

## **A Monte Carlo study on multiple output stochastic frontiers a comparison of two approaches**

**Henningsen, Geraldine; Henningsen, Arne; Jensen, Uwe**

*Published in:*  
Journal of Productivity Analysis

*Link to article, DOI:*  
[10.1007/s11123-014-0416-9](https://doi.org/10.1007/s11123-014-0416-9)

*Publication date:*  
2015

*Document Version*  
Peer reviewed version

[Link back to DTU Orbit](#)

*Citation (APA):*  
Henningsen, G., Henningsen, A., & Jensen, U. (2015). A Monte Carlo study on multiple output stochastic frontiers: a comparison of two approaches. *Journal of Productivity Analysis*, 44(3), 309-320. DOI: 10.1007/s11123-014-0416-9

## **DTU Library** Technical Information Center of Denmark

---

### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

## A Monte Carlo Study on Multiple Output Stochastic Frontiers: A Comparison of Two Approaches

Géraldine Henningsen · Arne Henningsen · Uwe Jensen

Received: date / Accepted: date

**Abstract** In the estimation of multiple output technologies in a primal approach, the main question is how to handle the multiple outputs. Often, an output distance function is used, where the classical approach is to exploit its homogeneity property by selecting one output quantity as the dependent variable, dividing all other output quantities by the selected output quantity, and using these ratios as regressors (OD). Another approach is the stochastic ray production frontier (SR), which transforms the output quantities into their Euclidean distance as the dependent variable and their polar coordinates as directional components as regressors. A number of studies have compared these specifications using real world data and have found significant differences in the inefficiency estimates. However, in order to get to the bottom of these differences, we apply a Monte-Carlo simulation. We test the robustness of both specifications for the case of a Translog output distance function with respect to different common statistical problems as well as problems arising as a consequence of zero values in the output quantities.

---

Géraldine Henningsen  
DTU Management Engineering  
Technical University of Denmark  
Fredensborgvej 399  
DK-4000 Roskilde  
Tel.: +45-46775159  
E-mail: [gehe@dtu.dk](mailto:gehe@dtu.dk)

Arne Henningsen  
Department of Food and Resource Economics  
University of Copenhagen  
Rolighedsvej 25  
DK-1958 Frederiksberg C  
Tel.: +45-35332274  
E-mail: [arne@ifro.ku.dk](mailto:arne@ifro.ku.dk)

Uwe Jensen  
Institute for Statistics and Econometrics  
University of Kiel  
Olshausenstrasse 40  
D-24118 Kiel  
Tel.: +49-431-8801627  
E-mail: [jensen@stat-econ.uni-kiel.de](mailto:jensen@stat-econ.uni-kiel.de)

Although our results show clear reactions to some statistical misspecifications, on average none of the approaches is clearly superior. However, considerable differences are found between the estimates at single replications. Taking average efficiencies from both approaches gives clearly better efficiency estimates than taking just the OD or the SR. In the case of zero values in the output quantities, the SR clearly outperforms the OD with observations with zero output quantities omitted and the OD with zero values replaced by a small positive number (ODz).

**Keywords** Multiple Outputs · SFA · Monte Carlo Simulation · Stochastic Ray Production Frontier · Output Distance Function

**JEL Classification:** C21 · C40 · D24

## 1 Introduction

Input and output distance functions and their parametric estimation form, foremost the stochastic frontier function (SFA), are widely applied instruments to measure productivity when technical inefficiency is taken into account. In the case of multiple outputs—given that the underlying production technologies differ significantly—it is common to use a dual approach and to estimate either a cost function, a profit function, or a system of shadow price equations. In cases where standard economic assumptions such as cost minimisation or profit maximisation do not hold, e.g. in some public sector services, or if price data are not available or unvarying, a primal approach to estimate multiple output production functions is an attractive option.

The difficulty with the estimation of output distance functions for multiple outputs is that there is no natural choice of a dependent variable (?). Therefore, it is common to select one output quantity, say  $y_M$ , as the dependent variable and to use the normalised other output quantities  $y_m/y_M$  as explanatory variables (in addition to the inputs). Alternatively, ? suggested using the output norm  $\|y\|$  as the dependent variable and the correspondingly normalised output quantities  $y_m/\|y\|$  as explanatory variables. Taking the latter approach into consideration, ?? introduced a further somewhat different concept to handle multiple outputs in the same framework by introducing the multiple-output stochastic ray frontier production function. In this specification, the simple ratios  $y_m/\|y\|$  are replaced by polar coordinates, i.e. replacing the Cartesian coordinates by polar coordinates. To the authors' knowledge, these different normalisation approaches have only been compared using real world data (e.g. ???), and although efficiency estimates showed considerable deviations (e.g. ? find mean deviations of up to 22 %), it is still unclear which approach performs better.

In order to get to the bottom of these empirical findings, we compare the performance of the classical normalisation approach with one output with the multiple-output stochastic ray frontier approach by means of a Monte Carlo simulation. We test the reaction of both approaches given several common data problems, e.g. endogeneity of the regressors, heteroscedasticity of the inefficiency term and noise term, or zero values in the output data of some observations.

The article is structured as follows: section two provides a short overview of the two normalisation concepts compared in the Monte Carlo simulation; section three describes the data generating process and the design of the Monte Carlo simulation; section four presents and discusses the results; and finally section five concludes.

## 2 Estimating multiple output distance functions in a primal approach

Following ? , chap. 3.2.3., the output distance function for multiple outputs can be estimated by applying the approach used in Stochastic Frontier Analysis (SFA). Starting from the SFA in the single-output case  $y = f(x; \beta) \cdot \exp\{v - u\}$ , where  $y$  is the output quantity,  $x$  is a vector of input quantities,  $\beta$  is a corresponding vector of parameters, and  $\exp\{v - u\}$  is an error term decomposed into a noise term  $v$  and an inefficiency term  $u$ . By exploiting the fact that in the single output case  $y/f(x) = \delta(x, y)$ , with  $\delta(x, y)$  the Shepardian distance function, one can rewrite the stochastic frontier model for the multiple-output case as

$$1 = \delta(x, y) \cdot \exp\{u - v\} \quad (1)$$

As  $\delta(x, y) \leq 1$ ,  $\exp\{u - v\} \geq 1$  can be used as a reciprocal measure of technical efficiency. There are now two possibilities to convert equation (1) into an estimable regression model:

1. By utilising the property of homogeneity of degree one in outputs  $\delta(x, \lambda y) = \lambda \delta(x, y) \forall \lambda > 0$  and setting  $\lambda$  to  $y_M^{-1}$  (e.g. ??), one yields  $\delta(x, y/y_M) = y_M^{-1} \cdot \delta(x, y)$  which leads to  $\delta(x, y) = y_M \cdot \delta(x, y/y_M)$ . By inserting the last equation into equation (1) and dividing by  $y_M$  the final estimation equation is denoted by

$$y_M^{-1} = \delta\left(x, \frac{y}{y_M}\right) \cdot \exp\{u - v\} \quad (2)$$

2. Another alternative is the stochastic ray production frontier developed by ?. Multiple outputs are modelled by decomposing the vector of  $M$  output quantities  $y = \|y\| \cdot p(\vartheta)$  into a scalar distance component, the Euclidean distance  $\|y\| = (\sum_{m=1}^M y_m^2)^{1/2}$ , and a vector of directional measures  $p(\vartheta)$  with  $\vartheta = (\vartheta_1, \dots, \vartheta_{M-1})$  a vector of polar coordinates where  $\vartheta_m \in [0, \pi/2]^{M-1} \forall m = 1, \dots, M-1$  and  $\sin(\vartheta_0) = \cos(\vartheta_M) = 1$ , and with  $p : [0, \pi/2]^{M-1} \rightarrow [0, 1]^M$  a function which transforms the polar-coordinate angle vector  $\vartheta$  to the output-mix vector  $p(\vartheta) = y/\|y\|$ , with norm  $\|p(\vartheta)\| = 1$ . The directional vector is measured as

$$p_m(\vartheta) = \frac{y_m}{\|y\|} = \cos(\vartheta_m) \prod_{j=0}^{M-1} \sin(\vartheta_j) \quad \forall m = 1, \dots, M$$

and  $\vartheta$  is recursively defined by

$$\vartheta_m(y) = \arccos\left(y_m / \left[\|y\| \prod_{j=0}^{M-1} \sin \vartheta_j\right]\right) \quad \forall m = 1, \dots, M.$$

The output distance function can then be expressed as

$$\omega(x, y) = \|y\| / \|f(x, \vartheta)p(\vartheta)\| = \|y\| / f(x, \vartheta)$$

with  $f(x, \vartheta) = \sup(\|y\| \mid \|y\| \cdot p(\vartheta) \in P(x))$ , with  $P(x)$  the output set defined by the technology  $P : x \rightarrow y$ . Inserting the upper definition into equation (1) yields

$$\|y\| = f(x, \vartheta(y)) \cdot \exp\{v - u\} \quad (3)$$

Expressing the output distance functions in (2), and (3) in parametric form by applying the flexible Translog form for  $n = 1, \dots, N$  inputs and  $m = 1, \dots, M$  outputs gives the following specifications

$$\begin{aligned} -\ln(y_M) = & \alpha_0 + \sum_{m=1}^{M-1} \alpha_m \ln(y_m/y_M) + 0.5 \sum_{m=1}^{M-1} \sum_{j=1}^{M-1} \alpha_{mj} \ln(y_m/y_M) \ln(y_j/y_M) \\ & + \sum_n \beta_n \ln(x_n) + 0.5 \sum_n \sum_l \beta_{nl} \ln(x_n) \ln(x_l) \\ & + \sum_{m=1}^{M-1} \sum_n \gamma_{mn} \ln(y_m/y_M) \ln(x_n) + u - v \end{aligned} \quad (4)$$

and

$$\begin{aligned} \ln(|y|) = & \alpha_0 + \sum_{m=1}^{M-1} \alpha_m \vartheta_m + 0.5 \sum_{m=1}^{M-1} \sum_{j=1}^{M-1} \alpha_{mj} \vartheta_m \vartheta_j + \sum_n \beta_n \ln(x_n) \\ & + 0.5 \sum_n \sum_l \beta_{nl} \ln(x_n) \ln(x_l) + \sum_{m=1}^{M-1} \sum_n \gamma_{mn} \vartheta_m \ln(x_n) - u + v, \end{aligned} \quad (5)$$

where  $\alpha_0$  is a scalar intercept,  $\alpha = [\alpha_m]$ ,  $m = 1, \dots, M-1$  and  $\beta = [\beta_n]$ ,  $n = 1, \dots, N$  are vectors of first-order parameters and  $A = [\alpha_{mj}]$ ,  $m, j = 1, \dots, M-1$ ,  $B = [\beta_{nl}]$ ,  $n, l = 1, \dots, N$ , and  $C = [\gamma_{mn}]$ ,  $m = 1, \dots, M-1$ ,  $n = 1, \dots, N$  are matrices of second-order parameters, where  $A$  and  $B$  are symmetric, i.e.  $\alpha_{mj} = \alpha_{jm} \forall m, j = 1, \dots, M-1$  and  $\beta_{nl} = \beta_{ln} \forall n, l = 1, \dots, N$ . As in the usual single-output stochastic production frontier, it is assumed that the noise term  $v$  follows a normal distribution with mean 0 and variance  $\sigma_v^2$  and the inefficiency term  $u$  follows a positively truncated normal distribution (unless other distributional forms are assumed).

### 3 Data generating process and design of the Monte Carlo simulation

It is apparent that the functional forms defined in equation (4) and (5) are not nested into each other. By choosing either equation (4) or (5) as the data generating process (DGP) one would discriminate against the other functional form. Therefore, following ? both functional forms, (4) and (5), are chosen for the DGP to test the robustness of either specification.

We run the simulation with a simple setting of two input variables  $x_n$  with  $n = 1, 2$ , and two output variables  $y_m$  with  $m = 1, 2$ . We follow the procedure suggested by ? to generate the data for the Monte Carlo (MC) simulation. In the first step, the parameters of functions (4) and (5) are chosen. The parameters of function (4) are taken from ? as a baseline scenario. These parameters—given that the explanatory variables are within specific ranges—fulfil the regularity conditions of the output distance function, i.e. homogeneous of degree one in outputs; convex and continuous in outputs and quasi-convex in inputs; and non-decreasing in  $y$  and non-increasing in  $x$ . With this set of parameters, the technology exhibits increasing returns to scale (IRS), whilst input-output separability (IOS) is not fulfilled. Further sets of parameters are chosen so that input-output separability (IOS) is fulfilled, i.e.  $\gamma_{mn} = 0 \forall m = 1, \dots, M-1, n = 1, \dots, N$ , and/or the technology exhibits constant returns to scale (CRS), i.e.  $\sum_n \beta_n = 1$  and  $\sum_n \beta_{nk} = 0 \forall k = 1, \dots, N$ . The parameters of these specifications as well as the parameters of the Translog ray production frontier (5) are chosen in a way that the levels and—as far as possible—the first and second derivatives are equal at the

**Table 1** Variation of DGP specifications

Nr	Approach	Form
1.1	eq. (2)	Translog (4), CRS, IOS
1.2		Translog (4), CRS, no IOS
1.3		Translog (4), IRS, IOS
1.4		Translog (4), IRS, no IOS
1.5	eq. (3)	Translog (5), CRS, IOS
1.6		Translog (5), CRS, no IOS
1.7		Translog (5), IRS, IOS
1.8		Translog (5), IRS, no IOS

sample mean among all 8 specifications. The eight specifications regarding the parameters are summarised in Table 1.

Secondly, following ?, the input quantities are sampled from a uniform distribution over the interval  $[5, 50]$ . This ensures that the regularity conditions are fulfilled, because they are fulfilled—given the chosen parameters—if the logarithmic input ratios lie in the interval  $|\ln x_2 - \ln x_1| \leq 2.5$ . The logarithmic output ratios are sampled from a uniform distribution in a way that they lie in the interval  $|\ln y_2 - \ln y_1| \leq 1.5$ . Thirdly, the inefficiency terms  $u$  are sampled from a half-normal distribution  $u \sim |N(0, \sigma_u^2)|$ . The noise terms  $v_m$  are sampled from a normal distribution  $v_m \sim N(0, \sigma_v^2)$ . Given the generated input quantities and output ratios and the chosen parameters, the “deterministic” fully efficient output quantities  $y^*$  are calculated using equation (4) and equation (5). The sampled output ratios enter  $\vartheta$  in equation (5) as

$$\vartheta_1 = \arccos \left( \frac{y_1}{\sqrt{y_1^2 + y_2^2}} \right) = \arccos \left( \frac{1}{\sqrt{1 + \exp(2(\ln(y_1) - \ln(y_2)))}} \right) \quad (6)$$

with  $\ln(y_1) - \ln(y_2)$  being the sampled logarithmic output ratios. Finally, the noise term and the inefficiency term are subjoined to the  $y^*$  in order to obtain the “observed” output quantities:

$$y_m = y_m^* \exp(v - u). \quad (7)$$

A total population of 2,500 observations is generated based on randomly drawn variables as described above. Then, in each replication of the Monte Carlo simulation, a new sample of 25, 100, or 200 observations is drawn from the population and used for the estimations.

We impose the following specifications on the basic setting to test the robustness of both approaches:<sup>1</sup>

1. Variation of sample size:

As the quality of the estimates varies with the sample size, we use three different sample sizes: 25, 100, and 200 observations.

2. Different ratios of the standard errors of the error terms:

$\sigma_u^2$  is set to 0.05 and 0.8 so that the average “true” efficiencies are around 86% and 56%, respectively. With  $\sigma_v^2 = 0.05$ , the ratio  $\sigma_u/\sigma_v$  is equal to 1 and 4, respectively (?).

<sup>1</sup> In earlier versions of the analysis we also tested other scenarios, i.e. omitted variables, multicollinearity, different distributions of the inefficiency term, different variances of the noise term, and using a CES-CET multiple-output production function as a data generating process instead of a standard Translog output distance function or a Translog ray production function. Both specifications performed equally well in all these scenarios.

3. Different distributions of the noise term:  
the noise term is simulated either with a normal distribution  $v \sim N(0, \sigma_v^2)$  or with a  $t$ -distribution  $v \sim t(0, \sigma_v^2, 15)$ .
4. Correlation of the output ratios with the noise term  $v$  and the inefficiency term  $u$ :  
A potential problem with the estimation form of the output distance function is that the output ratios appear as regressors in the estimation equation. This could lead to inconsistent parameter estimates, as the output ratios might well be endogenous regressors. For instance, this happens when inefficiency and noise affect the different outputs differently (?). Inconsistent estimates of the model parameters will in turn have an impact on the estimation of the efficiency term  $u$  and might lead to the under or overestimation of the efficiency. As this problem can have different occurrences given the individual observations, this effect will both inflict damage on the ranking as well as the level of the individual efficiencies. To test whether the endogeneity of the output-regressors has an impact on the estimation of the efficiency term, our MC simulation includes scenarios with different noise terms and inefficiency terms for the two output quantities.
5. Impact of returns to scale and input-output separability:  
Following ? non-constant returns to scale and missing input-output separability aggravate endogeneity problems. Therefore, the technology is modelled with constant returns to scale and variable returns to scale, as well as with fulfilled and unfulfilled input-output separability.
6. Heteroscedasticity of the noise term and inefficiency term:  
Following ? we either impose heteroscedasticity of the noise term  $v^*$  by  $v^* = v \cdot \exp\{\delta_0 + \delta_1 x_1 + \delta_2 x_2\}$ , or heteroscedasticity of the inefficiency term  $u^*$  by  $u^* = u \cdot \exp\{\delta_0 + \delta_1 x_1 + \delta_2 x_2\}$ , where  $\delta_0$ ,  $\delta_1$ , and  $\delta_2$  are chosen so that the mean of  $\exp\{\delta_0 + \delta_1 x_1 + \delta_2 x_2\}$  is approximately one.
7. Zero output quantities:  
Finally, to test the sensitivity of both specifications towards zeros in the output values, an increasing share of zero valued output observations is introduced into the data, i.e. 0.05, 0.1, 0.3. As it is not possible to generate zero valued outputs with the equation (4), the DGP with zero output quantities will be entirely based on equation (5).

As both approaches use different specifications, a direct comparison of parameter estimates is futile. Therefore, following Coelli and Perelman (2000), the comparison of the quality of the estimates is limited to the efficiency estimates of both approaches.<sup>2</sup> For the

<sup>2</sup> We calculate the efficiency estimates as  $E[\exp(-u)]$  (?). In cross-sectional settings—as in our simulation study—stochastic frontier approaches cannot produce consistent efficiency estimates (see e.g. ?). Therefore, it could be argued that we could estimate the models by corrected OLS (COLS) rather than by maximum likelihood (SFA). In cross-sectional settings, the performance of both COLS and SFA deteriorates with decreasing inefficiency-to-noise ratio  $\lambda = \sigma_u/\sigma_v$ . However, the simulation study of ?, p. 230/231 showed that the relative performance is a bit more complicated. Regarding absolute differences to the ‘true’ inefficiencies of the DGP, SFA outperforms COLS uniformly in the face of  $\lambda \leq 1$ , because COLS treats the—then relatively large—amount of noise as inefficiency. With larger  $\lambda$ , SFA needs larger sample sizes  $n$  to win this competition against the very simple COLS approach, e.g.  $\lambda = 8$  requires  $n \geq 500$ . Regarding inefficiency ranks (correlations with or absolute differences to the true ranks in the DGP), the performances of SFA and COLS are mostly rather similar. This difference in results between inefficiency levels and inefficiency ranks is due to the advantage of the SFA in estimating the constant (or its attempt to decompose the composed error into noise and inefficiency). As SFA panel data models can provide consistent efficiency estimates, we considered extending our simulation study to panel data, but we abstained from this extension, because many different panel data specifications for SFA models exist (see e.g. ?) so we would either have had to arbitrarily choose a specific panel data specification or our simulation study would have extended to an analysis of the many different panel data specifications for SFA models, which is clearly beyond the scope of our current study.

comparison between the true and the estimated efficiencies, we calculated the following performance measures for each replication:

1. the median absolute deviation (MAD) between the estimated and the true efficiencies:  $\text{median}(|\hat{e}_i - e_i|)$ , where  $\hat{e}_i$  is the estimated efficiency,  $e_i$  is the true efficiency, and subscript  $i$  indicates the observation.<sup>3</sup>
2. the average bias of the estimated efficiencies:  $\text{nObs}^{-1} \sum_i (\hat{e}_i - e_i)$ , where nObs is the number of observations.
3. the Spearman rank correlation coefficient between the estimated and the true efficiencies (e.g. ???), as the ranking of the efficiencies is often of more interest than the absolute values.

Finally, we calculated the average values of these performance measures over all replications within one scenario or within a group of scenarios.

Additionally, we generated WAIRDIPs (Weighted Absolute Inefficiency Rank Difference Plot), where the absolute differences in the ranking of the true and estimated efficiencies are averaged over replications and weighted with the sample size (?).

## 4 Results

The Monte Carlo experiment was conducted in the statistical programming language “R” (?) using the add-on package “frontier” (?) for the stochastic frontier estimations. The simulation includes 594 scenarios, and we conducted 500 replications per scenario.

First we look at the scenarios in which all output quantities are strictly positive so that the standard Translog output distance function (4) can be applied to all the observations without modifications. The median absolute deviations, the average biases, and the rank correlation coefficients between the estimated efficiencies and the “true” efficiencies are presented in Tables 2–4. These performance measures are presented as mean values over all 576 scenarios (288,000 replications) and as mean values over scenarios with specific properties of the data generating process—for efficiency estimates derived from the standard Translog output distance function (4), abbreviated as “OD,” for efficiency estimates derived from the Translog stochastic ray production frontier (5), abbreviated as “SR”, and for the arithmetic means of the efficiency estimates derived from the OD and the SR, abbreviated as “am”. Furthermore, Tables 2–4 present P-values for the tests of significant differences between different properties of the data generating process (obtained by an ANOVA with interaction terms) and P-values for two-sided “paired  $t$ -tests” for differences in the average performance of the efficiencies obtained by OD, SR, and am.

The functional form (OD or SR) that is used in the data generating process has only a very small influence on the precision of the estimated efficiencies. When the output quantities are generated by the OD, the estimated efficiencies are slightly more precise and less biased, but the ranking is less precise, regardless of whether the efficiencies are estimated by the OD or the SR. Although non-constant returns to scale and missing input-output separability should theoretically result in inconsistent estimates (?), we found that these properties of the technology did not have an effect on the precision of the estimated efficiencies.

<sup>3</sup> We also calculated the mean absolute deviation (MD) between the estimated and the true efficiencies (as, e.g., done by ?):  $\text{nObs}^{-1} \sum_i |\hat{e}_i - e_i|$ , where nObs is the number of observations. However, we do not present the calculated MDs, because the relative performance of the estimators according to the MDs was very similar to the relative performance of the estimators according to the MADs.



**Table 2** Median absolute deviations

	mean values			P values		
	OD	SR	am	OD-SR	OD-am	SR-am
all scenarios	0.1277	0.1278	0.1267	0.4053	0.0000	0.0000
OD	0.1274	0.1274	0.1264	0.2132	0.0000	0.0000
SR	0.1281	0.1281	0.1271	0.9583	0.0000	0.0000
P-value	0.0080	0.0232	0.0085			
CRS	0.1277	0.1277	0.1267	0.9941	0.0000	0.0000
IRS	0.1277	0.1278	0.1268	0.2414	0.0000	0.0000
P-value	0.8532	0.5928	0.6348			
IOS	0.1279	0.1279	0.1269	0.8292	0.0000	0.0000
non-IOs	0.1275	0.1276	0.1265	0.3346	0.0000	0.0000
P-value	0.1261	0.1908	0.1246			
no heterosced.	0.1260	0.1256	0.1250	0.0003	0.0000	0.0000
heterosc. in $u$	0.1238	0.1237	0.1228	0.2277	0.0000	0.0000
heterosc. in $v$	0.1333	0.1339	0.1325	0.0000	0.0000	0.0000
P-value	0.0000	0.0000	0.0000			
$v$ norm. distr.	0.1265	0.1262	0.1254	0.0001	0.0000	0.0000
$v$ t-distributed	0.1289	0.1293	0.1281	0.0000	0.0000	0.0000
P-value	0.0000	0.0000	0.0000			
$u$ & $v$ the same	0.1187	0.1187	0.1176	0.3581	0.0000	0.0000
$u$ & $v$ different	0.1367	0.1368	0.1359	0.7167	0.0000	0.0000
P-value	0.0000	0.0000	0.0000			
$\sigma_u^2 = 0.05$	0.0983	0.0985	0.0974	0.0000	0.0000	0.0000
$\sigma_u^2 = 0.8$	0.1571	0.1570	0.1561	0.4183	0.0000	0.0000
P-value	0.0000	0.0000	0.0000			
nObs = 25	0.1809	0.1810	0.1786	0.7172	0.0000	0.0000
nObs = 100	0.1074	0.1074	0.1069	0.1389	0.0000	0.0000
nObs = 200	0.0948	0.0948	0.0947	0.6041	0.0000	0.0000
P-value	0.0000	0.0000	0.0000			

While heteroscedasticity in the noise term  $v$  clearly reduces the precision of the efficiency estimates, heteroscedasticity in the inefficiency term  $u$  reduces the rank correlation coefficient, but also reduces the median absolute deviation and the bias, which indicates that heteroscedasticity in the noise term  $v$  is more severe than heteroscedasticity in the inefficiency term  $u$ .

As the estimation assumes normally distributed noise terms, generating the noise terms from a  $t$ -distribution clearly reduces the precision of the estimates. Similarly, when the noise term and the inefficiency term differ between the two output quantities, the precision of the efficiency estimates deteriorates, as this causes an endogeneity bias (?). This endogeneity problem has a very large effect on the (average) bias. Indeed, the over-estimation of the efficiencies almost disappears when this problem does not occur.

The variance of the inefficiency term  $u$  has a major influence on the performance of the efficiency estimates. While a larger variance of the inefficiency term  $u$ , and hence smaller average efficiencies, increases the median absolute deviation, the mean absolute deviation and the bias of the estimated efficiencies, it considerably improves the estimated ranking of the efficiencies. However, a larger variance of the inefficiency term  $u$  not only improves the ranking efficiencies, but also the *relative* precision of the estimated *inefficiencies*, measured by dividing the median absolute deviation, the mean absolute deviation and the bias by the average inefficiency in each replication.

Finally, the number of observations clearly has a significant influence on the precision of the efficiency estimates, whereas 25 observations are apparently insufficient to obtain reasonably precise efficiency estimates.

**Table 3** Average biases

	mean values			P values		
	OD	SR	am	OD-SR	OD-am	SR-am
all scenarios	0.0708	0.0704	0.0706	0.0000	0.0000	0.0000
OD	0.0696	0.0693	0.0695	0.0009	0.0009	0.0009
SR	0.0719	0.0715	0.0717	0.0000	0.0000	0.0000
P-value	0.0000	0.0000	0.0000			
CRS	0.0707	0.0703	0.0705	0.0000	0.0000	0.0000
IRS	0.0708	0.0705	0.0706	0.0007	0.0007	0.0007
P-value	0.8719	0.7230	0.7928			
IOS	0.0709	0.0705	0.0707	0.0000	0.0000	0.0000
non-IOS	0.0706	0.0703	0.0705	0.0022	0.0022	0.0022
P-value	0.4293	0.6809	0.5414			
no heterosced.	0.0772	0.0765	0.0769	0.0000	0.0000	0.0000
heterosc. in $u$	0.0602	0.0592	0.0597	0.0000	0.0000	0.0000
heterosc. in $v$	0.0748	0.0755	0.0752	0.0000	0.0000	0.0000
P-value	0.0000	0.0000	0.0000			
$v$ norm. distr.	0.0681	0.0674	0.0677	0.0000	0.0000	0.0000
$v$ t-distributed	0.0734	0.0735	0.0734	0.7318	0.7318	0.7318
P-value	0.0000	0.0000	0.0000			
$u$ & $v$ the same	0.0303	0.0299	0.0301	0.0000	0.0000	0.0000
$u$ & $v$ different	0.1113	0.1110	0.1111	0.0019	0.0019	0.0019
P-value	0.0000	0.0000	0.0000			
$\sigma_u^2 = 0.05$	0.0366	0.0367	0.0366	0.0973	0.0973	0.0973
$\sigma_u^2 = 0.8$	0.1049	0.1041	0.1045	0.0000	0.0000	0.0000
P-value	0.0000	0.0000	0.0000			
nObs = 25	0.1328	0.1315	0.1322	0.0000	0.0000	0.0000
nObs = 100	0.0419	0.0419	0.0419	0.2453	0.2453	0.2453
nObs = 200	0.0375	0.0378	0.0377	0.0000	0.0000	0.0000
P-value	0.0000	0.0000	0.0000			

While the average precision of the efficiency estimates clearly depends on the properties of the DGP, it is only marginally affected by the choice between the OD and the SR. Furthermore, the OD and the SR react in the same way to the modifications of the DGP. Although the average differences in the performance measures between OD and SR are statistically significant for several scenarios, they are so small—mostly 0.0005 or smaller—that the economic implications are negligible. We do not find that one of the approaches is clearly superior to the other with respect to the median absolute deviation and the bias, but our results indicate that efficiency rankings returned by OD are slightly more reliable than efficiency rankings returned by SR for practically all scenarios. Furthermore, it seems that the OD is slightly more robust regarding misspecifications in the two-sided error term  $v$ .

However, our results do not contradict previous empirical studies that found considerable differences in the efficiency estimates between the OD and SR (e.g. ???), because our results reported above only indicate that *on average* the *precision* of the efficiency estimates does not differ to a noteworthy extent between the OD and the SR. This is illustrated for a typical scenario in Figure 1. In some replications of our Monte Carlo simulation, the precision of the OD and the SR is similar (circles close to the horizontal line at MAD OD - MAD SR = 0), but the efficiency estimates considerably differ between the OD and the SR (MAD OD - MAD SR  $\gg$  0). In other replications, the OD clearly outperforms the SR (MAD OD - MAD SR < 0), while in further replications it is the other way round (MAD OD - MAD SR > 0) so that the average performance is the same for the OD and the SR (mean of MAD OD - MAD SR = 0). The average differences of the efficiency estimates between the OD and the SR (measured in terms of the median absolute deviation, and the rank correlation

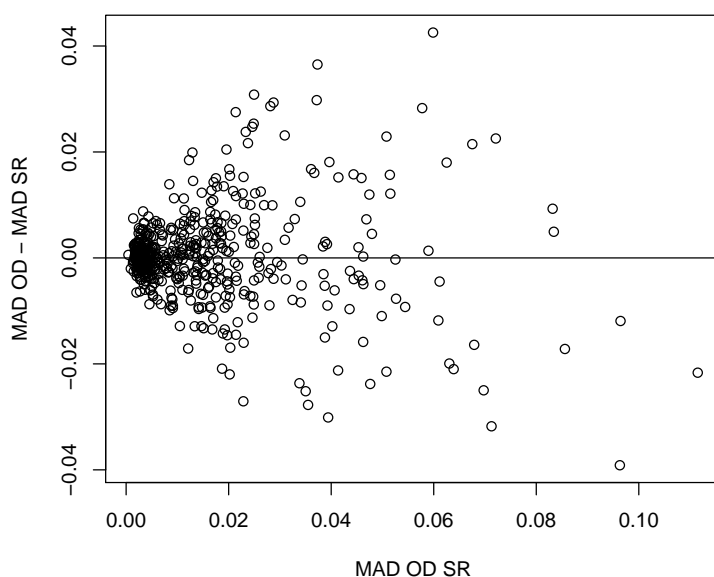
**Table 4** Rank correlation coefficients

	mean values			P values		
	OD	SR	am	OD-SR	OD-am	SR-am
all scenarios	0.5777	0.5772	0.5791	0.0000	0.0000	0.0000
OD	0.5751	0.5743	0.5763	0.0000	0.0000	0.0000
SR	0.5802	0.5800	0.5819	0.0473	0.0000	0.0000
P-value	0.0000	0.0000	0.0000			
CRS	0.5777	0.5772	0.5791	0.0000	0.0000	0.0000
IRS	0.5776	0.5772	0.5791	0.0000	0.0000	0.0000
P-value	0.9209	0.9601	0.9534			
IOS	0.5776	0.5772	0.5791	0.0004	0.0000	0.0000
non-IOs	0.5777	0.5771	0.5791	0.0000	0.0000	0.0000
P-value	0.7451	0.7938	0.8836			
no heterosced.	0.5951	0.5950	0.5966	0.2240	0.0000	0.0000
heterosc. in $u$	0.5755	0.5753	0.5775	0.1496	0.0000	0.0000
heterosc. in $v$	0.5623	0.5613	0.5633	0.0000	0.0000	0.0000
P-value	0.0000	0.0000	0.0000			
$v$ norm. distr.	0.5843	0.5843	0.5860	0.6236	0.0000	0.0000
$v$ t-distributed	0.5710	0.5700	0.5722	0.0000	0.0000	0.0000
P-value	0.0000	0.0000	0.0000			
$u$ & $v$ the same	0.5946	0.5946	0.5961	0.8729	0.0000	0.0000
$u$ & $v$ different	0.5607	0.5598	0.5621	0.0000	0.0000	0.0000
P-value	0.0000	0.0000	0.0000			
$\sigma_u^2 = 0.05$	0.3999	0.4001	0.4007	0.0184	0.0000	0.0000
$\sigma_u^2 = 0.8$	0.7554	0.7543	0.7575	0.0000	0.0000	0.0000
P-value	0.0000	0.0000	0.0000			
nObs = 25	0.4840	0.4822	0.4870	0.0000	0.0000	0.0000
nObs = 100	0.6140	0.6139	0.6148	0.0249	0.0000	0.0000
nObs = 200	0.6349	0.6355	0.6354	0.0000	0.0000	0.0429
P-value	0.0000	0.0000	0.0000			

between the efficiency estimates of the OD and the SR) are particularly significant when  $v$  is affected by either heteroscedasticity or non-normal distribution, in which case OD performs slightly better than SR. For an empirical production analyst who only has a single data set, it would therefore be desirable to know whether the OD or the SR specification is more suitable for his or her specific application. We investigated three potential criteria for choosing between the OD and the SR specification: (i) the spread of the ratio between the output quantities (measured as the variance of the logarithmic ratio of the output quantities), (ii) the log-likelihood values obtained by the estimations of the OD and the SR specification, and (iii) the test statistic from a (generalised) Regression Equation Specification Error Test (RESET) with squared and cubic fitted values as suggested by ? applied to the OD and the SR specifications. Unfortunately, none of these three criteria indicated whether the OD or the SR performs better when the true efficiencies are unknown.<sup>4</sup> However, an empirical production analyst could estimate both the OD and the SR and take the average efficiencies from both approaches. According to MADs and rank correlation coefficients,<sup>5</sup> these average efficiencies outperform both the OD and the SR approach in virtually all scenarios.

<sup>4</sup> For instance, in each replication, choosing the specification that gives the larger log-likelihood value or the lower  $\chi^2$  value of ?'s (1969) RESET does not result in performance measures that are significantly better than always taking the OD specification or always taking the SR specification.

<sup>5</sup> The average bias of the average efficiencies from both approaches, i.e.  $n\text{Obs}^{-1} \sum_i ((\hat{e}_i^{OD} + \hat{e}_i^{SR})/2 - e_i)$ , is equal to the mean of the two average biases of the two approaches, i.e.  $(n\text{Obs}^{-1} \sum_i (\hat{e}_i^{OD} - e_i) + n\text{Obs}^{-1} \sum_i (\hat{e}_i^{SR} - e_i))/2$ , and thus, cannot simultaneously outperform both approaches.



**Fig. 1** Relationship between the difference of the MAD of the OD and the MAD of the SR (MAD OD - MAD SR) and the MAD between the OD and the SR (MAD OD SR) for the scenario, where the DGP is an OD with CRS and IOS, no heteroscedasticity, the same  $u$  and  $v$  for both outputs, normally distributed  $v$ ,  $\sigma_u^2 = 0.8$ , and 100 observations

However, in the presence of zero output quantities, the OD and the SR do not generally perform equally well. As the OD specification cannot model zero output quantities, we estimated the OD only with the observations that have strictly positive output quantities. Additionally, we estimated the OD specification with a data set, where all zero values in the output quantities were replaced by a small positive number, in our case 0.01 (specification “ODz”). Although this approach has no theoretical foundation and the choice of the replacement value is arbitrary (see e.g. ?), it is still frequently used to deal with zero values in Translog functions. Tables 5, 6, and 7 report the results for the median absolute deviations, the average biases, and the Spearman rank correlation coefficients, respectively. Taking into account the fact that the OD can only estimate efficiencies of observations that have strictly positive output quantities, the comparisons between OD and ODz as well as between OD and SR are only based on efficiencies of observations with strictly positive output quantities. The efficiencies of all observations are used to compare the ODz to the SR.

The MAD of the OD and the ODz significantly increase in tact with the share of zeros in the output values. This stands in stark contrast to the SR which is virtually unaffected by the share of zeros. However, although both OD and ODz are significantly less robust in the presence of zero values, with the ODz having a slight advantage over the OD, in absolute terms, the differences are again rather small. We find comparable results for the influence of the share of zeros on the bias. While the SR is unaffected by the share of zero values, the bias of the OD significantly increases and—surprisingly—the bias of the ODz significantly decreases with the share of zeros. All three specifications show a tendency to overestimate the true inefficiency, especially when the sample size is small. However, on closer examination, for some scenarios the positive bias of the ODz turns to a distinct

**Table 5** Median absolute deviations for scenarios with zero output quantities

	only observations with non-zero output quantities				all observations			
	mean values		P values		mean values		P values	
	OD	ODz	SR	OD-ODz	OD-SR	ODz-SR	SR	ODz-SR
all scenarios	0.1193	0.1146	0.1115	0.0000	0.0000	0.1154	0.1115	0.0000
share = 0.05	0.1131	0.1133	0.1109	0.8014	0.0043	0.1141	0.1116	0.0014
share = 0.1	0.1156	0.1110	0.1107	0.0000	0.0000	0.1123	0.1109	0.1535
share = 0.3	0.1292	0.1196	0.1128	0.0000	0.0000	0.1198	0.1120	0.0000
P-value	0.0000	0.0000	0.3947			0.0000	0.7925	
share = 0.05, $\sigma_u^2 = 0.05$	0.0985	0.1012	0.0977	0.0002	0.0843	0.1010	0.0979	0.0000
share = 0.1, $\sigma_u^2 = 0.05$	0.0979	0.0997	0.0970	0.0295	0.1117	0.1000	0.0971	0.0001
share = 0.3, $\sigma_u^2 = 0.05$	0.1036	0.1014	0.0983	0.0361	0.0000	0.1011	0.0972	0.0000
share = 0.05, $\sigma_u^2 = 0.8$	0.1276	0.1254	0.1242	0.1606	0.0158	0.1271	0.1253	0.1934
share = 0.1, $\sigma_u^2 = 0.8$	0.1332	0.1224	0.1243	0.0000	0.0000	0.1246	0.1248	0.9007
share = 0.3, $\sigma_u^2 = 0.8$	0.1548	0.1378	0.1272	0.0000	0.0000	0.1385	0.1269	0.0000
P-value	0.0000	0.0001	0.7781			0.0002	0.7479	
share = 0.05, nObs = 25	0.1648	0.1656	0.1612	0.7334	0.0919	0.1663	0.1624	0.0578
share = 0.1, nObs = 25	0.1707	0.1582	0.1602	0.0000	0.0001	0.1599	0.1609	0.7007
share = 0.3, nObs = 25	0.1960	0.1736	0.1644	0.0000	0.0000	0.1753	0.1646	0.0013
share = 0.05, nObs = 100	0.0945	0.0929	0.0928	0.1087	0.0055	0.0939	0.0932	0.4922
share = 0.1, nObs = 100	0.0950	0.0938	0.0926	0.2573	0.0024	0.0952	0.0926	0.0020
share = 0.3, nObs = 100	0.1048	0.0993	0.0933	0.0001	0.0000	0.0988	0.0923	0.0000
share = 0.05, nObs = 200	0.0799	0.0813	0.0788	0.0179	0.0008	0.0821	0.0792	0.0000
share = 0.1, nObs = 200	0.0811	0.0810	0.0791	0.8260	0.0000	0.0817	0.0792	0.0000
share = 0.3, nObs = 200	0.0867	0.0860	0.0806	0.3955	0.0000	0.0853	0.0793	0.0000
P-value	0.0000	0.0668	0.9256			0.0215	0.8614	

**Table 6** Average biases for scenarios with zero output quantities

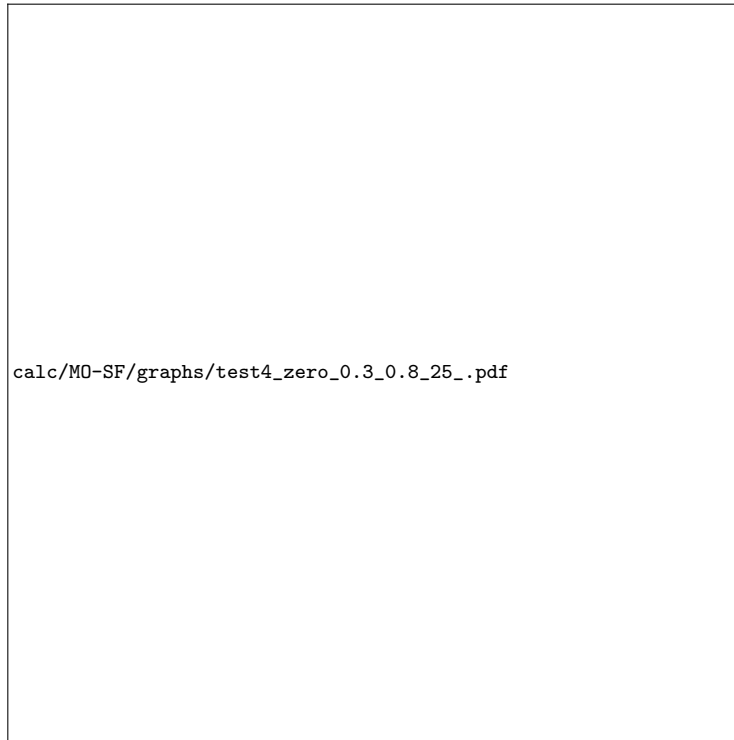
	only observations with non-zero output quantities				all observations			
	mean values		P values		mean values		P values	
	OD	ODz	SR	OD-ODz	OD-SR	ODz	SR	ODz-SR
all scenarios	0.0508	0.0368	0.0396	0.0000	0.0000	0.0370	0.0412	0.0000
share = 0.05	0.0421	0.0423	0.0397	0.8658	0.0138	0.0440	0.0408	0.0035
share = 0.1	0.0437	0.0346	0.0375	0.0000	0.0000	0.0367	0.0391	0.0743
share = 0.3	0.0665	0.0334	0.0417	0.0000	0.0000	0.0302	0.0437	0.0000
P-value	0.0000	0.0013	0.2550			0.0000	0.1809	
share = 0.05, $\sigma_u^2 = 0.05$	0.0297	0.0356	0.0297	0.0002	0.9767	0.0367	0.0305	0.0000
share = 0.1, $\sigma_u^2 = 0.05$	0.0283	0.0303	0.0273	0.2944	0.4891	0.0315	0.0282	0.0634
share = 0.3, $\sigma_u^2 = 0.05$	0.0448	0.0225	0.0329	0.0000	0.0000	0.0190	0.0323	0.0000
share = 0.05, $\sigma_u^2 = 0.8$	0.0545	0.0491	0.0497	0.0039	0.0039	0.0513	0.0511	0.8804
share = 0.1, $\sigma_u^2 = 0.8$	0.0592	0.0389	0.0476	0.0000	0.0000	0.0420	0.0500	0.0001
share = 0.3, $\sigma_u^2 = 0.8$	0.0882	0.0444	0.0505	0.0000	0.0000	0.0413	0.0551	0.0000
P-value	0.0009	0.0387	0.8490			0.0720	0.9069	
share = 0.05, nObs = 25	0.1080	0.1058	0.1019	0.4946	0.0290	0.1099	0.1036	0.0250
share = 0.1, nObs = 25	0.1151	0.0922	0.0973	0.0000	0.0000	0.0988	0.0992	0.9036
share = 0.3, nObs = 25	0.1628	0.1077	0.1075	0.0000	0.0000	0.1083	0.1098	0.7259
share = 0.05, nObs = 100	0.0109	0.0115	0.0100	0.7017	0.3012	0.0130	0.0111	0.1820
share = 0.1, nObs = 100	0.0099	0.0073	0.0087	0.1757	0.3496	0.0085	0.0104	0.2101
share = 0.3, nObs = 100	0.0245	-0.0112	0.0115	0.0000	0.0000	-0.0055	0.0137	0.0000
share = 0.05, nObs = 200	0.0075	0.0097	0.0072	0.0495	0.5473	0.0092	0.0077	0.1499
share = 0.1, nObs = 200	0.0063	0.0043	0.0065	0.1285	0.8022	0.0030	0.0077	0.0000
share = 0.3, nObs = 200	0.0122	-0.0063	0.0062	0.0000	0.0000	-0.0123	0.0076	0.0000
P-value	0.0000	0.0004	0.5508			0.0003	0.5272	

Table 7 Rank correlation coefficients for scenarios with zero output quantities

	only observations with non-zero output quantities				all observations			
	mean values		P values		mean values		P values	
	OD	ODz	SR	OD-ODz	OD-SR	ODz-SR	SR	ODz-SR
all scenarios	0.5917	0.6101	0.6142	0.0000	0.0000	0.6073	0.6165	0.0000
share = 0.05	0.6113	0.6205	0.6201	0.0000	0.0000	0.6106	0.6168	0.0000
share = 0.1	0.5999	0.6140	0.6138	0.0000	0.0000	0.6028	0.6123	0.0000
share = 0.3	0.5639	0.5958	0.6088	0.0000	0.0000	0.6086	0.6203	0.0000
P-value	0.0000	0.0000	0.0004			0.0079	0.0081	
share = 0.05, $\sigma_u^2 = 0.05$	0.4398	0.4404	0.4449	0.6629	0.0000	0.4345	0.4425	0.0000
share = 0.1, $\sigma_u^2 = 0.05$	0.4252	0.4286	0.4359	0.0576	0.0000	0.4233	0.4381	0.0000
share = 0.3, $\sigma_u^2 = 0.05$	0.4002	0.4281	0.4329	0.0000	0.0000	0.4417	0.4472	0.0006
share = 0.05, $\sigma_u^2 = 0.8$	0.7827	0.8006	0.7953	0.0000	0.0000	0.7866	0.7911	0.0013
share = 0.1, $\sigma_u^2 = 0.8$	0.7746	0.7994	0.7916	0.0000	0.0000	0.7823	0.7864	0.0132
share = 0.3, $\sigma_u^2 = 0.8$	0.7275	0.7634	0.7847	0.0000	0.0000	0.7755	0.7934	0.0000
P-value	0.0006	0.0000	0.6327			0.0000	0.8731	
share = 0.05, nObs = 25	0.4978	0.5223	0.5161	0.0000	0.0000	0.5044	0.5088	0.1240
share = 0.1, nObs = 25	0.4803	0.5189	0.5093	0.0000	0.0000	0.4911	0.5009	0.0034
share = 0.3, nObs = 25	0.4035	0.5022	0.5051	0.0000	0.0000	0.5115	0.5173	0.0906
share = 0.05, nObs = 100	0.6553	0.6607	0.6608	0.0000	0.0000	0.6517	0.6588	0.0000
share = 0.1, nObs = 100	0.6438	0.6511	0.6523	0.0000	0.0000	0.6454	0.6534	0.0000
share = 0.3, nObs = 100	0.6278	0.6342	0.6500	0.0001	0.0000	0.6469	0.6600	0.0000
share = 0.05, nObs = 200	0.6807	0.6785	0.6834	0.0000	0.0000	0.6756	0.6828	0.0000
share = 0.1, nObs = 200	0.6756	0.6720	0.6797	0.0000	0.0000	0.6719	0.6825	0.0000
share = 0.3, nObs = 200	0.6602	0.6509	0.6713	0.0000	0.0000	0.6674	0.6835	0.0000
P-value	0.0000	0.7829	0.9273			0.0020	0.1798	

negative bias, i.e. a tendency to underestimate the true efficiency. Furthermore, while the positive bias of the SR decreases with increasing sample size, the ODz clearly overshoots the mark for larger samples.

Surprisingly, the rank correlation of the inefficiencies for all three specifications is significantly affected by the presence of zero values. However, while the ODz and the SR seem to perform equally "badly", the deviation is not of the same magnitude as in the case of the OD, as is exemplified in figure 2.



**Fig. 2** WAIRDIP for the scenario, where the DGP is an SR with 30% of zeros in the output values, normally distributed  $v$ ,  $\sigma_v^2 = 0.8$ , and 25 observations; OD (—), ODz(···), SR (- - -)

The performance of the simple solution used in ODz shows that—despite the differences between ODz and SR being small in absolute terms—our results indicate that the SR should be the first choice for empirical applications with zero output quantities. Table 8 presents the performance of the ODz with three different replacement values (0.0001, 0.01, 1) as well as the performance of the SR based on average values over all scenarios. Although choosing different replacement values produces significantly different efficiency estimates (which supports the findings of ?), the essence of the outcome is unaltered as the SR in all cases performs as least as well as the ODz. In our simulation study, replacing zero output quantities by ones resulted on average in a significantly higher rank correlation than replacing zero output quantities by 0.01 or 0.0001. However, this result should not be generalised, because it depends on the units of measurement of the output variables and it is likely specific for the data generating process that we chose for our simulation study.



**Table 8** Different replacement values for zero output quantities

	mad	bias	rc
mean value: SR	0.1115	0.0412	0.6165
mean value: ODz (0.0001)	0.1158	0.0361	0.6050
mean value: ODz (0.01)	0.1154	0.0370	0.6073
mean value: ODz (1)	0.1145	0.0400	0.6107
P-value: SR - ODz (0.0001)	0.0000	0.0000	0.0000
P-value: SR - ODz (0.01)	0.0000	0.0000	0.0000
P-value: SR - ODz (1)	0.0000	0.0848	0.0000
P-value: ODz (0.0001) - ODz (0.01)	0.0378	0.0013	0.0000
P-value: ODz (0.0001) - ODz (1)	0.0002	0.0000	0.0000
P-value: ODz (0.01) - ODz (1)	0.0029	0.0000	0.0000

## 5 Conclusion

We compared the standard Translog output distance (OD) function (?) with the Translog stochastic ray (SR) production frontier (?) by means of a Monte Carlo simulation, exposing both specifications to common data problems. The results indicate that if a considerable share of output quantities is zero, the SR clearly outperforms the OD with observations with zero output quantities omitted and the OD with zero values replaced by a small positive number (ODz). However, in the case of strictly positive output quantities, on average both specifications perform rather equally robustly; the OD produces marginally more reliable efficiency rankings and seems to be slightly more capable of handling misspecifications in the two-sided error term  $v$ . Nevertheless, our results are in line with earlier empirical findings that show considerable differences between the inefficiency estimates of the OD and the SR at individual estimations. We investigated three potential criteria for choosing between the OD and the SR specification in specific empirical applications (spread of the ratio between the output quantities, log-likelihood values of the estimations of the OD and the SR models, test statistics from ?'s 1969 RESET test), but unfortunately, none of these criteria indicated whether the OD or the SR performs better when the true efficiencies are unknown. A first step could be to take the average efficiencies obtained by OD and SR, which generally outperforms both the OD and the SR. However, future research should focus on finding data-based criteria that indicate whether the OD or SR approach is more appropriate for a given data set.