Central European Journal of Economic Modelling and Econometrics

Bayesian Inference and Gibbs Sampling in Generalized True Random-Effects Models

Kamil Makieła*

Submitted: 16.11.2016, Accepted: 11.01.2017

Abstract

The paper investigates Bayesian approach to estimate generalized true random-effects models (GTRE). The analysis shows that under suitably defined priors for transient and persistent inefficiency terms the posterior characteristics of such models are well approximated using simple Gibbs sampling. No model re-parameterization is required. The proposed modification not only allows us to make more reasonable (less informative) assumptions as regards prior transient and persistent inefficiency distribution but also appears to be more reliable in handling especially noisy datasets. Empirical application furthers the research into stochastic frontier analysis using GTRE models by examining the relationship between inefficiency terms in GTRE, true random-effects, generalized stochastic frontier and a standard stochastic frontier model.

Keywords: generalized true random-effects model, stochastic frontier analysis, Gibbs sampling, Bayesian inference, cost efficiency, transient and persistent efficiency

JEL Classification: C11, C23, C51, D24

*Cracow University of Economics; e-mail: kamil.makiela@uek.krakow.pl

1 Introduction

Stochastic frontier application to panel data has led to a great deal of research into ways of modelling inefficiency variation. In recent years Colombi, Martini and Vittadini (2011) proposed a generalized true random-effects model (GTRE); see also Kumbhakar, Lien and Hardaker (2014) and Colombi *et al.* (2014). The model represents a very generalized form of a stochastic frontier model for panel data and it has caught some attention (see, e.g., Tsionas and Kumbhakar, 2014; Filippini and Greene, 2016; Badunenko and Kumbhakar, 2016). In the cost function framework it can be written as:

$$y_{it} = x'_{it}\beta + \varepsilon_{it} = x'_{it}\beta + \eta_i + u_{it} + \alpha_i + v_{it} \tag{1}$$

where y_{it} is the cost (in logs), x'_{it} is a k-element vector of independent variables (logs of prices, outputs etc.), β is a vector of model parameters, i ($i = 1, \ldots, n$) and t ($t = 1, \ldots, T$) are object and time indices. The composed error ε_{it} contains: i) two types of symmetric disturbances (α_i, v_{it}), one "standard" random disturbance v_{it} (with mean 0 and standard deviation σ_v), one firm-specific random-effect α_i to reflect firm heterogeneity (mean 0 and standard deviation σ_α); and ii) two types of nonnegative disturbances (η_i, u_{it}), one transient inefficiency u_{it} (mean 0 and standard deviation σ_u), one firm-specific persistent inefficiency η_i (mean 0 and standard deviation σ_η); traditionally based on these two we can acquire efficiency measures via direct transformation: efficiency=exp(-inefficiency) where efficiency $\in (0, 1]$. Special cases (simplifications) of the composed error term ε_{it} lead to models, which are already well known in the literature (see, e.g., Colombi, Martini and Vittadini, 2011; for a discussion). The stochastic components in ε_{it} are, in principle, statistically identifiable. Intuitively, however, it can be virtually impossible to, e.g., obtain good estimates of α_i, if variance of v_{it} is high and the other way around. Furthermore, variances of symmetric disturbances, α_i and v_{it}, also impact our ability to make precise inference about inefficiency component.

So far classical approach introduced by Filippini and Greene (2016) seems like a preferred choice to estimate GTRE models, no doubt due to significant difficulties with numerical implementation of the Bayesian model proposed by Tsionas and Kumbhakar (2014). In this paper we show that Bayesian inference can be successfully applied to GTRE models. Moreover, the sampling procedure for the modified model has good numerical properties and is relatively easy to implement.

The remaining part of the paper is as follows. In Section 2 the Bayesian model by Tsionas and Kumbhakar (2014) is discussed. We show that the model is based on very restrictive assumptions as regards prior efficiency distribution. This is likely to be the underlying cause of numerical problems Tsionas and Kumbhakar (2014) have faced when trying to sample from the posterior using "naive" Gibbs sampling (i.e., standard Gibbs sampling using the original model without re-parameterizations). We present how the Bayesian GTRE model can be modified in order to provide

more reasonable priors for persistent and transient efficiencies. We also present other possible variations of Bayesian GTRE models based on popular normal-half-normal and normal-exponential stochastic frontier (SF) specifications. In Section 3 we perform a series of simulations showing that the new Bayesian GTRE model outperforms its predecessors. We also discuss cases of very "noisy" datasets, where SF models find it difficult to yield satisfactory results and show that in all cases considered the new model is more reliable. In Section 4 we present an empirical study where we compare and contrast the results acquired using, among others, the newly constructed model with the findings in Tsionas and Kumbhakar (2014). Section 5 concludes with a discussion.

2 The modified Bayesian GTRE model

Let $\theta = (\beta, \sigma_v, \sigma_u, \sigma_\eta, \sigma_\alpha, u, \eta, \alpha)$ be a vector of structural parameters $(\beta, \sigma_v, \sigma_u, \sigma_\eta, \sigma_\alpha)$ and latent variables (u, η, α). We assume that all random components are independent of each other and of x_{it}. Independency assumption between η_i and u_{it} comes as a natural consequence of the two measuring two different dimensions of efficiency variation: η_i captures persistent cross-sectional efficiency variation between firms while u_{it} captures its fluctuation in time. We start with the model proposed by Tsionas and Kumbhakar (2014):

$$\begin{aligned} & p(\beta) p\left(\sigma_v^{-2}\right) p\left(\sigma_\alpha^{-2}\right) p\left(\sigma_u^{-2}\right) p\left(\sigma_\eta^{-2}\right) \\ & \quad \times \prod_{i=1}^{n} \prod_{t=1}^{T} f_N\left(y_{it} | x'_{it}\beta + \alpha_i + \eta_i + u_{it}, \sigma_v^2\right) f_N\left(\alpha_i | 0, \sigma_\alpha^2\right) \\ & \qquad f_N^+\left(\eta_i | 0, \sigma_\eta^2\right) f_N^+\left(u_{it} | 0, \sigma_u^2\right) \end{aligned} \tag{2}$$

where $f_N(.|a, c^{-1})$ denotes density function of the Normal distribution with mean a and precision c, $f_N^+(.|a, c^{-1})$ denotes density function of the half-Normal distribution, based on the Normal distribution with mean a and precision c. Informative prior on β is $p(\beta) = f_N\left(\beta \mid b, C^{-1}\right)$ with k-element vector b of prior mean and a k-by-k prior precision matrix C. Of course, a standard uninformative reference prior on β can be used if there is need. We focus our attention on priors on the variance components – $p\left(\sigma_v^{-2}\right) p\left(\sigma_\alpha^{-2}\right) p\left(\sigma_u^{-2}\right) p\left(\sigma_\eta^{-2}\right)$. In Tsionas and Kumbhakar (2014) we have that prior on the inverse variance σ_j^{-2}, i.e. precision, is $\sigma_j^{-2} Q_j \sim \chi^2(N_j)$, and that $Q_j = 10^{-4}, N_j = 1$ for $j = v, u, \eta, \alpha$. Alternatively we can rewrite this as $p\left(\sigma_j^{-2}\right) = f_G(\sigma_j^{-2} | 0.5 \cdot N_j, 0.5 \cdot Q_j)$, where $f_G(.|w, z)$ is the density function of the gamma distribution with mean $\frac{w}{z}$ and variance $\frac{w}{z^2}$. This formulation leads to a quite informative prior on the symmetric disturbances. The resulting marginal prior on, e.g., α_i is $p(\alpha_i) = f_S(\alpha_i | 1, 0, 10^4)$ where $f_S(.|v, a, c)$ is a Student's t-distribution with degrees of freedom v, location a, and precision c. Thus we have a considerably high precision of the marginal prior on α_i. The reader may find much less informative

priors on the symmetric disturbances in the Bayesian stochastic frontier literature (see, e.g., van den Broeck *et al.*, 1994). Our preliminary analysis have indicated that based on such less informative priors the model is unlikely to indicate any presence of symmetric individual effects. However, since the prior proposed by Tsionas and Kumbhakar (2014) does not seem to dominate the posterior (at least not within reasonable limits of σ_j variation; see *extreme cases* analysis, Section 3) we leave this prior unchanged. This comes with a note that by using this prior the researcher puts a rather strong, though not unreasonable, belief in the existence of symmetric individual effects (α_i).

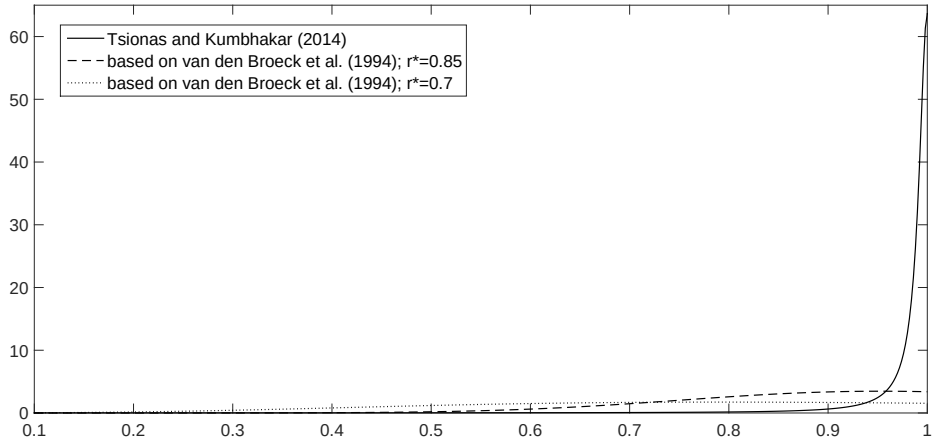
More importantly, however, the same prior also has been used for σ_u^{-2} and σ_η^{-2}, which yields very restrictive and unreasonable marginal priors on (in)efficiency components. In fact the median of the marginal prior distribution of efficiency in this model is about 0.99, quantile 0.25 is 0.976, quantile 0.75 is 0.996, the interquartile range (IQR) is only around 0.02 and the 95% highest prior density interval is (0.881,1]; corresponding characteristics of the marginal prior of inefficiency are: median=0.01, quantile(0.25)= 0.004, quantile(0.75)= 0.024, 95% highest prior density interval is about [0,0.127) (results acquired numerically via direct sampling from the prior). Clearly, this very tight informative prior may be strongly against information in the data leading to a very irregular (e.g., multimodal) posterior, which is difficult to sample from. Van den Broeck *et al.* (1994) have already discussed the problem of efficiency distribution and prior elicitation for model-specific parameters in SF models. Hence, following van den Broeck *et al.* (1994: pp. 286-7) we propose different priors on σ_u^{-2} and σ_η^{-2} in order to better reflect the initial knowledge about efficiency. For transient inefficiency, half-Normal distribution of u_{it} given precision τ, i.e. $p\left(u_{it} \mid \tau\right) = 2f_N^+\left(u_{it} \mid \tau\right) f_G\left(\tau \mid \frac{v_0}{2}, \frac{b}{2}\right)$, leads to a half-Student t marginal prior on u_{it}, i.e. $p\left(u_{it}\right) = 2f_S(u_{it}|v_0, 0, \frac{v_0}{b})I(u_{it} \geq 0)$. Hence, using z that follows an untruncated Student t distribution with location 0, scale 1 and degrees of freedom v_0 we can elicit quantiles (r_u^*) for the marginal prior transient efficiency. In particular, prior median corresponds to 0.75 quantile of z. Taking r_u^* equal to the prior median efficiency gives $z_{0.75}(v_0) = -\ln r_u^*/\tau_0$ and since $z_{0.75}(v_0)$ does not change much for $v_0 > 5$ we can set $v_0 = 10$, which yields $\tau_0^2 \approx 2\ln^2 r_u^*$ and $b = v_0\tau_0^2 \approx 20\ln^2\left(r_u^*\right)$; see van den Broeck *et al.* (1994). The same elicitation applies to η_i. This leads to the following modified Bayesian GTRE model:

$$\begin{aligned} & p\left(\beta\right) p\left(\sigma_v^{-2}\right) p\left(\sigma_\alpha^{-2}\right) f_G\left(\sigma_u^{-2}|5, 10\ln^2\left(r_u^*\right)\right) f_G\left(\sigma_\eta^{-2}|5, 10\ln^2\left(r_\eta^*\right)\right) \\ & \times \prod_{i=1}^{n}\prod_{t=1}^{T} f_N\left(y_{it}|x'_{it}\beta + \alpha_i + \eta_i + u_{it}, \sigma_v^2\right) f_N\left(\alpha_i|0, \sigma_\alpha^2\right) \\ & \qquad f_N^+\left(\eta_i|0, \sigma_\eta^2\right) f_N^+\left(u_{it}|0, \sigma_u^2\right) \end{aligned} \tag{3}$$

The reader should note that in fact this formulation does not represent one particular prior distribution but rather a range of possible priors more or less informative dependently on r^*. For this reason, the new model requires additional hyper-

parameters, r_u^* and r_η^* (prior medians), to be specified for transient and persistent efficiency. Since it seems intuitive to expect that a greater portion (if not all) of observed inefficiency is due to persistent differences between objects we may set, e.g., $r_u^* = 0.85$ and $r_\eta^* = 0.7$ as we do in simulation experiments (Section 3). This can be also interpreted that *a priori* we give slightly more chances for persistent inefficiency to exist in relation to transient efficiency, which captures time-varying component of total inefficiency variation. Prior simulation leads to the following characteristics of marginal priors for transient and persistent efficiency distribution: i) transient efficiency has median= 0.85, quantile(0.25)= 0.755, quantile(0.75)= 0.927, IQR= 0.172, mean=0.83; std.=0.122, and 95% highest prior density interval is (0.599,1]; 99% is (0.482,1]; persistent efficiency has median=0.7, quantile(0.25)=0.54, quantile(0.75)= 0.848, IQR= 0.308, mean= 0.683, std.= 0.2, and 95% highest prior density interval is $(0.325, 1]$; 99% is $(0.202, 1]$. Figure 1 shows plots of prior efficiency distributions for the two above prior medians and the prior proposed by Tsionas and Kumbhakar (2014: 3.2). It is now obvious that the proposed modification provides much more reasonable prior on efficiency, which can be additionally fine-tuned to better fit different applications (i.e., since we control location parameter of the prior efficiency we can easily try different values of r^*).

Figure 1: Prior efficiency distributions



Similarly to Tsionas and Kumbhakar (2014) conditional distributions are relatively straightforward to derive in this model and Gibbs sampling procedure can be used. We start with the conditional for a k-element vector β of the cost function parameters:

$$p\left(\beta \mid y, X, \theta_{-\beta}\right) = f_N^k \left(\left(C + \sigma_v^{-2} X'X\right)^{-1} \left(Cb + \sigma_v^{-2} X'\tilde{y}\right), \left(C + \sigma_v^{-2} X'X\right)^{-1} \right) \quad (4)$$

or in case of a reference uniform prior:

$$p\left(\beta \mid y, X, \theta_{-\beta}\right) = f_N^k\left(\left(X'X\right)^{-1}\left(X'\tilde{y}\right), \sigma_v^2\left(X'X\right)^{-1}\right) \tag{5}$$

where $\tilde{y} = y - \iota_T \bigotimes \alpha - \iota_T \bigotimes \eta - u$. For precision parameters σ_v^{-2} and σ_α^{-2} the conditionals are:

$$p\left(\sigma_v^{-2}\left(Q_v + \tilde{v}'\tilde{v}\right) \mid y, X, \theta_{-\sigma_v}\right) = f_{\chi^2}\left(\sigma_v^{-2}\left(Q_v + \tilde{v}'\tilde{v}\right)|nT + N_v\right) \tag{6}$$

$$p\left(\sigma_\alpha^{-2}\left(Q_\alpha + \alpha'\alpha\right) \mid y, X, \theta_{-\sigma_\alpha}\right) = f_{\chi^2}\left(\sigma_\alpha^{-2}\left(Q_\alpha + \alpha'\alpha\right)|n + N_\alpha\right) \tag{7}$$

where $\tilde{v}{=}y{-}X\beta{-}\iota_T{\bigotimes}\alpha{-}\iota_T{\bigotimes}\eta{-}u$, $Q_v{=}Q_\alpha{=}10^{-4}$, $N_\alpha{=}N_v{=}1$ and $f_{\chi^2}(.|\mathrm{s})$ denotes the χ^2 density function with s degrees of freedom. Conditionals σ_u^{-2} and σ_η^{-2} are:

$$p\left(\sigma_u^{-2} \mid y, X, \theta_{-\sigma_u}\right) = f_G\left(\sigma_u^{-2}|\frac{nT}{2} + 5, \frac{u'u}{2} + 10\ln^2\left(r_u^*\right)\right) \tag{8}$$

$$p\left(\sigma_\eta^{-2} \mid y, X, \theta_{-\sigma_\eta}\right) = f_G\left(\sigma_\eta^{-2}|\frac{n}{2} + 5, \frac{\eta'\eta}{2} + 10\ln^2\left(r_\eta^*\right)\right) \tag{9}$$

Moving on to latent variables, the conditional for an nT-element vector of transient inefficiencies is:

$$p(u|y, X, \theta_{-u}) \propto f_N^{nT}\left(u|\frac{\sigma_u^2}{\sigma_v^2 + \sigma_u^2}\tilde{u}, \frac{\sigma_v^2\sigma_u^2}{\sigma_v^2 + \sigma_u^2}I_{nT}\right) I\left(u \in R_+^{nT}\right) \tag{10}$$

where $\tilde{u} = y - X\beta - \iota_T \bigotimes \alpha - \iota_T \bigotimes \eta$. The reader should note that I_{nT} is an nT-by-nT identity matrix and that $I\left(u \in R_+^{nT}\right)$ truncates the normal distribution to only nonnegative values of u_{it}. This implicates that $f_N^{nT}(.|b, C^{-1})I(u \in R_+^{nT})$ is an nT-dimension truncated normal distribution function (with mean vector b and diagonal precision matrix C for the underlying untruncated distribution). This is a slightly different conditional than the one reported in Tsionas and Kumbhakar (2014; p. 119). Our analytical derivations have shown, however, that this is the appropriate formula for the conditional on u_{it} in the half-normal case; compare van den Broeck *et al.* (1994; p. 281), Makieła (2014; p. 198) and Tsionas (2002: 3.7). For n-element vector of persistent inefficiencies we have:

$$p\left(\eta|y, X, \theta_{-\eta}\right) \propto f_N^n\left(\eta|\frac{\sigma_\eta^2}{\frac{\sigma_v^2}{T} + \sigma_\eta^2}\widetilde{\eta}, \frac{\frac{\sigma_v^2\sigma_\eta^2}{T}}{\frac{\sigma_v^2}{T} + \sigma_\eta^2}I_n\right) I\left(\eta \in R_+^n\right) \tag{11}$$

where $\widetilde{\eta} = \overline{y} - \overline{X}\beta - \alpha - \overline{u}$ and symbol "$-$" denotes an n-element vector of n firm-wise averages for y, X, and u. The last but not least is the conditional for an n-element

vector of firm-specific random effects (α_i):

$$p\left(\alpha \mid y, X, \theta_{-\alpha}\right) = f_N^n\left(\alpha|\frac{\sigma_\alpha^2}{\frac{\sigma_v^2}{T}+\sigma_\alpha^2}\widetilde{\alpha}, \frac{\frac{\sigma_v^2\sigma_\alpha^2}{T}}{\frac{\sigma_v^2}{T}+\sigma_\alpha^2} I_n\right) \tag{12}$$

where this time $\widetilde{\alpha} = \overline{y} - \overline{X}\beta - \eta - \overline{u}$. Although the changes may seem small they are in fact groundbreaking. Unlike in Tsionas and Kumbhakar (2014), a straightforward (or "naive" as the authors refer to it) Gibbs sampling procedure constructed based on (4)-(12) has very good mixing properties.

The above discussed specification is a modification of the model proposed by Tsionas and Kumbhakar (2014), which generalizes normal-half-normal SF type models (Aigner, Lovell and Schmidt, 1977). Alternatively one can also wish to build a different Bayesian GTRE model, e.g., by generalizing the normal-exponential SF model proposed by Meeusen and van den Broeck (1977), i.e.,:

$$\begin{aligned} &p\left(\beta\right) p\left(\sigma_v^{-2}\right) p\left(\sigma_\alpha^{-2}\right) f_G\left(\lambda_u^{-1}|1, -\ln\left(r_u^*\right)\right) f_G\left(\lambda_\eta^{-1}|1, -\ln\left(r_\eta^*\right)\right) \\ &\times \prod_{i=1}^{n}\prod_{t=1}^{T} f_N\left(y_{it}|x'_{it}\beta + \alpha_i + \eta_i + u_{it}, \sigma_v^2\right) f_N\left(\alpha_i|0, \sigma_\alpha^2\right) \\ &\qquad f_G\left(\eta_i|1, \lambda_\eta^{-1}\right) f_G\left(u_{it}|1, \lambda_u^{-1}\right) \end{aligned} \tag{3a}$$

which leads to the following changes in equations (8)-(11) for conditionals on σ_u^{-2}(now :λ_u^{-1}), σ_η^{-2}(now :λ_η^{-1}), u_{it} and η_i:

$$p\left(\lambda_u^{-1} \mid y, X, \theta_{-\lambda_u}\right) = f_G\left(\lambda_u^{-1}|nT+1, u'\iota_{nT} - \ln\left(r_u^*\right)\right) \tag{8a}$$

$$p\left(\lambda_\eta^{-1} \mid y, X, \theta_{-\lambda_\eta}\right) = f_G\left(\lambda_\eta^{-1}|n+1, \eta'\iota_n - \ln\left(r_\eta^*\right)\right) \tag{9a}$$

$$p(u|y, X, \theta_{-u}) \propto f_N^{nT}\left(u|\tilde{u} - \sigma_v^2\lambda_u^{-1}\iota_{nT}, \sigma_v^2 I_{nT}\right) I\left(u \in R_+^{nT}\right) \tag{10a}$$

$$p(\eta|y, X, \theta_{-\eta}) \propto f_N^n\left(\eta|\widetilde{\eta} - \frac{\sigma_v^2\lambda_\eta^{-1}}{T}, \frac{\sigma_v^2}{T} I_n\right) I\left(\eta \in R_+^n\right) \tag{11a}$$

Furthermore, one can mix the different distributional assumptions about inefficiency terms in (3) and (3a), and construct "mixed" GTRE models. For example, we can assume η_i to be half-normal and u_{it} to be exponential:

$$\begin{aligned} &p\left(\beta\right) p\left(\sigma_v^{-2}\right) p\left(\sigma_\alpha^{-2}\right) f_G\left(\sigma_\eta^{-2}|5, 10\ln^2\left(r_\eta^*\right)\right) f_G\left(\lambda_u^{-1}|1, -\ln\left(r_u^*\right)\right) \\ &\times \prod_{i=1}^{n}\prod_{t=1}^{T} f_N\left(y_{it}|x'_{it}\beta + \alpha_i + \eta_i + u_{it}, \sigma_v^2\right) f_N\left(\alpha_i|0, \sigma_\alpha^2\right) \\ &\qquad f_N^+\left(\eta_i|0, \sigma_\eta^2\right) f_G\left(u_{it}|1, \lambda_u^{-1}\right) \end{aligned} \tag{3b}$$

or the other way around:

$$p(\beta)\, p\left(\sigma_v^{-2}\right) p\left(\sigma_\alpha^{-2}\right) f_G\left(\lambda_\eta^{-1}|1, -\ln\left(r_\eta^*\right)\right) f_G\left(\sigma_u^{-2}|5, 10\ln^2\left(r_u^*\right)\right) \times \prod_{i=1}^{n}\prod_{t=1}^{T} f_N\left(y_{it}|x'_{it}\beta + \alpha_i + \eta_i + u_{it}, \sigma_v^2\right) f_N\left(\alpha_i|0, \sigma_\alpha^2\right) f_G\left(\eta_i|1, \lambda_\eta^{-1}\right) f_N^+\left(u_{it}|0, \sigma_u^2\right) \tag{3c}$$

Construction of Gibbs sampler for such models only amounts to substituting blocks of posterior conditional distributions provided earlier.
On condition that we reasonably set the hyper-parameters (r_u^*, r_η^*), assigning an exponential distribution to η_i and u_{it}, or "mixing" the two SF types also leads to models with good properties as regards their computational applications. Moreover, in practice assigning different types of priors can potentially make it easier to separate the two inefficiency terms (perhaps due to slightly better identification of latent variables, η_i in particular).

3 Results based on simulation experiments

In order to analyse behaviour of the Gibbs sampler for the new half-normal GTRE model in (3) we generate datasets similar to the ones in Tsionas and Kumbhakar (2014: 4.2). Specifically, we set the number of observations as $n = 100$ and number of time periods as $T = 10$. We have a constant term and a covariate that is generated as independent standard normal and we set $\sigma_v = 0.1$, $\sigma_u = 0.2$, $\sigma_\alpha = 0.2$, $\sigma_\eta = 0.5$. The starting values are equal to the true parameter values (we would also initiate the sampler from the prior means to ensure that the results are not dependent on the starting points, i.e., too short burn-in phase). We run 150,000 Gibbs iterations, the first 50,000 being discarded. Following Tsionas and Kumbhakar (2014) proposition we then take every tenth draw to decrease autocorrelation in the chain and then calculate the posterior characteristics of model parameters and latent variables. The reader should note, however, that according to O'Hagan (1994) information about posterior characteristics of the model based on the full MCMC chain will always be higher than information based on any of its sub-chains. Even if autocorrelation between subsequent MCMC states is very high, a new state always yields some additional new information about the posterior. For this reason in the empirical application we use the whole MCMC chain. The last thing left to determine is the prior on β. Tsionas and Kumbhakar (2014) discuss both, informative as well as uninformative priors and note that they use an informative prior in their applications (with $b = 0_{k\times 1}$ and $C = 10^{-4}I_k$). Our preliminary results have indicated that numerically the biggest obstacle in using a "naive" Gibbs sampler (i.e., the standard Gibbs sampler for the original parametrization) for model in (2) is the prior on the intercept. If the prior is very informative (has very tight distribution around the true value) then "naive" Gibbs handles very well. This, however, is not a reasonable

assumption and once we move towards less informative prior we run into numerical difficulties when trying to sample from the posterior. For this reason we have decided to use the reference (uninformative) prior on β in our simulation experiments because numerically it represents the most challenging case for the Gibbs sampler to handle (in the empirical study we return to informative prior on β). Also, unlike Tsionas and Kumbhakar (2014: 4.2) we do not "re-generate" datasets of the same characteristics in this section (e.g., datasets generated M-times using the same values of T, n, β, and σ_j's). When estimating such M-times generated datasets (generated using the same data generating process, DGP hereafter) we have found that for a numerically stable sampling procedure with long MCMC runs the posterior characteristics exhibit hardly any differences showing that M different posteriors are very similar and centred around true values, even when MCMC chain autocorrelation is very high. For this reason we have decided to generate several datasets of slightly different characteristics each time (slightly different DGP) and use long MCMC runs. This has also allowed us to explore samplers' mixing properties under different conditions (experiments based on datasets re-generated 100 times are provided in Table 11 but are not discussed here). We do find particularly important, however, to check if the simulated values of the stochastic components (u, η, α, v) and explanatory variables (in X) that we generate are not "accidentally" empirically correlated. This could have some impact and incidentally change the posterior characteristics of the model. Fortunately none of the datasets we generated exhibits this problem. Numerical properties of the Gibbs sampler (stability, mixing speed etc.) have been monitored using cusum path plots (Yu and Mykland, 1998), multivariate potential scale reduction factors (MPSRF; see Brooks and Gelman, 1998) and autocorrelation functions. All datasets discussed in this section have been generated in MATLAB with restarted random number generator (zero seed), which allows their replication. Additional simulations have been made using randomized datasets (random seed) to check if the results are stable. Tables 1-3 show experiment results for Gibbs samplers constructed for 5 types of models:

1. Bayesian GTRE model based on equation (3) – labelled "new GTRE",

2. Bayesian GTRE model based on equation (2) and reparametrized as proposed in Tsionas and Kumbhakar (2014) – labelled "TK GTRE". We use slightly different conditionals for δ_i and u_{it} than the ones reported in the original paper. For $\delta_i = \alpha_i + \eta_i$ we take

$$\exp\left(-\frac{(R_i - \delta_i \iota_T)'(R_i - \delta_i \iota_T)}{2\sigma_v^2} - \frac{\delta_i^2}{2\sigma_\delta^2}\right) \Phi\left(\lambda_\delta \frac{\delta_i}{\sigma_\delta}\right),$$

where $\lambda_\delta = \frac{\sigma_\eta}{\sigma_\alpha}$, $R_i = [R_{i1}, \ldots, R_{iT}]'$, and $R_{it} = y_{it} - x_{it}'\beta - u_{it}$; for u_{it} see equation 10. Analytical derivation indicates that these are appropriate conditionals for this model.

3. Bayesian stochastic frontier true random-effects model, acquired as a simplification of model in (1) so that $\varepsilon_{it} = u_{it} + \alpha_i + v_{it}$ – labeled "TRE",

4. standard Bayesian SF model, which is a simplification of model in (1) so that $\varepsilon_{it} = u_{it} + v_{it}$ (see, e.g., Koop, Osiewalski and Steel, 1999, 2000a; Makieła 2009, 2014) – labelled "standard SF",

5. Bayesian GTRE model based on equation (2) with no re-parameterization (i.e., based on "naive" Gibbs sampling) – labelled "original TK GTRE".

For models in 3) and 4) we set $r^* = 0.7$ throughout the study (this also applies to the models used in the empirical study). Following propositions in Greene (2005, 2008) we have reported results for true random-effects model (*TRE*). This model, however, does not perform as well as a *standard SF* in identifying overall inefficiency ($\omega_{it} = \eta_i + u_{it}$) and thus we do not report its results further in this section. We return to this model in the empirical section where we show that *TRE* inefficiency estimates are closely related to transient inefficiency from *GTRE*. The reader should note that our report on simulation experiments could be extended by 3 other Bayesian GTRE models proposed in (3a)-(3c). However, we have found that results obtained with these models are very similarly to the model in (3), which also indicates that Bayesian inference in stochastic frontier models is not determined by the prior structure (which is often the critique).

Table 1: Basic results for *new GTRE*, *TK GTRE*, *TRE* and *standard SF*

	True values		new GTRE		TK GTRE		TRE		standard SF	
	Value	*Std*	$E(m)$	$D(m)$	$E(m)$	$D(m)$	$E(m)$	$D(m)$	$E(m)$	$D(m)$
β_0	1		1.023	0.053	0.890	0.130	1.401	0.034	1.244	0.035
β_1	1		1.003	0.005	1.003	0.005	1.004	0.005	1.004	0.012
σ_α	0.2		0.190	0.039	0.201	0.037	0.333	0.024		
σ_η	0.5		0.481	0.059	0.628	0.098				
σ_v	0.1		0.109	0.008	0.125	0.021	0.100	0.007	0.271	0.021
σ_u	0.2		0.190	0.015	0.136	0.073	0.212	0.011	0.406	0.042
α	0.000	0.200	0.004	0.156	0.030	0.186	0.000	0.058		
η	0.408	0.274	0.387	0.162	0.537	0.200				
u	0.160	0.120	0.152	0.080	0.109	0.090	0.167	0.080	0.325	0.176
ω	0.569	0.297	0.539	0.178	0.646	0.229	0.167	0.080	0.325	0.176
MPSRF			1.0244		1.3969		1.0143		1.0019	
Time			152		1207		103		85	

Note: β_0 is the intercept; β_1 is slope parameter; *Std* is the standard deviation calculated based on true values; $E(m)$ is posterior mean of m; $D(m)$ is posterior standard deviation of m; for α_i, η_i, u_{it} and ω_{it} we report an average of individual posterior mean and an average of individual posterior standard deviation; MPSRF is multivariate potential scale reduction factor; time is simulation duration given in seconds. *Source*: author's calculations.

As we can see in Table 1 Gibbs sampler for the *new GTRE* model handles better

Table 2: Basic results; correlations between posterior means and true values of latent variables; posterior means and standard deviations of the correlation coefficient

	new GTRE			TK GTRE			naive GTRE		
	$\rho_{\hat{m}}$	$\overline{\rho}_m$	$D(\rho_m)$	$\rho_{\hat{m}}$	$\overline{\rho}_m$	$D(\rho_m)$	$\rho_{\hat{m}}$	$\overline{\rho}_m$	$D(\rho_m)$
α	0.544	0.313	0.089	0.539	0.198	0.110	0.558	0.535	0.056
η	0.800	0.658	0.071	0.805	0.664	0.064	0.796	0.088	0.190
u	0.752	0.528	0.040	0.752	0.387	0.209	0.752	0.497	0.061
ω	0.792	0.647	0.061	0.690	0.637	0.064	0.487	0.252	0.113
	TRE			standard SF					
	$\rho_{\hat{m}}$	$\overline{\rho}_m$	$D(\rho_m)$	$\rho_{\hat{m}}$	$\overline{\rho}_m$	$D(\rho_m)$			
α	0.555	0.550	0.012						
η									
u	0.752	0.569	0.030	0.283	0.194	0.029			
ω	0.320	0.242	0.030	0.781	0.534	0.052			

Note: $\rho_{\hat{m}}$ is the correlation coefficient between posterior mean of "m" ($\hat{m}$) and true value of "m", i.e., $\rho_{\hat{m}} = \rho_m(\hat{m}, m_{true})$; $\overline{\rho}_m$ is the average value of the correlation coefficient between true value of "m" and each draw from the simulation, i.e., $\overline{\rho}_m = \frac{1}{S}\sum_{s=1}^{S} \rho_m(m^s, m_{true})$ and S is the number of draws; $D(\rho_m)$ is the standard deviation of the correlation coefficient between true value of "m" and each draw of "m" from the simulation. *Source*: author's calculations.

Table 3: Results for original TK GTRE under $Q_\eta = 10^{-4}$ and $Q_\eta = 10^{-2}$

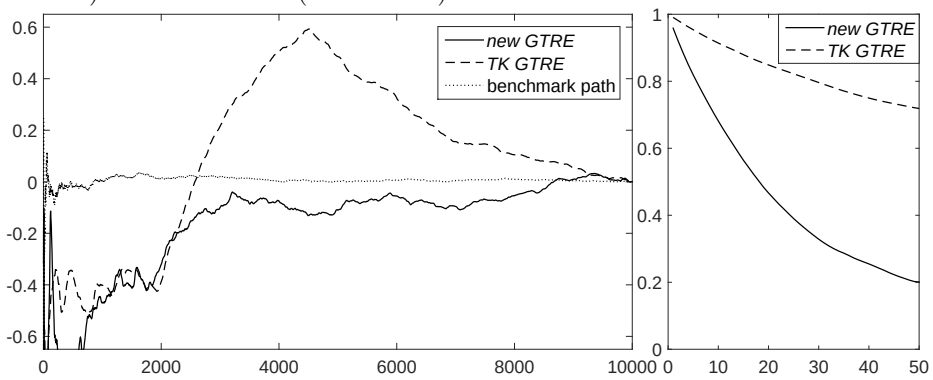
	True values		$Q_\eta = 10^{-4}$ 150 000 draws		$Q_\eta = 10^{-4}$ 300 000 draws		$Q_\eta = 10^{-2}$ 150 000 draws		$Q_\eta = 10^{-2}$ 300 000 draws	
	Value	*Std*	$E(m)$	$D(m)$	$E(m)$	$D(m)$	$E(m)$	$D(m)$	$E(m)$	$D(m)$
β_0	1		1.382	0.098	1.405	0.048	1.224	0.134	1.047	0.089
β_1	1		1.003	0.005	1.003	0.005	1.003	0.005	1.003	0.005
σ_α	0.2		0.322	0.040	0.331	0.025	0.275	0.063	0.209	0.057
σ_η	0.5		0.061	0.113	0.028	0.037	0.254	0.160	0.435	0.109
σ_v	0.1		0.115	0.011	0.115	0.010	0.115	0.011	0.110	0.009
σ_u	0.2		0.175	0.022	0.176	0.021	0.175	0.022	0.181	0.018
α	0.000	0.200	-0.001	0.091	-0.002	0.065	-0.001	0.151	0.001	0.162
η	0.408	0.274	0.049	0.106	0.022	0.039	0.203	0.180	0.348	0.177
u	0.160	0.120	0.140	0.079	0.141	0.079	0.140	0.079	0.144	0.078
ω	0.569	0.297	0.188	0.136	0.163	0.089	0.343	0.198	0.493	0.192
MPSRF			1.063		1.005		1.083		1.002	
Time			183		322		163		326	

Note: For 150 000 draws we discard first 50 thousand, for 300 thousand we discard first 100 thousand; see notes in Table 1 for notation. *Source*: author's calculations.

than *TK GTRE* in our simulations. Posterior means obtained using *new GTRE* are centred closely around the true values assumed in the DGP, while *TK GTRE* slightly overestimates σ_η and underestimates σ_u (see Table 1). More importantly, however, implementation of the new model is numerically much more efficient. The time needed to acquire results in MATLAB is nearly ten times shorter and the new sampler appears to have much better mixing properties, as measured by the multivariate potential scale reduction factor (MPRSF= 1.3969 vs. 1.0244; see Brooks

and Gelman, 1998). Other methods to compare samplers' performances (i.e. mixing speeds) are provided in Figure 2, which shows cusum path plots and autocorrelation functions of the intercept from the two simulations. We can clearly see that cusum in *new GTRE* stabilizes more quickly, has lower excursions and a more oscillatory path (less smooth) than its predecessor. Moreover, cusum path based on TK GTRE model shows a clear upward trend between 20 000 – 45 000 iteration and the MCMC autocorrelation is much higher. This means that if the *TK GTRE* sampler was to yield reliable results the simulation would require much longer MCMC runs with more iterations discarded in the burn-in stage.

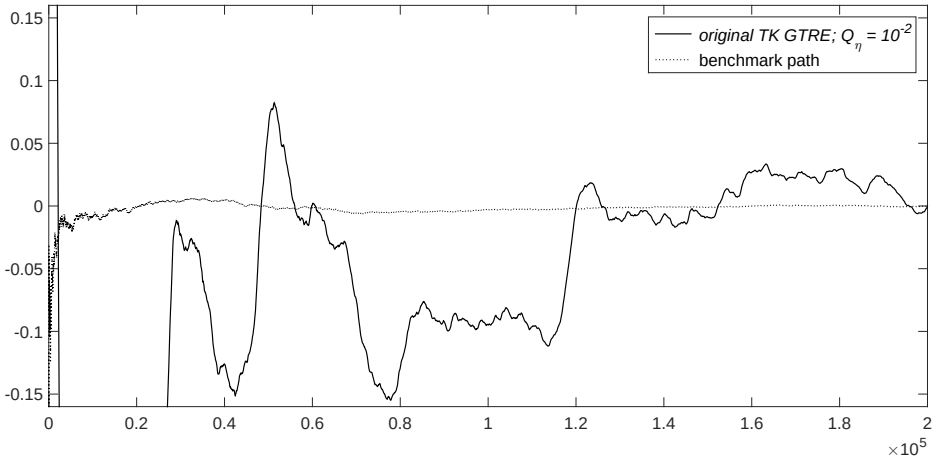
Figure 2: CUSUM path plots and autocorrelation functions for *new GTRE* model (solid line) and *TK GTRE* (dashed line)



Note: CUSUM paths (left plot) and autocorrelation function (for 50 lags, right plot) are made for the intercept based on 100 000 iterations (with every 10th taken). Solid line: *new GTRE* model; dashed line: *TK GTRE* model; dotted line (CUSUM graph only): benchmark path based on independent sampler. *Source:* author's calculations.

We now turn to simulation results from Gibbs sampler based on *original TK GTRE* with no re-parameterizations (Table 3). When we set $Q_\eta = 10^{-4}$, as in Tsionas and Kumbhakar (2014: p. 116), several marginal posteriors are nowhere near the values assumed in the simulation. The intercept estimate is too high, η_i estimate is very low and dispersion of posterior distribution of α_i is much larger than we would expect given the known DGP. Considering very tight informative prior on η_i this result should not be that surprising. In fact, once we set $Q_\eta = 10^{-2}$ and double the sampling run the marginal posterior distributions are considerably closer to values assumed in the simulation (see last column in Table 3; the reader should note that $Q_\eta = 10^{-2}$ still implicates a fairly tight informative prior with prior median efficiency about 0.9, quantile(0.25)= 0.78, quantile(0.75)= 0.96). This exercise shows that due to tight, unreasonable priors on transient and persistent efficiencies we may be dealing here with a very irregular posterior, one which is difficult to sample from (see cusum path plot in Figure 3).

Figure 3: CUSUM path plot for *original TK GTRE* model ($Q_\eta = 10^{-2}$); CUSUM path plot is for the intercept; the other (almost flat) line is a benchmark path based on independent sampler with the same mean and standard deviation



Prior medians of transient (r_u^*) and persistent (r_η^*) efficiency are additional hyper-parameters that need to be specified in the *new GTRE* model. In a standard Bayesian stochastic frontier analysis r^* is usually from 0.5-0.95 interval. Its value has only a marginal impact on the level of posterior mean efficiency in the sample and virtually no influence on relative differences in efficiency levels between observations (Makieła, 2014). Up to this point our prior assumption about transient and persistent efficiency distribution in the GTRE model was that transient efficiency is higher and less likely to exist than persistent (thus $r_u^* > r_\eta^*$). Although this seems like a reasonable assumption, we now set both prior medians equal and change them from 0.5 to 0.9. Table 4 presents estimation results for such cases. The results do not change significantly for fairly reasonable values of r_u^* and r_η^* that oscillate within 0.5-0.9 interval. Once r_u^* and r_η^* exceed 0.9 the priors on σ_u^{-2} and σ_η^{-2} become very diffused and thus the sampler's mixing speed may be low because high values of r^* (close to 1) give little prior chances that inefficiency terms exist (Koop, Steel and Osiewalski, 1995; Fernandez, Osiewalski and Steel, 1997; Ritter 1993). This also seems to be the case with the model proposed by Tsionas and Kumbhakar (2014).
The overall conclusion and recommendation for r_u^* and r_η^* does not change in relation to standard Bayesian SF models. Values for r_u^* and r_η^* should be reasonably set within 0.5-0.95 interval bearing in mind that values close to 0.95 implicate considerably tight informative prior and may cause numerical problems if information in the dataset does not support this idea. If we set highly unreasonable values for prior medians (e.g., very low prior median for transient and/or very high prior median for persistent) the results may turn out either over-optimistic or over-pessimistic with some signs of

Table 4: Simulation results for different values of r_u^* and r_η^* in new GTRE model

	True values		$r_u^* = r_\eta^* = 0.5$			$r_u^* = r_\eta^* = 0.6$			$r_u^* = r_\eta^* = 0.7$		
	Value	Std	$E(m)$	$D(m)$	$\rho_{\hat{m}}$	$E(m)$	$D(m)$	$\rho_{\hat{m}}$	$E(m)$	$D(m)$	$\rho_{\hat{m}}$
β_0	1		0.914	0.047		0.969	0.049		1.029	0.055	
β_1	1		1.007	0.006		1.006	0.005		1.005	0.005	
σ_α	0.200		0.161	0.033		0.178	0.035		0.201	0.040	
σ_η	0.500		0.600	0.055		0.528	0.055		0.458	0.060	
σ_ν	0.100		0.078	0.007		0.086	0.007		0.095	0.007	
σ_u	0.200		0.269	0.010		0.245	0.010		0.223	0.011	
α	0.000	0.200	0.000	0.144	0.521	0.000	0.151	0.537	0.000	0.159	0.552
η	0.408	0.274	0.449	0.153	0.806	0.409	0.158	0.803	0.364	0.164	0.799
u	0.160	0.120	0.206	0.077	0.747	0.191	0.079	0.750	0.176	0.080	0.751
ω	0.569	0.297	0.655	0.164	0.796	0.600	0.171	0.794	0.540	0.179	0.791
MPSRF			1.0467			1.0354			1.0346		
Time			226			214			246		

	True values		$r_u^* = r_\eta^* = 0.8$			$r_u^* = r_\eta^* = 0.9$		
	Value	Std	$E(m)$	$D(m)$	$\rho_{\hat{m}}$	$E(m)$	$D(m)$	$\rho_{\hat{m}}$
β_0	1		1.118	0.070		1.296	0.050	
β_1	1		1.004	0.005		1.003	0.005	
σ_α	0.200		0.242	0.047		0.313	0.028	
σ_η	0.500		0.363	0.076		0.169	0.044	
σ_ν	0.100		0.104	0.008		0.116	0.009	
σ_u	0.200		0.201	0.013		0.174	0.019	
α	0.000	0.200	-0.001	0.162	0.561	-0.001	0.112	0.558
η	0.408	0.274	0.292	0.167	0.796	0.135	0.104	0.797
u	0.160	0.120	0.159	0.080	0.752	0.139	0.078	0.752
ω	0.569	0.297	0.452	0.185	0.783	0.274	0.131	0.606
MPSRF			1.0343			1.0041		
Time			244			222		

numerical instability (poor mixing properties of the sampler). This seems especially important for persistent inefficiency. Fortunately, for reasonable-enough values of r_u^* and r_η^* we find hardly any impact on the posterior characteristics. Furthermore, the reader should note that in Bayesian GTRE models we can test different values of r_u^* and r_η^* using posterior probabilities of competing Bayesian models. Under equal prior odds we can compare competing model specifications with different prior median values or pool inference (e.g., about inefficiency terms) from them using marginal data density. Makieła (2014) shows how marginal data density can be estimated in stochastic frontier models via harmonic mean estimator with Lenk's (2009) correction.

In order to fully examine numerical efficiency (i.e., mixing speed) of the Gibbs sampler in the *new GTRE* model let us now explore other values for σ_α and σ_v in the DGP. As it has been mentioned in the introduction, practice shows that variance of α_i and v_{it} is crucial in acquiring good estimates of inefficiency components. Badunenko and Kumbhakar (2016) indicate that if variances of symmetric disturbances are relatively high it precludes our ability to satisfactory estimate transient and persistent

inefficiency terms in GTRE models. This is, however, not only obvious but also true for all kinds of SF models. The real question is if we can rely on GTRE models as much as on their simpler counterparts in such cases. Tables 5-8 report model's posterior characteristics once we increase σ_α, σ_v and both. For comparability we present results for *TK GTRE* and *standard SF*.

Table 5: Extreme case 1: estimations results when $\sigma_\alpha = 1$

	True values		new GTRE		TK GTRE		standard SF	
	Value	Std	$E(m)$	$D(m)$	$E(m)$	$D(m)$	$E(m)$	$D(m)$
β_0	1		0.940	0.131	1.438	0.125	1.186	0.109
β_1	1		1.004	0.005	1.003	0.005	1.017	0.033
σ_α	1.0		0.957	0.085	1.030	0.074		
σ_η	0.5		0.615	0.126	0.033	0.067		
σ_ν	0.1		0.109	0.008	0.123	0.019	0.991	0.037
σ_u	0.2		0.190	0.014	0.146	0.066	0.479	0.131
α	0.000	1.000	-0.010	0.356	-0.008	0.117		
η	0.408	0.274	0.488	0.355	0.026	0.068		
u	0.160	0.120	0.151	0.079	0.116	0.088	0.382	0.299
ω	0.569	0.297	0.639	0.364	0.142	0.114	0.382	0.299
MPSRF			1.0171		1.7469		1.0243	
Time			220		1082		66	

Table 6: Extreme case 2: estimations results when $\sigma_\nu = 0.8$

	True values		new GTRE		TK GTRE		standard SF	
	Value	Std	$E(m)$	$D(m)$	$E(m)$	$D(m)$	$E(m)$	$D(m)$
β_0	1		0.915	0.076	1.512	0.070	1.202	0.091
β_1	1		1.022	0.026	1.020	0.027	1.022	0.028
σ_α	0.2		0.031	0.041	0.332	0.039		
σ_η	0.5		0.567	0.064	0.030	0.048		
σ_ν	0.8		0.795	0.021	0.808	0.019	0.825	0.032
σ_u	0.2		0.252	0.064	0.036	0.056	0.459	0.109
α	0.000	0.200	0.000	0.050	0.001	0.208		
η	0.408	0.274	0.453	0.201	0.024	0.050		
u	0.160	0.120	0.201	0.161	0.029	0.060	0.366	0.277
ω	0.569	0.297	0.653	0.254	0.053	0.076	0.366	0.277
MPSRF			1.0026		1.1243		1.013	
Time			206		1052		64	

Two findings are worth noting. First, *new GTRE* handles extreme cases better than its predecessors. It is numerically more efficient and stable than *TK GTRE*, provides more accurate estimates of model parameters than both and, on average, its estimates have higher correlation with the true values of α_i, η_i, u_{it}, ω_{it} (especially when σ_v is high; see Table 8). Second, relatively high values of σ_v and σ_α make it extremely difficult to approximate inefficiency differences, regardless of the model used (GTRE or not). *New GTRE* model, however, does manage to identify average

Table 7: Extreme case 3: estimations results when $\sigma_\alpha = 1$ and $\sigma_\nu = 0.8$

	True values		new GTRE		TK GTRE		standard SF	
	Value	Std	$E(m)$	$D(m)$	$E(m)$	$D(m)$	$E(m)$	$D(m)$
β_0	1		0.840	0.172	1.536	0.119	1.201	0.108
β_1	1		1.022	0.027	1.021	0.027	1.036	0.041
σ_α	1		0.933	0.095	1.021	0.078		
σ_η	0.5		0.664	0.160	0.029	0.051		
σ_ν	0.8		0.793	0.021	0.807	0.021	1.264	0.037
σ_u	0.2		0.251	0.063	0.025	0.063	0.461	0.125
α	0.0	1.000	-0.001	0.434	-0.006	0.269		
η	0.408	0.274	0.530	0.388	0.023	0.053		
u	0.160	0.120	0.201	0.160	0.020	0.063	0.368	0.295
ω	0.569	0.297	0.731	0.421	0.043	0.082	0.368	0.295
MPSRF			1.0031		1.0284		1.0209	
Time			173		1049		67	

Table 8: Extreme cases 1-3: correlations between posterior means and true values of latent variables; posterior means and standard deviations of the correlation coefficient

	new GTRE			TK GTRE			standard SF		
	$\rho_{\hat{m}}$	$E(\rho_m)$	$D(\rho_m)$	$\rho_{\hat{m}}$	$E(\rho_m)$	$D(\rho_m)$	$\rho_{\hat{m}}$	$E(\rho_m)$	$D(\rho_m)$
	$\sigma_\alpha = 1$								
α	0.962	0.895	0.034	0.963	0.961	0.005			
η	0.239	0.091	0.095	0.278	0.005	0.102			
u	0.753	0.531	0.038	0.752	0.414	0.187	0.084	0.024	0.032
ω	0.328	0.141	0.083	0.226	0.158	0.086	0.252	0.072	0.036
	when $\sigma_\alpha = 1$								
α	0.396	0.030	0.108	0.416	0.327	0.059			
η	0.660	0.523	0.055	0.653	0.027	0.108			
u	0.132	0.024	0.031	0.143	0.004	0.032	0.108	0.013	0.031
ω	0.605	0.445	0.050	0.596	0.025	0.077	0.323	0.045	0.033
	when $\sigma_\alpha = 1$ and $\sigma_v = 0.8$								
α	0.930	0.826	0.049	0.931	0.903	0.011			
η	0.248	0.097	0.095	0.255	0.004	0.101			
u	0.140	0.025	0.032	0.125	0.002	0.032	0.060	0.013	0.031
ω	0.226	0.085	0.082	0.209	0.003	0.073	0.203	0.045	0.033

levels of posterior means for α_i, η_i, u_{it}, ω_{it} relatively well even in the most extreme scenario ($\sigma_\alpha = 1$ and $\sigma_v = 0.8$). In the same scenario inefficiency components are virtually undetected in *TK GTRE* (very low posterior estimates of σ_η, σ_u) and the model's implementation exhibits numerical instability (MPSRF= 1.7469).
In order to "help" *new GTRE* cope with low correlation in the most extreme scenario considered here ($\sigma_\alpha = 1$, $\sigma_v = 0.8$) one could try to fine-tune hyper-parameters r_u^* and r_η^* of the prior transient and persistent inefficiency. However, we have explored this concept and found that these hyper-parameters have little impact on posterior inefficiency estimates and virtually no influence as regards relative differences in inefficiency levels between observations.

Tsionas and Kumbhakar (2014: 4.1 & 4.3) have also explored other values for T, n, σ_α, σ_η, σ_v, σ_u and shorter Gibbs runs. We find that both models give good results for reasonable values of T, n, σ_α, σ_η, σ_v, σ_u. However, in all cases considered the new model numerically outperforms its predecessor. It takes significantly much less time to compute, its estimates are on average closer to the values set in the DGP and its implementation appears to be much more reliable when simulating from the posterior. The last remark becomes especially evident once we set $T = 5$ and consider more regression parameters (e.g., $k = 3$). In such datasets and comparable MCMC iterations the sampler based on *TK GTRE* significantly underestimates the intercept, overestimates σ_η and its implementation is numerically far less efficient in comparison to the new model (MPSRF= 1.2-1.3; see Table 9).

Table 9: Comparison between GTRE models; $T = 5$ and more parameters ($k = 2, 3$)

	True values		new GTRE		TK GTRE		True values		new GTRE		TK GTRE	
	$n = 100$, $T = 5$, $k = 2$						$n = 100$, $T = 5$, $k = 3$					
β_0	1		0.969	0.048	0.803	0.089	1		0.996	0.066	0.689	0.088
β_1	1		1.002	0.008	1.003	0.005	1		1.011	0.008	0.996	0.005
β_2	–		–	–	–	–	1		0.994	0.008	1.000	0.005
σ_α	0.2		0.198	0.039	0.196	0.038	0.2		0.206	0.041	0.189	0.041
σ_η	0.5		0.557	0.057	0.678	0.094	0.5		0.519	0.066	0.780	0.099
σ_ν	0.1		0.09	0.015	0.113	0.012	0.1		0.093	0.014	0.103	0.009
σ_u	0.2		0.208	0.021	0.179	0.025	0.2		0.22	0.019	0.203	0.016
α	0	0.2	0.001	0.164	0.033	0.185	0	0.2	0	0.169	0.030	0.184
η	0.418	0.319	0.437	0.171	0.587	0.199	0.411	0.301	0.411	0.177	0.672	0.200
u	0.155	0.119	0.166	0.082	0.143	0.080	0.17	0.126	0.175	0.085	0.162	0.080
ω	0.573	0.349	0.603	0.182	0.730	0.213	0.582	0.322	0.586	0.19	0.834	0.211
MPSRF			1.0541		1.330				1.0107		1.2079	
Time			13.5		111.8				13.7		114.4	

Note: Based on 15 thousand draws with initial 5 thousand discarded; example based on Tsionas and Kumbhakar (2014: p. 120).

Finally, we turn out attention to correlation coefficients between the true values of latent variables and their estimates. Tsionas and Kumbhakar (2014: 4.1) report that average correlation coefficient between η_i's and simulated $\eta_i^{(s)}$'s is 0.856 and between u_{it}'s and simulated $u_{it}^{(s)}$'s is about 0.754 (i.e., the average value of correlation coefficient between real values of latent variables η_i, u_{it} known from the DGP and each draw from the simulation $\eta_i^{(s)}, u_{it}^{(s)}$, where $s = 1, \ldots, S$ and S is the number of accepted draws; Tsionas and Kumbhakar label them "posterior means of the correlation coefficient", which they are not). Exact replication of the results based on Tsionas and Kumbhakar (2014: 4.1) is provided in Table 10. We find the correlation coefficients to be on average slightly lower for both *GTRE* models. Also, even though *GTRE* models give more in-depth analysis of efficiency, *standard SF* models still provide relatively good measures of overall inefficiency (ω_{it}). Thus, under the GTRE structure of the DGP a simple SF model is still quite useful in determining the overall efficiency ranking.

Table 10: Correlations between posterior means and true values of latent variables; posterior means and standard deviations of the correlation coefficient

	new GTRE			TK GTRE		
m	$\rho_{\hat{m}}$	$\overline{\rho}_m$	$D(\rho_m)$	$\rho_{\hat{m}}$	$\overline{\rho}_m$	$D(\rho_m)$
	$n = 100$, $T = 5$, $k = 2$					
α	0.536	0.166	0.104	0.559	0.039	0.101
η	0.808	0.676	0.057	0.824	0.760	0.055
u	0.751	0.508	0.068	0.656	0.609	0.017
ω	0.618	0.658	0.052	0.836	0.782	0.049
	$n = 100$, $T = 5$, $k = 3$					
α	0.560	0.328	0.091	0.441	0.026	0.103
η	0.830	0.670	0.077	0.810	0.730	0.068
u	0.732	0.559	0.049	0.721	0.671	0.016
ω	0.826	0.678	0.067	0.818	0.754	0.063

Note: Based on 15 thousand draws with initial 5 thousand discarded; results for two and three regression parameters ($k = 2, 3$); see notes for Table 2. *Source*: author's calculations.

Table 11: Sampling behavior of Bayes estimator in the new GTRE model

	α			η			u			ω		
	mean	median	std	mean	median	std	mean	median	std	mean	median	std
	$n = 50$, $T = 5$											
True	0.000222	0.000201	0.003129	0.403	0.400	0.046	0.159	0.158	0.007	0.562	0.561	0.048
Est.	-0.000267	0.000025	0.003167	0.490	0.479	0.069	0.179	0.176	0.018	0.669	0.561	0.072
	$n = 100$, $T = 5$											
True	-0.000071	-0.000010	0.002941	0.399	0.394	0.030	0.158	0.158	0.005	0.557	0.554	0.030
Est.	-0.000072	0.000103	0.002162	0.448	0.446	0.049	0.169	0.169	0.014	0.617	0.554	0.050
	$n = 100$, $T = 5$											
True	0.000083	0.000239	0.002886	0.405	0.403	0.028	0.399	0.398	0.014	0.804	0.805	0.031
Est.	-0.000330	-0.000060	0.002270	0.433	0.432	0.035	0.396	0.399	0.034	0.829	0.805	0.051
	$n = 100$, $T = 10$											
True	-0.000264	-0.000604	0.002990	0.396	0.394	0.025	0.400	0.399	0.009	0.796	0.794	0.026
Est.	-0.000011	-0.000002	0.000594	0.429	0.430	0.042	0.398	0.397	0.014	0.826	0.794	0.042
	σ_α			σ_η			σ_v			σ_u		
True	0.2			0.5			0.1			0.2		
	$n = 50$, $T = 5$											
Est.	0.147	0.154	0.083	0.546	0.546	0.052	0.081	0.092	0.029	0.224	0.223	0.028
	$n = 100$, $T = 5$											
Est.	0.154	0.164	0.058	0.560	0.561	0.051	0.091	0.092	0.014	0.212	0.211	0.017
True	0.1			0.5			0.1			0.5		
	$n = 100$, $T = 5$											
Est.	0.046	0.027	0.040	0.546	0.544	0.038	0.085	0.085	0.049	0.493	0.494	0.041
	$n = 100$, $T = 10$											
Est.	0.048	0.037	0.034	0.539	0.541	0.046	0.102	0.103	0.014	0.497	0.497	0.018

Note: "Est." is posterior mean; results are mean estimates calculated based on 100 datasets of the same characteristics (re-generated 100 times); simulation results based on 5000 burn-in and 5000 accepted draws; example similar to Tsionas and Kumbhakar (2014: p. 124). *Source*: author's calculations.

Table 12: Empirical results for the four models

	standard SF		TRE		GTRE		GSF	
	$E(m)$	$D(m)$	$E(m)$	$D(m)$	$E(m)$	$D(m)$	$E(m)$	$D(m)$
σ_ν	0.161	0.009	0.085	0.008	0.096	0.009	0.101	0.009
σ_u	0.271	0.020	0.190	0.017	0.162	0.019	0.159	0.018
σ_α			0.213	0.021	0.146	0.027		
σ_η					0.280	0.070	0.427	0.064
α			0.008	0.058	-0.001	0.021		
u	0.212	0.016	0.147	0.015	0.127	0.016	0.124	0.015
η					0.224	0.067	0.377	0.065
ω	0.212	0.016	0.147	0.015	0.351	0.076	0.502	0.073
$El(p_1)$	0.549	0.019	0.543	0.020	0.546	0.020	0.538	0.020
$El(p_2)$	0.401	0.012	0.378	0.012	0.376	0.011	0.384	0.012
$El(p_3)$	0.050	0.017	0.079	0.018	0.078	0.018	0.078	0.018
$El(y_1)$	0.108	0.007	0.086	0.010	0.085	0.010	0.087	0.010
$El(y_2)$	0.416	0.023	0.472	0.024	0.480	0.024	0.492	0.023
$El(y_3)$	0.216	0.018	0.213	0.020	0.213	0.020	0.198	0.021
$El(y_4)$	0.082	0.029	0.082	0.030	0.079	0.029	0.074	0.030
$El(y_5)$	0.098	0.011	0.083	0.014	0.080	0.014	0.090	0.014
TC	-0.047	0.004	-0.049	0.004	-0.049	0.004	-0.048	0.003
intercept	-0.904	1.040	0.097	0.995	-0.353	0.964	-0.096	0.972
RTS	1.086	0.008	1.068	0.013	1.069	0.013	1.063	0.012
MPSRF	1.002		1.014		1.043		1.007	

Note: $El\,(m)$ denotes cost elasticity of m; the table only provides average levels of elasticities due to space constrains; $E(m)$ and $D(m)$ are posterior mean and posterior standard deviation respectively; TC is technical change ($\partial \ln C\ /\partial t$); RTS are returns to scale; p_1 is wage rate for labor; p_2 is interest rate for borrowed funds; p_3 is price of capital; y_1 are consumer loans; y_2 are non-consumer loans; y_3 are securities; y_4 is financial equity capital; y_5 are non-traditional banking activities; see Feng and Serletis (2009) for more details.

4 Empirical application and results comparison

Empirical application is based on US banking data from 1998 to 2005 as in Feng and Serletis (2009). We use translog specification with eight input variables and a time trend (3 prices and 5 products; see notes in Table 12). Since one of the goals in this section is to compare and contrast the results obtained from the new model with those from Tsionas and Kumbhakar (2014) we also use "Group 1" from the dataset (very large banks) and focus on the main findings and differences. The reader should note, however, that the results are consistent for other bank groups as well. Tsionas and Kumbhakar (2014) find persistent inefficiency to be smaller than transient inefficiency. That is why *a priori* we do not favour any inefficiency component and set: $r_u^* = r_\eta^* = 0.8$. In order to impose economic regularity conditions (economic restrictions) on the cost function we modify the original minimally informative prior on β, i.e., we set $p\,(\beta) \propto \ f_N\left(\beta \mid b,\ C^{-1}\right) \times I_B\,(\beta)$, where $b = 0_{k\times 1},\ C = 10^{-4} I_k$. Economic regularity conditions are imposed always at the means (i.e., at variables' overall means in the dataset) and for the entire dataset through the support B of

the prior density $p(\beta)$. If met, $I_B(\beta) = 1$, zero otherwise. Each time (i.e., with each iteration of the MCMC scheme that meets the restrictions at the means) we take $l = 1000$ attempts to draw a configuration that meets restrictions globally for the entire dataset. This allows us to effectively probe the posterior even if probability of meeting the restrictions for the entire dataset at a specific draw are exceptionally low. Value of l determines a trade-off between computational speed (small l) and imposing strict global economic regularity restrictions for all data points and at each drawn configuration (large l). Exceptionally low probability of meeting economic regularity restriction for the entire dataset may occur in the MCMC scheme when we deal with translog functions with minimally informative priors and nuisance datasets, as it is the case here. A much simpler and straightforward solution would be to put a more informative prior on β, e.g., one that is based on the cost function theory. This, however, could preclude comparability with previous studies. The simulation is stopped once 100 thousand iterations are accepted, with initial 50 thousand discarded (sampler's burn-in phase).

Table 12 and Figures 4-6 compare results for four models: *GTRE,* true random-effects *(TRE), standard SF* and a *generalized SF* model, here labelled *GSF* (i.e.: $\varepsilon_{it} = u_{it} + \eta_i + v_{it}$). Similarly to Feng and Serletis (2009) we find overall annual reduction of total costs (technical progress), which is also partially in line with results from Tsionas and Kumbhakar (2014). Posterior estimates of returns to scale are between 1.063-1.086 indicating, on average, increasing returns to scale; see results for *RTS* in Table 12 for details. We also find an interesting pattern in terms of modelling inefficiency and individual effects in the analysed models. Since the *standard SF* model does not have individual effects, the posterior estimate of σ_v is (relatively) very high. Presence of a symmetric individual effect (α_i), which is quite relevant to the *TRE* specification, makes the posterior estimate of σ_v much smaller (in *TRE*) and there is also less inefficiency found there than in a *standard SF*. Furthermore, posterior standard deviation of the symmetric individual effect in the new *GTRE* is smaller in comparison to *TRE* (Figure 5), which is different to findings in Tsionas and Kumbhakar (2014). This can be attributed to very tight prior on η_i in the previous study. We have found that once we "tighten" the prior on η_i in our *GTRE* the posterior distribution of α_i also becomes more diffused (e.g., posterior standard deviation of α_i is 0.027 if prior median of η_i is set 0.9; compared to 0.021 when prior median of η_i is 0.8). Inefficiency components in *GSF* model are very similar to the ones from our *GTRE* with only persistent inefficiency being slightly higher. This difference is likely due to the lack of individual effects in *GSF*.

In general, we find inefficiency terms to be much larger than the ones reported by Tsionas and Kumbhakar (2014). The reader should note, however, that the previous model implied a very tight informative prior on efficiency centred around 0.99 value. Also, unlike in Tsionas and Kumbhakar (2014) we find that *a posteriori* persistent inefficiency distribution (η_i) is centered around considerably higher values and much more diffused than transient inefficiency (u_{it}). Thus, the resulting overall inefficiency

Figure 4: Posterior distributions of inefficiency components in the GTRE model

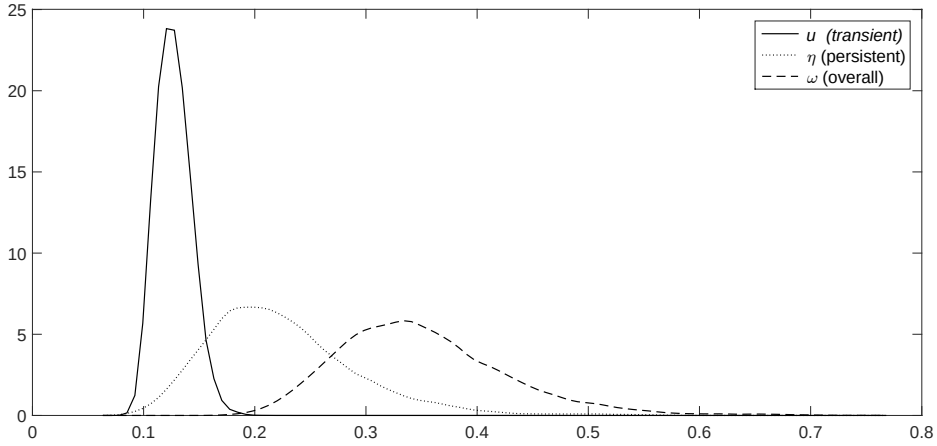
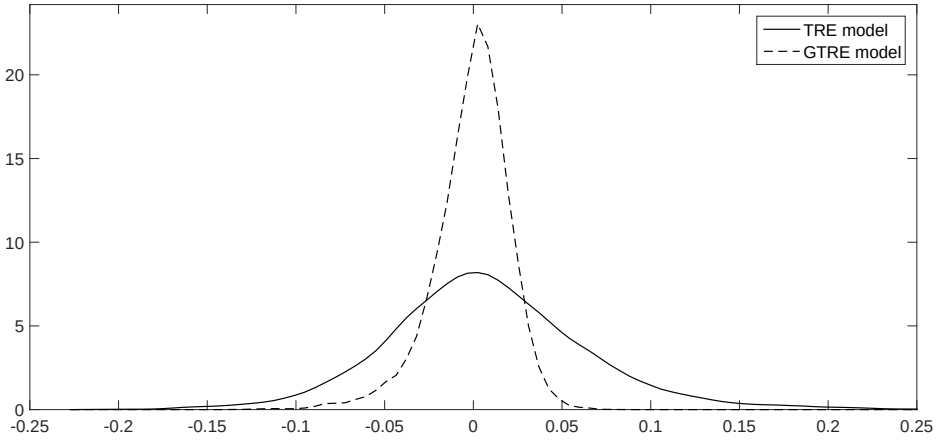
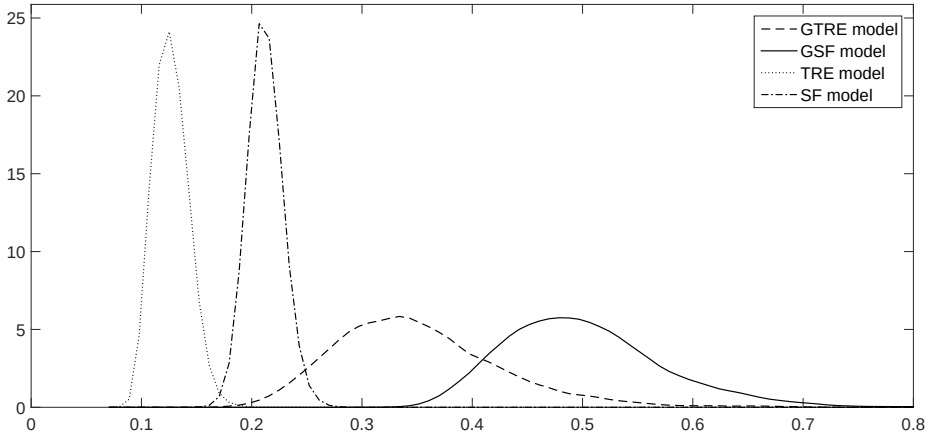


Figure 5: Posterior distribution of bank effects



scores ($\omega_{it} = u_{it} + \eta_i$) in the *GTRE* model are also considerably larger than in *TRE* or *standard SF* (see Figure 5). This would indicate that information in the data regarding persistent inefficiency is rather weak. Furthermore, inefficiency component in *TRE* model has very similar posterior characteristics to transient inefficiency from *GTRE*. Their density charts from Figures 4 and 6 nearly overlap and their posterior inefficiency rankings are almost identical (0.998 correlation between posterior means of inefficiency; see Table 13). This would indicate that inefficiency estimates that we acquire using *TRE* model should be treated as transient rather than overall inefficiency scores. Persistent inefficiency is likely captured via bank effects (α_i) in

Figure 6: Posterior distribution of overall inefficiency ω_{it} in GTRE, TRE, GSF and standard SF



the *TRE* model. We also find that posterior estimates of inefficiency scores in the *standard SF* are quite similar to overall inefficiency scores in the *GTRE* model (0.895 correlation between posterior means of inefficiency; see Table 13).

Table 13: Correlations between inefficiencies from four models

Correlation between overall inefficiency (ω)				
	standard SF	TRE	GTRE	GSF
standard SF	1	0.328 (0.039)	0.466 (0.058)	0.568 (0.040)
TRE	*0.606*	1	0.314 (0.051)	0.260 (0.044)
GTRE	*0.895*	*0.574*	1	0.638 (0.084)
GSF	*0.845*	*0.375*	*0.961*	1
Correlation between transient inefficiency (u)				
	standard SF	TRE	GTRE	GSF
standard SF	1	0.328 (0.039)	0.290 (0.042)	0.286 (0.040)
TRE	*0.606*	1	0.502 (0.058)	0.478 (0.059)
GTRE	*0.595*	*0.998*	1	0.432 (0.062)
GSF	*0.609*	*0.987*	*0.992*	1

Note: Lower triangles in the cross-tables (in *italic*) contain correlation coefficients between posterior means of inefficiencies in different models; upper triangles contain the posterior means and standard deviations (in brackets) of correlation coefficients between inefficiencies in different models (calculated using draws from simulations); for models *standard SF* and *TRE* overall inefficiency is equal to transient.

Since posterior distribution of η_i is relatively diffused and centered around significantly higher values than transient inefficiency (u_{it}) it is worth exploring how our prior belief,

expressed via prior median hyper-parameter r^*_η, influences posterior characteristics of the distribution of η_i. Sensitivity analysis provided in Table 14 indicates that for very high/low values of prior median information in the data pulls the posterior significantly away from the initially centred prior, even if the prior is relatively tight. More importantly, however, correlation coefficient of banks' persistent inefficiencies between models with prior median 0.6 and 0.9 is 0.993 (Spearman's rank correlation is 0.997). This indicates that prior median hyper-parameter of the *new GTRE* model has virtually no impact on relative differences in persistent inefficiency estimates between banks.

Table 14: Prior and posterior distribution of η_i under different prior median values

prior median	0.6		0.7		0.8		0.875		0.9	
	prior	posterior	prior	posterior	prior	posterior	prior	posterior	prior	posterior
	Efficiency distribution characteristics ($\exp(-\eta)$)									
mean	0.595	0.748	0.683	0.777	0.778	0.801	0.856	0.838	0.884	0.852
st.dev.	0.235	0.048	0.200	0.047	0.152	0.052	0.105	0.051	0.087	0.055
median	0.6	0.753	0.7	0.782	0.800	0.807	0.875	0.842328	0.900	0.850
	Inefficiency distribution characteristics (η)									
mean	0.625	0.292	0.436	0.254	0.273	0.224	0.163	0.179	0.129	0.173
st.dev.	0.513	0.065	0.358	0.061	0.224	0.067	0.134	0.063	0.106	0.066
median	0.506	0.283	0.352	0.247	0.221	0.214	0.132	0.172	0.104	0.161

To sum up, we find that inefficiency estimates (η_i, u_{it}, ω_{it}) from the GTRE model are highly correlated with their respective counterparts from simpler SF models. Whether one should consider the "full" GTRE model or one of its simplifications should be based on both, research needs at hand as well as each model's adequacy, e.g., as measured by the marginal data density. Unfortunately, despite recent breakthroughs in estimating the marginal data density (see e.g., Pajor 2016; Lenk 2009) its precise calculation in Bayesian SF models is still a challenging task. Moreover, it seems that due to a large number of latent variables GTRE models are particularly troublesome in this regard.

5 Concluding remarks

We have revisited Bayesian approach to estimating generalized true random-effects models (GTRE) and proposed some modifications to the earlier work by Tsionas and Kumbhakar (2014). Simulation exercises indicate that the new Bayesian model (and its numerical implementation) outperforms its predecessor, which is no doubt due to more reasonable priors on inefficiency terms. Both models manage to replicate true values assigned in the DGP when the conditions are favourable; that is: i) if the dataset is large-enough, ii) symmetric disturbances are relatively small in respect to inefficiency terms, and iii) we do not have that many regression parameters. Implementation of the new model, however, has better numerical properties, and thus

in more noisy datasets, or with more complex specifications (e.g., more parameters), advantages of the new model become evident. It is easier to implement, faster to compute and appears more reliable in use. Furthermore, we have proposed other types of Bayesian GTRE models (3a-c) that can also be relatively easily estimated using Gibbs sampling. All this makes Bayesian GTRE models with reasonable, interpretable priors a useful alternative to classical estimators in this field, which have only asymptotic properties.

The modification also allows us to conduct a more robust analysis because now we can easily control our prior beliefs about η_i and u_{it}. That is, we can trace how much information in the data alters the posterior in relation to the prior. This seems especially important in empirical applications of the GTRE model because we find that inefficiency terms can be considerably larger than reported in previous studies, persistent inefficiency in particular. We also find that posterior characteristics of bank-specific effects (i.e., object-specific η_i and α_i) are quite diffused and depended on the model prior specification. This, however, is to be expected since in the empirical example we deal with a relatively "short" panel ($T = 8$ and $n = 141$).

To conclude, the reader should consider the economic meaning of having a symmetric, object-specific effect (α_i) in a stochastic frontier framework. Inference about it is not only quite dependent on the priors on α_i and η_i, but also implies frontier heterogeneity to the extent where a notion of a common technical frontier, against which we benchmark our objects seems questionable (i.e., different technologies). It may be more suitable for a stochastic frontier framework to move away from the concept of purely individual "object-specific" effects in the GTRE model towards "cluster-specific" effects, (i.e., instead of α_i use α_j, $j = 1, \ldots, J$ and $J < N$) or a "cluster-specific" frontier framework (see, e.g., Koop, Osiewalski and Steel, 2000b), where one could indicate clusters (i.e., groups) of objects that operate under a common frontier.

Acknowledgements

The author acknowledges support from Foundation for Polish Science for young researchers (START 2014) and research funds granted to the Faculty of Management at Cracow University of Economics, within the framework of the subsidy for the maintenance of research potential. The author would like to thank Jacek Osiewalski, Anna Pajor and two anonymous referees for their comments and suggestions.

Supplementary materials

MATLAB codes needed to estimate models discussed in the article can be downloaded from: `http://www.mathworks.com/matlabcentral/fileexchange/61059-bayesian-gtre-models`

References

[1] Aigner D., Lovell C.K., Schmidt P. (1977). Formulation and Estimation of Stochastic Frontier Production Function Models. *Journal of Econometrics* 6(1), 21–37.

[2] Badunenko O., Kumbhakar S.C. (2016) When, Where and How to Estimate Persistent and Transient Efficiency in Stochastic Frontier Panel Data Models. *European Journal of Operational Research* 255(1): 272–287.

[3] van den Broeck J., Koop G., Osiewalski J., Steel M.F.J. (1994) Stochastic frontier models; a Bayesian perspective. *Journal of Econometrics* 61(2): 273–303.

[4] Brooks S, Gelman A. (1998) General Methods for Monitoring Convergence of Iterative Simulations. *Journal of Computational and Graphical Statistics* 7(4): 434–455.

[5] Colombi R., Martini G., Vittadini G. (2011) A Stochastic Frontier Model with short-run and long-run inefficiency random effects, No 1101, *Working Papers*, Department of Economics and Technology Management, University of Bergamo.

[6] Colombi R., Kumbhakar S.C., Martini G., Vittadini G. (2014) Closed-skew normality in stochastic frontiers with individual effects and long/short-run efficiency. *Journal of Productivity Analysis* 42(2): 123–136.

[7] Feng G., Serletis A. (2009) Efficiency and productivity of the US banking industry, 1998–2005: Evidence from the Fourier cost function satisfying global regularity conditions. *Journal of Applied Econometrics* 24(1): 105–138. DOI: 10.1002/jae.1021

[8] Fernandez C., Osiewalski J., Steel M.F.J. (1997) On the use of panel data in stochastic frontier models. *Journal of Econometrics* 79(1): 169–193.

[9] Filippini M., Greene W. (2016) Persistent and transient productive inefficiency: a maximum simulated likelihood approach. *Journal of Productivity Analysis* (45)2: 187–196. DOI: 10.1007/s11123-015-0446-y

[10] Greene W. (2005) Fixed and random effects in stochastic frontier models. *Journal of Productivity Analysis* 23(1): 7–32.

[11] Greene W. (2008) The Econometric Approach to Efficiency Analysis. In: H.O. Fried, C.A. Lovell, & S.S. Schmidt (ed) *The Measurement of Productive Efficiency and Productivity Growth*. Oxford University Press, New York, pp. 92–159.

[12] Koop G., Osiewalski J., Steel M.F.J. (1999) The Components of Output Growth: A Stochastic Frontier Analysis. *Oxford Bulleting of Economics and Statistics* 61(4): 455–487. DOI: 10.1111/1468-0084.00139

[13] Koop G., Osiewalski J., Steel M.F.J. (2000a) A Stochastic Frontier Analysis of Output Level and Growth in Poland and Western Economies. *Economics of Planning* 33(3): 185–202. DOI: 10.1023/A:1003919013358

[14] Koop G., Osiewalski J., Steel M.F.J. (2000b) Modeling the Sources of Output Growth in a Panel of Countries. *Journal of Business & Economic Statistics* 18(3): 284–299. DOI: 10.2307/1392262

[15] Koop G., Steel M.F.J., Osiewalski J. (1995) Posterior analysis of stochastic frontier models using Gibbs sampling. *Computational Statistics* 10(10): 353–373.

[16] Kumbhakar S.C., Lien G., Hardaker J. (2014) Technical efficiency in competing panel data models: a study of Norwegian grain farming, *Journal of Productivity Analysis* 41(2): 321–337.

[17] Lenk P. (2009) Simulation Pseudo-Bias Correction to the Harmonic Mean Estimator of Integrated Likelihoods. *Journal of Computational and Graphical Statistics* 18(4): 941–960. DOI: 10.1198/jcgs.2009.08022

[18] Makieła K. (2014) Bayesian Stochastic Frontier Analysis of Economic Growth and Productivity Change in the EU, USA, Japan and Switzerland. *Central European Journal of Economic Modelling and Econometrics* 6(3): 193–216.

[19] Makieła K. (2009) Economic Growth Decomposition. An Empirical Analysis Using Bayesian Frontier Approach. *Central European Journal of Economic Modelling and Econometrics* 1(4): 333–369.

[20] Meeusen W., van den Broeck J. (1977) Efficiency estimation from Cobb-Douglas Production Function with Composed Error. *International Economic Review* 18(2): 435–444.

[21] O'Hagan, A. (1994) *Kendall's Advanced Theory of Statistics Volume 2B, Bayesian Inference.* Edward Arnold, London.

[22] Pajor, A. (2016). Estimating the marginal likelihood using the arithmetic mean identity. *Bayesian Analysis*. DOI: 10.1214/16-BA1001.

[23] Ritter C. (1993) The Normal-Gamma frontier model under a common, vague prior does not produce a proper posterior. *Mimeo*. Université Catholique de Louvain, Louvain-la-Neuve.

[24] Tsionas M. (2002) Stochastic frontier models with random coefficients, *Journal of Applied Econometrics* 17(2): 127–147. DOI:10.1002/jae.637

[25] Tsionas M., Kumbhakar S.C. (2014) Firm Heterogeneity, Persistent And Transient Technical Inefficiency: A Generalized True Random-Effects model. *Journal of Applied Econometrics* 29(1): 110–132. DOI:10.1002/jae.2300

[26] Yu B., Mykland P. (1998) Looking at Markov samplers through cusum path plots: a simple diagnostic idea. *Statistical Computing* 8(3): 275–286.