On a model of environmental performance and technology gaps

Mike G. Tsionas*

Abstract

In this paper we consider a stochastic directional technology distance function to re-examine the results of recent research in which the authors estimate a generalized directional distance function using programming methods, derive technology gaps and, in a second stage, they fit a Markov process to the technology gaps. One problem is that in the second stage efficiencies and gaps are themselves estimated. Moreover, the authors consider two groups (Annex I and non-Annex I countries according to the Kyoto protocol). We allow for endogeneity of good and bad outputs and inputs, endogenously determined groups of countries, endogenous directions for each country and group, and a distribution of technological gaps (with respect to the meta-technology) which is based on a Markov process. We use a semi-parametric directional technology distance function and we propose stochastic envelopment of different frontiers allowing for its own "meta-inefficiency". All quantities of interest are estimated jointly using numerical Bayesian techniques.

Keywords: Environment and climate change; Efficiency; Metafrontier; Technology gaps; Bayesian analysis.

Acknowledgments: The author is grateful to Konstantinos Kounetas, Panagiotis Zervopoulos for providing the data and providing important comments. Thanks are also due to three anonymous reviewers for their many useful comments on an earlier version.

*Lancaster University Management School, LA1 4YX, U.K., m.tsionas@lancaster.ac.uk

1 Introduction

Kounetas & Zervopoulos (2019) proposed the estimation of a Generalized Directional Distance Function (GDDF) model and examine the convergence-divergence hypothesis for the technology gaps defined for each period. Efficiencies and technology gaps are taken as given from the estimation of the GDDF and, in turn, they assume that gaps follow a continuous-time Markov chain which can be used to study the convergence-divergence hypothesis. One potential problem with this approach is that stochastic properties of efficiencies and gaps are not taken into account in the second stage; see Simar and Wilson (1998, 1999, 2000, 2007).

Our contributions in this paper are as follows:

i) We use a Directional Technology Distance Function (DTDF) (Atkinson, Primont and Tsionas, 2018) where directions are parameters to be estimated instead of assigning arbitrary values to them.

ii) We show how theoretical constraints can be imposed on the DTDF.

iii) We introduce dynamics for group-specific inefficiency in the form of a log AR(1) processes, and we propose a novel way to estimate the meta-technology.

iv) The distribution of technological gaps is derived directly from the Markov process of group-specific technical efficiencies and efficiency of the meta-technology. This process has been used in Kounetas & Zervopoulos (2019) but it was implemented in a second stage following DEA analysis.

v) Possible endogeneity of good and bad outputs, as well as inputs, is explicitly taken into account.

vi) Directions for inputs are outputs are estimated from the data and they are country- as well as group-specific.

We propose a stochastic frontier model which accounts explicitly for the continuous-time Markov process for technology gaps defined with respect to a meta-technology. The Markov process for technology gaps is estimated inside the model thus taking fully into account uncertainty with respect to parameters and efficiencies. Of course, we have a completely different methodological approach comparing to Kounetas and Zervopoulos (2019). The two different approaches echo in two different "worlds" in efficiency and productivity literature (parametric vs non-parametric). The standard pros and cons apply in this instance. DEA methods do not specify a functional form the the group frontiers or the meta-frontier while stochastic frontier analysis depends on parametric (yet flexible) functional forms and distributional assumptions but do account for noise in the data which standard DEA does not. In this paper, we innovate within the context of stochastic frontier analysis and stochastic meta-frontiers a la Amsler, O'Donnell, and Schmidt (2017). Relative to previous studies (Huang et al. 2015; Lin et al. 2013) our formulation allows for *stochastic* envelopment of the group stochastic frontiers so it more in line with the state-of-the-art as in Amsler, O'Donnell, and Schmidt (2017). It is also worthy of attention that we assume different technologies for an unknown number of groups, effectively, relaxes the functional form assumption and makes it a flexible formulation.

Also worthy of mention is why it is important to consider different directions for inputs and outputs among countries/groups. In a sense, directions embody characteristics of production technology that increase/decrease inputs/outputs to achieve efficiency. Therefore, if the directions are data-driven and different among countries, it seems they use different technologies. However, we allow for different technologies among countries (classified endogenously into an unknown number of groups). So, in this sense, directions are not so much aspects of the technology (although they are, clearly, tied to them) but ways to achieve efficiency for a given group-frontier. These notions cannot be disentangled in a static model but in the dynamic model we use in this study, this separation is, at least, more transparent.

2 Directions and efficiency

Suppose we have inputs $x \in \mathbb{R}_+^K$, desirable outputs $y \in \mathbb{R}_+^M$, and undesirable outputs $b \in \mathbb{R}_+^L$. The DTDF (Atkinson, Primont and Tsionas, 2018) is defined as:

$$\overrightarrow{D}(x, y, b; g) = \sup\{\vartheta | (x + \vartheta g_x 1_K,\ y + \vartheta g_y 1_M,\ b + \vartheta g_b 1_L) \in \mathcal{T}(x, y, b)\}, \tag{1}$$

where $\mathcal{T}(x, y, b)$ is the set of all feasible inputs, outputs, and bads, $g = (g_x, g_y, g_b) \in \mathbb{R}^{K+M+L}$ is the set of directions, and 1_J denotes the $J \times 1$ vector of ones.[1] We define $g = [g_1, ..., g_K, g_{K+1}, ..., g_{K+M}, g_{K+M+1}, ..., g_{K+M+L}]'$. Cheng & Zervopoulos (2014) assume g_x, g_y are positive and g_b is a negative scalar. Here, we depart from this assumption and we allow directions to be different for each netput, decision-making-unit (DMU)-specific, and allowed to vary in the real line. Therefore, we do not impose an input- or output-oriented formulation; instead, the orientation is determined by the data.

Cheng and Zervopoulos (2014, p. 901) proposed to measure efficiency as follows:

$$TE(x, y, b) = \frac{1 - \frac{1}{K}\sum_{k=1}^{K} \vartheta g_k / x_k}{1 + \frac{1}{M+L}\left(\sum_{m=1}^{M} \vartheta g_m / y_m + \sum_{l=1}^{L} \vartheta g_l / b_l\right)}, \tag{2}$$

where $i \in \{1, ..., n\}$ denotes the ith DMU. Suppose there are certain technology groups, defined by $\mathbf{\Gamma} = \{1, ..., \Gamma\}$ so that DMUs belong to one of the Γ groups. Given the meta-technology ratio (MTR) (O'Donnell,Rao & Battese, 2008), defined as:

$$0 < MTR(x, y, b) = \frac{MTE(x, y, b)}{TE(x, y, b)} = \frac{\frac{1 - \frac{1}{K}\sum_{k=1}^{K} \vartheta^* g_k / x_k}{1 + \frac{1}{M+L}\left(\sum_{m=1}^{M} \vartheta^* g_m / y_m + \sum_{l=1}^{L} \vartheta^* g_l / b_l\right)}}{\frac{1 - \frac{1}{K}\sum_{k=1}^{K} \vartheta^\gamma g_k / x_k}{1 + \frac{1}{M+L}\left(\sum_{m=1}^{M} \vartheta^\gamma g_m / y_m + \sum_{l=1}^{L} \vartheta^\gamma g_l / b_l\right)}} \leq 1, \tag{3}$$

where MTE is technical efficiency of a DMU with respect to the meta-technology, and TE represents the technical

[1] For positive directions, we could write: $\overrightarrow{D}(x, y, b; g) = \sup\{\vartheta | (x + \vartheta g_x 1_K,\ y + \vartheta g_y 1_M,\ b + \vartheta g_b 1_L) \in \mathcal{T}(x, y, b)\}$. This is a way to explain that the objective for inputs is minimization while that for outputs is maximization (even if the undesirable outputs should be minimized).

efficiency of a DMU with respect to the γ group frontier ($\gamma \in \mathbf{\Gamma}$). Moreover, ϑ^* is similar to ϑ but refers to the meta-technology (Cheng & Zervopoulos, 2014). An MTR value close to unity indicates less technology heterogeneity, and a value close to zero reveals greater technology heterogeneity. The technology gap is defined as:

$$TG(x, y, b) = TE(x, y, b) \cdot (1 - MTR(x, y, b)) = TE(x, y, b) - MTE(x, y, b). \tag{4}$$

3 Statistical model

3.1 Roadmap

Before proceeding, it is important to outline what our models is about in non-technical terms.

1. We use a family of stochastic frontiers to describe heterogeneity among countries. Heterogeneity is captured by classifying endogenously different countries into groups whose membership and number are estimated from the data.

2. Moreover, the two-sided and one-sided error components are are group and country specific. The assumption of different technologies for an unknown number of groups, effectively, relaxes the functional form assumption and makes it a flexible formulation.

3. To account for endogeneity, each DTDF is supplemented with a reduced form for the remaining endogenous variables.

4. We follow Amsler, O'Donnell & Schmidt (2017) in defining the meta-technology. This involves *stochastic* envelopment of the group-specific frontiers. The Bayesian implementation of the meta-technology and technology gaps (in point 6 below) appear to be novel to the best of my knowledge.

5. The corresponding estimate of the meta-technology can be implemented easily via simulation and for meta-technology inefficiency we assume a log AR(1) process for inefficiency.

6. To address the technology gap problem in Kounetas & Zervopoulos (2019) the technology gap is a continuous-time Markov process which is estimated endogenously or inside the model.

3.2 General

In this paper, we use a family of stochastic frontiers to describe heterogeneity:

$$\overrightarrow{D}(x_{it}, y_{it}, b_{it}; g^{\gamma}, \beta^{\gamma}) = v_{it}^{\gamma} - u_{it}^{\gamma}, i = 1, ..., n, t = 1, ..., T, \gamma \in \mathbf{\Gamma}, \tag{5}$$

where $\beta^\gamma \in \mathcal{B} \subset \mathbb{R}^p$ is a vector of unknown parameters, v_{it}^γ is a two-sided error term, and $u_{it}^\gamma \geq 0$ is another error component. Efficiency is

$$\vartheta_{it}^\gamma = e^{-u_{it}^\gamma},\ i = 1, ..., n, t = 1, ..., T, \gamma \in \mathbf{\Gamma}. \tag{6}$$

Clearly, our stochastic assumption must involve:

$$u_{it}^\gamma \in \mathbb{R}_+,\ i = 1, ..., n, t = 1, ..., T, \gamma \in \mathbf{\Gamma}. \tag{7}$$

We delay the specification of stochastic assumptions for later. The probability that unit i belongs to group γ is $p_{i\gamma}$ ($p_{i\gamma} \geq 0 \forall i, \gamma$, $\sum_{\gamma \in \Gamma} p_{i\gamma} = 1 \forall i$). To account for endogeneity of the variables in (5), each DTDF is supplemented with a reduced form for the remaining endogenous variables, viz. $\tilde{Y}_{it} = [x_{it,2}, ..., x_{it,K}, y'_{it}, b'_{it}]'$. The reduced form is as follows:

$$\tilde{Y}_{it} = W_{it}\pi^\gamma + \tilde{V}_{it}^\gamma,\ i = 1, ..., n, t = 1, ..., T,\ \gamma \in \mathbf{\Gamma}, \tag{8}$$

where π^γ is a $d \times 1$ parameter vector, $W_{it} = \begin{bmatrix} w'_{it} & & & \\ & w'_{it} & & \\ & & \ddots & \\ & & & w'_{it} \end{bmatrix}$ is a $(K - 1 + M + L) \times d$ matrix, w_{it} is a $d \times 1$ vector of instruments, and d is the number of instruments. Here, we use as instruments $x_{i,t-1,1}$, $Y_{i,t-1}$, time trend, and squares and interactions of all these variables. The functional form for the DTDF is the same as in Atkinson & Tsionas (2016) and Atkinson, Primont & Tsionas (2018). To ease notation let $z_{it} = [x'_{it}, y'_{it}, b'_{it}]'$. Specifically, we use a quadratic functional form for each group:

$$\overrightarrow{D}(z_{it}; g_i^\gamma, \beta^\gamma) = \sum_{j=1}^J \beta_j^\gamma \left(z_{it,j} + g_{i,j}^\gamma\right) + \frac{1}{2}\sum_{j=1}^J \sum_{j'=1}^J \beta_{jj'} \left(z_{it,j} + g_{i,j}^\gamma\right) \left(z_{it,j} + g_{i,j'}^\gamma\right)' = v_{it}^\gamma - u_{it}^\gamma, \\ i = 1, ..., n, t = 1, ..., T,\ \gamma \in \mathbf{\Gamma}, \tag{9}$$

where $J = K + M + L$. The restrictions that must be imposed are translation, g-homogeneity of degree minus one, concavity and monotonicity, stated in equations (3)–(6) of Atkinson, Primont & Tsionas (2018). *We wish to emphasize that the adoption of different technologies for an unknown number of groups, effectively, relaxes the functional form assumption and makes it a flexible formulation.* Moreover, we should notice that βs in (9) do not have an economic interpretation as they are, merely, coefficients in a second-order Taylor series expansion for the unknown DTDF. However, as we will see in Section 4, certain functions of them, do have an economic interpretation.

In turn, the theoretical properties imply nonlinear restrictions among β and g, stated in equation (27) of

Atkinson, Primont & Tsionas (2018). These restrictions can be stated as:

$$\mathcal{A}(g_i^\gamma)\beta^\gamma = \mathbf{0}_{(K+M+L+1)},\ i = 1, ..., n,\ \gamma \in \mathbf{\Gamma}, \tag{10}$$

where $\mathcal{A}(g)$ is an $r \times (K + M + L)$ matrix. Here, $r = K + M + L + 1$ is the number of restrictions that must be imposed to satisfy the theoretical properties of a DTDF. Atkinson, Primont & Tsionas (2018) impose these equality restrictions using the following device:

$$\mathcal{A}(g_i^\gamma)\beta^\gamma = \boldsymbol{\xi}_{(r\times 1)}, \tag{11}$$

where $\boldsymbol{\xi} \sim \mathcal{N}_r\left(-\boldsymbol{\epsilon},\ c^2\mathbf{I}\right)$, where $\mathbf{I}$ denotes the identity matrix, $c = 10^{-4}$, and $\boldsymbol{\epsilon} = [1, 0, 0, ..., 0]'$. The small variance practically enforces exactly the equality restrictions.[2] The system of equations (9), (8), and (11) can be jointly estimated using Markov Chain Monte Carlo (MCMC) provided $\Gamma = 1$. In the more general case ($\Gamma > 1$) we have a multivariate mixture-of-normals distribution. Effectively, as we will allow Γ to be determined by the data, the parametric functional form in (9) is relaxed in a flexible way.

3.3 Stochastic assumptions

To develop our stochastic assumption for u_{it}^γ, we follow Cheng & Zervopoulos (2014) and Kounetas & Zervopoulos (2019). We use (2) to define:

$$r_{it}^\gamma \equiv TE^\gamma(x_{it}, y_{it}, b_{it}; g_i^\gamma) = \frac{1 - \frac{1}{K}\sum_{k=1}^{K}\vartheta_{it}^\gamma g_{i,k}/x_{it,k}}{1 + \frac{1}{M+L}\left(\sum_{m=1}^{M}\vartheta_{it}^\gamma g_{i,m}/y_{it,m} + \sum_{l=1}^{L}\vartheta_{it}^\gamma g_{i,l}/b_{it,l}\right)}, \tag{12}$$

which is "generalized efficiency". For a given $\gamma \in \mathbf{\Gamma}$, the stochastic properties of ϑ_{it}^γ are tied to the properties of $TE^\gamma(x_{it}, y_{it}, b_{it}; g_i^\gamma, \beta^\gamma)$ which, in turn, satisfies the Markov property in (29). We assume

$$\log u_{it}^\gamma = \delta_1^\gamma + \delta_2^\gamma \log u_{i,t-1}^\gamma + w_{it}'\delta^\gamma + \varepsilon_{it}^\gamma,\ i = 1, ..., n, t = 1, ..., T,\ \gamma \in \mathbf{\Gamma}, \tag{13}$$

where w_{it} is the $d \times 1$ vector of instrumental variables defined in (8), and

$$\varepsilon_{it}^\gamma \sim \mathcal{N}\left(0,\ (\sigma_\varepsilon^\gamma)^2\right),\ i = 1, ..., n, t = 1, ..., T,\ \gamma \in \mathbf{\Gamma}. \tag{14}$$

So, in (13) we have a dynamic process of inefficiency which is, in this way, disentangled from directions g_i^γ as in (9), and (11). As we stated in the introductory, section, these notions cannot be disentangled in a static model but in the context of the dynamic model we use in this study, this separation is, at least, more transparent.

In turn, we define $\vartheta_{it}^\gamma = e^{-u_{it}^\gamma},\ i = 1, ..., n, t = 1, ..., T,\ \gamma \in \mathbf{\Gamma}$. The reason for adopting (13) is that we want

[2] We have tried smaller values like 10^{-6} and 10^{-8} but all results remained the same.

to derive transition kernels for efficiencies and technological gaps relative to the meta-technology -a construction that would be incompatible with $\delta_2^\gamma = 0$ or $\delta_2^* = 0$ in (16) below.

4 Techniques

4.1 Meta-technology

Kerstens, O'Donnell & Van de Woestyne (2019), in a seminal paper, argue that estimates of efficiency are potentially unreliable and develop a refined methodology for nonparametric envelopment of non-convex meta-sets. In an empirical application, they find that convexification assuming a convex meta-set generally leads to erroneous results. Instead, we follow Amsler, O'Donnell & Schmidt (2017) in defining the meta-technology:

$$\overrightarrow{D}_{\text{meta}}(x, y, b; g, \beta) = \mathbb{E}\left\{\max_{\gamma \in \mathbf{\Gamma}} \overrightarrow{D}(x, y, b; g^\gamma, \beta^\gamma) + v^\gamma\right\} + v_{it}^* - u_{it}^*, \tag{15}$$

where the expectation $\mathbb{E}(\cdot)$ is taken with respect to v^γ. Here, v_{it}^* is an error term in the meta-technology, and $u_{it}^* \geq 0$ is meta-efficiency. The specification in (15) appears to be novel in the sense that the *stochastic* envelopment of the different frontiers is allowed to have its own "meta-inefficiency" u_{it}^* . The corresponding estimate of the meta-technology can be implemented easily via simulation as we will see below. Moreover, for meta-technology inefficiency we assume a process similar to (13):

$$\log u_{it}^* = \delta_1^* + \delta_2^* \log u_{i,t-1}^* + w_{it}'\delta^* + \varepsilon_{it}^*,\ i = 1, ..., n, t = 1, ..., T, \tag{16}$$

where

$$\varepsilon_{it}^* \sim \mathcal{N}\left(0,\ \sigma_{\varepsilon^*}^2)\right),\ i = 1, ..., n, t = 1, ..., T. \tag{17}$$

4.2 Compact representation of the system

The system (9), (11), and (8) can be written compactly as follows:

$$\begin{bmatrix} z_{it}' & \mathbf{0} \\ \mathbf{0} & \tilde{Y}_{it} \\ [\mathcal{A}(g_i^\gamma) + \boldsymbol{\epsilon}] \otimes \iota_{\mathbf{T}} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \beta^\gamma \\ \pi^\gamma \end{bmatrix} = \begin{bmatrix} v_{it}^\gamma - u_{it}^\gamma \\ \tilde{V}_{it}^\gamma \\ \boldsymbol{\xi}_{it} \end{bmatrix},\ i = 1, ..., n, t = 1, ..., T, \tag{18}$$

where $\alpha^\gamma = \begin{bmatrix} \beta^\gamma \\ \pi^\gamma \end{bmatrix}$, and z_{it} denotes all observed data in (9), $\boldsymbol{\epsilon} = [1, 0, ..., 0]'$, and ι_T is a $T \times 1$ vector of ones. Moreover, we have the meta-technology:

$$\mathbb{E}\left\{\max_{\gamma \in \boldsymbol{\Gamma}} \overrightarrow{D}(x_{it}, y_{it}, b_{it}; g_i^\gamma, \beta^\gamma) + v^\gamma\right\} = v_{it}^* - u_{it}^*,\ i = 1, ..., n, t = 1, ..., T. \tag{19}$$

The expectation is approximated as:

$$\mathbb{E}\left\{\max_{\gamma \in \boldsymbol{\Gamma}} \left[\overrightarrow{D}(x_{it}, y_{it}, b_{it}; g_i^\gamma, \beta^\gamma) + v_{it}^\gamma\right]\right\} \simeq N^{-1} \sum_{\kappa=1}^{N} \left\{\max_{\gamma \in \boldsymbol{\Gamma}} \left[\overrightarrow{D}(x_{it}, y_{it}, b_{it}; g_i^\gamma, \beta^\gamma) + v_{it}^{\gamma,(\kappa)}\right]\right\}. \tag{20}$$

Moreover, $v_{it}^{\gamma,(\kappa)} \sim \mathcal{N}(0, \sigma_{11}^\gamma), \kappa = 1, ..., N$, and σ_{11}^γ is the first element of Σ^γ defined in (23) below. This estimator is consistent as $N \to \infty$. The equation in (19) is nonlinear in both the parameters and the data.

We write this system as:

$$\mathcal{X}_{it}(g_i^\gamma)\alpha^\gamma = \mathbf{v}_{it}^\gamma - u_{it}\boldsymbol{\iota}, \tag{21}$$

where $\mathbf{v}_{it}^\gamma = \begin{bmatrix} v_{it}^\gamma \\ \boldsymbol{\xi}_{it} \\ \tilde{V}_{it}^\gamma \end{bmatrix}$, $\boldsymbol{\iota} = [1, 0, 0, ..., 0]'$. This is supplemented with the nonlinear equation:

$$\mathbb{E}\left\{\max_{\gamma \in \boldsymbol{\Gamma}} \overrightarrow{D}(x_{it}, y_{it}, b_{it}; g_i^\gamma, \beta^\gamma) + v^\gamma\right\} \equiv \mathcal{M}(x_{it}, y_{it}, b_{it}; g_i, \beta) = v_{it}^* - u_{it}^*, i = 1, ..., n, t = 1, ..., T, \tag{22}$$

where $\beta = (\beta^\gamma, \gamma \in \boldsymbol{\Gamma})$, $g_i = (g_i^\gamma, \gamma \in \boldsymbol{\Gamma})$, $i = 1, ..., n$.

Our stochastic assumptions are:

$$V_{it}^\gamma = \begin{bmatrix} v_{it}^\gamma \\ \tilde{V}_{it}^\gamma \end{bmatrix} \sim \mathcal{N}_{K+M+L}\left(\mathbf{0}, \Sigma^\gamma\right),\ \boldsymbol{\xi}_{it} \sim \mathcal{N}(\mathbf{0}, c^2\mathbf{I}),\ c = 10^{-4},\ v_{it}^* \sim \mathcal{N}(0, \sigma_{v^*}^2), u_{it}^* \sim |\mathcal{N}(0, \sigma_{u^*}^2)|, \\ i = 1, ..., n, t = 1, ..., T,\ \gamma \in \boldsymbol{\Gamma}. \tag{23}$$

All error components in (23) are assumed to be independent. The variance of the meta-technology, $\sigma_{v^*}^2$, can be either treated as an unknown parameter or set to a small value like c^2. We believe that the metafrontier could have its own variability so, we prefer to treat $\sigma_{v^*}^2$ as an unknown parameter. The correlation of errors in DTDF (v_{it}) and the errors in reduced form ($\tilde{V}_{it}$) is essential in order to model endogeneity.

4.3 Likelihood and posterior

First, we derive the likelihood / posterior for a given observation conditional on past values of $\{u_{it}^{\gamma}\}, \{u_{it}^{*}\}$:

$$\begin{gathered}\mathscr{L}_{it}(\beta, g, \{\Sigma^{\gamma}\}, \{u_{it}^{\gamma}\}, \{u_{it}^{*}\}, \sigma_{v^*}^2, \sigma_{u^*}^2, \sigma_{\varepsilon}^2, \sigma_{\varepsilon^*}^2, \delta, p; x_{it}, y_{it}, b_{it}) \propto \\ \textstyle\sum_{\gamma\in\mathbf{\Gamma}} \{p_{\gamma}|\Omega^{\gamma}|^{-nT/2} \exp\left\{-\frac{1}{2}\left[\mathcal{X}_{it}(g_i^{\gamma})\alpha^{\gamma} + u_{it}^{\gamma}\boldsymbol{\iota}\right]' (\Omega^{\gamma})^{-1}\left[\mathcal{X}_{it}(g_i^{\gamma})\alpha^{\gamma} + u_{it}^{\gamma}\boldsymbol{\iota}\right]\right\} \cdot, \\ \sigma_{\varepsilon^{\gamma}}^{-nT/2} \exp\left\{-\frac{1}{2\sigma_{\varepsilon^{\gamma}}^2}\left(\log u_{it}^{\gamma} - \delta_1^{\gamma} - \delta_2^{\gamma}\log u_{i,t-1}^{\gamma} - w_{it}'\delta^{\gamma}\right)\right\} \cdot \\ \sigma_{\varepsilon^*}^{-nT/2} \exp\left\{-\frac{1}{2\sigma_{\varepsilon^*}^2}\left(\log u_{it}^{*} - \delta_1^{*} - \delta_2^{*}\log u_{i,t-1}^{*} - w_{it}'\delta^{*}\right)\right\}\} \cdot \\ \sigma_{v^*}^{-nT/2} \exp\left\{-\frac{1}{2\sigma_{v^*}^2}\left(\mathcal{M}(x_{it}, y_{it}, b_{it}; g_i, \beta) + u_{it}^{*}\right)\right\} \cdot \textstyle\prod_{\gamma\in\mathbf{\Gamma}} I(u_{it}^{\gamma} \geq u_{it}^{*}),\end{gathered} \tag{24}$$

where $\Omega^{\gamma} = \begin{bmatrix} \Sigma^{\gamma} & \mathbf{O} \\ \mathbf{O} & c^2\mathbf{I} \end{bmatrix}$, $\delta = \{\delta_1^{\gamma}, \delta_2^{\gamma}, \delta^{\gamma}, \delta_1^{*}, \delta_2^{*}, \delta^{*}\}$, $p = \{p_1, ..., p_{\Gamma}\}$, and $\mathbf{O}$ denotes a matrix of zeros. The initial conditions $u_{i0}^{\gamma}, u_{i0}^{*}$ are treated as parameters with a flat prior in $\mathbb{R}_+$ but we suppress them in the interest of clarity. The overall likelihood function is:

$$\begin{gathered}\mathscr{L}(\beta, g, \{\Sigma^{\gamma}\}, \{u_{it}^{\gamma}\}, \{u_{it}^{*}\}, \sigma_{v^*}^2, \sigma_{u^*}^2, \sigma_{\varepsilon}^2, \sigma_{\varepsilon^*}^2, \delta, p; \mathscr{D}) \propto \\ \textstyle\prod_{i=1}^{n}\prod_{t=1}^{T} \mathscr{L}_{it}(\beta, g, \{\Sigma^{\gamma}\}, \{u_{it}^{\gamma}\}, \{u_{it}^{*}\}, \sigma_{v^*}^2, \sigma_{u^*}^2, \sigma_{\varepsilon}^2, \sigma_{\varepsilon^*}^2, \delta, p; x_{it}, y_{it}, b_{it}),\end{gathered} \tag{25}$$

where $\mathscr{D} = \{(x_{it}, y_{it}, b_{it}), i = 1, ..., n, t = 1, ..., T\}$ denotes the data. Given a prior $\mathscr{P}(\beta, \{g_i\}_{i=1}^{n}, p, \{\Sigma^{\gamma}\})$, the posterior is provided by Bayes' theorem:

$$\begin{gathered}\mathscr{P}(\beta, g, \{\Sigma^{\gamma}\}, \sigma_{v^*}^2, \sigma_{u^*}^2, \sigma_{\varepsilon}^2, \sigma_{\varepsilon^*}^2, \delta, p|\mathscr{D}) \propto \\ \mathscr{L}(\beta, g, \{\Sigma^{\gamma}\}, \{u_{it}^{\gamma}\}, \{u_{it}^{*}\}, \sigma_{v^*}^2, \sigma_{u^*}^2, \sigma_{\varepsilon}^2, \sigma_{\varepsilon^*}^2, \delta, p; \mathscr{D}) \cdot \mathscr{P}(\beta, \{g_i\}_{i=1}^{n}, \{\Sigma^{\gamma}\}, \sigma_{v^*}^2, \sigma_{u^*}^2, \sigma_{\varepsilon}^2, \sigma_{\varepsilon^*}^2, p).\end{gathered} \tag{26}$$

Our priors are as follows:

$$\mathscr{P}(\beta, g, \{\Sigma^{\gamma}\}, \sigma_{v^*}^2, \sigma_{u^*}^2, \sigma_{\varepsilon}^2, \sigma_{\varepsilon^*}^2, \delta, p) \propto \left\{\prod_{\gamma\in\mathbf{\Gamma}} |\Sigma^{\gamma}|^{-(K+M+L+1)/2} \cdot \mathbb{I}(p^{\gamma} \in \mathscr{S})\right\} \cdot \prod_{i=1}^{n} \mathbb{I}(g_i^{\gamma} \in \mathscr{G}), \tag{27}$$

where $\mathbb{I}(\cdot)$ is the indicator function, $\mathscr{G} = \{g \in \mathbb{R}^{K+M+L}| \, g'g = 1\}$, $\mathscr{S} = \{x \in \mathbb{R}_+^C| \, \sum_{i=1}^{C} x_i = 1)$. This is a diffuse (flat) prior. For a symmetric positive definite matrix $\mathbf{\Phi}$ whose dimensionality is $N \times N$, the Jeffreys' reference prior is known to be $\mathscr{P}(\mathbf{\Phi}) \propto |\mathbf{\Phi}|^{-(N+1)/2}$. Moreover, we normalize the directions for each DMU to have unit length. This avoids fixing (as a normalization) one direction to 1 or -1, which is arbitrary.

5 Data and empirical results

5.1 General

We use the same data as in Kounetas & Zervopoulos (2019). The data set is a balanced panel consisting of 103 countries during 1995–2011, and the final panel has 1,751 observations. There is a single output variable, the inputs are capital, labor and energy, and the bad output is CO_2 emissions. In turn, we define $\vartheta_{it}^* = e^{-u_{it}^*}, \ i = 1, ..., n, t = 1, ..., T, \ \gamma \in \mathbf{\Gamma}$. Therefore, the technology gap is estimated using the difference between technologies in $\mathbf{\Gamma}$ and the meta-technology. This measure depends on the parameters of the model so, it is computed for each MCMC iteration and, in turn, it is averaged across all MCMC draws to provide a simulation-consistent estimator.

From (12), (13), and (16), given ϑ_{it}^* and $\{\vartheta_{it}^{\gamma}, \gamma \in \mathbf{\Gamma}\}$, we can obtain efficiencies r_{it}^* and $r_{it}^{\gamma}, \ \gamma \in \mathbf{\Gamma}$. In turn, technology gap is:

$$TG_{it}^{\gamma} = r_{it}^{\gamma} - r_{it}^* \, i = 1, ..., n, t = 1, ..., T, \ \gamma \in \mathbf{\Gamma}, \tag{28}$$

subject to $r_{it}^{\gamma} \geq r_{it}^* \, \forall \gamma \in \mathbf{\Gamma}$. Since $r_{it}^{\gamma}, r_{it}^*$ follow nonlinear autoregressive process, the technology gap also follows a nonlinear autoregressive process. To address the technology gap problem in Kounetas & Zervopoulos (2019) the technology gap defined in (4), say $\{\Psi(t), t \geq 0\}$ is a continuous-time Markov process. Suppose the state space is $\mathscr{E} \subseteq \mathbb{R}$. Then from the Markov property we have: $Pr\left(\Psi(t+\tau) \in \mathscr{A} | \Psi(j) = \psi, j \leq t\right) = Pr^{(\tau)}\left(\psi, \mathscr{A}\right), \ \mathscr{A} \subseteq \mathscr{E} \subseteq \mathbb{R}$. Clearly, we have:

$$f_{t+\tau}(\omega) = \int_{-\infty}^{\infty} f_{\tau}(\omega|\psi) f_t(\psi) d\psi \, \forall t, \tau \geq 0. \tag{29}$$

The joint distribution $f_{\tau}^{(s)}(\omega|\psi)$ can be obtained from the bivariate distribution of $\left(TG_{i,t_o}^{(s)}, TG_{i,t_1}^{(s)}\right)$ where $t_o = 1970$ and $t_1 = 2011$ for a particular MCMC draw $s \in \{1, ..., S\}$. The bivariate distribution is approximated using the technique of Wand (1994) given the MCMC draws. The conditional distribution is computed as:

$$f_{\tau}^{(s)}(\omega|\psi) = \frac{f^{(s)}(\omega, \psi)}{f^{(s)}(\psi)} \propto f^{(s)}(\omega, \psi), \ s = 1, ..., S. \tag{30}$$

In turn, the final estimate of the transition kernel is:

$$\hat{f}_{\tau}(\omega|\psi) = S^{-1} \sum_{s=1}^{S} f_{\tau}^{(s)}(\omega|\psi) \, \forall \omega, \psi \in \mathscr{E}, \tag{31}$$

which is consistent as $S \to \infty$. We apply MCMC to (26) for all combinations in $\Gamma \in \left\{1, ..., \overline{\Gamma}\right\}$, where $\overline{\Gamma} = 5$. To impose the prior notion of parsimony we enhance the model with the following prior[3]:

$$\mathscr{P}(\Gamma) \propto \frac{1}{\Gamma}. \tag{32}$$

The marginal likelihood is defined as:

$$\mathfrak{M}^{\gamma}(\mathscr{D}) = \int p(\theta^{\gamma}, \lambda^{\gamma}; \mathscr{D}) d\theta^{\gamma}\, d\lambda^{\gamma}\ \forall \gamma \in \mathbf{\Gamma}. \tag{33}$$

where λ denotes all latent variables and θ denotes all structural parameters. The integral is not available analytically but an unbiased estimator can be provided using Particle Filtering (see Appendix A). In turn, posterior model probabilities for each group can be estimated as:

$$\pi^{\gamma}(D) = \frac{\mathfrak{M}^{\gamma}(\mathscr{D})}{\sum_{\gamma' \in \mathbf{\Gamma}} \mathfrak{M}^{\gamma'}(\mathscr{D})}\ \forall \gamma \in \mathbf{\Gamma}. \tag{34}$$

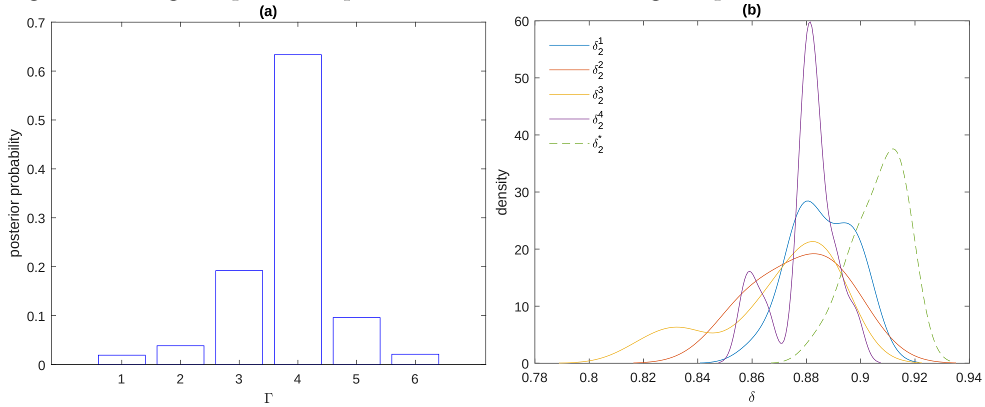
Based on the marginal or integrated likelihood of each configuration we determine $\Gamma = 4$. The posterior probabilities of Γ are presented in panel (a) of Figure 1. Although posterior model probabilities favor $\Gamma = 4$ groups, the other models have a total probability of 36.67%. As it does not seem proper to select only the results corresponding to $\Gamma = 4$, we perform Bayesian Model Averaging (BMA) for all quantities of interest, like efficiencies and technological gaps.

We also present the marginal posterior densities of δ_2^{γ} and δ_2^{*}, viz. the autoregressive parameters in (13) and (16) in panel (b) of Figure 1.These marginal posteriors show considerable persistence and they have a (dominant) mode near 0.90.

---

[3]We have examined more general priors of the form $\mathscr{P}(\Gamma) \propto (1+\Gamma)^{-(\underline{\nu}+1)}\, e^{-\underline{q}\Gamma}, \Gamma \in \{1, 2, ...\}$, where $\underline{\nu}, \underline{q} \geq 0$ using random values in the intervals $\underline{\nu} \in [0, 10]$ and $\underline{q} \in [0, 100]$ without noticing significant changes in marginal likelihoods and posterior model probabilities. The reason is that this prior favors relatively low values of Γ as well. A Poisson prior $\mathscr{P}(\Gamma) = e^{-\underline{\lambda}} \underline{\lambda}^{\Gamma} / \Gamma!$, $\Gamma \in \{1, 2, ...\}$ provided the same results for values of $\underline{\lambda}$ less than about 20.

Figure 1: Marginal posterior probabilities of Γ and marginal posterior densities of δs

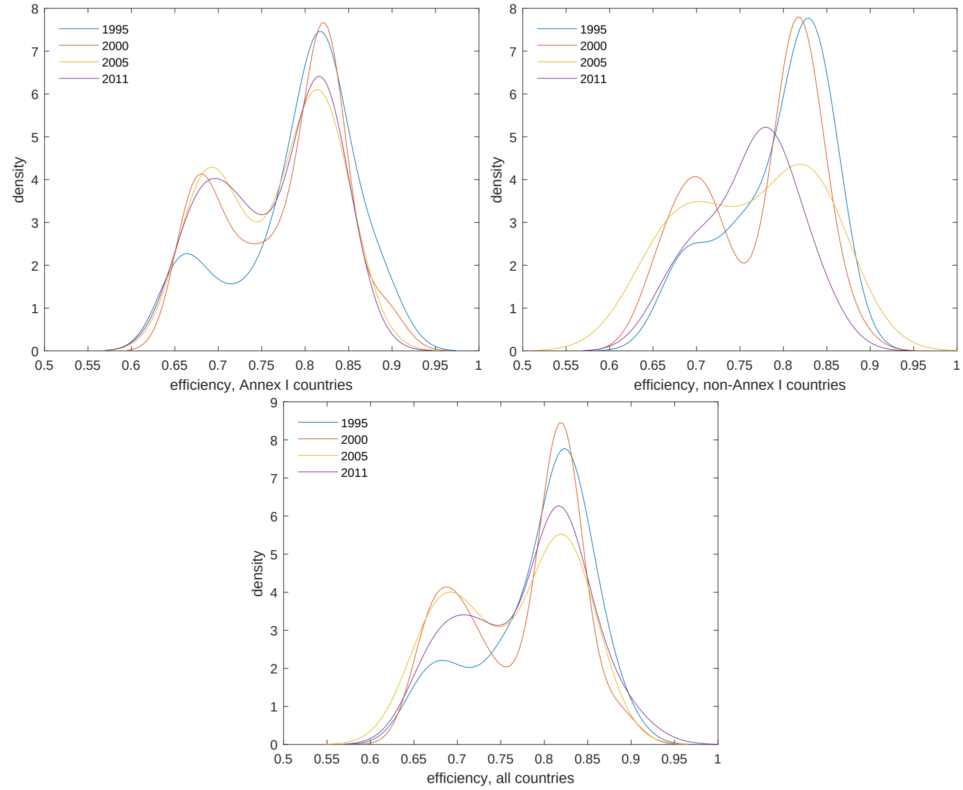


To implement MCMC, we use a Gibbs sampler for parameters and a Particle Filter (Sequential Monte Carlo) to integrate out latent inefficiencies $\{\{u_{it}^{\gamma}, \gamma \in \mathbf{\Gamma}\}, \{u_{it}^{*}\}\}$. We use 150,000 passes the first 50,000 of which are discarded to mitigate possible start up effects, and 1,000 particles per MCMC draw following the suggestion of Creal and Tsay (2015). The number of simulations (N) to obtain the expected value in (22) is set to 5,000 for each MCMC iteration. Technical details are presented in Appendix A. Parameter estimates for the GTDF are reported in the Electronic Supplement (Table B5). Tables B1 - B4 provide posterior mean estimates for efficiency and technology gaps.

5.2 Efficiency, directions and meta-technology

The results for technical efficiency are reported in Figure 2 for the years 1995, 2000, 2005 and 2011. These densities are estimated based on BMA, using the posterior model probabilities which are given in panel (b) of Figure 1. Evidently, the densities are multimodal and show certain variation over time. Notice that these densities are far from normal, showing that asymptotic-based inferences are not valid in this instance.

Figure 2: Posterior densities of technical efficiency (BMA)

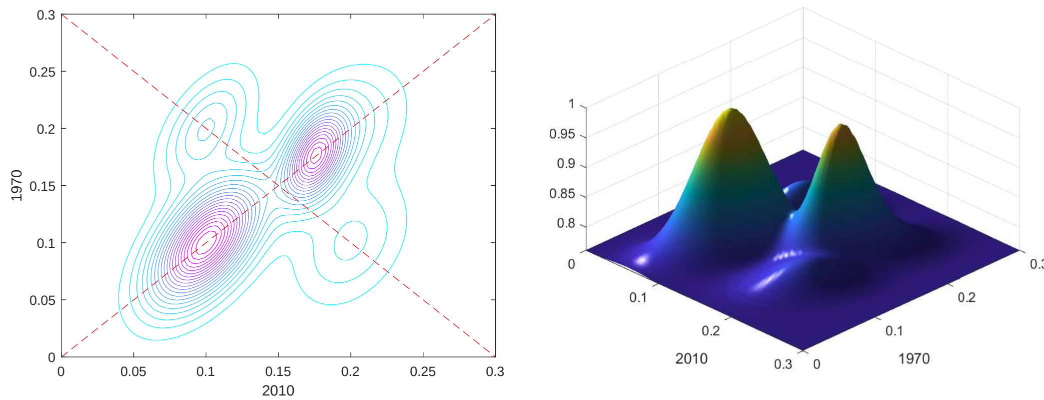


Bivariate posteriors[4] of technology gaps are shown in Figure 3 (along with the 45° line and its reflection in the left panel). The densities show four distinct modes. The modes along the 45° line indicate individual convergence clubs for all 103 countries. More specifically, there are two local maxima in the low and the high technological gaps, but not a third in the middle of relative gaps as in Kounetas & Zervopoulos (2019).[5] There are two modes on and below the 45° line and another two minor modes, located along the reflected 45° line (left panel of Figure 3); the first one is located, approximately at (0.08, 0.20) and the second near (0.2, 0.13) indicating a decrease and increase, respectively, of technology gaps between 1970 and 2010. The dominant modes are located near (0.08, 0.12) and (0.17, 0.19) respectively. Compared to Figures 5 and 6 in Kounetas and Zervopoulos (2019) technology gaps are much smaller, owing to the presence of noise in the meta-technology (19).

[4]All bivariate posteriors are normalized to one in the dominant mode.

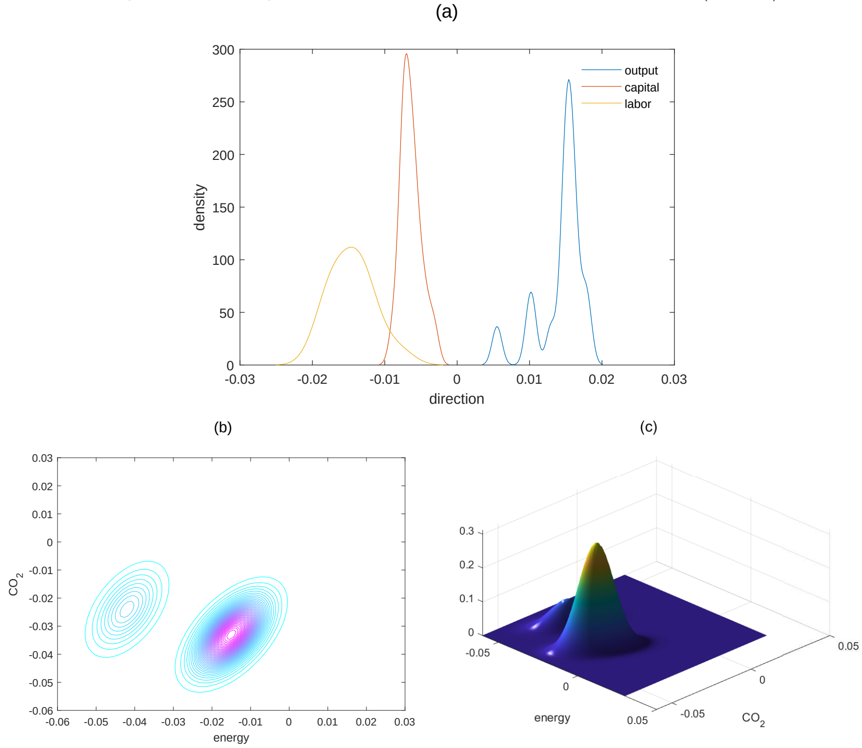
[5]Of course, some differences may be attributed to the fact that these authors used a GGDF whereas we use a DTDF, although our inefficiency measures are in the spirit of Kounetas & Zervopoulos (2019). An anonymous referee rightly argues that the direction selected in a stochastic directional distance function will affect the functional estimates because deviations from the estimated function are minimized in the specified direction (Layer et al. 2020; Kuosmanen and Johnson, 2017). Moreover, it changes the technical efficiency scores (Layer et al. 2020). In this paper the selected direction corresponds to an interpretation of the inefficiency measure, based on the distance to the economically efficient point. Of course this is correct but on the positive side, directions are estimated from the data and, in this sense, efficiency estimation and direction estimation are disentangled.

Figure 3: Bivariate posteriors of technology gaps (BMA)



In panel (a) of Figure 4, we report marginal posterior densities of directions for output, capital, and labor across countries for the particular group $\gamma \in \mathbf{\Gamma}$ to which they belong. The output directions are positive, while the directions for capital and labor take negative values for the most part. The marginal posterior of the capital directions is also negative, for the most part, although its spread is much smaller compared to other directions, showing that adjustments for this input are rather small (as it is quite expensive to decrease capital in the short-run). The notable feature of these posteriors is that *the data are quite informative about input- and output-orientation and, therefore, fixing the directions in advance is not necessary, even when precise prior information about them is not imposed*.

Figure 4: Marginal posterior densities of directions (BMA)



In panels (b) and (c) of Figure 4, we report posterior contours and the (normalized) posterior density of energy and CO_2 directions. The joint distribution is bimodal with the two modes located in negative values. The minor mode is located also at negative values but shows more flexibility in changing energy and emissions. This result shows that, in the posterior, energy and CO_2 directions are positively correlated, there are two different groups in the data in terms of directions in the energy-CO_2 emissions space and the two modes are, practically, totally separated. Two important questions are: If each country is allowed to choose inputs/outputs directions, do they tend to choose "easy" ways to achieve short-term efficiency and avoid employing optimal technology for long-term perspectives? How we can consider or differentiate such directional differences in efficiency evaluation? As our directions are normalized to have unit norm, we can back out directions in inputs and good / bad outputs in original units, as a percentage of corresponding inputs / outputs. Posterior means and standard deviations are reported in Table 1.

Table 1: Directions

	capital	labor	energy	CO_2	output
Annex I countries					
1995-2005	-0.51%	-3.14%	-2.57%	-0.32%	0.32%
	(0.022)	(0.014)	(0.013)	(0.010)	(0.021)
2006-2011	-0.48%	-3.10%	-5.34%	-1.32%	0.45%
	(0.018)	(0.013)	(0.013)	(0.032)	(0.019)
Non-Annex I countries					
1995-2005	-0.44%	-3.32%	-2.33%	-0.21%	0.43%
	(0.013)	(0.019)	(0.012)	(0.013)	(0.031)
2006-2011	-0.52%	-3.20%	-2.44%	-0.19%	0.49%
	(0.016)	(0.011)	(0.012)	(0.017)	(0.029)

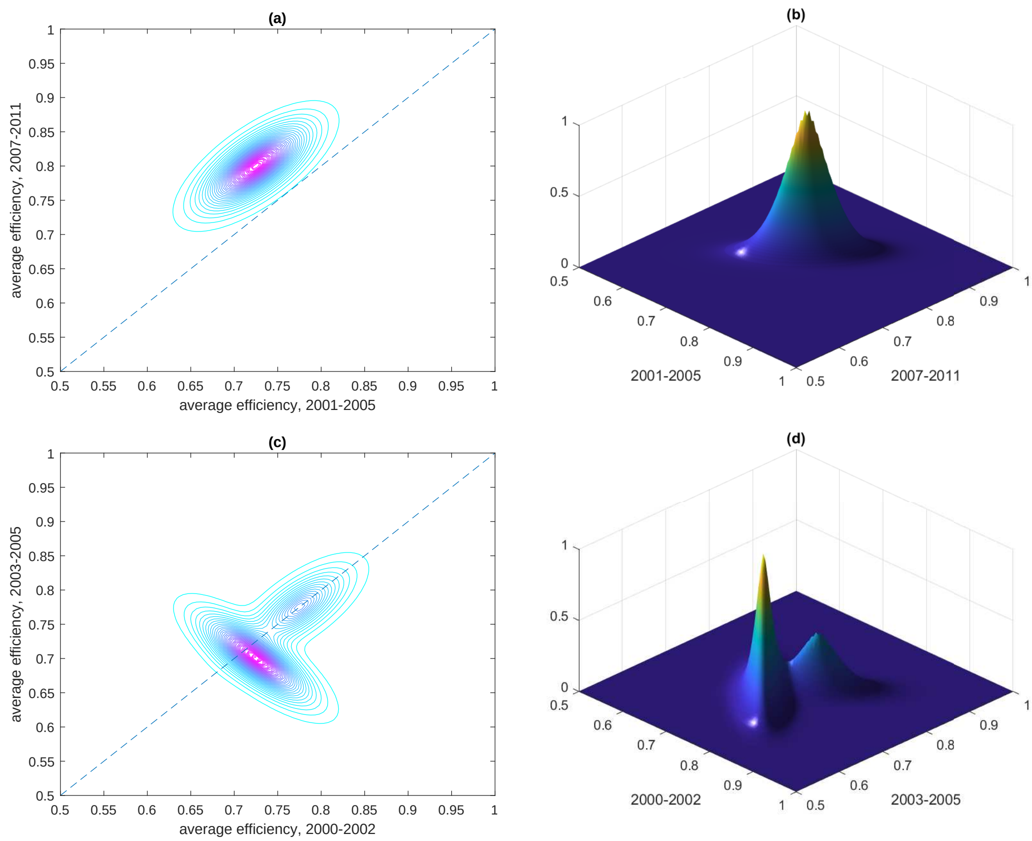
Notes: Directions are in percentage terms relative to inputs / (good and bad) outputs. Posterior standard deviations appear in parentheses and are also percentages.

From Table 1, it appears that Annex I and non-Annex I countries are quite similar in terms of output, labor, capital and energy. One important difference is that between the two periods (1995-2005 and 2006-2011) there has been more reduction of energy (from 2.57% to 5.34%, relative to the levels of this input) as well as CO_2 (from 0.32% to 1.32%, relative to the levels of this input. Labor is an input that can be adjusted downwards more easily compared to capital; relative to energy we notice no difference between Annex I and non-Annex I countries except for the period 2006-2011 for Annex I countries. From these results it turns out that: i) Annex I countries have been reducing CO_2 at the rate of 1.32% per year in the period 2006-2011 (although, clearly, they did not have to based on the obligations of Kyoto protocol), ii) both groups (Annex I and non-Annex I countries) tend to choose "easy" ways to achieve short-term efficiency in the sense that labor and energy are more easily adjusted compared to other inputs and outputs. Another way of examining the same question is the following evidence. First, the sample densities of (posterior mean) meta-efficiencies are rather similar across periods (see Figure 2). Second, marginal posterior densities of the autoregressive parameters of the AR(1) for meta-inefficiency (equation (16)) indicate considerable persistence. Therefore, on the one hand efficiency is rather "sticky" but so are capital, CO_2 (for the most part except Annex-I countries after 2005 or 2006) and output and, additionally, meta-inefficiency is rather persistent. In this sense, it is clear that efficiency relative to the meta-technology is, at best, something that countries leave for the long-run.

Regressions (17)–(24) in Kounetas & Zervopoulos (2019) could be, potentially, a source of problems as efficiencies are themselves estimated. The purpose of their regressions is to examine whether the improvement of environmental efficiency of Annex I countries from 2006 until 2011 reflects the positive effect of the adoption of the Kyoto Protocol (signed in 2006 by Annex I countries). To examine the possible improvement of environmental efficiency, for each MCMC draw, we average environmental efficiencies for 2007–2011 and 2001–2005, that is five years prior to and after 2006. The averaging is performed over all Annex I countries and 2001–2005, 2007–2011. Effectively, we have $S = 100,000$ MCMC draws for these two quantities.[6]

[6] An anonymous reviewer correctly points out that this study separates periods between 2001-2005 and 2007-2011 due to the adoption of the Kyoto Protocol (signed in 2006 by Annex I countries) (page 13), but the Kyoto Protocol was adopted in 1997 and entered into force in 2005 for the first commitment period (2008-2012). The choice for 2005 is motivated by Kounetas and Zervopoulos (2019), and

Figure 5: Posterior kernels of environmental efficiencies (BMA)



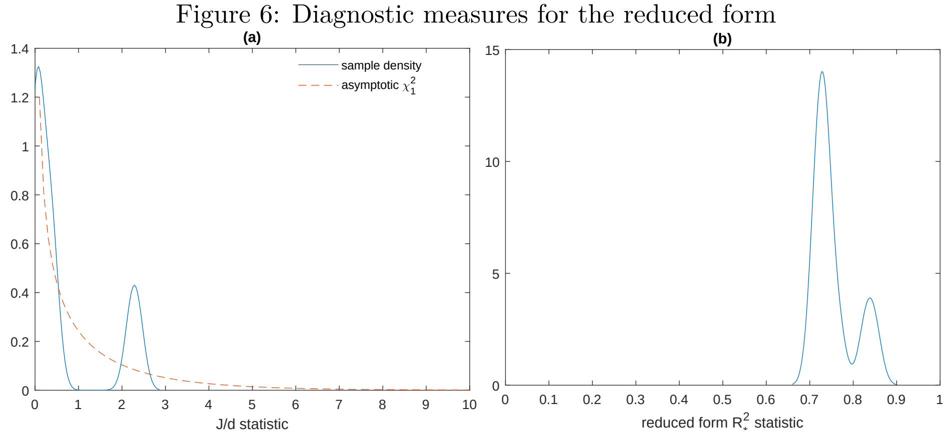
Contours of the bivariate posteriors of environmental efficiencies for 2001–2005 versus 2007–2011, are presented in panel (a) of Figure 5. The mode of the distribution indicates an improvement from nearly 0.72 to 0.82, and the contours lie above the 45° line, indicating that the Kyoto protocol had a significant positive effect in environmental efficiencies averaging, approximately, ten percentage points. In panel (b) we report the corresponding (normalized) density. In panel (c) we report the contours of average efficiencies of 2000–2002 versus 2003–2005, viz. prior to the Kyoto protocol. In panel (d) we present the bivariate posterior density for 2000–2002 versus 2003–2005. The density is clearly bimodal and the two modes lie along the 45° line, showing no convergence of environmental efficiencies prior to the Kyoto protocol. The dominant mode is located at nearly 0.70 and the other mode at 0.80, showing the persistence of environmental efficiency at relatively low levels. From panel (b) it is evident that the distribution is unimodal so convergence has taken place after the Kyoto protocol in Annex I countries.

particularly their definition of dummy variable D, preceding their equations (20) and (21). Although 2008 or 2009 could have been adopted for the second period instead of 2005, we use 2005 as in Kounetas and Zervopoulos (2019), equations (21) – (24). Moreover, Kounetas and Zervopoulos (2019, p. 1143) write: "The improvement of the environmental efficiency of Annex I countries from 2006 until 2011 reflects the positive effect of the adoption of the Kyoto Protocol. For instance, Annex I countries presented a reduction in CO2 emissions by 0.61% on average for the period 2006–2011." Of course, 2006 could have been the "critical" year per their equation (24).

5.3 Instrument validity

We address another important issue, viz. the validity of instruments in (8).[7] We proceed in the standard fashion of developing a Sargan-Hansen J-statistic for the validity of over-identifying restrictions (related to the instruments in w_{it}) resulting from (8) in the context of Generalized Method of Moments (GMM). We present the details of formulating this statistic in the Technical Appendix. The J-statistic follows, asymptotically, the *chi*-square distribution, χ^2_d , where d is the number of over-identifying restrictions. In turn $J/d \sim \chi^2_1$. The J-statistic is computed for each MCMC draw so, finally, we can recover its (finite sample) distribution and compare it to the asymptotic distribution or simply to its critical value at the 5% level of significance (which is, approximately, 3.84). The distributions are presented in panel (a) of Figure 6. In panel (b) we deal with another important issue, viz. that the instruments should be correlated with the endogenous variables. This reduces to the question of finding an R^2-like measure for the reduced form. This measure, which we denote by R^2_*, is derived as the system-R^2 (Buse, 1973) of the reduced form (for each MCMC replication) and its density is presented in panel (b) of Figure 6.

Figure 6: Diagnostic measures for the reduced form



Interestingly, the finite sample distribution of J/d-statistic is bimodal with the dominant mode near 0.5 and another mode near 2.5. In comparison to the asymptotic χ^2_1 distribution, it is clear that the the finite sample distribution is well within the bounds specified by the asymptotic distribution. In this sense, the reduced form is valid in the sense that we do not have evidence against the use of suggested instruments. In panel (b) of Figure 6, we present the (finite sample) density of R^2_*. This measure is between 0.7 and 0.9 so we do not have evidence against the hypothesis that the fit of the reduced form is satisfactory.

5.4 Policy implications

As we have seen, countries have a low probability of changing their relative position within a year in terms of technological gaps. So, mobility is rather limited during the seventeen year period we examine here. This is

[7]The author is grateful to an anonymous reviewer for pointing this out.

consistent with the findings of Kounetas and Zervopoulos (2019, p. 1145–1146). It is, of course, quite possible that this persistence as well as country-specific and meta-technology efficiency or directions are influenced by industrial compositions, and institutional or regulatory differences on undesirable outputs for each country. The vast amount of heterogeneity across countries evidenced here, besides even club formation and meta-technology, is a strong indication that this may be true. Moreover, as argued in Kounetas and Zervopoulos (2019, p. 1145–1146) "[t]his could be explained in terms of factor accumulation deformations, factor price changes that induce the introduction of new technologies [...] and localized technological change [...] Factor accumulation distortions [...] of the examined countries in both physical and human capital terms could be important to facilitate the objective of creating the three clubs. For instance, physical capital investment may embody new energy-saving technologies to help in catching up with the frontier, but this is not the case for all countries". Mainly due to data limitations, incorporation of these variables into our instrumental variables is, presently, impossible. However, it is clear that this is the background into which the results should be interpreted in broad terms.

The important difference of clusters, in terms of technological gaps and other things (efficiency, etc.), could depend on differences in country-specific technologies and / or on spillovers from the meta-frontier production function (Kounetas and Zervopoulos, 2019, p. 1146) so the investigation of differences among country-specific technologies and the meta-technology deserve further attention, especially in the lack of data regarding the "background" variables in this study.

The fact that, in this study, we find different cluster allocations is not surprising in its own right, as cluster identification is very much dependent on the model. There are cluster techniques that can be applied "universally" without reference to a given model, and there are model-specific cluster techniques, as the one in this paper. Both approaches have their advantages and disadvantages. A universal technique does not depend on any particularly model, which is good, but on the other hand it can miss important features of the data. For example, if the data satisfy a linear regression model like $y_i = \alpha + \beta x_i + u_i$ any universal clustering technique will find several clusters in the data (x_i, y_i). A model-specific technique will find just one cluster but the model itself is open to criticism. This criticism is healthy in the sense that it only through this criticism that models can be improved to account for other features of the data that were not originally intended to be accounted for. Choosing between the model in this study and the model proposed by Kounetas and Zervopoulos (2019) is not easy and we leave the matter to be decided in future research. In fact, choosing more broadly between a DEA model and a stochastic frontier model is an important issue that is not addressed in the literature, although many comparisons have been made. As usual, both approaches have their strong and weak aspects and, one could argue that combining the two should yield better inferences about efficiency, productivity, etc. How to combine remains, of course, an open question that we do not address in this paper as it is beyond its scope. In a similar vein, it is currently ambiguous whether we can formally compare and contrast cluster allocations delivered by different techniques, which is why we did not go along this path in this study. Since we have an econometric model, we account for noise which is not captured by DEA-like techniques. On the other hand, we have parametric (although flexible) assumptions about the data generating mechanism of the data. Any (formal) comparison between cluster allocations should credit each procedure according to its own merits and discount it based on its weaknesses. Currently, we simply do not know how to do this in a satisfactory way but we do suggest that it is an important avenue for future research.

In terms of policy, capital investment that embodies energy-saving and cleaner technologies may help in faster convergence to the meta-technology and, at the same time, increase environmental efficiency. However, this depends, again, on the country-specific background which is unobserved but evidenced in heterogeneity across clubs and countries. The policy implications of this study, *conditionally* on the technologies we observe (which may not be neither energy-saving or cleaner) energy-saving and CO_2 reductions by other means, appears to be the only practical way. This is especially so in view of the fact that what is documented in this study, as well as in Kounetas and Zervopoulos (2019) is that there are differences of technology, directions, technology gaps, etc. These differences can be explained by the background for each country and club, but in the absence of data regarding this background, there is not much more we can say except that changes in this background should be implemented via institutional means in the absence of wide access to markets for CO2 emissions. Of course, factor accumulation deformations, factor price changes that induce the introduction of new technologies etc., have a role to play but such variables or explanations are endogenous to the problem as they depend on markets for factors which, again, is background in terms of estimating production frontiers. One feature of such markets could be their price instability, and dependence on political decisions or negotiations among suppliers (Grubb, 2003). An important insight is that "much as the oil markets involve a high degree of government-industry interaction (though now somewhat less than formerly), the Kyoto system is likely to involve the same. Some governments at least wish to protect and support emergent industries that can deliver, and profit from, lower carbon futures" (Grubb, 2003, p. 174).

Clearly, technical change and therefore technology gaps can be affected by the Kyoto protocol bundle of measures, in the sense that the Kyoto targets are costly , and moreover subject to a "re-configuration" of emissions if some industries migrate to countries without emission caps (Barrett, 2001). As there is vast heterogeneity in terms of industrial composition among countries, one does not expect a uniform reduction in CO_2 emissions due to the Kyoto protocol. For example, in Germany and Japan, emissions were reduced even before the Kyoto protocol mostly based on government policies, regulations, and changes in public opinion (Wang et al., 2019). On the other hand, the CO_2 emission patterns in Italy show that emissions tend to be high in economies dominated by heavy industry. On the other hand, in Taiwan, which is neither an Annex I or non-Annex I country, the predominance of export-oriented manufacturing industries caused a sharp increase in CO_2 emissions following economic development (*op. cit*). India, is an economy whose industrial composition is, mainly, agriculture and software design, thus associated with low emissions, confirms that the structure—rather than the development stage and income—may be responsible for its carbon emissions. Therefore, "[p]rudent reduction policies and low-carbon development can effectively reduce CO_2 emissions. Thus, the Kyoto Protocol can promote reduction of CO_2 emissions in Annex I countries" (*op cit.*, p.

17).

From this corroborating evidence, it seems certain that industrial composition is an essential determinant of emissions when heavy industry is predominant and requires high-emissions technology. As a result, enforcement of reductions is likely to decrease income until cleaner technologies become available so, in the short-run there is an important trade-off between emissions and output (per capita or otherwise). In this perspective, the rising trend of emissions in Italy, for example, would partially at, least, revert if the Italian nuclear program had been implemented. This program stopped right after the global financial crisis of 2018-19. This indicates that global economic conditions are an important element of the “background” becomes, except institutional changes, legislation, etc., mainly industrial composition and the need for each country to maintain standard of living and GDP growth. As a general, but certainly not universal conclusion one can say that the Kyoto protocol and, especially, the Paris Agreement (2015), *can* promote reduction of CO_2 emissions at least in Annex I countries but given *slowly* changing industrial composition, based on comparative advantage in international trade, for *most* countries, whether Annex I or non-Annex I.

Concluding remarks

In this paper, we resolve a technical issue in Kounetas & Zervopoulos (2019) where a Markov chain for technological gaps is estimated in a second stage, following preliminary estimation of a generalized directional distance function. We embed a Markov process in a new model (based on a flexible directional technology distance function) to avoid two–stage estimations. We provide several methodological innovations: We estimate a data-based group-specific directional technology distance function. We provide a novel way of modeling and estimating the meta-technology. We allow for endogeneity of all inputs and good and bad outputs. Directions are estimated from the data and they are allowed to be *a priori* different for each country and each group, which allows us to adopt a broader view regarding the notion of directions in a distance function. The proposed approach relies on Bayesian modeling organized around Markov Chain Monte Carlo methods, and especially the Sequential Monte Carlo – Gibbs algorithm. Relative to Kounetas & Zervopoulos (2019) we find that the 1970–2011 transition kernel for technological gaps has four (instead of three) modes, Annex I countries converge in terms of environmental efficiency after the Kyoto protocol but not before, and that the transition kernel before the Kyoto agreement (2009–2005 and 2007–2011) is bimodal indicating club-convergence around sub-optimal environmental performance levels. Instead, relative to prior the Kyoto agreement, post-Kyoto environmental performance is enhanced.

Appendix A.

Our MCMC is based on particle filtering to integrate out latent inefficiencies from the joint posterior and a

Gibbs sampler to integrate the other parameters. Specifically, we use a recent advance in Sequential Monte Carlo methods known as the particle Gibbs (PG) sampler, see Andrieu et al. (2010). The algorithm allows us to draw paths of the state variables in large blocks. Particle filtering is a simulation based algorithm that sequentially approximates continuous, marginal distributions using discrete distributions. This is performed by using a set of support points called "particles" and probability masses; see Creal (2012) for a review.

The PG sampler draws a single path of the latent or state variables from this discrete approximation. As the number of particles [8] M goes to infinity, the PG sampler draws from the exact full conditional distribution. As mentioned in Creal and Tsay (2015, p. 339): "The PG sampler is a standard Gibbs sampler but defined on an extended probability space that includes all the random variables that are generated by a particle filter. Implementation of the PG sampler is different than a standard particle filter due to the "conditional" resampling algorithm used in the last step. Specifically, in order for draws from the particle filter to be a valid Markov transition kernel on the extended probability space, Andrieu et al. (2010) note that there must be positive probability of sampling the existing path of the state variables that were drawn at the previous iteration. The pre-existing path must survive the resampling steps of the particle filter. The conditional resampling step within the algorithm forces this path to be resampled at least once. We use the conditional multinomial resampling algorithm from Andrieu et al. (2010), although other resampling algorithms exist, see Chopin and Singh (2013)."

We follow Creal and Tsay (2015). Suppose the posterior is $p(\theta, \lambda_{1:T}|\mathbf{y}_{1:T})$ where

$$\lambda_{1:T} = \left[\{u^{\gamma}_{i,1:T}, \gamma \in \mathbf{\Gamma}, i = 1, ..., n\}, \{u^{*}_{i,1:T}, i = 1, ..., n\}\right],$$

denotes the latent variables whose prior can be described by $p(\lambda_t|\lambda_{t-1}, \theta)$ given by (13) and (16). We maintain throughout the constraint

$$u^{*}_{it} \leq \min_{\gamma \in \mathbf{\Gamma}} u^{\gamma}_{it}, \tag{1}$$

by rejection sampling. In the PG sampler we can draw the structural parameters $\theta|\lambda_{1:T}, \mathbf{y}_{1:T}$ as usual, from their posterior conditional distributions. This is important because, in this way, we can avoid mixture approximations or other Monte Carlo procedures that need considerable tuning and may not have good convergence properties. As such posterior conditional distributions we omit the details and focus on drawing the latent variables.

Suppose we have $\lambda^{(1)}_{1:T}$ from the previous iteration. The particle filtering procedure consists of two phases.

Phase I: Forward filtering (Andrieu et al., 2010).

- Draw a proposal $\lambda^{(m)}_{it}$ from an importance density $q(\lambda_{it}|\lambda^{(m)}_{i,t-1}, \theta), m = 2, ..., M$.

[8] M is not to be confused with number of outputs (which is one) in main text.

- Compute the importance weights:

$$\omega_{it}^{(m)} = \frac{p(y_{it}; \lambda_{it}^{(m)}, \theta) p(\lambda_{it}^{(m)} | \lambda_{i,t-1}^{(m)}, \theta)}{q(\lambda_{it} | \lambda_{i,t-1}^{(m)}, \theta)}, m = 1, ..., M. \quad (2)$$

- Normalize the weights: $\tilde{\omega}_{it}^{(m)} = \frac{\omega_{it}^{(m)}}{\sum_{m'=1}^{M} \omega_{it}^{(m')}}, m = 1, ..., M.$
- Resample the particles $\{\lambda_{it}^{(m)}, m = 1, ..., M\}$ with probabilities $\{\tilde{\omega}_{it}^{(m)}, m = 1, ..., M\}$.

In the original PG sampler, the particles are stored for $t = 1, ..., T$ and a single trajectory is sampled using the probabilities from the last iteration. An improvement upon the original PG sampler was proposed by Whiteley (2010), who suggested drawing the path of the latent variables from the particle approximation using the backwards sampling algorithm of Godsill et al. (2004). In the forwards pass, we store the normalized weights and particles and we draw a path of the latent variables as we detail below (the draws are from a discrete distribution).

Phase II: Backward filtering (Chopin and Singh, 2013, Godsill et al., 2004).

- At time $t = T$ draw a particle $\lambda_{iT}^{*} = \lambda_{iT}^{(m)}$.
- Compute the backward weights: $\omega_{t|T}^{(m)} \propto \tilde{w}_{t}^{(m)} p(\lambda_{i,t+1}^{*} | \lambda_{it}^{(m)}, \theta)$.
- Normalize the weights: $\tilde{\omega}_{t|T}^{(m)} = \frac{\omega_{t|T}^{(m)}}{\sum_{m'=1}^{M} w_{t|T}^{(m')}}, m = 1, ..., M.$
- Draw a particle $\lambda_{it}^{*} = \lambda_{it}^{(m)}$ with probability $\tilde{\omega}_{t|T}^{(m)}$.

Therefore, $\lambda_{i,1:T}^{*} = \{\lambda_{i1}^{*}, ..., \lambda_{iT}^{*}\}$ is a draw from the full conditional distribution. The backwards step often results in dramatic improvements in computational efficiency. For example, Creal and Tsay (2015) find that $M = 100$ particles is enough. There remains the problem of selecting an importance density $q(\lambda_{it} | \lambda_{i,t-1}, \theta)$. We use an importance density implicitly defined by $\lambda_{it} = a_{it} + \sum_{p=1}^{P} b_{it} \lambda_{i,t-1}^{p} + h_{it} \xi_{it}$ where ξ_{it} follows a standard (zero location and unit scale) Student-t distribution with $\nu = 5$ degrees of freedom. That is, we use polynomials in $\lambda_{i,t-1}$ of order P. We select the parameters a_{it}, b_{it} and h_{it} during the burn-in phase (using $P = 1$ and $P = 2$) so that the weights $\{\tilde{w}_{it}^{(m)}, m = 1, ..., M\}$ and $\{\tilde{\omega}_{t|T}^{(m)}, m = 1, ..., M\}$ are approximately not too far from a uniform distribution.

Chopin and Singh (2013) have analyzed the theoretical properties of the PG sampler, and proved that the sampler is uniformly ergodic. They also prove that the PG sampler with backwards sampling strictly dominates the original PG sampler in terms of asymptotic efficiency.

Alternatively, when the dimension of the state vector is large, we can draw $\lambda_{i,1:T}$, conditional on all other paths $\lambda_{-i,1:T}$ that are not path i. Therefore, we can draw from the full conditional distribution $p(\lambda_{i,1:T} | \lambda_{-i,1:T}, \mathbf{y}_{1:T}, \theta)$.

Finally, for purposes of identification we follow Geweke (2007) and set

$$p^1 < p^2 < ... < p^\Gamma. \tag{3}$$

The performance of MCMC can be assessed using relative numerical efficiency (RNE) and Geweke's (1992) convergence diagnostic (CD). RNE is also proposed by Geweke (1992) and shows how close to i.i.d is MCMC from the posterior. Values closer to one indicate that MCMC performs like i.i.d sampling. The CD is asymptotically normal and values less than two in absolute value, indicate successful convergence. Our results are reported in Table A1. From the results, it turns out that RNE is acceptably large and CD indicates convergence of MCMC.

Table 1: MCMC diagnostics

	RNE	|CD|
θ	0.414	1.17
$\{u_{it}^*\}$	0.382	1.26
$\{u_{i0}^*\}$	0.615	1.33
$\{u_{it}^\gamma,\ \gamma \in \mathbf{\Gamma}\}$	0.402	1.36
$\{u_{i0}^\gamma,\ \gamma \in \mathbf{\Gamma}\}$	0.617	1.44
technology gaps	0.515	1.23

Notes: Structural parameters are denoted by θ, $\{u_{it}^*\}$ is meta-technology inefficiency, $\{u_{it}^\gamma,\ \gamma \in \mathbf{\Gamma}\}$ is group-specific inefficiencies, and technology gaps are defined in (28). Moreover, $\{u_{i0}^*\}$ and $\{u_{i0}^\gamma,\ \gamma \in \mathbf{\Gamma}\}$ are initial conditions for meta-technology and group-specific inefficiencies, respectively, see (13) and (16). For all quantities, RNE and |CD| are maxima across DMUs, time periods, groups (where appropriate) and parameters indicated in the Table.

References

Albrecht, J.; François, D.; Schoors, K. A Shapley (2002). Decomposition of Carbon Emissions without Residuals. Energy Policy 30, 727–736.

Amsler, C., O'Donnell, C. J. & Schmidt, P., 2017. Stochastic metafrontiers. Econometric Reviews, 36 (6–9), 1007–1020.

Andrieu, C., Doucet, A., Holenstein, R., 2010. Particle Markov chain Monte Carlo methods (with discussion). Journal of the Royal Statistical Society, Series B 72 (2), 1–33.

Atkinson, S. E., & Tsionas, M. G. (2016). Directional distance functions: Optimal endogenous directions. Journal of Econometrics, 190 (2) 301–314.

Atkinson, S. E., Primont, D., & M. G. Tsionas (2018). Statistical inference in efficient production with bad inputs and outputs using latent prices and optimal directions. Journal of Econometrics, 204 (2), 131–146.

Barrett, S. (2001), 'Towards a better climate treaty', Policy Matters, 01-29. Washington, DC: AEI/Brookings Joint Center for Regulatory Studies. November. Reprinted in World Economics, 3(2), pp. 35–45

Buse, Adolf (1973). Goodness of Fit in Generalized Least Squares Estimation, The American Statistician 27

(1973), 106–108.

Cheng, G. , & Zervopoulos, P. D. (2014). Estimating the technical efficiency of health care systems: a cross-country comparison using the directional distance function. European Journal of Operational Research, 238 (3), 899–910.

Chopin, N., Singh, S.S., 2013. On the particle Gibbs sampler. Working paper, ENSAE. http://arxiv.org/abs/1304.1887.

Creal, D.D., 2012. A survey of sequential Monte Carlo methods for economics and finance. Econometric Rev. 31 (3), 245–296.

Creal, D., & R. Tsay (2015). High dimensional dynamic stochastic copula models. Journal of Econometrics 189 (2), 335–345.

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

Geweke, J. (2007). Interpretation and inference in mixture models: Simple MCMC works. Computational Statistics & Data Analysis 51(7), 3529–3550.

Godsill, S.J., Doucet, A., & West, M., 2004. Monte Carlo smoothing for nonlinear time series. Journal of the American Statistical Association, 99 (465), 156–168.

Grubb, M. (2003). The Economics of the Kyoto Protocol. World Economics 4 (3), 143–189.

Huang, T. H., Chiang, D. L., & Tsai, C. M. (2015). Applying the new metafrontier directional distance function to compare banking efficiencies in Central and Eastern European countries. Economic Modelling 44, 188–199.

Kerstens, K., C. O'Donnell & I. Van de Woestyne (2019). Metatechnology frontier and convexity: A restatement. European Journal of Operational Research 275 (2), 780–792.

Kounetas, K., & Zervopoulos, P. D. (2014). A cross-country evaluation of environmental performance: Is there a convergence-divergence pattern in technology gaps? European Journal of Operational Research, 273, 1136–1148.

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

Lin, E. Y. Y., Chen, P. Y., & Chen, C. C. (2013). Measuring the environmental efficiency of countries: a directional distance function metafrontier approach. Journal of environmental management, 119, 134-142.

Layer, K., Johnson, A. L., Sickles, R. C., & Ferrier, G. D. (2020). Direction selection in stochastic directional distance functions. European Journal of Operational Research, 280(1), 351–364.

O'Donnell, C. J., Rao, D. P., & Battese, G. E. (2008). Metafrontier frameworks for the study of firm-level efficiencies and technology ratios. Empirical Economics, 34 (2), 231–255.

Simar, L., & Wilson, P. W. (1998). Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models. Management Science, 44(1), 49–61.

Simar, L., & Wilson, P. W. (1999). Estimating and bootstrapping Malmquist indices. European Journal of Operational Research, 115(3), 459–471.

Simar, L., & Wilson, P. W. (2000). A general methodology for bootstrapping in nonparametric frontier models. Journal of Applied Statistics, 27(6), 779–802.

Simar, L., & Wilson, P. W. (2007). Estimation and inference in two-stage, semi-parametric models of production processes. Journal of Econometrics, 136 (1), 31–64.

Wand, M.P., (1994). Fast Computation of Multivariate Kernel Estimators. Journal of Computational and Graphical Statistics 3 (4), 433–445.

Wang C-H, Ko, M-H, Chen, W-J (2019). Effects of Kyoto Protocol on CO2 Emissions: A Five-Country Rolling Regression Analysis. Sustainability 2019, 11, 744, 1–20.

c:\cross-country