Robust Bayesian Inference in Stochastic Frontier Models

Mike G. Tsionas*

Abstract

We use the concept of coarsened posteriors to provide robust Bayesian inference via coarsening in order to robustify posteriors arising from stochastic frontier models. These posteriors arise from tempered versions of the likelihood when at most a pre-specified amount of data is used, and are robust to changes in the model. Specifically, we examine robustness to changes in the distribution of the composed error in SFM. Moreover, coarsening is a form of regularization, reduces overfitting and makes inferences less sensitive to model choice. The new techniques are illustrated using artificial data as well as in a substantive application to large U.S banks.

Keywords: Productivity and Efficiency; Bayesian Analysis; Robustness; Stochastic Frontier Models.

---

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

1   Introduction

The stochastic frontier model (SFM) is a standard tool in estimation of efficiency from observed data. Robustness of SFM has not been examined thoroughly in the literature although many alternative distributional assumptions have been proposed for the error components of the model. [1] Bayesian analysis of SFM is widely used due to the convenience allowed by Markov Chain Monte Carlo in dealing with latent inefficiencies which are present in the model, particularly under alternative distributional assumptions. Feng et al. (2019) proposed a semiparametric model for stochastic frontier models: Specifically, the one-sided error term is approximated by a log-transformed Rosenblatt-Parzen kernel density estimator. In a Monte Carlo study they find that that the kernel-based semiparametric model performs better than the commonly-used exponential stochastic frontier model. Their study also indicates that the kernel model shows similar performance to a non-parametric model.

Our motivation in this paper is to take into account the previous literature with an eye towards making the posterior (and, therefore, statistical inferences) more robust to misspecification. For example, standard inference in stochastic frontier models does not take into account not only outliers but, perhaps more importantly, deviations of the assumed distributions of two-sided and one-sided error terms from their actual counterparts. To the extent that the actual distributions are unknown (see, for example, Feng et al., 2019 for details) misspecification is quite likely so, in practice, statistical inferences are likely to be misleading.

A robust posterior, in the general case, has been proposed by Miller and Dunson (2019). Specifically, rather than conditioning on the observed data assumed to be generated by the model, we condition on the event that the model generates data that are *distributionally close* to the observed data. This technique allows to examine robustness to changes in the distribution of the composed error in SFM. Additionally, coarsening is a form of regularization, reduces overfitting and makes inferences less sensitive to model choice. When we are interested in estimation of efficiency, returns to scale, productivity growth etc., this is clearly a desirable goal.

2   Model

Suppose we have observed data $x \in \mathcal{X} \subseteq \mathbb{R}^n$ and the "ideal" data is X^*-ideal in the sense that it is a random sample from true data generating process (DGP). We focus on the case of i.i.d data whose distribution has density $p(x_i|\theta)$, where $\theta \in \Theta \subseteq \mathbb{R}^M$ is a parameter and Θ is the parameter space. The usual posterior $p(\theta|x)$ may not be robust if there are outliers and / or we are not certain that the density $p(x_i|\theta)$ is the true one.

A robust posterior is defined by Miller and Dunson (2019) in the i.i.d case as follows: $p(\theta|d(\hat{P}_{X^*}, \hat{P}_x) < r)$, where $\hat{P}_x = n^{-1} \sum_{i=1}^n \delta_{x_i}$ is the empirical distribution of x and similarly for $\hat{P}_{X^*}$, for some discrepancy measure $\mathscr{D}(\cdot, \cdot)$ and a given $r > 0$. Therefore, we condition on the event that the empirical distribution of actual data is close to the empirical distribution of data generated by the model, using a certain discrepancy function between probability measures.

---

[1] For excellent surveys, see Greene (1999) and Parmeter and Kumbhakar (2014).

If P_o and P_θ for $\theta \in \Theta$ have densities p_o and p_θ, respectively, the Kullback–Leibler divergence is defined as:

$$\mathscr{D}(P_\theta, P_o) = {}_\Theta\, p_o(\chi) \log \frac{p_o(\chi)}{p_\theta(\chi)} \lambda(d\chi), \tag{1}$$

where λ is the Lebesgue measure. Miller and Dunson (2019) have derived the following approximation to the posterior under the assumption that r has an exponential distribution with parameter α:

$$p(\theta|d(\hat{P}_X, \hat{P}_x) < r) \propto p(\theta) \prod_{i=1}^{n} p(x_i|\theta)^\varsigma \equiv p_\varsigma(\theta|x), \tag{2}$$

where $\varsigma = \frac{\alpha}{\alpha+n}$ and $p(\theta)$ is the prior. The approximation is accurate when $n \gg \alpha$ or $\alpha \gg n$. The main objective of such "coarsened" posteriors is robustness to small changes in the shape of the distribution of the data, i.e. the data generating process. The approximation avoids altogether computation of $\mathscr{D}(\hat{P}_X, \hat{P}_x)$ or the term ${}_\Theta\, p_o \log p_o$ which is independent of θ. The interpretation of the coarsened posterior is that it adjusts the sample size from n to $n\varsigma$, so effectively we have a smaller sample size.

In this paper we consider the production SFM[2]:

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

where $x_i \in \mathbb{R}^k$ is a vector of regressors, and $v_i \sim iid\mathcal{N}(0, \sigma_v^2)$, $u_i \sim iid\mathcal{N}_+(0, \sigma_u^2), i = 1, ..., n$, $(v_i, u_i)\perp x_i, i = 1, ..., n$. Let $\sigma^2 = \sigma_v^2 + \sigma_u^2$ and $\lambda = \frac{\sigma_u}{\sigma_v}$ and define the parameter vector $\theta = [\beta', \sigma_v, \sigma_u]'$. The augmented posterior is:

$$p(y_i, u_i|x_i, \theta) \propto \sigma_v^{-1}\sigma_u^{-1} \exp\left\{ -\frac{1}{2\sigma_v^2} (y_i + u_i - x_i'\beta)^2 - \frac{1}{2\sigma_u^2}u_i^2 \right\}, i = 1, ..., n. \tag{4}$$

The marginal density of y_i is $p(y_i|x_i, \theta) = {}_0^\infty\, p(y_i, u_i|x_i, \theta)du_i$ and is available in closed form:

$$p(y_i|x_i, \theta) \propto \sigma^{-1} \exp\left\{ -\frac{(y_i - x_i'\beta)^2}{2\sigma^2} \right\} \Phi\left(-\frac{\lambda}{\sigma}(y_i - x_i'\beta) \right), i = 1, ..., n, \tag{5}$$

where $\Phi(\cdot)$ is the standard normal distribution function; see Kumbhakar and Lovell (2000, p. 78).

Suppose we have a prior $p(\theta)$ and $y = [y_i; i = 1, ..., n]$, $X = [x_i; i = 1, ..., n]$ denote the data. A coarsened posterior that uses at most ςn out of n observations is:

$$p_\varsigma(\theta|y, X) \propto p(\theta) \prod_{i=1}^{n} p(y_i|x_i, \theta)^\varsigma. \tag{6}$$

Similarly, we can define the coarsened augmented posterior:

---

[2]The cost frontier is obtained by taking $-y_i$ and $-x_i$. Moreover, the extension to the case of panel data is straightforward.

Algorithm 1 MCMC draws from posterior conditional distributions

---

1. Draw regression parameters from

$$\beta|\sigma_v, \sigma_u, u, y, X \sim \mathcal{N}_k\left(b, \frac{\sigma_v^2}{\zeta}(X'X)^{-1}\right), \tag{11}$$

where $b = (X'X)^{-1}X'(y+u)$.

2. Draw the scale parameter σ_v^2 as follows:

$$\frac{\underline{q} + (y - X\beta + u)'(y - X\beta + u)}{\sigma_v^2/\zeta}|\beta, \sigma_u, u, y, X \sim \chi^2(\underline{n} + \zeta n). \tag{12}$$

3. Draw the scale parameter σ_u^2:

$$\frac{u'u}{\sigma_u^2/\zeta}|\beta, \sigma_v, u, y, X \sim \chi^2(n). \tag{13}$$

4. Draw technical inefficiencies:

$$u_i|\beta, \sigma_v, \sigma_u, u, y, X \sim \mathcal{N}_+(\hat{u}_i, \sigma^2), i = 1, .., n, \tag{14}$$

where $\hat{u}_i = -\frac{\sigma^2(y_i - x_i'\beta)}{\sigma_v^2}, i = 1, ..., n.$

$$p_\zeta(\theta, u|y, X) \propto p(\theta) \prod_{i=1}^{n} p(y_i, u_i|x_i, \theta)^\zeta, \tag{7}$$

where $u = [u_1, ..., u_n]'$. For the SFM it becomes:

$$p_\zeta(\theta, u|y, X) \propto p(\theta)\sigma_v^{-\zeta n}\sigma_u^{-\zeta n} \exp\left\{-\frac{\zeta}{2\sigma_v^2}\sum_{i=1}^{n}(y_i + u_i - x_i'\beta)^2 - \frac{\zeta}{2\sigma_u^2}\sum_{i=1}^{n}u_i^2\right\}. \tag{8}$$

With a nearly flat prior of the form[3]:

$$p(\theta) \propto \sigma_v^{-(\underline{n}+1)} \exp\left\{-\frac{\underline{q}}{2\sigma_v^2}\right\}, \quad p(\sigma_u) \propto \sigma_u^{-1}, \quad p(\beta) \propto \text{const.}, \tag{9}$$

the posterior becomes:

$$p_\zeta(\theta, u|y, X) \propto p(\theta)\sigma_v^{-\zeta n - \underline{n} - 1}\sigma_u^{-\zeta n - 1} \exp\left\{-\frac{\zeta}{2\sigma_v^2}\left[\underline{q} + \sum_{i=1}^{n}(y_i + u_i - x_i'\beta)^2\right] - \frac{\zeta}{2\sigma_u^2}\sum_{i=1}^{n}u_i^2\right\}. \tag{10}$$

Inferences can be implemented using Gibbs sampling with data augmentation based on drawing random numbers from the posterior conditional distributions summarized in Table 1.

We implement the Gibbs sampler using 15,000 iterations, the first 5,000 of which are discarded to mitigate possible start up effects. In all computations we set $\underline{n} = 0$ and $\underline{q} = 0.1$. In the neighborhood of these values we did not notice sensitivity

---

[3]The prior for σ_v with $\underline{n} = 0$ has been proposed by Fernandez et al. (1997).

of posteriors.[4] It should be mentioned that the coarsened posterior will have better mixing properties corresponding to $\zeta = 1$ (which already mixes well) as the likelihood is tempered. Convergence was diagnosed successfully using Geweke's (1992) diagnostics.

Suppose $\left\{u_i^{(s)}, s = 1, ..., S\right\}$ denotes Gibbs draws for technical inefficiencies (i.e. a sample from the posterior), and set $R_i^{(s)} = \exp(-u_i^{(s)})$. Then an estimate of firm-specific efficiency is:

$$\hat{R}_i = S^{-1} \sum_{s=1}^{S} R_i^{(s)}, i = 1, ..., n. \tag{15}$$

Clearly, such measures depend on the amount of coarsening (ζ) and, therefore, there is the possibility of sensitive dependence on alternative model specifications.

Sometimes, we may have prior information on the regression parameters, β, summarized in the form:

$$\beta | \sigma_v \sim \mathcal{N}_k \left(\beta_o, \frac{\sigma_v^2}{\zeta} V_o\right), \tag{16}$$

where the prior mean is β_o and the prior covariance matrix is given by $\frac{\sigma_v^2}{\zeta} V_o$, where β_o and V_o are known. In this case, we have to modify (11) as follows:

$$\beta | \sigma_v, \sigma_u, u, y, X \sim \mathcal{N}_k \left(\bar{b}, \overline{V}\right), \tag{17}$$

where $\bar{b} = \left(X'X + V_o^{-1}\right)^{-1} X'(y + u)$, and $\overline{V} = \frac{\sigma_v^2}{\zeta} \left(X'X + V_o^{-1}\right)^{-1}$.

3   Illustration

We give an illustration using two models. **Model I** is: $y_i = \beta_1 + \beta_2 x_{i1} + \beta_3 x_{i2} + v_i - u_i, i = 1, ..., n$, where $\beta_1 = -1$, $\beta_2 = \beta_3 = \frac{1}{2}$, with $n = 1000$ observations, when $v_i \sim$ i.i.d $\mathcal{N}(0, 0.1^2)$ and, independently, $u_i \sim i.i.d\mathcal{N}_+(0, 0.5^2)$. Here, x_{i1} and x_{i2} are generated from standard normal distributions. **Model II** has 900 observations generated from the first model, but the last 100 are generated as $y_i = v_i - u_i, i = 1, ..., n$, where $v_i \sim$ i.i.d $\mathcal{N}(0, 0.1^2)$ and $u_i \sim i.i.d\mathcal{N}_+(0, 0.1^2)$. Therefore, in the second model, part of the data is is generated from a process *without* systematic part and lower signal-to-noise ratio.

In the left panel of Figure 1, we report the true density of efficiency scores along with estimates of the density for $\zeta = 1$, 0.95, 0.98 and 0.9. The densities corresponding to different values of ζ are sample densities of posterior mean scores across the sample. In the right panel of Figure 1, we report 100 random efficiency scores corresponding to posteriors with different values of ζ.

From the right panel of Figure 1, the differences between efficiency scores are not so marked as in Model II whose results are reported in the upper panels of Figure 2. In the lower panel we report marginal posterior densities of the parameters

---

[4]The neighborhood is defined as $\underline{n} + \nu$ and $\underline{q} + s$, where $\nu \in \left[10^{-4}, 3\right]$ and $s \in \left[10^{-4}, 1\right]$ following a uniform distribution.
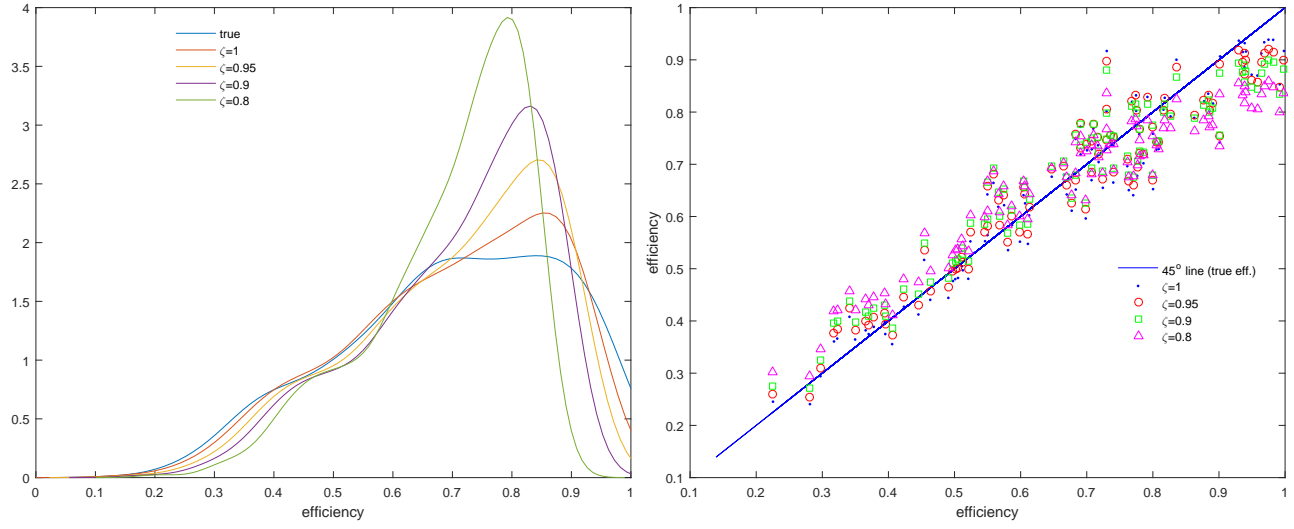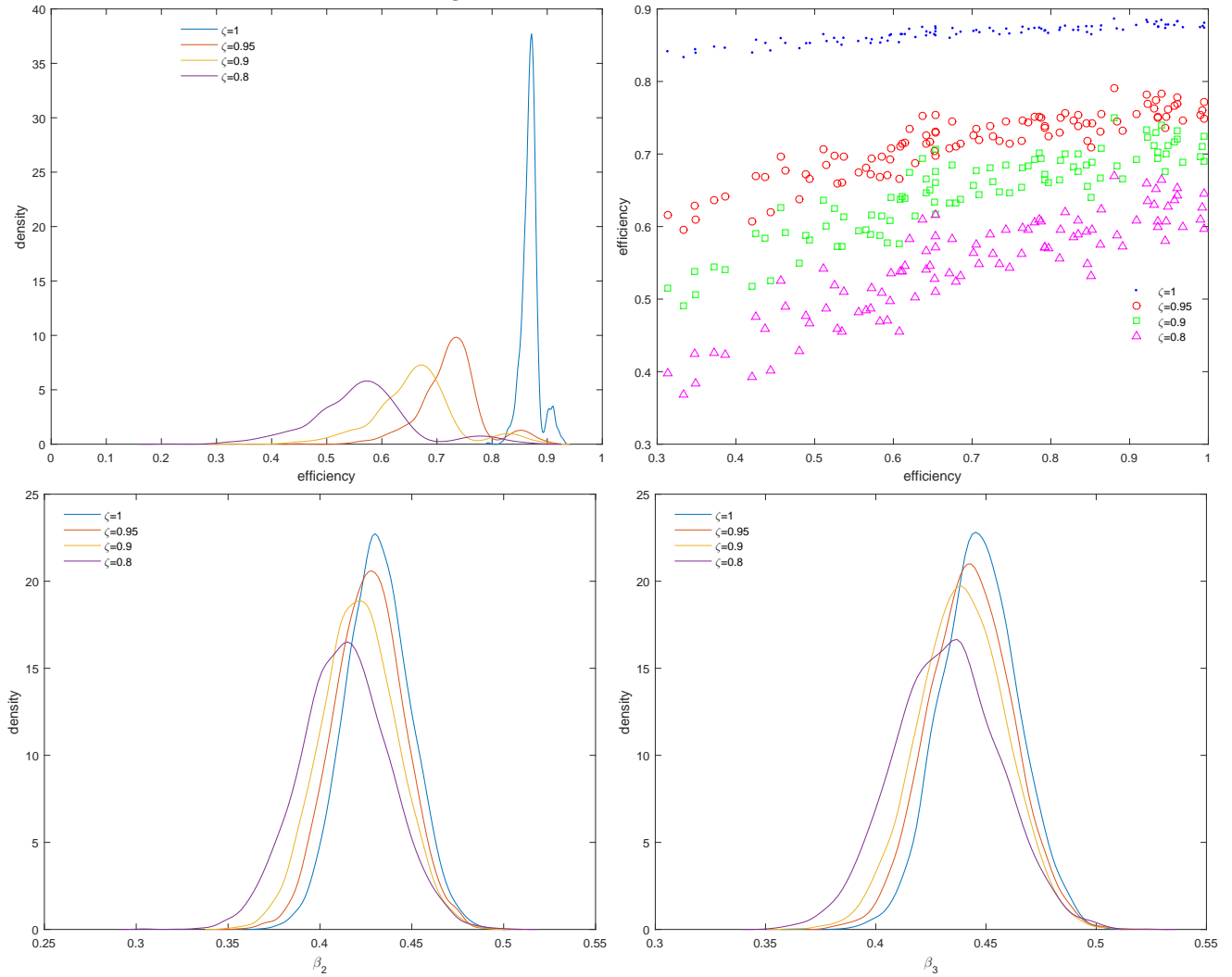
Figure 1: Artificial data, Model I

Figure 2: Artificial data, Model II

β_2 and β_3. Efficiency scores are markedly different compared to $\zeta = 1$ and as ζ decreases to 0.8 the efficiency distributions move to the left. The marginal posteriors of β_2 and β_3 also seem to be different although they are not far from the posterior mean when $\zeta = 1$.

4   Empirical application

We use the banking data of Malikov, Kumbhakar and Tsionas (2015) to estimate a translog cost frontier with five input prices, five outputs, a bad output (non-performing loans) and equity included as quasi-fixed input. The data is an unbalanced panel with 2,397 bank-year observations for 285 large, relatively homogeneous US banks (2001:I-2010:IV). We refer the reader to Malikov et al. (2015) for further details on the data. Our results are reported in Figures 3 and 4. As ζ increases, the distribution of efficiency scores shifts to the left, efficiency scores remain, however, highly correlated (upper right panel) and the posterior densities of both σ and λ shift to the right. Therefore, changes in distributional assumptions are likely to increase the variance of the error term but also the signal-to-noise ratio λ from 1.2 to 1.8, on the average. From the results in Figure 4, posterior densities of output cost elasticity (e_{cy}), elasticity with respect to non-performing loans (e_{cb}), technical change (e_{ct}) and elasticity with respect to quasi-fixed equity ($e_{c,eq}$) remain robust as ζ changes, although this is less so for output cost elasticity and, therefore, returns to scale.

Finally, to address the question of selecting ζ, Miller and Dunson (2019) propose a measure of fit and a measure of model complexity. A measure of fit is given by the average log-likelihood and the complexity measure is given by the coarsened posterior. Instead, one case use the log marginal likelihood:

$$\mathcal{M}(y, X) = \frac{p(\theta)p_\zeta(y|X, \theta)}{p_\zeta(\theta|y, X)} \; \forall \theta \in \Theta, \tag{18}$$

where $p_\zeta(y|X, \theta)$ is the tempered likelihood. Since this is an identity, we can use $\theta = \overline{\theta}_\zeta$, the posterior mean, and the denominator can be approximated with a multivariate normal distribution: $p_\zeta(\theta|y, X) \simeq (2\pi)^{-M/2}|V_\zeta|^{-1/2}$, where $V_\zeta = S^{-1}\sum_{s=1}^{S}(\theta_\zeta^{(s)} - \overline{\theta}_\zeta)(\theta_\zeta^{(s)} - \overline{\theta}_\zeta)'$ is the posterior covariance matrix, and $M = dim(\theta)$. This is the Laplace approximation to log marginal likelihood (LML), see DiCiccio et al. (1997). We report the values of LML for ten values of ζ in Figure 5.

Since LML stabilizes at $\zeta = 0.55$ one would be safe to average efficiency distributions over the different values of ζ shown in Figure 3. This is valid as alternative models (corresponding to different values of ζ) have approximately the same posterior model probability given by: $p_\zeta(y, X) = \frac{\mathcal{M}_\zeta(y, X)}{\sum_{\zeta'} \mathcal{M}_{\zeta'}(y, X)}$. Moreover, from the upper right panel of Figure 4 these efficiency scores are highly correlated and slight changes in the data generating process affect only their location. Averaged efficiency densities and densities of output cost elasticity are presented in Figure 6.

The results indicate that efficiency ranges from 60% to slightly less than 100% with a median near 85%. As output cost elasticity averages 1.2, there seem to exist decreasing returns to scale in U.S large banks (the returns to scale measure

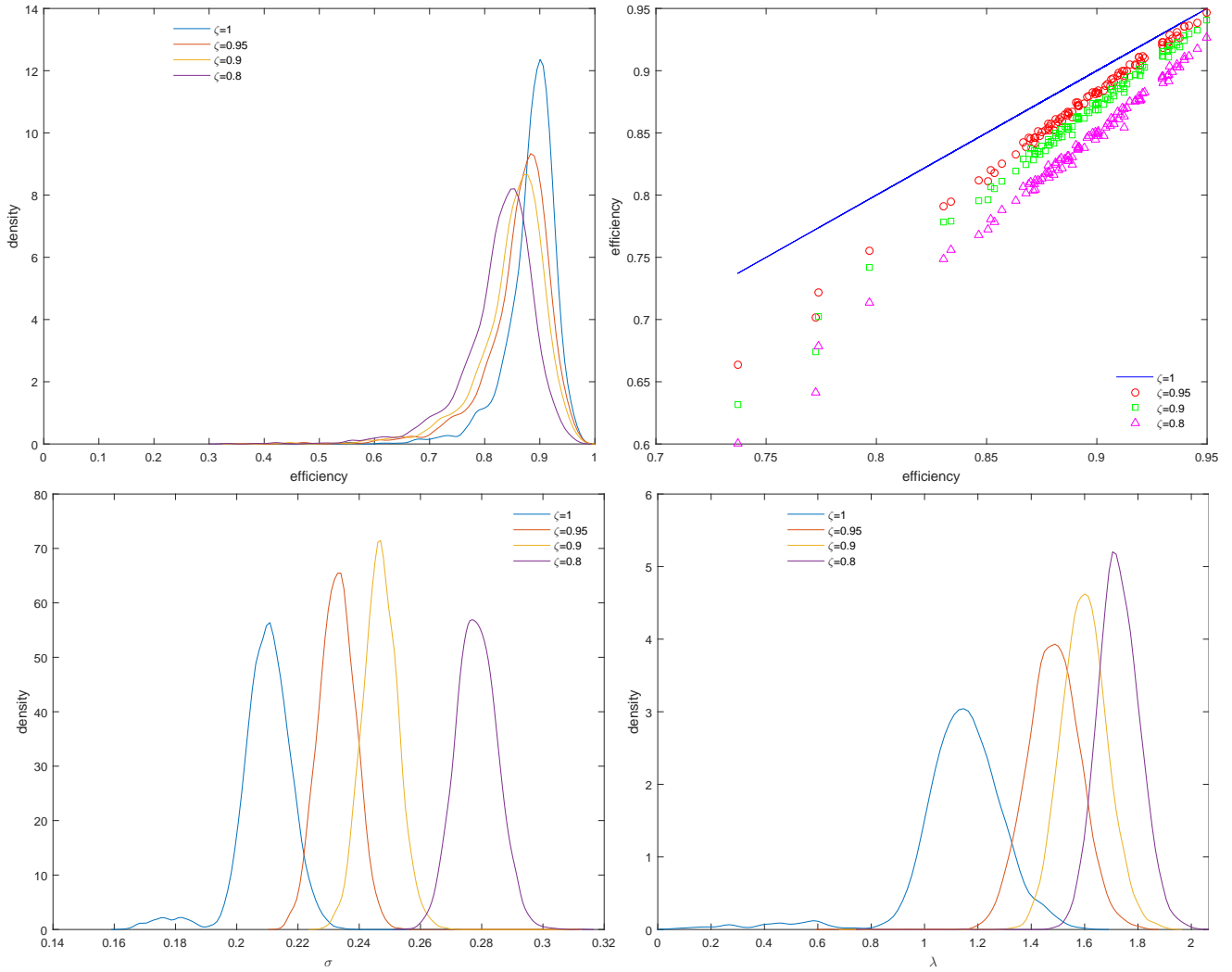Figure 3: U.S Banking data, efficiency scores, σ and λ

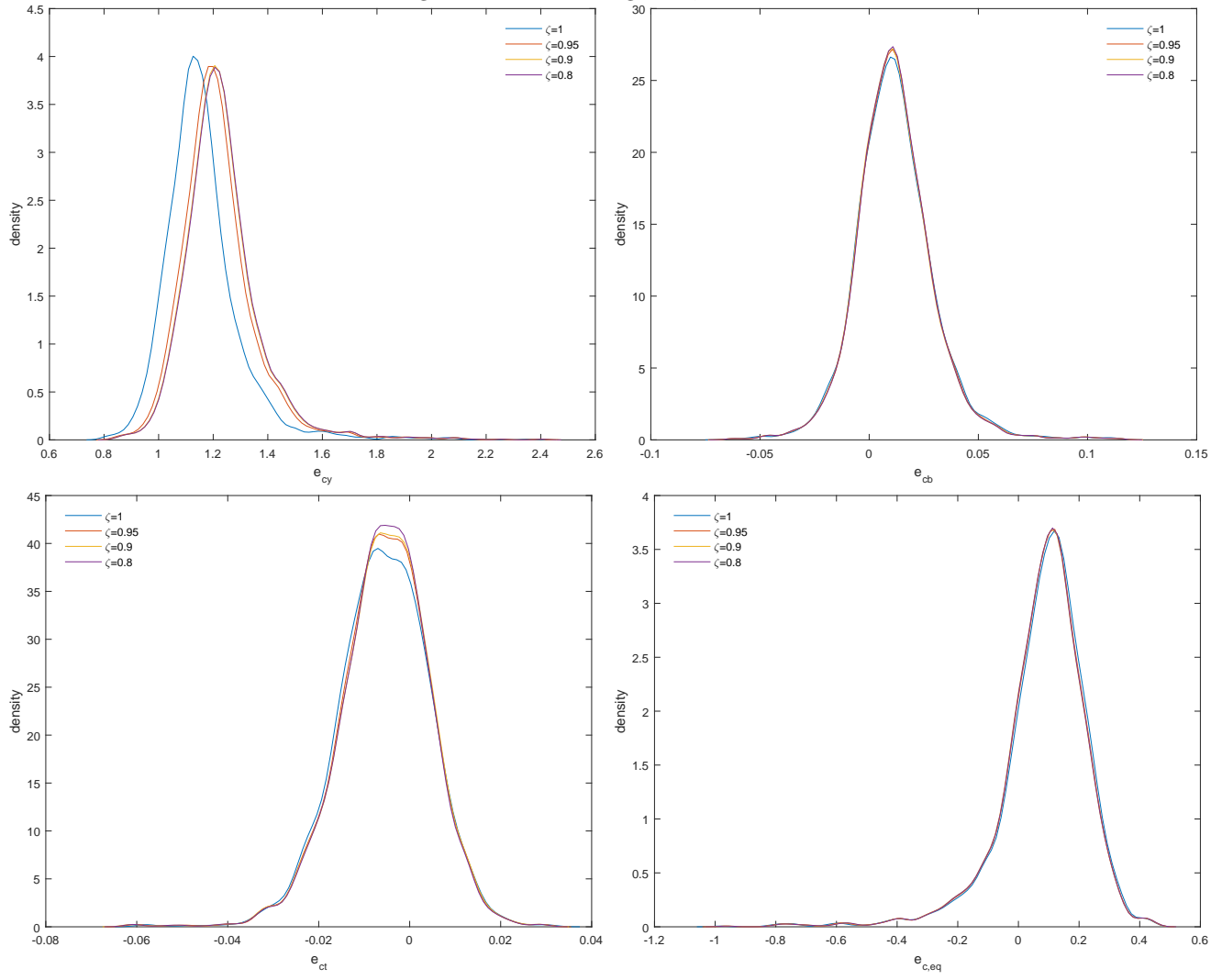Figure 4: U.S Banking data, elasticities

Figure 5: Normalized log marginal likelihood
Note: Log marginal likelihood is normalized so that its value at $\zeta = 0.5$ is zero.
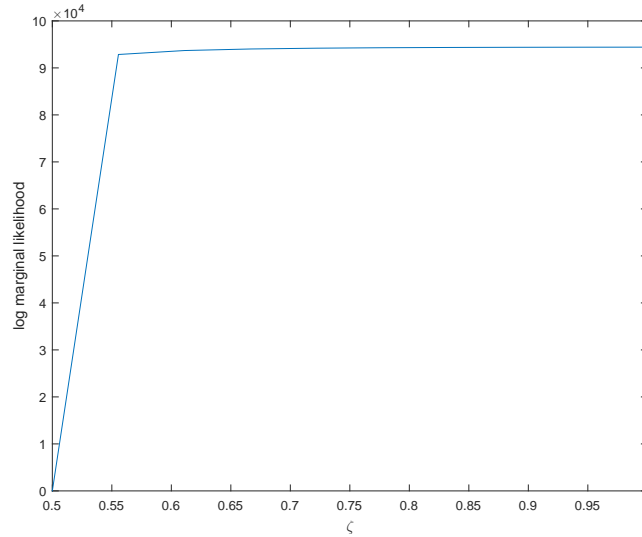
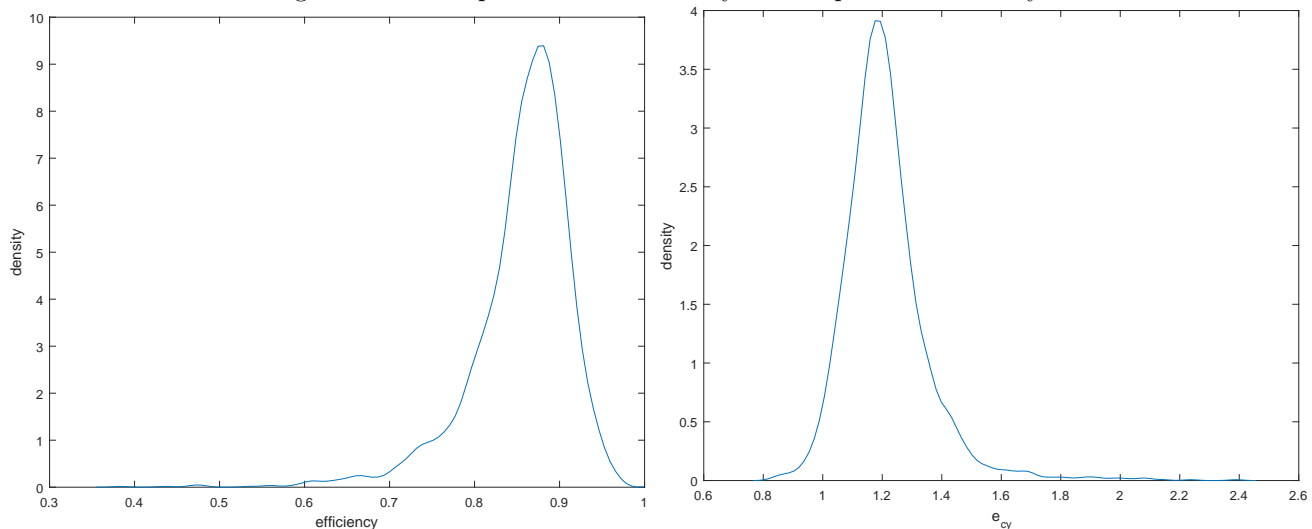Figure 6: Robust posteriors of efficiency and output cost elasticity

Table 1: MCMC convergence diagnostics

	GCD	RNE	acf(50)
β	1.619	0.582	0.392
σ_v	1.303	0.618	0.403
σ_u	1.202	0.588	0.380
u	1.719	0.449	0.344

Notes: GCD is Geweke's (1992) convergence diagnostic which follows (asymptotically in the number of MCMC draws) a standard normal distribution. RNE denotes "relative numerical efficiency" which is equal to one under i.i.d sampling from the posterior. Moreover, acf(50) denotes autocorrelation of MCMC draws at lag 50. More specifically, "GCD" is the maximum absolute value across parameters β or u. We use the same strategy for RNE and acf(50).

is $1/e_{cy}$) with constant returns enjoyed by a sizable number of banks, and increasing returns being only exceptional. Finally, as model uncertainty increases (corresponding to lower values of ζ), e_{cy} increases slightly and its posterior shifts to the right (upper left panel of Figure 4).

Finally, in Table 1 we provide diagnostic measures to ensure that our MCMC draws provide access to the true posterior.

Concluding remarks

We have proposed the concept of coarsened posteriors to robustify inferences from SFM. In an application to U.S banks, we find that economies of scale and technical change can be estimated in a relatively robust way, including elasticities with respect to equity and non-performing loans but efficiency inferences are less robust to the amount of data we use to robustify the posterior. This causes some concern as small changes in the distributional assumptions may change efficiency scores considerably. Fortunately, efficiency scores remain highly correlated across the robustness parameter (ζ), at least in this application. In terms of future research, it would be interesting to examine the performance of robust posteriors in more complicated SFM and reconcile differences that arise from different approaches to modeling inefficiency. Moreover, an interesting avenue for future research would be based on recent work by Guo et al. (2018) who examined whether a parametric production frontier function is suitable in the analysis. Guo et al. (2018) developed two test statistics based on

local smoothing and an empirical process, respectively and suggested also residual-based wild bootstrap versions of these two test statistics. As coarsening provides more robust results it is likely that the procedures in Guo et al. (2018) would tend to be in favor of the parametric specification although one has to resolve the issue of applying the Guo et al. (2018) procedures in a Bayesian context.

References

DiCiccio, T. J., R. E. Kass, A. Raftery, and L. Wasserman (1997). Computing Bayes factors by combining simulation and asymptotic approximations. Journal of the American Statistical Association 92, 903–915.

Feng, G., Wang, C. and Zhang, X. (2019), Estimation of Inefficiency in Stochastic Frontier Models: A Bayesian Kernel Approach, Journal of Productivity Analysis 51, 1–19.

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 (1), 169–193.

Geweke J. (1992). Evaluating the accuracy of sampling based approaches to the calculation of posterior moments. Bayesian Statistics 4, J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith (eds.), pp. 169–193.

Greene, W. (1999). Frontier Production Functions. In Handbook of Applied Econometrics. Volume II: Microeconomics, M.H. Pesaran and Schmidt, P. (Eds.), Oxford: Blackwell, chapter 3, pp. 75–153.

Guo, X., Li, G.-R., McAleer, M., Wong, W.K. (2018), Specification Testing of Production in a Stochastic Frontier Model, Sustainability, 10, 3082; doi:10.3390/su10093082.

Kumbhakar, S. C., and C. A. K. Lovell (2000). Stochastic frontier analysis. Cambridge, Cambridge University Press.

Malikov, E., S. C. Kumbhakar, and E. 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.

Miller, J. W., and D. B. Dunson (2019). Robust Bayesian inference via coarsening. Journal of the American Statistical Association 114 (527), 1113–1125.

Parmeter C., and S. C. Kumbhakar (2014). Efficiency Analysis: A Primer on Recent Advances. Foundations and Trends in Econometrics 7 (3–4), 191–385.