Quantile Stochastic Frontiers

Mike G. Tsionas*

Abstract

Following the recent literature we propose a novel quantile Stochastic Frontier Model (SFM) and develop Markov Chain Monte Carlo techniques for numerical Bayesian inference. In an empirical application to US large banks we document important differences between the Quantile and the traditional SFM, in terms of several aspects of the data. We also document considerable heterogeneity among different quantiles in terms of returns to scale, technical change, efficiency change, technical efficiency, as well as productivity growth.

Keywords: Productivity and Competitiveness; Efficiency; Quantile Stochastic Frontier Model; Bayesian Inference.

Acknowledgment: The author is grateful to three anonymous reviewers for many useful remarks on an earlier version.

*Lancaster University Management School, LA1 4YX, UK, m.tsionas@lancaster.ac.uk

1 Introduction

Jradi and Ruggiero (2019) and Jradi, Parmeter and Ruggiero (2019a,b) propose a Stochastic Data Envelopment Analysis (SDEA) problem, and they introduce the concept of the most likely quantile. As they mention: "The usefulness of the SDEA estimator depends on the ability to choose the appropriate τ [quantile]. The model can be solved for any quantile but additional information is needed to determine which quantile is most likely." Their approach is based on the Afriat inequalities so that the functional form of the frontier is left unspecified. Of course, the approach relies on a known distribution for the composed error term, say a normal/half-normal convolution. Quantile regression models are based on conditional quantiles (including the median) as a function of certain covariates and, in this sense, they are more flexible than least squares which focuses exclusively on the conditional mean. Certainly, although less general than nonparametric regression, it is an interesting alternative as economic restrictions can be imposed easily, the method handles outliers in a transparent manner, and heterogeneity is allowed as we focus attention on *modeling the entire distribution* of the dependent variables given the covariates.

Previous work includes Wang and Wang (2013), who applied shape-restricted support vector regression with pinball loss, Aragon, Daouia, and Thomas-Agnan (2005), Daouia and Simar (2007) and Martins-Filho and Yao (2008), who proposed the use of conditional quantiles of random output. An alternative is provided by Wang et al. (2014) in which a two-step approach is proposed: first, identifying fitted values that minimize asymmetric absolute loss under shape restrictions; and second, constructing an estimator that links these fitted values. Of course, quantiles and quantile estimation have a long history of use in applied studies (Rychlik, 2017; Ye et al., 2017; Batur et al., 2018; Chang, 2015, Ng et al., 2017; Batur and Choobineh, 2010; Chen and Kelton, 2006; Taylor, 2007, and Somers and Whittaker, 2007).

As Behr (2010) noted in the context of banking: "[T]he quantile regression approach allows efficient or almost efficient banks to apply production relations which may differ strongly from average or low efficiency banks. Indeed, the empirical regression results hint for strong differences in regression parameters at different quantiles, that is for less or more efficient banks. Additionally, conditional quantile regressions are very robust compared to conditional mean regressions against outliers."

In this paper, we consider explicitly quantile regression in the context of stochastic frontier modeling. Specifically, for each quantile of the conditional distribution of the dependent variable (output, say) we allow for the possibility that there is quantile-specific inefficiency. In particular, the quantile-specific model consists of a specification that relates inputs to output but there is a quantile-specific one-sided error term as well, which accounts for technical inefficiency. Therefore, the quantile-specific model is a convolution of asymmetric Laplace-half-normal errors rather than a model based exclusively on the asymmetric Laplace model. This provides more flexibility and, at the same time, it allows for heterogeneity as the technology is allowed to be different across different quantiles. The model unifies the problem of selecting the most likely quantile of Jradi and Ruggiero (2019) and estimation of frontier parameters, as well as technical inefficiency in a common statistical framework. Moreover, as Jradi and Ruggiero (2019) suggest, we impose the no-quantile-crossing property of Wang et al. (2014). As the problem of the functional form remains, we propose the use of flexible and globally concave Symmetric

Generalized McFadden form (Kumbhakar, 1994). Moreover, regarding the Bayesian approach, "The use of Bayesian inference in generalized linear and additive models is quite standard these days. The relative ease with which Markov chain Monte Carlo (MCMC) methods may be used for obtaining the posterior distributions, even in complex situations, has made Bayesian inference very useful and attractive. Unlike conventional methods, Bayesian inference provides one with the entire posterior distribution of the parameter of interest. In addition, it allows for parameter uncertainty to be taken into account when making predictions." (Yu and Moheed, 2001, p. 438). Additionally, by treating the quantile as a parameter we have access to its posterior distribution and, therefore, the most likely quantile can be estimated using the posterior mean, median or mode. Of course, access to the complete set of quantile estimates is retained (for fixed quantile q ranging from, say, 0.1 to 0.9 with step 0.1 or lower).

2 Model

Quantile regression was introduced by Koenker and Bassett (1978) and is based on the idea that given the linear model:

$$y_i = x_i'\beta + v_i, i = 1, ..., n, \tag{1}$$

then the qth regression quantile ($0 < q < 1$) is any solution to the linear programming problem:

$$\min_{\beta} \; q \sum_{\{i|y_i - x_i'\beta \geq 0\}} |y_i - x_i'\beta| + (1-q) \sum_{\{i|y_i - x_i'\beta < 0\}} |y_i - x_i'\beta|. \tag{2}$$

Here, x_i is a $k \times 1$ vector of regressors, and β is a $k \times 1$ parameter vector. We denote observations of the dependent variable as $y = [y_1, ..., y_n]'$, $X = [x_1', ..., x_n']'$. In this paper, we extend the model to allow for efficiency estimation:

$$y_i = x_i'\beta + v_i - u_i, i = 1, ..., n, \tag{3}$$

where u_i is a non-negative error component representing technical inefficiency. We denote $u = [u_1, ..., u_n]'$. Roughly, what we want to do is represent a stochastic frontier model in terms of quantiles of the distribution of the dependent variable:

$$y_i = x_i'\beta(q) + v_i(q) - u_i(q), i = 1, ..., n, \tag{4}$$

with some abuse of notation to indicate that parameters β, noise and inefficiency differ by quantile. This is a general form of quantile regression and allows parameters and inefficiency to vary by quantile without necessarily using the concept of most likely quantile. This is done in the interest of users who are interested in the behavior of parameters, inefficiency, productivity growth, etc., for all quantiles of the distribution of the dependent variable.

Quantile regression can be given a statistical interpretation if we consider the asymmetric Laplace density:

$$p(v) \propto \tau^{-1} \exp\left\{-\tau^{-1}|v| \left[qI_{[0,\infty)}(v) + (1-q)I_{(-\infty,0]}(v)\right]\right\}. \tag{5}$$

The likelihood function is:

$$L_q(\beta, \tau; y, X) \propto \tau^{-n} \exp\left\{-\tau^{-1} \sum_{i=1}^{n} |y_i - x_i'\beta| \left[qI_{[0,\infty)}(y_i - x_i'\beta) + (1-q)I_{(-\infty,0]}(y_i - x_i'\beta)\right]\right\}. \tag{6}$$

Next we consider the normal mixture representation of the Laplace distribution. Specifically, we consider the following probability density for the disturbance, with the new scale parameter $\sigma_v = (2\tau)^{1/2}$:

$$p(v_i|w_i) \propto \left(\sigma_v^2 w_i\right)^{-1/2} \exp\left\{-\frac{v_i^2}{2\sigma_v^2 w_i}\left[qI_{[0,\infty)}(v_i) + (1-q)I_{(-\infty,0]}(v_i)\right]\right\}, \tag{7}$$

where w_i follows a standard exponential distribution, viz. $p(w_i) = \exp(-w_i)$. For Bayesian inference in quantile regression, see Tsionas (2003), Yang and He (2010) and Yang et al. (2016).

We assume that, independently of v_i and x_i, u_i follows a half-normal distribution, viz. $u_i \sim \mathcal{N}_+(0, \sigma_u^2)$. The composed error, $\varepsilon_i = v_i - u_i$ has density:

$$p(\varepsilon_i) \propto \tau^{-1}\sigma_u^{-1} \int_0^\infty \exp\left\{-\tau^{-1}|\varepsilon_i| \left[qI_{[0,\infty)}(\varepsilon_i) + (1-q)I_{(-\infty,0]}(\varepsilon_i)\right] - \tfrac{1}{2\sigma_u^2}u_i^2\right\} du_i. \tag{8}$$

Based on the likelihood, we have the following posterior:

$$\begin{gathered} p_q(\beta, \tau, \sigma_u|y, X) \propto \tau^{-n}\sigma_u^{-n} \cdot p(\beta, \tau, \sigma_u)\cdot \\ \int_0^\infty \exp\left\{-\tau^{-1} \textstyle\sum_{i=1}^n |y_i - x_i'\beta + u_i| \left[qI_{[0,\infty)}(y_i - x_i'\beta + u_i) + (1-q)I_{(-\infty,0]}(y_i - x_i'\beta + u_i)\right] - \frac{1}{2\sigma_u^2}\sum_{i=1}^n u_i^2\right\} du_i, \end{gathered} \tag{9}$$

where $p(\beta, \tau, \sigma_u)$ is the prior. Alternatively, we consider the augmented posterior:

$$\begin{gathered} p_q(\beta, \tau, \sigma_u, \{u_i\}_{i=1}^n|y, X) \propto \tau^{-n}\sigma_u^{-n} \cdot p(\beta, \tau, \sigma_u)\cdot \\ \exp\left\{-\tau^{-1} \textstyle\sum_{i=1}^n |y_i - x_i'\beta + u_i| \left[qI_{[0,\infty)}(y_i - x_i'\beta + u_i) + (1-q)I_{(-\infty,0]}(y_i - x_i'\beta + u_i)\right] - \frac{1}{2\sigma_u^2}\sum_{i=1}^n u_i^2\right\}. \end{gathered} \tag{10}$$

Notice that since q is considered as a parameter, MCMC draws become available and, therefore, we can have access to the marginal posterior density of q. As a result, the mean or mode of the marginal posterior distribution of q provides direct access to the most likely quantile.

Based on (7) we can consider an alternative augmented posterior where the weights $\{w_i\}_{i=1}^n$ are treated as parameters:

$$p_q(\beta,\sigma_v,\sigma_u,\{u_i\}_{i=1}^n,\{w_i\}_{i=1}^n|y,X)\propto\sigma_v^{-n}\sigma_u^{-n}\cdot p(\beta,\sigma_v,\sigma_u)\cdot$$
$$\exp\left\{-\tfrac{1}{2\sigma_v^2}\sum_{i=1}^n\tfrac{(y_i-x_i'\beta+u_i)^2}{w_i}\left[qI_{[0,\infty)}(y_i-x_i'\beta+u_i)+(1-q)I_{(-\infty,0]}(y_i-x_i'\beta+u_i)\right]-\tfrac{1}{2\sigma_u^2}\sum_{i=1}^n u_i^2-\sum_{i=1}^n w_i\right\}. \tag{11}$$

Our prior is as follows:

$$\begin{aligned} p(\beta|\sigma_v,\sigma_u)&\propto 1,\\ p(\sigma_v)&\propto\sigma_v^{(\underline{n}_v+1)}\exp\left(-\tfrac{\underline{q}_v}{2\sigma_v^2}\right)\quad,\\ p(\sigma_u)&\propto\sigma_u^{(\underline{n}_u+1)}\exp\left(-\tfrac{\underline{q}_u}{2\sigma_u^2}\right), \end{aligned} \tag{12}$$

where $\underline{n}_v,\underline{n}_u,\underline{q}_v,\underline{q}_u\geq 0$ are prior parameters. The priors are conditionally conjugate and the prior for β is flat. In our empirical work, we set $\underline{n}_v,\underline{n}_u=1,\underline{q}_v,\underline{q}_u=10^{-4}$ which corresponds to an almost flat prior. Finally, assuming that we have a flat prior on q, viz.

$$p(q)\propto 1,\ 0\leq q\leq 1, \tag{13}$$

it is easy to see from (11) that we have the following conditional posterior:

$$p(q|\beta,\tau,\sigma_u,\{u_i,w_i\}_{i=1}^n,y,X)\propto\exp\left(Dq\right),\ 0\leq q\leq 1, \tag{14}$$

where $D=-\tau^{-1}\sum_{i=1}^n\left\{I_{[0,\infty)}(y_i-x_i'\beta+u_i)-I_{(-\infty,0]}(y_i-x_i'\beta+u_i)\right\}$. Typical conditional densities are shown in Figure 1 to motivate our choice for sampling in section 3.2.

It should be noted that treating q as a parameter in (13) and (14), allows the most likely quantile (say q^*) to be found using the median, mode or mean of its posterior distribution via standard numerical methods based on MCMC.

3 Numerical Bayesian inference

3.1 General

We use MCMC methods of inference, and especially the Gibbs sampler with data augmentation. The Gibbs sampler provides access to the posterior by drawing from the posterior conditional distributions.

Based on (11) we derive the following posterior conditional distributions:

$$p_q(\sigma_v|\beta,\sigma_u,\{u_i\}_{i=1}^n,\{w_i\}_{i=1}^n,y,X)\propto\sigma_v^{-(n+\underline{n}_v+1)}\exp\left\{-\frac{Q_v}{2\sigma_v^2}\right\}, \tag{15}$$

where $Q_v=\underline{q}_v+\sum_{i=1}^n\tfrac{(y_i-x_i'\beta+u_i)^2}{w_i}\left[qI_{[0,\infty)}(y_i-x_i'\beta+u_i)+(1-q)I_{(-\infty,0]}(y_i-x_i'\beta+u_i)\right]$,

$$p_q(\sigma_u|\beta,\sigma_v,\{u_i\}_{i=1}^n,\{w_i\}_{i=1}^n,y,X)\propto\sigma_u^{-(n+\underline{n}_u+1)}\exp\left\{-\frac{Q_u}{2\sigma_u^2}\right\}, \tag{16}$$

where $Q_u = \underline{q}_u + \sum_{i=1}^n u_i^2$,

$$p_q(w_i|\beta, \sigma_v, \sigma_u, \{u_i\}_{i=1}^n, y, X) \propto \exp\left\{-w_i - A_i w_i^{-1}\right\}, w_i > 0, \tag{17}$$

where $A_i = \frac{1}{2\sigma_v^2}(y_i - x_i'\beta + u_i)^2 \left[qI_{[0,\infty)}(y_i - x_i'\beta + u_i) + (1-q)I_{(-\infty,0]}(y_i - x_i'\beta + u_i)\right]$,

$$p_q(u_i|\beta, \sigma_v, \sigma_u, \{w_i\}_{i=1}^n, ; y, X) \propto \exp\left\{-\frac{(r_i + u_i)^2}{2\sigma_v^2 w_i}\left[qI_{[0,\infty)}(r_i + u_i) + (1-q)I_{(-\infty,0]}(r_i + u_i)\right] - \frac{1}{2\sigma_u^2}u_i^2\right\}, u_i \geq 0, \tag{18}$$

where $r_i = y_i - x_i'\beta$.

The conditional posterior distribution of β is given by:

$$\begin{gathered} p_q(\beta|\sigma_v, \sigma_u, \{u_i\}_{i=1}^n, \{w_i\}_{i=1}^n, y, X) \propto \\ \exp\left\{-\frac{1}{2\sigma_v^2}\sum_{i=1}^n \frac{(y_i - x_i'\beta + u_i)^2}{w_i}\left[qI_{[0,\infty)}(y_i - x_i'\beta + u_i) + (1-q)I_{(-\infty,0]}(y_i - x_i'\beta + u_i)\right]\right\}. \end{gathered} \tag{19}$$

Therefore,

$$\begin{gathered} p_q(\beta|\sigma_v, \sigma_u, \{u_i\}_{i=1}^n, \{w_i\}_{i=1}^n, y, X) \propto \\ \exp\left\{-\frac{1}{2\sigma_v^2}(y + u - X\beta)' W^{-1}(y + u - X\beta)\left[q^{n_o(\beta,u)} + (1-q)^{n-n_o(\beta,u)}\right]\right\}, \end{gathered} \tag{20}$$

where $n_o(\beta, u) = \sum_{i=1}^n I_{[0,\infty)}(y_i - x_i'\beta + u_i)$, $W = diag[w_1, ..., w_n]$

The distribution in (20) resembles a multivariate normal, since when $q = 1$ (or zero) we have:

$$\beta|\sigma_v, \sigma_u, \{u_i\}_{i=1}^n, \{w_i\}_{i=1}^n, y, X \sim \mathcal{N}_k\left(\hat{\beta}, V\right), \tag{21}$$

where $\hat{\beta} = (X'W^{-1}X)^{-1}X'W^{-1}(y + u)$, $V = \sigma_v^2(X'W^{-1}X)^{-1}$. The "scale" factor of (20) depends on β through $n_o(\beta, u)$. In practice, we can use (21) as a proposal distribution. Suppose the proposed candidate is β^c and MCMC is currently at state $\beta^{(s)}$. Then we accept the candidate with the Metropolis-Hastings probability:[1]

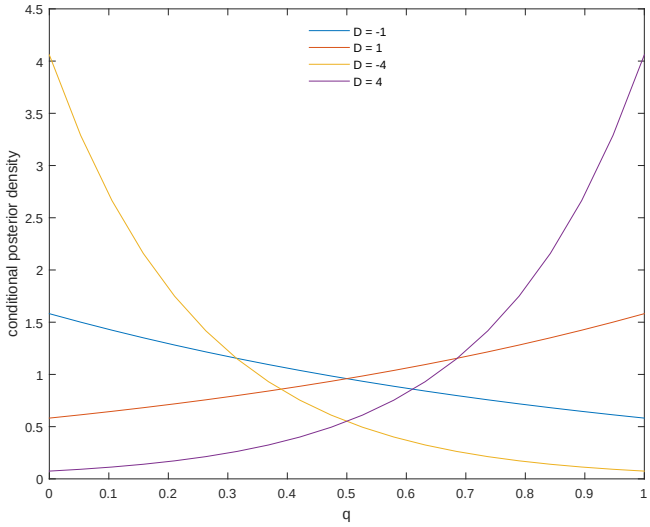
$$\min\left\{1, \frac{\exp\left\{-\frac{1}{2\sigma_v^2}(y + u - X\beta^c)' W^{-1}(y + u - X\beta^c)\left[q^{n_o(\beta^c,u)} + (1-q)^{n-n_o(\beta^c,u)} - 1\right]\right\}}{\exp\left\{-\frac{1}{2\sigma_v^2}\left(y + u - X\beta^{(s)}\right)' W^{-1}(y + u - X\beta^{(s)})\left[q^{n_o(\beta^{(s)},u)} + (1-q)^{n-n_o(\beta^{(s)},u)} - 1\right]\right\}}\right\}.$$

3.2 Drawing from the conditional posterior of q

Random draws from the posterior conditional of q, can be realized as follows. When $D < 0$, the posterior conditional is exponential restricted in the interval $(0, 1)$ so random draws can be obtained easily. When $D > 0$, we use acceptance sampling from the density $f(x) = 2x$, $0 \leq x \leq 1$, from which random draws are realized as: $x = U^{1/2}$, where U is a standard uniform draw. The draw is accepted with probability: $\frac{\exp(Dq)}{2q}$. From these draws, we can reconstruct the marginal posterior

[1] For a different algorithm, see Tsionas (2003).

Figure 1: Typical conditional posterior densities of q



density $p(q|y, X)$. Then we have two options: Either we use the draws for $\beta, \{u_i\}_{i=1}^n, q$ (which means that uncertainty with respect to q is explicitly taken into account), or we find the posterior mean $\bar{q} = E[p(q|y, X)] \doteq S^{-1} \sum_{s=1}^{S} q^{(s)}$, we condition on $q = \bar{q}$, and we run again the Gibbs sampler. In the second case, we condition on the "most likely" value of q. In the first case, which is the one that we recommend, q is *explicitly* treated as a parameter and, as we remarked, uncertainty is taken into account.

3.3 Imposition of no-quantile-crossing property and non-parametric functional form

Regarding the no-quantile-crossing property (NQCP) of Wang et al. (2014), this can be imposed by using the constraints:

$$x_i'\beta(q) \geq x_i'\beta(q+1) \, \forall i = 1, ..., n \, \forall q \in \mathcal{Q} \equiv \{q_1, ..., q_{H-1}\}, \tag{22}$$

where $0 < q_1 < ... < q_{H-1} < 1$, where H is the number of quantiles we consider. For any given value of q in Bayesian analysis, the NCQP can be enforced at all observed points via rejection sampling (i.e. simply discard the draws for which NCQP in (22) is not satisfied). Alternatively, one can impose that (22) holds at a small number of observations, and then check whether it holds at all other points. In practice, this results in considerable savings of time relative to naive rejection.

When q is a parameter, the enforcement of NQCP is more complicated but can be implemented easily as follows. Suppose we have the MCMC draws $\left\{\beta^{(s)}, q^{(s)}, s = 1, ..., s'\right\}$ where s' is the current draw. If $q^{(s')} > q^{(s'-1)}$, we need to have:

$$x_i'\beta^{(s')} < x_i'\beta^{(s)} \, \forall i = 1, ..., n, \tag{23}$$

otherwise we need: $x_i'\beta^{(s')} > x_i'\beta^{(s)}$ (the posterior probability of exact equality is zero). This restriction can be imposed easily as it involves only two sets of βs. If we allow the burn-in phase of the Gibbs sampler to be long enough, so that we have convergence to the posterior, then imposition of (22) at each observed point is *much* easier and holds at *all* data points.

Regarding the specification of the functional form, we notice that (3) can be a translog or any other functional form which is linear in the parameters. Often, this assumption is restrictive and, as Jradi and Ruggiero (2019) remarked, SDEA performs much better under misspecification relative to a linear model, as it does not make assumptions about the functional form. As SDEA, like the CNLS of Kuosmanen and Johnson (2010) and Kuosmanen and Kortelainen (2012)allows for monotonicity and curvature, we propose to use the Symmetric Generalized McFadden (SGM) functional form, proposed by Kumbhakar (1994) based on McFadden (1978) and introduced by Diewert and Wales (1987). Since in the next section we will use a cost function, we have:

$$C(w,Y) = a'w + g(w)(b'Y)^2 + w'\Lambda\tilde{Y} + \tfrac{1}{2}(\vartheta'w)\tilde{Y}'\Gamma\tilde{Y}, \tag{24}$$

where $a, \vartheta \in \mathbb{R}^K$, $b \in \mathbb{R}^M$, $g(w) = \frac{1}{2}\frac{w'\mathbb{D}w}{w'e}$, $\mathbb{D} \equiv [d_{kk'}, k, k' = 2, ..., K]$ is a symmetric, negative semidefinite matrix, $e \in \mathbb{R}^K_{++}$, Λ and Γ ($\Gamma = \Gamma'$) are $K \times M'$ and $M' \times M'$ matrices, respectively, $\tilde{Y}$ is the vector of outputs, possibly augmented with other variables (like a time trend), and $M' \geq M$ is the dimensionality of Y.This function is linearly homogeneous in prices, and globally concave with respect to prices. In practice, we set $e = [1, ..., 1]'$ as the only purpose of $w'e$ is to make sure that linear homogeneity holds without choosing arbitrarily one of the prices as numeraire. The aim of the cost function is mainly to estimate returns to scale, technical change, efficiency change, and productivity growth. Any cost function is dual to a production function or more accurately a production transformation function and the characteristics of the second can be derived using duality via the cost function.

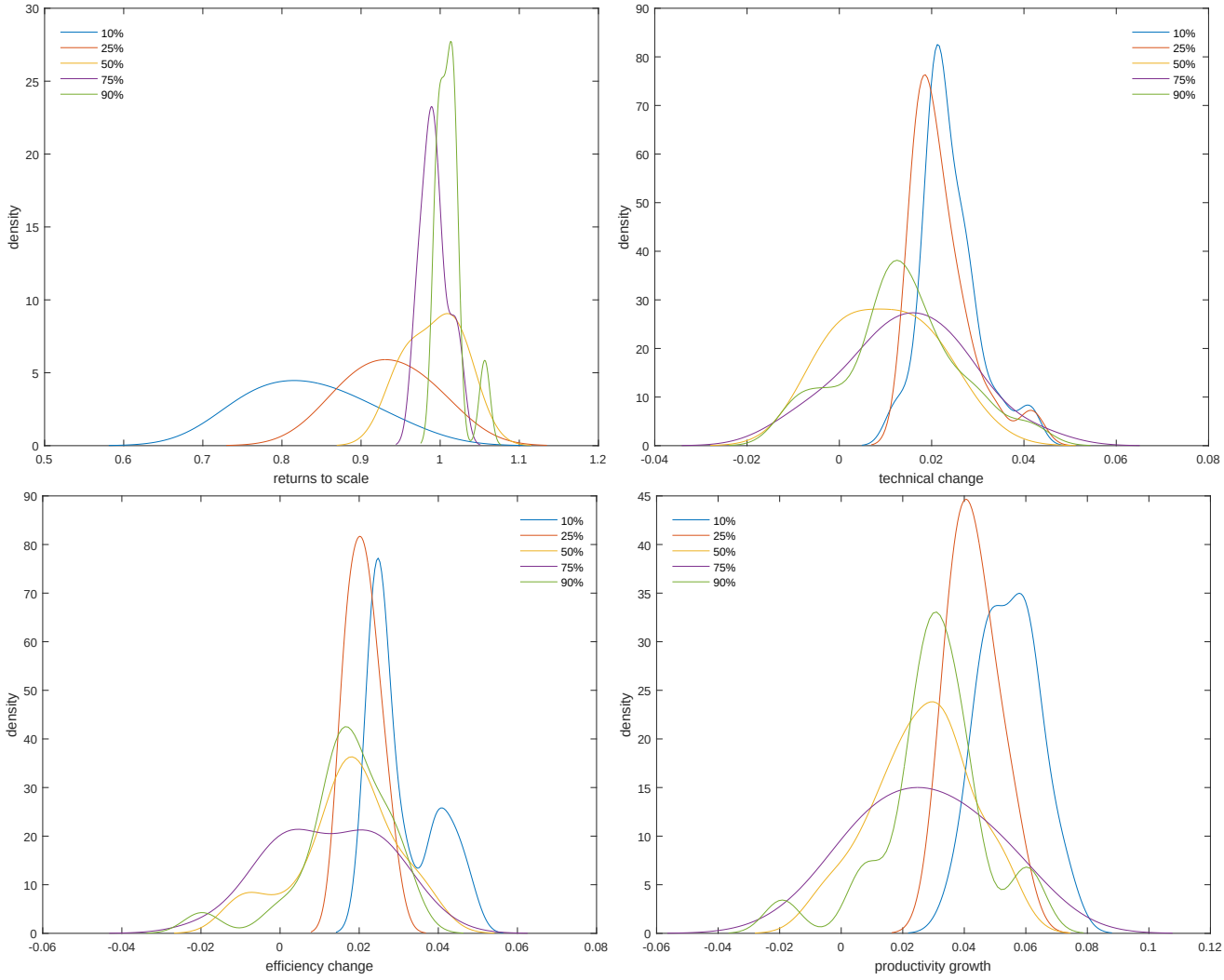
4 Empirical application

We apply the new techniques to a US banking data set, previously analyzed by Malikov, Kumbhakar, and Tsionas (2014). The data is an unbalanced panel with 2,397 bank-year observations for 285 large commercial banks operating in 2001-2010, whose total assets were in excess of one billion dollars (in 2005 U.S. dollars) in the first three years of observation. The data come from Call Reports available from the Federal Reserve Bank of Chicago. There are five inputs (whose prices, W, are available) and five outputs (Y). We estimate a cost function of the form $\ln C = \ln C(\ln W, \ln Y, Z) + v + u$. The functional form is translog and logs of non-performing loans as well as equity are included as quasi-fixed netputs (Z). We also include a time trend to measure technical change.[2] We implement MCMC using 150,000 passes the first 50,000 of which are discarded to mitigate possible start-up effects.

Reported in Figure 2 are sample distributions of posterior means for returns to scale (RTS), technical change (TC), efficiency change (EC) and productivity growth ($PG = TC + EC$). RTS is defined as the inverse cost elasticity, e_{cy}^{-1} (see Hanoch, 1975), where $e_{cy} = \sum_{m=1}^{M} \frac{\partial C(W,Y)}{\partial Y_m}\frac{Y_m}{C(W,Y)}$. Technical change is defined as $TC = \frac{\partial C(W,Y)}{\partial t}\frac{1}{C(W,Y)}$. Notice that $e_{cy} = \sum_{m=1}^{M} \frac{\partial \log C(W,Y)}{\partial \log Y_m}$, so it is an elasticity and measures the percentage effect on cost when all outputs increase by 1%.

[2] We impose monotonicity and concavity at 100 different points randomly chosen from the support of explanatory variables $\ln W, \ln Y$. In turn, we verified that the constraints hold at all points in the sample. It was not possible to enforce the restrictions using less than about 50 points, in this instance.

Figure 2: Aspects of the model, I



Similarly, $TC = \frac{\partial \log C(W,Y)}{\partial t}$ measures the percentage change in cost arising from "exogenous factors" or time (viz. neutral technical change).

Clearly, all measures differ substantially by quantile. Specifically, PG seems to be positive in the 10%, 25% and 50% quantiles, while for the 90% quantile the posterior is multimodal and there is evidence that PG can be negative for some banks. This effect is more pronounced for banks in the 75% quantile. The distributions of EC are also multimodal, and there is evidence of positive EC only for banks in the 10% and 25% quantile.

Overall cost efficiency distributions by different quantile, are reported in the upper-left panel of Figure 3. In the upper-right panel reported are the efficiency distributions for the entire sample from quantile and traditional SFM. In the lower-left panel, we present the distribution of rank correlation coefficients between efficiency estimates across different quantiles. In the lower-right panel, we report the posterior distributions of p-values for a heteroskedasticity test for the classical and quantile SFM. These p-values indicate that the traditional SFM clearly suffers from heteroskedasticity, while this is successfully resolved via the quantile SFM. To implement these tests in a Bayesian framework, the cost function

Figure 3: Aspects of the model, II

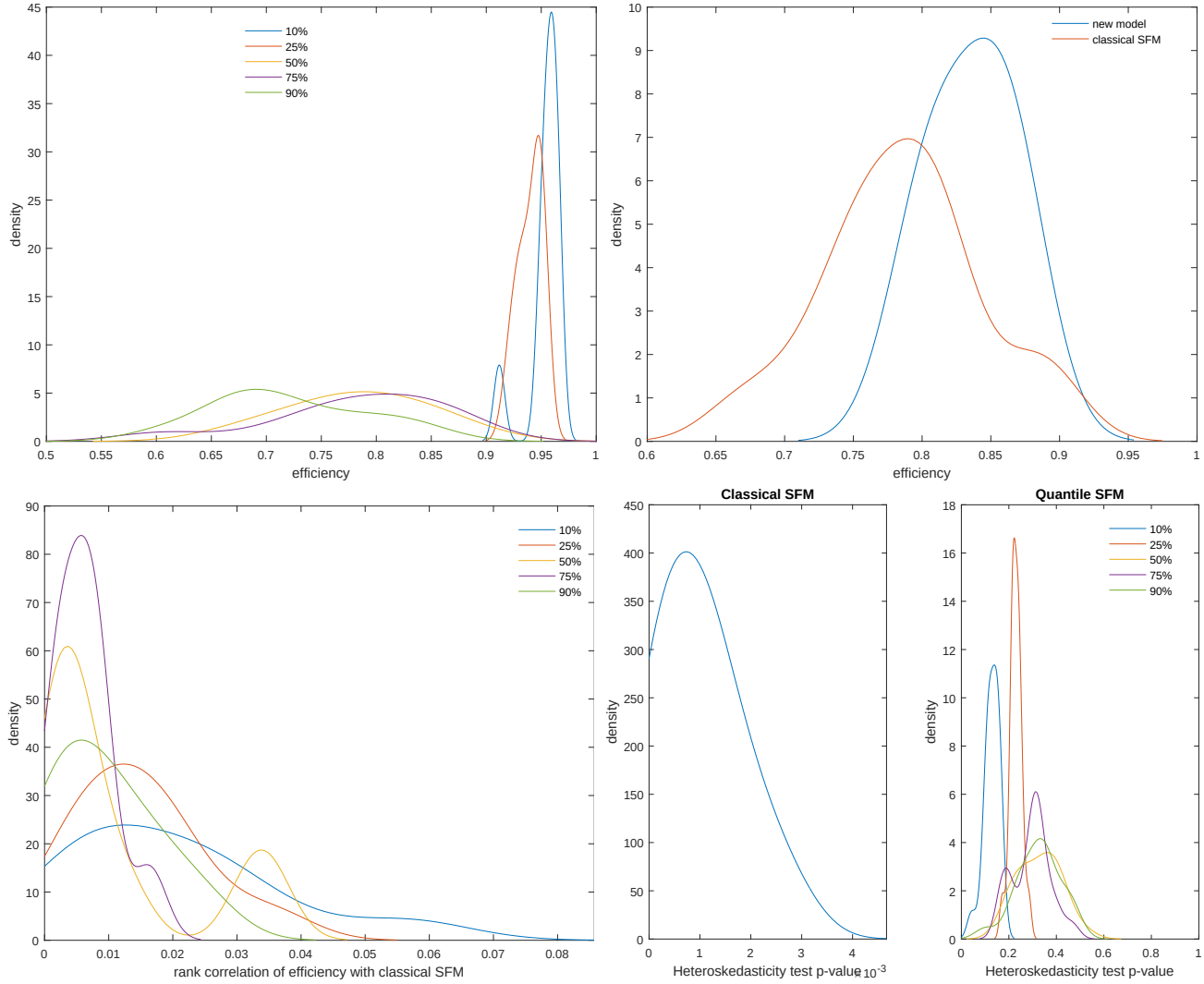
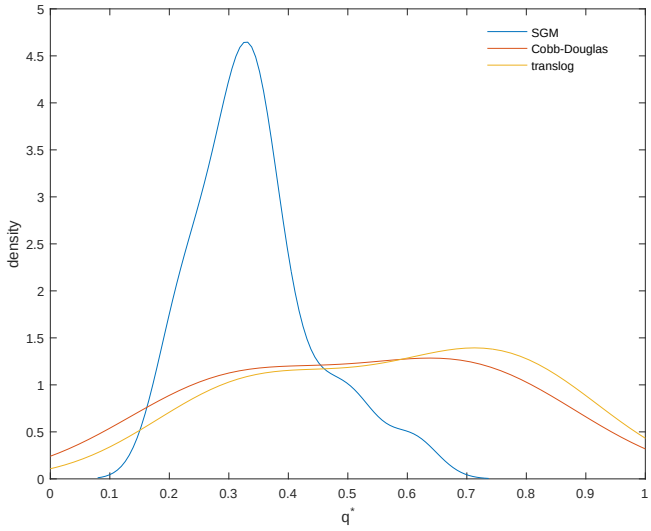


Figure 4: Marginal posterior densities of the optimal or most likely quantile, q^*



residuals squared are regressed on all explanatory variables for each MCMC draw. In turn, we save the p-value for the F-test that all regressors, except the intercept, are zero. The distribution of these p-values is reported in the lower-right panel of Figure 3. Clearly, heteroskedasticity should be interpreted as specification error in this context. Finally, we report marginal posterior densities of q, in Figure 4. We report marginal posterior densities under the assumption that the functional form is SGM, Cobb-Douglas or translog. The SGM favors values close to 0.4 on the average. The marginal posterior densities for the Cobb-Douglas are quite similar but, practically, flat showing that identification of q^* is quite difficult, in this instance. We can attribute this fact to misspecification of the functional form. The most likely quantile can be estimated using the posterior mean or mode and, in this instance, it is slightly over 0.30 for the SGM.

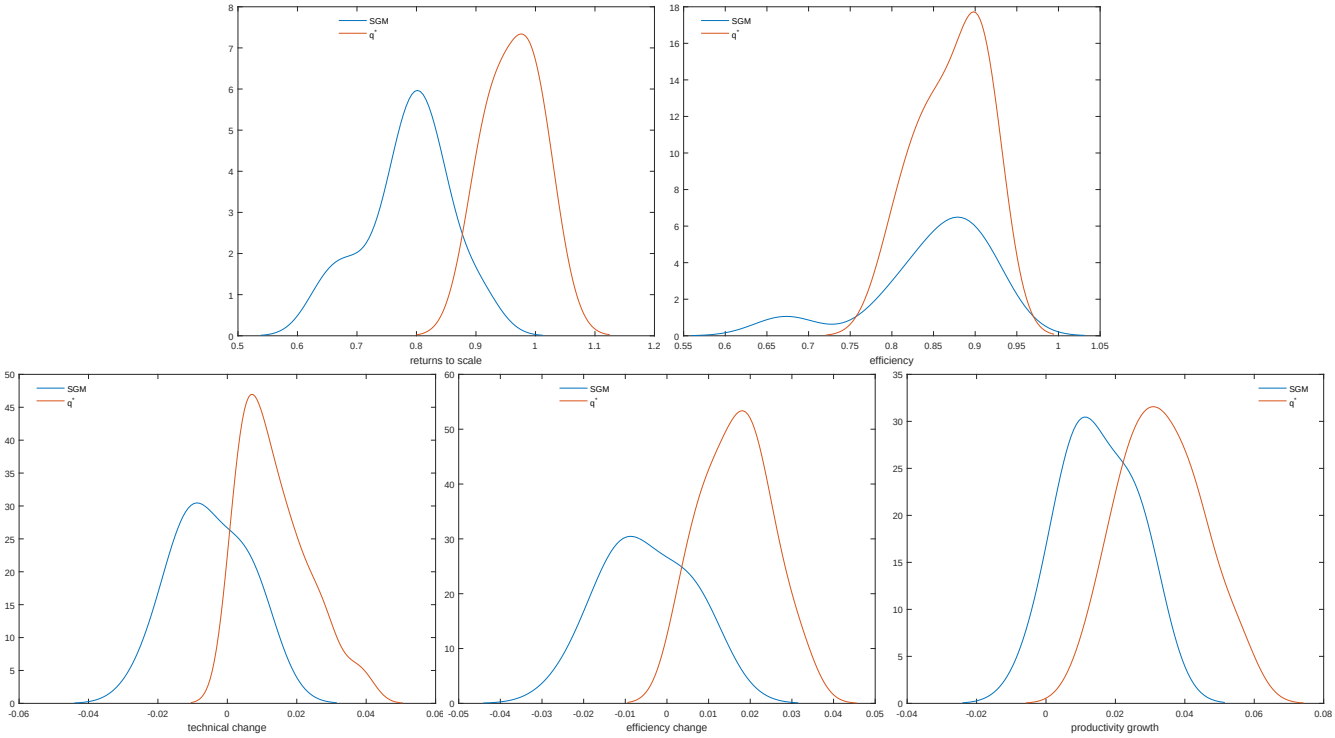
Another interesting issue is the comparison of the various measures between classical stochastic frontier (represented here by the SGM functional form) and the corresponding measures for the optimal quantile (q^*). This is done in Figure 5.

Evidently, returns to scale are much lower in the classical frontier model (averaging 0.8) whereas they are close to unity, on the average, for the optimal quantile frontier. In terms of efficiency, the optimal quantile frontier has a much large concentration of posterior probability around 0.90 whereas SGM has a long left tail from about 0.55 to 0.70. For technical change, efficiency change and productivity growth the optimal quantile frontier yields systematically larger values compared to SGM, and productivity growth averages near 3% whereas the corresponding value for SGM is near zero.

Concluding Remarks

In this study, we proposed a Quantile SFM and developed MCMC techniques for numerical Bayesian inference, extending Jradi and Ruggiero (2019). In an empirical application to US large banks we document important differences between the Quantile and the traditional SFM, in terms of several aspects of the data. We also document considerable heterogeneity among different quantiles in terms of returns to scale, technical change, efficiency change, technical efficiency, as well as

Figure 5: Comparison of classical stochastic frontier and optimal quantile frontier



productivity growth. Our functional form is based on the Symmetric Generalized McFadden which is flexible and globally concave in input prices, by construction. The Quantile SFM can be extended to allow for determinants of inefficiency, also known as environmental variables. This is a simple, yet important generalization that could be taken up in future work.

References

Aragon, Y., Daouia, A., & Thomas-Agnan, C. (2005). Nonparametric frontier estimation: A conditional quantile-based approach. Econometric Theory, 21, 358–389.

Behr, A. (2010). Quantile regression for robust bank efficiency score estimation. European Journal of Operational Research, 200 (2), 568–581.

Daouia, A., & Simar, L. (2007). Nonparametric efficiency analysis: A multivariate conditional quantile approach. Journal of Econometrics 140, 375–400.

Diewert, W.E., and T. J. Wales. 1987. Flexible Functional Forms and Global Curvature Conditions. Econometrica 55 (1), 43–68.

Hanoch, G. (1975). The Elasticity of Scale and the Shape of Average Costs. The American Economic Review 65 (3), 492–497.

Jradi, S., Parmeter, C., & Ruggiero, J. (2019a). Quantile estimation of the stochastic frontier model. Economics Letters, 182, 15-18.

Jradi, Parmeter and Ruggiero (2019b). Quantile Estimation of Stochastic Frontiers with the Normal-Exponential

Specification, submitted to EJOR.

Jradi, M., J. Ruggiero (2019). Stochastic Data Envelopment Analysis: A Quantile Regression Approach to Estimate the Production Frontier. European Journal of Operational Research, 278 (2), 385–393.

Koenker, R. W. and Bassett, G. W. (1978). Regression quantiles. Econometrica 46, 33–50.

Kumbhakar, S. C. (1994). A Multiproduct Symmetric Generalized McFadden Cost Function. Journal of Productivity Analysis 5, 349–357.

Kuosmanen, T. and A. Johnson (2010). Data envelopment analysis as nonparametric least square regression. Operations Research 58, 149–160.

Kuosmanen, T., and M. Kortelainen (2012). Stochastic non-smooth envelopment of data: Semi-parametric frontier estimation subject to shape constraints. Journal of Productivity Analysis 38, 11–28.

Malikov, E., S. C. Kumbhakar, and M. G. Tsionas (2016). A Cost System Approach to the Stochastic Directional Technology Distance Function with Undesirable Outputs: The Case of U.S. Banks in 2001-2010. Journal of Applied Econometrics 31 (7), 1407–1429.

Martins-Filho, C., & Yao, F. (2008). A smooth nonparametric conditional quantile frontier estimator. Journal of Econometrics 143, 317–333.

McFadden, D. L. (1978). Cost, Revenue, and Profit Functions. In *Production Economics*, edited by D. L. McFadden and M. Fuss. Amsterdam: North-Holland.

Tsionas, E. G. (2003). Bayesian quantile inference. Journal of Statistical Computation and Simulation 73 (9), 659–674.

Wang, Y., & Wang, S. (2013). Estimating α-frontier technical efficiency with shape restricted kernel quantile regression. Neurocomputing, 101, 243–251.

Wang, Y., Wang, S., Dang, C., & Ge, W. (2014). Nonparametric quantile frontier estimation under shape restriction. European Journal of Operational Research 232, 671–678.

Yang, Y.; He, X. (2010). Bayesian empirical likelihood for quantile regression. Annals of Statistics 40 (2), 1102–1131.

Yang, Y.; Wang, H.X.; He, X. (2016). Posterior Inference in Bayesian Quantile Regression with Asymmetric Laplace Likelihood. International Statistical Review 84 (3): 327–344.

Yu, K.; R. A. Moyeed (2001). Bayesian quantile regression. Statistics & Probability Letters 54 (4), 437–447.