Centre for Efficiency and Productivity Analysis

Working Paper Series
No. WP182021

Goodness-of-fit in Optimizing Models of Production:

A Generalization with a Bayesian Perspective

Mike Tsionas and Valentin Zelenyuk

Date: December 2021

School of Economics
University of Queensland
St. Lucia, Qld. 4072
Australia

ISSN No. 1932 - 4398

Goodness-of-fit in Optimizing Models of Production: A Generalization with a Bayesian Perspective

Mike Tsionas* Valentin Zelenyuk†

December 9, 2021

Abstract

We propose a very general approach for modeling production technologies that allows for modeling both inefficiency and noise that are specific for each input and each output. The approach is based on amalgamating ideas from nonparametric activity analysis models for production and consumption theory with stochastic frontier models. We do this by effectively re-interpreting the activity analysis models as simultaneous equations models in Bayesian compression and artificial neural networks frameworks. We make minimal assumption about noise in the data and we allow for flexible approximations to input- and output-specific slacks. We use compression to solve the problem of an exceeding number of parameters in general production technologies and we also incorporate environmental variables in the estimation. We present Monte Carlo simulation results and empirical illustration and comparison of this approach for US banking data.

*Mike Tsionas, Montpellier Business School & Lancaster University Management School, LA1 4YX, U.K.

†Valentin Zelenyuk, School of Economics, The University of Queensland, Brisbane, Australia; v.zelenyuk@uq.edu.au. The support from the Australian Research Council (FT170100401) and The University of Queensland is acknowledged. Feedback from Bao Hoang Nguyen, Evelyn Smart and Zhichao Wang is appreciated. These individuals and organizations are not responsible for the views expressed here.

1 Introduction

In his seminal work, Afriat (1967) launched a flourishing stream of literature about non-parametric testing of regularity conditions in economics. Originally cast in a consumer economics context, it paved the way for non-parametric demand analysis and, in particular for testing the rationality of consumer behavior (Varian (1982, 1983) and, more recently, Cherchye et al. (2007), Kitamura and Stoye (2018) and Smeulders et al. (2021), etc.).

In another seminal work, Afriat (1972) developed a similar approach for the context of production economics, in parallel with Hanoch and Rothschild (1972). This generated a series of articles, e.g., Diewert (1973), Varian (1984), Diewert and Parkan (1985), that took these ideas further, developing various non-parametric tests of consistency with the optimizing behavior of firms using demand and supply data, as well as various restrictions on technology, such as returns to scale, separability, homotheticity, etc. In a follow up paper, Varian (1990) proposed a "goodness-of-fit" measure and illustrated it for the context of aggregate demand estimation, arguing that:

> "... a more fruitful approach to testing optimizing behavior is to measure the departure from optimization using the estimated objective function, and see whether this departure [measured via a goodness-of-fit measure] is significant in an economic sense ..."

Leveraging on these works, Färe and Grosskopf (1995) unveiled the more explicit links between the Afriat-Hanoch-Rothschild-Diewert-Parkan-Varian approach and the works on the duality of Shephard (1953, 1970) and the efficiency analysis literature pioneered by Farrell (1957), concluding that:

> "... one may view Farrell efficiency measures ... when computed using the programming approach as providing nonparametric tests of technology which are also measures of goodness-of-fit."

The goal of this paper is to revisit this literature with recently developed insights into efficiency analysis and in Bayesian statistics, with a goal to develop more robust goodness-of-fit measures that will complement the previously developed analogues, and try to overcome their limitations, such as sensitivity to outliers.

Our paper is structured as follows: Section 2 briefly describes the related literature. Section 3 outlines the economic theory underpinnings. Section 4 describes the econometric model and the estimation approaches, starting with the deterministic convex hulls and then generalizing it to the stochastic convex hulls representation. It also describes how such models can be estimated within a Bayesian compression framework and proposes several measures of goodness-of-fit (economic and statistical). Section 5 describes how to model the determinants of inefficiency via flexible Bayesian artificial neural networks (ANN). Section 6 presents an empirical illustration of the introduced approach to real data, for US banking. Section 7 gives concluding remarks. Monte Carlo evidence can be found in Supplement 1 while the further technical details can be found in Supplement 2.

2 Related Literature in a Nutshell

Most of the literature on non-parametric modeling and testing of regularity conditions in production contexts is based on the linear programming approach, via a local-linear approximation of technology. This approach started as "activity analysis" models (Debreu (1951), Koopmans (1951), Farrell (1957)) and these were extended and popularized in business and the OR/MS community as Data Envelopment Analysis (DEA) by Charnes et al. (1978), with many variants developed since then.[1] Its key advantages include relative simplicity and absence of parametric assumptions on the frontier or on the inefficiency. The main caveat of such an approach is not accounting for statistical noise and, in particular, possibly high sensitivity to outliers. Addressing this caveat was the main idea behind the origin of Stochastic

[1] E.g., for a textbook style description, see Ray (2004) or Sickles and Zelenyuk (2019, Chapters 8-10) and many references therein.

Frontier Analysis (SFA): inspired by Farrell (1957) and the discussion articles that followed, SFA came to light thanks to Aigner et al. (1977) and Meeusen and van den Broeck (1977) and was developed further in many studies.[2]

As with any method, parametric SFA also has its caveats—they are typically related to the requirements for the parametric assumptions about the frontier, or about the inefficiency, or about the statistical noise, or about all of these three aspects. In turn, these caveats inspired a series of studies that attempted to address some or all of these limitations in various ways, bridging the two approaches by proposing various alternatives. Examples of these include various versions of semiparametric and nonparametric SFA approaches,[3] order-m and order-α quantile frontier approaches[4] and stochastic DEA and StoNED.[5]

Our current paper is positioned in the latter stream—it also proposes a hybrid, where we try to retain flexibility as well as allow for statistical noise. To do so, we formulate the production problem as simultaneous equations models, and then cast it in the Bayesian compression framework of Guhaniyogi and Dunson (2015) as well as the artificial neural networks (ANN) approach (Hornik et al. (1989) and White (1989, 1990)). This allows minimal assumptions about noise in the data and yet also provides flexibility of approximations to input- and output-specific slacks. The use of Bayesian compression also helps to solve the problem of the large number of parameters in the general production technology formulation. We also consider the case of "environmental variables" that can be used to model and explain the inefficiency within such an approach. To obtain the posteriors, we use state-of-the-art Markov Chain Monte Carlo (MCMC) methods based on Metropolis Adjusted Langevin Algorithm (Durmus et al. (2017)). The Monte Carlo results and the empirical illustration suggest the proposed approach gives a promising path to improving the toolbox for the non-parametric testing of regularity conditions and the related goodness-of-fit measures.

[2]E.g., see Kumbhakar and Schmidt (2016) or a textbook style description in Sickles and Zelenyuk (2019, Chapters 11-16), and Kumbhakar et al. (2021a,b) for recent reviews and many references therein.

[3]E.g., see Fan et al. (1996), Kumbhakar et al. (2007), Martins-Filho and Yao (2015), Simar et al. (2017) and Parmeter and Zelenyuk (2019) for a recent review and comparison.

[4]E.g., see Cazals et al. (2002), Daouia and Simar (2007) and Bădin et al. (2012).

[5]E.g., see Simar and Zelenyuk (2011) and Kuosmanen and Johnson (2017).

3 Economic Theory Underpinnings

The economic theory foundations for the non-parametric production analysis were described in a few classical papers, e.g., Afriat (1972), Hanoch and Rothschild (1972), Varian (1984), Chavas and Cox (1990), Färe and Grosskopf (1995), etc., with different emphases and in different styles. Here we will be closer to the style of Färe and Grosskopf (1995), with important generalizations, as we find it more convenient for our purposes of engaging it with the efficiency analysis approach (via a hybrid of DEA and SFA). Specifically, recall that Färe and Grosskopf (1995) focused on the cost-minimization approach, while we will focus on profit-maximization and then decompose it all the way to a Debreu-Farrell-type technical efficiency measure by adapting the recent developments in Färe et al. (2019) and Färe and Zelenyuk (2020). The latter decomposition is particularly useful as it connects the profit maximization paradigm with the Debreu-Farrell-type measurement that is taken when prices are not observed (as in our case).

To facilitate formal discussion, let $x = (x_1, ..., x_N)' \in \mathbb{R}^N_+$ and $y = (y_1, ..., y_M)' \in \mathbb{R}^M_+$ denote inputs and outputs, respectively, and recall that the profit maximization problem of a firm can be stated as

$$\text{Maximal Profit} := \sup_{x,y}\{p(y)y - w(x)x \ : \ (x,y) \in \Psi\}, \tag{1}$$

i.e., such a firm will aim for an optimal allocation of (x, y) that will give the highest profit given the input prices, denoted with a row vector $w(x) \in \mathbb{R}^N_+$, the output prices, denoted with a row vector $p(y) \in \mathbb{R}^M_+$, and the technology available to it, characterized by a set Ψ, defined in generic terms as

$$\Psi = \{(x,y) \ : \ \ x \in \mathbb{R}^N_+ \text{ can produce } y \in \mathbb{R}^M_+\} \tag{2}$$

with standard regularity conditions or axioms of production imposed on this set.

Mainstream economic theory teaches us that, under certain assumptions, "perfect competition" even if hardly ever achievable in practice, provides a useful benchmark—a theoretically "efficient outcome" for the industry and, in particular, implies that the prices for all outputs and all inputs would be exogenous to an individual firm.[6] Therefore, (1) simplifies to the following profit function

$$\mathcal{PF}(p, w|\Psi) = \sup_{x,y}\{py - wx \ : \ (x, y) \in \Psi\}. \tag{3}$$

Recently, Färe et al. (2019) proposed a general measure of *profit efficiency*, which, for a particular allocation of interest (x_i, y_i), is defined as

$$\begin{aligned}\mathcal{GPE}(x_i, y_i; w, p|\Psi) = \sup_{\theta^{(x)},\theta^{(y)},x,y} \big\{\psi(\theta^{(x)}, \theta^{(y)}) : \psi_y(p, y_i, \theta^{(y)}) - \psi_x(w, x_i, \theta^{(x)}) \leq py - wx,\\ (\theta^{(y)}, \theta^{(x)}) \in \Theta, \ (x, y) \in \Psi\big\}\end{aligned} \tag{4}$$

where $\psi(\theta^{(x)}, \theta^{(y)})$ is an objective function, $\theta^{(y)} = (\theta_1^{(y)}, \ldots, \theta_M^{(y)})$ and $\theta^{(x)} = (\theta_1^{(x)}, \ldots, \theta_N^{(x)})$ are optimization variables (and Θ is the set specifying their possible values) along with (x, y), while $\psi_x(w, x, \theta^{(x)})$ and $\psi_y(p, y, \theta^{(y)})$ are the functions specifying the way (e.g., orientation) the measurement of the efficiency is to be done w.r.t. each element of input and output vectors, selected by a researcher based on the purpose of measurement.

Many efficiency measures in the literature can be derived as a special case of this measure. Following Färe et al. (2019) and Färe and Zelenyuk (2020), we will focus on the case when $\psi_y(p, y_i, \theta^{(y)}) = \sum_{m=1}^{M} p_m\theta_m^{(y)}y_{im}$, $\psi_x(w, x_i, \theta^{(x)}) = \sum_{l=1}^{N} w_l\theta_l^{(x)}x_{il}$, while $\theta_1^{(x)} = \theta_2^{(x)} = ... = \theta_N^{(x)} = 1$, $\theta_1^{(y)} = \theta_2^{(y)} = ... = \theta_M^{(y)} = \theta$ and $\psi(\theta^{(x)}, \theta^{(y)}) = \theta$ and assuming $py_i \neq 0$. Then, we have the *output-oriented profit efficiency measure*

$$\mathcal{OPE}(x_i, y_i; w, p|\Psi) = \sup_{\theta}\left\{\sup_{x,y}\left\{\frac{py - wx + wx_i}{py_i} \ : \ (x, y) \in \Psi\right\} \geq \theta\right\}, \tag{5}$$

[6] E.g., see Mas-Colell et al. (1995) for details.

which has various interesting decompositions that can be useful for instrumenting a "goodness-of-fit" measure.

In particular, for any allocation $(x_i, y_i) \in \Psi$ and any strictly positive price composition (w, p), and assuming $py^i \neq 0$, we have the following decomposition:

$$\mathcal{OPE}(x_i, y_i; w, p|\Psi) = \mathcal{OTE}(x_i, y_i|\Psi) \times \mathcal{OAE}(x_i, y_i, p|\Psi) \times \mathcal{RAE}(x_i, y_i, p, w|\Psi), \tag{6}$$

where the first component is the output oriented Debreu-Farrell measure of technical efficiency, the second component is the output oriented allocative efficiency measure and the third component is yet another allocative efficiency measure that gauges the gap between profit maximization and revenue maximization. Also, the product of the first two components gives the revenue efficiency, which we denote as $\mathcal{RE}(x_i, y_i, p|\Psi)$, while the product of the last two components is the total allocative efficiency that gauges the gap between the Debreu-Farrell-type technical efficiency and the output oriented profit efficiency, which we will denote as $\mathcal{AE}(x_i, y_i, p, w|\Psi)$.

To be more precise, $\mathcal{OTE} : \mathbb{R}^N_+ \times \mathbb{R}^M_+ \to \mathbb{R}_+ \cup \{+\infty\}$, is defined as

$$\mathcal{OTE}(x, y|\Psi) \equiv \sup_{\theta}\{\theta > 0 : (x, \theta y) \in \Psi\}, \tag{7}$$

which is the reciprocal of the output oriented Shephard distance function, and thus dual to the revenue function, defined as

$$\mathcal{RF}(x, p|\Psi) \;\; = \;\; \sup_{y}\{py \;\; : \;\; (x, y) \in \Psi\}, \tag{8}$$

Moreover, from Shephard's duality, we have, for any $p \in \mathbb{R}^M_{++}$:

$$\mathcal{RE}(x, y, p|\Psi) := \frac{\mathcal{RF}(x, p|\Psi)}{py} \geq \mathcal{OTE}(x, y|\Psi), \;\; \forall (x, y) \in \Psi \tag{9}$$

and the gap between the two efficiency measures in (9) gives rise to the output oriented

allocative efficiency measure, $\mathcal{OAE}(x, y, p|\Psi)$, so that for all $(x, y) \in \Psi$, we have

$$\mathcal{RE}(x, y, p|\Psi) = \mathcal{OTE}(x, y|\Psi) \times \mathcal{OAE}(x, y, p|\Psi). \tag{10}$$

Finally,

$$\mathcal{RAE}(x, y, p, w|\Psi) = \frac{\mathcal{OPE}(x, y; w, p|\Psi)}{\mathcal{RE}(x, y, p|\Psi)}. \tag{11}$$

Similarly, taking the input oriented analogue of (5), will give analogous decomposition, where the first component would be the input oriented Debreu-Farrell measure of technical efficiency, the second component would be the input oriented allocative efficiency measure and their product would give the cost efficiency—exactly the same as those considered in Färe and Grosskopf (1995) for their goodness-of-fit measures. Moreover, in this framework we also would have an additional, third, component that gives another allocative efficiency measure that gauges the gap between profit maximization and cost minimization, which can also be used as a goodness-of-fit measure.

Any of the components in (6) or their analogues (e.g., input/cost orientation) can be viewed as measures of goodness-of-fit, in the same spirit as the input/cost oriented Debreu-Farrell-type measures can be viewed so, according to Färe and Grosskopf (1995), just with different perspectives. Specifically, $\mathcal{OPE}$ can be viewed as a measure of overall goodness-of-fit with respect to profit maximization criterion, while $\mathcal{RE}$ can be viewed as a measure of goodness-of-fit with respect to revenue maximization criterion, and $\mathcal{OTE}$ can be used as a measure of goodness-of-fit with respect to the technology frontier only, radially in the output-space. Which of these measures are feasible depends on the data and on the prior beliefs about optimizing behavior, e.g., if the researcher has input-output data but does not have price data, then only $\mathcal{OTE}$ of these can be used. Such measures of goodness-of-fit can be used on an individual basis, for each firm or for each allocation, or in an aggregate sense, by looking at their averages or at the averages of their squares (in the spirit of Cramer-von-Mises criterion). Such averaging could be simple (i.e., equally weighted) or weighted

with some economically-relevant weights (e.g., as in Simar and Zelenyuk (2007), Simar and Zelenyuk (2018)).

Importantly, note that the measures of goodness-of-fit considered so far are based on technical efficiency of radial type measurement, either in the output space or in the input space. This is a simplification of the reality (which gives some advantages and some sacrifice), and if needed can be relaxed by employing the so-called *slack-based measures* of technical efficiency, defined as

$$\mathcal{SBTE}(x,y|\Psi) \equiv \sup_{u_1^{(x)},...,u_N^{(x)},u_1^{(y)},...,u_M^{(y)}} \Big\{ \psi(u_1^{(x)},...,u_N^{(x)},u_1^{(y)},...,u_M^{(y)}) : \\ (x_1 - u_1^{(x)},...,x_N - u_N^{(x)}, y_1 + u_1^{(y)},...,y_M + u_M^{(y)}) \in \Psi, \\ u_1^{(x)},...,u_N^{(x)},u_1^{(y)},...,u_M^{(y)} \geq 0 \Big\}, \quad (12)$$

where $\psi(u_1^{(x)},...,u_N^{(x)},u_1^{(y)},...,u_M^{(y)})$ is a suitable aggregator function of the individual slacks selected by a researcher (e.g., see Charnes et al. (1985), Tone (2001), Fukuyama and Weber (2009), Färe et al. (2015) for some examples). In the next section we will discuss the estimation matters for these measures.

4 Estimation Approaches

4.1 Deterministic Hulls

Consider the estimator of the production set characterized by

$$\widehat{\Psi} = \Big\{ (x,y) | x \geq X'\lambda,\ y \leq Y'\lambda,\ \lambda \in \widetilde{\Lambda} \Big\}, \quad (13)$$

where $X \in \mathbb{R}_+^{n\times N}$ and $Y \in \mathbb{R}_+^{n\times M}$ are matrices containing all the observations on inputs and outputs ($x \in \mathbb{R}_+^N$ and $y \in \mathbb{R}_+^M$), respectively, while $\lambda = (\lambda_1,...,\lambda_n)'$ is the vector of the so-called intensity variables, with $\widetilde{\Lambda}$ being the set of its permissible values.

Characterization (13) satisfies the standard regularity conditions or axioms of production theory and gives the tightest convex free disposal cone of the data (X, Y) if and only if $\widetilde{\Lambda} = \mathbb{R}^n_+$. Such technology exhibits constant returns to scale (CRS) and we will start with this convenient and popular in economics paradigm, e.g., as did Debreu (1951), Farrell (1957), Charnes et al. (1978), Färe et al. (1994), Kumar and Russell (2002), as well as Kitamura and Stoye (2018) and Smeulders et al. (2021) for consumer analysis, among many others. Further generalizations are possible by imposing various constraints onto Λ, e.g., requiring $\sum_{j=1}^n \lambda_j \leq 1$ or $\sum_{j=1}^n \lambda_j = 1$ will model non-increasing returns to scale (NIRS) or variables returns to scale (VRS), respectively, while requiring $\sum_{j=1}^n \lambda_j = 1$ and $\lambda_j \in \{0, 1\}$ will remove convexity, modeling the tightest free disposal hull technology. In operations research literature these approaches are known as Data Envelopment Analysis (DEA) estimators. Statistical properties of the estimators are now well-established due to rich statistical literature on the subject (see Korostelev et al. (1995), Kneip et al. (1998), Park et al. (2000), Kneip et al. (2008), Park et al. (2010), Kneip et al. (2015), and a brief review in Simar and Wilson (2015)).

It is also worth noting that (13) with $\widetilde{\Lambda} = \mathbb{R}^n_+$ is the so-called H-representation of a convex cone and an equivalent (due to Minkowski-Weyl theorem) characterization can be given via the so-called V-representation involving equalities rather than inequalities (e.g., see Kitamura and Stoye (2018) and Smeulders et al. (2021) for related discussion). While our approach can be also described via both representations, for our purposes the H-representation is more convenient for explicitly incorporating the one-sided inefficiency terms that force agents (firms, banks, hospitals, etc.), whose technology we model from the data, to be below the frontier of the technology characterized by such a polyhedron ((13) with $\widetilde{\Lambda} = \mathbb{R}^n_+$).

The estimator of the output oriented Debreu-Farrell measure of technical efficiency associated with (13), for an allocation (x_i, y_i), takes the following form

$$\widehat{\mathcal{OTE}}(x_i, y_i|\widehat{\Psi}) \equiv \max_{\theta,\lambda_1,\ldots,\lambda_n} \{\theta : \sum_{j=1}^n \lambda_j y_j \geq \theta y_i,\ \sum_{j=1}^n \lambda_j x_j \leq x_i, \theta \geq 0, \lambda \in \widetilde{\Lambda}\}, \tag{14}$$

while the corresponding estimator of revenue function is given by

$$\widehat{\mathcal{RF}}(x_i, p|\widehat{\Psi}) \equiv \max_{y,\lambda_1,...,\lambda_n} \{py : \sum_{j=1}^{n} \lambda_j y_j \geq y, \sum_{j=1}^{n} \lambda_j x_j \leq x_i, \lambda \in \widetilde{\Lambda}\} \tag{15}$$

and, provided that $py_i \neq 0$, yielding the corresponding estimator of revenue efficiency

$$\widehat{\mathcal{RE}}(x_i, y_i, p|\widehat{\Psi}) = \widehat{\mathcal{RF}}(x_i, p|\widehat{\Psi})/py_i. \tag{16}$$

Meanwhile, the corresponding estimator of the profit function is given by[7]

$$\widehat{\mathcal{PF}}(p, w|\widehat{\Psi}) \equiv \max_{x,y,\lambda_1,...,\lambda_n} \{py - wx : \sum_{j=1}^{n} \lambda_j y_j \geq y, \sum_{j=1}^{n} \lambda_j x_j \leq x, \lambda \in \widetilde{\Lambda}\} \tag{17}$$

and thus yielding the corresponding estimator of the output oriented profit efficiency

$$\widehat{\mathcal{OPE}}(x_i, y_i; w, p|\widehat{\Psi}) = \frac{\widehat{\mathcal{PF}}(p, w|\widehat{\Psi})}{py_i} + \frac{wx_i}{py_i}, \tag{18}$$

i.e., it is the sum of the estimated maximal profit margin and the observed cost-revenue ratio.

Furthermore, the ratio of (16) and (14) gives the estimate of output oriented allocative efficiency $\mathcal{OAE}(x, y, p|\Psi)$, while the ratio of (18) and (16) gives the estimate of $\mathcal{RAE}(x, y, p, w|\Psi)$. When price data is not available, then researchers typically choose to focus on the Debreu-Farrell measure of technical efficiency (14).

Finally, note that these optimization problems are specific for an observation $i \in \{1, ..., n\}$ and so for a sample of size n will involve the estimation of n of such problems for each type of inefficiency, thus implying i-specific intensity variables, which we will denote as $\lambda_i = (\lambda_{i1}, ..., \lambda_{ij}, ..., \lambda_{in})'$.

[7]Here it is worth noting that this optimization problem can be unbounded or yield zero when technology is a cone (or any non-decreasing returns to scale) and so additional constraints might be needed in such circumstances (e.g., placing limits on some inputs or outputs, or constraints on λ of the type as for NIRS or VRS, etc.).

4.2 Stochastic Hulls

As mentioned in the introduction, the main caveat of the approach described in the previous section is not accounting for statistical noise and, in particular, possibly high sensitivity to outliers. So our first goal is to add stochasticity into this framework. Before proceeding, it is worth noting again that many interesting studies proposed different ways to do that. This includes the work on adapting convex non-parametric least squares (Kuosmanen (2008), Kuosmanen and Johnson (2010), Kuosmanen and Johnson (2017)), order-m and order-α quantile frontier approaches (Cazals et al. (2002), Daouia and Simar (2007) and Bădin et al. (2012)) and stochastic DEA (Simar and Zelenyuk (2011) and Olesen and Petersen (2016)). Here we take a different and (to our knowledge) novel approach, using a Bayesian paradigm. To do so, first note that, conditional on the data, (X, Y), if $\widetilde{\Lambda} = \mathbb{R}^n_+$, then (13) is a *deterministic* convex free disposal cone, where a for particular observation i we have

$$x_i \geq X'\lambda_i,\ y_i \leq Y'\lambda_i,\ \lambda_i \geq \mathbf{0},\ i = 1, ..., n, \tag{19}$$

where note that $\lambda_i = (\lambda_{i1}, ..., \lambda_{ij}, ..., \lambda_{in})'$.

Now, inspired by the SFA literature and the slack-based measurement of efficiency described above, we can express these inequalities in (19) as in the following model:

$$x_i = X'\lambda_i + v_i^{(x)} + u_i^{(x)},\ y_i = Y'\lambda_i + v_i^{(y)} - u_i^{(y)},\ \lambda_i \geq \mathbf{0},\ i = 1, ..., n, \tag{20}$$

where $v_i^{(x)} \in \mathbb{R}^N, v_i^{(y)} \in \mathbb{R}^M$ represent two-sided symmetric statistical errors, $u_i^{(x)} \in \mathbb{R}^N_+, u_i^{(y)} \in \mathbb{R}^M_+$ represent one-sided asymmetric random errors or inefficiencies (slacks in inputs and outputs respectively). Importantly, note that unlike in a typical SFA approach (except for Kumbhakar and Tsionas (2021)), both the noise and the inefficiency are multivariate vectors, with elements corresponding to each input and each output in the data.

Therefore, we have

$$\begin{bmatrix} x_i \\ -y_i \end{bmatrix} = \begin{bmatrix} X' \\ -Y' \end{bmatrix} \lambda_i + \begin{bmatrix} v_i^{(x)} \\ v_i^{(y)} \end{bmatrix} + \begin{bmatrix} u_i^{(x)} \\ u_i^{(y)} \end{bmatrix} \Leftrightarrow z_i = Z'\lambda_i + v_i + u_i,\ \lambda_i \geq \mathbf{0},\ i = 1, ..., n, \quad (21)$$

in obvious notation. In particular, note that the inefficiency term is a vector of dimension $d = N + M$:

$$u_i = \begin{bmatrix} u_i^{(x)} \\ u_i^{(y)} \end{bmatrix},\ i = 1, ..., n. \quad (22)$$

The model (21) is a system of simultaneous equations model, with $N + M$ equations, which gives a stochastic representation of the deterministic (conditional on (X, Y)) convex free disposal cone. That is, it is a stochastic version of (13) when $\widetilde{\Lambda} = \mathbb{R}^n_+$, where the stochasticity enters in a very general manner, through individual inefficiencies (slacks) and statistical noise, corresponding to each element of input and each element of output.

The Jacobian of transformation for (21) can be shown to be

$$\mathscr{J}(\lambda) = ||I_{n\times n} - \Lambda||, \quad (23)$$

where $\Lambda = (\lambda'_1, \lambda'_2, ..., \lambda'_n)$, where λ_{ij} is the j^{th} element of λ_i.[8]

Note that the only restriction on these variables so far is that they are non-negative, i.e., $\lambda_i \geq \mathbf{0}$, $i = 1, ..., n$, to satisfy the requirement of the conical structure or the CRS technology. Meanwhile, restricting $\sum_{j=1}^n \lambda_{ij} \leq 1$ will model non-increasing returns to scale, while strengthening it to $\sum_{j=1}^n \lambda_{ij} = 1$ will model the variable returns to scale for the observation of interest i.

Since, in general, $z_i \in \mathbb{R}^{N+M}$ and $\lambda_i \in \mathbb{R}^n$ it is clear that the number of parameters exceeds by far the number of observations so we have the "small n, large p" paradigm to deal with, which is typically a challenging problem and, in certain cases, are examples of the

[8]This is because we have $\underset{(n\times d)}{Z} = \underset{(n\times n)}{\Lambda}\underset{(n\times d)}{Z} + \underset{(n\times d)}{V} + \underset{(n\times d)}{U}$, where $d = N + M$.

so-called NP-hard problems. While there are different ways to approach the "small n, large p" paradigm, a Bayesian approach offers a way with an appealing simplicity, which we will follow here. Specifically, we will deploy the Bayesian compression to address this problem, by writing

$$\lambda_i = \Phi_i \gamma,\ i = 1, ..., n, \tag{24}$$

where $\Phi_i \in \mathbb{R}^{n \times s_i}$, $(i = 1, ..., n)$, $\gamma \in \mathbb{R}^{s_i \times 1}$ and s_i is the dimensionality that we must choose optimally along with the elements of Φ.

In turn, we have

$$z_i = Z' \Phi_i \gamma + v_i + u_i,\ i = 1, ..., n. \tag{25}$$

To proceed with the estimation, we adapt the recent developments on Bayesian compressed regression by Guhaniyogi and Dunson (2015), who suggested drawing the elements of matrix $\Phi^{(i)} = \left[\Phi^{(i)}_{i'j'},\ \ i', j' \in \{1, ..., n\}\right]$ randomly as follows:

$$\begin{aligned} \Pr(\Phi^{(i)}_{i'j'} = \tfrac{1}{\sqrt{\phi_i}}) &= \phi_i^2, \\ \Pr(\Phi^{(i)}_{i'j'} = 0) &= 2\phi(1 - \phi_i), \\ \Pr(\Phi^{(i)}_{i'j'} = -\tfrac{1}{\sqrt{\phi_i}}) &= (1 - \phi_i)^2, \end{aligned} \tag{26}$$

where ϕ_i and s_i $(i = 1, ..., n)$ are parameters.[9] Matrices Φ_i are orthonormalized before proceeding. Parameters s_i can be drawn as integers from a uniform distribution and ϕ_i from a uniform distribution in $(0.1, 0.9)$. This is because, (i) setting ϕ_i to 0 or 1 results in zero / one probabilities in (26), and (ii) ϕ_i should necessarily be between 0 and 1 but given (i) it is good to avoid values close to 0 or 1, e.g., can be drawn from a uniform distribution in $(0.1, 0.9)$ as we do here.

Besides orthonormalizing, various restrictions can be imposed onto (24). For example, to model non-increasing returns to scale, we can keep only the draws that satisfy $\sum_{j=1}^{n} \lambda_{ij} \leq 1$ for each i; a form of acceptance-rejection sampling. Alternatively, to model the variable

[9] Note that we allow for different dimensions s_i.

returns to scale, we can impose $\sum_{j=1}^{n} \lambda_{ij} = 1$ for each i by re-normalizing the draws so that they sum up to unity.

Suppose $v_i \sim \mathcal{N}_d(\mathbf{0}, \Sigma)$ and let the u_is be independently distributed with density $p(u_i; \delta)$ supported in $\mathbb{R}_+^d$, where δ is a finite-dimensional parameter vector. Suppose $\mathbb{Z}$ denotes all available data. The joint posterior of the parameters and the latent variables is given by Bayes' theorem:

$$\begin{aligned} p(\gamma, \delta, \{u_i\}_{i=1}^n | \mathbb{Z}) \propto |\Sigma|^{-n/2} \exp\left\{-\tfrac{1}{2} \sum_{i=1}^{n} (z_i - Z'\Phi_i\gamma - u_i) \Sigma^{-1} (z_i - Z'\Phi_i\gamma - u_i)\right\} \cdot \\ \{\prod_{i=1}^{n} p(u_i; \omega)\} \cdot \mathscr{J}(\Lambda) \cdot p(\gamma, \delta), \end{aligned} \tag{27}$$

for given values of $\{\phi_i\}_{i=1}^n, s$, where $p(\gamma, \delta)$ is a prior. Apparently, the u_is will have to be integrated out of this posterior, a task that is analytically impossible. For $p(u_i; \delta)$ we can choose a truncated normal distribution:

$$u_i \sim \mathcal{N}_d^+(\mu, \Omega), \ i = 1, ..., n, \tag{28}$$

where $\mathcal{N}_d^+(\mu, \Omega)$ denotes the truncated normal distribution in $\mathbb{R}^d$ with location vector μ and scale matrix Ω. A deterministic approximation results if we set $\mu = \mathbf{0}$ and let the elements of Ω converge to zero. In the sequel, we denote the density corresponding to (28) by $p(u_i; \omega)$, where ω is a finite dimensional vector that contains all unknown parameters.

Flexible approximations can be introduced by making the μs flexible function of environmental variables, in case they are available. Moreover, we do so in the interest of incorporating *environmental variables* $w_i \in \mathbb{R}^{d_w}$ into the model in the next section.

Our prior for γ and δ is

$$p(\gamma, \delta) \sim \mathcal{N}(0, q^2 I), \tag{29}$$

where q is a parameter that we set (in the benchmark case) to $q = 10$ so that the prior is proper but diffuse.

4.3 Efficiency-based measures of Goodness-of-fit

Once the estimates of the efficiency scores, $\widehat{\mathcal{OTE}}(x^j, y^j|\widehat{\Psi})$, $\widehat{\mathcal{RE}}(x^j, y^j, p|\widehat{\Psi})$, $\widehat{\mathcal{OPE}}(x^j, y^j; w, p|\widehat{\Psi})$, are obtained, they can be used as the measures of goodness-of-fit as they are or as their reciprocals (to convert them on the scale $[0, 1]$).

For the more general modeling approach that models inefficiency via the individual slacks, their estimates need to be converted into the same units of measurement and then aggregated into a scalar measure. While there are different ways to do so, here we will consider and adapt two ideas from OR literature. The first one is due to Tone (2001), who proposed using the following aggregating function

$$TGoF = \left(1 - \frac{1}{N}\sum_{l=1}^{N}\frac{u_l^{(x)}}{x_l}\right) / \left(1 + \frac{1}{M}\sum_{m=1}^{M}\frac{u_m^{(y)}}{y_m}\right) \tag{30}$$

provided that $x_l > 0, \forall l$ and $y_m > 0, \forall m$. Note that this TGoF is measured on the scale $[0, 1]$, as one would typically desire for a goodness-of-fit measure.

The second idea is due to Fukuyama and Weber (2009), who proposed the so-called directional slack-based (in)efficiency aggregating function

$$FWGoF = \frac{1}{2}\left(\frac{1}{N}\sum_{l=1}^{N}\frac{u_l^{(x)}}{g_{x_l}} + \frac{1}{M}\sum_{m=1}^{M}\frac{u_m^{(y)}}{g_{y_m}}\right) \tag{31}$$

where $g_x = (g_{x_1}, ..., g_{x_N}) > 0_N$ and $g_y = (g_{y_1}, ..., g_{y_M}) > 0_N$ and so $(-g_x, g_y)$ is a direction in (x, y)-space chosen by the researcher depending on the aims (or preferences) of the measurement of (in)efficiency (e.g., a natural choice would be $g_x = x$ and $g_y = y$, provided that $x > 0$ and $y > 0$). Note that this FWGoF is measured on the scale $[0, \infty]$ and so may need to be converted to the scale $[0, 1]$, to be commensurate with other goodness-of-fit measures, e.g., which can be done as

$$\widetilde{FWGoF} = (1 + FWGoF)^{-1}. \tag{32}$$

Once these measures are obtained for each firm i, they can be averaged over all firms

within the sample (or a sub-sample of interest) to obtain a measure of goodness-of-fit for that sample.

4.4 Statistical Measures of Goodness-of-fit

The goodness-of-fit measures discussed above are based on the economic theory rationale, as conceptualized by Varian (1990) and elaborated on further by Färe and Grosskopf (1995).

Since we are also operating with statistical tools, it is also natural to consider statistical measures of goodness-of-fit. A natural candidate here would be the R-squared for systems (Carter and Nagar (1977)). Specifically, suppose we have a system with $\mathfrak{p}$ equations,

$$Y = YB + X\Gamma + U,$$

where Y is the $n \times \mathfrak{p}$ matrix of observations on the dependent variable, X is the $n \times \mathfrak{k}$ matrix of observations on the $\mathfrak{k}$ exogenous variables, B and Γ are matrices containing unknown parameters. The reduced form is

$$y = (Z \otimes I)\,\vartheta + v$$

where $y = [Y_1', \ldots, Y_{\mathfrak{p}}']'$, and Z is the matrix of observations on the exogenous variables. Carter and Nagar (1977) proposed the measure

$$R^2 = \frac{\hat{\vartheta}' Z_*' Z_* \hat{\vartheta}}{\hat{\vartheta}' Z_*' Z_* \hat{\vartheta} + \hat{v_*}' \hat{v_*}},$$

where $\hat{v}$ denotes residuals, $\hat{\vartheta}$ denotes any estimate of the system's structural parameters, and

$$Z_*' Z_* = \hat{Z}' \left(\hat{\Omega}^{-1} \otimes I \right) \hat{Z} \quad \text{and} \quad \hat{v}_*' \hat{v}_* = \hat{v}' \left(\hat{\Omega}^{-1} \otimes I) \right) \hat{v}$$

where

$$\hat{Z} = \begin{bmatrix} [X\hat{\Pi}_1, X_1] & & \\ & \ddots & \\ & & [X\hat{\Pi}_{\mathfrak{p}}, X_{\mathfrak{p}}] \end{bmatrix},$$

where $\hat{\Pi}_1, \ldots, \hat{\Pi}_{\mathfrak{p}}$ are estimates of the reduced form parameters for each equation, and $X = [X_1, \ldots, X_{\mathfrak{p}}]$.

Moreover, based on BIC we can obtain an approximate marginal likelihood as $ML \simeq \exp\left(-\frac{1}{2}BIC\right)$. Specifically, suppose the parameter vector is $\theta = [\gamma', \delta']' \in \mathbb{R}^{\mathfrak{d}}$, then the BIC in our context is defined as

$$BIC = -2\log L(\theta) + \mathfrak{d}\log(n), \tag{33}$$

where $\mathfrak{d}$ is the total number of parameters and n is the sample size. An alternative is, of course, to use the log marginal likelihood (LML) derived from a Laplace approximation to the log posterior as

$$\log ML \simeq \log L(\bar{\theta}; Y) + \log p(\bar{\theta}) + \tfrac{n}{2}\log|S|, \tag{34}$$

where $\bar{\theta}$ is the posterior mean of θ and S denotes the posterior covariance of θ, quantities that can be easily computed from MCMC draws. As a matter of fact we can compute the value of BIC for each MCMC draw instead of computing it at a specific value of θ (e.g., the posterior mean). From this we obtain an approximation to the posterior of $-\frac{1}{2}$BIC, viz. to the marginal likelihood (ML). In our application we will present some approximations to log marginal likelihood and report Bayes factors for model selection (see Table 5).

5 Modeling Determinants of Inefficiency via Bayesian ANN

In the current literature, the inefficiency is usually formulated as a scalar measure in general and for modeling determinants of such inefficiency in particular. It is therefore natural to

start our description with the scalar case too. While there are many measures of efficiency, the most popular appears to be the output oriented technical efficiency measure, which can be formulated as $r_i = e^{-U_i}$. Since $r_i \in (0, 1]$ it also seems natural to assume $r_i = F(w_i'\delta)$ where $\delta \in \mathbb{R}^{d_w}$ is a vector of parameters, $w_i \in \mathbb{R}^{d_w}$ is a vector of determinants, and F is any suitable function that will ensure $r_i \in (0, 1]$, like a distribution function, for example the normal or logistic.[10] To make this approach flexible we extend to the case

$$r_i = \sum_{g=1}^{G} p_g F(w_i'\delta_g),\ i = 1, ..., n, \tag{35}$$

where $\delta_g \in \mathbb{R}^{d_w}$ is a vector of parameters $(1 \leq g \leq G)$, G is the number of terms that we consider to arrive at a flexible approximation, and p_gs are probabilities. Flexibility arises from the well-known global approximation properties of artificial neural networks (e.g. Hornik et al. (1989) and White (1989, 1990)).

In turn, we have

$$U_i = -\ln \sum_{g=1}^{G} p_g F(w_i'\delta_g) + \epsilon_i,\ i = 1, ..., n, \tag{36}$$

where ϵ_i is an error term which is used to introduce stochasticity into this approach, $p_g \geq 0$ $(g = 1, ..., G)$, $\sum_{g=1}^{G} p_g = 1$; the p_gs are mixing probabilities introduced to make (35), and (36) tighter approximations to the inefficiency process.[11] We will examine two cases, viz. with and without the presence of ϵ_is in (36).

[10] E.g., see Parmeter et al. (2017) and Paul and Shankar (2018) as well as Tsionas and Mamatzakis (2019) for related discussions with alternative estimators.

[11] The major difference between U_is and $u_i^{(x)}, u_i^{(y)}$ (in (20) or (21)) is that the former are in logarithm of percentage terms due to (35), while the latter are in the same units as the corresponding inputs and outputs and so they need to be converted to percentage terms by dividing with the corresponding variables.

The extension to our case with N inputs and M outputs is as follows.

$$
\begin{gathered}
\text{Input slacks:} \\
U_{i1}^{(x)} = -\ln \textstyle\sum_{g=1}^{G} p_{g1}^{(x)} F(w_{i1}^{(x)\prime} \delta_{g1}^{(x)} + \epsilon_{i1}),\ i = 1, ..., n \\
U_{i2}^{(x)} = -\ln \textstyle\sum_{g=1}^{G} p_{g2}^{(x)} F(w_{i2}^{(x)\prime} \delta_{g2}^{(x)} + \epsilon_{i2}),\ i = 1, ..., n \\
\vdots \\
U_{iN}^{(x)} = -\ln \textstyle\sum_{g=1}^{G} p_{gN}^{(x)} F(w_{iN}^{(x)\prime} \delta_{gN}^{(x)} + \epsilon_{iN}^{(x)}),\ i = 1, ..., n \\
\text{Output slacks:} \\
U_{i1}^{(y)} = -\ln \textstyle\sum_{g=1}^{G} p_{g1}^{(y)} F(w_{i1}^{(y)\prime} \delta_{g1}^{(y)} + \epsilon_{i,N+1}^{(y)}),\ i = 1, ..., n \\
U_{i2}^{(y)} = -\ln \textstyle\sum_{g=1}^{G} p_{g2}^{(y)} F(w_{i2}^{(y)\prime} \delta_{g2}^{(y)} + \epsilon_{i,N+2}^{(y)}),\ i = 1, ..., n \\
\vdots \\
U_{iM}^{(y)} = -\ln \textstyle\sum_{g=1}^{G} p_{gM}^{(y)} F(w_{iM}^{(y)\prime} \delta_{gM}^{(y)} + \epsilon_{i,N+M}^{(y)}),\ i = 1, ..., n
\end{gathered}
\tag{37}
$$

where $\delta_{gj}^{(x)} \in \mathbb{R}^{d_w}$ $(j = 1, ..., N)$, $\delta_{gj}^{(y)} \in \mathbb{R}^{d_w}$ $(j = 1, ..., M)$ are parameter vectors and, similarly, we have the probability parameters $p_{gj}^{(x)} \in \mathbb{R}^{d_w}$ $(j = 1, ..., N)$, $p_{gj}^{(y)} \in \mathbb{R}^{d_w}$ $(j = 1, ..., M)$.[12]

For the random errors in (37) we assume

$$
\epsilon_i = \left[\epsilon_{i1}, ..., \epsilon_{iN}, \epsilon_{i,N+1}, ..., \epsilon_{i,N+M}\right]' \sim \mathcal{N}_{N+M}\left(\mathbf{0}, \Sigma_\epsilon\right), \tag{38}
$$

where Σ_ϵ is a diagonal matrix, $\Sigma_\epsilon = \begin{bmatrix} \sigma_{\epsilon 1}^2 & & & & & & \\ & \ddots & & & & & \\ & & \sigma_{\epsilon N}^2 & & & & \\ & & & \sigma_{\epsilon,N+1}^2 & & & \\ & & & & \ddots & & \\ & & & & & \sigma_{\epsilon,N+M}^2 \end{bmatrix}$.

[12] Including the errors in this way is novel relative to Paul and Shankar (2018) and Tsionas and Mamatzakis (2019). The errors cannot be introduced additively into these equations as they cannot be discriminated from v_is. They can be interpreted as DMU-specific random effects associated with each input and each output.

To facilitate posterior inference, we amend (37) by introducing small artificial errors:

$$
\begin{gathered}
\text{Input slacks:} \\
U_{i1}^{(x)} = -\ln \textstyle\sum_{g=1}^{G} p_{g1}^{(x)} F(w_{i1}^{(x)\prime}\delta_{g1}^{(x)} + \epsilon_{i1}) + \xi_{i1},\ i = 1, ..., n, \\
U_{i2}^{(x)} = -\ln \textstyle\sum_{g=1}^{G} p_{g2}^{(x)} F(w_{i2}^{(x)\prime}\delta_{g2}^{(x)} + \epsilon_{i2}) + \xi_{i2},\ i = 1, ..., n, \\
\vdots \\
U_{iN}^{(x)} = -\ln \textstyle\sum_{g=1}^{G} p_{gN}^{(x)} F(w_{iN}^{(x)\prime}\delta_{gN}^{(x)} + \epsilon_{iN}^{(x)}) + \xi_{iN},\ i = 1, ..., n, \\
\text{Output slacks:} \\
U_{i1}^{(y)} = -\ln \textstyle\sum_{g=1}^{G} p_{g1}^{(y)} F(w_{i1}^{(y)\prime}\delta_{g1}^{(y)} + \epsilon_{i,N+1}^{(y)}) + \xi_{i,N+1},\ i = 1, ..., n, \\
U_{i2}^{(y)} = -\ln \textstyle\sum_{g=1}^{G} p_{g2}^{(y)} F(w_{i2}^{(y)\prime}\delta_{g2}^{(y)} + \epsilon_{i,N+2}^{(y)}) + \xi_{i,N+2},\ i = 1, ..., n, \\
\vdots \\
U_{iM}^{(y)} = -\ln \textstyle\sum_{g=1}^{G} p_{gM}^{(y)} F(w_{iM}^{(y)\prime}\delta_{gM}^{(y)} + \epsilon_{i,N+M}^{(y)}) + \xi_{i,N+M},\ i = 1, ..., n,
\end{gathered}
\tag{39}
$$

where

$$
\xi_{ij} \sim \mathcal{N}\left(0, \bar{\sigma}_o^2\right),\ i = 1, ..., n,\ j = 1, ..., N + M. \tag{40}
$$

We set, on purpose, $\sigma_o = 10^{-4}$ so that these errors are indeed artificial, i.e. "sufficiently small".[13] Therefore, our new posterior becomes

$$
\begin{aligned}
p(\gamma, \delta, \{U_i\}|\mathbb{Z}) \propto\ & |\Sigma|^{-n/2} \exp\left\{-\tfrac{1}{2}\textstyle\sum_{i=1}^{n}\left(z_i - Z'\Phi_i\gamma - U_i\right)\Sigma^{-1}\left(z_i - Z'\Phi_i\gamma - U_i\right)\right\} \cdot \\
& |\Sigma_\epsilon|^{-n/2} \exp\left\{-\tfrac{1}{2}\textstyle\sum_{j=1}^{N}\frac{1}{\sigma_{\epsilon j}^2}\sum_{i=1}^{n}\left(U_{ij}^{(x)} + \ln\sum_{g=1}^{G} p_{gj}^{(x)} F(w_{ij}^{(x)\prime}\delta_{gj}^{(x)} + \epsilon_{ij})\right)^2\right\} \cdot \\
& \exp\left\{-\tfrac{1}{2}\textstyle\sum_{j=1}^{M}\frac{1}{\sigma_{\epsilon,N+j}^2}\sum_{i=1}^{n}\left(U_{i,j}^{(y)} + \ln\sum_{g=1}^{G} p_{g,j}^{(y)} F(w_{ij}^{(y)\prime}\delta_{g,j}^{(y)} + \epsilon_{i,N+j})\right)^2\right\} \cdot \\
& \bar{\sigma}_o^{-(N+M)n/2} \exp\left\{-\tfrac{1}{2\bar{\sigma}_o^2}\textstyle\sum_{i=1}^{n}\sum_{j=1}^{N+M}\epsilon_{i,j}^2\right\} \cdot \\
& \mathscr{J}(\Lambda) \cdot p(\gamma, \delta) \cdot \textstyle\prod_{j=1}^{N+M} p(\sigma_{\epsilon j}),
\end{aligned}
\tag{41}
$$

[13] The technical reason for introducing the artificial errors is because we avoid formal application of the change of variables theorem from εs to Us which, due to the nonlinearity of the transformation, and the imposition of global monotonicity, would increase the heavily computational burden. Another reason to avoid global monotonicity is that we are not sure that the effect of environmental variables on inefficiencies is, indeed, monotonic. Moreover, given the artificial errors, the ϵ_is become standard random effects and can be drawn during MCMC in standard ways.

given the definitions in (37), and for a fixed value of $\bar{\sigma}_o$ as in (39). Note that, because (37) are stochastic, we have to integrate the high-dimensional inefficiency variables out of the posterior. The prior remains the same as in (29), and additionally we assume the proper but diffuse priors

$$p(\sigma_{\epsilon j}) \propto \sigma_{\epsilon j}^{-(\underline{n}+1)} e^{-\underline{q}_j/(2\sigma_{\varepsilon j}^2)},\ j = 1, ..., N+M, \tag{42}$$

where $\underline{n}$ and $\underline{q}_j$ are prior parameters (Zellner, 1971, p. 371, equation A.37b). The priors are proper but diffuse provided $\underline{n} = 1$ and, say, $\underline{q}_j = 10^{-4}$. For $\underline{n} = 0$ the priors are improper.

To provide numerical access to the posteriors in (27) or (41) we use state-of-the-art Markov Chain Monte Carlo (MCMC) methods known as fast Metropolis Adjusted Langevin Algorithm (fMALA), see Durmus et al. (2017); the methods are described in the Technical Appendix. Convergence and numerical performance of MCMC can be tested using the standard diagnostics of Geweke (1992). After the system has been estimated, statistical inferences about inefficiencies can be made using (37) by taking, for example, averages across MCMC draws. Monte Carlo evidence for this approach can be found in Supplement 1 while the further technical details can be found in Supplement 2.

6 Empirical Illustration

To apply the new techniques we use the data of Malikov et al. (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 outputs $(y_1, ..., y_5)$, five inputs $(x_1, ..., x_5)$, and a quasi–fixed input (equity capital). All nominal stock variables are deflated to 2005 U.S. dollars using the consumer price index for all urban consumers. The list of included variables is 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 use MCMC with 150,000 iterations discarding the first 50,000 to mitigate possible start up effects. We monitor convergence and numerical performance using Geweke's (1992) convergence diagnostics. All results for the proposed approach are computed under the assumption of stochastic inefficiency and compared to a few alternatives from the literature. We also impose $\sum_{j=1}^{n} \lambda_{ij} = 1$ for each i.

In Tables 1, 2 and 3, we present the results for the approach proposed in this paper, for DEA-CRS, DEA-NIRS, DEA-VRS and from the approach of Kumbhakar and Tsionas (2021).[14] The latter approach is perhaps the closest (albeit still very different) in nature to our approach from the SFA paradigm to the approach proposed in this paper, in the sense that it also allows for both output-specific and input-specific inefficiency terms (slacks) and stochastic noise terms, as well as the environmental variables. While having similar objectives, the latter approach is different in terms of assumptions on the model. Specifically, recall that Tsionas and Kumbhakar (2021, henceforth TK) presented a model that accommodates both output-specific and input-specific inefficiency components (input and output slacks). They used a translog function to represent the technology in which the input slacks are generalized to have both deterministic (functions of exogenous variables) and stochastic components. Here, we apply their MCMC-based techniques to a more general model when the slacks depend on the same variables as we described above.[15]

Table 1 summarizes the estimated slacks and their standard errors for each input and each output. Table 2 presents the estimated marginal effects of the environmental variables onto slacks and their standard errors, again for each input and each output, while the corresponding Z-statistics that help evaluating statistical significance (from zero) of these marginal effects are presented in Table 3. From these tables, one can see that different environmental variables often may have a very different association with different slacks, not just quanti-

[14] Also see Supplement 3 for the estimation results for the densities of efficiency, slacks and marginal effects.
[15] Our MCMC configuration is the same as in TK.

Table 1: Estimated slacks and their standard deviations

Notes: We report posterior means with posterior standard deviations (using MCMC draws in the post-burn in period with a White covariance matrix with ten lags). For purposes of comparisons, slacks in (37) are expressed as percentages relative to the respective variables.

	this paper		DEA-CRS		DEA-NIRS		DEA-VRS		TK (2021)	
	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
slack x_1	0.257	0.016	0.050	0.020	0.158	0.006	0.090	0.003	0.099	0.007
slack x_2	0.307	0.009	0.105	0.013	0.027	0.027	0.125	0.001	0.103	0.008
slack x_3	0.176	0.011	0.058	0.016	0.063	0.001	0.038	0.006	0.084	0.031
slack x_4	0.197	0.023	0.011	0.008	0.049	0.008	0.068	0.010	0.147	0.010
slack x_5	0.116	0.021	0.071	0.010	0.004	0.012	0.111	0.023	0.083	0.002
slack y_1	0.023	0.003	0.075	0.015	0.110	0.025	0.008	0.004	0.060	0.001
slack y_2	0.086	0.013	0.251	0.022	0.136	0.008	0.105	0.009	0.115	0.009
slack y_3	0.064	0.009	0.102	0.008	0.011	0.003	0.123	0.016	0.104	0.008
slack y_4	0.040	0.004	0.146	0.021	0.130	0.024	0.038	0.007	0.039	0.020
slack y_5	0.102	0.021	0.265	0.003	0.050	0.005	0.136	0.010	0.051	0.004

Table 2: Estimated marginal effects of determinants onto slacks, and their standard errors

Notes: w_j represents the jth relative price, that is $w_j = \log(W_j/W_1)$.

	equity	equity/TA	TA	trend	trend^2	w_1	w_2	w_3	w_4
					coefficients				
slack x_1	0.096	-0.171	0.177	0.062	0.158	-0.111	0.099	0.102	-0.209
slack x_2	-0.123	-0.102	-0.154	-0.145	-0.061	-0.156	-0.027	-0.067	-0.020
slack x_3	0.161	-0.173	-0.058	0.051	-0.015	-0.051	-0.015	-0.029	-0.078
slack x_4	0.089	0.106	0.124	0.003	-0.036	-0.254	-0.003	0.056	0.166
slack x_5	0.194	-0.039	-0.165	-0.006	-0.039	0.081	0.091	0.017	-0.057
slack y_1	-0.019	0.052	-0.165	0.169	0.014	0.132	-0.001	0.043	0.098
slack y_2	0.038	-0.078	0.030	0.096	0.063	-0.084	-0.077	-0.015	0.035
slack y_3	0.118	-0.111	0.045	-0.040	-0.069	0.010	0.071	-0.155	-0.109
slack y_4	0.038	-0.171	-0.209	-0.078	0.060	0.013	-0.026	0.086	-0.146
slack y_5	0.056	-0.027	0.078	0.066	0.017	-0.083	-0.027	-0.086	0.124
					standard errors				
slack x_1	-0.013	-0.038	-0.116	0.067	0.032	-0.042	-0.041	0.031	0.044
slack x_2	-0.132	0.024	-0.043	0.014	0.079	-0.031	-0.025	0.064	-0.033
slack x_3	-0.047	-0.049	0.010	-0.029	0.026	-0.039	0.001	-0.082	-0.084
slack x_4	0.007	-0.033	0.038	-0.045	-0.007	-0.048	-0.041	0.117	-0.035
slack x_5	0.008	-0.112	-0.026	0.076	0.049	-0.090	-0.098	0.015	0.045
slack y_1	0.033	0.025	0.023	0.015	-0.056	0.026	0.135	0.078	-0.017
slack y_2	0.104	-0.039	0.142	-0.004	0.021	-0.070	-0.026	0.003	0.014
slack y_3	-0.044	-0.038	0.014	-0.038	0.044	0.016	-0.059	-0.093	0.095
slack y_4	-0.068	-0.044	-0.015	-0.064	-0.109	-0.023	0.035	0.064	-0.059
slack y_5	0.073	-0.029	-0.147	0.006	-0.024	0.002	0.106	-0.078	-0.025

tatively but also qualitatively: they can be very different in terms of magnitude, sign and statistical (in)significance.

For example, one can see that for this data and model specification, the equity variable has a significant negative association with the slacks of x_1 (Labor), x_3 (Purchased Funds) and y_3 (Commercial & Industrial Loans) but also has significant positive association with the slacks of x_4 (Interest-Bearing Transaction Accounts) and x_5 (Non-Transaction Accounts). Meanwhile, we also see statistically insignificant (relative to zero) association with x_2 (Physical Capital), y_1 (Consumer Loans), y_2 (Real Estate Loans), y_4 (Securities) and y_5 (Off-Balance Sheet Activities Income). Analogous information can be obtained for all the other environmental variables. It is worth noting that all other standard approaches in DEA and SFA typically do not provide such detailed information, usually focusing on an scalar-type efficiency measure in a particular direction. A particular advantage of the proposed approach is that besides the point-estimates of each specific slack, we also obtain the marginal effects of environmental variables on each of these slacks, as well as the standard errors for each slack and for each marginal effect (associated with each slack). In turn, these standard errors can be used for interval-estimates or for significance testing about these slacks or about the marginal effects.

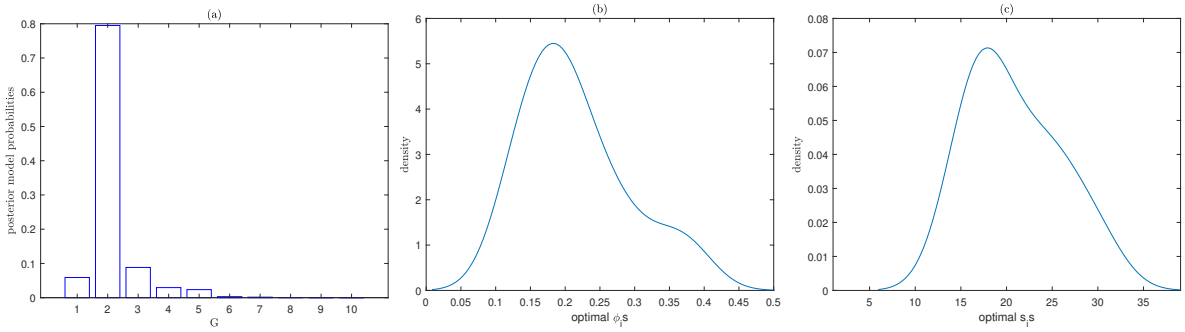
All in all, information like this, in Tables 1-3, can be very useful for researchers as well as practitioners wishing to analyze and understand which environmental variable (from the many considered) is a relevant predictor for each specific slack (for each input or output) and how substantial it is in terms of its magnitude relative to other relevant predictors. Therefore we hope this approach can complement the existing approaches very well, as another valuable instrument in the overall toolbox for performance analysis and testing of optimizing behavior.

Table 3: Z-statistics for the marginal effects of determinants onto slacks

Notes: w_j represents the jth relative price, that is $w_j = \log(W_j/W_1)$.

	equity	equity/TA	TA	trend	trend^2	w_1	w_2	w_3	w_4
slack x_1	-7.19	4.55	-1.52	0.93	4.92	2.66	-2.43	3.27	-4.72
slack x_2	0.94	-4.33	3.56	-10.48	-0.77	5.07	1.06	-1.06	0.60
slack x_3	-3.39	3.50	-5.56	-1.72	-0.57	1.31	-24.44	0.35	0.93
slack x_4	12.14	-3.21	3.24	-0.06	4.94	5.33	0.07	0.48	-4.80
slack x_5	24.38	0.35	6.31	-0.08	-0.78	-0.90	-0.92	1.14	-1.26
slack y_1	-0.57	2.05	-7.21	11.26	-0.25	4.98	0.00	0.55	-5.74
slack y_2	0.36	1.99	0.21	-22.44	3.04	1.19	3.00	-4.49	2.52
slack y_3	-2.66	2.91	3.13	1.06	-1.57	0.64	-1.19	1.67	-1.15
slack y_4	-0.56	3.87	13.52	1.22	-0.55	-0.57	-0.74	1.34	2.50
slack y_5	0.77	0.96	-0.53	10.33	-0.70	-52.05	-0.25	1.10	-5.03

Figure 1: Results related to compression and G



In Figure 1 we report posterior model probabilities for selecting G in (37) (see panel (a)), and distributions of optimal ϕ_is and dimensionality s_is in panels (b) and (c), respectively. From panel (a), G in (37) has a mode value of 2 and ranges, roughly, from 1 to 5. The optimal ϕ_is of Bayesian compression average 0.20 and range between zero and 0.5 (panel (b)). Finally, the optimal compression parameters s_i are close to 20 (which is much less than the sample size indicating significant compression) and range, roughly from 7 to 37.

From a qualitative perspective, it is worth noting that the estimated efficiency scores and the slacks on average from the proposed approach are fairly similar to those from other approaches we compare to. In particular, observing Table 4, note that, on average, the Debreu-Farrell type efficiency level suggested by the proposed approach is around 89% while those from DEA-CRS, DEA-NIRS, DEA-VRS are around 87%. Similar estimates and con-

Table 4: Measures of goodness of fit

Notes: Standard deviations are reported in parentheses where appropriate.

	this paper	DEA-CRS	DEA-NIRS	DEA-VRS	TK
Debreu-Farrell $\left(\frac{1}{\mathcal{OTE}(x^j,y^j\|\Psi)}\right)$	0.887	0.872	0.873	0.873	—
	(0.015)	(0.014)	(0.015)	(0.014)	
$\mathcal{OTE}(x^j, y^j\|\Psi)$	1.127	1.147	1.146	1.146	—
	(0.015)	(0.014)	(0.015)	(0.015)	
TGoF	0.870	0.893	0.872	0.872	0.874
FWGoF	0.869	0.893	0.871	0.872	0.875
System R^2	0.872	0.872	0.865	0.864	0.824

Table 5: log Bayes factors

Notes: Reported are LMLs in favor of a given model and against DEA. The Bayes factor is the LML of a given model minus the LML of DEA whose exponential is the evidence in favor of a given model against DEA. The LML for comparing "this paper" and TK is the difference of the two LMLs.

	Based on BIC	Based on Laplace approximation
this paper	282.32	283.55
TK	277.40	278.00
this paper against TK	95.85	96.71

clusions are for TGoF and FWGoF and for the System-R^2. While there is always a degree of subjectivity in a selection of models, we think some more confidence might be placed on the estimates of the present paper as they perform better in terms of log marginal likelihood and Bayes factors than TK.

From Table 5 it turns out that "this paper's" LML is much better compared to TK as the odds are overwhelming.

7 Concluding Remarks

In this paper we proposed a fairly general approach for modeling production technologies, which allows for modeling inefficiency and noise that are specific for each input and each output. The approach is based on merging ideas from nonparametric activity analysis models for production and consumption theory with stochastic frontier literature. We implement

this by re-casting the activity analysis models as simultaneous equations models in Bayesian compression and ANN frameworks. With minimal assumptions about noise in the data, the proposed approach allows for flexible approximations to input- and output-specific slacks. We also deploy Bayesian compression to solve the problem of an exceeding number of parameters for such general production technologies and describe a way to model determinants of inefficiency corresponding to each input and each output, to obtain marginal effects and their standard errors. Our Monte Carlo simulations provide encouraging results. We also provided an empirical illustration, where we compared the proposed approach with other alternatives, for the context of US banking data.

Finally, a fruitful avenue for future research would be to adapt the approach proposed in this paper to a consumer economics context, leveraging on the seminal works of Afriat (1967) and Varian (1982), Färe and Grosskopf (1995) and more recent works of Cherchye et al. (2007), Kitamura and Stoye (2018) and Smeulders et al. (2021), among others.

References

Afriat, S. N. (1967). The construction of utility functions from expenditure data. *International Economic Review*, 8(1):67–77.

Afriat, S. N. (1972). Efficiency estimation of production functions. *International Economic Review*, 13(3):568–598.

Aigner, D., Lovell, C., and Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics*, 6(1):21–37.

Bădin, L., Daraio, C., and Simar, L. (2012). How to measure the impact of environmental factors in a nonparametric production model. *European Journal of Operational Research*, 223(3):818–833.

Carter, R. A. and Nagar, A. L. (1977). Coefficients of correlation for simultaneous equation systems. *Journal of Econometrics*, 6(1):39–50.

Cazals, C., Florens, J.-P., and Simar, L. (2002). Nonparametric frontier estimation: A robust approach. *Journal of Econometrics*, 106(1):1–25.

Charnes, A., Cooper, W., Golany, B., Seiford, L., and Stutz, J. (1985). Foundations of data envelopment analysis for Pareto-Koopmans efficient empirical production functions. *Journal of Econometrics*, 30(1–2):91–107.

Charnes, A., Cooper, W., and Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2(6):429–444.

Chavas, J.-P. and Cox, T. L. (1990). A non-parametric analysis of productivity: The case of u.s. and japanese manufacturing. *The American Economic Review*, 80(3):450–464.

Cherchye, L., De Rock, B., and Vermeulen, F. (2007). The collective model of household consumption: A nonparametric characterization. *Econometrica*, 75(2):553–574.

Daouia, A. and Simar, L. (2007). Nonparametric efficiency analysis: A multivariate conditional quantile approach. *Journal of Econometrics*, 140(2):375 – 400.

Debreu, G. (1951). The coefficient of resource utilization. *Econometrica*, 19(3):273–292.

Diewert, W. (1973). Afriat and revealed preference theory. *The Review of Economic Studies*, 40(3):419–425.

Diewert, W. and Parkan, C. (1985). Tests for the consistency of consumer data. *Journal of Econometrics*, 30(1):127–147.

Durmus, A., Roberts, G., Vilmart, G., and Zygalakis, K. (2017). Fast langevin based algorithm for mcmc in high dimensions. *Annals of Applied Probability*, 27(4):2195–2237.

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

Färe, R., Fukuyama, H., Grosskopf, S., and Zelenyuk, V. (2015). Decomposing profit efficiency using a slack-based directional distance function. *European Journal of Operational Research*, 247(1):335–337.

Färe, R. and Grosskopf, S. (1995). Nonparametric tests of regularity, farrell efficiency, and goodness-of-fit. *Journal of Econometrics*, 69(2):415–425.

Färe, R., Grosskopf, S., Norris, M., and Zhang, Z. (1994). Productivity growth, technical progress, and efficiency change in industrialized countries. *American Economic Review*, 84(1):66–83.

Färe, R., He, X., Li, S. K., and Zelenyuk, V. (2019). A unifying framework for Farrell efficiency measurement. *Operations Research*, 67(1):183–197.

Färe, R. and Zelenyuk, V. (2020). Profit efficiency: Generalization, business accounting and the role of convexity. *Economics Letters*, 196:109483.

Farrell, M. J. (1957). The measurement of productive efficiency. *Jornal of the Royal Statistical Society. Series A (General)*, 120(3):253–290.

Fukuyama, H. and Weber, W. (2009). A directional slacks-based measure of technical inefficiency. *Socio-Economic Planning Sciences*, 43(4):274–287.

Geweke, J. (1992). *Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments*, pages 169–193. Clarendon Press, Oxford, UK.

Guhaniyogi, R. and Dunson, D. B. (2015). Bayesian compressed regression. *Journal of the American Statistical Association*, 110(512):1500–1514.

Hadjidoukas, P., Angelikopoulos, P., Voglis, C., Papageorgiou, D., and Lagaris, I. (2014). Ndl-v2.0: A new version of the numerical differentiation library for parallel architectures. *Computer Physics Communications*, 185(7):2217–2219.

Hanoch, G. and Rothschild, M. (1972). Testing the assumptions of production theory: A nonparametric approach. *The Journal of Political Economy*, 80(2):256–275.

Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109.

Hornik, K., Stinchcombe, M., and White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366.

Kitamura, Y. and Stoye, J. (2018). Nonparametric analysis of random utility models. *Econometrica*, 86(6):1883–1909.

Kneip, A., Park, B. U., and Simar, L. (1998). A note on the convergence of nonparametric DEA estimators for production efficiency scores. *Econometric Theory*, 14(6):783–793.

Kneip, A., Simar, L., and Wilson, P. W. (2008). Asymptotics and consistent bootstraps for DEA estimators in nonparametric frontier models. *Econometric Theory*, 24(6):1663–1697.

Kneip, A., Simar, L., and Wilson, P. W. (2015). When bias kills the variance: Central limit theorems for DEA and FDH efficiency scores. *Econometric Theory*, 31(2):394–422.

Koopmans, T. (1951). *Activity Analysis of Production and Allocation*. New York, NY: Wiley.

Korostelev, A., Simar, L., and Tsybakov, A. B. (1995). Efficient estimation of monotone boundaries. *Annals of Statistics*, 23(2):476–489.

Kumar, S. and Russell, R. R. (2002). Technological change, technological catch-up, and capital deepening: Relative contributions to growth and convergence. *American Economic Review*, 92(3):527–548.

Kumbhakar, S. and Schmidt, P. (2016). *Endogeneity Problems in Econometrics, Special Issue of the Journal of Econometrics*, volume 190. Amsterdam, Netherlands: Elsevier.

Kumbhakar, S. C., Park, B. U., Simar, L., and Tsionas, E. G. (2007). Nonparametric Stochastic Frontiers: A Local Maximum Likelihood Approach. *Journal of Econometrics*, 137(1):1–27.

Kumbhakar, S. C., Parmeter, C. F., and Zelenyuk, V. (2021a). *Stochastic Frontier Analysis: Foundations and Advances I*, page forthcoming. Springer Singapore, Singapore.

Kumbhakar, S. C., Parmeter, C. F., and Zelenyuk, V. (2021b). *Stochastic Frontier Analysis: Foundations and Advances II*, page forthcoming. Springer Singapore, Singapore.

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

Kuosmanen, T. (2008). Representation theorem for convex nonparametric least squares. *The Econometrics Journal*, 11(2):308–325.

Kuosmanen, T. and Johnson, A. (2017). Modeling joint production of multiple outputs in StoNED: Directional distance function approach. *European Journal of Operational Research*, 262(2):792–801.

Kuosmanen, T. and Johnson, A. L. (2010). Data envelopment analysis as nonparametric least-squares regression. *Operations Research*, 58(1):149–160.

Liu, J. S. (2008). *Monte Carlo strategies in scientific computing / Jun S. Liu.* Springer Series in Statistics. Springer, New York.

Malikov, E., Kumbhakar, S. C., and Tsionas, M. G. (2016). 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.

Martins-Filho, C. and Yao, F. (2015). Semiparametric stochastic frontier estimation via profile likelihood. *Econometric Reviews*, 34(4):413–451.

Mas-Colell, A., Whinston, M. D., and Green, J. R. (1995). *Microeconomic Theory.* Oxford University Press.

Meeusen, W. and van den Broeck, J. (1977). Efficiency estimation from Cobb-Douglas production functions with composed error. *International Economic Review*, 18(2):435–44.

Olesen, O. B. and Petersen, N. C. (2016). Stochastic data envelopment analysis–a review. *European Journal of Operational Research*, 251(1):2–21.

Park, B. U., Jeong, S.-O., and Simar, L. (2010). Asymptotic distribution of conical-hull estimators of directional edges. *Annals of Statistics*, 38(3):1320–1340.

Park, B. U., Simar, L., and Weiner, C. (2000). The FDH estimator for productivity efficiency scores: Asymptotic properties. *Econometric Theory*, 16(6):855–877.

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

Parmeter, C. F. and Zelenyuk, V. (2019). Combining the virtues of stochastic frontier and data envelopment analysis. *Operations Research*, 67(6):1628–1658.

Paul, S. and Shankar, S. (2018). On estimating efficiency effects in a stochastic frontier model. *European Journal of Operational Research*, 271(2):769–774.

Ray, S. C. (2004). *Data Envelopment Analysis: Theory and Techniques for Economics and Operations Research*. New York, NY: Cambridge University Press.

Robert, C. P., Robert, C., and Casella, G. (2004). *Monte Carlo statistical methods / Christian P. Robert, George Casella.* Springer texts in statistics. Springer, New York, 2nd ed. edition.

Shephard, R. W. (1953). *Cost and Production Functions*. Princeton, NJ: Princeton University Press.

Shephard, R. W. (1970). *Theory of Cost and Production Functions*. Princeton studies in mathematical economics. Princeton, NJ: Princeton University Press.

Sickles, R. and Zelenyuk, V. (2019). *Measurement of Productivity and Efficiency: Theory and Practice*. New York, NY: Cambridge University Press.

Simar, L., Van Keilegom, I., and Zelenyuk, V. (2017). Nonparametric least squares methods for stochastic frontier models. *Journal of Productivity Analysis*, 47(3):189–204.

Simar, L. and Wilson, P. W. (2015). Statistical approaches for nonparametric frontier models: A guided tour. *International Statistical Review*, 83(1):77–110.

Simar, L. and Zelenyuk, V. (2007). Statistical inference for aggregates of Farrell-type efficiencies. *Journal of Applied Econometrics*, 22(7):1367–1394.

Simar, L. and Zelenyuk, V. (2011). Stochastic FDH/DEA estimators for frontier analysis. *Journal of Productivity Analysis*, 36(1):1–20.

Simar, L. and Zelenyuk, V. (2018). Central limit theorems for aggregate efficiency. *Operations Research*, 166(1):139–149.

Smeulders, B., Cherchye, L., and Rock, B. D. (2021). Nonparametric Analysis of Random Utility Models: Computational Tools for Statistical Testing. *Econometrica*, 89(1):437–455.

Tone, K. (2001). A slacks-based measure of efficiency in data envelopment analysis. *European Journal of Operational Research*, 130(3):498–509.

Varian, H. R. (1982). The nonparametric approach to demand analysis. *Econometrica*, 50(4):945–973.

Varian, H. R. (1983). Non-parametric tests of consumer behaviour. *Review of Economic Studies*, 50(1):99–110.

Varian, H. R. (1984). The nonparametric approach to production analysis. *Econometrica*, 52(3):579–597.

Varian, H. R. (1990). Goodness-of-fit in optimizing models. *Journal of Econometrics*, 46(1):125–140.

Voglis, C., Hadjidoukas, P., Lagaris, I., and Papageorgiou, D. (2009). A numerical differentiation library exploiting parallel architectures. *Computer Physics Communications*, 180(8):1404–1415.

White, H. (1989). Learning in artificial neural networks: A statistical perspective. *Neural Computation*, 1(4):425–464.

White, H. (1990). Connectionist nonparametric regression: Multilayer feedforward networks can learn arbitrary mappings. *Neural Networks*, 3(5):535–549.

Supplement 1: Monte Carlo Evidence

Consider the production technology of the following form

$$\Psi \;=\; \left\{(x,y)\;:\;\left(\sum_{m=1}^{M}\beta_m(y_m)^2\right)^{1/2}\leq\left(\prod_{l=1}^{N}(x_l)^{\alpha_l}\right)\right\} \tag{A.1}$$

where $\beta_m \geq 0$, $\alpha_l \geq 0$ and to assure linear homogeneity of the distance functions we let $\sum_{l=1}^{N}\alpha_l = 1$. Moreover, without loss of generality, we normalize $\sum_{m=1}^{M}\beta_m = 1$.

We chose the coefficient randomly in each Monte Carlo (MC) replication as follows: for each $l \in \{2,...,N\}$ we set $\alpha_l = \left(\widetilde{\alpha}_l / \left(\sum_{l=2}^{N}\widetilde{\alpha}_l\right)\right)(1-\alpha_1)$ where $\widetilde{\alpha}_l \overset{iid}{\sim}$ Uniform$(0,1)$, while α_1 is normalized to some arbitrary value in $(0,1)$, e.g., 0.1, and for each $m \in \{1,...,M\}$ we set $\beta_m = \widetilde{\beta}_m / \sum_{m=1}^{M}\widetilde{\beta}_m$ where $\widetilde{\beta}_m \overset{iid}{\sim}$ Uniform$(0,1)$.

We first generate efficient outputs for each $k \in \{1,...,n\}$, e.g., as $\widetilde{y}_{im} \overset{iid}{\sim}$ Uniform$(0.1,1)$ for each $m \in \{1,...,M\}$, and then generate $N-1$ of actual (i.e., inefficient) inputs, e.g., as $x_l = x_l^* + u_l^{(x)}$, where for each i, we have $x_{il} \overset{iid}{\sim}$ Uniform$(0.1,1)$ and $u_{il} \overset{iid}{\sim} |N(0,\sigma_{lu^{(x)}})|$, for each $l \in \{2,...,N\}$, and then define the remaining efficient input, x_{i1}, in terms of the other generated inputs and all the outputs via (A.1), i.e., for each $i \in \{1,...,n\}$, set

$$x_{i1} := \left(\left(\left(\sum_{m=1}^{M}\beta_m(\widetilde{y}_{im})^2\right)^{1/2}\right)/\left(\prod_{l=2}^{N}(x_{il})^{\alpha_l}\right)\right)^{1/\alpha_1}. \tag{A.2}$$

Finally, we get the inefficient input x_1 as $x_1 = x_1^* + u_1^{(x)}$ where $u_{i1} \overset{iid}{\sim} |\mathcal{N}(0,\sigma_{1u^{(x)}})|$.

Furthermore, we assume that a researcher observes the inefficient inputs and instead of the maximal or efficient outputs $\tilde{y}_i$ the researcher observes inefficient outputs defined as

$$y_i = \widetilde{y}_i \times \exp(-u_i^{(y)}), \tag{A.3}$$

where $u_i^{(y)} \overset{iid}{\sim} |\mathcal{N}_m(0,\Omega_u)|$, for $i = 1,...,n$. Note that in this setup, the inefficiencies are purely random and below we explain how to have them dependent on determinants (or

environmental variables).

Also note that the Shephard's output distance function for this scenario is given by

$$\begin{aligned} D_o(x,y) &= \inf\left\{\theta>0 \ : \ \left(\sum_{m=1}^{M}\beta_m(y_m/\theta)^2\right)^{1/2} \leq \left(\prod_{l=1}^{N}(x_l)^{\alpha_l}\right)\right\} \\ &= \left(\sum_{m=1}^{M}\beta_m(y_m)^2\right)^{1/2} / \left(\prod_{l=1}^{N}(x_l)^{\alpha_l}\right), \end{aligned} \tag{A.4}$$

which is a univariate measure of inefficiency (and is a reciprocal of the output oriented Debreu-Farrell type measure of technical (in)efficiency), with the convenient property that $D_o(x,y) \in [0,1]$, where 1 represents the 100% technical efficiency score.

Finally, to introduce the environmental variables into this data generating process, recall that the proposed approach models them as follows:

$$\begin{gathered} \text{Input slacks:} \\ U_{i1}^{(x)} = -\ln\textstyle\sum_{g=1}^{G} p_{g1}^{(x)} F(w_{i1}^{(x)\prime}\delta_{g1}^{(x)} + \epsilon_{i1}),\ i=1,...,n \\ U_{i2}^{(x)} = -\ln\textstyle\sum_{g=1}^{G} p_{g2}^{(x)} F(w_{i2}^{(x)\prime}\delta_{g2}^{(x)} + \epsilon_{i2}),\ i=1,...,n \\ \vdots \\ U_{iN}^{(x)} = -\ln\textstyle\sum_{g=1}^{G} p_{gN}^{(x)} F(w_{iN}^{(x)\prime}\delta_{gN}^{(x)} + \epsilon_{iN}^{(x)}),\ i=1,...,n \\ \text{Output slacks:} \\ U_{i1}^{(y)} = -\ln\textstyle\sum_{g=1}^{G} p_{g1}^{(y)} F(w_{i1}^{(y)\prime}\delta_{g1}^{(y)} + \epsilon_{i,N+1}^{(y)}),\ i=1,...,n \\ U_{i2}^{(y)} = -\ln\textstyle\sum_{g=1}^{G} p_{g2}^{(y)} F(w_{i2}^{(y)\prime}\delta_{g2}^{(y)} + \epsilon_{i,N+2}^{(y)}),\ i=1,...,n \\ \vdots \\ U_{iM}^{(y)} = -\ln\textstyle\sum_{g=1}^{G} p_{gM}^{(y)} F(w_{iM}^{(y)\prime}\delta_{gM}^{(y)} + \epsilon_{i,N+M}^{(y)}),\ i=1,...,n \end{gathered} \tag{A.5}$$

where $\delta_{gj}^{(x)} \in \mathbb{R}^{d_w}$ $(j=1,...,N)$, $\delta_{gj}^{(y)} \in \mathbb{R}^{d_w}$ $(j=1,...,M)$ are parameter vectors and, similarly, we have the probability parameters $p_{gj}^{(x)} \in \mathbb{R}^{d_w}$ $(j=1,...,N)$, $p_{gj}^{(y)} \in \mathbb{R}^{d_w}$ $(j=1,...,M)$.

Hence, we will generalize the functional technology characterization (A.4) as following

$$\begin{aligned} D_o(x,y|F(w_i'\delta)) &= \inf\left\{\theta>0\ :\ \left(\sum_{m=1}^{M}\beta_m(y_m/\theta)^2\right)^{1/2}\leq\left(\prod_{l=1}^{N}(x_l)^{\alpha_l}\right)F(w_i'\delta)\right\} \\ &= \left(\sum_{m=1}^{M}\beta_m(y_m)^2\right)^{1/2}/\left(F(w_i'\delta)\prod_{l=1}^{N}(x_l)^{\alpha_l}\right) \end{aligned} \tag{A.6}$$

where, as before, $F(w_i'\delta)$ is a cumulative distribution function (cdf) of, say, the normal or the logistic distribution, and each element of w_i is generated randomly, e.g., from standard uniform, while the vector of parameters δ is fixed at some level. Note that there are two types of inefficiencies in (A.6), the purely random inefficiencies, $u_{il}^{(x)}$ and $u_{im}^{(y)}$, for $l=1,...,N$ and $m=1,...,M$ when generating the data, and the deterministic inefficiency $F(w_i'\delta)$ which results if we set $\Sigma_\epsilon=\mathbf{O}$, that is there are no random error terms in (A.5).

Alternatively, the environmental variables can be also introduced more generally, by setting

$$U_{il}^{(x)}=-\ln F(w_{il}^{(x)\prime}\delta_l^{(x)}+\epsilon_{il}^{(x)}),\ i=1,...,n, \tag{A.7}$$

for each $l=1,...,N$ and

$$U_{im}^{(y)}=-\ln F(w_{im}^{(y)\prime}\delta_m^{(y)}+\epsilon_{im}^{(y)}),\ i=,...,n, \tag{A.8}$$

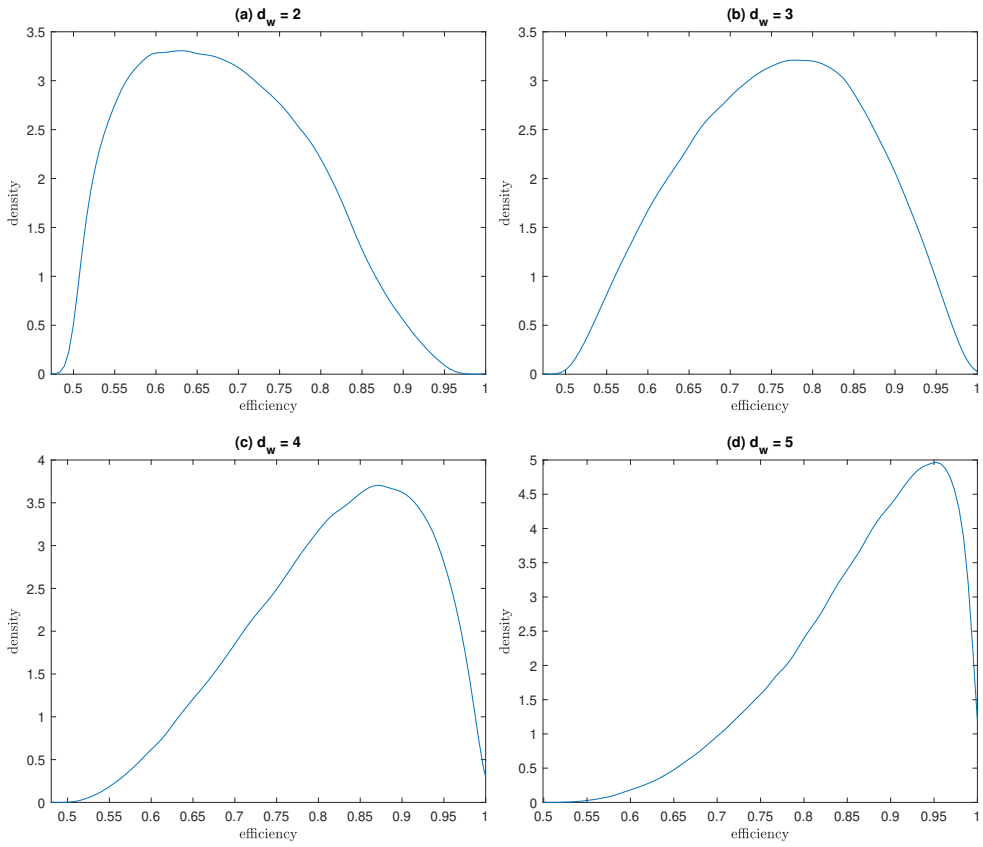
for each $m=1,...,M$, and where each element of $w_{il}^{(x)}$ and $w_{im}^{(y)}$ is generated randomly, e.g., from standard uniform, while the vector of parameters $\delta_l^{(x)}$ and $\delta_m^{(y)}$ are fixed at some levels. Then, these inefficiencies can be used in the data generation as described above in place of the purely random $u_{il}^{(x)}$ and $u_{im}^{(y)}$, for $l=1,...,N$ and $m=1,...,M$. Moreover, for all $i=1,...,n$ we first set $\epsilon_{il}^{(x)}=\epsilon_{im}^{(y)}=0$, for each $l=1,...,N$ and $m=1,...,M$.

We assume that Σ is a diagonal matrix whose diagonal elements are all equal to $\sigma_v>0$. If we assume there are no environmental variables we generate $U_i^{(x)}$, $U_i^{(y)}\sim$ i.i.d $\mathcal{N}_+(0,\sigma_u^2)$, a half-normal distribution. In this case we can set the signal-to-noise ratio $\lambda=\frac{\sigma_u}{\sigma_v}$ to alternative

values. In the presence of environmental variables this is no longer possible but we can still vary σ_v. We use two environmental variables. In our Monte Carlo we pretend that G in (A.5) is unknown so it has to be selected using the marginal likelihood criterion (DiCiccio et al., 1997). Moreover, we follow the compression scheme that we described in the main text of the paper. Additionally, we can experiment with the presence or absence of random errors ϵ_i in (A.5) or (A.7)-(A.8). We also impose $\sum_{j=1}^{n} \lambda_{ij} = 1$ for each i.

If we generate the two w_is from standard uniform distributions and their coefficients from a uniform distribution in $(0, 1)$, in Figure A.1 we display the density of efficiency resulting from 10,000 different combinations of the two w_is (with sample size $n = 50$) and their coefficients, when the errors $\epsilon_{il}^{(x)} = \epsilon_{im}^{(y)} = 0$. To examine the stochastic inefficiency case we draw the errors $\epsilon_{il}^{(x)}, \epsilon_{im}^{(y)}$ from a normal distribution with zero mean and common variance s^2 which we adjust so that the average sample signal-to-noise ratio has a given value (λ).

Figure A.1: Efficiency distributions with environmental variables



The standard deviations of inefficiency range from 0.148 ($d_w = 2$) to 0.110 (for $d_w = 5$) so it seems reasonable to examine values of σ_v like $0.05, 0.10, 0.15, 0.20, 0.25$ and 0.50. Without environmental variables, we will examine the same configurations of these parameters for consistency.

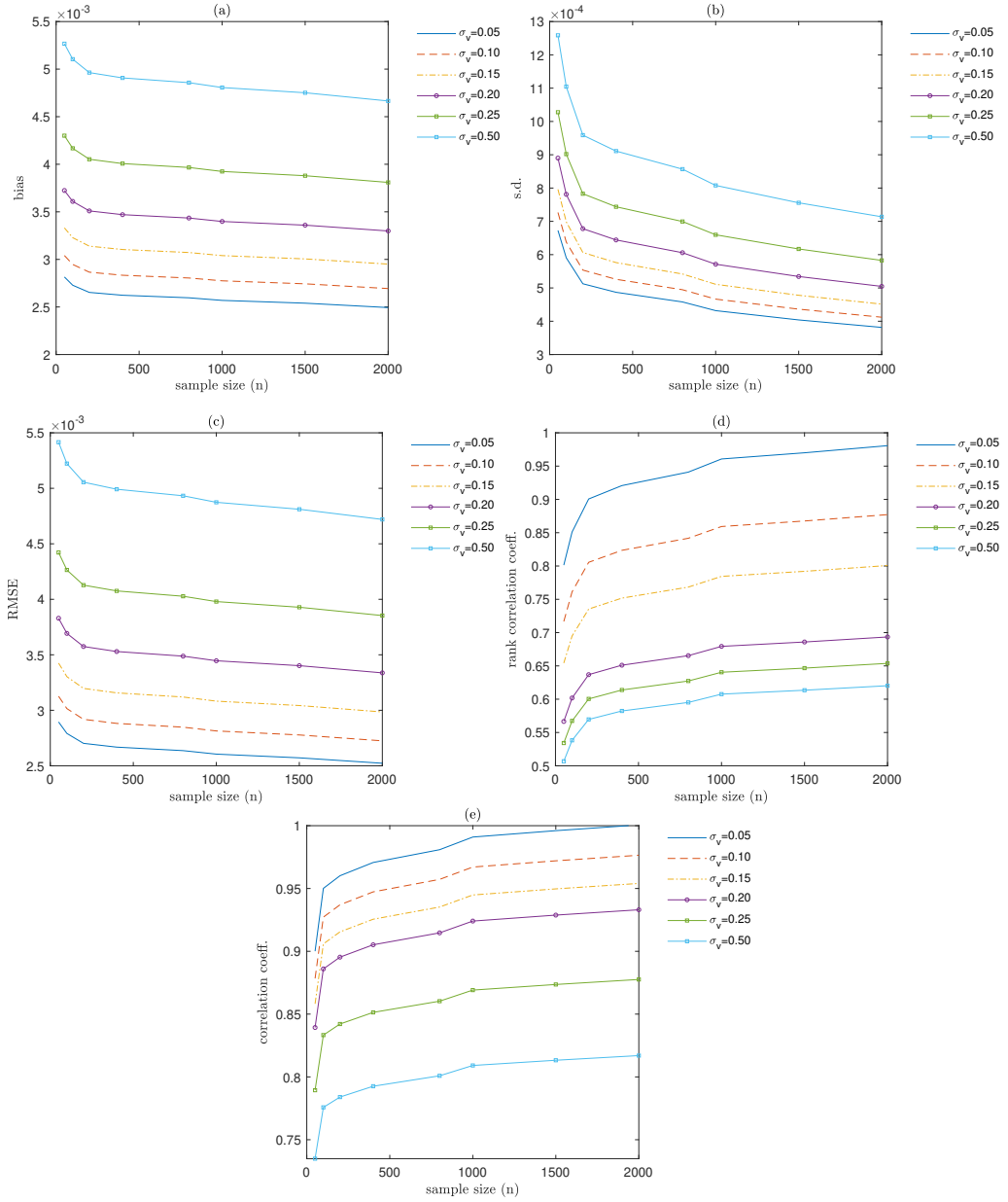
We use 5,000 Monte Carlo simulations each of which employs MCMC using 15,000 iterations omitting the first 5,000 to mitigate possible start up effects. Our interest focuses on bias, standard deviation and RMSE of inefficiency estimates.

Table 6: Monte Carlo results (without environmental variables)

$\sigma_v \downarrow \setminus n \rightarrow$	50	100	200	400	800	1000	1500	2000
Bias								
0.05	0.2815	0.2728	0.2653	0.2623	0.2597	0.2569	0.2540	0.2494
0.10	0.3041	0.2947	0.2866	0.2833	0.2805	0.2775	0.2743	0.2694
0.15	0.3331	0.3228	0.3139	0.3104	0.3072	0.3040	0.3005	0.2951
0.20	0.3724	0.3609	0.3510	0.3470	0.3435	0.3399	0.3360	0.3299
0.25	0.4300	0.4168	0.4053	0.4007	0.3966	0.3924	0.3880	0.3809
0.50	0.5267	0.5104	0.4964	0.4908	0.4858	0.4806	0.4752	0.4665
Standard deviation								
0.05	0.0673	0.0590	0.0513	0.0487	0.0458	0.0432	0.0404	0.0382
0.10	0.0727	0.0638	0.0554	0.0526	0.0495	0.0466	0.0436	0.0412
0.15	0.0796	0.0699	0.0606	0.0576	0.0542	0.0511	0.0478	0.0451
0.20	0.0890	0.0781	0.0678	0.0644	0.0606	0.0571	0.0534	0.0505
0.25	0.1028	0.0902	0.0783	0.0744	0.0700	0.0660	0.0617	0.0583
0.50	0.1259	0.1105	0.0959	0.0911	0.0857	0.0808	0.0756	0.0714
RMSE (Root Mean Squared Error)								
0.05	0.2895	0.2792	0.2702	0.2668	0.2637	0.2605	0.2572	0.2523
0.10	0.3127	0.3015	0.2919	0.2882	0.2848	0.2814	0.2778	0.2725
0.15	0.3425	0.3303	0.3197	0.3157	0.3120	0.3083	0.3043	0.2985
0.20	0.3829	0.3693	0.3575	0.3530	0.3488	0.3446	0.3402	0.3337
0.25	0.4422	0.4264	0.4128	0.4076	0.4027	0.3980	0.3928	0.3854
0.50	0.5415	0.5223	0.5055	0.4992	0.4933	0.4874	0.4811	0.4720
Rank Correlation Coefficient with true inefficiency								
0.05	0.8013	0.8511	0.9007	0.9207	0.9410	0.9607	0.9700	0.9807
0.10	0.7167	0.7613	0.8056	0.8235	0.8416	0.8593	0.8676	0.8772
0.15	0.6543	0.6949	0.7354	0.7518	0.7683	0.7844	0.7920	0.8007
0.20	0.5666	0.6018	0.6369	0.6511	0.6654	0.6793	0.6859	0.6935
0.25	0.5342	0.5674	0.6004	0.6138	0.6273	0.6405	0.6467	0.6538
0.50	0.5068	0.5383	0.5696	0.5823	0.5951	0.6076	0.6135	0.6202
Correlation Coefficient with true inefficiency								
0.05	0.9001	0.9501	0.9602	0.9707	0.9809	0.9910	0.9960	0.9975
0.10	0.8784	0.9272	0.9370	0.9473	0.9572	0.9671	0.9720	0.9765
0.15	0.8582	0.9059	0.9155	0.9255	0.9352	0.9448	0.9497	0.9540
0.20	0.8393	0.8860	0.8953	0.9052	0.9147	0.9241	0.9288	0.9331
0.25	0.7894	0.8333	0.8421	0.8514	0.8603	0.8691	0.8736	0.8776
0.50	0.7349	0.7758	0.7840	0.7926	0.8009	0.8091	0.8133	0.8170

Notes: Results for biases, standard deviations and RMSEs are multiplied by 10.

Figure A.2: Monte Carlo results (without environmental variables)



Notes: Results for biases, standard deviations and RMSEs are not multiplied by 10.

Table 7: Monte Carlo results (with environmental variables, deterministic inefficiency)

Notes: Results for biases, standard deviations and RMSEs are multiplied by 10. "Deterministic inefficiency" corresponds to the case when the errors $\epsilon_i$set to zero. In this case we do not need the ξ_is and, instead, we use the formulation in (A.5).

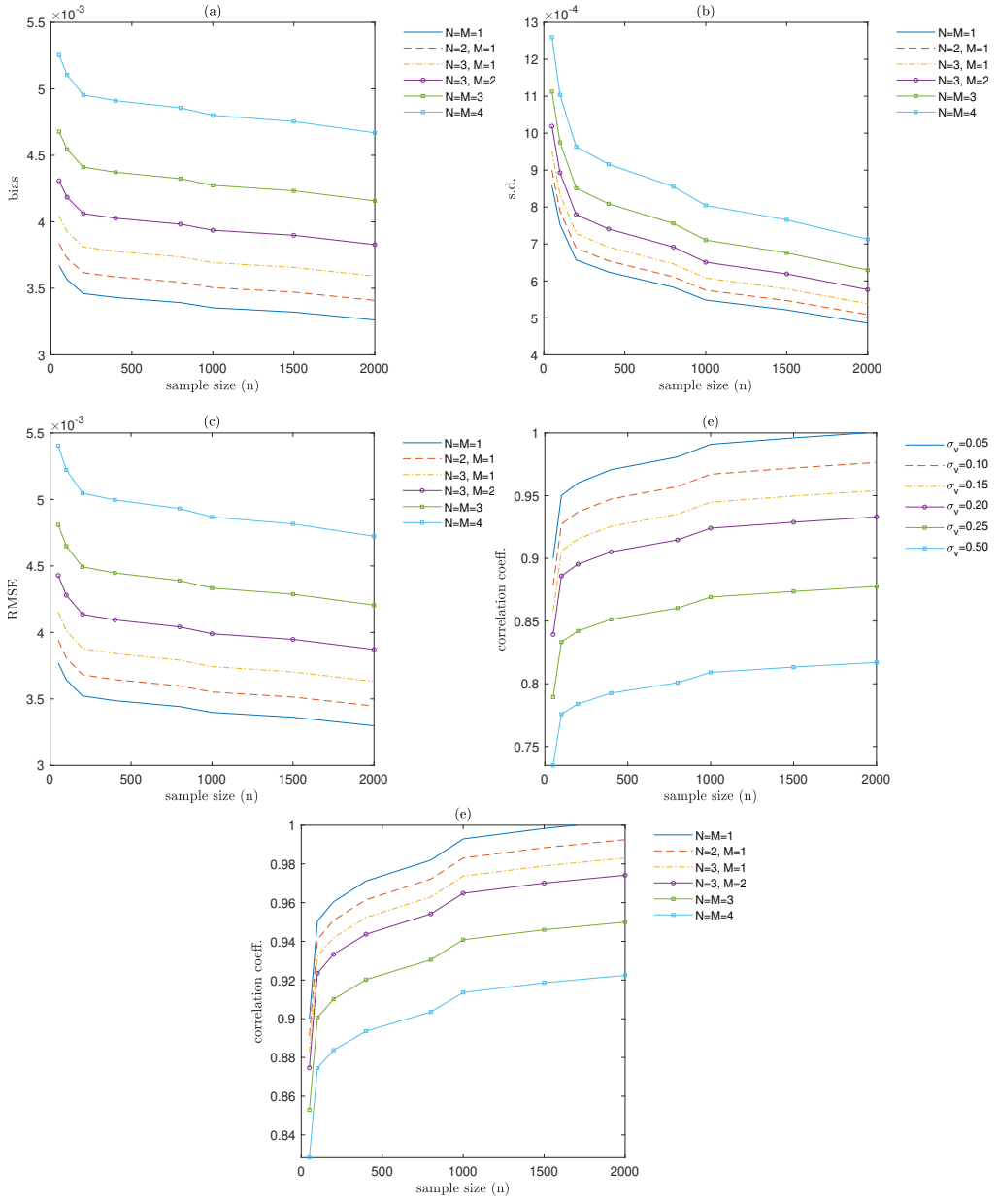
$(N,M)\downarrow \setminus n \rightarrow$	50	100	200	400	800	1000	1500	2000
Bias								
(1,1)	0.3841	0.3732	0.3622	0.3586	0.3554	0.3514	0.3473	0.3410
(2,1)	0.3992	0.3879	0.3764	0.3726	0.3694	0.3652	0.3610	0.3544
(3,1)	0.4178	0.4060	0.3940	0.3900	0.3866	0.3823	0.3778	0.3710
(3,2)	0.4418	0.4293	0.4166	0.4124	0.4088	0.4042	0.3995	0.3922
(3,3)	0.4747	0.4613	0.4477	0.4432	0.4393	0.4343	0.4293	0.4215
(4,4)	0.5253	0.5105	0.4954	0.4904	0.4861	0.4807	0.4751	0.4665
Standard Deviation								
(1,1)	0.0921	0.0816	0.0702	0.0661	0.0631	0.0596	0.0552	0.0515
(2,1)	0.0957	0.0848	0.0729	0.0687	0.0656	0.0620	0.0574	0.0535
(3,1)	0.1001	0.0888	0.0763	0.0719	0.0687	0.0649	0.0601	0.0560
(3,2)	0.1059	0.0939	0.0807	0.0760	0.0726	0.0686	0.0635	0.0592
(3,3)	0.1138	0.1009	0.0867	0.0817	0.0780	0.0737	0.0682	0.0636
(4,4)	0.1259	0.1116	0.0960	0.0904	0.0864	0.0816	0.0755	0.0704
RMSE (Root Mean Squared Error)								
(1,1)	0.3950	0.3821	0.3690	0.3646	0.3610	0.3564	0.3517	0.3449
(2,1)	0.4105	0.3971	0.3834	0.3789	0.3752	0.3704	0.3655	0.3584
(3,1)	0.4296	0.4156	0.4013	0.3966	0.3926	0.3877	0.3826	0.3752
(3,2)	0.4543	0.4394	0.4244	0.4193	0.4152	0.4100	0.4045	0.3967
(3,3)	0.4882	0.4722	0.4560	0.4506	0.4461	0.4405	0.4347	0.4263
(4,4)	0.5402	0.5226	0.5046	0.4987	0.4937	0.4875	0.4810	0.4717
Rank Correlation Coefficient								
(1,1)	0.8015	0.8512	0.9003	0.9207	0.9406	0.9606	0.9704	0.9804
(2,1)	0.7580	0.8051	0.8515	0.8708	0.8895	0.9085	0.9178	0.9272
(3,1)	0.7243	0.7692	0.8135	0.8320	0.8499	0.8680	0.8769	0.8859
(3,2)	0.6740	0.7158	0.7571	0.7743	0.7909	0.8078	0.8160	0.8244
(3,3)	0.6544	0.6950	0.7351	0.7518	0.7680	0.7844	0.7923	0.8005
(4,4)	0.6374	0.6770	0.7160	0.7322	0.7480	0.7640	0.7717	0.7797
Correlation Coefficient								
(1,1)	0.9000	0.9501	0.9601	0.9704	0.9806	0.9907	0.9959	0.9965
(2,1)	0.8891	0.9386	0.9485	0.9586	0.9687	0.9787	0.9839	0.9884
(3,1)	0.8789	0.9277	0.9375	0.9475	0.9575	0.9674	0.9725	0.9770
(3,2)	0.8691	0.9175	0.9271	0.9371	0.9469	0.9567	0.9617	0.9662
(3,3)	0.8429	0.8898	0.8991	0.9088	0.9183	0.9278	0.9327	0.9370
(4,4)	0.8133	0.8585	0.8675	0.8768	0.8861	0.8952	0.8999	0.9041

Notes: Results for biases, standard deviations and RMSEs are multiplied by 10.

Generally, RMSEs do not scale as $\sqrt{n}$. Previously developed asymptotic theory is also in

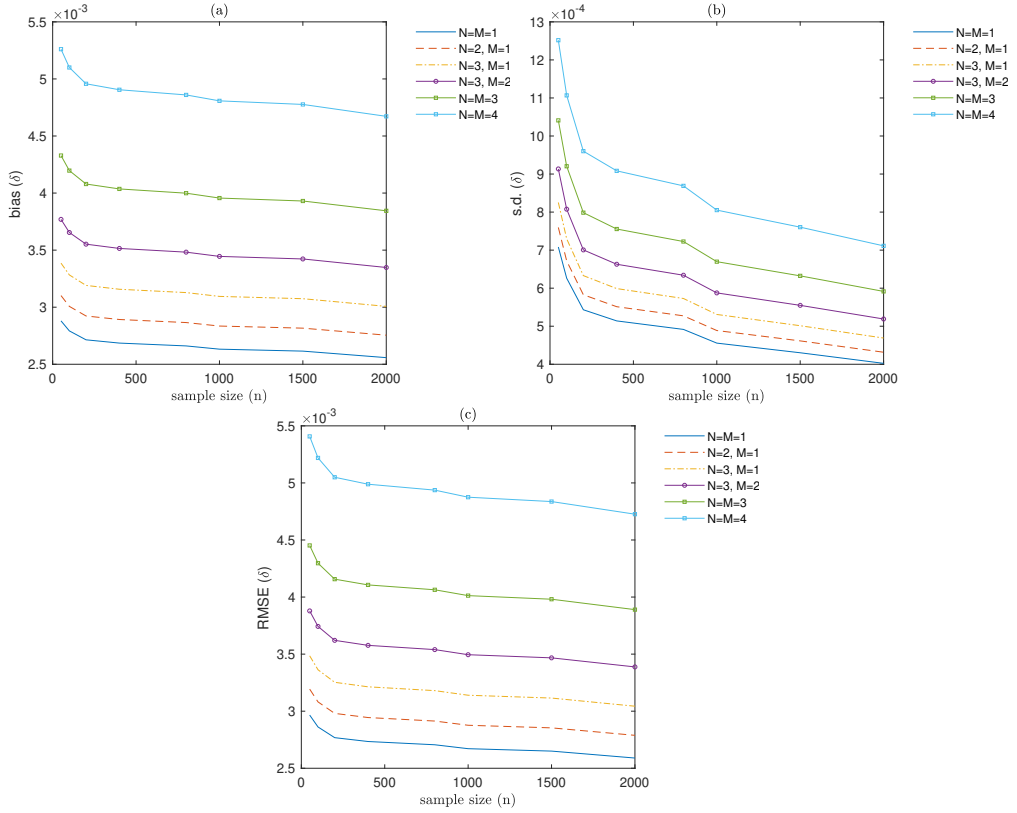
Figure A.3: Monte Carlo results (with environmental variables, deterministic inefficiency)

Notes: Results for biases, standard deviations and RMSEs are not multiplied by 10. "Deterministic inefficiency" corresponds to the case when the errors $\epsilon_i$set to zero. In this case we do not need the ξ_is and, instead, we use the formulation in (A.5).



line with this finding. In Figure A.4, we report the estimates of biases, standard deviations and RMSEs for the two δ inefficiency parameters in (A.5). As a function of the sample size, the estimates of bias are presented in panel (a), standard deviations in panel (b) and RMSEs in panel (c), for a variety of values of N and M. Due to the parameterizations of inefficiency in (A.5) it seems that RMSEs scale approximately as $\sqrt{n}$ for all N and M. Finally, as we have two parameters in δ, we report maximum bias, standard and RMSE.

Figure A.4: Monte Carlo results for inefficiency parameters (δ), deterministic inefficiency



Notes: In Figure A.4, we report the estimates of biases, standard deviations and RMSEs for the two δ inefficiency parameters in (A.5). As a function of the sample size, the estimates of bias are presented in panel (a), standard deviations in panel (b) and RMSEs in panel (c), for a variety of values of N and M. As we have two parameters in δ, we report maximum bias, standard deviation and RMSE.

The case with stochastic inefficiency does not yield qualitatively different results.[16]

[16]The exact results are available in Supplement 2 in in Tables OS-1 and OS-2.

Supplement 2. Technical Appendix

MCMC

We use a recent advance on the Metropolis Adjusted Langevin Algorithm (MALA) called fast MALA (fMALA), see Durmus et al. (2017). Suppose we have a parameter vector $\boldsymbol{\theta} \in \mathbb{R}^d$, where now d denotes the dimensionality of the parameter vector, and we target the non-normalized density $\pi(\boldsymbol{\theta})$ which represents the posterior, omitting the dependence on data to ease notation. We consider a Langevin diffusion defined by:

$$\mathrm{d}\boldsymbol{\theta}_t = \tfrac{1}{2}\boldsymbol{\Sigma} \cdot \nabla \ln \pi(\boldsymbol{\theta}_t) + \boldsymbol{\Sigma}^{1/2}\mathrm{d}\boldsymbol{W}_t, \tag{A.9}$$

where $\{\boldsymbol{W}_t, t \geq 0\}$ is a standard d-dimensional Brownian motion, and $\boldsymbol{\Sigma}$ is a given positive definite self-adjoint matrix. Under appropriate assumptions on π one can show that the dynamics generated by (A.9) are ergodic and result in $\pi(\boldsymbol{\theta})$ as the unique invariant distribution. A standard approach is to discretize (A.9) using a one step integrator, and sample using the averages over the numerical trajectories. This approach introduces a bias because the posterior does not coincide in general with the exact π.

An alternative way of sampling from π exactly, i.e., such that it is not biased by discretizing (A.9), is by using the Metropolis-Hastings algorithm (Hastings (1970)). The idea is to construct a Markov chain $\{\boldsymbol{\theta}_j\}$, where at each step j, given $\boldsymbol{\theta}_j$, a new sample proposal $\boldsymbol{\theta}^c$ is generated from the Markov chain with a transition kernel $q(\boldsymbol{\theta}, \cdot)$. This proposal is then accepted ($\boldsymbol{\theta}_{j+1} = \boldsymbol{\theta}^c$) with probability $\alpha(\boldsymbol{\theta}_j, \boldsymbol{\theta}^c)$ and rejected ($\boldsymbol{\theta}_{j+1} = \boldsymbol{\theta}_j$) otherwise. If we have

$$\alpha(\boldsymbol{\theta}, \boldsymbol{\theta}^c) = \min\left\{1, \frac{\pi(\boldsymbol{\theta}^c)q(\boldsymbol{\theta}^c, \boldsymbol{\theta})}{\pi(\boldsymbol{\theta})q(\boldsymbol{\theta}, \boldsymbol{\theta}^c)}\right\}, \tag{A.10}$$

then the resulting Markov chain $\{\theta_j\}$ is π-invariant and will, for large j generate samples from π under mild ergodicity assumptions (Liu (2008), Robert et al. (2004)). In general, a

candidate is generated as:

$$\boldsymbol{\theta}^c = \boldsymbol{\mu}(\boldsymbol{\theta}, h) + \boldsymbol{S}(\boldsymbol{\theta}, h)\boldsymbol{\zeta}, \tag{A.11}$$

where $\boldsymbol{\zeta} \sim \mathcal{N}_d(\mathbf{0}, \mathbf{I}_d)$. The specific fMALA proposal has

$$\boldsymbol{\mu}(\boldsymbol{\theta}, h) = x + \tfrac{h}{2} f(\boldsymbol{\theta}) - \tfrac{h^2}{24} \nabla f(\boldsymbol{\theta}) \cdot f(\boldsymbol{\theta}) + \{\boldsymbol{\Sigma} : \nabla^2 f(\boldsymbol{\theta})\}, \tag{A.12}$$

$$\boldsymbol{S}(\boldsymbol{\theta}, h) = \left(h^{1/2}\mathbf{I}_d + \tfrac{h^{3/2}}{12} Df(\boldsymbol{\theta})\right) \boldsymbol{\Sigma}^{1/2}, \tag{A.13}$$

where $f(\boldsymbol{\theta}) \triangleq \boldsymbol{\Sigma} \cdot \nabla \ln \pi(\boldsymbol{\theta})$, $\nabla f(\boldsymbol{\theta})$ and $\nabla^2 f(\boldsymbol{\theta})$ are the $d \times d$ Jacobian and $d \times d^2$ Hessian of $f(\boldsymbol{\theta})$, respectively, and $\boldsymbol{\Sigma} = \boldsymbol{S}(\boldsymbol{\theta}, h)$. Let $\nabla^2 f(\boldsymbol{\theta}) = [\boldsymbol{H}_1(\boldsymbol{\theta}), \ldots, \boldsymbol{H}_d(\boldsymbol{\theta})]$ where $[\boldsymbol{H}_i(\boldsymbol{\theta})]_{jk} = \frac{\partial^2 f_i(\theta)}{\partial \theta_k \partial \theta_j}$, then $\{\boldsymbol{\Sigma} : \nabla^2 f(\boldsymbol{\theta})\}_i \triangleq tr\left[\boldsymbol{\Sigma}' \boldsymbol{H}_i(\boldsymbol{\theta})\right]$. The scaling constant has received detailed attention in Durmus et al. (2017) and it is related directly to the discretization of (A.9). Specifically, Durmus et al. (2017) recommend $h = \varepsilon d^{-1/5}$ for some positive constant, ε. The optimal acceptance rate maximizing the first-order efficiency is very close to the limiting value of 0.704 predicted in Theorem 3.2 of Durmus et al. (2017). Therefore, one can calibrate the constant ε (during the burn-in phase) so that the acceptance rate is close to 0.70.

This approach has been found to perform excellently once ε and h are calibrated correctly during the burn-in phase. All derivatives are computed numerically[17] during the burn-in phase, and they are interpolated[18] in the main phase of the MCMC algorithm. This results in dramatic computational savings and, as a matter of fact, a different chain can be run in parallel in computers with multiple nodes.

Our transition density $q(\theta, \theta^c)$ is a d-dimensional Student-t distribution with five degrees of freedom. We monitor convergence using the standard diagnostics of Geweke (1992).

[17] We use the `Fortran77` subroutines in package `NDL` of Voglis et al. (2009). Specifically we use version 2.0 of Hadjidoukas et al. (2014), `https://data.mendeley.com/datasets/j2fhmszg85/1`, see also `http://cpc.cs.qub.ac.uk/summaries/AEDG_v1_0.html`

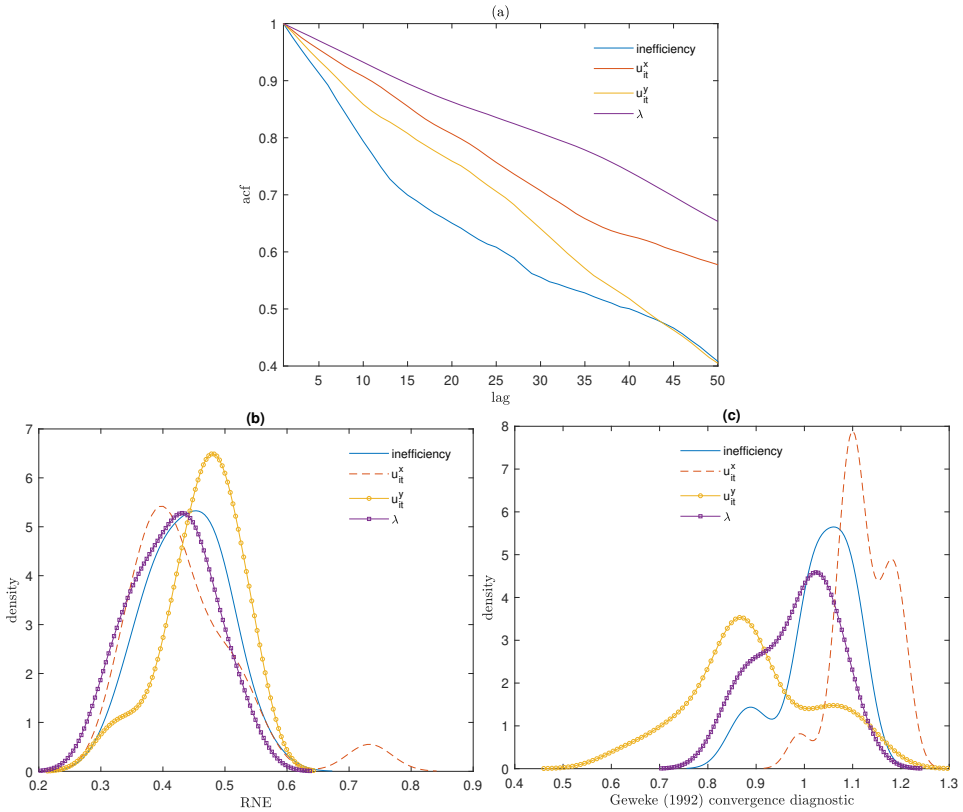
[18] We use the `Fortran` subrourines in `finterp` by Jacob Williams in `https://github.com/jacobwilliams/finterp/blob/master/README.md`. Alternatively, we use for comparison `RBF_INTERP_ND` in `https://people.sc.fsu.edu/~jburkardt/f_src/rbf_interp_nd/rbf_interp_nd.html`. `RBF_INTERP_ND` is a `Fortan90` library by John Burkardt which defines and evaluates radial basis function (RBF) interpolants to multidimensional data.

Numerical performance and convergence of MCMC

To assess numerical performance of MCMC we focus on autocorrelation in MCMC, Relative Numerical Efficiency (RNE), and Geweke's (1992) convergence diagnostic. RNE should be equal to 1 if i.i.d sampling from the posterior were possible. The convergence diagnostic is a z-statistic that compares the posterior means in the first and last 25% of MCMC draws (so if it is less 1.96 in absolute value, we cannot reject convergence). The results are reported in Figure A.5.

Figure A.5: Numerical performance of MCMC

Notes: In panel (a) we report autocorrelation functions for inefficiency, input and output slacks as well as λs. For input and output slacks as well as λs we report the maximum autocorrelation coefficient in this case. In panel (b) we report Relative Numerical Efficiency (RNE) for inefficiency, input and output slacks as well as λs. For input and output slacks as well as λs we report the minimum RNE in this case. In panel (c) we report Geweke's (1992) convergence diagnostic for inefficiency, input and output slacks as well as λs. For input and output slacks as well as λs we report the maximum absolute convergence diagnostic in this case.

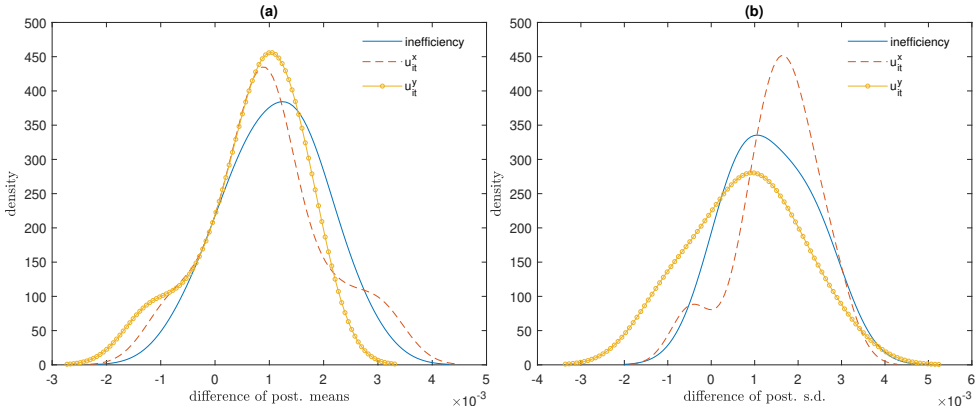


Autocorrelation functions show that MCMC has some autocorrelation but this is not destructively large so as to make problematic a thorough exploration of the posterior. RNEs

are close to 0.5 with a distinct mode near 0.7 and Geweke's (1992) convergence diagnostics indicate that convergence in MCMC chains has occurred. Posterior sensitivity analysis corresponding to 1,000 different values of q is performed and the results are shown in Figure A.6. In panel (a) we report distributions of differences of posterior means relative to the benchmark case. In panel (b) we report distributions of differences of posterior standard deviations relative to the benchmark case.

Figure A.6: Posterior sensitivity analysis

Notes: In panel (a) we report sample distributions of difference of posterior means relative to the benchmark case. In panel (b) we report sample distributions of difference of posterior standard deviations relative to the benchmark case. The distributions arise by considering 1,000 different values of q uniformly drawn in the interval (1,100).



Supplement 3. Empirical Illustration Extras

Figure A.7: Empirical results for large U.S. banks

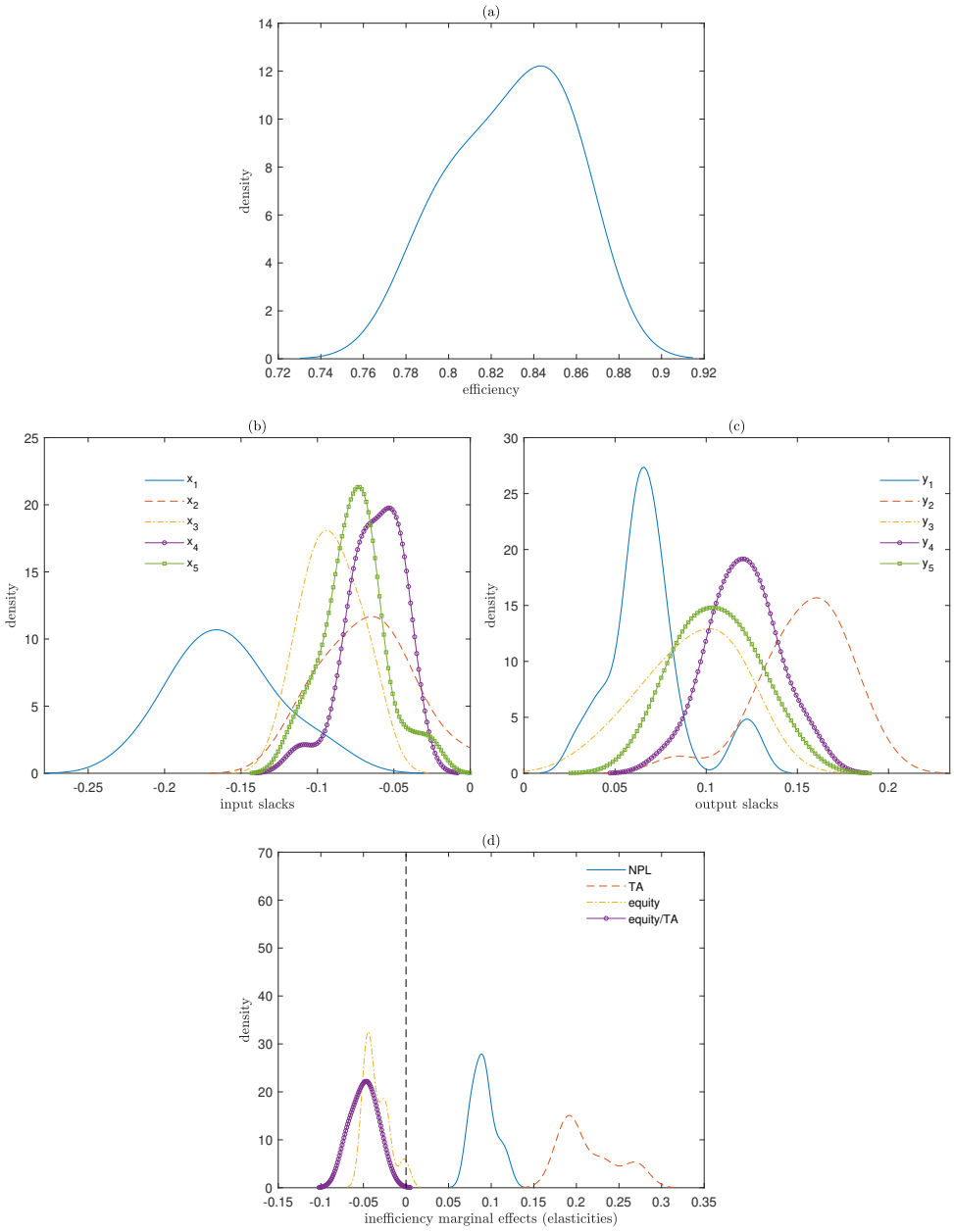


Figure A.8: CRS-DEA results