Optimal Combinations of Stochastic Frontier and Data Envelopment Analysis Models

Abstract

Recent research has shown that combination approaches, such as taking the maximum or the mean over different methods of estimating efficiency scores, have practical merits and offer a useful alternative to adopting only one technique. This recent research shows that taking the maximum minimizes the risk of underestimation, and improves the precision of efficiency estimation. In this paper, we propose and implement a formal criterion of weighting based on maximizing proper criteria of model fit (viz. log predictive scoring) and show how it can be applied in Stochastic Frontier as well as in Data Envelopment Analysis models, where the problem is more difficult. Monte Carlo simulations show that the new techniques perform very well and a substantive application to large U.S. banks shows some important differences with traditional models. The Monte Carlo simulations are also substantive as it is for the first time that proper and coherent optimal model pools are subjected to extensive testing in finite samples.

Key Words: Productivity and Competitiveness; Data Envelopment Analysis; Stochastic Frontier Analysis; Efficiency Analysis; Predictive Distributions.

1 Introduction

Andor, Parmeter and Sommer (2019, APS) proposed an interesting combination approach in efficiency analysis. As they write "combination approaches, such as taking the maximum or the mean over Data Envelopment Analysis (DEA) and Stochastic Frontier Analysis (SFA) efficiency scores, have certain practical merits and might offer a useful alternative to strict reliance on a singular method. In particular, the results highlight that taking the maximum not only minimizes the risk of underestimation, but can also improve the precision of efficiency estimation" (p. 240). In effect, APS proposed the following: "An interesting extension would be to consider alternative weighting schemes for taking the average or a theoretical criterion, perhaps based on model fit, to develop weights. Future research could also focus on alternative metrics of combination, such as a median, or another quantile of interest" (p. 251).

APS build on the idea that combining results from different estimators of efficiency is preferable to relying on any single estimator. They cite earlier literature that is based on this notion although, of course, "combination approaches might be best characterized as ad hoc and without a proper theoretical foundation to rest on" (APS, p. 241). APS, in their Monte Carlo experiments find that taking the maximum of DEA and SFA estimates is the best alternative in terms of estimating individual efficiency scores. In contrast, taking the mean of DEA and SFA estimates does not perform particularly well. As they notice, "taking the maximum not only minimizes the risk of underestimation of regulated firms but also maximizes the accuracy of estimating the level of inefficiency" (op. cit., pp. 241–242).

In effect, APS consider six different models (two stochastic frontier models using the Cobb–Douglas and translog functional forms, two DEA methods using constant or variable returns to scale, and the average and maximum of efficiency scores). Another reason that "taking the max" produces good results is that all methods seemed to underestimate efficiency for most firms although this depends on the parameterization, the importance of noise and the extent of inefficiency in the data.

From the perspective of a regulatory authority (which the focus of the APS study) it turns out that combining different estimators and using Monte Carlo simulations holds considerable promise and should be a viable alternative to as to exclusive reliance on a single set of individual estimates. APS specifically "advise that for practical regulatory applications, Monte Carlo simulations should be conducted under concrete real-world conditions and considering a variety of estimation methods before deciding on the final estimation method" (op. cit. p. 251). Based on this advice, one can envision that more elaborate weighting schemes could perform well as "For example, while the arithmetic mean did not perform well in our simulations, this is a *direct* result of the naive weighting scheme that was deployed. An interesting extension would be to consider *alternative weighting schemes* for taking the average or a *theoretical criterion*, perhaps based on model fit, to develop weights" (op. cit., p. 251, emphasis added).

Consequently, APS make a good case in favor of a combination approach when it comes to an array of different efficiency estimates. Such estimates cannot be judged on their own but only i) in relation to a theoretical criterion (that provides a sense of security in terms of optimizing a certain metric), and ii) in relation to the specific data at hand. Two obvious questions suggest themselves: What to optimize and how to handle the specificity of any given data set. Although a universal answer would be too risky, it seems reasonable to proceed based on a predictive approach to take

account of the specific data. As, at least for the most part, the essence of all science is prediction, this approach is predicated on sufficiently solid ground, provided the context of prediction is articulated (that is, is it out-of-sample or in-sample prediction?). The question of the metric to optimize seems to be more difficult to answer. For example, APS make a forceful case in their Footnote 9 that bias, root mean squared error and percentage of firms whose efficiency is underestimated, are the most important metrics for regulators but they also mention that "different metrics could be used by researchers and practitioners depending upon their aims and needs from the analysis".

In this paper, *we put forward a formal selection method for weighting different estimates in a coherent framework.* The approach relies on obtaining posterior predictive distributions of each model / method and then considering a linear combination of these distributions, known as linear opinion pooling (Good, 1953, Geweke and Amisano, 2011 a, b). In turn, the optimal weights can be obtained via straightforward maximization of a criterion function.The proper criterion function (which defines what is "coherent" about this approach) in our instance is log predictive scoring, that is the log of the predictive distribution which has received attention in the forecasting literature (e.g. Geweke and Amisano, 2011 b).

Additionally, the paper contributes to the derivation and calculation of the posterior predictive distribution of (in)efficiency in SFA models but also show how the new techniques can be used in DEA-like models. Therefore, the new techniques can be used by both the SFA and DEA community to find relevant models and determine an optimal pool of models that perform well out-of-sample, no matter whether they are in the SFA, DEA or an intermediate family.

2 Background

Following Hall and Mitchell (2007) suppose the true density of data y_t is $f(y_t)$. This is, of course, unknown. However, a number of forecasters (indexed $m = 1, \ldots, M$) have their own forecast densities, say $f_{tm}(y_t)$ $(m = 1, \ldots, M)$. The question is how to aggregate these probability densities, instead of aggregating say, their means. A linear opinion pool solves this problem by taking a weighted average:

$$p_t(y_t) = \sum_{m=1}^{M} w_m p_{tm}(y_t), \tag{1}$$

where $w_m \geq 0$, $m = 1, \ldots, M$, and $\sum_{m=1}^{M} w_m = 1$. The question then becomes how to derive the optimal weights. The Kullback-Leibler information criterion is:

$$KL = \int f_t(y_t) \ln \frac{f_t(y_t)}{p_t(y_t)} \, \mathrm{d}y_t = \mathbb{E}\left[\ln f_t(y_t) - \ln p_t(y_t)\right], \tag{2}$$

where the expectation is taken with respect to the true but unknown distribution whose density is $f_t(y_t)$. If we have a sample of data, say $\{y_t\}_{t=1}^{T}$. Under broad assumptions, we have

$$KL \xrightarrow{p} T^{-1} \sum_{t=1}^{T} \left[\ln f_t(y_t) - \ln p_t(y_t)\right]. \tag{3}$$

In turn, minimizing KL is equivalent to maximizing $T^{-1}\sum_{t=1}^{T}\ln p_t(y_t)$, that is:

$$\begin{aligned}&\max_{\{w_m\}_{m=1}^{M}} T^{-1}\sum_{t=1}^{T}\ln\sum_{m=1}^{M} w_m p_{tm}(y_t),\\ &w_m\geq 0,\ m=1,\ldots,M;\ \sum_{m=1}^{M} w_m=1.\end{aligned} \tag{4}$$

From this problem it is clear that, in the forecasting literature, log scoring is well-known in the combination of forecast densities. The point of this paper is that we can apply the same idea when it comes to combination (or selection) of (in)efficiency scores as different models can be considered as different agents that deliver forecasts. The models that we will consider include both stochastic frontiers as well as data envelopment analysis models.

3 Methods

3.1 Models

Alternative models provide predictive distributions for a vector $\mathbf{y}_o\in\mathbb{R}^{d_o}$ given a set of observations

$$\mathbf{Y}=\{\mathbf{Y}_{it},\ i=1,\ldots,n,\ t=1,\ldots,T\},$$

which are, typically, inputs and outputs for certain firms (indexed by i) over time (indexed by t). Models are indexed by $m\in\mathscr{M}=\{1,\ldots,M\}$ and, for each model, parameters are $\theta_m\in\Theta_m\subseteq\mathbb{R}^{d_m}$, $\forall m\in\mathscr{M}$. The likelihood function of each model is $\mathscr{L}_m(\theta_m;\mathbf{Y})$ and the prior is $p(\theta_m)$. Although priors can be improper, we assume that the posteriors are finitely integrable. By Bayes' theorem the posterior of each model is: $p_m(\theta_m|\mathbf{Y})\propto\mathscr{L}_m(\theta_m;\mathbf{Y})p(\theta_m)\,\forall m\in\mathscr{M}$. The posterior predictive distribution for $\mathbf{y}_o$ is:

$$p_m(\mathbf{y}_o|\mathbf{Y})=\int_{\Theta_m}p_m(\mathbf{y}_o,\theta_m|\mathbf{Y})\mathrm{d}\theta_m=\int_{\Theta_m}p_m(\mathbf{y}_o|\mathbf{Y},\theta_m)p_m(\theta_m|\mathbf{Y})\,\mathrm{d}\theta_m\,\forall m\in\mathscr{M}. \tag{5}$$

Access to the posterior $p_m(\theta_m|\mathbf{Y})$ is often provided by Markov Chain Monte Carlo (MCMC) methods. Additionally, $p_m(\mathbf{y}_o|\mathbf{Y},\theta_m)$ is often available in closed form or it is easy to calculate. As an illustration, let us consider the stochastic frontier model (SFM):[1]:

$$y_{it}=\mathbf{x}_{it}'\boldsymbol{\beta}+v_{it}-u_{it},\ i=1,\ldots,n,\ t=1,\ldots,T, \tag{6}$$

where $\mathbf{x}_{it}\in\mathbb{R}^k$ is a vector of inputs, $\boldsymbol{\beta}\in\mathbb{R}^k$ is a vector of parameters, v_{it} denotes the two-sided error term and u_{it} is a random variable with non-negative support representing technical inefficiency. A standard assumption is that

$$v_{it}\overset{iid}{\sim}\mathcal{N}(0,\sigma_v^2),\ u_{it}\overset{iid}{\sim}\mathcal{N}_+(0,\sigma_u^2), \tag{7}$$

independently of each other, as well as the covariates $\mathbf{x}_{it}$. The likelihood function of the model is available in closed form

[1] Cost and production frontiers can be accommodated using this model. Moreover, input and output distance functions can be accommodated using appropriate definition of inputs and outputs. For a summary and comparisons with DEA see Parmeter and Zelenyuk (2020).

and it is given by

$$\mathscr{L}_1(\theta_1; \mathbf{Y}) \propto \sigma^{-nT} \prod_{i=1}^{n} \prod_{t=1}^{T} \left\{ \phi\left(\tfrac{y_{it}-\mathbf{x}_{it}'\boldsymbol{\beta}}{\sigma}\right) \Phi\left(-\lambda \tfrac{y_{it}-\mathbf{x}_{it}'\boldsymbol{\beta}}{\sigma}\right) \right\}, \tag{8}$$

where $\sigma^2 = \sigma_v^2 + \sigma_u^2$ and $\lambda = \frac{\sigma_u}{\sigma_v}$, where $\phi(\cdot)$ and $\Phi(\cdot)$ denote, respectively, the standard normal density and distribution functions. In this instance, the parameter vector is $\theta_1 = [\boldsymbol{\beta}', \sigma_v, \sigma_u]'$, and we agree that this is **model 1**. An alternative specification (Kumbhakar and Lovell, 2000, pp 80–82) is that inefficiency follows an exponential distribution with density $p(u_{it}|\alpha) = \alpha^{-1} e^{-u_{it}/\alpha}$ $(\alpha > 0)$. In this case, which we can call **model 2**, the likelihood function is:

$$\mathscr{L}_2(\theta_2; \mathbf{Y}) \propto \alpha^{-nT} \exp\left\{ nT \tfrac{\sigma_v^2}{2\alpha^2} + \tfrac{1}{\alpha} \sum_{i=1}^{n} \sum_{t=1}^{T} (y_{it} - \mathbf{x}_{it}'\boldsymbol{\beta}) \right\} \prod_{i=1}^{n} \prod_{t=1}^{T} \Phi\left(\tfrac{y_{it}-\mathbf{x}_{it}'\boldsymbol{\beta}}{\sigma_v} - \tfrac{\sigma_v}{\alpha}\right), \tag{9}$$

where the parameter vector is $\theta_2 = [\boldsymbol{\beta}', \sigma_v, \alpha]'$. Additionally, one may consider a truncated normal distribution for the one-sided error (Kumbhakar and Lovell, 2000, pp. 82–86) or a *gamma* distribution (op. cit., pp. 86–89).

From (6) it is clear that

$$y_{it}|\mathbf{x}_{it}, u_{it}, \theta \sim \mathcal{N}\left(\mathbf{x}_{it}'\boldsymbol{\beta} - u_{it}, \sigma_v^2\right). \tag{10}$$

Given any distribution with density $p(u_{it}|\theta)$ the posterior predictive distribution has density:

$$p(y_{it}|\mathbf{x}_{it}) = \int_0^\infty \int_\Theta (2\pi\sigma_v^2)^{-1/2} \exp\left\{ -\tfrac{(y_{it}-\mathbf{x}_{it}'\boldsymbol{\beta}+u_{it})^2}{2\sigma_v^2} \right\} p(u_{it}|\theta) p(\theta|\mathbf{Y}) \, \mathrm{d}u_{it} \, \mathrm{d}\theta. \tag{11}$$

For example, in the normal-half–normal case, we have:

$$p(y_{it}|\mathbf{x}_{it}) \propto \int \sigma^{-1} \phi\left(\tfrac{y_{it}-\mathbf{x}_{it}'\boldsymbol{\beta}}{\sigma}\right) \Phi\left(-\lambda \tfrac{y_{it}-\mathbf{x}_{it}'\boldsymbol{\beta}}{\sigma}\right) p(\beta, \sigma_v, \sigma_u|Y) \, \mathrm{d}\beta \, \mathrm{d}\sigma_v \, \mathrm{d}\sigma_u. \tag{12}$$

Although the integral cannot be computed in closed form, one can use MCMC methods to draw a sequence $\left\{\theta^{(s)}, s = 1, \ldots, S\right\}$ from the posterior and approximate the posterior predictive using the following highly accurate approximation:

$$p(y_{it}|\mathbf{x}_{it}) \simeq S^{-1} \sum_{s=1}^{S} \tfrac{2}{\sigma^{(s)}} \phi\left(\tfrac{y_{it}-\mathbf{x}_{it}'\boldsymbol{\beta}^{(s)}}{\sigma^{(s)}}\right) \Phi\left(-\lambda \tfrac{y_{it}-\mathbf{x}_{it}'\boldsymbol{\beta}^{(s)}}{\sigma^{(s)}}\right). \tag{13}$$

In the normal–exponential case, we have

$$p(y_{it}|\mathbf{x}_{it}) \simeq \text{const.} \tfrac{1}{\alpha^{(s)}} \exp\left\{ \tfrac{\sigma_v^{(s)2}}{2\alpha^{(s)2}} + \tfrac{1}{\alpha^{(s)}} \left(y_{it} - \mathbf{x}_{it}'\boldsymbol{\beta}^{(s)}\right) \right\} \Phi\left(\tfrac{y_{it}-\mathbf{x}_{it}'\boldsymbol{\beta}^{(s)}}{\sigma_v^{(s)}} - \tfrac{\sigma_v^{(s)}}{\alpha^{(s)}}\right), \tag{14}$$

where $\left\{\beta^{(s)}, \sigma_v^{(s)}, \alpha^{(s)}, s = 1, \ldots, S\right\}$ converges to the distribution whose density is the product of the prior and the expression in (9).

3.2 Optimal model pools

Given a set of models whose posterior predictive densities are $p_m(\mathbf{y}_o|\mathbf{Y})$, we consider predictive densities of the form:

$$\sum_{m\in\mathscr{M}} w_m p_m(\mathbf{y}_o|\mathbf{Y}_{1:t});\ \sum_{m\in\mathscr{M}} w_m = 1;\ w_m \geq 0\ \forall m \in \mathscr{M}, \tag{15}$$

which are known as *linear opinion pools*. The data up to and including period t are denoted $\mathbf{Y}_{1:t}$. We consider using the log predictive score function:

$$\sum_{i=1}^{n}\sum_{t=1}^{T} \log\left[\sum_{m=1}^{M} w_m p_m(\mathbf{y}_o|\mathbf{Y}_{1:t})\right], \tag{16}$$

as in Geweke and Amisano (2011 a, b). The log predictive score function is a measure of the out-of-sample prediction record of the model. Other rules are, of course, possible, like mean square prediction error. Given any such rule, the methods proposed here are generic. Any scoring rule is said to be **proper** if the modeler is led to report a predictive density that is coherent and consistent with his subjective probabilities (Good, 1952). "Coherent", in this context, means that prior and data evidence are strictly separated so, for example, one does not use a prior that is based on the same data that is used for performing statistical inference.

Bernardo (1979) showed that *the only proper local scoring rule takes the form of* (16), although in the case of a discrete support for $\mathbf{y}_o$ the result is known since de Finetti and Savage (1963). Maximizing (16) subject to (15) is quite feasible given nonlinear programming software.

Therefore, the problem is:

$$\begin{gathered}\max_{\{w_m, m\in\mathscr{M}\}}\ \sum_{i=1}^{n}\sum_{t=1}^{T} \log\left[\sum_{m=1}^{M} w_m p_m(\mathbf{y}_o|\mathbf{Y}_{1:t})\right], \\ \text{subject to} \\ \sum_{m\in\mathscr{M}} w_m = 1;\ w_m \geq 0\ \forall m \in \mathscr{M}.\end{gathered} \tag{17}$$

Therefore, the purpose of this optimization program is to maximize weighted log predictive scoring (or, equivalently, maximize out-of-sample performance) subject to the usual portfolio-like constraints on the weights attached to different models. *It must be emphasized that this approach is quite general in that it allows for different, general models not necessarily sharing the same parameters although, apparently, all models are conditioned on (that is use) the same data.*[2]

There are at least two uses into which we can put the new techniques.

First, one can estimate all models using all the available data and determine the optimal weights in (16) and (15) using (17). In turn, inefficiency estimates can be obtained by averaging the different inefficiency estimates using the optimal weights.

Second, one can estimate all models using a portion of the observations and, in turn, compute the posterior predictive inefficiency (or efficiency) distributions, average them (using the optimal weights from the portion of the sample that was

[2] In our computations we used the `fortran 77` version of library `lbfgs` in `netlib`. The problem can be converted to unconstrained optimization by reparametrizing $w_m = \frac{\exp\left(-\omega_m^2\right)}{1+\sum_{m'>1}\exp\left(-\omega_{m'}^2\right)}$, and the ω_is are defined on the real line. Convergence is quite fast given any initial conditions and no other local optima were found.

used to estimate the models), and compare "actual" inefficiency[3] (or efficiency) with inefficiency estimates from individual models as well as the optimally weighted inefficiency.

Although one might be tempted to proceed as follows we need to warn the reader that this is not necessarily good practice. Suppose inefficiency estimates for a particular firm (that we can "o") are $\hat{u}_m$, $m \in \mathscr{M}$. Then, a possible estimator is $\tilde{u} = \sum_{m \in \mathscr{M}} w_m \hat{u}_m$. The problem is that the posterior predictive distribution of inefficiency (or efficiency) is not necessarily symmetric. The correct procedure is the following.

We are interested explicitly in the distribution with density $p(u_o|\mathbf{Y}, \mathbf{x}_o)$ or, alternatively, $p(r_o|\mathbf{Y}, \mathbf{x}_o)$, where $r_o = e^{-u_o}$ and u_o denotes inefficiency of firm "o". Moreover, suppose $\mathbf{y}_o \equiv (y_o, \mathbf{x}_o)$ denotes the vector of output–inputs for this firm which is *not* part of $\mathbf{Y}$. Simple calculations show that:

$$p(u_o|\mathbf{Y}, \mathbf{y}_o) = \int_{\Theta} p(u_o|\mathbf{Y}, \mathbf{y}_o, \theta) p(\theta|\mathbf{Y}, \mathbf{y}_o) \, \mathrm{d}\theta, \tag{18}$$

The density $p(u_o|\mathbf{Y}, \mathbf{y}_o, \theta)$ is, usually, not available in closed form, with the exception of normal-half–normal and normal–exponential models in which cases we know that it is a truncated normal distribution. More generally, since, at least under normality of v_{it} (and thus v_o) we have[4]:

$$\begin{aligned} p(u_o|\mathbf{Y}, \mathbf{y}_o, \theta) = (2\pi\sigma_v^2)^{-1/2} \exp\left\{-\tfrac{(y_o - \mathbf{x}_o'\boldsymbol{\beta} + u_o)^2}{2\sigma_v^2}\right\} p(u_o|\theta) p(\theta) \propto \\ \exp\left\{-\tfrac{(y_o - \mathbf{x}_o'\boldsymbol{\beta} + u_o)^2}{2\sigma_v^2}\right\} p(u_o|\theta). \end{aligned} \tag{19}$$

For any distribution of u_o this can be easily evaluated point-wise. The other element in (18) is $p(\theta|\mathbf{Y}, \mathbf{y}_o)$ which is the posterior of the parameter given the observed data $\mathbf{Y}$ *and* the additional data $\mathbf{y}_o$. What we have readily available is a MCMC sample from $p(\theta|\mathbf{Y})$. The difference between $p(\theta|\mathbf{Y}, \mathbf{y}_o)$ and $p(\theta|\mathbf{Y})$ is the additional observation $\mathbf{y}_o$. In principle, this requires an additional MCMC. However, we have

$$p(\theta|\mathbf{Y}, \mathbf{y}_o) \propto p(\theta|\mathbf{Y}) p(\mathbf{y}_o|\theta), \tag{20}$$

where $p(\mathbf{y}_o|\theta)$ is the likelihood contribution of the additional observation. Therefore, a MCMC sample $\left\{\theta^{(s)}, s = 1, \ldots, S\right\}$ from $p(\theta|\mathbf{Y})$ can be transformed to an approximate sample from $p(\theta|\mathbf{Y}, \mathbf{y}_o)$ using the Sampling-Importance-Resampling (SIR) procedure (Rubin, 1988). In this instance, the procedure amounts to accepting each existing draw with probability proportional to $p(\mathbf{y}_o|\theta)$. In turn, given a sequence of draws from the correct posterior, which we still denote by $\left\{\theta^{(s)}, s = 1, \ldots, S\right\}$ to avoid cluttering the notation, we have:

$$p(u_o|\mathbf{Y}, \mathbf{y}_o) \simeq \text{const.} \, S^{-1} \sum_{s=1}^{S} \exp\left\{-\tfrac{(y_o - \mathbf{x}_o'\boldsymbol{\beta}^{(s)} + u_o)^2}{2\sigma_v^{(s)2}}\right\} p(u_o|\theta^{(s)}). \tag{21}$$

Of course, normality is not a serious restriction and more general models can be considered in future research. We

[3] "Actual" inefficiency is the estimate from a given model once the data for the firm whose inefficiency is sought, become available.

[4] The formula is derived using the joint distribution $p(\mathbf{Y}, \mathbf{y}_o, u_o, \mathbf{u}; \theta)$ where $\mathbf{u} = [u_{it}]$ assuming u_o is not an element of $\mathbf{u}$. Notice that $p(u_o|\mathbf{Y}, \mathbf{y}_o, \theta) = p(\mathbf{Y}, \mathbf{y}_o, u_o|\mathbf{u}, \theta) p(\mathbf{u}|\theta) p(\theta) \propto p(\mathbf{Y}, \mathbf{y}_o, u_o|\mathbf{u}, \theta) \propto p(u_o|\mathbf{Y}, \mathbf{y}_o)$.

Table 1: Cases and scenarios

Case	σ_v	Distribution of u	True cost function	σ_u	Sample size (n)
1 SGA-CD	0.01, 0.05	$\mathcal{N}_+(0,\sigma_u^2)$	CD	0.01, 0.05, 0.15	50, 100, 200, 400, 800
2	0	$\mathcal{N}_+(0,\sigma_u^2)$	CD	0.01, 0.05, 0.15	50, 100, 200, 400, 800
3	0.01, 0.05	$\Gamma(1,\sigma_u)$	CD	0.01, 0.05, 0.15	50, 100, 200, 400, 800
4	0	$\Gamma(1,\sigma_u)$	CD	0.01, 0.05, 0.15	50, 100, 200, 400, 800
5	0.01, 0.05	$\mathcal{N}_+(0,\sigma_u^2)$	TL	0.01, 0.05, 0.15	50, 100, 200, 400, 800
6	0	$\mathcal{N}_+(0,\sigma_u^2)$	TL	0.01, 0.05, 0.15	50, 100, 200, 400, 800
7 SFA-TL	0.01, 0.05	$\Gamma(1,\sigma_u)$	TL	0.01, 0.05, 0.15	50, 100, 200, 400, 800
8	0	$\Gamma(1,\sigma_u)$	TL	0.01, 0.05, 0.15	50, 100, 200, 400, 800
9	0.01, 0.05	$\mathcal{N}_+(0,\sigma_u^2)$	TL & DEA	0.01, 0.05, 0.15	50, 100, 200, 400, 800
10	0	$N_+(0,\sigma_u^2)$	TL & DEA	0.01, 0.05, 0.15	50, 100, 200, 400, 800
11 TL-DEA	0.01, 0.05	$\Gamma(1,\sigma_u)$	TL & DEA	0.01, 0.05, 0.15	50, 100, 200, 400, 800
12	0	$\Gamma(1,\sigma_u)$	TL & DEA	0.01, 0.05, 0.15	50, 100, 200, 400, 800
13	0.01, 0.05	$\mathcal{N}_+(0,\sigma_u^2)$	DEA	0.01, 0.05, 0.15	50, 100, 200, 400, 800
14	0	$N_+(0,\sigma_u^2)$	DEA	0.01, 0.05, 0.15	50, 100, 200, 400, 800
15	0.01, 0.05	$\Gamma(1,\sigma_u)$	DEA	0.01, 0.05, 0.15	50, 100, 200, 400, 800
16 DEA	0	$\Gamma(1,\sigma_u)$	DEA	0.01, 0.05, 0.15	50, 100, 200, 400, 800

should notice that solving (16) subject to (15) solves the problem of "maximizing out-of-sample fit" in terms of the data for y_{it}; not in terms of predicting inefficiency itself as actual inefficiency is unknown.

4 Monte Carlo simulations

In this section, we provide Monte Carlo results to examine the performance of the new techniques. We adopt the same set up as in APS. Since the results are likely to depend on the underlying DGP (data generating process) APS propose four DGPs. The first cost function is Cobb–Douglas and the parameters are equal to 0.39, 0.36, and 0.23, yielding returns of scale of 0.98. The second cost function is translog estimated from the data in APS. The remaining two scenarios provide an advantage to DEA-like estimators. The third scenario results if we estimate the true cost frontier based on the translog, and then monotonize and convexify the estimates by using DEA-VRS on the fitted values. This approach is consistent with stochastic DEA as proposed in Simar and Zelenyuk (2011). In the fourth scenario, the cost frontier is estimated by application of DEA-VRS in the APS data set, and using the stochastic frontier estimates to generate the data. Another advantage is that monotonicity and concavity hold automatically on the generated frontier. In turn, values of σ_v and σ_u can be assigned different values for the four different DGPs. We reproduce Table 3 in APS as Table 1 to summarize the different cases that we examine in our Monte Carlo simulations.

The results from our Monte Carlo simulation are reported in Tables 2–6.

The message from the Monte Carlo evidence in Tables 2–6 is that the new techniques perform extremely well, certainly better than APS, even when the DGPs favor the application of DEA-like methods. The impressive behavior of the new techniques is, certainly, attributed to the propriety of the scoring rule (in the particular sense we have defined) and the ability of the Geweke and Amisano (2011 a,b) optimization problem to disentangle relevant models from a constellation of possibly irrelevant models. Certainly, these results are quite encouraging, and imply that f*ormal, likelihood-based scoring*

Table 2: Case 1 (SFA-CD), Root mean squared errors (RMSE)

	n	σ_v	σ_u	λ	APS	*this paper*
I	$n = 50$	0.01	0.01	1	0.0097	*0.0045*
	$n = 100$	0.01	0.01	1	0.0096	*0.0042*
	$n = 200$	0.01	0.01	1	0.0101	*0.0039*
	$n = 400$	0.01	0.01	1	0.0115	*0.0036*
	$n = 800$	0.01	0.01	1	0.0121	*0.0032*
II	$n = 50$	0.05	0.01	0.2	0.0100	*0.0084*
	$n = 100$	0.05	0.01	0.2	0.0100	*0.0080*
	$n = 200$	0.05	0.01	0.2	0.0098	*0.0078*
	$n = 400$	0.05	0.01	0.2	0.0097	*0.0074*
	$n = 800$	0.05	0.01	0.2	0.0095	*0.0070*
III	$n = 50$	0.01	0.05	5	0.0221	*0.0098*
	$n = 100$	0.01	0.05	5	0.0163	*0.0075*
	$n = 200$	0.01	0.05	5	0.0124	*0.0068*
	$n = 400$	0.01	0.05	5	0.0103	*0.0061*
	$n = 800$	0.01	0.05	5	0.0093	*0.0057*
IV	$n = 50$	0.05	0.05	1	0.0411	*0.0112*
	$n = 100$	0.05	0.05	1	0.0365	*0.0092*
	$n = 200$	0.05	0.05	1	0.0303	*0.0088*
	$n = 400$	0.05	0.05	1	0.0273	*0.0080*
	$n = 800$	0.05	0.05	1	0.0258	*0.0077*
V	$n = 50$	0.01	0.15	15	0.0535	*0.0131*
	$n = 100$	0.01	0.15	15	0.0369	*0.0124*
	$n = 200$	0.01	0.15	15	0.0249	*0.0102*
	$n = 400$	0.01	0.15	15	0.0171	*0.0098*
	$n = 800$	0.01	0.15	15	0.0132	*0.0092*
VI	$n = 50$	0.05	0.15	3	0.0640	*0.0152*
	$n = 100$	0.05	0.15	3	0.0508	*0.0137*
	$n = 200$	0.05	0.15	3	0.0433	*0.0131*
	$n = 400$	0.05	0.15	3	0.0396	*0.0127*
	$n = 800$	0.05	0.15	3	0.0379	*0.0123*

Notes: The Table reports root mean squared error (RMSE) results from the MAX estimator of Andor, Parmeter and Sommer (2019, Table 5, p. 245).

Table 3: Case 16 (DEA), Root mean squared errors (RMSE)

	n	σ_v	σ_u	λ	APS	this paper
I	$n = 50$	0.01	0.01	1	0.0086	*0.0033*
	$n = 100$	0.01	0.01	1	0.0058	*0.0022*
	$n = 200$	0.01	0.01	1	0.0037	*0.0018*
	$n = 400$	0.01	0.01	1	0.0029	*0.0014*
	$n = 800$	0.01	0.01	1	0.0027	*0.0010*
II	$n = 50$	0.05	0.01	0.2	0.0337	*0.0132*
	$n = 100$	0.05	0.01	0.2	0.0226	*0.0127*
	$n = 200$	0.05	0.01	0.2	0.0162	*0.0123*
	$n = 400$	0.05	0.01	0.2	0.0131	*0.0098*
	$n = 800$	0.05	0.01	0.2	0.0120	*0.0092*
III	$n = 50$	0.01	0.05	5	0.0631	*0.0144*
	$n = 100$	0.01	0.05	5	0.0433	*0.0131*
	$n = 200$	0.01	0.05	5	0.0295	*0.0122*
	$n = 400$	0.01	0.05	5	0.0209	*0.0118*
	$n = 800$	0.01	0.05	5	0.0170	*0.0097*

Notes: The Table reports root mean squared error (RMSE) results from the MAX estimator of Andor, Parmeter and Sommer (2019, Table 6, p. 246).

Table 4: Case 7 (SFA-TL), Root mean squared errors (RMSE)

	n	σ_v	σ_u	λ	APS	*this paper*
I	$n = 50$	0.01	0.01	1	0.0105	*0.0082*
	$n = 100$	0.01	0.01	1	0.0096	*0.0044*
	$n = 200$	0.01	0.01	1	0.0087	*0.0032*
	$n = 400$	0.01	0.01	1	0.0083	*0.0030*
	$n = 800$	0.01	0.01	1	0.0082	*0.0024*
II	$n = 50$	0.05	0.01	0.2	0.0143	*0.0072*
	$n = 100$	0.05	0.01	0.2	0.0140	*0.0068*
	$n = 200$	0.05	0.01	0.2	0.0138	*0.0061*
	$n = 400$	0.05	0.01	0.2	0.0137	*0.0055*
	$n = 800$	0.05	0.01	0.2	0.0133	*0.0052*
III	$n = 50$	0.01	0.05	5	0.0272	*0.0078*
	$n = 100$	0.01	0.05	5	0.0199	*0.0072*
	$n = 200$	0.01	0.05	5	0.0170	*0.0069*
	$n = 400$	0.01	0.05	5	0.0174	*0.0062*
	$n = 800$	0.01	0.05	5	0.0189	*0.0059*
IV	$n = 50$	0.05	0.05	1	0.0449	*0.0120*
	$n = 100$	0.05	0.05	1	0.0400	*0.0117*
	$n = 200$	0.05	0.05	1	0.0377	*0.0115*
	$n = 400$	0.05	0.05	1	0.0373	*0.0098*
	$n = 800$	0.05	0.05	1	0.0535	*0.0093*
V	$n = 50$	0.01	0.15	15	0.0573	*0.0122*
	$n = 100$	0.01	0.15	15	0.0389	*0.0117*
	$n = 200$	0.01	0.15	15	0.0265	*0.0097*
	$n = 400$	0.01	0.15	15	0.0200	*0.0091*
	$n = 800$	0.01	0.15	15	0.0170	*0.0088*
VI	$n = 50$	0.05	0.15	3	0.0638	*0.0132*
	$n = 100$	0.05	0.15	3	0.0517	*0.0129*
	$n = 200$	0.05	0.15	3	0.0483	*0.0121*
	$n = 400$	0.05	0.15	3	0.0503	*0.0118*
	$n = 800$	0.05	0.15	3	0.0516	*0.0105*

Notes: The Table reports root mean squared error (RMSE) results from the MAX estimator of Andor, Parmeter and Sommer (2019, Table 7, p. 247).

Table 5: Case 11 (SFA-TL), Root mean squared errors (RMSE)

	n	σ_v	σ_u	λ	APS	this paper
I	$n = 50$	0.01	0.01	1	0.0104	*0.0072*
	$n = 100$	0.01	0.01	1	0.0096	*0.0068*
	$n = 200$	0.01	0.01	1	0.0092	*0.0061*
	$n = 400$	0.01	0.01	1	0.0093	*0.0056*
	$n = 800$	0.01	0.01	1	0.0095	*0.0050*
II	$n = 50$	0.05	0.01	0.2	0.0141	*0.0070*
	$n = 100$	0.05	0.01	0.2	0.0138	*0.0067*
	$n = 200$	0.05	0.01	0.2	0.0139	*0.0064*
	$n = 400$	0.05	0.01	0.2	0.0139	*0.0060*
	$n = 800$	0.05	0.01	0.2	0.0133	*0.0057*
III	$n = 50$	0.01	0.05	5	0.0277	*0.0114*
	$n = 100$	0.01	0.05	5	0.0194	*0.0089*
	$n = 200$	0.01	0.05	5	0.0147	*0.0081*
	$n = 400$	0.01	0.05	5	0.0131	*0.0076*
	$n = 800$	0.01	0.05	5	0.0136	*0.0071*
IV	$n = 50$	0.05	0.05	1	0.0445	*0.0102*
	$n = 100$	0.05	0.05	1	0.0397	*0.0096*
	$n = 200$	0.05	0.05	1	0.0377	*0.0091*
	$n = 400$	0.05	0.05	1	0.0375	*0.0088*
	$n = 800$	0.05	0.05	1	0.0419	*0.0082*
V	$n = 50$	0.01	0.15	15	0.0582	*0.0106*
	$n = 100$	0.01	0.15	15	0.0393	*0.0101*
	$n = 200$	0.01	0.15	15	0.0265	*0.0097*
	$n = 400$	0.01	0.15	15	0.0190	*0.0092*
	$n = 800$	0.01	0.15	15	0.0165	*0.0088*
VI	$n = 50$	0.05	0.15	3	0.0640	*0.0082*
	$n = 100$	0.05	0.15	3	0.0513	*0.0078*
	$n = 200$	0.05	0.15	3	0.0475	*0.0074*
	$n = 400$	0.05	0.15	3	0.0491	*0.0070*
	$n = 800$	0.05	0.15	3	0.0505	*0.0067*

Notes: The Table reports root mean squared error (RMSE) results from the MAX estimator of Andor, Parmeter and Sommer (2019, Table 8, p. 248).

Table 6: Overall performance of methods

Scenarios	Performance measure	MAX	this paper
I All	RMSE	0.0215	*0.0104*
	Bias	0.0052	*0.0013*
	PU	0.3150	*0.045*
II No noise	RMSE	0.0142	*0.0077*
	Bias	0.0084	*0.0032*
	PU	0.0775	*0.0100*
III With noise	RMSE	0.0267	*0.0084*
	Bias	0.0025	*0.0004*
	PU	0.4650	*0.0851*

Notes: The MAX measures are taken from Andor, Parmeter and Sommer (2019, Table 4, p. 245). RMSE is root mean squared error and PU is percentage of firms whose efficiency is underestimated.

rules along with optimization as in Geweke and Amisano (2011 a, b) can produce model pools of relevant models that can, effectively, deal with the problems of optimal weighting schemes in APS.

5 Empirical results

To apply the new techniques we use the data of Malikov, Kumbhakar and Tsionas (2016). Specifically, we have an unbalanced panel with 2,397 bank–year observations for 285 banks from Call Reports available from the Federal Reserve Bank of Chicago and include all FDIC-insured commercial banks with reported data for 2001:Q1–2010:Q4. We have five desirable outputs ($y_1 - y_5$), five inputs ($x_1 - x_5$), and a quasi–fixed input (equity capital). The bank's undesirable output is total non-performing loans (b). We estimate a translog input distance function which includes a time trend as an additional variable to capture technical change. Bank effects are included as well.

We consider the following models:

Model 1, is a normal-half–normal SFM.

Model 2, is a normal–exponential SFM.

Model 3, is a normal-*gamma*–normal SFM. The density of the gamma distribution is $f(u|P,\alpha) = \frac{\alpha^P}{(P-1)!}u^{P-1}e^{-\alpha u}$ ($u \geq 0$, $\alpha, P > 0$) and we consider the case of an unknown shape parameter P whose prior is $p(P) \propto P^{-1}$ (Tsionas, 2000).

Model 4 deviates from the previous models in that we allow for dynamic inefficiency (Tsionas, 2006):

$$\log u_{it} = z_{it}'\gamma + \rho \log u_{i,t-1} + e_{it}. \tag{22}$$

It remains to consider methods in the Stochastic DEA tradition. There are compelling reasons to assume noisy data (Simar and Zelenyuk, 2011; Parmeter and Zelenyuk, 2020).

Specifically, we use the DEA formulation in Keshvari (2017) (see for example Allon, Beenstock, Hackman, Passy, and Shapiro (2007), Keshvari and Kuosmanen (2013), Kuosmanen and Kortelainen (2012), and Kuosmanen (2008)) as formulated in Tsionas and Izzeldin (2018) which is our Model 5.

Our Model 6 relies on the alternative interpretation of DEA as provided in Tsionas (2020). Our Model 7 uses DEA scores as a prior to calibrate prior parameters of a normal-truncated–nomal specification (Tsionas, 2003). All MCMC computations are based on 150,000 iterations discarding the first 50,000 to mitigate possible start up effects.

We present technical inefficiency estimates in panel (a) of Figure 1 and optimal weights in panel (b).

The optimal model weights are 0.0192, 0.0207, 0.0162, 0.1146, 0.2884, 0.3163, and 0.2247. The first three models, unsurprisingly, deliver similar inefficiency distributions and their weights are almost the same. The last four models account for almost 94% of total weights and their inefficiency distributions are different; the largest rank correlation coefficient being 0.17 between models 5 and 7.

This solves in a formal way the problem posed by APS. There is, however, an additional exercise that we can consider. Suppose $\mathbf{y}_o$ corresponds to a particular bank–year block observation (i.e. all observations of a given bank) that

Figure 1: Technical inefficiency estimates

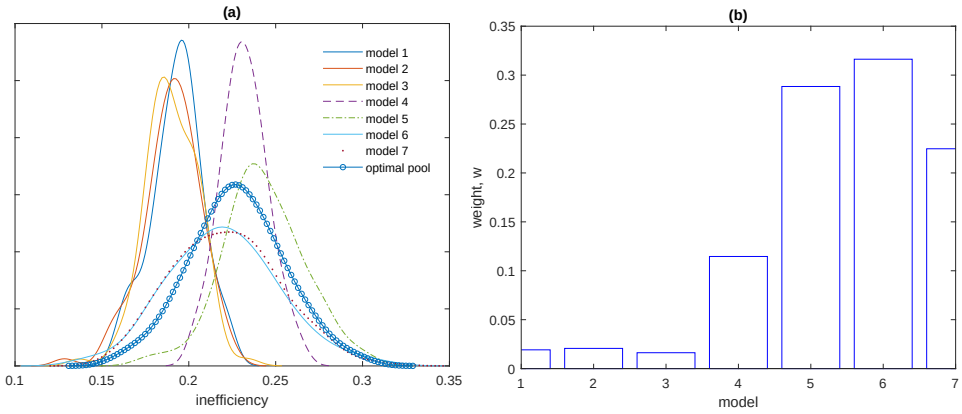
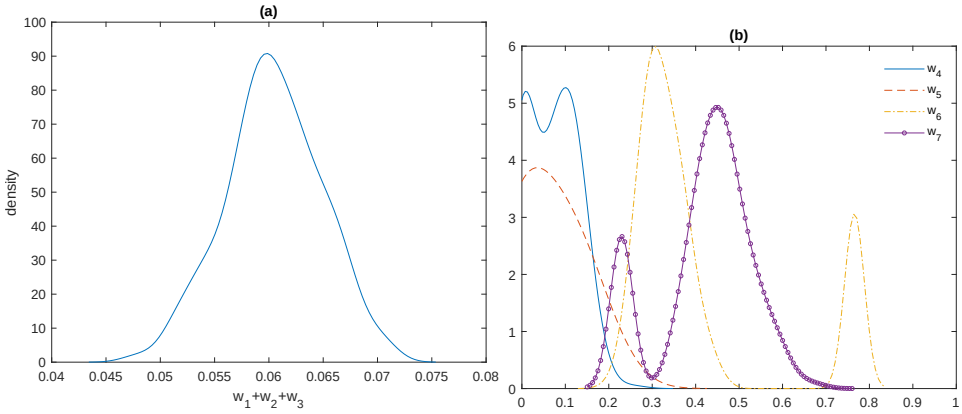


Figure 2: Sample distributions of optimal pool weights



we omit followed by MCMC in the remaining sample and solution of (16) subject to (15).[5]

Our interest is in the resulting distribution of weights. Some evidence is provided in Figure 2. In panel (a) we report the sample distribution of weights for the first three models which is concentrated around 0.06 and extends from nearly 0.045 to 0.075.

Optimal weights for the other models are reported in panel (b). The interesting feature of the sample distributions in panel (b) is, certainly, the fact that these densities have considerable probability density in the neighborhood of zero, for models 1, 5, and 7 but not so for model 6; this model has a distinct mode nearly 0.5 and less pronounced but important mode around, approximately, 0.65 and 0.93, and weights for this model are never zero. As a matter of fact, there is some density concentration around unity and it turned out that in 47 out of 285 banks, model 6 had an optimal weight equal to one. The implication is that Bayesian extensions of DEA receive considerable support in the light of the data from the point of view of optimal predictive pools and, therefore, they could prove to be valuable tools in future research. Somewhat surprisingly, a normal-truncated–normal model with the prior calibrated using DEA information (viz. Model 7, Tsionas, 2003) which is, essentially, an empirical Bayes procedure, performs well although not as well as the Bayes DEA

[5]Despite the fact that we use full MCMC, if computational resources are scarce one can resort to the SIR approximation that we described above.

procedure (Model 6, Tsionas, 2020).

6 Optimal pooling for DEA models

The models we considered so far depend on parametric assumptions to the extent that the posterior predictive distribution should be amenable to point-wise evaluation. Many models, including DEA, produce (in)efficiency scores by relying on linear programming or other procedures. Therefore, given a set of data, $\mathbf{Y}$, there is a procedure $\mathscr{P}$ (or operator) that generates inefficiency scores: $\hat{u} = \mathscr{P}(\mathbf{Y}; \boldsymbol{\xi})$ where $\boldsymbol{\xi} \in \boldsymbol{\Xi} \subseteq \mathbb{R}^{d_\xi}$ is a parameter. In DEA we do not have such parameters but we can imagine that the data generating process (DGP) is, for example, a multivariate normal distribution for $\mathbf{Y}$, in which case $\boldsymbol{\xi}$ represents the mean vector and covariance matrix. An extension to a multivariate mixture of normal distributions is straightforward and various clustering procedures may be used to define clusters[6] (the definition thus becoming part of $\boldsymbol{\xi}$) and their associated moments.

Given a *sampling* distribution for the parameters whose density is $f_{\boldsymbol{\xi}}(\boldsymbol{\xi})$, one can generate a sequence of draws $\left\{\xi^{(l)}, l = 1, \ldots, L\right\}$, simulate a data set $\mathbf{Y}^{(l)}$ and, in turn, provide efficiency scores $\hat{u}^{(l)} = \mathscr{P}(\mathbf{Y}^{(l)}; \boldsymbol{\xi}^{(l)})$. Often, the sampling distribution is asymptotically normal or it can be approximated by a bootstrap or jackknife procedure (to reduce bias in the latter case). This procedure accounts for uncertainty in the data. Therefore, an approximation to the distribution of inefficiency scores reduces to approximating the distribution of draws $\left\{\hat{u}^{(l)}, l = 1, \ldots, L\right\}$ point-wise. Standard kernel density procedures can be used in this instance (Parmeter and Kubmhakar, 2014; Parmeter and Racine, 2012; Parmeter, Wang and Kumbhakar, 2017).

Additionally, in DEA it is straightforward to compute an inefficiency score, say $\hat{u}_o$, for a unit whose data is $\mathbf{y}_o$, that is $\hat{u}_o = \mathscr{P}([\mathbf{Y}\, \mathbf{y}_o]; \boldsymbol{\xi})$ and, in turn, an approximation to the distribution of $\hat{u}_o$ can be obtained through the draws $\left\{\hat{u}_o^{(l)}, l = 1, \ldots, L\right\}$. Therefore, for this class of models, it is possible to use maximization of (16) subject to (15) and obtain optimal predictive pools.

Following APS, we use the standard CCR (Charnes, Cooper, & Rhodes, 1978) and BCC (Banker, Charnes, & Cooper, 1984) DEA formulations that allow for constant returns to scale (CRS) and variable returns to scale (VRS), with output orientation. Output orientation can be motivated on theoretical grounds as we use the financial intermediation approach in the previous section. Additionally, we use the Stochastic DEA (SDEA) approach, see Simar (2007), Simar and Zelenyuk (2011) as applied in Parmeter and Zelenyuk (2020).

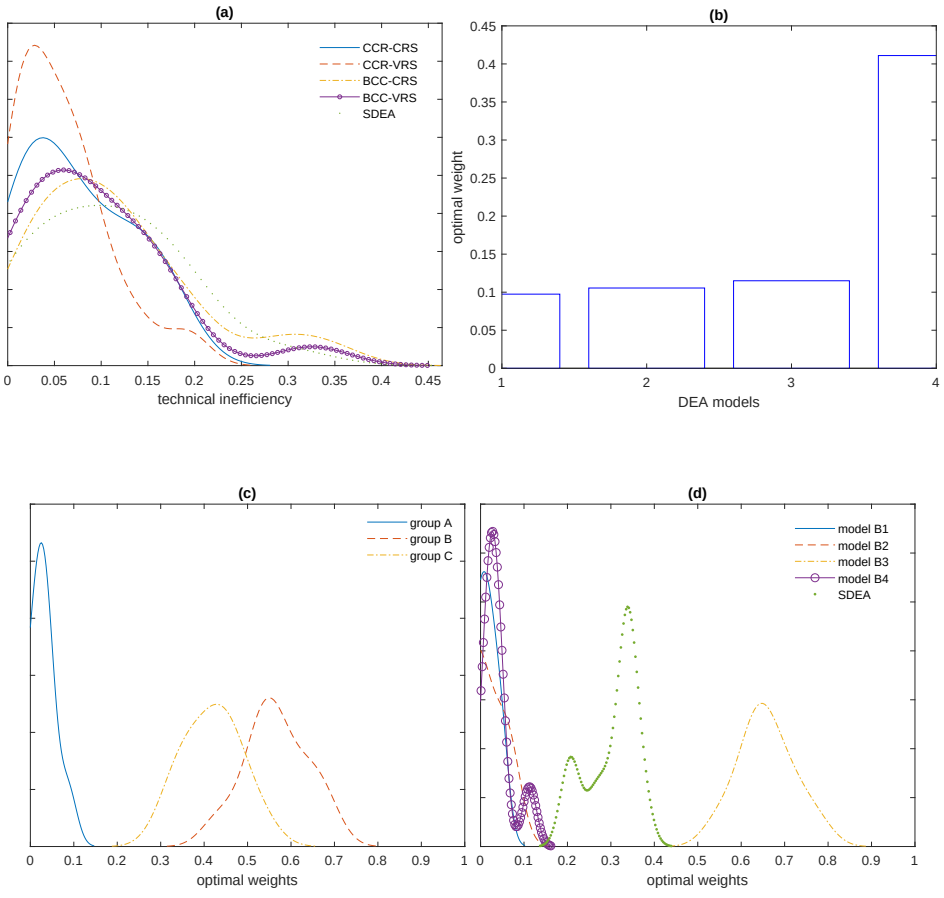
We report distributions of inefficiency scores in panel (a) of Figure 3 and optimal weights in panel (b).

The predictive optimal weights for the DEA models are 0.1337, 0.1447, 0.1578, and 0.5638. Therefore, the BCC-VRS model receives the greatest weight and inefficiency distributions seem to be the same and their rank correlation coefficients are large (0.85 on the average). Interestingly, *none of the weights is zero* so, the *different DEA models seem to convey independent and useful information on inefficiency scores.* The empirically relevant question is whether the DEA models receive non-zero weights when the predictive pool includes Models 1–7.

We examine their performance in the same context as in Figure 2, and we report results in Figure 4. As there are,

[6] See Po, Guh and Yu (2009) and Krüger (2010).

Figure 3: DEA models



clearly, three groups of models (Models 1–3, 4–7 and the four DEA models), we first report the sampling distributions of sum of optimal weights in panel (c) of Figure 3.

Clearly, model groups B and C (consisting of models 4 through 7 and the five DEA models) receive the lion's share in terms of optimal weights so, from the point of view of optimal predictive pooling, the main contribution comes from these groups. Closer examination of the results in group C, showed that it is SDEA (Simar and Zelenyuk, 2011) which receives the lion's share within this group so, there seems to be great potential for SDEA.

The optimal weights of each model separately, within group B, *given that all twelve models are considered*, have a sampling distribution which is presented in panel (d) of Figure 3 along with SDEA which comes out as a clear winner within group C. For concreteness, models in group B are called **Models B1, B2, B3, and B4**.

Interestingly, the lion's share now goes to model B3 (the Bayes DEA model, Tsionas, 2020) which receives optimal weights in excess of 0.6 and 0.921 on the average, with considerable probability density concentration around unity but also to the SDEA model (whose optimal weights range from 0.15 to 0.43, averaging 0.302). The other models have optimal weights with considerable probability density concentration around zero and in no case they exceed, roughly, 0.20. We reiterate that this result is obtained using all twelve models which are both in the DEA and SFA tradition.

Concluding remarks and further directions

In this paper, we elaborated on the derivation and computation of posterior predictive distributions for stochastic frontier models and provided a formal answer to the question set out in APS regarding optimal weighting of inefficiency measures to obtain an overall optimal estimator. As posterior predictive distributions are, more often than not, asymmetric, using standard location measures like the mean, median or other quantiles seems difficult. We use the concept of optimal predictive pools to address the problem in APS. Optimal predictive pools (Geweke and Amisano, 2011 a, b) yield optimally weighted predictive distributions to "maximize fit". The optimal weights can, in turn, be used to obtain an optimal distribution of inefficiency estimates. In this context, the only proper predictive rule is log predictive scoring, a fact that greatly simplifies the answer to the important question posed by APS.

The new techniques are applied successfully to a US banking data set, and we find that two methods receive the greatest support in the light of the data, in terms of log predictive scoring, viz. the Bayes DEA model (Tsionas, 2020) and the Stochastic DEA model of Simar (2007), and Simar and Zelenyuk (2011); see also Parmeter and Zelenyuk (2020). "Simple" SFA analysis (normal-half–normal, normal–exponential, etc.) do not seem to be associated with sizable optimal predictive weights although a dynamic stochastic frontier model (Tsionas, 2006) and an empirical Bayes SFA model (Tsionas, 2003) have non-negligible weights depending on the pool of models we consider. Overall, when all models are considered (twelve in total), Bayes DEA and SDEA dominate other models. Despite this dominance, other models have non-zero weights so, final inferences on (in)efficiency scores should also take them into account.

The investigation of other interesting models is necessarily left for future research along with some other issues. These interesting models include FLW (Fan, Li, & Weersink, 1996), SVKZ (Simar, Van Keilegom, & Zelenyuk, 2017), and PSZ (Park, Simar, & Zelenyuk, 2015). Our framework is general enough to accommodate such models and, as we showed,

computation of predictive distributions (although not always *posterior* predictive distributions) can be accomplished in a routine and embarrassingly parallel way (Delgado and Parmeter, 2013) within each model, for most models. Two issues can be considered in future research. The first, relates to Monte Carlo studies regarding the finite–sample behavior of optimal predictive pools. The second and, perhaps more important, concerns the issue of whether "simple" SFA and DEA models can be safely ignored in favor of their more advanced versions (for example dynamic inefficiency models in SFA and SDEA in the DEA tradition). The issue is empirically relevant in the sense that reduction of model pools apparently minimizes computational burden. Theoretical or Monte Carlo insights into this question would be highly relevant in efficiency estimation.

References

[1] Allon, G., M. Beenstock, S. Hackman, U. Passy, & A. Shapiro, 2007. Nonparametric estimation of concave production technologies by entropic methods. Journal of Applied Econometrics 22 (4), 795–816.

[2] Andor, M. A., C. Parmeter, & S. Sommer, 2019. Combining uncertainty with uncertainty to get certainty? Efficiency analysis for regulation purposes. European Journal of Operational Research 274, 240–252.

[3] Banker, R. D. , Charnes, A. , & Cooper, W. W., 1984. Some models for estimating technical and scale inefficiencies in data envelopment analysis. Management Science 30 (9), 1078–1092 .

[4] Bernardo, J.M., 1979. Expected information as expected utility. The Annals of Statistics 7, 686–690.

[5] Charnes, A., Cooper, W. W. , & Rhodes, E., 1978. Measuring the efficiency of decision making units. European Journal of Operational Research 2 (6), 429–444 .

[6] de Finetti, B., & Savage, L.J., 1963. The elicitation or personal probabilities, unpublished manuscript.

[7] Delgado, M. S., & Parmeter, C. F., 2013. Embarrassingly easy embarrassingly parallel processing in R. Journal of Applied Econometrics 28 (6), 1224–1230.

[8] Fan Y., Li Q., & Weersink, A., 1996. Semiparametric estimation of stochastic production frontier models. Journal of Business and Economic Statistics 14 (4), 460–468.

[9] Geweke, J., & G. Amisano, 2011a. Optimal prediction pools. Journal of Econometrics 164, 130–141.

[10] Geweke, J., & Amisano, G., 2011b. Hierarchical Markov normal mixture models with applications to financial asset returns. Journal of Applied Econometrics 26, 1–29.

[11] Good, I. J., 1952. Rational decisions. Journal of the Royal Statistical Society Series B 14, 107–114.

[12] Hall, S. G., Mitchell, J., 2007. Combining density forecasts. International Journal of Forecasting 23, 1–1

[13] Keshvari, A., & T. Kuosmanen, 2013. Stochastic non-convex envelopment of data: Applying isotonic regression to frontier estimation European Journal of Operational Research 231 (2), 481–491.

[14] Krüger, J. J., 2010. Comment on "A new clustering approach using data envelopment analysis" European Journal of Operational Research, 206 (1), 269–270.

[15] Kumbhakar, S. C., & C. A. K. Lovell, 2000. Stochastic frontier analysis. Cambridge University Press.

[16] Kumbhakar, S. C., Park, B. U., Simar, L. & Tsionas, M. G., 2007. Nonparametric stochastic frontiers: A local maximum likelihood approach. Journal of Econometrics 137 (1), 1–27.

[17] Kuosmanen, T., 2008. Representation theorem for convex nonparametric least squares. Econometrics Journal 11 (2), 308–325.

[18] Kuosmanen, T., & M. Kortelainen, 2012. Stochastic non-smooth envelopment of data: Semiparametric frontier estimation subject to shape constraints. Journal of Productivity Analysis, 38 (1), 11–28.

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

[20] Park, B. U., Simar, L., & Zelenyuk, V., 2015. Categorical data in local maximum likelihood: Theory and applications to productivity analysis. Journal of Productivity Analysis 43 (1),199–214.

[21] Parmeter, C. F., & Kumbhakar, S. C., 2014. Efficiency analysis: A primer on recent advances. Foundations and Trends in Econometrics 7 (3–4), 191–385.

[22] Parmeter, C. F., & Racine, J. S., 2012. Smooth constrained frontier analysis. Chen X, Swanson N, eds. Recent Advances and Future Directions in Causality, Prediction, and Specification Analysis: Essays in Honor of Halbert L. White Jr. (Springer, New York), pp. 463–489.

[23] Parmeter, C. F., Wang H. - J., & Kumbhakar, S. C., 2017. Nonparametric estimation of the determinants of inefficiency. Journal of Productivity Analysis 47 (3), 205–221.

[24] Parmeter, C. F., & V. Zelenyuk, 2020. Combining the Virtues of Stochastic Frontier and Data Envelopment Analysis, Operations Research, forthcoming.

[25] Po, R-W, Y-Y Guh, & M-S Yang, 2009. A new clustering approach using data envelopment analysis European Journal of Operational Research, 199 (1), 276–284.

[26] Rubin, D. B., 1988. Using the SIR algorithm to simulate posterior distributions. In Bayesian Statistics 3, ed. J. M. Bernardo, M. H. DeGroot, D. V. Lindley, & A. F. M. Smith, 395–402. Oxford: Oxford University Press.

[27] Simar, L., 2007. How to improve the performances of DEA/FDH estimators in the presence of noise. Journal of Productivity Analysis 28 (2), 183–201.

[28] Simar, L., Van Keilegom I., & Zelenyuk, V., 2017. Nonparametric least squares methods for stochastic frontier models. J. Productivity Anal. 47(3), 189–204.

[29] Simar, L., & Zelenyuk, V., 2011. Stochastic FDH/DEA estimators for frontier analysis, Journal of Productivity Analysis 36 (1), 1–20.

[30] Tsionas, M. G. 2000, Full Likelihood Inference in Normal-Gamma Stochastic Frontier Models. Journal of Productivity Analysis 13 (3), 183–205.

[31] Tsionas, M. G., 2003, Combining DEA and stochastic frontier models: An empirical Bayes approach. European Journal of Operational Research 147, 499–510.

[32] Tsionas, M. G., 2006, Inference in dynamic stochastic frontier models. Journal of Applied Econometrics 21 (5), 669–676.

[33] Tsionas, M. G., 2020. A coherent approach to Bayesian Data Envelopment Analysis. European Journal of Operational Research, 281 (2), 439–448.

[34] Tsionas, M. G., & M. Izzeldin, 2018. Smooth approximations to monotone concave functions in production analysis: An alternative to nonparametric concave least squares European Journal of Operational Research, 271 (3), 797–807.