Darmstadt
Discussion Papers
in ECONOMICS

TECHNISCHE
UNIVERSITÄT
DARMSTADT

Estimating the Materials Balance Condition: A Stochastic Frontier Approach

Benjamin Hampf

Nr. 226

Arbeitspapiere der
Volkswirtschaftlichen Fachgebiete der TU Darmstadt

ISSN: 1438-2733

Estimating the Materials Balance Condition: A Stochastic Frontier Approach

by

Benjamin Hampf*†

Department of Law and Economics, Darmstadt University of Technology,

Hochschulstraße 1, D-64289 Darmstadt, Germany

April 2015

Abstract

In this paper we propose a stochastic formulation of the materials balance condition which imposes physical constraints on production technologies. The estimation of the model involves a composed error term structure that is commonly applied in the literature on stochastic frontier analysis of productive efficiency. Moreover, we discuss how OLS, maximum likelihood and Bayesian methods can be used to estimate the proposed model. In contrast to previous approaches our model allows to estimate the physical limitations to production possibilities in the presence of statistical noise and depends on substantially weaker data requirements. We demonstrate the applicability of our new approach by estimating the materials balance condition for SO_2 and CO_2 using a sample of fossil-fueled power plants in the United States.

JEL classification: Q53, C51, Q40, D24

Keywords: Materials balance condition; Abatement efficiency; Stochastic frontier analysis; Laws of thermodynamics; Applied econometrics; Environmental economics

*Tel.: +49 6151 165319, Fax: +49 6151 163897, E-Mail: hampf@vwl.tu-darmstadt.de

†I am grateful to Jens Krüger and Kenneth Rødseth for valuable comments.

1 Introduction

Accounting for the physical limits on production possibilities imposed by thermodynamic laws has become an important issue in the literature on environmental production economics. For example, Baumgärtner et al. (2001) discuss the physical limits to joint production of desirable outputs and pollution while Lauwers (2009) emphasizes the importance of physical constraints when estimating the productive efficiency of decision making units (DMUs, e.g. firms). In particular, researchers in production economics are concerned with the first law of thermodynamics, the law of energy conservation.[1] In its reduced form as the the law of mass conservation, it has been introduced to the economic literature by Ayers and Kneese (1969) and has become an important research object commonly referred to as the materials balance condition (MBC). This condition states that materials cannot vanish during a production or, more general, transformation process. For example, in the production of wooden chairs the initial amount of the input wood has to be equal to the weight of the good output (wooden chairs) as well as the production residual (shavings).

In the economic literature the MBC provides an important tool for both theoretical and empirical research with a focus on environmental pollution.[2] Theoretical research is concerned with modeling production technologies which account for the restrictions imposed by the MBC. While Ebert and Welsch (2007) discuss the implications of the MBC for production economics, Førsund (2009) and Murty et al. (2012) criticize conventional models on production technologies including environmental pollution (e.g. Färe et al. (1989)) for not accounting for the MBC. Since these models allow for physically impossible changes in the production structure of the DMUs they are likely to lead to biased and implausible results, e.g. when analyzing the productive efficiency of DMUs based on these estimated technologies. To overcome this issue, Hampf (2014) and Rødseth (2014) propose theoretical production models which allow to estimate the production technology accounting for the MBC. While Hampf (2014) models the overall environmental production technology set as the intersection of a conventional technology set and a set defined by the MBC, Rødseth (2014) proposes a modified set of axioms for production technologies based on the concepts of weak g-disposability and output essentiality. Empirical studies on environmental efficiency (see Coelli et al. (2007), Welch and Barnum (2009), Hoang and Rao (2010), Hoang and Coelli (2011), Hoang and Alauddin (2012) and Hoang and Nguyen (2013)) apply the MBC to evaluate minimal emissions for a fixed amount of outputs accounting for factor substitution according to the MBC.

Although the MBC has become a frequently applied tool in economic studies, its current formulation imposes critical limits for empirical research. When incorporated into economic models, the MBC is included as a deterministic equation without accounting for any random elements. For example, stochastic effects which may affect the generation of emissions (e.g. weather conditions) are not captured in this modelling. While this approach to the MBC reflects the original idea of a physical law which can be studied under laboratory conditions, it is less suitable for

[1] See Çengel (2008) for an introduction to the laws of thermodynamics.

[2] In this paper we focus on the implications of the MBC for microeconomic theory and applications. For studies on macroeconomic consequences of the MBC see van den Bergh and Nijkamp (1994) and Akao and Managi (2007).

empirical economic research where predominantly data are used which are obtained from a non-controllable environment, hence are faced with stochastic influences. To our knowledge, the only study which accounts for stochastic effects in a model including the MBC is Hoang and Nguyen (2013). In this study differences in environmental efficiency, which is estimated using the MBC, are explained by applying second-stage regressions. However, as pointed out by Simar and Wilson (2007) the two-stage approach provides statistically valid results only under very restrictive assumptions, e.g. a separability condition which ensures that the second-stage regressors affect the efficiency results but not the shape of the technology. Therefore, the approach by Hoang and Nguyen (2013) addresses differences in environmental efficiency assuming that the coefficients of the MBC remain known, deterministic constants.

Furthermore, all currently available models which include the MBC depend on very detailed data regarding the production process (e.g. material content of inputs) which are only available in a very small number of datasets. For example, all of the above cited studies are either focusing on the energy sector in the United States or datasets from agricultural economics.[3] This dependency on detailed data is particularly unfortunate since it forces applied researchers to rely on conventional approaches to environmental production technologies which violate the MBC and, thus, are based on physically infeasible production possibilities.[4]

To overcome these problems, we propose a stochastic formulation of the MBC and discuss methods to estimate this model in the presence of unobserved abatement of pollutants. Our proposed estimation strategy is based on recent methodological advances in the literature on stochastic frontier analysis (SFA) accounting for a composed error term, heterogeneity and doubly-truncated distributions. We further discuss how ordinary least squares (OLS), maximum likelihood (ML) and Bayesian methods can be used to estimate the parameters of the stochastic materials balance condition (SMBC). The empirical applicability of our model is demonstrated by an estimation of the SMBC for sulfur dioxide (SO_2) and carbon dioxide (CO_2) using a well-studied dataset on U.S. power plants. Since for this sample very detailed data (e.g. on abatement efficiencies) are available, we are able to compare the estimated results of our novel model with the reported values which provide benchmark results.

Our results indicate that for an analysis of pollutants for which abatement technologies are available (e.g. SO_2) our new methodology estimated by ML or Bayesian methods provides reasonable results for both the coefficients of the MBC as well as the abatement efficiency of the plants. Moreover, we find that in this setting conventional OLS leads to seriously biased estimates. However, for pollutants without abatement possibilities (e.g. CO_2) we find that OLS performs better in small samples than the ML or Bayesian estimation of a misspecified model allowing for abatement.

This paper is structured as follows: Section 2 presents the formulation and the estimation of the stochastic materials balance condition while section 3 presents the data and results of our analysis of U.S. power plants. Finally, section 4 concludes the paper.

[3] This difficulty for applied research has been discussed e.g. during a panel session on the modeling of bad outputs at the European Workshop on Efficiency and Productivity Analysis in Verona in 2011.

[4] See Hampf and Rødseth (forthcoming) for a comparison of a conventional and a materials balance approach to environmental efficiency.

2 Theory of the stochastic materials balance condition

In this section we present the theoretical approach to stochastic materials balances. We start by presenting the formulation of the SMBC and discuss distributional assumptions on the composed error term. Based on this formulation we discuss possibilities to estimate this stochastic model.

2.1 Formulating the stochastic materials balance condition

In the following discussion we assume a production process where m inputs $\boldsymbol{x} \in \mathbb{R}_+^m$ are used to produce k good (or desirable) outputs $\boldsymbol{y} \in \mathbb{R}_+^k$ jointly with a single pollutant $p \in \mathbb{R}_+$. We denote by T the technology set of the production process containing all technically (and physically) possible production points. Therefore, T can be defined as:

$$T = \left\{ (\boldsymbol{x}, \boldsymbol{y}, p) \in \mathbb{R}_+^{m+k+1} : \boldsymbol{x} \text{ can produce } (\boldsymbol{y}, p) \right\}. \tag{2.1}$$

T is restricted by the MBC in a way that it only includes production points which ensure that the amount of materials bound in the inputs is equal to the amount of materials bound in the good outputs and the pollutant. Formally, the MBC for the above defined technology set is given by:

$$\boldsymbol{s}_x^T \boldsymbol{x} = \boldsymbol{s}_y^T \boldsymbol{y} + p \tag{2.2}$$

Here, $\boldsymbol{s}_x^T \in \mathbb{R}_+^m$ denotes the transpose of the $m \times 1$ vector of emission factors indicating the amount of materials bound in one unit of each input while $\boldsymbol{s}_y^T \in \mathbb{R}_+^k$ denotes the transpose of the $k \times 1$ vector of recuperation factors indicating the amount of materials bound in one unit of each output. Both vectors may include zeros for inputs or outputs which do not contain any materials (e.g. labor input or electricity output).

In this formulation of the MBC we have assumed that no pollution control in form of abatement activities takes place. Therefore, all materials which are not contained in the good outputs are released as pollution p to the environment. However, in many production processes abatement technologies (e.g. scrubbers) are used to reduce the amount of pollution emitted to the environment. Denoting $a \in \mathbb{R}_+$ the amount of abatement of pollutant p the MBC which accounts for abatement processes can be formulated as:

$$\boldsymbol{s}_x^T \boldsymbol{x} = \boldsymbol{s}_y^T \boldsymbol{y} + p + a. \tag{2.3}$$

As discussed in the introduction, empirical analyses use equation (2.3) as a deterministic restriction on the production possibilities. Therefore, precise data on inputs, outputs, pollution and emission as well as recuperation factors are needed for applied research.[5] Usually, the amount of abatement is unknown to the researcher and is estimated as the residual of the MBC (see Hampf (2014)). Therefore, the application of the model depends on the availability of very detailed data and, even if these data are available, does not account for any stochastic elements.

[5] While precise data on emission factors are available for energy economic studies, studies in agricultural economics are often based on rough rule-of-thumbs approaches to these factors (see the discussion on nitrogen and phosphorus content in Hoang and Coelli (2011)).

To overcome these issues, we propose a stochastic formulation of the MBC. Assuming a sample of $i = 1, \ldots, n$ observations (e.g. plants) the stochastic materials balance condition can be defined as:

$$p_i = \boldsymbol{\beta}_x^T \boldsymbol{x}_i - \boldsymbol{\beta}_y^T \boldsymbol{y}_i - a_i + u_i \tag{2.4}$$

where we have replaced the known emission and recuperation factors by unknown, to be estimated vectors of regression coefficients $\boldsymbol{\beta}_x$ and $\boldsymbol{\beta}_y$. In order to interpret these coefficients as emission and recuperation factors we additionally impose the restrictions $\boldsymbol{\beta}_x \geq \mathbf{0}$ and $\boldsymbol{\beta}_y \geq \mathbf{0}$ indicating that neither inputs nor outputs can be associated with a negative material content. In this model p_i can be interpreted as the net emissions released to the environment, while $\boldsymbol{\beta}_x^T \boldsymbol{x}_i - \boldsymbol{\beta}_y^T \boldsymbol{y}$ represents the gross emissions generated during the production of good outputs. The random error term u_i captures the stochastic influences on the regression model (e.g. measurement errors). Combining the stochastic error term with the (unknown) amount of abatement a_i leads to the composed error term of the SMBC model:

$$\epsilon_i = -a_i + u_i. \tag{2.5}$$

In our model the composed error term ϵ_i influences the emissions in two ways. The gross emissions are reduced by the amount of abatement a_i. However, the final amount of net emissions released to the environment is also affected by random components (e.g. weather conditions) captured by u_i.

This structure of the proposed regression model is similar to the stochastic frontier models studied in the literature on the measurement of productive efficiency (see Kumbhakar and Lovell (2000) for an overview of this literature). These models are used to estimate the coefficients of production functions (e.g. partial elasticities of production) accounting for productive inefficiency of the observations, which corresponds to our abatement term, and random noise. In the following we exploit and adapt concepts and assumptions from the SFA literature to obtain an estimation procedure for our SMBC model.

In line with the SFA literature we assume that the random effects are normally distributed with mean equal to 0 and variance σ_u^2. Moreover, we assume that they are independent from the abatement term a_i.[6] Hence, we impose the distributional assumption:

$$u_i \overset{iid}{\sim} N\left(0, \sigma_u^2\right) \tag{2.6}$$

To estimate the emission and recuperation factors of the SMBC and to obtain results for the amount of abatement, distributional assumptions need to be imposed on the abatement term a_i. In contrast to the noise component u_i, the abatement a_i does not follow a symmetric distribution due to a truncation from below at zero. This truncation follows from the non-negativity assumption regarding the amount of abatement (or inefficiency in the SFA models).

Since the seminal papers on SFA by Aigner et al. (1977) and Meeusen and van den Broeck (1977) a large number of distributional assumptions regarding the inefficiency term has been proposed accounting for the the special structure of this term. In addition to the truncation at zero, it

[6] See Smith (2008) for a SFA model accounting for dependencies in the composed error term.

is assumed that the distribution is right-skewed implying that a large share of the probability mass is located close to zero. Therefore, most observations are assumed to be only slightly inefficient. Using equation (2.5) and combining the (symmetric) normal distribution for u_i with a right-skewed distribution for the inefficiency term leads to a left-skewed distribution of the composed error term ϵ_i. Following Waldman (1982) the skewness of the residuals of a simple OLS regression can be used as an indicator on whether the SFA model contains any inefficiency. These residuals should be left-skewed if the output of the observations is systematically reduced due to inefficiencies. To obtain a right-skewed distribution for the inefficiency term truncated from below at zero, Aigner et al. (1977) propose to use a half-normal distribution or an exponential distribution. Stevenson (1980) proposes a more flexible formulation based on a truncated normal distribution allowing for a non-zero mode of the distribution. Greene (1990) presents a model based on the gamma distribution.[7] Note that all of these models assume that the mean and the variance of the inefficiency distribution are identical across the observations.

Although the basic formulation of SFA models is similar to our SMBC model, the conventional distributional assumptions regarding the inefficiency term cannot be readily imposed on the abatement term a_i in our model. The abatement term differs from the inefficiency term of SFA models in the following points:

1. a_i is bounded from below *and* above.
2. $E(a_i)$ and $Var(a_i)$ are not identical for all $i = 1, \dots, n$.
3. No assumption regarding the skewness of the distribution of a_i can be imposed.

The first point captures the fact that abatement (like inefficiency) can not be negative and hence a_i is truncated from below at zero. However, in contrast to inefficiency the abatement term a_i is also truncated from above. This upper boundary is equal to the gross emissions $(\boldsymbol{\beta}_x^T \boldsymbol{x}_i - \boldsymbol{\beta}_y^T \boldsymbol{y}_i)$ since the maximal possible amount of abatement cannot be larger than the gross emissions which implies non-negative net emissions.[8]

The second point addresses the issue that in our model a_i denotes the total amount of abated gross emissions. Since these gross emissions depend on the individual amount of inputs and outputs of the DMUs, the mean of the distribution on a_i cannot be constant across all observations. Moreover, in line with Wang (2002) we do not impose the rather strict constraint that the variance of a_i is constant while the mean varies. Therefore, we assume that the distribution of a_i exhibits a mean and a variance varying among observations.

The last point captures the fact that the skewness of a_i depends on the abatement efficiency of the DMUs. If, in line with the conventional SFA models, one assumes that most DMUs are very efficient, hence abate nearly all of their emissions, then $a_i \approx \boldsymbol{\beta}_x^T \boldsymbol{x}_i - \boldsymbol{\beta}_y^T \boldsymbol{y}$ for most observations indicating a mode of the distribution close to the upper boundary given by the gross emissions. This leads to a left-skewed distribution for a_i and, thus, a right-skewed distribution for the

[7] See Greene (2008) for an overview of other distributional assumptions.

[8] From a theoretical point of view an upper boundary is also present in SFA models since the inefficiency cannot be larger than the output according to the production function. However, in these models the inefficiency is often concentrated at zero so that the upper boundary has no effect on the distribution or the results.

composed error term.[9] Therefore, the conventional SFA assumptions leading to a right-skewed distribution and the diagnostic by Waldman (1982) can not be applied in our model.

To obtain a distributional assumption which accounts for all points discussed above we combine several recently proposed methodological advances in the literature of SFA. To our knowledge, the only SFA model which accounts for a doubly-truncated inefficiency term has been proposed by Almanidis et al. (2014). They extend the truncated normal model by Stevenson (1980) and assume that the inefficiency term is bounded from below at zero and from above at a fixed constant $B > 0$. In our model of abatement their approach results in the distributional assumption:

$$a_i \overset{iid}{\sim} \left|N(\omega, \sigma_a^2)\right|^+, \;\; 0 \leq a_i \leq B. \tag{2.7}$$

where ω denotes the mean of the normal distribution truncated at 0 and B, while σ_a^2 denotes the variance.[10] Assuming $B \geq \omega$ the mean of the untruncated distribution is equal to the mode of the truncated distribution. As discussed above, the assumption of a fixed mean and variance of the distribution cannot be uphold in our model. Several formulations accounting for varying means (see e.g. Kumbhakar et al. (1991)) or variances (see e.g. Caudill and Ford (1993)) based on parametric specifications have been proposed. Wang (2002) presents a unifying approach including functional forms for both parameters. In addition, Kumbhakar and Tsionas (2008) propose a model based on Kumbhakar et al. (2007) which estimates individual means and variances of the truncated normal distribution by using local maximum likelihood estimation.

To obtain a suitable and identifiable distribution for the abatement term a_i, we combine the approach by Almanidis et al. (2014) with the scaling property for the mean and variance of a distribution as introduced in the SFA literature by Wang and Schmidt (2002). As discussed in Alvarez et al. (2006), the scaling property implies that the individual distribution of a_i can be obtained by individual scaling of a basic random variable. This basic variable is assumed to follow the same distribution for all observations implying a fixed mean and variance of the basic distribution. In our approach to the SMBC we define the basic variable a_i^* by:

$$a_i^* = \frac{a_i}{\boldsymbol{\beta}_x^T \boldsymbol{x}_i - \boldsymbol{\beta}_y^T \boldsymbol{y}_i}. \tag{2.8}$$

a_i^* can be interpreted as the abatement efficiency of observation i. Hence, $100\% \times a_i^*$ represents the percentage reduction in gross emissions due to abatement. For this basic variable we assume the basic distribution:

$$a_i^* \overset{iid}{\sim} \left|N(\mu, \sigma_{a^*}^2)\right|^+, \;\; 0 \leq a_i^* \leq 1. \tag{2.9}$$

Here, μ denotes the mean (mode) of the untruncated (doubly-truncated) distribution and $\sigma_{a^*}^2$ the variance of the untruncated distribution. Since the abatement of one unit of gross emissions cannot be negative or larger than the gross emissions, the distribution of a_i^* is doubly-truncated at zero and one.

This specification of the distribution for the basic random variable implies that all observation

[9] See Almanidis and Sickles (2011) for further discussions on the "wrong" skewness issue in SFA models.

[10] In the terms of Amsler et al. (forthcoming) ω represents the pre-truncation mean. They further discuss whether the pre- or the post-truncation mean should be the focus of SFA modeling. In this paper we follow the conventional approaches and focus on the pre-truncation mean.

are assumed to have the same distribution regarding their abatement efficiency a_i^*. Based on this basic distribution all heterogeneity in the mean and variances of the distribution of the absolute amount of abatement a_i is caused by differences in the amount of gross emissions (differences in scale). This assumption leads to the following distribution for the abatement a_i:

$$a_i \sim \left(\boldsymbol{\beta}_x^T \boldsymbol{x}_i - \boldsymbol{\beta}_y^T \boldsymbol{y}_i\right) \times \left|N\left(\mu, \sigma_{a^*}^2\right)\right|^+, \ \ 0 \leq a_i^* \leq 1. \tag{2.10}$$

Since scaling of the doubly truncated distribution is feasible (see Horrace (2005)), a_i follows the distribution:

$$a_i \sim \left|N\left(\left(\boldsymbol{\beta}_x^T \boldsymbol{x}_i - \boldsymbol{\beta}_y^T \boldsymbol{y}_i\right)\mu, \left(\boldsymbol{\beta}_x^T \boldsymbol{x}_i - \boldsymbol{\beta}_y^T \boldsymbol{y}_i\right)^2 \sigma_{a^*}^2\right)\right|^+, \ \ 0 \leq a_i \leq \boldsymbol{\beta}_x^T \boldsymbol{x}_i - \boldsymbol{\beta}_y^T \boldsymbol{y}_i. \tag{2.11}$$

To visualize the scaling property, figure 1 depicts two density functions. The solid line indicates the density of the basic distribution $a_i^* \sim |N(1, 0.2)|^+, 0 \leq a_i^* \leq 1$ while the dashed line indicates the density of the scaled distribution $a_i \sim |N(2, 0.8)|^+, 0 \leq a_i \leq 2$ where $a_i = 2a_i^*$. While both density functions show the same shape (or structure) they differ with regard to their scale and, hence, exhibit different means and variances.

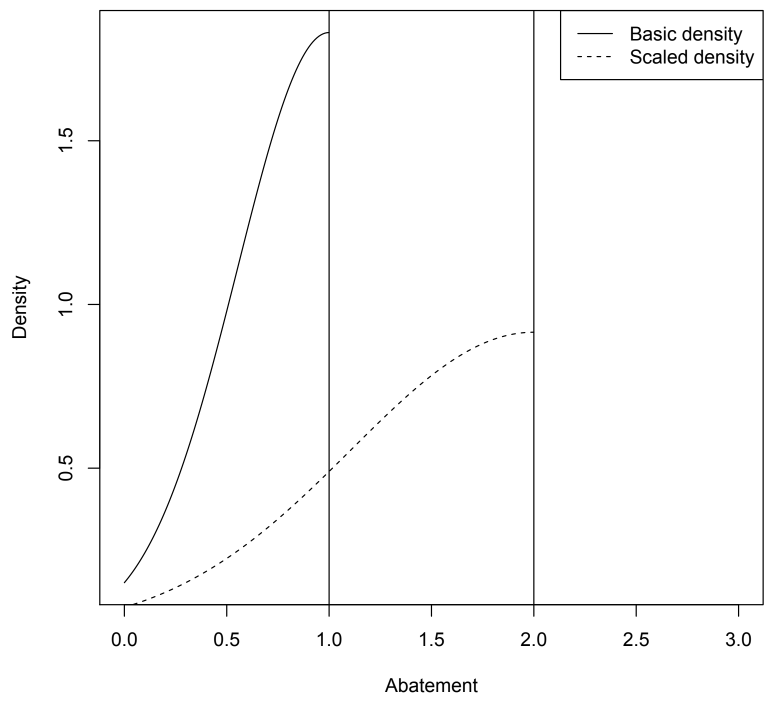

Figure 1: Example of the scaling property

Note that in contrast to the original publication on the scaling property by Wang and Schmidt (2002) we do not account for the influence of exogenous variables (so-called $\boldsymbol{z}$-variables).[11] By assuming a parametric specification of the mean of the distribution for a_i based on the inputs and outputs of the model, we follow the approach by Huang and Liu (1994) where a parametric specification of the mean based on both endogenous and exogenous variables is proposed. However, the model by Huang and Liu (1994) does not account for heterogeneity of the variance or the scaling property of the distribution. Modeling the variance as a function of the endogenous variables has been proposed by Caudill and Ford (1993) but without using the scaling property. Moreover, in contrast to Wang and Schmidt (2002) we do not rely on an exponential functional form which precludes a negative mean of the function. We refrain from this restriction in order to make our distributional assumption more closely connected to the discussion on the SMBC. If negative results are an issue in the estimation, the model can be adapted to account for an exponential structure.

Finally, we want to mention that our model assumes that the emission (recuperation) factors $\boldsymbol{\beta}_x$ ($\boldsymbol{\beta}_y$) are homogeneous across the observations.[12] Models that include heterogeneous coefficients have been proposed by Tsionas (2002) and Huang (2004). While, in principle, our model could be modified to account for heterogeneity, due to the simultaneous inclusion of the heterogeneous coefficients in the distribution on a_i the identification of the parameters may be difficult. For example, Galán et al. (2014) note that models with heterogeneous coefficients tend to show only very small inefficiencies. Intuitively, it may be impossible in a modified SMBC accounting for heterogeneous coefficients to identify econometrically whether a small amount of net emissions of an observation is caused by a small material content of the inputs of this observation or whether the observation abates a large amount of the gross emissions.

2.2 Estimating the stochastic materials balance condition

In the literature on stochastic frontier models with composed error terms various econometric methods to estimate the models have been discussed. Greene (1980) proved that ordinary least squares provide consistent estimates of the coefficients of the frontier function. More precisely, in a frontier model including a constant and slope parameters, OLS provide a consistent estimation of the slope parameters while the constant needs to be adjusted by adding the maximum value of residuals to obtain an unbiased estimator. This corrected ordinary least squares (COLS) approach has been proposed by Winsten (1957). In our model of the SMBC no constant is included since pollution can not be positive if no inputs or outputs are included. Therefore, simple OLS could be applied to estimate the SMBC. However, due to the correlation of the residuals and the regressors which is caused by the mean and the variance of the distribution of a_i being functions of the regressors, the standard assumptions for linear regression models are violated and hence OLS does not lead to consistent estimates. Only in case that $a_i = 0 \ \forall i$, thus in a situation without any abatement activities, OLS can be expected to provide statistically valid results. However, even if $a_i = 0 \ \forall i$ the non-negativity restrictions on the emission and

[11] However, the proposed model could be easily extended to include these variables.
[12] This implies, for example, that the coal used by power plants is homogeneous across the plants.

recuperation factors need to be accounted for. If OLS does not lead to positive coefficients, the restrictions can be accounted for by using inequality restricted least squares (see Liew (1976)).

Assuming $a_i \geq 0 \ \forall i$, we follow Huang and Liu (1994) and use maximum likelihood estimation to obtain the parameters of the stochastic materials balance condition, the distribution of the abatement efficiency and the total amount of abatement. Modifying the results for the doubly-truncated normal model by Almanidis et al. (2014) the log-likelihood function of our model reads as:

$$
\begin{aligned}
lnL = &-ln\left[\Phi\left(\frac{1-\mu}{\sigma_{a^*}}\right)-\Phi\left(\frac{-\mu}{\sigma_{a^*}}\right)\right]-\sum_{i=1}^{n} ln\left(\sqrt{\sigma_u^2+\left(\boldsymbol{\beta}_x^T\boldsymbol{x}_i-\boldsymbol{\beta}_y^T\boldsymbol{y}_i\right)^2\sigma_{a^*}^2}\right)\\
&-\frac{n}{2}ln\left(2\pi\right)-\sum_{i=1}^{n}\frac{\left(p_i+(\mu-1)\left(\boldsymbol{\beta}_x^T\boldsymbol{x}_i-\boldsymbol{\beta}_y^T\boldsymbol{y}_i\right)\right)^2}{2\left(\sigma_u^2+\left(\boldsymbol{\beta}_x^T\boldsymbol{x}_i-\boldsymbol{\beta}_y^T\boldsymbol{y}_i\right)^2\sigma_{a^*}^2\right)}\\
&+\sum_{i=1}^{n} ln\left[\Phi\left(\frac{\frac{p_i\left(\boldsymbol{\beta}_x^T\boldsymbol{x}_i-\boldsymbol{\beta}_y^T\boldsymbol{y}_i\right)\sigma_{a^*}}{\sigma_u}+\frac{(1-\mu)\sigma_u}{\sigma_{a^*}}}{\sqrt{\sigma_u^2+\left(\boldsymbol{\beta}_x^T\boldsymbol{x}_i-\boldsymbol{\beta}_y^T\boldsymbol{y}_i\right)^2\sigma_{a^*}^2}}\right)-\Phi\left(\frac{\frac{\left(p_i-\boldsymbol{\beta}_x^T\boldsymbol{x}_i+\boldsymbol{\beta}_y^T\boldsymbol{y}_i\right)\left(\boldsymbol{\beta}_x^T\boldsymbol{x}_i-\boldsymbol{\beta}_y^T\boldsymbol{y}_i\right)\sigma_{a^*}}{\sigma_u}-\frac{\mu\sigma_u}{\sigma_{a^*}}}{\sqrt{\sigma_u^2+\left(\boldsymbol{\beta}_x^T\boldsymbol{x}_i-\boldsymbol{\beta}_y^T\boldsymbol{y}_i\right)^2\sigma_{a^*}^2}}\right)\right]
\end{aligned}
\tag{2.12}
$$

where $\Phi(.)$ denotes the standard normal distribution function.

Maximizing this function leads to ML estimators for $\boldsymbol{\beta}_x, \boldsymbol{\beta}_y, \mu, \sigma_{a^*}^2$ and σ_u^2. Note that suitable transformations of the parameters are needed to ensure non-negativity. In particular, μ is bounded in the interval $[0,1]$. This can be achieved, e.g., by the hyperbolic transformation of parameters (see Nash (2014) for a discussion).[13]

The results of the ML estimation can be used to calculate the abatement efficiency a_i^* using a modified version of the Jondrow et al. (1982) estimator of technical inefficiency for SFA models. The estimator of the abatement efficiency in our SMBC model is given by the expected value of a_i^* conditional on $\frac{\epsilon_i}{\boldsymbol{\beta}_x^T\boldsymbol{x}_i-\boldsymbol{\beta}_y^T\boldsymbol{y}_i}$ and reads as:

$$
E\left(a_i^*\left|\frac{\epsilon_i}{\boldsymbol{\beta}_x^T\boldsymbol{x}_i-\boldsymbol{\beta}_y^T\boldsymbol{y}_i}\right.\right)=\mu^*+\sigma^*\frac{\phi\left(-\frac{\mu^*}{\sigma^*}\right)-\phi\left(\frac{1-\mu^*}{\sigma^*}\right)}{\Phi\left(\frac{1-\mu^*}{\sigma^*}\right)-\Phi\left(-\frac{\mu^*}{\sigma^*}\right)}
\tag{2.13}
$$

where $\phi(.)$ denotes the standard normal density function. Moreover,

$$
\mu^*=\frac{\mu^*\frac{\sigma_u^2}{\left(\boldsymbol{\beta}_x^T\boldsymbol{x}_i-\boldsymbol{\beta}_y^T\boldsymbol{y}_i\right)^2}-\frac{\epsilon_i}{\boldsymbol{\beta}_x^T\boldsymbol{x}_i-\boldsymbol{\beta}_y^T\boldsymbol{y}_i}\sigma_{a^*}^2}{\sigma_{a^*}^2+\frac{\sigma_u^2}{\left(\boldsymbol{\beta}_x^T\boldsymbol{x}_i-\boldsymbol{\beta}_y^T\boldsymbol{y}_i\right)^2}},\quad \sigma^*=\frac{\sigma_{a^*}\sqrt{\frac{\sigma_u^2}{\left(\boldsymbol{\beta}_x^T\boldsymbol{x}_i-\boldsymbol{\beta}_y^T\boldsymbol{y}_i\right)^2}}}{\sqrt{\sigma_{a^*}^2+\frac{\sigma_u^2}{\left(\boldsymbol{\beta}_x^T\boldsymbol{x}_i-\boldsymbol{\beta}_y^T\boldsymbol{y}_i\right)^2}}}.
$$

Based on the estimates of a_i^* and the model parameters, the total amount of abatement of observation i can be estimated using:

$$
\widehat{a}_i=\widehat{a}_i^*\left(\widehat{\boldsymbol{\beta}}_x^T\boldsymbol{x}_i-\widehat{\boldsymbol{\beta}}_y^T\boldsymbol{y}_i\right).
\tag{2.14}
$$

[13] In the statistical software R the `dfoptim` package provides an optimization algorithm based on the hyperbolic transformation allowing for interval bounded parameters.

However, the ML estimator suffers from the fact that the doubly-truncated SFA model is only locally identified (see Almanidis et al. (2014) for further discussions). Hence, multiple local maxima may exist making it necessary to check multiple starting values or use global optimization methods (e.g. simulated annealing) to find the global optimum which may not exist (see Ritter and Simar (1997) for further discussions on this issue in SFA models). One solution to this problem is the reduction of the number of parameters and hence the flexibility of the model. For example, following the conventional SFA models one may assume that most observations abate nearly all their gross emissions and, thus, are abatement efficient. Therefore, the number of parameters can be reduced by setting $\mu = 1$.

Another solution to this issue are Bayesian methods which do not rely on the maximization of the likelihood function. Bayesian stochastic frontier models have been introduced by van den Broeck et al. (1994).[14] In their paper, van den Broeck et al. (1994) focus on deriving priors for the basic stochastic frontier models as discussed above. Since our model is not based on the classical SFA models we cannot rely on these informative priors. However, as shown by Fernández et al. (1997) the posteriori distribution exists given that the applied priors are proper (integrating to one) even if vague (or non-informative) priors are used. To estimate the above derived model of a stochastic materials balance condition we propose a set of priors extending the ideas of Griffin and Steel (2007):[15]

$$\begin{aligned} &\beta_{x_r} \sim \left| N\left(0, \sigma^2_{\beta_{x_r}}\right)\right|^+ \quad \forall r = 1, \ldots, m \\ &\beta_{y_s} \sim \left| N\left(0, \sigma^2_{\beta_{y_s}}\right)\right|^+ \quad \forall s = 1, \ldots, k \\ &\sigma^{-2}_{a^*} \sim \Gamma(a, b) \\ &\sigma^{-2}_{u} \sim \Gamma(c, d) \\ &\mu \sim U(0, 1) \end{aligned} \tag{2.15}$$

To ensure non-negativity of the emission and recuperation factors we use independent truncated normal priors with truncation point zero. The independence assumption is based on the theoretical concept of emission factors, e.g. in our empirical example of a SMBC for SO_2 of U.S. power plants the sulfur content of coal is independent from the sulfur content of natural gas. However, in case that this assumption cannot be imposed the dependencies between the factors could be taken into account by using multivariate truncated normal priors. However, this distribution is not implemented in standard software for Bayesian econometrics (e.g. WinBUGS) but has to be manually accounted for (see Wilhelm and Manjunath (2010) for a presentation of the multivariate truncated normal distribution for the statistical software R). In addition to the priors for the emission and recuperation factors, we assume the conventional gamma priors for the inverted variances (precisions) of the distributions of the random noise and the abatement efficiency term. Finally, a uniform prior spanning the interval [0,1] is assumed for the mean of the truncated distribution for a_i^*.

[14] For an introduction to Bayesian SFA see Tsionas (2001). For practical implementations using WinBUGS see Griffin and Steel (2007).

[15] See the following section for a discussion on the choice of the parameters of the prior distributions in an empirical analysis.

3 Empirical application to U.S. power plants

In this section we present the data used to illustrate the above proposed SMBC model and discuss the results for U.S. power plants comparing the estimates obtained by OLS, ML and Bayesian analysis.

3.1 Data of U.S. power plants

To demonstrate the empirical applicability of our model we estimate the stochastic materials balance condition for two pollutants, sulfur dioxide (SO_2) and carbon dioxide (CO_2), using a sample of U.S. power plants. These plants have been analyzed in numerous empirical studies on production economics using both conventional (see e.g. Färe et al. (2007) and Mekaroonreung and Johnson (2012)) as well as materials balance based models (see Welch and Barnum (2009), Hampf (2014) and Hampf and Rødseth (forthcoming)).[16] The deterministic MBC models are applied to analyze these plants since very detailed data on the fuel consumption, emission output as well as emission factors are freely available.

We apply our new stochastic materials balance model to these plants to show how valid results can be obtained using only a subset of the data needed in previous studies while simultaneously accounting for random noise. Moreover, we use the reported material content as well as the abatement efficiencies with regard to SO_2 as a benchmark to compare the results of our model.[17] Finally, the estimation of the SMBC for CO_2 and SO_2 allows us to analyze the effects of mispecifying the SMBC. For SO_2 (CO_2) abatement technologies are (not) available and hence assuming zero (positive) abatement may lead to biased results.

In our study we focus on generating units which only use bituminous coal. Moreover, we also include units which in addition to bituminous coal use natural gas. Although our model can account for multiple fuels (e.g. different types of oil, coal etc.) we restrict our sample to these two inputs. On the one hand, this approach allows to model a more homogeneous dataset with a limited number of observations as frequently analyzed in empirical studies (see e.g. Mekaroonreung and Johnson (2012)).[18] On the other hand, including one fuel type which contains sulfur (coal) and one fuel type which does not contain any sulfur (natural gas) allows to evaluate the quality of the estimates in terms of indicating an emission factor of sulfur dioxide equal to zero for natural gas. Since electricity generation is not associated with the production of a material containing output, no good output is included in our estimation.

We construct our sample using data on the electricity generating units of the plants provided by the U.S. Energy Information Administration (EIA) and the U.S. Environmental Protection Agency (EPA).[19] The analysis could be also performed using aggregated data on the plant level.

[16] See Zhou et al. (2008) for a survey on energy-related studies in efficiency analysis.

[17] To remove SO_2 emissions power plants use flue gas desulfurization units (FGDs). For further discussions on abatement by FGDs see Srivastava and Josewicz (2001).

[18] Note that in contrast to studies on productive efficiency our model does not need to be based on a homogeneous production technology since the materials balance is not affected by the efficiency with regard to transforming inputs into good outputs, e.g. differences among gas and coal fired power plants in generating electricity do not influence the material content of the inputs.

[19] See Woodruff et al. (2012) for further details on power plants structures (e.g. generators, boilers etc.)

However, some information e.g. on the abatement efficiency are only reported on the unit level. Moreover, typical studies on the efficiency of plants use unit level data in order to obtain a larger sample size and account for potential heterogeneity among the units which could be not accounted for if plant level data are used. Unit level data on the fuel consumption are provided by the file EIA-923. File EIA-860 includes detailed data on the FGD units. In particular, the abatement efficiency (on an annual basis) of the FGDs is reported. Finally, the EPA provides the Clean Air Markets database which contains data on the net SO_2 and CO_2 emissions of the electricity generating units.

Our analyzed sample contains 87 generating units which were in operation during 2012. While this number seems rather small it is in line with the number of observations in typical empirical analyses and provides more realistic opportunities to evaluate our model compared to the simulation of a large number of artificial data. The descriptive statistics of the included variables are presented in table I.

Table I: Descriptive statistics of the variables

	Min	Mean	Median	Max	St.dev.
Bituminous coal (short tons)	100.00	492327.03	374572.00	2132220.00	523251.65
Natural gas (1000 cubic feet)	0.00	363516.49	67230.00	7673302.00	1040011.29
Sulfur dioxide (short tons)	4.02	2389.62	1294.84	17448.63	3084.46
Carbon dioxide (short tons)	1844.48	1232139.18	903058.48	5527276.51	1294754.51

To evaluate whether the emission factors of bituminous coal and natural gas for SO_2 and CO_2 estimated by our model are realistic, we calculate approximations of the true values based on data of the sulfur content provided by the EIA and natural constants from physical chemistry. For the basics on physical chemistry as used in the following derivation of the emission factors see Atkins and de Paula (2014). The results of our calculations are summarized in table II.

Table II: Approximate emission factors

Fuel type	Emission factor (in short tons)	
	Sulfur dioxide (SO_2)	Carbon dioxide (CO_2)
Bituminous coal	$0.008 - 0.12$	–
Natural gas	0	0.056

In addition to the fuel input of the generating units, the EIA (2012, table 1) reports an interval of the sulfur content for each fuel type based on the minimal and maximal weight proportion of sulfur. For natural gas no sulfur content is reported since methane (CH_4) does not contain any sulfur. For bituminous coal the weight share of sulfur lies within the interval of $0.4 - 6.0\%$ leading to $3.6 - 54.4$ kilogram of sulfur per short ton of bituminous coal (1 short ton equals 0.907 metric tons which equal 1000 kilograms). The molar mass of sulfur is 32 gram per mol.[20] Hence, the burning of 1 short ton of bituminous coal leads to $112.5 - 1700$ mol of sulfur. Using the molar mass of SO_2 (64 gram per mol) this corresponds to $7200 - 108800$ gram of SO_2. Therefore,

[20] Mol is a chemical unit to express the amount of particles. One mol corresponds to 6.022×10^{23} particles.

the burning of one short ton of bituminous coal leads to approximately $0.008 - 0.12$ short tons of SO_2.

Unfortunately, the EIA does not report the carbon content of fuels. Thus, we cannot calculate an approximation of the CO_2 emission factor for bituminous coal. However, assuming that the natural gas input is pure methane we can calculate the theoretical emission factor of CO_2 for natural gas. As noted in table I, the EIA reports the natural gas input in 1000 cubic feet. One cubic feet is equal to 0.028 cubic meters and the density of methane is 0.66 kilogram per cubic meter. Therefore, 1000 cubic feet of methane weigh $1000 \times 0.028 \times 0.66 = 18.48$ kilogram. Since the molar mass of methane is 16 gram per mol this amounts to 1155 mol of methane. The molar mass of carbon dioxide is 44 gram per mol leading to 50820 gram of CO_2 per 1000 cubic feet of burned natural gas. This is equal to 0.056 short tons providing an approximation of the true CO_2 emission factor of natural gas.

In our analysis we will use the calculated results for the emission factors as a benchmark for the results estimated based on our novel methodology described above.

3.2 Results for U.S. power plants

In the following we present the results of our estimation of the SMBC using the data on 87 electricity generating units in the United States. We compare the results for each of the two pollutants (SO_2 and CO_2) using the three different estimation methods (OLS, ML and Bayesian analysis) discussed above. During our analysis we found that the log-likelihood function indeed shows multiple local maxima (see the discussion on the ML estimation in section 2). The global optimum has been obtained by a grid search using various combinations of starting values. Bayesian methods are applied using the priors specified above with parameter specification $a = b = c = d = 0.001$ leading to diffuse priors for the inverse variances (see Griffin and Steel (2007)). The Bayesian analysis is performed using 10000 iterations with a burn-in of 1000 iterations and a thinning rate of 5 draws leading to a fast mixing and converging of the chains. The reported results for the Bayesian analysis are the posterior median results.

The results of our estimations are presented in table III which contains the estimated emission factors for bituminous coal ($\widehat{\beta}_{BIT}$) and natural gas ($\widehat{\beta}_{NG}$), the estimated mode of the distribution of the abatement efficiency a^* ($\widehat{\mu}$) and the coefficient of determination (R^2). Since our model does not contain an intercept we follow Kvåseth (1985) and apply the uncentered R^2 value to compare the models.[21] The uncentered R^2 value for our SMBC model is defined as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (p_i - \widehat{p}_i)^2}{\sum_{i=1}^{n} p_i^2} \tag{3.1}$$

Note that the definition of the fitted values of the net pollution ($\widehat{p}_i$) differ among SO_2 and CO_2. For SO_2 the fitted values are calculated as: $\widehat{p}_i = \widehat{\beta}_{BIT} \times BIT_i + \widehat{\beta}_{NG} \times NG_i - \widehat{a}_i$, where BIT_i (NG_i) denotes the amount of bituminous coal (natural gas) input of generating unit i. In case

[21] Kvåseth (1985) discusses and compares various definitions of R^2 for standard and non-standard regression models.

of CO_2 the fitted values are obtained by: $\widehat{p_i} = \widehat{\beta}_{BIT} \times BIT_i + \widehat{\beta}_{NG} \times NG_i$. The difference is due to the fact that no abatement activities are present for CO_2 emissions.

In addition to the estimated coefficients we present standard errors in parentheses. Statistical significance is indicated by asterisks. Note that in order to make the OLS and ML results comparable to the Bayesian results we use the posterior median results as point estimators and derive statistical significance based on the estimated standard errors.[22]. Our approach therefore differs from usual Bayesian analyses which use the deviance information criterium (DIC) by Spiegelhalter et al. (2002) or Bayes factors to address statistical significance of coefficients. However, since our main interests is the comparison of results among different estimation methods we apply the frequentists methodology on the Bayesian estimates as well.

Table III: Results for the SMBC of U.S. power plants

	Sulfur dioxide (SO_2)			Carbon dioxide (CO_2)		
Coefficient	OLS	ML	Bayesian	OLS	ML	Bayesian
$\widehat{\beta}_{\mathrm{BIT}}$	0.0036***	0.056***	0.056***	2.472***	3.231***	3.18***
	(0.0004)	(0.0011)	(0.0034)	(0.018)	(0.035)	(0.1281)
$\widehat{\beta}_{\mathrm{NG}}$	0.000065	0.0002	0.00033	0.057***	0.087***	0.088***
	(0.0003)	(0.0002)	(0.0002)	(0.012)	(0.0062)	(0.0082)
$\widehat{\mu}$		1***	0.988***		0.239***	0.2213***
		(0.056)	(0.0153)		(0.016)	(0.0657)
R^2	0.44	1	1	0.99	0.90	0.91

Note: Standard errors in parentheses. For the Bayesian estimates posterior median results are reported. *** indicates statistical significance on the 1%-level.

We start by discussing the results for SO_2, a pollutant for which abatement technologies in form of flue gas desulfurization units exist. The results for the emission factors show that no model estimates a statistically significant content of sulfur for natural gas. Therefore, all models are able to correctly indicate that natural gas does not have an influence on SO_2 emissions. However, we find large differences in the numerical values of the statistically significant coefficient for bituminous coal across the estimation methods. While ML and Bayesian analyses both lead to an estimated coefficient of 0.056 the OLS estimate is 0.0036 and hence approximately 15 times smaller. Comparing the results with the SO_2 interval as obtained based on the reported sulfur content by the EIA ($0.008 - 0.12$) we find that while ML and Bayesian analyses lead to estimates which are close to the center of the interval, the OLS estimate lies far below the interval indicating a biased estimation of the emission factor in the presence of abatement. This poor performance is also indicated by the R^2 value which for OLS (0.44) is far smaller than the value for ML or Bayesian analyses (1).[23]

The large differences between the OLS and the ML and Bayesian estimates can be attributed to the incorporation of abatement efficiency in the latter models. The estimated mode of 1 (0.988)

[22] Due to the use of truncated normal priors for the emission factors the confidence intervals can by definition not include negative values. Therefore, the statistical significance can not be evaluated based on these intervals.

[23] Note that the small differences in the estimates among the ML and the Bayesian method do not lead to differences in the rounded R^2 value.

of the ML (Bayesian) estimation indicates that a large amount of the gross emissions is reduced by abatement. In contrast, OLS attributes the low net emissions to smaller emission factors of the inputs, in particular bituminous coal, and hence provides biased results for these coefficients.

Applying the estimation methods to CO_2 for which no abatement activities are present we obtain different results regarding the ability to estimate the emission factors. The R^2 values of 0.90 and 0.91 indicate that ML and Bayesian analyses perform well when estimating the net emissions. Nonetheless, they are outperformed by OLS with a R^2 of 0.99. This becomes even more obvious when comparing the results for the emission factors. All methods find a statistically significant CO_2 content of bituminous coal and natural gas. However, as can be seen from comparing the estimates for $\widehat{\beta}_{NG}$ with the theoretical emission factor of methane (0.056), the OLS estimate (0.057) is much closer providing a nearly perfect match to this value while the ML (0.085) or the Bayesian median posteriori estimate (0.088) are farther away from the theoretical values.

The larger emission factors of the ML and Bayesian estimations can be explained by the abatement term which should be zero for CO_2. However, both methods find a mode of the abatement efficiency that is statistically significantly larger than zero indicating that the misspecified models lead to biased estimates. While we assume that the estimation of a non-zero mode is largely due to the small sample size of our analysis, it nonetheless shows that OLS may lead to better results compared to a misspecified application of the doubly-truncated SFA model for the SMBC.

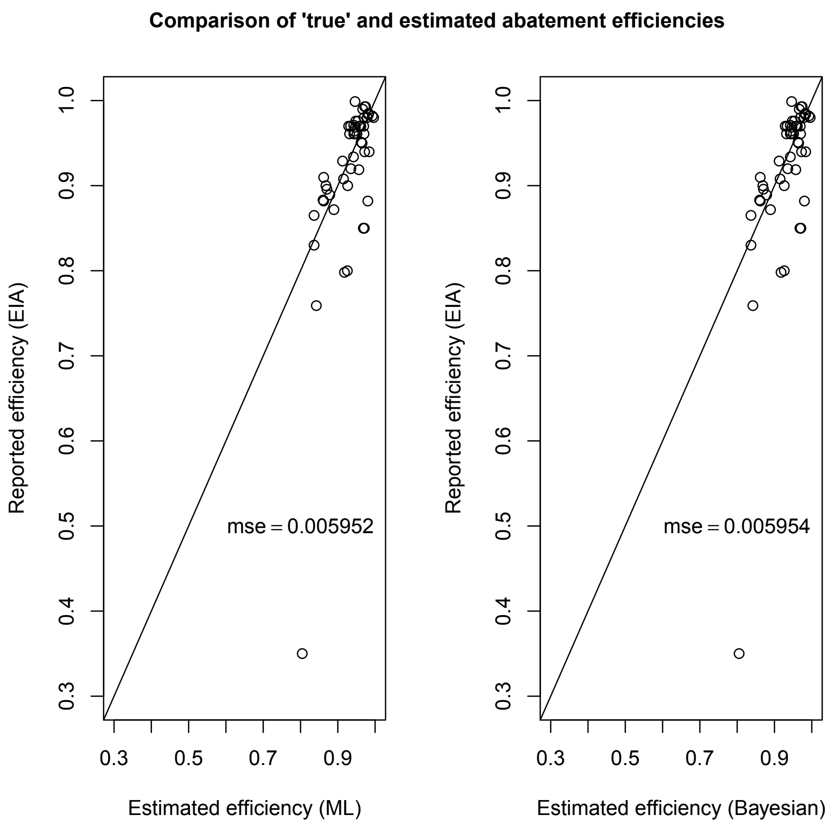

Figure 2: Abatement efficiencies for sulfur dioxide

To visualize whether the abatement efficiency a^* of SO_2 is correctly estimated by ML (using the Jondrow et al. (1982) type estimator defined above) and the Bayesian analysis (where the efficiency is simultaneously computed with the remaining parameters), figure 2 presents the abatement efficiencies reported by the EIA (which we use as the "true" value) and the estimated efficiencies according to the ML approach (left panel) and the Bayesian analysis (right panel). The diagonal line with slope 1 indicates all points for which estimated and true values are identical.

The figure shows that both approaches provide very similar estimates of the abatement efficiencies. Moreover, the majority of points is located close to the diagonal line indicating that most of the estimated efficiencies differ only slightly from their true values. This is also supported by the mean squared errors (mse) of the abatement efficiencies, which is 0.005952 for the ML and 0.005954 for the Bayesian approach. However, we also find that a few points are located farther below the diagonal line indicating that the accuracy of the estimation decreases for observations located farther away from the mode of the efficiency distribution. This becomes particularly obvious from the single outlier observation with a reported abatement efficiency of 0.35 and an estimated abatement efficiency of 0.8. Combining the findings from the figure we can conclude that both approaches (ML and Bayesian analysis) are capable of estimating the abatement efficiency correctly given homogeneous abatement efficiencies. However, the precision of the estimates declines if the abatement efficiencies are very heterogeneous among the observations. In this case the assumption of a common basic distribution of the efficiencies may not be supported by the data.

4 Conclusion

In this paper we have addressed the issue of estimating the materials balance condition in the presence of statistical noise and an unknown amount of abatement. Therefore, we have formulated a stochastic version of the materials balance condition and have shown how econometric methods based on stochastic frontier models can be applied to estimate our model. By accounting for random influences and allowing to estimate abatement efficiencies using substantially weaker data requirments than previous approaches, our model provides a possibility to incorporate fundamental physical laws in a larger number of applied research in production economics.

To demonstrate the applicability of model we estimated the stochastic materials balance condition for sulfur dioxide and carbon dioxide using a sample of U.S. power plants. Comparing estimations based on OLS, maximum likelihood and Bayesian methods we found that applying simple OLS leads to seriously biased results if emissions resulting from the production process are reduced by using abatement technologies. Comparing maximum likelihood and Bayesian methods we found that the latter provide a simpler approach to the estimation since maximum likelihood estimations are faced with multiple local maxima needing extensive grid searches or global optimization methods to obtain valid results. Comparing the abatement efficiencies as predicted by our model with the efficiencies reported by the plants we found that our model allows to estimate efficiencies correctly if they are homogeneous across the observations, i.e.

follow the same distribution. Outliers can be far less well explained in our model indicating that large heterogeneity of the efficiencies may limit the applicability of our model.

The proposed model addresses the estimation of a stochastic materials balance condition in the case of a single pollutant. To estimate the materials balance for two pollutants we implicitly assumed separability of the pollutants in our empirical example. Future research may address the possibility to estimate the stochastic materials balance condition simultaneously for multiple pollutants taking potential interdependencies into account. Such extensions could be based on stochastic frontier models of multiple outputs accounting for the production of bad outputs as proposed by Fernández et al. (2002). Moreover, in our model we assumed that the emission factors (regression parameters) are homogeneous across the analyzed observations. While stochastic frontier models of heterogeneous parameters lead to identification issues when combined with our parametric specifications of the mean and the variance of the abatement distribution, the local maximum likelihood approach of Kumbhakar and Tsionas (2008) could be combined with our model to allow for locally differentiated emission factors.

References

Aigner, D., C. A. K. Lovell, and P. Schmidt (1977). "Formulation and Estimation of Stochastic Frontier Production Function Models". In: *Journal of Econometrics* 6, pp. 21–37.

Akao, K.-I. and S. Managi (2007). "Feasibility and Optimality of Sustainable Growth under Materials Balance". In: *Journal of Economic Dynamics & Control* 31, pp. 3778–3790.

Almanidis, P., J. Qian, and R. C. Sickles (2014). "Stochastic Frontier Models with Bounded Inefficiency". In: *Festschrift in Honor of Peter Schmidt: Econometric Methods and Applications*. Ed. by R. C. Sickles and W. C. Horrace. New York: Springer, pp. 47–81.

Almanidis, P. and R. C. Sickles (2011). "The Skewness Issue in Stochastic Frontier Models: Fact or Fiction?" In: *Exploring Research Frontiers in Contemporary Statistics and Econometrics: A Festschrift for Léopold Simar*. Berlin: Springer, pp. 201–227.

Alvarez, A., C. Amsler, L. Orea, and P. Schmidt (2006). "Interpreting and Testing the Scaling Property in Models where Inefficiency Depends on Firm Characteristics". In: *Journal of Productivity Analysis* 25, pp. 201–212.

Amsler, C., P. Schmidt, and W.-J. Tsay (forthcoming). "A Post-Truncation Parameterization of Truncated Normal Technical Inefficiency". In: *Journal of Productivity Analysis*. DOI: 10.1007/s11123-014-0409-8.

Atkins, P. and J. de Paula (2014). *Atkins' Physical Chemistry*. 10th ed. Oxford University Press.

Ayers, R. U. and A. V. Kneese (1969). "Production, Consumption, and Externalities". In: *American Economic Review* 59, pp. 282–297.

Baumgärtner, S., H. Dyckhoff, M. Faber, J. Proops, and J. Shiller (2001). "The Concept of Joint Production and Ecological Economics". In: *Ecological Economics* 36, pp. 365–372.

Caudill, S. B. and J. M. Ford (1993). "Biases in Frontier Estimation Due to Heteroscedasticity". In: *Economics Letters* 41, pp. 17–20.

Çengel, Y. A. (2008). *Introduction to Thermodynamics and Heat Transfer*. 2nd ed. New York: McGraw-Hill.

Coelli, T., L. Lauwers, and G. Van Huylenbroeck (2007). "Environmental Efficiency Measurement and the Materials Balance Condition". In: *Journal of Productivity Analysis* 28, pp. 3–12.

Ebert, U. and H. Welsch (2007). "Environmental Emissions and Production Economics: Implications of the Materials Balance". In: *American Journal of Agricultural Economics* 89, pp. 287–293.

EIA (2012). *Form EIA-923 Power Plant Operations Report Instructions*.

Färe, R., S. Grosskopf, C. A. K. Lovell, and C. Pasurka (1989). "Multilateral Productivity Comparisons when some Outputs are Undesirable: A Nonparametric Approach". In: *Review of Economics and Statistics* 71, pp. 90–98.

Färe, R., S. Grosskopf, and C. Pasurka (2007). "Environmental Production Functions and Environmental Distance Functions". In: *Energy* 32, pp. 1055–1066.

Fernández, C., G. Koop, and M. F. J. Steel (2002). "Multiple-Output Production with Undesirable Outputs: An Application to Nitrogen Surplus in Agriculture". In: *Journal of the American Statistical Association* 97, pp. 432–442.

Fernández, C., J. Osiewalski, and M. F. J. Steel (1997). “On the Use of Panel Data in Stochastic Frontier Models with Improper Priors”. In: *Journal of Econometrics* 79, pp. 169–193.

Førsund, F. R. (2009). “Good Modelling of Bad Outputs: Pollution and Multiple-Output Production”. In: *International Review of Environmental and Resource Economics* 3, pp. 1–38.

Galán, J. E., H. Veiga, and M. P. Wiper (2014). “Bayesian Estimation of Inefficiency Heterogeneity in Stochastic Frontier Models”. In: *Journal of Productivity Analysis* 42, pp. 85–101.

Greene, W. H. (1980). “Maximum Likelihood Estimation of Econometric Frontier Functions”. In: *Journal of Econometrics* 13, pp. 27–56.

— (1990). “A Gamma-Distributed Stochastic Frontier Model”. In: *Journal of Econometrics* 46, pp. 141–163.

— (2008). “The Econometric Approach to Efficiency Analysis”. In: *The Measurement of Productive Efficiency and Productivity Growth*. Ed. by H. O. Fried, C. A. K. Lovell, and S. S. Schmidt. Oxford University Press, pp. 92–250.

Griffin, J. E. and M. F. J. Steel (2007). “Bayesian Stochastic Frontier Analysis using WinBUGS”. In: *Journal of Productivity Analysis* 27, pp. 163–176.

Hampf, B. (2014). “Separating Environmental Efficiency into Production and Abatement Efficiency: A Nonparametric Model with Application to US Power Plants”. In: *Journal of Productivity Analysis* 41, pp. 457–473.

Hampf, B. and K. L. Rødseth (forthcoming). “Carbon Dioxide Emission Standards for U.S. Power Plants: An Efficiency Analysis Perspective”. In: *Energy Economics*. DOI: 10.1016/j.eneco.2015.04.001.

Hoang, V.-H. and T. T. Nguyen (2013). “Analysis of Environmental Efficiency Variation: A Materials Balance Approach”. In: *Ecological Economics* 86, pp. 37–46.

Hoang, V.-N. and M. Alauddin (2012). “Input-Orientated Data Envelopment Analysis Framework for Measuring and Decomposing Economic, Environmental and Ecological Efficiency: An Application to OECD Agriculture”. In: *Environmental and Resource Economics* 51, pp. 431–452.

Hoang, V.-N. and T. Coelli (2011). “Measurement of Agricultural Total Factor Productivity Growth Incorporating Envionmental Factors: A Nutrients Balance Approach”. In: *Journal of Environmental Economics and Management* 62, pp. 462–474.

Hoang, V.-N. and D. S. P. Rao (2010). “Measuring and Decomposing Sustainable Efficiency in Agricultural Production: A Cumulative Exergy Balance Approach”. In: *Ecological Economics* 69, pp. 1765–1776.

Horrace, W. C. (2005). “On Ranking and Selection from Independent Truncated Normal Distributions”. In: *Journal of Econometrics* 126, pp. 334–354.

Huang, C. J. and J.-T. Liu (1994). “Estimation of a Non-Neutral Strochastic Frontier Production Function”. In: *Journal of Productivity Analysis* 5, pp. 171–180.

Huang, H.-C. (2004). “Estimation of Technical Inefficiencies with Heterogeneous Technologies”. In: *Journal of Productivity Analysis* 21, pp. 277–296.

Jondrow, J., C. A. K. Lovell, I. S. Materov, and P. Schmidt (1982). “On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Function Model”. In: *Journal of Econometrics* 19, pp. 233–238.

Kumbhakar, S. C., S. Ghosh, and J. T. McGuckin (1991). “A Generalized Production Frontier Approach for Estimating Determinants of Inefficiency in U.S. Dairy Farms”. In: *Journal of Business & Economic Statistics* 9, pp. 279–286.

Kumbhakar, S. C. and C. A. K. Lovell (2000). *Stochastic Frontier Analysis*. Cambridge (Mass.): Cambridge University Press.

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

Kumbhakar, S. C. and E. G. Tsionas (2008). “Scale and Efficiency Measurement using a Semiparametric Stochastic Frontier Model: Evidence from the U.S. Commercial Banks”. In: *Empirical Economics* 34, pp. 585–602.

Kvåseth, T. O. (1985). “Cautionary Note about R^2”. In: *American Statistician* 39, pp. 279–285.

Lauwers, L. (2009). “Justifying the Incorporation of the Materials Balance Principle into Frontier-Based Eco-Efficiency Models”. In: *Ecological Economics* 68, pp. 1605–1614.

Liew, C. K. (1976). “Inequality Constrained Least-Squares Estimation”. In: *Journal of the American Statistical Association* 71, pp. 746–751.

Meeusen, W. and J. van den Broeck (1977). “Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error”. In: *International Economic Review* 18, pp. 435–444.

Mekaroonreung, M. and A. L. Johnson (2012). “Estimating the Shadow Prices of SO_2 and NO_x for U.S. Coal Power Plants: A Convex Nonparametric Least-Squares Approach”. In: *Energy Economics* 34, pp. 723–732.

Murty, S., R. R. Russell, and S. B. Levkoff (2012). “On Modeling Pollution-Generating Technologies”. In: *Journal of Environmental Economics and Management* 64, pp. 117–135.

Nash, J. C. (2014). *Nonlinear Parameter Optimization using R Tools*. New York: Wiley.

Ritter, C. and L. Simar (1997). “Pitfalls of Normal-Gamma Stochastic Frontier Models”. In: *Journal of Productivity Analysis* 8, pp. 167–182.

Rødseth, K. L. (2014). “Axioms of a Polluting Technology: A Materials Balance Approach”. In: *Environmental and Resource Economics, revise and resubmit*.

Simar, L. and P. W. Wilson (2007). “Estimation and Inference in Two-Stage, Semi-Parametric Models of Production Processes”. In: *Journal of Econometrics* 136, pp. 31–64.

Smith, M. D. (2008). “Stochastic Frontier Models with Dependent Error Components”. In: *Econometrics Journal* 11, pp. 172–192.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. van der Linde (2002). “Bayesian Measures of Model Complexity and Fit”. In: *Journal of the Royal Statistical Society B* 64, pp. 583–639.

Srivastava, R. K. and W. Josewicz (2001). “Flue Gas Desulfurization: The State of the Art”. In: *Journal of the Air & Waste Management Association* 51, pp. 1676–1688.

Stevenson, R. E. (1980). “Likelihood Functions for Generalized Stochastic Frontier Estimation”. In: *Journal of Econometrics* 13, pp. 57–66.

Tsionas, E. G. (2001). “An Introduction to Efficiency Measurement using Bayesian Stochastic Frontier Models”. In: *Global Business & Economics Review* 3, pp. 287–311.

— (2002). “Stochastic Frontier Models with Random Coefficients”. In: *Journal of Applied Econometrics* 17, pp. 127–147.

van den Bergh, J. C. J. M. and P. Nijkamp (1994). “Dynamic Macro Modelling and Materials Balance: Economic-Environmental Integration for Sustainable Development”. In: *Economic Modelling* 11, pp. 283–307.

van den Broeck, J., G. Koop, J. Osiewalski, and M. F. J. Steel (1994). “Stochastic Frontier Models: A Bayesian Perspective”. In: *Journal of Econometrics* 61, pp. 273–303.

Waldman, D. M. (1982). “A Stationary Point for the Stochastic Frontier Likelihood”. In: *Journal of Econometrics* 18, pp. 275–279.

Wang, H.-J. (2002). “Heteroscedasticity and Non-Monotonic Efficiency Effects of a Stochastic Frontier Model”. In: *Journal of Productivity Analysis* 18, pp. 241–253.

Wang, H.-J. and P. Schmidt (2002). “One-Step and Two-Step Estimation of the Effects of Exogenous Variables on Technical Efficiency Levels”. In: *Journal of Productivity Analysis* 18, pp. 129–144.

Welch, E. and D. Barnum (2009). “Joint Environmental and Cost Efficiency Analysis of Electricity Generation”. In: *Ecological Economics* 68, pp. 2336–2343.

Wilhelm, S. and B. G. Manjunath (2010). “tmvtnorm: A Package for the Truncated Multivariate Normal Distribution”. In: *R Journal* 2, pp. 25–29.

Winsten, C. B. (1957). “Discussion on Mr. Farrell’s Paper”. In: *Journal of the Royal Statistical Society* 120, pp. 282–284.

Woodruff, E. B., H. B. Lammers, and T. F. Lammers (2012). *Steam Plant Operation.* New York: McGraw-Hill.

Zhou, P., B. W. Ang, and K. L. Poh (2008). “A Survey of Data Envelopment Analysis in Energy and Environmental Studies”. In: *European Journal of Operational Research* 189, pp. 1–18.