Minimax regret priors for efficiency estimation

Mike Tsionas*

January 23, 2023

Abstract

We propose a minimax regret empirical prior for inefficiencies in a stochastic frontier model and for its other parameters. The class of priors over which we consider minimax regret is given by DEA interval scores and, for the parameters, the class of priors induced by maximum likelihood estimates. The new techniques are shown to perform well in a Monte Carlo study as well as in real data for large U.S. data banks.

Key Words: Productivity and Competitiveness; Stochastic Frontier Models; Minimax Regret Prior; Data Envelopment Analysis.

---

*Montpellier Business School & Lancaster University Management School, LA1 4YX, U.K., m.tsionas@montpellier-bs.com

1   Introduction

The problem of specifying a distribution for inefficiency in stochastic frontier models (SFMs) is long-standing and well-known. Typical choices are the half-normal, the exponential, and the truncated-normal. A partial list includes Chen and Lin (2009), Kao et al., (2019), Kumbhakar and Tsionas (2020), Tabak et al. (2010), Tran et al. (2022), Tsionas (2021, 2022), Papadopoulos and Parmeter (2021), Jradi et al. (2021). Tsionas (2003) proposed using DEA scores for crafting a prior for SFMs, an approach that was refined in Tsionas (2023).

In this paper we propose the use of a minimax regret empirical prior relative to the sample distribution of efficiency scores from Data Envelopment Analysis (DEA). Despite the fact that DEA does not account for noisy data, it is likely that interval estimators of efficiency (a quantity naturally defined in $(0, 1]$) will be more robust and thus they offer a natural candidate for crafting or benchmarking an (in)efficiency prior. These interval priors are closer to the idea of "thick frontiers" (Berger and Humphrey, 1992; Tsionas, 2019) in which classes rather than point inferences are used. The minimax prior is taken with respect to all distributions that are consistent with DEA interval estimators of inefficiency and they can be used to mitigate concerns about use of a single prior (be it, for example, a half-normal or DEA itself). For related work that uses DEA in SFA with non-standard efficiency distributions, see Campbell, Rogers and Rezek (2008), Rezek, Campbell and Rogers (2011), Macedo and Scotto (2014), Macedo, Silva and Scotto (2014).

The term "prior" in conjunction with inefficiency is used as the latter represent missing data. We consider the minimax regret prior as an empirical prior for inefficiencies as well

as other SFM parameters and we use DEA and maximum likelihood estimation to craft this empirical (Bayes) prior. Given that inefficiencies are not observed and they are coarse, any given interval is compatible with different "data" on inefficiency, making the construction of a single prior problematic. When observations are coarse and may overlap, it is not clear how to define the likelihood function: several options exist, according to whether we take into account the measurement process governing the incompleteness or not. In this paper, we focus on optimizing the likelihood function that we should have observed, had observations been precise. Due to incomplete observations, this likelihood function is imprecise, since there are several possible precise datasets compatible with the coarse observations. Two approaches have been proposed: one considers an optimistic point of view aiming to disambiguate the data, by maximizing the maximum likelihood value across candidate datasets. Another more cautious one tries to maximize the minimum likelihood value across candidate datasets thus adopting a robust optimization approach. Both approaches have their weaknesses and can be criticized as being extreme ones, yielding either too deterministic or too dispersed distribution functions. In this paper we propose an alternative criterion that tries to define a trade-off between the two previous approaches, and can be seen as minimizing a maximal regret criterion.Finally, we apply the new techniques to a dataset of large U.S. banks to showcase the applicability of minimax regret priors.

2 Model

2.1 Formulation

We consider a stochastic frontier model (SFM)[1]

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

where the regressors $x_i \in \mathbb{R}^p$ explain the dependent variable y_i, $\beta \in \mathbb{R}^p$ is a parameter vector, $v_i | x_i, u_i \sim \mathcal{N}(0, \sigma_v^2)$ represents measurement error (noise), and u_i stands for technical inefficiency. We denote efficiency by $r_i = e^{-u_i} \in (0, 1]$ $(i = 1, ..., n)$. More specifically, we consider the class of input distance functions (IDFs). Suppose the inputs are denoted $X_i = (X_{i1}, ..., X_{iI}) \in \mathbb{R}_+^I$ not to be confused with x_is, and outputs are denoted $Y_i = (Y_{i1}, ..., Y_{iJ}) \in \mathbb{R}_+^J$. When there is no risk of confusion we omit the observation index i $(i = 1, ..., n)$. Let the set of feasible production plans be

$$\mathcal{T} = \{(X, Y) : X \text{ can produce } Y). \tag{2}$$

The IDF is defined as

$$D(X, Y) = \max\{\lambda | (X/\lambda, Y) \in \mathcal{T}\}, \tag{3}$$

and it has several properties: It is homogeneous of degree one in inputs, increasing in inputs, decreasing in outputs and concave in inputs. It can be shown that a translog *approximation*

---

[1] Although we have panel data we omit the time index in the interest of simplicity.

to an IDF takes the following form (see also O'Donnell, 2018, p. 284):[2]

$$\log X_{i1} = \alpha_0 + \sum_{k=2}^{I} \alpha_k \log \tilde{X}_{ik} + \sum_{m=1}^{J} \beta_m \log Y_{im}$$

$$+ \frac{1}{2} \sum_{k=2}^{I} \sum_{k'=2}^{I} \alpha_{kk'} \log \tilde{X}_{ik} \log \tilde{X}_{ik'} + \frac{1}{2} \sum_{m=1}^{J} \sum_{m'=1}^{J} \beta_{kk'} \log Y_{im} \log Y_{im'}$$

$$+ \sum_{k=2}^{I} \sum_{m=1}^{J} \gamma_{km} \log \tilde{X}_{ik} \log Y_{im} + \sum_{k=2}^{I} \sum_{l=1}^{L} \delta_{kl} w_{il} \log \tilde{X}_{ik'} \tag{4}$$

$$+ \sum_{m=1}^{M} \sum_{l=1}^{L} \zeta_{ml} w_{il} \log Y_{im} + v_i + u_i, \ \forall i = 1, ..., n,$$

where $\tilde{X}_{ik} = \frac{X_{ik}}{X_{i1}}$, α_0, $(\alpha_k, \alpha_{kk'}, \beta_{mm'}, \gamma_{km}, \delta_{kl}, \zeta_{ml})$ are unknown parameters, $\mathbf{w}_i = (w_{il}, l = 1, ..., L)$ $(\forall i = 1, ..., n)$ is an $L \times 1$ vector of exogenous variables that includes firm-specific time effects, a time trend and possibly quasi-fixed inputs. A full translog functional form is used to incorporate this vector into the analysis. Notice that (4) is in the form of (1). To benchmark a prior, we use DEA with the given sets of inputs and outputs. DEA provides a set of inefficiency scores or estimates u_i^{DEA} for u_i ($R_i = e^{-u_i}$ and, therefore, $R_i^{\text{DEA}} = e^{-u_i^{\text{DEA}}}$) which ignore the presence of noise or any simplifications arising from knowing approximately, at least, the functional form. For example, the translog is known to be second-order accurate for an arbitrary IDF. Our purpose is to use the efficiency estimates (R_i^{DEA}) for benchmarking a prior $p_R(R_i)$ from which a prior $p_u(u)$ on inefficiency $u_i = -\log r_i$ can be computed using change of variables. Although DEA delivers numerical estimates we prefer to use the concept of "thick frontiers" to mitigate the problem of noise in DEA. Specifically, we assume that efficiency scores can be in r distinct groups, defined by $(0, \bar{a}_1)$, $[\bar{a}_1, \bar{a}_2)$, ..., $[\bar{a}_{r-2}, \bar{a}_{r-1})$, $(\bar{a}_r, 1]$. Despite the influence of noise, the thick frontier is less affected by it so intervals estimators of efficiency should be closer to reality. As we are dealing with crafting a prior, we need not worry about true inefficiency falling in the correct interval all the time. The efficiency groups

---

[2]No risk of confusion should arise between index k used here and those in the next section.

can be transformed to inefficiency intervals and we have $(-\infty, a_1)$, $[a_1, a_2)$, ..., $[a_{r-2}, a_{r-1})$, $(a_r, 1]$, where $a_j = -\log \bar{a}_j$. The intervals are fixed but their number $1 + r$ can vary so, we can choose r. In selecting r, we must keep in mind that finer intervals are less robust to noise while coarser intervals are more robust but less informative in terms of locating inefficiency. Our purpose is to adopt a prior which has minimax regret with respect to all priors compatible with the DEA estimates.

2.2 The minimax regret empirical prior

To summarize, for a random variable u representing inefficiency we consider an experiment where u is unobserved but there is an estimator $\hat{u}$ which provides incomplete observations for inefficiency:[3]

$u : \Omega \to \mathcal{U}$, is the outcome of a random experiment. Suppose $\mathcal{U} = \{-\infty, a_1, ..., a_M\}$.

$\hat{u} : \Omega \to \hat{\mathcal{U}} \subseteq \mathfrak{s}(\mathcal{U})$, where $\mathfrak{s}(\mathcal{U})$ is the set of all subsets of $\mathcal{U}$, and $\hat{u}$ represents the estimator. We assume $\hat{\mathcal{U}} = \{A_1, ..., A_r\}$, $A_i \subseteq \mathcal{U}$ albeit in our problem $r = M$. Our question is what is the best choice for the distribution $P(u)$ given the estimates. This choice concerns the selection of a prior probability distribution to be used subsequently in more elaborate analysis. In other words, the estimator $\hat{u}$ is considered as a rough guide. We will write our choice as $P(u|\theta)$ conditional on a parameter $\theta \in \Theta \subseteq \mathbb{R}^d$. Let $\hat{\mathbf{u}} = \{G_1, .., G_n\}$ be a sample of observations and $G_j \in \hat{\mathcal{U}}$ representing estimating inefficiency to be in group or interval j; $\hat{U} = G_j \Leftrightarrow u_j \in G_j$, where u_j is the jth unobserved outcome of $\mathbf{u} = (u_1, ..., u_n)$, generated through the random process u. Notice that G_j merely provides a group or interval to which inefficiency belongs. This is broadly consistent with the idea of "thick frontiers". Let also

---

[3]Similarly, we can consider the problem in terms of efficiencies.

$\mathcal{U}(\hat{\mathbf{u}})$ be the set of samples of u consistent with the estimates $\hat{\mathbf{u}}$.

As usual in cases involving latent data we have three likelihood functions, namely:

1. The observed sample in $\hat{\mathcal{U}}$: $P(\hat{\mathbf{u}}|\theta) = \prod_{i=1}^{n} p(\hat{u}_i|\theta)$, where $p(\hat{u}_i|\theta)$ is a density.

2. The latent sample in $\mathcal{U}$: $P(\mathbf{u}|\theta) = \prod_{i=1}^{n} p(u_i|\theta)$.

3. The complete sample of $z_i = (u_i, G_i)$: $L^{\mathbf{z}}(\theta) = \prod_{i=1}^{n} p(z_i|\theta)$, $\mathbf{z} = ((u_1, G_1), ..., (u_n, G_n))$,

$u_j \in G_j$, $\forall j = 1, ..., n$.

In this discussion we are interested in the second option viz., the latent sample in $\mathcal{U}$. The desired distribution is a member of $\{P(\mathbf{u}|\theta) : \mathbf{u} \in \mathcal{U}(\hat{\mathbf{u}})\}$. Based on the set of estimates $\hat{\mathbf{u}} = (\hat{u}_1, ..., \hat{u}_n) \in \hat{\mathcal{U}}^n$ we have two options:

(i) maximax strategy: find $(\mathbf{u}^*, \theta^*)$ that maximizes $\overline{\Lambda}(\theta) = \max_{\mathbf{u} \in \mathcal{U}(\hat{\mathbf{u}})} P(\mathbf{u}|\theta)$.

(ii) maximin strategy: find $\theta^* \in \Theta$ that maximizes $\underline{\Lambda}(\theta) = \min_{\mathbf{u} \in \mathcal{U}(\hat{\mathbf{u}})} P(\mathbf{u}|\theta)$.

We will not consider these problems in detail and we will only refer to Guillaume & Dubois (2019) where the problems are analyzed albeit not in the context of efficiency estimation. It suffices to say that both approaches are extreme, yielding either "too deterministic" or "too dispersed" distribution functions. Suppose n_k is the number of appearances of a_k in $\mathbf{u}$, q_j is the number of observations of A_j in $\hat{\mathbf{u}}$, and n_{ik} is the number of times that (a_i, A_k) is in the complete sample $\mathbf{z}$. It follows that $\mathbf{u} \in \mathcal{U}(\hat{\mathbf{u}})$ if an only if the following conditions are satisfied (Guillaume & Dubois, 2019):

$$\sum_{k=1}^{r} n_k = \sum_{j=1}^{r} q_j = n,$$

$$n_k = \sum_{j=1}^{r} n_{kj} \;\forall k = 1, ..., M,$$

$$q_j = \sum_{k=1}^{M} n_{kj} \;\forall j = 1, ..., r,$$

$$n_{kj} = 0 \text{ if } a_k \notin A_j \;\forall k, j.$$

(5)

Using log-likelihoods and letting $p_k^\theta = \Pr(U = a_k | \theta)$ the two problems can be written as :

$$\max_{\mathbf{p}} \max_{\mathbf{n}} \sum_{k=1}^{M} n_k \log p_k,$$

$$\max_{\mathbf{p}} \min_{\mathbf{n}} \sum_{k=1}^{M} n_k \log p_k,$$

$$\text{subject to constraints (5) and} \tag{6}$$

$$\sum_{k=1}^{M} p_k = 1,$$

$$n_k, n_{kj} \in \mathbb{N}, \ p_k > 0 \ \forall k = 1, ..., m.$$

If we let $\theta = (p_1, ..., p_M)$ be a probability mass function, we have the probability mass assignment $\mathrm{pmf}(A_j) = \frac{q_j}{n}$ $(j = 1, ..., r)$, the optimal solution to the maximin likelihood problem subject to the first three constraints of (5) and $n_k \in \mathbb{R}_+$, is the distribution with maximal entropy, namely the solution to $\max_{\mathbf{n}} - \sum_{k=1}^{M} \frac{n_k}{n} \log \frac{n_k}{n}$. The solution to the second problem in (6) is the solution with minimal entropy, namely the solution to $\min_{\mathbf{n}} - \sum_{k=1}^{M} \frac{n_k}{n} \log \frac{n_k}{n}$ subject to the first three constraints of (5). The maximax approach tends to resolve uncertainty very strongly yielding Dirac distributions consistent with the estimates. The maximin approach yields distributions with high variances interpreting incomplete information as the result of extreme randomness: the less information the larger the variance, which is not fully satisfactory either. Suppose $\hat{\theta}_\mathbf{u}$ is the maximum likelihood estimator of θ provided $\mathbf{u}$ is observed. The minimax relative regret problem is

$$\max_{\theta \in \Theta} \min_{\mathbf{u} \in \mathcal{U}(\hat{\mathbf{u}})} \frac{P(\mathbf{u}|\theta)}{P(\mathbf{u}|\hat{\theta}_\mathbf{u})}. \tag{7}$$

The interpretation is that we try to find the value of θ that reaches the best compromise between the various ideal values $\hat{\theta}_\mathbf{u}$ for all $\mathbf{u}$ in agreement with the estimates in $\hat{\mathbf{u}}$. The

problem of minimizing maximal regret is

$$\min_{\theta \in \Theta} \max_{\mathbf{u} \in \mathcal{U}(\hat{\mathbf{u}})} L(\mathbf{u}|\hat{\theta}_{\mathbf{u}}) - L(\mathbf{u}|\theta), \tag{8}$$

which is equivalent to the following mathematical programming model:

$$\min_{\mathbf{p}} \alpha$$

subject to:

(i) $\sum_{k=1}^{M} n_k \log n_k - \sum_{k=1}^{M} n_k \log p_k \leq \alpha, \ \forall u \in \mathcal{V}(\mathcal{U}(\mathbf{y}))$, \hfill (9)

(ii) $\sum_{k=1}^{M} p_k = 1,$

(iii) $p_k \geq 0, \ \forall k = 1, ..., M,$

where $\mathcal{V}(\cdot)$ signifies the vertices of a set. The problem can be solved only numerically using standard solvers.[4] In our application we set $r = 10$. This is the value that maximizes the log marginal likelihood in our estimated model. The model is estimated using Bayesian techniques organized around Markov Chain Monte Carlo (MCMC) using 15,000 iterations the first 5,000 of which are omitted in the interest of mitigating possible start up effects. Flat priors are used on the translog parameters and the two-sided error standard deviation, σ_v.[5]

---

[4]We use `conmax.f` from the `netlib` distribution.

[5]If β denotes all translog parameters then this prior is $p(\beta, \sigma_v) \propto \sigma_v^{-1} e^{-\bar{q}/(2\sigma_v^2)}$ where $\bar{q}$ is set to machine precision. For a general minimax regret prior, MCMC is a random-walk Metropolis algorithm (Tierney, 1994) fine tuned so that approximately 20–25% of all candidates are accepted. Marginal likelihood for different values of r (ranging from 3 to 20) is computed using the Laplace approximation (DiCiccio et al., 1997). Convergence and numerical performance is monitored using Geweke's (1992) diagnostics.

3 The minimax regret prior for all other parameters

In the SFM given by (1), parameters β and σ_v can be assigned a minimax regret prior conditional on inefficiencies $U = (u_1, ..., u_n)$. With the exception of the intercept, we know that least squares (LS) estimates of β are consistent. Suppose the vector b contains LS parameter estimates ignoring technical inefficiency and S is the vector of LS standard errors. Them we know that the intervals $b \pm 2S$ contain the true parameters with sampling probability 95% (element-wise). We assume that for the intercept, a consistent estimator and its standard error are available in the first elements of b and S, viz., b_1 and S_1. Such a consistent estimator is provided by $b_1 := b_1 + \min_{i=1,...,n} e_i$ where $e_i = y_i - x_i'b$ are LS residuals. An alternative is to estimate (1) by maximum likelihood using any conventional distributional assumption e.g., half-normality, which provides consistent standard errors, S, as well. Using the 95% confidence intervals $b_j \pm 2S_j$ $(j = 1, ..., d)$ it is possible to place r points in them (the two extremes being the endpoints of the intervals). In turn, using the methodology of Section 2.2 it is possible to compute the minimax regret empirical prior of β_j conditional on all other βs denoted β_{-j} as well as σ_v and U, viz., we obtain the prior

$$p(\beta_j | \beta_{-j}, \sigma_v, U). \tag{10}$$

The prior of σ_v can be constructed in the same way. In connection to (4) we can use the same methodology or argue that, in this instance, the βs do not have a structural interpretation and therefore it is better to place priors in terms of functions of interest like returns-to-scale (RTS), technical change (TC), and efficiency change (EC). These functions are linear in β

and provide a set of linear functions that we denote $\mathbf{A}\beta$ where $\mathbf{A}$ has dimensionality $F \times d$ where F is the number of functions of interest ($F=4$ in this case) and β is $d \times 1$. A class of possible priors is

$$\mathbf{A}\beta \sim \mathcal{N}_F(\mathbf{r},\ h^2\mathbf{I}_F), \tag{11}$$

where $\mathbf{r}$ is an $F \times 1$ vector and $\mathbf{I}_F$ denotes the $F \times F$ identity matrix. As the properties of a prior like (11) are well understood we will proceed with the former option. We assume that $r^{(\beta,\sigma_v)}$ is the common value of r corresponding to all conditional priors of β and σ_v in (10). We assume a common value merely for convenience and in the empirical application this has been found to be 8 when we maximize the log marginal likelihood of the model. With a full minimax regret empirical prior on β, σ_v, and U we proceed to showcase the new techniques in actual data from large U.S. banks.

4   Monte Carlo study

To examine the properties of the minimax regret empirical prior we consider a Monte Carlo study using model (4) retaining only the first-order terms. Therefore, we have a Cobb-Douglas *approximation* to an IDF (O'Donnell, 2018, p. 284 on this point, where it is argued that if there is more than one output and output sets are bounded, then IDFs cannot be *exact* Cobb-Douglas functions). We have five inputs and five outputs as in our empirical application in the next Section. The regressors are generated by bootstrapping with replacement the data set in the next Section so that we have a more realistic scenario. Coefficients α_k are set to $\frac{1}{4}$ and coefficients β_j to $\frac{1}{5}$ and we consider values of $\sigma_v \in \{0.01, 0.05.0.1, 0.3\}$. The sample size is $n \in \{100, 250, 500, 1000, 2500, 5000\}$. Inefficiency is generated from a *gamma*

distribution, $\mathcal{G}(a,c)$ whose mean is $\frac{a}{c}$ with $a = \frac{2}{3}$ and $c = 1$, which has mean $\frac{2}{3}$. The density of efficiency is reported in Figure 1.

Figure 1: Density of $\mathcal{G}(\frac{2}{3}, 1)$
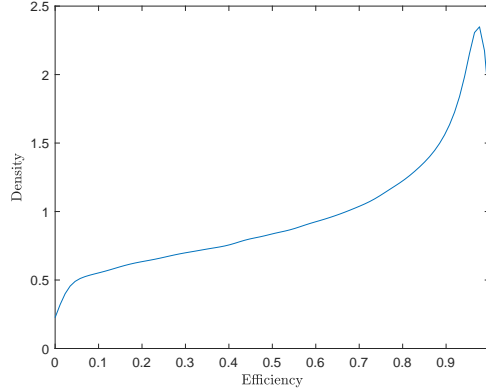
We use $r=10$ supported points for both inefficiencies as well as the parameters of the SFM as this value turned out quite reasonable in the empirical application. We consider 5,000 Monte Carlo replications. For each replication we use 15,000 MCMC passes omitting the first 5,000 and starting values are obtained from maximum likelihood (ML) with half-normally distributed inefficiencies. For ML estimates of inefficiency we know that the estimates are generally consistent unless we have time-invariant fixed effects (Greene, 2005). For DEA we know that efficiency scores converge very slowly at a rate $O(n^{-\frac{1}{I+J}})$. This information is enough to compare our own estimates. Before proceeding, however, it is interesting to examine how the minimax regret priors look like. In Figure 2 we present the joint density of regression parameters α_1 and α_2 (panels (a) and (b)), and their marginals in panels (c) and (d). Clearly, the two parameters presented here are positively correlated as they are based on the data. Moreover, the marginals and the joint density show that the empirical prior is not symmetric and, therefore, cannot be close to a bivariate normal. More evidence on

non-normality is provided by the marginal prior of σ_v reported in panel (e) which is, clearly, multimodal. Multimodality is the result of the compromising role of the minimax regret prior which balances the maximax and maximin priors defined in the previous Section. As the minimax regret empirical prior balances two extremes (uniformity on the one hand and Dirac masses on the other) multimodality is not overly surprising, and it is also evident in densities of α_3 and α_4 presented in panels (f), (g), (h), and (i).

Figure 2: Minimax regret empirical priors of selected parameters
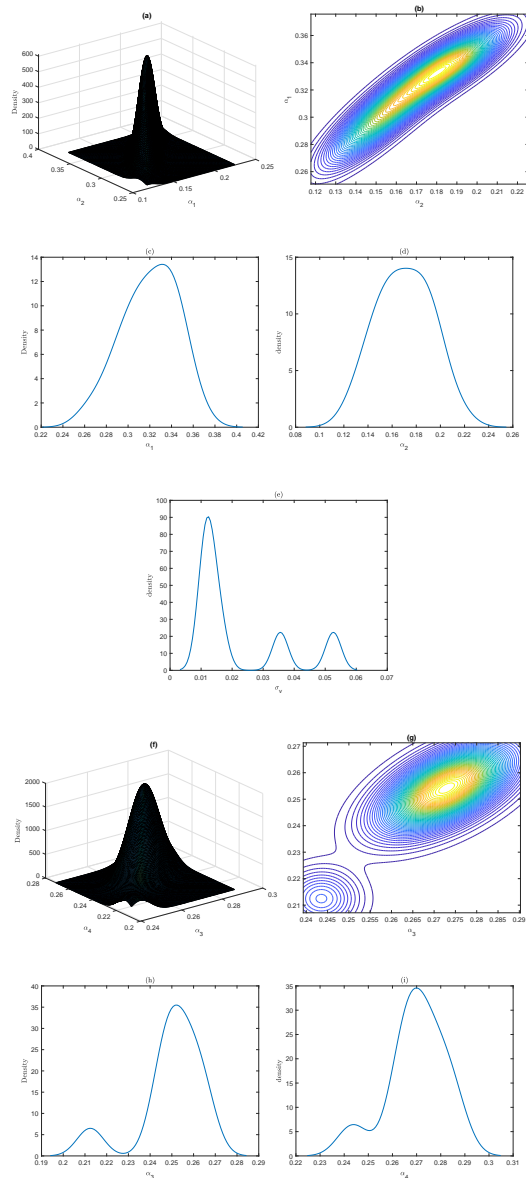
Our Monte Carlo results are reported in Table 1 where shown are mean squared errors of inefficiency estimates. The surprising feature of these results is that the mean squared error of inefficiency estimates, for a given value of σ_v, is that it goes down at the rate n instead of the usual $\sqrt{n}$ –a fact that can be justified if we recall that the prior is empirical, that is it is base on the same data as the posterior.

14

Table 1: Monte Carlo results I

		σ_v			
		0.01	0.05	0.10	0.30
n	100	0.017	0.025	0.032	0.044
	250	0.0068	0.010	0.0128	0.018
	500	0.0034	0.005	0.0065	0.090
	1000	0.0017	0.0025	0.0031	0.0043
	2500	0.0007	0.0010	0.0015	0.0018
	5000	0.0003	0.0005	0.0007	0.0008

As $\sigma_U = \frac{\sqrt{b}}{c} \simeq 1.22$ one may argue that the results in Table 1 are too optimistic. To show that this is not the case, we report additional results in Table 2 with values of σ_v closer to σ_U. Although mean-squared errors are higher, the general tendency of decreasing at rate n rather than $\sqrt{n}$ continues to apply. From the Monte Carlo study, we may conclude the efficiency estimates converge rather quickly (as opposed to DEA which converges very slowly relative to the total number of inputs and outputs, $I + J$) and they are essentially unbiased and consistent. This important property means that minimax regret priors inherit the advantages of DEA being a nonparametric estimator and the advantages of SFA in terms of handling noise. Additionally, the efficiency estimators converge as a function of the sample size rather than $I + J$. This shows that the minimax regret prior can combine effectively the virtues of both DEA and SFA in providing accurate estimates of technical inefficiency.

5   Data

To illustrate the new techniques we use U.S. banking data previously used by Malikov, Kumbhakar and Tsionas (2015). As in their paper, we focus on a selected subsample of relatively homogeneous large banks, i.e. those with total assets in excess of \$1 billion dollars

Table 2: Monte Carlo results II

		σ_v			
		1.0	2.0	3.0	5.0
n	100	0.057	0.062	0.079	0.092
	250	0.0069	0.0250	0.032	0.037
	500	0.0033	0.126	0.016	0.019
	1000	0.0018	0.0064	0.0080	0.0092
	2500	0.0009	0.0025	0.0032	0.0038
	5000	0.0004	0.0012	0.0016	0.0020

(in 2005 US dollars) in the first three years of observation. The data on commercial banks come from Call Reports available from the Federal Reserve Bank of Chicago and include all FDIC-insured commercial banks with reported data for 2001:Q1–2019:Q4. We have an unbalanced panel with 3,897 bank–year observations for 285 banks. We deflate all nominal stock variables to 2005 US dollars using the consumer price index for all urban consumers. We have the following outputs: consumer loans (y1), real estate loans (y2), commercial and industrial loans (y3), securities (y4), and off-balance-sheet income (y5) . See Berger and Mester (1997, 2003) and Hughes and Mester (1998, 2013). The inputs are labor, i.e. the number of full-time equivalent employees (x1), physical capital (x2), purchased funds (x3), interest-bearing transaction accounts (x4) and non-transaction accounts (x5). For comparison purposes we estimate also a translog output distance function (ODF) using the same inputs and outputs. To account for technical change we include a time trend.
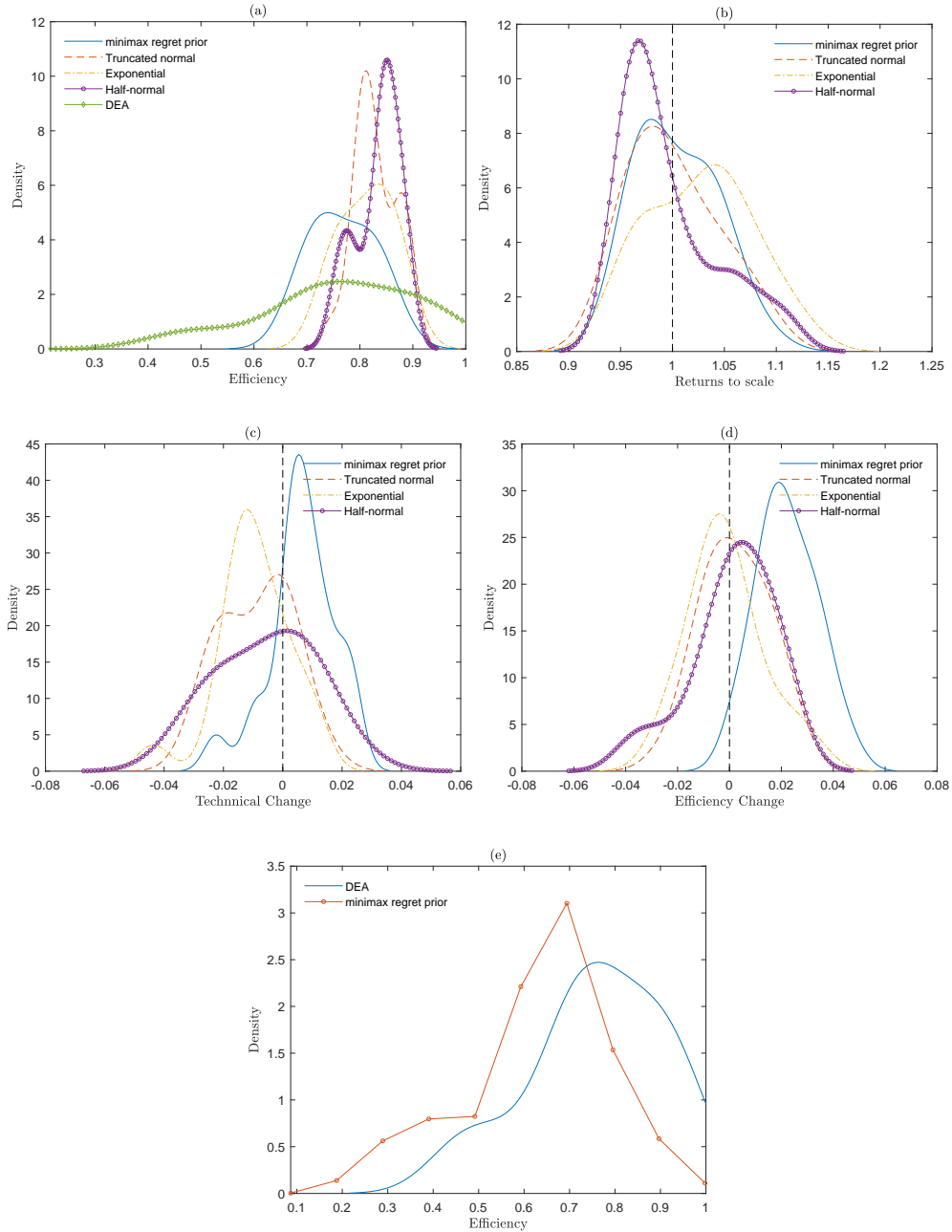
In panel (a) of Figure 3 reported are sample distributions of posterior-mean estimated technical efficiencies. DEA corresponds to input-oriented Data Envelopment Analysis. We also consider SFMs with exponentially, half-normally, and truncated-normally distributed inefficiencies (for the priors of β and σ_v in this case, see Footnote 5; for the shape parame-

ters of the distributions we assume flat priors). In panel (b) reported are sample distributions of posterior-mean estimated returns to scale (RTS). In panel (c) reported are sample distributions of posterior-mean estimated technical change (TC). In panel (d) we present sample distributions of posterior-mean estimated of efficiency change (EC). Finally, in panel (e) we present the actual minimax regret prior (using linear interpolation)[6] compared to the DEA sample distribution of efficiency scores. Clearly, there are important differences when a minimax regret prior is placed on the inefficiencies. RTS, TC, and EC are different and from panel (f) the minimax regret prior is strikingly different compared to DEA despite the fact that the former is based on the latter. In turn, the choice of a minimax regret prior affects materially all measures of interest like RTS, TC, etc.

In terms of policy implications, the new model is likely to be useful as we do not rely on a parametric distribution for inefficiency but rather we anchor on DEA (a popular approach in most policy discussions and empirical applications) and a minimax regret approach to construct the (prior) model. Therefore, inefficiency estimates share some of the good properties of DEA (nonparametric specification) but avoiding some of its drawbacks like slow convergence in the total number of inputs and outputs. In this example, as DEA overestimates efficiency, we find that the minimax regret prior for incomplete data provides lower efficiency and brings results closer to what is known from previous studies. Specifically, bank efficiencies are in the neighborhood of 80–85%. Additionally, we provide a perspective with respect to the following idea. Although DEA may be biased, it is nevertheless biased in the right direction as it underestimates inefficiencies, at least in this example; therefore, policy recommendations to reduce inefficiency will not surprise the decision maker in the

---

[6]The true prior is a stepwise function.

Figure 3: Distributions of interest

Notes: In panel (a) reported are sample distributions of posterior-mean estimated technical efficiencies. DEA corresponds to input-oriented Data Envelopment Analysis. In panel (b) reported are sample distributions of posterior-mean estimated returns to scale. In panel (c) reported are sample distributions of posterior-mean estimated technical change (TC). In panel (d) we present sample distributions of posterior-mean estimated of efficiency change (EC). In panel (e) we present the actual minimax regret prior (using linear interpolation) compared to the DEA sample distribution of efficiency scores.

wrong direction. In other words, if according to DEA inefficiency is 30% and in fact it is 20%, if policy measures are taken to reduce inefficiency by the true amount (20%) then the decision maker will not be disappointed. In this sense DEA is a useful benchmark for policy analysis. However, SFA often performs much better due to better statistical properties so, a combination of both SFA and DEA is to be recommended in practice. One way to accomplish this task is to draft a prior based on the minimax regret criterion for incomplete data and use this prior specification as a distribution for use in SFA. To conclude this discussion, having access to the most reliable information (e.g., from this new unbiased and consistent estimator) allows the use of adequate and proportional policy measures to reduce the true amount of inefficiency.

6 Concluding remarks

In this paper we construct a minimax regret prior for inefficiencies in SFMs (i) to avoid specific distributional assumptions which are made invariably in the literature, and (ii) to avoid the use of a single prior on unobserved or "missing data" inefficiencies. The minimax regret prior is taken with respect to all priors consistent with interval estimates of inefficiencies from DEA. Interval estimation is used as it makes DEA more robust to the presence of noise. The minimax regret inefficiency prior is extended to include the parameters of the SFM. MCMC methods are developed and the new techniques are applied successfully to a dataset on large U.S. banks.

References

[1] Berger, A.N., & Humphrey, D.B. (1992). Measurement and efficiency issues in commercial banking. In Z. Griliches (Ed.), Output measurement in the service sector, NBER, Chicago, IL.

[2] Berger A. N., & Mester, L. J. (1997). Inside the black box: what explains differences in the efficiencies of financial institutions? Journal of Banking and Finance 21(7): 895–947.

[3] Berger A. N., & Mester, L. J. (2003). Explaining the dramatic changes in performance of US banks: technological change,deregulation, and dynamic changes in competition. Journal of Financial Intermediation 12 (1), 57–95.

[4] Campbell, R., K. Rogers & J. Rezek (2008). Efficient frontier estimation: a maximum entropy approach. Journal of Productivity Analysis, 30, 213–221.

[5] Chen, Y. H., Lin, W. T. (2009). Analyzing the relationships between information technology, inputs substitution and national characteristics based on CES stochastic frontier production models. International Journal of Production Economics 120 (2), 552–569.

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

[7] Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In *Bayesian Statistics 4* (J. M. Bernado, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.). Clarendon Press, Oxford, UK, 169–193.

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

[9] Guillaume, R., & Dubois, D. (2019). A Maximum Likelihood Approach to Inference Under Coarse Data Based on Minimax Regret. In: Destercke, S., Denoeux, T., Gil, M., Grzegorzewski, P., & Hryniewicz, O. (eds) Uncertainty Modelling in Data Science. SMPS 2018. Advances in Intelligent Systems and Computing, vol 832. Springer, Cham. https://doi.org/10.1007/978-3-319-97547-4_14, pp. 99–106.

[10] Hughes J. P., & Mester, L. J. (1998). Bank capitalization and cost: evidence of scale economies in risk management and signaling. Review of Economics and Statistics 80 (2), 314–325.

[11] Hughes J. P., & Mester, L. J. (2013). Who said large banks don't experience scale economies? Evidence from a risk–return-driven cost function. Journal of Financial Intermediation 22 (4), 559–585.

[12] Jradi, S., Parmeter, C.F., & Ruggiero, J. (2021). Quantile estimation of stochastic frontiers with the normal-exponential specification. European Journal of Operational Research 295 (2), 475–483.

[13] Kao, T-W (D), Su, H-C., & Chen, Y-S (2020). The curvilinear relationships between structural embeddedness and productive efficiency: An exploratory study. International Journal of Production Economics 212, 176–185.

[14] Kumbhakar, S. C., & Tsionas, M. G. (2020). On the estimation of technical and allocative efficiency in a panel stochastic production frontier system model: Some new

formulations and generalizations. European Journal of Operational Research 287 (2), 762–775.

[15] Kumbhakar, S. C., & Tsionas, M. G. (2021). Dissections of input and output efficiency: A generalized stochastic frontier model. International Journal of Production Economics Volume 232, February 2021, 107940.

[16] Macedo, P., & M. Scotto (2014). Cross-entropy estimation in technical efficiency analysis. Journal of Mathematical Economics, 54, 124–130.

[17] Macedo, P., E. Silva & M. Scotto (2014). Technical efficiency with state-contingent production frontiers using maximum entropy estimators. Journal of Productivity Analysis, 41, 131–140.

[18] Malikov, E., S.C. Kumbhakar & M. G. Tsionas (2015). A Cost System Approach to the Stochastic Directional Technology Distance Function with Undesirable Outputs: The Case of us Banks in 2001–2010. Journal of Applied Econometrics 31 (7), 1407–1429.

[19] O'Donnell, C. (2018). Productivity and Efficiency Analysis: An Economic Approach to Measuring and Explaining Managerial Performance. Springer Nature, Singapore.

[20] Papadopoulos, A., & Parmeter, C. F. (2021). Type II failure and specification testing in the Stochastic Frontier Model. European Journal of Operational Research Volume 293 (3), 990–1001.

[21] Rezek, J., R. Campbell & K. Rogers (2011). Assessing Total Factor Productivity Growth in Sub-Saharan African Agriculture. Journal of Agricultural Economics, 62, 357–374.

[22] Tabak, B. M., & Tecles, P. L. (2010). Estimating a Bayesian stochastic frontier for the Indian banking system. International Journal of Production Economics 125 (1), 96–110.

[23] Tierney, L. (1994). Markov chains for exploring posterior distributions. Annals of Statistics 22 (4), 1701–1728.

[24] Tran, K. C., Tsionas, M. G., & Prokhorov, A. B. (2022). Semiparametric estimation of spatial autoregressive smooth-coefficient panel stochastic frontier models. European Journal of Operational Research, in press.

[25] Tsionas, M. G. (2003). Combining DEA and stochastic frontier models: An empirical Bayes approach. European Journal of Operational Research Volume 147 (3), 499–510.

[26] Tsionas, M., G. (2019). Bounded rationality and thick frontiers in stochastic frontier analysis. European Journal of Operational Research Volume 284 (2), 762–768.

[27] Tsionas, M., G. (2022). Combining data envelopment analysis and stochastic frontiers via a LASSO prior. European Journal of Operational Research, in press.

c:\incomplete_data_inefficiency