Combining Data Envelopment Analysis and Stochastic Frontiers via a LASSO prior

May 10, 2022

Abstract

Technical inefficiencies in stochastic frontier models can be thought of as non-negative parameters. Since, however, their number along with other parameters exceeds the sample size, an adaptive LASSO estimator is a reasonable way to overcome the problem, especially in view of the oracle properties of the estimator under broad conditions. It is shown that the adaptive LASSO estimator can be thought of as the posterior mean of a usual stochastic frontier model with a special prior that benchmarks inefficiencies on known quantities. We take these quantities from DEA scores to obtain technical inefficiencies having oracle properties. The LASSO parameters can be estimated routinely in the Bayesian context without the need for cross-validation. In an application to a data set of large U.S. banks we find that adaptive LASSO outperforms significantly traditional stochastic frontier models.

Key Words: Data Envelopment Analysis; Stochastic Frontier Analysis; Technical Efficiency; Adaptive LASSO; Bayesian Analysis.

Acknowledgments: The author is indebted to three anonymous reviewers for constructive comments on an earlier version.

1 Introduction

Despite recent advances in Stochastic Frontier Models (SFM) and Data Envelopment Analysis (DEA) it is still true, at least for the most part, that SFMs make parametric assumptions and can handle noise in data, whereas DEA does not make parametric assumptions at the cost of not allowing for noise.[1] In this paper, we combine SFM and DEA via the Least Absolute Shrinkage and Selection operator (LASSO, see Tibshirani, 1996). Specifically, we show how a SFM can exploit information from DEA in the form of a LASSO prior, producing an estimator that has so-called "oracle properties", see Fan and Li (2001). With the number of regressors (say p) exceeding the number of observations (n) the fact that LASSO has the oracle property is shown in Zou (2006). In this paper, we propose to combine the SFM with the right LASSO penalty to produce an estimator which is the posterior mean in a suitably defined likelihood function / posterior distribution. The contribution of this paper is that we focus on SFA but we do not specify an ad hoc distribution for inefficiency. Instead, we rely on arguments from the LASSO literature to produce this distribution by benchmarking also on the DEA results. Although Tsionas (2003) used DEA scores to benchmark a half-normal distribution for SFA here we follow a different path which is related directly to Zou (2006), Chernozhukov et al. (2019) and Belloni et al. (2012) to convert a LASSO problem to an SFA model with the equivalent prior. The prior is then benchmarked using the DEA scores while at the same time accounting for an important shortcoming in SFA, viz. that inefficiencies are never zero.

Suppose the model is

$$y_i = \boldsymbol{x}_i' \boldsymbol{\beta} + v_i + u_i, \; i = 1, \ldots, n, \tag{1}$$

where $\boldsymbol{x}_i \in \mathbb{R}^k$ is a vector of regressors with coefficients $\boldsymbol{\beta} \in \mathbb{R}^k$,$v_i$ is a two-sided error term, y_i

[1] In DEA, the contributions of Simar and Wilson (2007, 2011, 2019) and Simar and Zelenyuk (2020) have allowed to take account of noise. In SFM, flexible functional forms have been proposed that make the parametric assumptions considerably less stringent, e.g. Michaelides et al. (2010), Tsionas (2003, 2019), Tsionas and Izzeldin (2018), Tsionas and Mallick (2019), Tsionas and Andrikopoulos (2020), Kumbhakar and Tsionas (2020). An early attempt to combine DEA and SFM is described in Tsionas (2003) where information about DEA scores is incorporated into a parametrically specified distribution of inefficiency.

denotes the $i$the observation on the dependent variable, and $u_i \geq 0$ represents technical inefficiency ($\boldsymbol{u} = [u_1, \ldots, u_n]' \in \mathbb{R}^n_+$). Moreover, let Y denote the available data. viz. $Y = [y_i, \boldsymbol{x}_i, i = 1, \ldots, n]$, $\boldsymbol{y} = [y_1, \ldots, y_n]'$ be the vector containing all observations on the dependent variable, and $\boldsymbol{X} = [x_i', i = 1, ..., n]$ be the $n \times k$ matrix containing all observations on the explanatory variables.

We treat technical inefficiencies as potentially zero for some (potentially several firms) and, therefore, as parameters to be estimated. Since there as many inefficiencies as observations, often a distributional assumption is made on the u_is, for example exponential, half-normal, etc. Such assumptions can often lead to misleading results as they impose structure on things for which very little is known.

Clearly, the number of unknown parameters which include $\boldsymbol{\beta}$ as well as technical inefficiencies (u_is) exceeds the sample size so, we are in the statistical paradigm of so-called "large p, small n", where p is the number of unknown parameters and n is the sample size. A natural solution to the problem is to use a LASSO prior which is known for its good statistical properties. Although the LASSO is often applied to $\boldsymbol{\beta}$ in this paper it is applied to the inefficiencies which constitute the main reason why the number of parameters exceeds the sample size.

The adaptive LASSO has been used in the time-invariant case with panel data but this framework it too restrictive (Jung, 2018, ch. 3). In this paper, we draw on the adaptive LASSO case (e.g. Zou, 2008) to draft a prior that benchmarks on DEA scores. The LASSO is supported by much theoretical work. Donoho, Johnstone, Kerkyacharian, and Picard (1995) proved near-minimax optimality; see also Donoho and Huo (2002), Donoho and Elad (2002), Donoho (2004), Donoho and Johnston (1994), and Meinshausen and Bühlmann (2004). Therefore, based on statistical theory it is expected to provide estimators with good asymptotic properties and, often, with very good finite sample properties. The use of the LASSO to estimate inefficiency is a novel approach to uncovering this latent variable. Specifically, if you look at table 2, even though the Normal-Exponential model is correctly specified it still is outperformed by the adaptive LASSO. This is because the Jondrow et al. (1982; JLMS) estimator is inconsistent as an estimator for u_i (but consistent as an estimator for $E[u_i|\varepsilon_i]$). This is a well known result but we would like to stress its importance here. There is recent work by Zeebari et al. (2021) that discusses some issues with shrinkage and JLMS that offer some interesting connections with the work here. Zeebari et al. (2021) notice that the standard SFM approach shrinks inefficiency towards its mean or mode. This generates a distribution that is different from the distribution of the *unconditional* inefficiency and, therefore, the accuracy of estimated inefficiency is negatively correlated with the difference between inefficiency and its mean or mode. The type of shrinkage that LASSO provides versus JLMS can be explained in this framework. One reason the LASSO performs well is that the shrinkage is towards zero regardless of the values provided by the LASSO and it is towards the mean regardless of the values provided by JLMS. In our simulations the value of the mean implies that the shrinkage works in favor of LASSO initially, and a larger mean actually works more in favor of JLMS. So, the LASSO is quite flexible and works well in this framework.

2 Models

Referring only to regression parameters β, the adaptive LASSO of Zou (2006) relies on the optimization problem:

$$\min_{\beta \in \mathbb{R}^k} \sum_{i=1}^{n} (y_i - \boldsymbol{x}_i' \boldsymbol{\beta})^2 + \lambda_n \sum_{j=1}^{k} \hat{w}_j |\beta_j|, \tag{2}$$

where $\hat{w}_j = |\hat{\beta}_j|^{-\gamma}$ (for some $\gamma > 0$), $\hat{\beta}$ is a consistent estimator like least squares (LS), and λ_n is a constant such that $\frac{\lambda_n}{\sqrt{n}} \to 0$. Constants λ_n and γ are selected by cross-validation. Another method relies on the same optimization problem but using

$$\hat{w}_j = \sqrt{n^{-1} \sum_{i=1}^{n} x_{i,j}^2}, \tag{3}$$

(e.g. Chenozhukov, Hansen, and Spindler, 2019), see also Belloni, Chernozhukov, and Hansen (2012). Applying LASSO to $\boldsymbol{\beta}$ is often the case when the number of regressors is large. If this is not a concern, but interest focuses instead on technical inefficiencies, the problem becomes

$$\min_{\beta \in \mathbb{R}^k, u \in \mathbb{R}^n_+} \sum_{i=1}^{n} (y_i - u_i - \boldsymbol{x}_i' \boldsymbol{\beta})^2 + \lambda_n \sum_{i=1}^{n} \hat{w}_i u_i, \tag{4}$$

where a typical choice would be to set $\hat{w}_i = \tilde{u}_i^{-\gamma}$, given some "initial estimates" $\tilde{u}_i$. These initial estimates are the DEA scores. This is based on the prior notion that inefficiencies are likely to be "small" so we begin by giving the data the "benefit of the doubt" in the sense that we are willing to assume a mostly efficient sector or set of decision-making-units.

In the following we define $\lambda_n = \frac{\lambda}{\sqrt{n}}$ This problem is equivalent to assuming (1) with the assumption $v_i \sim \mathcal{N}(0, \sigma_v^2)$ and $p(u_i) = \frac{\lambda}{\sqrt{n}} \tilde{w}_i e^{-\frac{\lambda}{\sqrt{n}} \tilde{w}_i u_i}$ ($\lambda > 0$, $u_i, \tilde{w}_i \geq 0$) which is an exponential distribution with parameter $\lambda \tilde{w}_i$. However, some DEA benchmarks can be zero indicating full efficiency so we modify the weights to

$$\tilde{w}_i = (c + \tilde{u}_i)^{-\gamma}, \tag{5}$$

for some $c > 0$. This constant can be used to "re-benchmark" the zero scores.

A flat prior can be specified, for example, as:

$$p(\boldsymbol{\beta}, \lambda, \sigma_v, c, \gamma) \propto \lambda^{-1} \sigma_v^{-1} e^{-c^2/(2\bar{h})}. \tag{6}$$

The prior is flat on all parameters except c for which we assume a half-normal distribution, viz. $c \sim \mathcal{N}_+(0, h^2)$ where we set $\bar{h} = 0.1$ to take account of the notion that this parameter is likely to be "small".

First, we omit the dependence on c and γ so, the augmented posterior distribution has density

$$\begin{gathered} p(\boldsymbol{\beta}, \sigma_v, \lambda, \boldsymbol{u}|Y) \propto \\ \sigma_v^{-(n+1)} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n \left(y_i - u_i - \boldsymbol{x}_i'\boldsymbol{\beta}\right)^2\right\} \lambda^{n-1}\sum_{i=1}^n \exp\left\{-\frac{\lambda}{\sqrt{n}}\tilde{w}_i u_i\right\}. \end{gathered} \tag{7}$$

The posterior conditional distribution of β is

$$\boldsymbol{\beta}|\sigma_v, \lambda, u, Y \sim \mathcal{N}_k\left(\boldsymbol{b}, \boldsymbol{V}\right), \tag{8}$$

where $\boldsymbol{b} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{y}-\boldsymbol{u})$, $\boldsymbol{V} = \sigma_v^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$. The posterior conditional distribution of σ_v^2 is

$$\frac{(\boldsymbol{y}-\boldsymbol{u}-\boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y}-\boldsymbol{u}-\boldsymbol{X}\boldsymbol{\beta})}{\sigma_v^2}|\boldsymbol{\beta}, \lambda, \boldsymbol{u}, Y \sim \chi_n^2, \tag{9}$$

see also Tsionas (2003).

The posterior conditional distribution of λ is:

$$\lambda|\boldsymbol{\beta}, \sigma_v, u, Y \sim \mathcal{G}\left(n,\ \frac{1}{\sqrt{n}}\sum_{i=1}^n \tilde{w}_i u_i\right), \tag{10}$$

where $\mathcal{G}$ denotes the gamma distribution. Finally, the posterior conditional distribution of each u_i is given by

$$u_i|\boldsymbol{\beta}, \sigma_v, \lambda, Y \sim \mathcal{N}_+\left(-\frac{1}{2}\left(e_i - \frac{\lambda}{\sqrt{n}}\sigma_v^2\tilde{w}_i\right),\ \sigma_v^2\right),\ i = 1, \ldots, n, \tag{11}$$

where $e_i = y_i - \boldsymbol{x}_i'\boldsymbol{\beta}$, and $\mathcal{N}_+$ denotes the truncated normal distribution in $\mathbb{R}_+$.

If we have to treat c and γ as unknown parameters, their posterior conditional distributions are as follows. For c we have

$$p(c|\beta, u, \sigma_v, \lambda, Y) \propto \exp\left\{-\frac{\lambda}{\sqrt{n}}\sum_{i=1}^n u_i(c+\tilde{u}_i)^{-\gamma} - \frac{c^2}{2h}\right\},\ c > 0, \tag{12}$$

The posterior conditional distribution of γ is

$$p(\gamma|\boldsymbol{\beta}, \boldsymbol{u}, \sigma_v, \lambda, c, Y) \propto \exp\left\{-\frac{\lambda}{\sqrt{n}}\sum_{i=1}^{n} W_i^{-\gamma} u_i\right\},\ \gamma > 0, \tag{13}$$

where $W_i = c + u_i$. Although (12) and (13) are not in known families, both are log-concave and, therefore, highly efficient specialized algorithms can be used to provide random drawings (e.g. Gilks and Wild, 1992).

Cycling through the various posterior conditional distributions produces a sample from the distribution whose density is proportional to (7), a procedure known as Gibbs sampling, see Gelfand and Smith (1990).

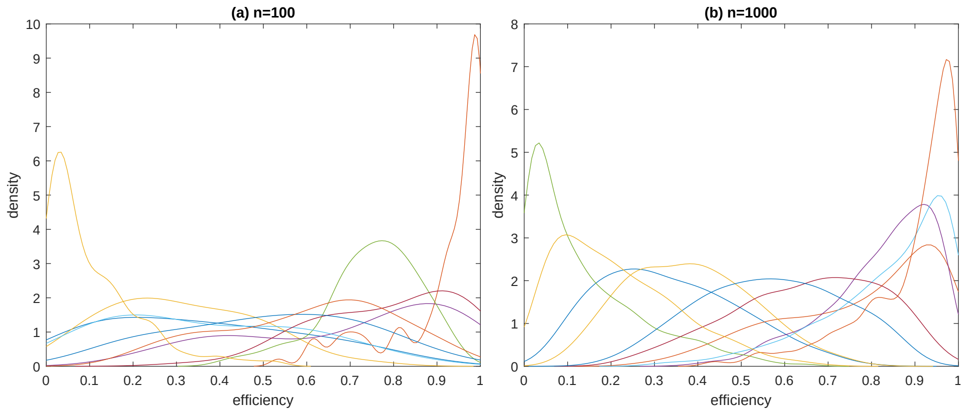
3 Monte Carlo experiment

To validate the new techniques we consider a Monte Carlo experiment with five regressors in $\boldsymbol{x}_i$ generated as standard uniform, coefficients $\boldsymbol{\beta}$ all equal to 0.25, and measurement error $v_i \overset{iid}{\sim} \mathcal{N}(0, \sigma_v^2)$. We generate inefficiencies from a data generating process (DGP) which does not correspond to half-normal or exponential models but rather it is

$$\begin{aligned} r_i &= \Phi(\boldsymbol{z}_i'\boldsymbol{\gamma}) \\ u_i &= -\log r_i,\ i = 1, \ldots, n, \end{aligned} \tag{14}$$

where $\boldsymbol{z}_i$ is an $n \times m$ vector containing standard uniform random variates, $\boldsymbol{\gamma}$ is an associated parameter vector whose elements are generated from standard normal distributions, r_i denotes efficiency, viz. $r_i = e^{-u_i}$, and $\Phi(\cdot)$ denotes the standard normal distribution function. This DGP corresponds to the efficiency modeling of Paul and Shankar (2018), further developed in Tsionas and Mamatzakis (2019). By construction, efficiency is in the interval (0,1). Typical efficiency distributions resulting from this

Figure 1: Typical efficiency distributions



Notes: Presented are typical efficiency distributions from the DGP in (14) for $n = 100$ in panel (a) and $n = 1000$ in panel (b).

DGP are presented in Figure 1, for $n = 100$ in panel (a) and $n = 1000$ in panel (b). Clearly, many different shapes are allowed by this DGP. The average standard deviations of technical inefficiency are 0.56 and 0.46 in the two cases, respectively. Therefore we use three values for σ_v; viz. 0.1, 0.25 and 0.5 for low, moderate and excessive noise.

The Monte Carlo experiments are replicated 10,000 times and the results are presented in Table 1 where we report mean absolute errors (MAE) between true and estimated inefficiency for the normal-half-normal (NHN) model, the normal-exponential (NEX) model, DEA, and adaptive LASSO.

From the Monte Carlo results, it turns out that (i) MAEs are roughly the same for normal-half-normal and the normal-exponential models, (ii) DEA becomes more inaccurate as noise levels increase and is, generally, more prone to estimation bias compared to the SFA models which is, of course, expected, and (iii) the adaptive LASSO has the smallest MAE compared to the other three methods although its performance deteriorates as the noise levels increase. The bias of adaptive LASSO for a

Table 1: Monte Carlo results: Mean Absolute Errors

		$n = 100$	$n = 250$	$n = 500$	$n = 1000$
$\sigma_v = 0.1$	NHN	0.071	0.068	0.057	0.050
	NEX	0.067	0.061	0.054	0.048
	DEA	0.085	0.072	0.065	0.058
	Adaptive LASSO	0.073	0.067	0.055	0.044
$\sigma_v = 0.25$	NHN	0.107	0.115	0.099	0.096
	NEX	0.103	0.114	0.110	0.094
	DEA	0.154	0.149	0.135	0.120
	Adaptive LASSO	0.094	0.085	0.081	0.077
$\sigma_v = 0.50$	NHN	0.180	0.170	0.168	0.158
	NEX	0.181	0.170	0.169	0.160
	DEA	0.187	0.172	0.168	0.155
	Adaptive LASSO	0.114	0.103	0.098	0.088

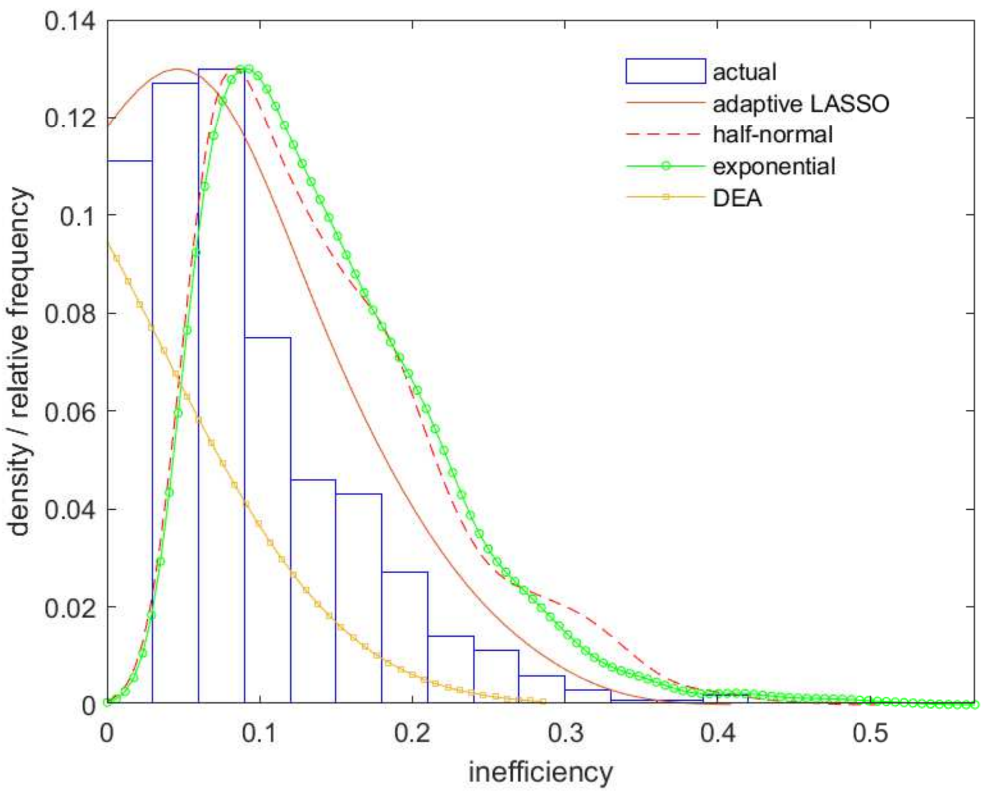
Notes: Reported are mean absolute errors (MAE) between true and estimated inefficiency for the normal-half-normal (NHN) model, the normal-exponential (NEX) model, DEA, and adaptive LASSO.

given sample size is quite acceptable as, for example, when $m = 1000$ its MAE is 0.044, 0.077, and 0.088 as σ_v increases from 0.1 and 0.25 to 0.5, respectively. For the same sample size, the other three methods deliver MAEs close to 0.05, 0.10 and 0.16 for the different values of σ_v, respectively. n this sense, the adaptive LASSO performs well in a DGP that does not favor any particular model on prior grounds.

Another interesting finding is that the adaptive LASSO model provides estimates that are, evidently, different from DEA despite the fact that it anchors on DEA scores. Scores from SFA are quite close but their rank correlations with DEA become lower as σ_v increases. The reason is that the adaptive LASSO model anchors on DEA scores but this prior is revised by the data and the final results depend not only on the prior but the data as well –the extent of noise more precisely. The prior is also quite flexible as (i) it depends on two parameters (c and γ) and it adapts to the data through posterior inference, (ii) it is itself a data-based prior (or an empirical Bayes prior) that anchors on DEA scores. A typical example from Monte Carlo simulations is reported in Figure 2, with $\sigma_v = 0.25$ (so that noise is moderate), and $n = 1000$ where, indeed, the results from the adaptive LASSO are closer to the true inefficiency distribution as opposed to normal-half-normal, normal-exponential, and DEA.

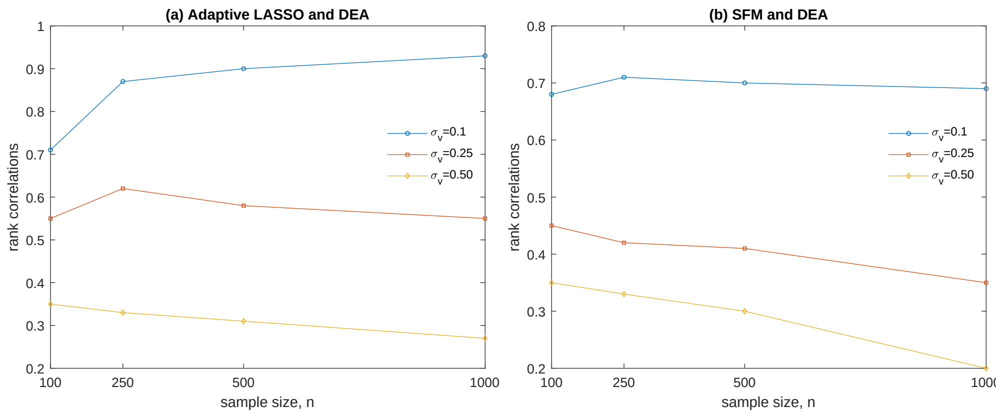
Regarding rank correlations, we present them graphically in Figure 3 in the interest of not overcrowding Table 1. Rank correlations (medians across Monte Carlo simulations in Table 1) between Adaptive LASSO and DEA (reported in panel (a)) are generally high and increase with the sample size when $\sigma_v = 0.1$ but this is reversed for higher values of σ_v and , in fact, correlations decrease with the sample size. Rank correlations between DEA and SFM are reported in panel (b). We choose the normal-half-normal model as the normal-exponential model provided similar results. These rank correlations are generally high when $\sigma_v = 0.1$ but decrease rapidly with the sample size as σ_v increases to 0.25 and 0.50.

Figure 2: Typical example from Monte Carlo simulations



Notes: Presented is a typical example from Monte Carlo simulations with $\sigma_v = 0.25$ (so that noise is moderate), and $n = 1000$.

Figure 3: Rank correlations from Monte Carlo experiment



Notes: Rank correlations between Adaptive LASSO and DEA are reported in panel (a). Rank correlations between DEA and SFM are presented in panel (b). For SFM, we choose the normal-half-normal model as the normal-exponential model provided similar results. Rank correlations are medians across all Monte Carlo simulations in Table 1.

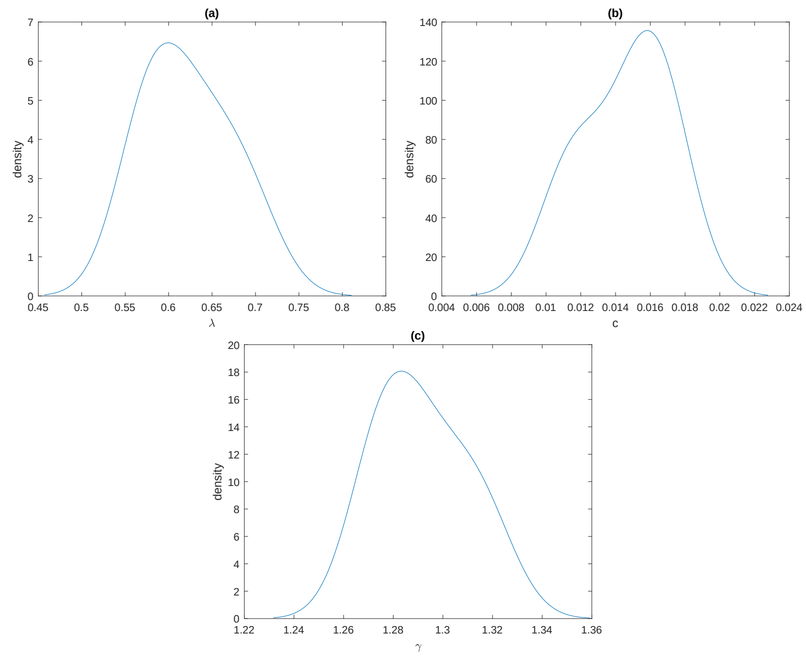
4 Empirical application

We have an unbalanced panel with 2,397 bank-year observations for 285 large U.S. commercial banks operating in 2001-2010, whose total assets were more than one billion dollars (in 2005 U.S. dollars) in the first three years of observation. We use the data for 2009 to focus on the effects of the financial crisis. The data come from Call Reports available from the Federal Reserve Bank of Chicago and they have been used in Malikov et al. (2016). For detailed description of the data construction, see Section 5 of their paper.

Inputs and outputs are as follows: y_1 Consumer Loans, y_2 Real Estate Loans, y_3 Commercial & Industrial Loans, y_4 Securities, y_5 Off-Balance Sheet Activities Income, x_1 Labor, number of full-time employees, x_2 Physical Capital (Fixed Assets), x_3 Purchased Funds, x_4 Interest-Bearing Transaction Accounts, x_5 Non-Transaction Accounts.

We estimate a translog cost function with five input prices, five outputs, and we include a time trend to capture the effects of technical change. We benchmark on the score from input-oriented DEA

Figure 4: Marginal posterior densities



Notes: Reported are marginal posterior densities of λ in panel (a), c in panel (b), and γ in panel (c).

with variable returns to scale (Banker et al., 1984).

In Figure 3 we report marginal posterior densities of the key LASSO parameters λ (panel (a)), c (panel (b)), and γ (panel (c)). Parameter λ ranges from 0.45 to 0.82, c ranges from 0.006 to 0.023 and γ from 1.23 to 1.36 and they seem to be asymmetric, thus making asymptotic-based normal inferences somewhat misleading.

Posterior moments are reported in Table 2. From Table 2, the posterior means (posterior standard deviations) of λ, c, γ are respectively 0.621 (0.047), 0.015 (0.002), and 1.291 (0.017). Therefore, from (5), modified weights are, approximately, $\tilde{w}_i = (0.015 + \tilde{u}_i)^{-1.291}$ where $\tilde{u}_i$ are the DEA scores. The weights are critical in both (4) and (7).

In Figure 4 we report sample densities of efficiency estimates from a half-normal SFM, an exponential SFM, scores from DEA as well as from the novel adaptive LASSO approach.

Efficiency estimates from SFM using either the exponential or half-normal distributions are very different compared to DEA scores. This can be seen from panel (a) of Figure 4. In panel (b) we provide a comparison with three other popular functional forms, the normalized Quadratic (Diewert and, Wales, 1987), the Symmetric Generalized McFadden (SGM, Diewert and, Wales, 1987 and Kumbhakar, 1994), the Generalized Leontief (GL, Diewert and, Wales, 1987) and the flexible Fourier form due to Gallant (1984). For the Fourier flexible form we follow the implementation in Feng and Serletis (2008).

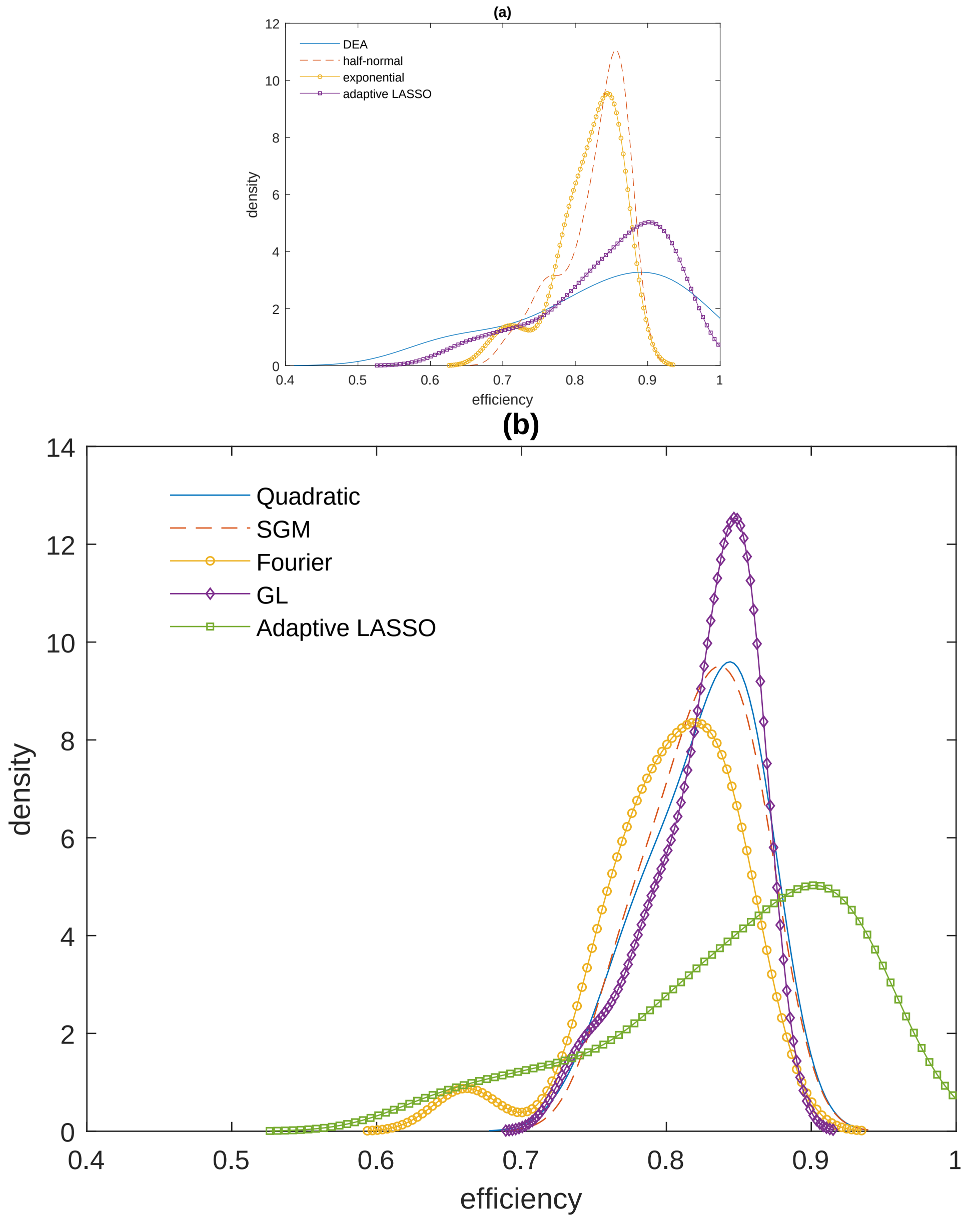
Therefore, our results are robust to alternative cost functions.[2] The adaptive LASSO model seems to be anchoring more on DEA but its estimates are quite different, the rank correlation coefficient being, approximately, 0.65. Scores from SFMs are quite close but their rank correlations with DEA are below 0.50. The adaptive LASSO prior is essentially an empirical Bayes estimator that utilizes the data by benchmarking on DEA so, it is expected to perform better than SFM models. To find whether this is true more generally, a Monte Carlo experiment is now included in the revised version. However, it is worth mentioning that as $\sigma_v \to 0$ we do expect SFM and DEA to provide similar results. When this is not the case, DEA that does not account for noise will not deliver scores similar to the SFM estimates. These results hold more generally as we saw in the Monte Carlo experiments.

In Figure 5, we present sample distributions of posterior mean estimates for technical change, efficiency change, and productivity growth. The models have quite different implications for technical change (panel (a)), efficiency change (panel (b)) and productivity growth (estimated as the sum of the other two, in panel (c)). DEA and adaptive LASSO estimates are quite close to zero ranging between $\pm 1\%$, whereas estimates for the SFMs are two to three times larger in magnitude.

An interesting feature of the Bayesian approach is that we can compute marginal likelihoods and

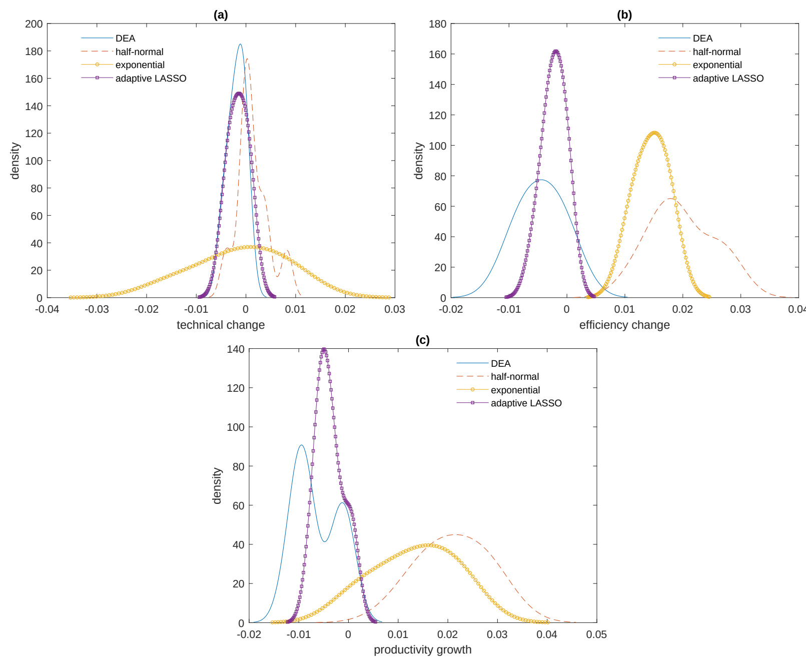
[2] We use 150,000 iterations of the Gibbs sampler omitting the first 50,000 to mitigate possible start up effects. Convergence and performance of MCMC are monitored using the standard Geweke (1992) diagnostics.

Figure 5: Efficiency scores / estimates



Notes: Reported in panel (a) are sample distributions of efficiency scores DEA, a nonrmal-half-normal SFM, a normal-exponential model, and the adaptive LASSO model. In panel (b) we provide a comparison with three other popular functional forms, the normalized Quadratic (Diewert and, Wales, 1987), the Symmetric Generalized McFadden (SGM, Diewert and Wales, 1987; Kumbhakar, 1994), the Generalized Leontief (GL, Diewert and, Wales, 1987) and the flexible Fourier form due to Gallant (1984).

Figure 6: Quantities of interest



Notes: Reported are sample distributions of technical change (TC, panel (a)), efficiency change (EC) in panel (b) and productivity growth (PG=TC+EC) in panel (c), from DEA, the normal-half-normal SFM, the normal-exponential SFM and the adaptive LASSO model.

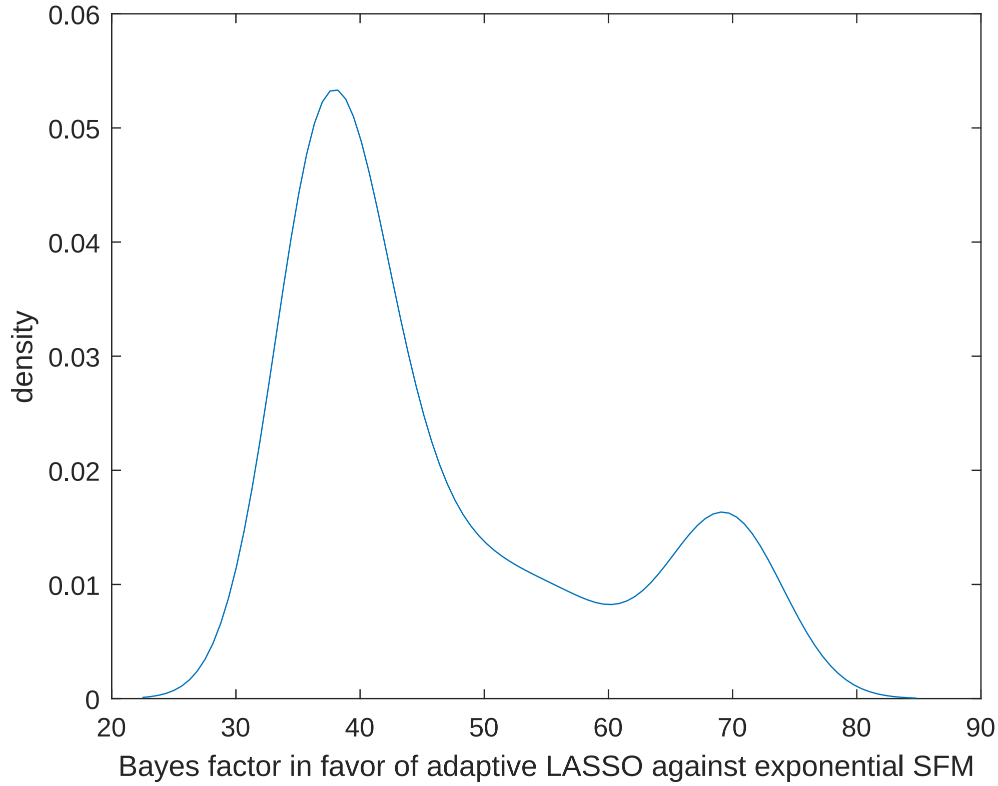
Bayes factors quite easily. Before proceeding, we should mention that marginal likelihoods and Bayes factors are standard tools in model comparison from the Bayesian viewpoint. From the practical perspective, we think it necessary to provide verbatim the following recent result from Fong and Holmes (2020). "In Bayesian statistics, the marginal likelihood, also known as the evidence, is used to evaluate model fit as it quantifies the joint probability of the data under the prior. In contrast, non-Bayesian models are typically compared using cross-validation on held-out data, either through k-fold partitioning or leave-p-out subsampling. We show that the marginal likelihood is formally equivalent to exhaustive leave-p-out crossvalidation averaged over all values of p and all held-out test sets when using the log posterior predictive probability as the scoring rule" (Fong and Holmes, 2020, p. 489). This result is quite strong even for convinced Bayesians and shows that the marginal likelihood criterion and, therefore Bayes factors as well, has quite important properties related to exhaustive subsampling.

For a model with parameters $\boldsymbol{\theta} \in \Theta$, data Y, likelihood function $L(\boldsymbol{\theta}; Y)$ and prior $p(\boldsymbol{\theta})$, the marginal likelihood is $M(Y) = \int_{\Theta} L(\boldsymbol{\theta}; Y) p(\boldsymbol{\theta})\, \mathrm{d}\boldsymbol{\theta}$ and can be estimated using the method of Chib (1995). For any two models the Bayes factor (or posterior odds ratio when the prior odds are 1:1) is $BF_{1:2} = \frac{M_1(Y)}{M_2(Y)}$. The Bayes factors are reported in the lower panel of Table 2.

Relative to the SFM with exponentially distributed one-sided error term, the half-normal specification has a Bayes factor of 1.817 (indicating that it is slightly better but not significantly so) whereas the adaptive LASSO model has a Bayes factor of 35.42, approximately, which is quite significant (roughly Bayes factors in excess of 10 are considered as indicating considerable support in the light of the data, e.g. Kass and Raftery, 1995).

To make sure the comparison is fair, we omit at random B blocks of banks (viz. all temporal observations for a given bank) at a time, where B is itself randomly chosen between 1 and 10. We repeat this exercise 1,000 times, re-computing Bayes factors in each instance and we report the results in Figure 6.

Figure 7: Sample distribution of Bayes factors



Notes: Reported is the sample distribution of Bayes factors of adaptive LASSO against the normal-exponential SFM. The procedure is as follows. We omit at random B blocks of banks (viz. all temporal observations for a given bank) at a time, where B is itself randomly chosen between 1 and 10. We repeat this exercise 1,000 times, re-computing Bayes factors in each instance.

Table 2: Posterior moments and Bayes factors

Parameter / Model	
posterior moments	
λ	0.621
	(0.047)
c	0.015
	(0.002)
γ	1.291
	(0.017)
Bayes factors	
Exponential	1.000
Half-Normal	1.817
adaptive LASSO	35.420

Notes: Reported are posterior means (with posterior standard deviations in parentheses) for parameters λ, c, γ (upper panel) as well as Bayes factors of the normal-half-normal and the adaptive LASSO relative to the normal-exponential model.

From Figure 6, it turns out that a Bayes factor of 35 is somewhat lower than the median 41.2 obtained by using this "bootstrap" procedure. In fact, the distribution of Bayes factors is bimodal with the two modes near 40 and 70. Of course, the results are still in favor of the adaptive LASSO model.

Concluding remarks

In this paper we propose a Bayesian alternative to estimating a adaptive LASSO model which is a way of combining SFM and DEA. SFM accounts for noise in the data but makes parametric assumptions whereas most DEA models do not make parametric assumptions but allowing for noise is problematic. One way to combine the methods, is to use an adaptive LASSO model which is shown to be equivalent to a SFM with a special prior on technical efficiencies. This allows the routine use of standard Bayesian methods organized around Gibbs sampling as well as formal model comparison based on marginal likelihoods and Bayes factors. An application to large U.S. banks shows that adaptive LASSO delivers results that are between those from DEA and SFMs and that the different models have substantively different implications in terms of technical change, efficiency change, and productivity growth. The reason is that (i) the adaptive LASSO uses a two-parameter prior that adapts well to the data through posterior inference, and (ii) that it is, essentially, an empirical Bayes prior that uses the data in the form of DEA scores but at the same time accounts for noise or measurement error in the data. Monte Carlo evidence further suggests that this conclusion holds more generally and, in fact, the adaptive LASSO performs very well for moderately sized samples as well as for low to high variance values of the two-sided error term.

References

[1] 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.

[2] Belloni, A., Chen, D., Chernozhukov, V., & Hansen, C. (2012). Sparse Models and Methods for Optimal Instruments with an Application to Eminent Domain, Econometrica, 80, 2369–2429.

[3] Chernozhukov, V., Hansen, C., & Spindler, N. (2019). High-dimensional metrics in R. Unpublished manuscript, https://arxiv.org/abs/1603.01700

[4] Chib, S. (1995). Marginal likelihood from the Gibbs output. Journal of the American Statistical Association 90 (432), 1313–1321.

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

[6] Donoho, D. (2004). For Most Large Underdetermined Systems of Equations, the Minimal 1-Norm Solution is the Sparsest Solution. Technical Report, Stanford University, Dept. of Statistics.

[7] Donoho, D., & Elad, M. (2002), Optimally Sparse Representation in General (Nonorthogonal) Dictionaries via 1-Norm Minimizations. Proceedings of the National Academy of Science USA, 1005, 2197–2202.

[8] Donoho, D., & Huo, X. (2002). Uncertainty Principles and Ideal Atomic Decompositions. IEEE Transactions on Information Theory, 47, 2845–2863.

[9] Donoho, D., &d Johnstone, I. (1994), Ideal Spatial Adaptation via Wavelet Shrinkages. Biometrika 81, 425–455.

[10] Donoho, D., Johnstone, I., Kerkyacharian, G., & Picard, D. (1995). Wavelet Shrinkage: Asymptopia? (with discussion), Journal of the Royal Statistical Society, Series B 57, 301–337.

[11] Fan, J., & Li, R. (2001). Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties. Journal of the American Statistical Association 96, 1348–1360.

[12] Feng , H., and A. Serletis (2008). 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.

[13] Fong, E., & Holmes, C. C. (2020). On the marginal likelihood and cross-validation. Biometrika 107 (2), 489–496.

[14] Gallant, S. G. M. (1984). The Fourier Flexible Form. American Journal of Agricultural Economics 66 (2), 204–208.

[15] Gelfand, A. E., & Smith, A. F. M. (1990). Sampling-Based Approaches to Calculating Marginal Densities, Journal of the American Statistical Association, 85, 398–409.

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

[17] Gilks, W.R., & Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. Applied Statistics 41, 337–348.

[18] Jondrow, J., Lovell, C.A.K., Materov, I.S., Schmidt, P. (1982). On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Function Model. Journal of Econometrics, 19, 233–238.

[19] Jung, H. (2018). Production, efficiency and managerial selection of workers into peer networks. Ph.D. thesis, Syracuse University, ALL.856.

[20] Kass, R. E., & A. E. Raftery (1995). Bayes Factors. Journal of the American Statistical Association 90 (430), 773–795.

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

[22] 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.

[23] Malikov, E., Kumbhakar, S. C., & Tsionas, M. G. (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.

[24] Michaelides, P. G., Vouldis, A. T., & Tsionas, M. G. (2010). Globally flexible functional forms: The neural distance function. European Journal of Operational Research 206 (2), 456–469.

[25] Paul, S., & Shankar, S. (2018). On estimating efficiency effects in a stochastic frontier model. European Journal of Operational Research 271 769–774.

[26] Simar L., & Wilson, P. W. (2007). Estimation and inference in two-stage, semi-parametric models of productive efficiency. Journal of Econometrics 136, 31–64.

[27] Simar, L., & Wilson, P. W. (2011). Two-stage DEA: caveat emptor. Journal of Productivity Analysis 36, 205–218.

[28] Simar L., and Wilson, P. W. (2019). Technical, allocative and overall efficiency: Estimation and inference. European Journal of Operational Research 28 (3), 1164–1176.

[29] Simar, L., and Zelenyuk, V. (2020). Improving finite sample approximation by central limit theorems for estimates from Data Envelopment Analysis. European Journal of Operational Research 284 (3), 1002–1015.

[30] Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso, Journal ofthe Royal Statistical Society, Series B 58, 267–288.

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

[32] Tsionas, M. G. (2019). A coherent approach to Bayesian Data Envelopment Analysis. European Journal of Operational Research 281 (2), 439–448.

[33] Tsionas, M. G., & Andrikopoulos, A. (2020). On a High-Dimensional Model Representation method based on Copulas. European Journal of Operational Research 284 (3), 967–979.

[34] Tsionas, M. G., & Izzeldin, M. (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.

[35] Tsionas, M. G., & Mallick, S. (2018). A Bayesian semiparametric approach to stochastic frontiers and productivity. European Journal of Operational Research 274 (1), 391–402.

[36] Tsionas, M. G., & Mamatzakis, E. (2019). Further results on estimating inefficiency effects in stochastic frontier models. European Journal of Operational Research 275 (3), 1157–1164.

[37] Zeebari, Z., Månsson, K., Sjölander, P., Söderberg, M. (2021) (2021). Regularized Conditional Estimators of Unit Inefficiency in Stochastic Frontier Analysis, with Application to Electricity Distribution Market. https://ideas.repec.org/p/hhs/ratioi/0345.html

[38] Zou, H. (2006). Adaptive Lasso and its Oracle Properties. Journal of the American Statistical Association 101 (476), 1418–1429.