Jacek Osiewalski, Jerzy Marzec, Katedra Ekonometrii i Badań Operacyjnych, Uniwersytet Ekonomiczny w Krakowie

Bayesian Analysis of Cost Efficiency With an Application to Bank Branches[1]

Jacek Osiewalski, Jerzy Marzec

Department of Econometrics
Academy of Economics
31-510 Kraków, Poland

ABSTRACT: In this paper we describe the use of the Bayesian statistical methodology for making inferences in composed error stochastic frontier models for panel data or individual cross-sections. We argue that one of modern Monte Carlo methods known as Gibbs sampling can greatly reduce the computational difficulties involved in analysing such models. We use the Bayesian approach to estimate a short-run cost frontier for N=58 branches of one of Polish commercial banks.

1. Introduction

The stochastic frontier or composed error framework was first introduced in Meeusen and van den Broeck (1977) and Aigner, Lovell and Schmidt (1977) and has been used in many empirical applications. In particular, stochastic frontier models have recently been used in many studies of production and cost efficiency in the banking sector; see Bauer, Hancock (1993) and Cebenoyan, Cooperman, Register and Hudgins (1993), Ferrier and Lovell (1990), Kaparakis, Miller and Noulas (1994), Mester (1993). All these empirical studies used the sampling-theory (classical) methods of inference.

Van den Broeck, Koop, Osiewalski and Steel (1994), hereafter BKOS, Koop, Steel and Osiewalski (1995), and Koop, Osiewalski and Steel, hereafter KOS (1994, 1997a, 1997b, 1997c) used Bayesian methods to analyse stochastic frontier models and argued that such methods had several advantages over their classical counterparts in the treatment of these models. Most importantly, the Bayesian methods enable to provide exact finite sample results for any feature of interest and to take fully into account parameter uncertainty. The Bayesian approach has been successfully applied in various empirical problems, ranging from hospital efficiencies [KOS (1997c)] to analyses of the growth of countries in KOS (1997a,b). In this paper we review the Bayesian stochastic frontier approach to

[1] This work was supported by the Polish Committee for Scientific Research (KBN; grant no. 1-H02B-

efficiency analysis and apply it to measurement of cost efficiency of bank branches.

2. Stochastic cost frontiers

The basic stochastic frontier long-run cost model assumes that a firm's observed total cost will deviate from the theoretical microeconomic cost function because of random noise and possible inefficiency. That is, for N firms in the sample (each observed over T periods),

$$\ln\left(\frac{TC_{ti}}{w_{ti,J}}\right) = \ln\left(C(\frac{w_{ti,1}}{w_{ti,J}},\frac{w_{ti,2}}{w_{ti,J}},K,\frac{w_{ti,J-1}}{w_{ti,J}},Q_{ti,1},K,Q_{ti,G};\beta)\right) - v_{ti} + z_{ti}, \qquad \textbf{(1)}$$

$$i = 1,K,N, \quad t = 1,K,T,$$

where TC_{ti} denotes observed total cost for firm i at time t, $w_{ti,1},...,w_{ti,J}$ are factor prices, $Q_{ti,1},...,Q_{t1,G}$ are observed quantities of G products of the analysed industry (in our case, the banking sector), v_{ti} is a symmetric disturbance capturing the random nature of the frontier itself (due to, e.g., measurement error); z_{ti} is a nonnegative disturbance capturing the level of inefficiency of firm i at time t, and β is a vector of unknown parameters of the cost function. Following the standard microeconomic theory of the firm, the cost frontier C(**.**) in (1) imposes homogeneity in factor prices; other regularity conditions (like monotonicity with respect to quantities and prices) should be ensured through parametric restrictions. These restrictions can be either global (for all possible levels of exogenous variables) or local (for the values observed in the sample). The most popular functional form of the cost frontier is translog, which is a second order local approximation and thus is called a (locally) flexible functional form, but simpler, non-flexible forms (like Cobb-Douglas) are still in use. An alternative functional form is based on the Müntz-Szatz expansion and is called the Asympotically Ideal Model (AIM) by Barnett, Geweke and Wolfe (1991). The AIM cost function is globally flexible and can be made locally regular by imposing appropriate restrictions, but its estimation is nontrivial, as shown by Barnett, Geweke and Wolfe (1991) and KOS (1994).

In cases where some inputs cannot be easily changed (are considered fixed) a short-run cost model is employed. If we assume that only one input is fixed, and we denote it by K, the following stochastic cost frontier model can be specified:

$$\ln\left(\frac{VC_{ti}}{w_{ti,J}}\right) = \ln\left(C_{Var}(K,\frac{w_{ti,2}}{w_{ti,J}},K,\frac{w_{ti,J-1}}{w_{ti,J}},Q_{ti,1},K,Q_{ti,G};\beta)\right) - v_{ti} + z_{ti}, \qquad \textbf{(2)}$$

$$i = 1,K,N, \quad t = 1,K,T,$$

where VC denotes total variable cost and K is the input of the first of J factors of production. Throughout the methodological part of the paper, we shall label the individual units as "firms", but the

015-11).

same models can, of course, be used in other contexts.

3. The Sampling Model and Prior Distribution

The basic stochastic frontier sampling model considered here can be written as:

$$y_{ti} = h(x_{ti}, \beta) + v_{ti} - z_{ti}, \qquad \textbf{(3)}$$

where y_{ti} is the negative of the natural logarithm of cost for firm i at time t (i=1,..,N; t=1,...,T); x_{ti} is a row vector of exogenous variables; h, a known measurable function, and β, a vector of k unknown parameters, define the deterministic part of the frontier; and v_{ti} and z_{ti} are random terms, one symmetric about zero and the other nonnegative. We assume that z_{ti} and v_{ti} are independent of each other and across firms and time. Thus, we do not allow for autocorrelated errors. Generally, efficiency will be measured as $r_{ti}=\exp(-z_{ti})$, which is an easily interpretable quantity in (0,1]. In the case of a cost function, z_{ti} captures the overall cost inefficiency, reflecting cost increases due to both technical and allocative inefficiency of the firm i at time t. For the translog cost function, Kumbhakar (1997) derives the exact relationship between allocative inefficiency in the cost share equations and in the cost function, which indicates that z_{ti}s in (3) cannot be independent of the exogenous variables and the parameters in the cost function. However, this independence assumption will usually be maintained as a crude approximation because it leads to simpler posterior analysis.

Note that our framework is suitable for panel data, but the case of just one cross-section is easily covered as it corresponds to T=1. Here we make the assumption that z_{ti} is independent (conditionally upon whatever parameters are necessary to describe its sampling distribution) over both i and t, as in KOS (1997a,b); see also Pitt and Lee (1981, Model II). KOS (1997c) follow an alternative modeling strategy and assume that the inefficiency level is an individual (firm) effect, i.e. $z_{ti}=z_i$ (t=1,...,T); see also Pitt and Lee (1981, Model I) and Schmidt and Sickles (1984).

In general, a fully parametric Bayesian analysis requires specifying

(i) a sampling distribution parameterized by a finite-dimensional vector (say, $\theta \in \Theta$),

(ii) a prior distribution for that θ.

In order to satisfy (i) and obtain the likelihood function we assume that v_{ti} is $N(0,\sigma^2)$, i.e. Normal with zero mean and constant variance σ^2, and z_{ti} is Exponential with mean (and standard deviation) λ_{ti} which can depend on some (say, m-1) exogenous variables explaining possible systematic differences in efficiency levels. In particular, we assume

$$\lambda_{ti} = \prod_{j=1}^{m} \phi_j^{-w_{tij}} \qquad \textbf{(4)}$$

where $\phi_j>0$ are unknown parameters and $w_{ti1}=1$. If m>1, the distribution of z_{ti} can differ for different t or i or both and thus in KOS(1997c) this case is called the Varying Efficiency Distribution (VED) model. If m=1, then $\lambda_{ti}=\phi_1^{-1}$ and all inefficiency terms constitute independent draws from the same distribution. This case is called the Common Efficiency Distribution (CED) model. Note that the sampling density of the observable y_{ti} given x_{ti}, $w_{ti}=(w_{ti1},...,w_{tim})$ and $\theta=(\beta', \sigma^{-2},\phi_1,...,\phi_m)'$ is a location mixture of Normals with the Exponential density of the inefficiency term as the mixing density:

$$p(y_{ti}|x_{ti}, w_{ti}, \theta) = \int_0^{\infty} f_N^1\left(y_{ti}|h(x_{ti},\beta) - z_{ti}, \sigma^2\right) f_G\left(z_{ti}|1, \prod_{j=1}^{m} \phi_j^{w_{tij}}\right) dz_{ti}, \qquad \textbf{(5)}$$

where $f_G(.|a,b)$ indicates the Gamma density with mean a/b and variance a/b^2, and a=1 corresponds to an Exponential distribution. Alternatively, the sampling density can be represented as

$$p(y_{ti}|x_{ti}, w_{ti}, \theta) = \lambda_{ti}^{-1} \exp[-\lambda_{ti}^{-1}(m_{ti} + \frac{1}{2}\sigma^2\lambda_{ti}^{-1})]\Phi(m_{ti}/\sigma), \qquad \textbf{(6)}$$

where

$$m_{ti} = h(x_{ti},\beta) - y_{ti} - \sigma^2/\lambda_{ti},$$

$\Phi(.)$ denotes the distribution function of N(0,1), and λ_{ti} is given by (4). See Greene (1990) and BKOS for similar expressions. The likelihood function, $L(\theta|data)$, is the product of the densities (6) over t and i. As a result of integrating out the inefficiency terms z_{ti}, the form of (6) is quite complicated indeed, and even the numerical evaluation of the ensueing likelihood function is nontrivial, as shown in BKOS.

An important aspect of any efficiency analysis is making inferences on individual efficiencies of observed firms. It is easy to show that, conditionally on the parameters and the data, the unobserved inefficiency term z_{ti} of an observed y_{ti} has a truncated Normal distribution with density

$$p(z_{ti}|y_{ti}, x_{ti}, w_{ti}, \theta) = [\Phi(m_{ti}/\sigma)]^{-1} f_N^1(z_{ti}|m_{ti}, \sigma^2) I(z_{ti} \geq 0), \qquad \textbf{(7)}$$

see Greene (1990) and BKOS. In (7), I(.) denotes the indicator function.

In principle, the prior distribution of θ can be any distribution, but it is usually preferred not to introduce too much subjective information about the parameters. Therefore, we use the following prior structure

$$p(\theta) = p(\sigma^{-2})p(\beta)p(\phi_1,...,\phi_m) \propto f_G(\sigma^{-2}|\frac{n_0}{2},\frac{s_0}{2})f(\beta)\prod_{j=1}^{m} f_G(\phi_j|1,g_j), \qquad \textbf{(8)}$$

which reflects lack of prior knowledge about the frontier parameters β, possibly except for regularity conditions imposed by economic theory. That is, we assume $f(\beta)\equiv 1$ if there are no regularity constraints, and, if such conditions are imposed, $f(\beta)=1$ for all β satisfying them and $f(\beta)=0$ for all other β. Alternatively, we could use a proper prior distribution on β, possibly truncated to the region of regularity. Typically, we shall choose the prior hyperparameters $n_0>0$ and $s_0>0$ so as to represent very weak prior information on the precision of the stochastic frontier. Note that we cannot take as the prior density for σ^{-2} the kernel of the limiting case where $s_0=0$, because this would result in the lack of existence of the posterior distribution [see Fernández, Osiewalski and Steel (1997)]. Thus, the use of the usual Jeffreys type prior for σ^{-2} (which corresponds to the Gamma kernel with $n_0=s_0=0$) is precluded, unless we put some panel structure on the inefficiency terms by assuming that, e.g., they are time-invariant individual effects as in KOS (1997c). For the m parameters of the efficiency distribution we take proper, independent Exponential priors in order to avoid the pathology described by Ritter (1993) and discussed in more general terms by Fernández, Osiewalski and Steel (1997). Following KOS (1997b, 1997c), we use $g_j=1$ for $j>1$ and $g_1=-\ln(r^*)$, where $r^*\in(0, 1)$ is the hyperparameter to be elicited. In the CED model ($m=1$), r^* can be interpreted as the prior median efficiency, because it is exactly the median of the marginal prior distribution of firm efficiency $r_{ti}=\exp(-z_{ti})$; see BKOS. In the VED case ($m>1$) our prior for $\phi=(\phi_1,...,\phi_m)'$ is quite noninformative and centered over the prior for the CED model. Note that the prior on ϕ, a parameter which is common to all firms, induces prior links between the firm-specific inefficiency terms.

4. Bayesian Inference Using Gibbs Sampling

The Bayesian approach combines all the information about the model parameters in their posterior density

$$p(\theta|\text{data}) \propto p(\theta)L(\theta|\text{data}). \qquad \textbf{(9)}$$

As this is a multivariate density, the crucial task of any applied Bayesian study is "to calculate relevant summaries of the posterior distribution, to express the posterior information in a usable form, and to serve as formal inferences if appropriate. It is in the task of summarizing that computation is typically needed." [O'Hagan (1994), p.205].

Using $g(\theta;\text{data})$ as a generic notation for any function of interest, we can represent our Bayesian

inference problem as the ratio of integrals

$$E[g(\theta;data)|data)] = \frac{\int_{\Theta} g(\theta;data)p(\theta)L(\theta|data)d\theta}{\int_{\Theta} p(\theta)L(\theta|data)d\theta}. \qquad \textbf{(10)}$$

The main numerical difficulty amounts to evaluating this ratio. In the case of the stochastic frontier model, the likelihood is too complex to analytically calculate any such posterior summary. Numerical integration methods are unavoidable. Most quantities of interest such as moments of the parameters or of functions of the parameters, probabilities of certain regions for the parameters, etc. can be expressed as expectations of some g(.;.) in (10). The integrating constant of the posterior exactly corresponds to the integral in the denominator of (10).

Note that traditional quadratures, like Cartesian product rules, are not helpful because they are feasible only when the dimension of θ, equal to k+m+1, is very small. However, our model is inherently "non-standard" and the dimension of θ is already 5 in the simplest case of the CED model (m=1) with a Cobb-Douglas frontier depending on $\beta=(\beta_1\ \beta_2\ \beta_3)'$ (k=3), and will be much higher for more complicated models. This effectively renders these type of numerical integration procedures quite useless. Thus we resort to a popular modern Monte Carlo integration method, known as Gibbs sampling.

Gibbs sampling is a technique for obtaining a sample from a joint distribution of a random vector α by taking random draws from only full conditional distributions. A detailed description of the technique can be found in e.g. Casella and George (1992), and Tierney (1994).

Suppose we are able to partition α into $(\alpha_1',...,\alpha_p')'$ in such a way that sampling from each of the conditional distributions (of α_i given the remaining subvectors; i=1,...,p) is relatively easy. Then the Gibbs sampler consists of drawing from these distributions in a cyclical way. That is, given the qth draw, $\alpha^{(q)}$, the next draw, $\alpha^{(q+1)}$, is obtained in the following pass through the sampler:

$\alpha_1^{(q+1)}$ is drawn from $p(\alpha_1|\alpha_2=\alpha_2^{(q)},...,\alpha_p=\alpha_p^{(q)})$,

$\alpha_2^{(q+1)}$ is drawn from $p(\alpha_2|\alpha_1=\alpha_1^{(q+1)},\alpha_3=\alpha_3^{(q)},...,\alpha_p=\alpha_p^{(q)})$,

...

$\alpha_p^{(q+1)}$ is drawn from $p(\alpha_p|\alpha_1=\alpha_1^{(q+1)},...,\alpha_{p-1}=\alpha_{p-1}^{(q+1)})$.

Note that each pass consists of p steps, i.e. drawings of the p subvectors of α. The starting point, $\alpha^{(0)}$, is arbitrary. Under certain general conditions [irreducibility and aperiodicity as described in e.g. Tierney (1994)], the distribution of $\alpha^{(q)}$ converges to the joint distribution, $p(\alpha)$, as q tends to infinity. Thus, we draw a sample directly from the joint distribution in an asymptotic sense. In practical applications we have to discard a (large) number of passes before convergence is reached. As the drawings in Gibbs sampling are (asymptotically) from the actual posterior distribution, which is properly normalized, there is no need to evaluate the integrating constant in the denominator of (10) separately. So we do not

require the evaluation of a ratio of integrals as in (10) in order to "normalize" our estimates of g(θ;data).

In order to efficiently use Gibbs sampling to make posterior inferences on both the parameters and firm efficiencies, we have to consider the joint posterior density of θ and z, p(θ,z|data), where z is a TNx1 vector of all the z_{ti}s. Now, instead of integrating out z, which was shown to lead to a very nonstandard likelihood function in Section 3, we shall consider θ given z and the data, which is quite easy to deal with. On the other hand, this implies that we also need to include z in the Gibbs sampler. Note that the dimension is then NT+k+m+1, greater than the number of observations. Despite this high dimensionality, the steps involved in the Gibbs are very easy to implement, and the resulting sampler is found in Koop, Steel and Osiewalski (1995) to have very good numerical properties, and to be far preferable to Monte Carlo with Importance Sampling in the particular application used. The conditional posterior density p(z|data,θ) is the product of the TN independent truncated Normal densities given by (7), so we can very easily draw z_{ti}s given the data and the parameters. These draws are immediately transformed into efficiency indicators defined as $r_{ti}=\exp(-z_{ti})$. Thus, this NT-dimensional step of each pass of our Gibbs sampler is quite simple. It is worth stressing at this stage that the Gibbs sampler, unlike the Monte Carlo approach outlined in BKOS, yields a draw of the whole vector z at each pass, and that for this reason, the efficiency measures for all N firms and all T periods are obtained as a by-product of our Gibbs sampling methodology.

However, the main difference with respect to Monte Carlo - Importance Sampling used in BKOS is in drawing θ. Now, the unwieldy form of the marginal posterior, p(θ|data), is not used at all, and we focus instead on the conditional posterior densities of subvectors of θ given the remaining parameters and z. Given z, the frontier parameters (β,σ^{-2}) are independent of φ and can be treated as the parameters of the (linear or nonlinear) Normal regression model in (3). Thus we obtain the following full conditionals for σ^{-2} and β:

$$p(\sigma^{-2}|\text{data},z,\beta) = f_G(\sigma^{-2}|\frac{n_0+TN}{2},\frac{1}{2}\{s_0+\sum_{t,i}[y_{ti}+z_{ti}-h(x_{ti},\beta)]^2\}), \tag{11}$$

$$p(\beta|\text{data},z,\sigma^{-2}) \propto f(\beta)\exp[-\frac{1}{2}\sigma^{-2}\sum_{t,i}(y_{ti}+z_{ti}-h(x_{ti},\beta))^2]. \tag{12}$$

The full conditional posterior densities of ϕ_j (j=1,...,m) have the general form

$$p(\phi_j|data, z, \phi_{(-j)}) \propto f_G(\phi_j|1+\sum_{t,i} w_{tij}, g_j) \exp(-\phi_1 \sum_{t,i} z_{ti} D_{ti1}), \quad \textbf{(13)}$$

where

$$D_{tir} = \prod_{j \neq r} \phi_j^{w_{tij}}$$

for r=1,...,m (D_{ti1}=1 when m=1) and $\phi_{(-j)}$ denotes ϕ without its jth element. Since w_{ti1}=1, the full conditional of ϕ_1 is just Gamma with parameters 1+NT and $g_1+z_{11}D_{111}+... +z_{TN}D_{TN1}$. Of course, the full conditionals for the z_{ti}'s are given by (7).

Depending on the form of the frontier and on the values of w_{tij}s for j>2, the full conditionals for β and for ϕ_j (j=2,...,m) can be quite easy or very difficult to draw from. Drawing from nonstandard conditional densities within the Gibbs sampler requires special techniques, like rejection methods or the Metropolis-Hastings algorithm [see e.g. Tierney (1994) or O'Hagan (1994)]. These procedures imply a substantial added complexity in the numerical integration and require additional input from the user. Therefore, we stress two important special cases where considerable simplifications are possible:

(i) linearity of the frontier,

(ii) 0-1 dummies for w_{tij} (j=2,...,m).

If $h(x_{ti},\beta)=x_{ti}\beta$ then (12) is a k-variate Normal density, possibly truncated due to regularity conditions. That is, we have

$$p(\beta|data, z, \sigma^{-2}) \propto f(\beta) f_N^k(\beta|\hat{\beta}, \sigma^2 (X'X)^{-1}), \quad \textbf{(14)}$$

where

$$\hat{\beta} = (X'X)^{-1} X'(y + z),$$

and y and X denote a NTx1 vector of y_{ti}s and a NTxk matrix with x_{ti}s as rows, respectively.

Cobb-Douglas or translog frontiers serve as examples of linearity in β; see Koop, Steel and Osiewalski (1995) and KOS (1997a,1997c).

The dichotomous (0-1) character of the variables explaining efficiency differences (w_{tij}; j=2,...,m) greatly simplifies (13) which simply becomes a Gamma density:

$$p(\phi_j|data, z, \phi_{(-j)}) = f_G(\phi_j|1+\sum_{t,i} w_{tij}, g_j + \sum_{t,i} w_{tij} z_{ti} D_{tij}), \quad \textbf{(15)}$$

From the purely numerical perspective, it pays to dichotomize these original variables in w_{ti} which are not 0-1 dummies.

The above discussion confirms that the stochastic frontier cost models considered in this paper

can be analyzed using Gibbs sampling methods. That is, even though the marginal posteriors of θ and the z_{ti}s are unwieldy, the conditionals for a suitable partition of the set of parameters should be much easier to work with. By taking a long enough sequence of successive draws from the conditional posterior densities, each conditional on previous draws from the other conditional densities, we can create a sample which can be treated as coming from the joint posterior distribution. The posterior expectation of any arbitrary function of interest, $g(\theta,z;data)$, can be approximated by its sample mean, g^*, based on M passes.

5. An Empirical Application: Cost Effciency of Bank Branches

We illustrate the Bayesian stochastic frontier analysis using the data from N=58 branches of one of Polish commercial banks. We use only cross-sectional data (T=1; first quarter of 1997) and thus the time subscript t will be omitted. Our model is

$$\ln\left(\frac{VC_i}{w_{i,Bor}}\right) = \alpha_0 + \alpha_1 \ln\left(\frac{w_{i,Com}}{w_{i,Bor}}\right) + \alpha_2 \ln\left(\frac{w_{i,Dep}}{w_{i,Bor}}\right) + \alpha_3 \ln\left(\frac{w_{i,Lab}}{w_{i,Bor}}\right) \qquad \textbf{(16)}$$
$$+\alpha_4 \ln(K_i) + \alpha_5 \ln(Q_i) + \alpha_6 \ln^2(Q_i) - v_i + z_i, \qquad i = 1,K\ ,N,$$

where the following notation is adopted:

VC = cost of labor (personnel expenses) + cost of computers, software and other goods and services purchased from outside suppliers + cost of financial capital,

w_{Lab} = price of labor = (personnel expenses)/(number of full-time equivalent employees),

w_{Dep} = price of deposits = (interest expense on deposits)/(volume of deposits),

w_{Bor} = price of money "borrowed" from other branches,

w_{Com} = average price of computers, software and other goods and services purchased by the bank from outside suppliers,

K = book value of buildings,

Q = aggregate volume of different loans + the excess of deposits over loans (if positive).

In the specification given above, we follow the microeconomic analysis of Sealey and Lindley (1977) who view the bank as using labor, physical capital, and deposits to produce earning assets. Thus we use deposits and other borrowed money as inputs (financial capital), and income generating money as the aggregate product of the bank. This approach (often called the intermediation approach) has been adopted in many empirical studies, using econometric as well as mathematical programming tools; see Akhaiven, Swamy, Taubman and Singamsetti (1997), Berger (1993), Berger and Humphrey (1991), Cebenoyan, Cooperman, Register and Hudgins (1993), Dietsch (1993), English, Grosskopf, Hayes and Yaiswarng (1993), Grabowski, Ragan and Rezvanian (1993), Hughes and Mester (1993),

Humphrey (1993), Kaparakis, Miller and Noulas (1994), Mester (1993), Muldur and Sassenou (1993).

As a consequence of the approach we follow, the total variable cost VC includes both operating and interest costs. Our product variable Q comprises loans to individuals, commercial and industrial loans, and the excess of deposits over loans (if positive). The latter component of our aggregate product reflects the fact that the branches operate within the bank, and their excessive deposits can be used by those branches which lack funds for potential loans. In fact, branches tend to specialize either in the acquisition of financial capital from depositors or in loaning funds. Branches from the first group ("depository" branches) provide funds, which are used by branches from the other group. These funds are provided at a constant price, fixed by the bank and only used to correct the calculation of the operating profit of a branch. Thus, for a "depository" branch, the volume of its excess funds can be treated as a product, because it increases the calculated profit of that branch. On the other hand, this money is an input for the branches which lack funds for loans; its price (w_{Bor}) is constant over branches. Also, the price of computers, software and other goods and services purchased by the bank from outside suppliers (w_{Com}) can be treated as constant over branches as main purchases are decided on the level of the bank which chooses the supplier (of e.g. hardware or software); this input price is negociated by the management of the bank and is the same for all branches.

In (16) we assume that some part of physical capital (represented by buildings) is a fixed input, which enters our cost frontier directly (and not through its price). Hence (16) is a short-run specification which tends to explain the variable cost (VC) of a branch. Since two input prices (w_{Com}) i (w_{Bor}) are constant over units (branches), the estimated form of our stochastic frontier model is

$$\ln(VC_i) = \beta_1 + \beta_2 \ln(w_{i,Dep}) + \beta_3 \ln(w_{i,Lab}) + \beta_4 \ln(K_i) + \beta_5 \ln(Q_i) + \beta_6 \ln^2(Q_i) - v_i + z_i, \quad i = 1,K\,,N, \qquad \textbf{(17)}$$

where:

$\beta_j = \alpha_j$ for j=2,…,6; $\beta_1 = \alpha_0 + \alpha_1 \ln(w_{Com}) + [1-(\alpha_1 + \alpha_2 + \alpha_3)]\ln(w_{Bor})$.

The functional form in (16) was suggested by Nerlove (1963) and used by Christensen and Greene (1976) in modeling costs of electric utility companies in the US in 1970. It is based on the Cobb-Douglas specification, but permits returns to scale to vary with Q. The returns to scale coefficient can be calculated from (17) as

$$RTS_i = \varepsilon_i^{-1} = \left(\beta_5 + 2\beta_6 \ln(Q_i)\right)^{-1} \qquad i = 1,K\,,N, \qquad \textbf{(18)}$$

and it relates to scale changes caused by proportional increases in variable inputs alone. We also tried the full translog form (with 15 free coefficients) instead of the specification in (17), but we faced serious identification problems due to the limited number of observations (N=58).

As regards the economic regularity conditions, homogeneity in factor prices is imposed in (16) and thus is assured in (17); monotonicity in prices would require α_1, α_2, α_3 >0 and $\alpha_1+\alpha_2+\alpha_3$ <1 in (16), which leads to β_2, β_3 >0 and $\beta_2+\beta_3$ <1 in (17). Given Q and factor prices, an increase in the amount of K,

the fixed input, should be compensated by a decrease in other inputs, leading to a smaller variable cost; thus we expect $\beta_4<0$. Obviously, we also require that $RTS_i>0$ for all i, but this condition is always met in practice. We impose all these regularity conditions through the prior density for $\beta= (\beta_1 \ldots\beta_6)'$. That is, we assume the prior structure in (8), where $f(\beta)$ is the indicator function of the restrictions on β.

The first model we estimate corresponds to the assumption that all branches have the same prior distribution of inefficiency, i.e. we have the CED specification (m=1, no exogenously modelled differences in cost efficiency). We take $r^*=0.7$ as the prior median of efficiency. As regards the prior hyperparameters for precision, we assume $n_0=s_0=10^{-6}$ which leads to a very diffuse Gamma distribution (with mean 1 and variance $2*10^6$) reflecting little prior knowledge about this parameter.

The final posterior results for our 66-dimensional vector of unknown parameters and efficiencies were obtained using $2*10^6$ Gibbs passes; the sampler converged quickly and was very stable, indicating that the results could have been based on a much shorter run. Table 1 presents the posterior means and standard deviations of the parameters of the cost frontier. As regards factor prices, the interest rate on deposits (i.e. the price of financial capital) exerts the strongest influence on variable cost. The role of the price of labor is less evident because of the much smaller posterior mean and the significantly larger standard deviation. We can indirectly estimate the sum of the elasticities with respect to the other two factor prices (which are constant over branches) as $1-(\beta_2+\beta_3)$; the posterior mean and standard deviation of this parameter is 0.205 and 0.110, respectively. Figure 1 shows the marginal posterior densities of the elasticities with respect to factor prices.

The regularity restrictions imposed through the prior distribution were binding for β_3 and, especially, for β_4. The role of the fixed factor (buildings) in explaining the variable cost is very small (if any). When β_4 was left unrestricted, its posterior mean was positive, but this had almost no consequence for our inference on other quantities of interest.

Our specification permits the elasticity of variable cost with respect to Q, ε_i, to vary with Q. Table 2 presents the posterior means and standard deviations of $\varepsilon_i = \beta_5 + 2\beta_6 \ln(Q_i)$ and of efficiency indicators $r_i = \exp(-z_i)$ for all branches ordered by increasing production. In addition, Table 2 provides the information whether a given branch had more deposits than loans ($w_{i,2}=0$) or not.

Our returns to scale measure (18), evaluated at the posterior means of β_5 and β_6, is plotted in Figure 2 against the values of our aggregate product in the range covering the sample points. 1% increase of variable cost is associated with about 1.23 % increase of production for the smallest branch, 1.08 % increase of production for the two medium branches, and about 0.98 % increase for the largest branch. For almost all branches, our estimates of RTS_i are greater than one.

We assumed that, a priori, there is about 50% chance that cost efficiency of any given branch is below 0.7. Our data set points at much higher efficiency and leads to the average posterior mean of r_i equal 0.924 with 0.044 as the average posterior standard deviation. However, the individual posterior

means are quite spread, ranging from 0.694 to 0.972. Figure 3 presents the marginal posterior densities $p(r_i|data)$ for the branches with the maximum, minimum and average posterior means of r_i. The densities are a bit diffuse as our inference is based an individual cross-section and, in a sense, we have only one observation for each estimated efficiency. This shows one of advantages of our Bayesian approach which gives us a direct and easily understandable evaluation of uncertainty in terms of marginal posterior densities. Using just one cross section also means that our inference on individual efficiency levels is somewhat sensitive to our prior assumptions. Taking $r^*=0.95$ (instead of 0.70) leads to much more uniformly spread posterior means, ranging from 0.809 to 0.975, but with almost the same average (0.933) and the same ranking of branches as for $r^*=0.70$.

A closer look at Table 2 reveals that "depository" branches tend to be more efficient than the others. This suggests that $w_{i,2}$ may be an important variable for explaining systematic differences in cost efficiency. We used the methodology outlined in Sections 3 and 4 to estimate the VED specification with m=2, $w_{i,1}\equiv 1$ and $w_{i,2}$ as in Table 2. The most probable values of the parameter capturing systematic differences in efficiency, ϕ_2 (see (4)), lie below 1. The posterior mean and standard deviation of $\ln(\phi_2)$ is -1.17 and 0.32, respectively, indicating that "depository" branches are significantly more efficient. The results for our VED case are very similar to the ones presented above for the CED model, except that the VED specification shows more variability in efficiency levels and much less in the RTS_i's. Our empirical conclusions should be treated with caution as we base them on only one cross section of 58 units. We expect to gain insight into the complicated issues of measuring cost efficiency of bank branches by using panel data and a more sophisticated frontier model.

6. Conclusion

In this paper, we have reviewed a Bayesian analysis of stochastic frontier models, arguing that Gibbs sampling can be used to greatly reduce the computational burden inherent to this analysis. Following KOS (1994, 1997a, 1997b, 1997c), we have shown how the posterior conditional densities can be used to set up a Gibbs sampler. The structure of the Gibbs sampler follows naturally from viewing the inefficiency terms as additional parameters in a regression model, see Fernández, Osiewalski and Steel (1997). In important special cases all conditionals are either truncated Normal, Normal or Gamma distributions which leads to enormous computational gains.

We have applied the Bayesian methodology to make posterior inference on the technology and cost efficiency of 58 branches of one of Polish banks. Our preliminary results, based on one cross-section and a relatively simple cost frontier, indicate increasing returns to scale (varying with the branch output level) and systematic differences in efficiency, explained by the relation between the volume of

loans and the financial capital input.

The statistical rigour and empirical flexibility of our Bayesian approach encourage us to consider a much more complicated cost frontier model and to estimate it using panel data.

References

Aigner, D., C.A.K. Lovell, and P. Schmidt (1977). "Formulation and estimation of stochastic frontier productions function models," **Journal of Econometrics**, 6.

Akhaiven J., P.A.V.B. Swamy, S.B. Taubman, R.N. Singamsetti (1997), "A general method of deriving the inefficiencies of bank from a profit function", **Journal of Productivity Analysis**, 8.

Barnett W., J. Geweke, M. Wolfe (1991), "Seminonparametric Bayesian Estimation of the Asymptotically Ideal Production Model" **Journal of Econometrics**, 49.

Bauer W.B., D. Hancock (1993), "The efficiency of the Federal Reserve in providing check processing services", **Journal of Banking and Finance**, 17.

Berger A. N. (1993), "'Distribution-free' estimates of efficiency in the U.S. banking industry and tests of the standard distributional assumptions", **Journal of Productivity Analysis**, 4.

Berger A. N., D. Hancock, D.B. Humphrey (1993), "Bank efficiency derived from the profit function", **Journal of Banking and Finance**, 17.

Berger A. N., G. Hanweck, D. Humphrey (1987), "Competitive viability in banking. Scale, scope, and mix economies", **Journal of Monetary Economics**, 20.

Berger A. N., D.B. Humphrey (1991), "The dominace of inefficiencies over scale and product mix economies in banking", **Journal of Monetary Economics**, 28.

van den Broeck, J., G. Koop, J. Osiewalski, and M. Steel (1994). "Stochastic frontier models: A Bayesian perspective," **Journal of Econometrics**, 61.

Casella, G. and E. George (1992). "Explaining the Gibbs sampler," **The American Statistician**, 46.

Cebenoyan A.S., E.S. Cooperman, C.A. Register, S.C. Hudgins (1993), "The relative efficieny of stock versus Mutual S&Ls: A stochastic cost frontier approach", **Journal of Financial Services Research**.

Christensen L.R., W.H. Greene (1976), Economics of scale in U.S. electric power generation, **Journal of Political Economy**, 84.

Dietsch M. (1993), "Economies of scale and scope in French commercial banking industry", **Journal of Productivity Analysis**, 4.

English M., S. Grosskopf, K. Hayes., S. Yaiswarng (1993), "Output allocative and technical efficiency of banks", **Journal of Banking and Finance**, 17.

Fernández, C., J. Osiewalski and M.F.J. Steel (1997). "On the use of panel data in stochastic frontier models with improper priors," **Journal of Econometrics**, 79.

Ferrier G.D., C.A.K. Lovell (1990), "Measuring cost efficiency in banking: econometric and linear programming evidence", **Journal of Econometrics**, 46.

Grabowski R., N. Ragan, R. Rezvanian (1993), "Organizational forms in banking: An empirical investigation of cost efficiency", **Journal of Banking and Finance**, 17.

Greene, W.H. (1990). "A gamma-distributed stochastic frontier model," **Journal of Econometrics**, 46.

Hughes J., L.J. Mester (1993), "A quality and risk-adjusted cost function for banks: Evidence on the

"too-big-to-fail" doctrine", **Journal of Productivity Analysis**, 4.

Humphrey D. (1985), "Costs and scale economies in bank intermediation", w: **Handbook of Banking Strategy** (red.: R.C. Aspinwall, R. Eisenbeis), J. Wiley, New York.

Humphrey D. (1993), "Cost and technical change: Effects from bank deregulation", **Journal of Productivity Analysis**, 4.

Kaparakis E., S. M. Miller, A.G. Noulas, (1994), "Short-run cost inefficiency of commercial banks: A flexible stochastic frontier approach", **Journal of Money, Credit, and Banking**, 26.

Koop, G, J. Osiewalski and M.F.J. Steel (1997a). "The components of output growth: A stochastic frontier analysis," manuscript.

Koop, G, J. Osiewalski and M.F.J. Steel (1997b). "Measuring the sources of output growth in a panel of countries," manuscript.

Koop, G., J. Osiewalski and M.F.J. Steel (1994). "Bayesian efficiency analysis with a flexible form: The AIM cost function," **Journal of Business and Economic Statistics**, 12.

Koop, G., J. Osiewalski and M.F.J. Steel (1997c). "Bayesian efficiency analysis through individual effects: Hospital cost frontiers," **Journal of Econometrics**, 76.

Koop, G., M.F.J. Steel and J. Osiewalski (1995). "Posterior analysis of stochastic frontier models using Gibbs Sampling," **Computational Statistics**, 10.

Kumbhakar, S.C. (1997). "Modeling allocative inefficiency in a translog cost function and cost share equations: An exact relationship," **Journal of Econometrics**, 76.

Meeusen, W. and J. van den Broeck (1977). "Efficiency estimation from Cobb-Douglas production functions with composed error," **International Economic Review**, 8.

Mester L.J. (1987), "A multiproduct cost study of savings and loans", **Journal of Finance**, 42.

Mester L.J. (1993), "Efficiency in the savings and loan industry", **Journal of Banking and Finance**, 17.

Muldur U., Sassenou M. (1993), "Economies of scale and scope in French banking and savings insitutions", **Journal of Productivity Analysis**, 4.

Nerlove, M. (1963), "Return to scale in electricity supply", in. C.F Christ, ed. **Measurement in Economics** (Stanford University Press, Stanford, CA).

Noulas A.G., C.R. Subhash, S.M. Miller (1990), "Returns to scale and input substitution for large U.S. banks", **Journal of Money, Credit, and Banking**, 22.

O'Hagan A. (1994), **Bayesian Inference**, Edward Arnold, London.

Pitt M.M., L.F. Lee (1981), "The measurement and sources of technical inefficiency in the Indonesian weaving industry," **Journal of Development Economics**, 9.

Ritter C. (1993), "The Normal-Gamma frontier model under a common vague prior does not produce a proper posterior," manuscript.

Schmidt P., R.C. Sickles (1984), "Production frontiers and panel data," **Journal of Business and Economic Statistics**, 2.

Sealey C.W., J.T. Lindley (1977), "Inputs, outputs, and a theory of production and cost at depository financial institutions", **Journal of Finance**, 32.

Tierney, L. (1994). "Markov chains for exploring posterior distributions" (with discussion), **Annals of Statistics**, 22.

Table 1. The posterior means and standards deviation of the parameters of model (17).

Parameter	E(•|data)	D(•|data)
β_1	4.034	8.713
β_2	0.632	0.062
β_3	0.163	0.103
β_4	-0.006	0.005
β_5	-0.057	0.946
β_6	0.027	0.026
$Var(v_i)= \sigma^2$	0.0103	0.0044
$E(z_i)=\lambda_i=\phi_1^{-1}$	0.121	0.031

Table 2. The posterior means and standards deviation of branch-specific characteristics.

i	w_{i2}	$E(\varepsilon_i \mid data)$	$D(\varepsilon_i \mid data)$	$E(r_i \mid data)$	$D(r_i \mid data)$
1	1	0.813	0.113	0.794	0.114
2	1	0.824	0.103	0.894	0.081
3	1	0.862	0.068	0.958	0.038
4	1	0.863	0.067	0.811	0.089
5	0	0.869	0.062	0.939	0.048
6	0	0.872	0.059	0.945	0.045
7	0	0.875	0.057	0.942	0.047
8	0	0.877	0.055	0.937	0.050
9	1	0.877	0.055	0.790	0.090
10	0	0.883	0.050	0.900	0.066
11	0	0.884	0.049	0.935	0.050
12	1	0.892	0.043	0.843	0.081
13	1	0.892	0.043	0.970	0.029
14	0	0.893	0.042	0.944	0.045
15	0	0.894	0.042	0.953	0.040
16	0	0.898	0.038	0.906	0.063
17	0	0.899	0.038	0.940	0.047
18	1	0.899	0.037	0.891	0.070
19	0	0.902	0.036	0.948	0.043
20	1	0.908	0.032	0.942	0.046
21	0	0.912	0.030	0.906	0.063
22	0	0.913	0.029	0.949	0.042
23	0	0.916	0.028	0.945	0.045
24	1	0.916	0.028	0.834	0.082
25	1	0.918	0.028	0.904	0.064
26	0	0.918	0.027	0.945	0.044
27	1	0.922	0.026	0.905	0.064
28	1	0.922	0.026	0.694	0.090
29	1	0.923	0.026	0.799	0.087
30	1	0.923	0.026	0.895	0.066
31	1	0.924	0.026	0.852	0.078
32	0	0.925	0.026	0.948	0.043

i	w_{i2}	$E(\varepsilon_i \mid data)$	$D(\varepsilon_i \mid data)$	$E(r_i \mid data)$	$D(r_i \mid data)$
33	0	0.926	0.026	0.936	0.049
34	0	0.929	0.026	0.917	0.058
35	0	0.936	0.027	0.943	0.046
36	1	0.936	0.027	0.820	0.085
37	1	0.936	0.027	0.731	0.090
38	0	0.938	0.028	0.926	0.054
39	1	0.940	0.029	0.779	0.089
40	0	0.944	0.030	0.940	0.048
41	0	0.948	0.032	0.952	0.041
42	0	0.949	0.033	0.916	0.059
43	1	0.951	0.034	0.877	0.074
44	0	0.954	0.036	0.894	0.068
45	0	0.955	0.037	0.918	0.058
46	0	0.960	0.040	0.832	0.088
47	0	0.961	0.042	0.925	0.055
48	1	0.962	0.042	0.860	0.078
49	0	0.967	0.046	0.945	0.045
50	0	0.970	0.048	0.917	0.059
51	1	0.975	0.052	0.764	0.091
52	0	0.977	0.054	0.972	0.027
53	0	0.979	0.056	0.927	0.055
54	0	0.985	0.061	0.929	0.054
55	1	0.988	0.063	0.932	0.052
56	0	0.997	0.072	0.925	0.057
57	0	1.009	0.083	0.954	0.041
58	1	1.019	0.092	0.815	0.098

Figure 1. The marginal posterior densities of the price elasticities.

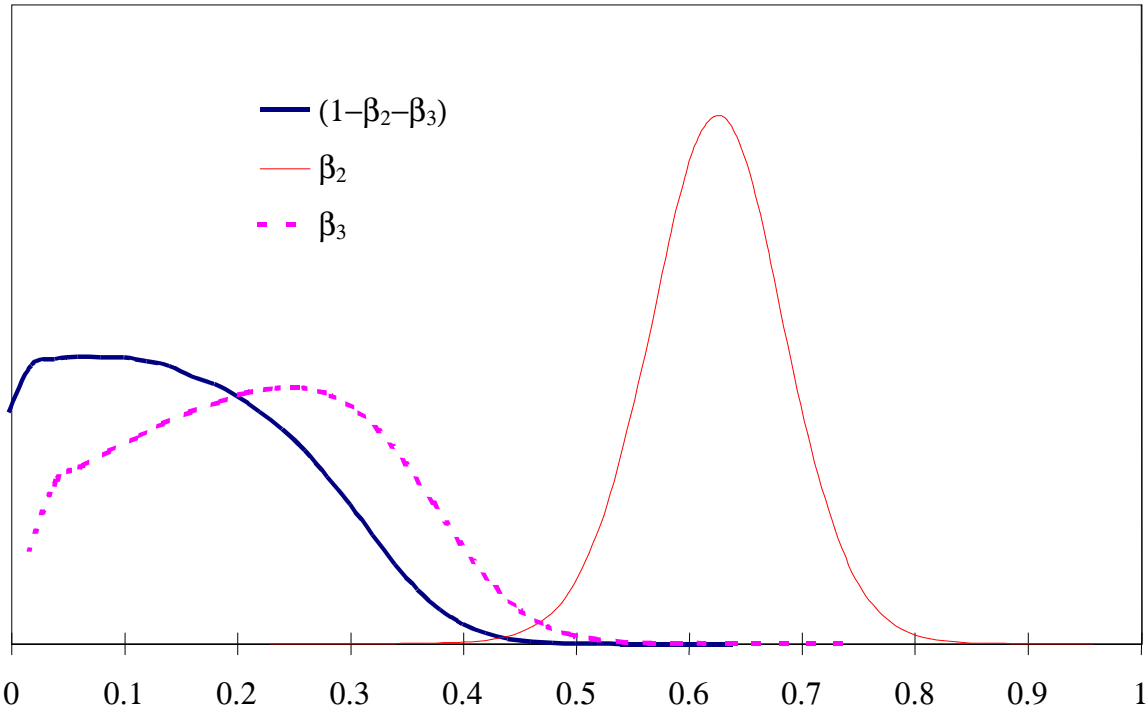
(1−β2−β3)
β2
β3
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

Figure 2. The returns to scale estimate as a fuction of the output level.

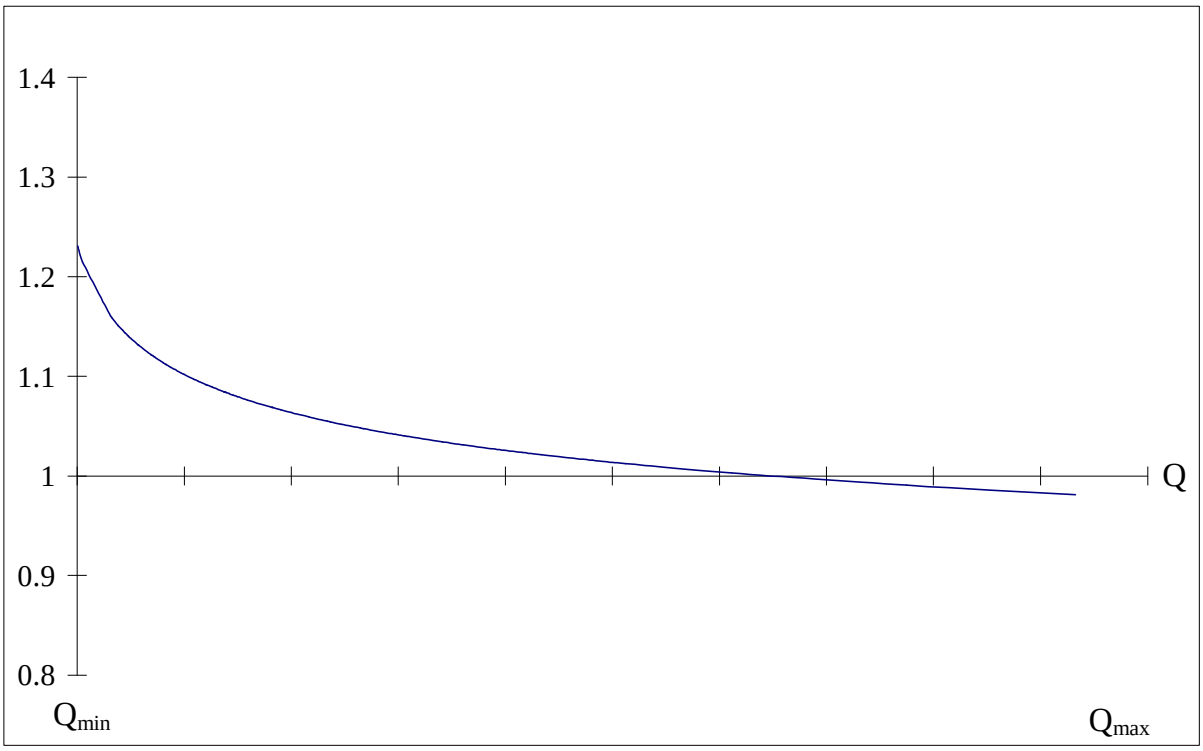

Figure 3. The marginal posterior densities of cost efficiency r_i.