows for hospital characteristics to affect efficiencies. We have considered a number of candidate variables for $W$. The characteristics we finally retain are nonprofit and for-profit status dummies and a dummy based on the number of clinical workers per patient as $w_{12}, w_{13},$ and $w_{14}$, which are used to explain efficiencies.

However, we can also use the other individual effects models to shed some light on factors which may affect efficiency. Table 4 presents average posterior means
and standard deviations of individual efficiencies for for-profit, nonprofit, and government-run hospitals. Once again, the one-sided efficiency cases yield very similar results for \( r^* = 0.8 \). However, results for the SIE model are quite different. For the random effects models a clear pattern emerges: for-profit hospitals are less efficient than nonprofit or government-run hospitals.
Table 4
Averages of posterior means of efficiencies for hospital subgroups (averages of posterior standard deviations in parentheses)

<table>
<thead>
<tr>
<th></th>
<th>Nonprofit</th>
<th>For-profit</th>
<th>Govt.-run</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>SIE</td>
<td>0.456</td>
<td>0.485</td>
<td>0.502</td>
<td>0.467</td>
</tr>
<tr>
<td></td>
<td>(0.028)</td>
<td>(0.029)</td>
<td>(0.029)</td>
<td>(0.028)</td>
</tr>
<tr>
<td>MIED</td>
<td>0.853</td>
<td>0.819</td>
<td>0.821</td>
<td>0.843</td>
</tr>
<tr>
<td></td>
<td>(0.026)</td>
<td>(0.027)</td>
<td>(0.028)</td>
<td>(0.026)</td>
</tr>
<tr>
<td>VED</td>
<td>0.866</td>
<td>0.793</td>
<td>0.871</td>
<td>0.855</td>
</tr>
<tr>
<td></td>
<td>(0.026)</td>
<td>(0.027)</td>
<td>(0.026)</td>
<td>(0.026)</td>
</tr>
<tr>
<td>CED</td>
<td>0.861</td>
<td>0.796</td>
<td>0.870</td>
<td>0.851</td>
</tr>
<tr>
<td></td>
<td>(0.026)</td>
<td>(0.027)</td>
<td>(0.026)</td>
<td>(0.026)</td>
</tr>
</tbody>
</table>

Table 5
Posterior means and standard deviations of $\gamma$

<table>
<thead>
<tr>
<th></th>
<th>$\gamma_1$</th>
<th>$\gamma_2$</th>
<th>$\gamma_3$</th>
<th>$\gamma_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>2.05</td>
<td>-0.02</td>
<td>-0.50</td>
<td>-0.25</td>
</tr>
<tr>
<td>Std. dev</td>
<td>0.16</td>
<td>0.16</td>
<td>0.19</td>
<td>0.11</td>
</tr>
</tbody>
</table>

The varying efficiency distribution model allows us to investigate directly the effect of the $w_i$'s on efficiency. Posterior means and standard deviations for the elements of $\gamma$ are given in Table 5. The posterior means of $\gamma$ given in Table 5 are consistent with government-run hospitals being most efficient followed by non-profit and for-profit hospitals. An approximate Bayesian Highest Posterior Density interval test (see Zellner, 1971, pp. 298–302) favours the VED model. Remember that the CED model is equal to the VED model with $\gamma_2 = \cdots = \gamma_m = 0$. If the posterior for the $\gamma_j$’s $(j = 2, \ldots, m)$ is approximately Normal, then the distribution of the inner product of the standardized $\gamma_j$’s is approximately $\chi^2_{m-1}$. For our data, this quantity evaluated at zero, is 17.50. Since $m = 4$, this indicates that the VED model provides explanatory power beyond the CED model.

The strongest finding is that for-profit status tends to decrease hospital efficiency. In addition, the posterior mean of $\gamma_4$, the coefficient on the dummy for high worker to patient ratios, indicates that having a greater number of workers tends to decrease efficiency as well. The posterior means of $\gamma_3$ and $\gamma_4$ are both more than two standard deviations from zero. We find the result that for-profit hospitals tend to be less efficient than other hospitals counter-intuitive. One possible explanation is that for-profit hospitals tend to compete by offering higher quality service. Since it is hard to directly observe quality, low efficiency might actually be capturing higher quality. However, one way in which hospitals can offer higher quality is by providing more clinical personnel. The fact that $\gamma_3$ is
still significant when $w_4$ is added indicates that this latter mechanism for improving quality does not suffice to explain the failure of for-profit hospitals to achieve high efficiency.

In Koop, Osiewalski, and Steel (1994a), a relative predictive measure of lack of fit is developed, a detailed justification of which is given therein. If we let $e_{ft} = u_f + v_f$, then we advocate using $E(e^2_{ft}|y,X) = E(\sigma^2 + 2\mu^2|y,X)$ as a measure of fit for the CED model. For the VED model, $\exp(-w_f y)$ is analogous to $\mu$, so we use $E(\sigma^2 + 2\exp(-2\tilde{w} y)|y,X,W)$, where $\tilde{w}$ is the average of the $w_f$'s, which we use for $w_f$ corresponding to the average firm. The measure of fit is 0.061 for the CED model and 0.056 for the VED model. In other words, in terms of our measure of fit, the VED model does better. We cannot construct a comparable measure for the other models. However, note the posterior mean of $\sigma^2$ is 0.0034 for the SIE model, 0.0042 for MIED, and 0.0043 for both random effects models. The fact that the prior of the SIE model strongly favours low efficiencies implies that measurement error is smaller for this latter model, as much more of the total error is allocated to inefficiency.

Fig. 9 plots the posterior density of $r_f$ for nonprofit, for-profit, and government-run hospitals based on the VED model. Corresponding means (standard deviations) are 0.871 (0.114), 0.806 (0.161), and 0.872 (0.114) for nonprofit, for-profit, and government-run hospitals, respectively. This merely reinforces our previous conclusions, viz. that for-profit hospitals tend to be less efficient than other hospitals.
5. Conclusion

In this paper we have described and analyzed Bayesian models for inference on firm-specific efficiencies. We show how, by using different prior structures, we can derive Bayesian analogues to the classical fixed and random individual effects models. The fixed effects models are characterized by the absence of links between individual effects, and thus do not require a hierarchical prior structure. Within this class, we define the standard individual effects (SIE) model, which puts an improper uniform prior on the firm-specific intercepts and measures relative efficiencies in terms of differences between the intercepts, and the marginally independent efficiency distribution (MIED) model, where independent proper one-sided priors are used for individual effects, and efficiencies are thus defined in absolute terms. Bayesian random effects models do link individual effects through the hierarchical structure of the prior, parameterizing the $n$ effects in terms of a small number $m \ll n$ of additional parameters. The common efficiency distribution (CED) model takes $m = 1$ and assigns a common exponential prior distribution to inefficiencies; the varying efficiency distribution (VED) model allows the mean of the exponential prior to vary according to $m - 1$ hospital characteristics.

The seemingly innocuous flat prior on individual intercepts associated with the SIE model enables us to reinterpret many classical results in a Bayesian context. However, adoption of this model necessarily implies a strong prior belief in low firm efficiencies, which is very different from those commonly held. Furthermore, the fixed effects nature of this model makes it difficult for the data to correct this prior information when $T$ is small. For this reason, we would advocate using the one-sided individual effects models for inference on efficiencies in a stochastic frontier context. Whereas the MIED model allows us to capture our prior beliefs through any proper distribution, its fixed effects structure still implies a large sensitivity to the choice of the particular prior when $T$ is not large. Furthermore, the lack of links between individual effects inherent to fixed effects models make prediction of these effects for unobserved firms a useless exercise.

We apply our methods to a panel of U.S. hospitals and obtain reasonable results for both random effects models, which display an impressive robustness with respect to large changes in our prior hyperparameter.

Appendix

*SIE model*

The SIE model can be analyzed using Monte Carlo integration. The marginal posterior for the parameters of the cost frontier, $\beta$, is the $k$-variate Normal
distribution with
\[ E(\beta|y,X) = \hat{\beta} = S^{-1} \sum_{i=1}^{n} (X_i - t_T \bar{x}_i)^T (y_i - \bar{y}_i t_T), \]  
(A.1)

where
\[ \bar{x}_i = \frac{1}{T} X_i t_T, \quad \bar{y}_i = \frac{1}{T} t_T y_i, \]
\[ S_i = (X_i - t_T \bar{x}_i)'(X_i - t_T \bar{x}_i), \quad S = \sum_{i=1}^{n} S_i. \]

Eq. (A.1) is the so-called 'within estimator' from the panel data literature. The covariance matrix for the marginal posterior for \( \beta \) is given by
\[ V(\beta|y,X) = \hat{\sigma}^2 S^{-1}, \]  
(A.2)

where
\[ \hat{\sigma}^2 = \frac{1}{n(T-1) - k} \sum_{i=1}^{n} (y_i - \hat{\alpha}_i t_T - X_i \hat{\beta})' (y_i - \hat{\alpha}_i t_T - X_i \hat{\beta}). \]

The marginal posterior of \( \alpha \) is the \( n \)-variate Normal distribution with means
\[ E(\alpha_i|y,X) = \hat{\alpha}_i = \bar{y}_i - \bar{x}_i \hat{\beta}, \quad i = 1, \ldots, n, \]  
(A.3)

and covariances
\[ \text{cov}(\alpha_i, \alpha_j|y,X) = \hat{\sigma}^2 \left( \frac{\delta(i,j)}{T} + \bar{x}_i S^{-1} \bar{x}_j \right), \quad i, j = 1, \ldots, n, \]  
(A.4)

where \( \delta(i,j) = 1 \) if \( i = j \) and 0 otherwise.

Since the \( \alpha_i \)'s are Normally distributed, it follows that the marginal posterior of \( \eta^{(j)} \) is the \((n-1)\)-variate Normal distribution with means
\[ E(\eta^{(j)}_i|y,X) = \hat{\alpha}_i - \hat{\alpha}_j, \]
and covariances
\[ \text{cov}(\eta^{(j)}_i, \eta^{(j)}_h|y,X) = \hat{\sigma}^2 \left[ (\bar{x}_i - \bar{x}_j)' S^{-1} (\bar{x}_h - \bar{x}_j) + \frac{1 + \delta(i,h)}{T} \right], \]
for \( i, h = 1, \ldots, n, i \neq j, h \neq j \), where \( \eta^{(j)}_i \) is \( \alpha_i - \alpha_j \).

**MIED model**

The conditional distributions used in the Gibbs sampler have simple forms. In particular, conditional on \( u \) and \( \lambda^{-1} = (\lambda^{-1}_1 \ldots \lambda^{-1}_n)' \), the posterior density of the frontier parameters and precision, \( \sigma^{-2} \), has the usual Normal-Gamma form:
\[ p(\alpha_0, \beta, \sigma^{-2}|y,X,u,\lambda^{-1}) = p(\sigma^{-2}|y,X,u) p(\alpha_0, \beta|y,X,u, \sigma^{-2}), \]  
(A.5)
where

\[
p(\sigma^{-2}|y,X,u) = f_G \left( \sigma^{-2} \left| \frac{nT - k - 1}{2} \right. \right),
\]

\[
\frac{1}{2} \left[ y - (t_n: X) \left( \begin{array}{c} \alpha^* \\ \beta^* \end{array} \right) - (I_n \otimes I_T)u \right]' \times \left[ y - (t_n: X) \left( \begin{array}{c} \alpha^* \\ \beta^* \end{array} \right) - (I_n \otimes I_T)u \right] 
\]

(A.6)

and

\[
p(\alpha_0, \beta|y,X,u, \sigma^{-2}) = f_{N}^{k+1} \left( \begin{array}{c} \alpha_0 \\ \beta \end{array} \right) \left. \left| \left( \begin{array}{c} \alpha^* \\ \beta^* \end{array} \right), \sigma^{-2} \left[ \frac{nT}{nTX^*} \frac{nTX^*}{X'X} \right]^{-1} \right. \right), \]

(A.7)

In the previous equations, \( f_G(.|a,b) \) denotes the density function of the Gamma distribution with mean \( a/b \) and variance \( a/b^2 \), \( y = (y_1 \ldots y_n)' \), an \( nT \times 1 \) vector, \( X = (X_1^' \ldots X_n^') \), an \( nT \times k \) matrix, and

\[
\left( \begin{array}{c} \alpha^* \\ \beta^* \end{array} \right) = \left( \begin{array}{c} \frac{nT}{nTX^*} \\ \frac{nTX^*}{X'X} \end{array} \right)^{-1} \left( \begin{array}{c} nT(\bar{y}^* - \bar{u}) \\ X'X - X'(I_n \otimes I_T)u \end{array} \right).
\] Here,

\[
\bar{y}^* = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} y_{it}, \quad \bar{u} = \frac{1}{n} \sum_{i=1}^{n} u_i,
\]

and

\[
\bar{X}^* = \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} x_{it}.
\]

The conditional posterior for the inefficiencies takes the form:

\[
p(u|y,X,\alpha_0, \beta, \sigma^{-2}, \lambda^{-1}) 
\]

\[
\propto f_{N}^{n} \left( u|\bar{y} - (t_n: \bar{X}) \left( \begin{array}{c} \alpha_0 \\ \beta \end{array} \right) - \frac{\sigma^2}{T} \lambda^{-1}, \frac{\sigma^2}{T} I_n \right) \times I^+(u), \quad (A.8)
\]

where

\[
\bar{y} = \left( \begin{array}{c} \bar{y}_1 \\ \vdots \\ \bar{y}_n \end{array} \right), \quad \bar{X} = \left( \begin{array}{c} \bar{x}_1^' \\ \vdots \\ \bar{x}_n^' \end{array} \right),
\]

and \( I^+(u) \) is the indicator function for \( R^n_+ \). In other words, the \( u_i \)'s are independently Normally distributed, but truncated to the positive orthant. The conditional
posterior for $\lambda^{-1}$ becomes

$$p(\lambda^{-1}|y, X, u, \alpha_0, \beta, \sigma^{-2}) = p(\lambda^{-1}|u) = \prod_{i=1}^{n} f_G(\lambda_i^{-1}|2, u_i - \ln(r^*)). \quad (A.9)$$

A Gibbs sampler can be set up in terms of Eqs. (A.5), (A.8), and (A.9). Random sampling from all these densities is standard. Note that, despite the high dimensionality of the problem (2n + k + 2 = 803 in our problem), three steps suffice for each Gibbs draw. It is worth stressing that this specification assumes a separate efficiency distribution for each firm. Thus, conditionally upon the parameters describing the frontier, the $u_i$'s are posterior independent and are only updated by the $T$ observations for firm $i$, similarly as the $\alpha_i$'s in the SIE model.

**VED model**

The conditional posterior for the frontier parameters is identical to that given previously in Eqs. (A.5), (A.6), and (A.7). The conditional posterior for the inefficiencies takes the form:

$$p(\lambda|y, X, W, \alpha_0, \beta, \sigma^{-2}, \gamma) \propto f_N^n \left( u_i - \lambda_i \right) \left( \frac{\alpha_0}{\beta} - \frac{\sigma^2}{T} \zeta, \frac{\sigma^2}{T} I_n \right) \times I^+(u), \quad (A.10)$$

where $\zeta = (\exp(w_1'\gamma) \ldots \exp(w_n'\gamma))'$. In other words, the $u_i$'s are again independently Normally distributed, truncated to the positive orthant. The conditional distribution of any of the $\phi_h$'s ($\phi_h = \exp(\gamma_h), h = 1, \ldots, m$) depends only on $u$ and $\phi^{(-h)} = (\phi_1, \ldots, \phi_{h-1}, \phi_{h+1}, \ldots, \phi_m)'$ and is given by

$$p(\phi_h|W, u, \phi^{(-h)}) = f_G \left( \phi_h | a_h + \sum_{i=1}^{n} w_{ih} g_h + \sum_{i=1}^{n} w_{ih} u_i \prod_{j \neq h} \phi_j^{w_{ij}} \right). \quad (A.11)$$

With the Gibbs sampler we now have to deal with numerical integration in only $n + k + m + 2$ dimensions (423 in our empirical application).

**CED model**

The Gibbs sampler can be implemented by cyclical drawings from (A.5) and (A.8), where $\lambda^{-1} = \mu^{-1}t_n$, and from the full conditional for $\mu^{-1}$ which is

$$p(\mu^{-1}|y, X, u, \alpha_0, \beta, \sigma^{-2}) = p(\mu^{-1}|u) = f_G(\mu^{-1}|(n + 1), n\tilde{u} - \ln(r^*)). \quad (A.12)$$
References


Koop, G., J. Osiewalski, and M. Steel, 1994b, Hospital efficiency analysis with individual effects: A Bayesian approach, Center for Economic Research discussion paper no. 9447 (Tilburg University, Tilburg).


