Sensitivity analysis of efficiency rankings to distributional assumptions: applications to Japanese water utilities

Yane, Shinji and Berg, Sanford
St. Andrew's University-Momoyama Gakuin University , University of Florida

2011

Sensitivity Analysis of Efficiency Rankings to Distributional Assumptions: Applications to Japanese Water Utilities

Shinji YANE[i]
Professor of Economics, St. Andrew's University
Momoyama Gakuin University,
Osaka, Japan

Sanford Berg (corresponding author e-mail: sberg @ ufl.edu)
Distinguished Service Professor, Economics
University of Florida
Gainesville, FL 32611

Abstract: This paper examines the robustness of efficiency score rankings across four distributional assumptions for trans-log stochastic production-frontier models, using data from 1,221 Japanese water utilities (for 2004 and 2005). One-sided error terms considered include the half-normal, truncated normal, exponential, and gamma distributions. Results are compared for homoscedastic and doubly heteroscedastic models, where we also introduce a doubly heteroscedastic variable mean model, and examine the sensitivity of the nested models to a stronger heteroscedasticity correction for the one-sided error component. The results support three conclusions regarding the sensitivity of efficiency rankings to distributional assumptions. When four standard distributional assumptions are applied to a homoscedastic stochastic frontier model, the efficiency rankings are quite consistent. When those assumptions are applied to a doubly heteroscedastic stochastic frontier model, the efficiency rankings are consistent when proper and sufficient arguments for the variance functions are included in the model. When a more general model, like a variable mean model is estimated, efficiency rankings are quite sensitive to heteroscedasticity correction schemes.

Running Title: Sensitivity of Efficiency Rankings to Distributional Assumptions: Japanese Water Utilities

Keywords: stochastic production frontier models, Japanese water utilities, heteroscedasticity

[i] Yane initiated and completed most of this research as Visiting Scholar at the Public Utility Research Center (University of Florida), 2009-2010. Berg is PURC Director of Water Studies. Both are grateful to PURC for providing support. Chunrong Ai provided useful comments on earlier drafts but is absolved from remaining errors.

Sensitivity Analysis of Efficiency Rankings to Distributional Assumptions: Applications to Japanese Water Utilities

1. Introduction

Efficient frontier techniques, including both stochastic frontier analysis (SFA) and data environment analysis (DEA), are widely used to identify high and low performing organizations. The application of sophisticated yardstick comparisons and associated benchmarking incentive schemes can improve efficiency.[1] However, as Kumbhakar and Lovell (2000, p.90) conclude that, even within a parametric approach, " . . . it is unclear whether a ranking of producers by their efficiency scores is sensitive to distributional assumptions, although it is clear that sample mean efficiencies are sensitive." Since a distributional assumption is essential for SFA, especially in the context of cross-sectional models, this empirical problem presents issues for the application of efficiency scores in the context of benchmarking. The purpose of this paper is to examine the sensitivity of efficiency rankings to distributional assumptions regarding the one-sided efficiency error term for SFA.

In his analysis of stochastic cost frontiers for 123 U.S. electric utilities, Greene (1990, p.157) used four types of models where one-sided error components are assumed, using half normal, truncated normal, exponential, and gamma distributions. The reported sample mean (in)efficiencies are 0.8839 (0.1234), 0.9013 (0.1039), 0.9058 (0.0989) and 0.9002 (0.1051) respectively. Based on these results, Green (pp. 155-8) also concluded that the frontier parameter estimates were roughly similar for the four models; however, the gamma model yielded a different inefficiency distribution.

Kumbhakar and Lovell (2000, p.90) used the same data and calculated the correlation coefficients for rankings; the highest was 0.9803, between the half normal and truncated normal models, whereas the lowest was 0.7467 between the exponential and gamma models. These correlations suggest that rankings can be somewhat sensitive to distributional assumptions.

Greene (2008, p.182) also presents new results based on the same data but on a full translog model; he concludes that mean inefficiency estimates are almost identical, although there are differences in the parameter estimates. The reported sample mean (in)efficiencies are 0.9240 (0.0790), 0.9281 (0.0746), 0.9279 (0.0748) and 0.9368 (0.0653) respectively. Hence, in contrast with the initial conclusion by Kumbhakar and Lovell (2000), the mean efficiency scores no longer seem to be sensitive to distributional assumptions in the translog case. In fact, the lowest correlation coefficient is 0.9116 between the half normal and gamma models. In the context of ranking correlations, the highest is 0.9999, between the truncated normal and exponential models, and the lowest is 0.9554 between the half normal and gamma models. These new results suggest that not only efficiency rankings but also mean efficiencies are consistent among different assumed distributions. Thus, Greene (2008, p.114) concludes that the overall pictures drawn by SFA and DEA are similar, although the evidence is mixed due to different efficiency evaluations of financial institutions (the industry from which data were obtained). Here, we will focus on consistency within SFA models, where different error distribution assumptions are considered.

As Greene argues (2008, p.180), the issue of robustness to different error distribution assumptions does not have an analytical solution. However, it is useful to explore the extent of consistency of efficiency scores (and utility rankings) under different distributional assumptions, since that can provide sign-posts for analysts conducting performance studies. Furthermore, the reported correlations are derived from a homoscedastic frontier model. That model (which neglects heteroscedasticity) faces serious problems in the context of SFA. Previous empirical studies conclude that estimated parameters and efficiency scores are sensitive to specification of the one-sided (inefficiency) error component and/or the two-sided (idiosyncratic) error component. A number of approaches have been suggested to address these problems: Caudill, Ford and Gropper (1995) use a half-normal one-sided heteroscedastic frontier model; Hadri (1999) and Hadri, Guermat and Whittaker (2003) develop a half-normal doubly heteroscedastic frontier model; Greene (2004, and 2005a,b) applies a truncated-normal heterogeneous mean model as well as true fixed or random model; and Wang and Schmidt (2002) and Alvarez et al. (2006) propose scaling-function models. To the extent that correcting for heteroscedasticity affects estimates of frontier parameters and efficiency scores, an appropriate heteroscedasticity correction presents a serious technical issue. Unless the sensitivity to specification is addressed, the policy-relevance of estimates will be called into question.

Therefore, it is useful to examine the consistency among heteroscedastic frontier models that have different distributional assumptions. In the present study, we combine the above mentioned four types of distributional assumptions with homoscedastic and doubly heteroscedastic stochastic production-frontier models, utilizing a sample of 1,221 Japanese water utilities, pooled for two years. Here, the dispersion in the size distribution of utilities suggests that the homogeneity assumption is violated. Thus, we also introduce a doubly heteroscedastic variable mean model, and examine the sensitivity of nested models to a more comprehensive heteroscedasticity correction for the one-sided error component.

Our estimated results suggest three possibilities regarding the sensitivity of efficiency ranking is sensitive to distributional assumptions. When we apply the four types of distributional assumptions to a homoscedastic stochastic frontier model, an efficiency ranking will be clearly consistent. When they apply them to a doubly heteroscedastic stochastic frontier model, analysts will be able to make an efficiency ranking consistent whenever they can find proper and sufficient arguments for the variance functions. When a more general model, like a variable mean model, is estimated, the efficiency ranking is quite sensitive to heteroscedasticity correction schemes. In general, controlling for heteroscedasticity is very important for efficiency rankings; getting the correct specification of the heteroscedasticity form is just as important. Therefore one must conduct sensitivity tests before making policy recommendations. If results are sensitive to the error specification, one must use a more flexible specification, such as nonparametric specification for the heteroscedasticity.

The remainder of the paper is organized as follows. In Section 2, we briefly describe our data and models, and present estimates of parameters, mean efficiencies and efficiency rankings of the homoscedastic translog production-frontier models with different distributional assumptions. In Section 3, we show the corresponding results of doubly heteroscedastic frontier models with different distributional assumptions. We also examine estimates of three nested models which consists of a doubly heteroscedastic half-normal, truncated-normal and variable mean models when we increase significant arguments for the one-sided error component. The last section presents some implications of the study.

2. Homoscedastic Stochastic Production-Frontier Models

Data and Models

We use two-year pooled data which consists of 2,442 observations (1,221 utilities) in the Japanese water industry in fiscal years 2004 and 2005. The data are from *Annual Statistics of*

Public Enterprises (*Chihou Kouei-Kigyou Nenkan).* The largest single cost items except for capital and labor expenditures are outsourcing and purchased water expenditures. In addition, we use intake water capacity as a proxy for actual intake water volume because we have only intake water volume but not purchased water volume. Since the correlation between intake water volume and capacity is high (0.99), we use "purchasing water capacity" plus "other intake water capacity." We also calculate the number of virtual staff based on outsourcing by dividing outsourcing expenditures by payment per employee in each prefecture. Then our output and input variables for a production function are defined as follows:

Y: total delivered water volume in a year (1,000 m^3)

K: length of all pipes (1,000 m)

L: total number of staff, including estimated number of staff from outsourcing

O: intake water volume without purchased water volume (1,000 m^3)

P: purchased water volume (1,000 m^3)

Table 1 summarizes descriptive statistics and it shows that our data exhibit considerable size dispersion.[2]

Table 1: Descriptive Statistics of 2442 Observations in FY 2004-05

Variable	Skewness	Kurtosis	S.D.	Mean	Min	Median	Max
Y	22	597	55,295	12,313	222	3,922	1,624,602
K	15	330	1,017	443	17	224	25,914
L	20	504	314	68	1	19	8,876
O	23	657	84,245	14,881	0	4,282	2,586,888
P	12	198	21,649	5,784	0	77	404,137

As Greene (2008, p.181) suggests, consistency is also affected by the functional form adopted. Thus, we use a translog production function rather than a restricted Cobb-Douglas function.

When we denote each output observation by $\boldsymbol{y_i}$ and inputs K, L, O and P by $\boldsymbol{x^m}$ or $\boldsymbol{x^n}$,for *m,n* = *1(K), 2(L),3(I),4(P),* then our stochastic production-frontier model is written as follows.

$$\ln yi = \alpha + \sum_{m=1}^{4} \beta_m \ln x_i^m + \frac{1}{2} \sum_{m=1}^{4} \sum_{n=1}^{4} \beta_{mn} \ln x_i^m \ln x_i^n + \varepsilon_i\,, \quad \text{where } \boldsymbol{\varepsilon_i = v_i + u_i} \tag{1}$$

$$v_i \sim N(0, \sigma_{vi}^2) \tag{2}$$

$$u_i \sim N^+(\mu_i, \sigma_{ui}^2) \quad \text{or} \quad u_i \sim G(\theta_i, \mathrm{P}) \tag{3}$$

$$e_i = \exp(-E(u_i \mid \varepsilon_i)) \tag{4}$$

The two-sided error component for each utility $\boldsymbol{i}$, $\boldsymbol{v_i}$, and the nonnegative one-sided error component, $\boldsymbol{u_i}$, are assumed to be distributed independently of each other and of the regressors. The technical efficiency of each utility, $\boldsymbol{e_i}$, is measured by the mean of the conditional distribution of $\boldsymbol{u_i}$ given the total error term, $\boldsymbol{\varepsilon_i}$.

The one-sided disturbance is assumed to be a truncated normal or Gamma distribution; assuming homoscedasticity results in a constant term of $\boldsymbol{\sigma_{ui} = \sigma_u}$ or $\boldsymbol{\theta_i = \theta_0}$ in (3) respectively, as well as $\boldsymbol{\sigma_{vi} = \sigma_v}$ in (2). A half normal model is a restricted form of a truncated normal model because $\boldsymbol{\mu_i = 0}$ for all $\boldsymbol{i}$, whereas an exponential model is a special case of a Gamma model when $\boldsymbol{P = 1}$. In addition, a truncated normal model is a restricted form of a variable mean model in the sense that $\boldsymbol{\mu_i = \mu_0}$ for all $\boldsymbol{i}$ and then a half normal, truncated normal and variable mean models are nested.

Homoscedastic Stochastic Production-Frontier Models

Table 2 presents estimates of homoscedastic frontier parameters based on four types of distributional assumptions; half-normal (H), truncated normal (T), exponential (X) and gamma (G) distributions.[3] As expected, the estimated parameters are not substantially different from estimates using ordinary least squares (OLS). The estimates among these four frontier models are much closer to each other than to the OLS estimates, although several estimates of the half normal model are slightly different from others.

Table 2: Homoscedastic Stochastic-Production-Frontier Models

	OLS	Half	Trunc	eXpo	Gamma
Constant	1.9293***	2.0594***	2.0433***	2.0437***	2.0368***
	(0.1927)	(0.1822)	(0.1800)	(0.1805)	(0.1798)
Log(K)	0.2968**	0.3093**	0.3001**	0.2999**	0.2994**
	(0.1045)	(0. 0994)	(0.1002)	(0.0975)	(0.0998)
Log(L)	0.2284**	0. 1626*	0.1686*	0.1688*	0.1698*
	(0.0816)	(0.0772)	(0.0772)	(0.0759)	(0.0767)
Log(O)	0.2654***	0.2916***	0.2845***	0.2844***	0.2837***
	(0.0221)	(0.0206)	(0.0200)	(0.0204)	(0.0199)
Log(P)	0.2468***	0.2769***	0.2718***	0.2717***	0.2711***
	(0.0172)	(0.0159)	(0.0159)	(0.0159)	(0.0159)
L(K)L(K)	-0.0587	-0.0704*	-0.0718*	-0.0717*	-0.0717*
	(0.0320)	(0.0305)	(0.0312)	(0.0298)	(0.0311)
L(L)L(L)	-0.0435	-0.0369	-0.0399	-0.0399	-0.0401
	(0.0232)	(0.0220)	(0. 0206)	(0.0218)	(0.0205)
L(O)L(O)	0.0450***	0.0413***	0.0417***	0.0417***	0.0417***
	(0.0028)	(0.0025)	(0.0022)	(0.0026)	(0.0022)
L(P)L(P)	0.0494***	0.0452***	0.0463***	0.0463***	0.0464***
	(0.0023)	(0.0020)	(0.0018)	(0.0021)	(0.0018)
L(K)L(L)	-0.0167	-0.0182	-0. 0155	-0.0155	-0.0153
	(0.0252)	(0.0240)	(0.0232)	(0.0235)	(0.0232)
L(K)L(O)	0.0177***	0.0216***	0.0227***	0.0227***	0.0227***
	(0.0041)	(0.0039)	(0.0040)	(0.0038)	(0.0040)
L(K)L(P)	0.0148***	0.0189***	0.0188***	0.0187***	0.0187***
	(0.0031)	(0.0029)	(0.0030)	(0.0029)	(0.0030)
L(L)L(O)	0.0158***	0.0195***	0.0185***	0.0185***	0.0184***

	(0.0043)	(0.0040)	(0.0041)	(0.0039)	(0.0041)
L(L)L(P)	0.0229***	0.0259***	0.0250***	0.0250***	0.0250***
	(0.0031)	(0.0028)	(0.0029)	(0.0028)	(0.0029)
L(O)L(P)	-0.0631***	-0.0681***	-0.0678***	-0.0677***	-0.0676***
	(0.0023)	(0.0021)	(0.0020)	(0.0022)	(0.0019)
R^2 / LL	0.9701	380.9279	395.4415	395.4672	395.5657

Standard errors in parentheses. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

The likelihood ratio (LR) test strongly rejects the restriction of the half normal model, but it cannot reject the restriction of the exponential model. Thus, we can say that the estimates of the frontier parameters are roughly similar: only the estimates of the half normal model whose restriction is rejected by the LR test are slightly different. Several Tables provide evidence regarding the consistency of the results. **Table 3** confirms that efficiency estimates are also quite similar for the different error models, except for the half-normal model. In particular, the truncated normal and exponential models have almost the same efficiency distribution, which is the same result found by Greene (2008, p.182).[4] In his earlier work, Greene (1990, p.158) also suggests that a restricted model produces smaller values of estimated efficiencies than a more general model for most of the sample observations: a conclusion that is consistent with our results, shown in Table 3.[5]

Table 3: Estimated Efficiency Distributions from Homoscedastic Frontier Models

Model	Skewness	Kurotsis	S.D.	Mean	Min	Median	Max
Half	-0.9748	3.5862	0.0969	0.8121	0.4905	0.8328	0.9662
Trunc	-2.0816	8.5865	0.0844	0.8671	0.4018	0.8929	0.9681
eXpo	-2.1418	9.1015	0.0846	0.8675	0.3552	0.8934	0.9681
Gamma	-2.2673	9.8223	0.0834	0.8764	0.3589	0.9024	0.9718

Table 4 shows that the lowest correlation coefficient is 0.9603 between the half normal and gamma models, supporting the consistency of estimated efficiency scores for the four error

distribution specifications. None of the efficiency rankings are sensitive to distributional assumptions: the lowest ranking correlation coefficient is 0.999 (between the half normal and gamma models again). Therefore, we can conclude that both efficiency scores and their rankings are consistent among these four types of models.

Table 4: Correlations for Estimated Efficiencies from Homoscedastic Frontier Models[a]

Model	Half	Trunc	eXpo	Gamma
Half	**1**	**0.9697**	**0.9673**	**0.9603**
Trunc	**0.9998**	**1**	**0.9998**	**0.9991**
eXpo	**0.9998**	**1**	**1**	**0.9995**
Gamma	**0.9991**	**0.9993**	**0.9993**	**1**

a) Spearman rank correlations below diagonal and Pearson correlations above diagonal.

However, these high correlations do not necessarily imply a simple linear relationship between efficiency scores. For example, **Figure 1** suggests that the estimated efficiency distribution from the normal half model is convex when compared with the distribution associated with the truncated normal model. Interestingly, the normal half model also takes a similar convex form relative to the exponential and gamma models; except for the half normal model, these three models have a close linear relationship with each other. However, the correlation coefficients of the half normal model are relatively low on Table 4.

Figure 1: Estimated Efficiencies: Half against Trunc

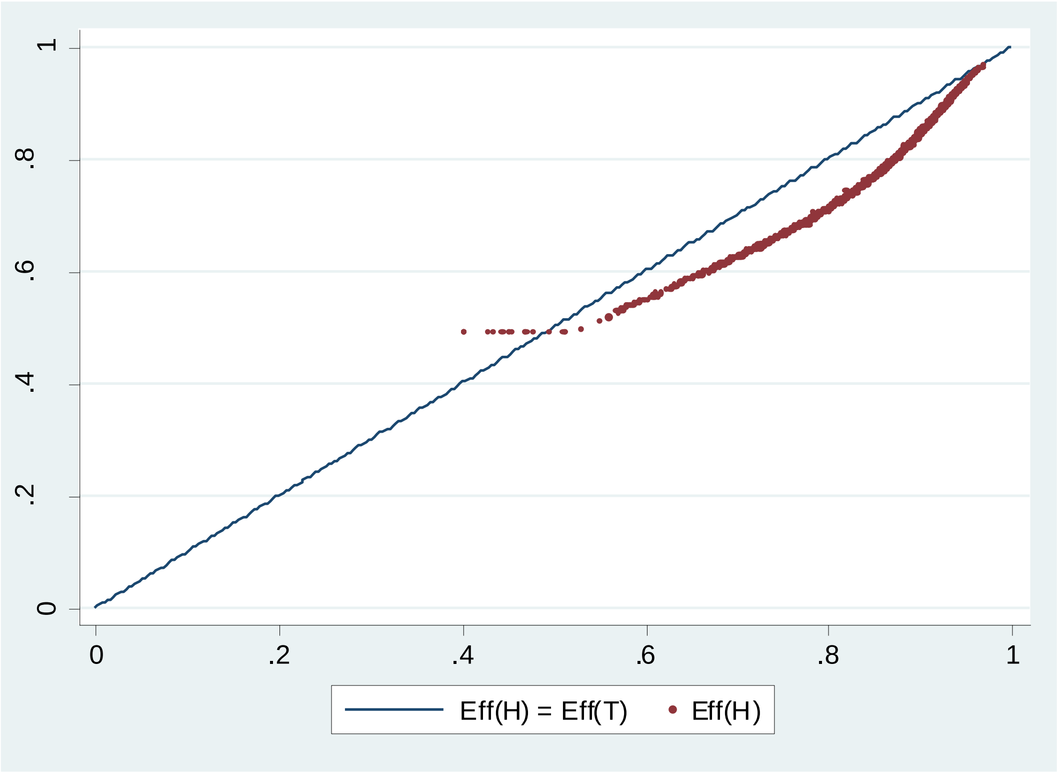
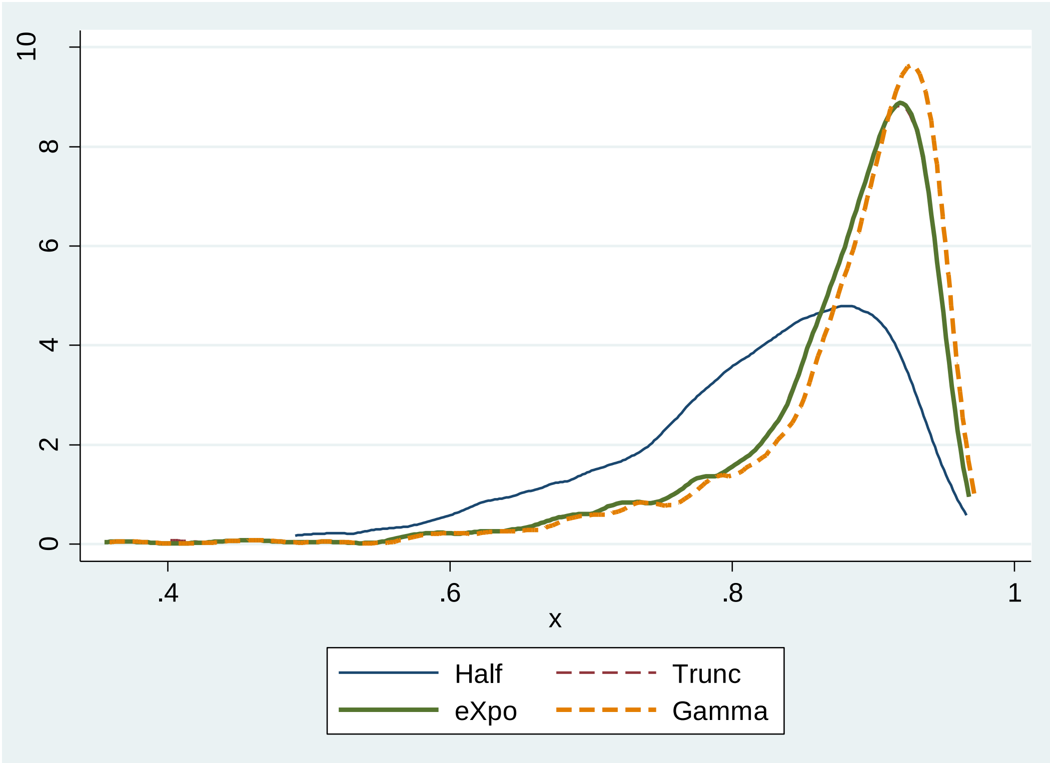


Figure 2 depicts the estimated efficiency distributions for the four homoscedastic models. While the half normal model has a peak at a lower efficiency level, the other three distributions share a long and thin tail on the left side of a relatively higher efficiency peak. Therefore, as was suggested by patterns in Figure 1, the half normal model is apt to be able to distinguish more efficient utilities in some detail; the other models do not have this capability in this case.

Figure 2: Estimated Efficiency Distributions from Four Homoscedastic Models



3. Doubly Heteroscedastic Stochastic Frontier Models

Doubly Heteroscedastic Stochastic Production-Frontier Models

A half normal doubly heteroscedastic model developed by Hadri (1999) and Hadri et al. (2003) allows heteroscedasticity for both error components. A homoscedastic assumption on each error component in the last section can be examined using the likelihood ratio (LR) tests. We can also apply not only half normal model but also other three models by assuming that the two-sided and one-sided error terms take the following multiplicative heteroscedasticity form:[6]

$$\sigma_{vi}^2 = \sigma_v^2 \exp(Z_i^v \gamma) = \exp(\gamma_0 \ln(\sigma_v^2) + Z_i^v \gamma) \qquad (5)$$

$$\sigma_{ui}^2 = \sigma_u^2 \exp(Z_i^u \delta) = \exp(\delta_0 \ln(\sigma_u^2) + Z_i^u \delta) \text{ or } \theta_i = \theta \exp(-Z_i^u \delta) \quad (6)$$

where Z_i^v and Z_i^u are vectors of conventional size-related exogenous variables (like firm size) and efficiency-related environmental variables (like firm management) respectively, and $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$ capture the corresponding unknown parameters respectively. Since we introduce two types of four distributions for the one-sided error term in (3), the heteroscedastic corrections for the half normal and truncated normal models take a different form in the exponential and gamma models as shown in (6).

In this paper, the conventional size-related exogenous variables for the two-sided error component, Z_i^v, are

diwv1-diwv6: size dummy variables, based on intake water volume (diwv1=1 represents the smallest group),

and the efficiency-related environmental variables for the one-sided error component, Z_i^u, are

rraw: a proxy for raw water ratio defined by chemical expenditures per intake water volume,

rout: outsourcing ratio defined by the ratio of the number of staff based on outsourcing to the number of total staff, and

uprice: unit price defined by water supply revenue divided by total billed water volume.

Now we can examine four types of heteroscedastic stochastic production-frontier models for their inefficiency error components: half normal (**H**), truncated normal (**T**), exponential (**X**) and gamma (**G**) distributions. To do so, we estimate a one-sided heteroscedastic model (**u**), a two-sided heteroscedastic model (**v**) and a doubly heteroscedastic model (**uv**) for each type of distributional assumptions. For all types of the models, the likelihood ratio (LR) tests strongly reject the restriction of homoscedasticity for the one-sided and two-sided error components. Thus, we focus on a doubly heteroscedastic model, which is statistically more appropriate than the other three models (including the homoscedastic model introduced in the previous section).

Table 5 presents estimates of doubly heteroscedastic frontier parameters based on four types of distributional assumptions as well as feasible general least squares (**FGLS**) by using the same arguments of z_i^v and z_i^{u1}. **Huv**, **Tuv**, **Xuv** and **Guv** denote doubly heteroscedastic models (**uv**) with half-normal (H), truncated normal (T), exponential (X) and gamma (G) distributions, respectively. The agreement between Huv and Tuv is striking, whereas FGLS estimates seem closer to them than Guv. Since the LR tests strongly reject the restriction of the half normal and exponential models, we can say that the estimates of the frontier parameters are (at most) only roughly similar.

Table 5: Doubly Heteroscedastic Production-Frontier Models

	FGLS	Huv	Tuv	Xuv	Guv
Constant	**1.9911*****	**2.3731*****	**2.3672*****	**2.2949*****	**2.4879*****
	(0.1837)	**(0.1717)**	**(0.1730)**	**(0.1710)**	**(0.1634)**
Log(K)	**0.2655****	**0.2429***	**0.2437***	**0.2505****	**0.2997*****
	(0.0985)	**(0. 0958)**	**(0.0953)**	**(0.0952)**	**(0.0886)**
Log(L)	**0.2377****	**0. 2692*****	**0.2679*****	**0.2496*****	**0.2167****
	(0.0783)	**(0.0740)**	**(0.0735)**	**(0.0735)**	**(0.0704)**
Log(O)	**0.2767*****	**0.2700*****	**0.2697*****	**0.2727*****	**0.2821*****
	(0.0221)	**(0.0188)**	**(0.0188)**	**(0.0188)**	**(0.0184)**
Log(P)	**0.2553*****	**0.2541*****	**0.2541*****	**0.2576*****	**0.2694*****
	(0.0161)	**(0.0143)**	**(0.0143)**	**(0.0144)**	**(0.0143)**
L(K)L(K)	**-0.0599***	**-0.0423**	**-0.0427**	**-0.0481**	**-0.0573***
	(0.0302)	**(0.0298)**	**(0.0296)**	**(0.0296)**	**(0.0276)**
L(L)L(L)	**-0.0477***	**-0.0250**	**-0.0247**	**-0.0302**	**-0.0215**
	(0.0224)	**(0.0195)**	**(0. 0194)**	**(0.0194)**	**(0.0188)**
L(O)L(O)	**0.0461*****	**0.0325*****	**0.0323*****	**0.0335*****	**0.0288*****

	(0.0028)	**(0.0022)**	**(0.0022)**	**(0.0022)**	**(0.0020)**
L(P)L(P)	**0.0467*** **	**0.0404*** **	**0.0403*** **	**0.0413*** **	**0.0364*** **
	(0.0022)	**(0.0016)**	**(0.0016)**	**(0.0016)**	**(0.0015)**
L(K)L(L)	**-0.0144**	**-0.0272**	**-0. 0270**	**-0.0239**	**-0.0225**
	(0.0244)	**(0.0226)**	**(0.0225)**	**(0.0225)**	**(0.0212)**
L(K)L(O)	**0.0180*** **	**0.0200*** **	**0.0201*** **	**0.0207*** **	**0.0206*** **
	(0.0041)	**(0.0039)**	**(0.0039)**	**(0.0038)**	**(0.0037)**
L(K)L(P)	**0.0186*** **	**0.0164*** **	**0.0165*** **	**0.0170*** **	**0.0178*** **
	(0.0028)	**(0.0027)**	**(0.0028)**	**(0.0027)**	**(0.0026)**
L(L)L(O)	**0.0143** **	**0.0170*** **	**0.0170*** **	**0.0178*** **	**0.0197*** **
	(0.0043)	**(0.0041)**	**(0.0041)**	**(0.0041)**	**(0.0041)**
L(L)L(P)	**0.0213*** **	**0.0213*** **	**0.0213*** **	**0.0219*** **	**0.0227*** **
	(0.0028)	**(0.0026)**	**(0.0026)**	**(0.0026)**	**(0.0025)**
L(O)L(P)	**-0.0642*** **	**-0.0602*** **	**-0.0601*** **	**-0.0616*** **	**-0.0614*** **
	(0.0021)	**(0.0017)**	**(0.0018)**	**(0.0018)**	**(0.0017)**
R^2 / LL	**0.9715**	**709.2707**	**714.5512**	**672.3141**	**740.4062**

Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 6 shows that the sample mean efficiencies become considerably different when comparing a restricted model to an unrestricted model; Huv and Xuv have relatively higher mean efficiencies than Tuv and Guv. In contrast with homoscedastic mean efficiencies presented in Table 2, the two unrestricted models indicate lower mean efficiencies among the sample of Japanese water utilities.

Table 6: Estimated Efficiency Distributions from Doubly Heteroscedastic Frontier Models

Model	**Skewness**	**Kurtosis**	**S.D.**	**Mean**	**Min**	**Median**	**Max**
Huv	**-1.7274**	**6.4203**	**0.0987**	**0.8666**	**0.3698**	**0.9002**	**0.9824**
Tuv	**-0.9973**	**4.8775**	**0.0806**	**0.7127**	**0.2972**	**0.7252**	**0.8949**
Xuv	**-2.1102**	**8.2741**	**0.1019**	**0.8876**	**0.3174**	**0.9247**	**0.9911**
Guv	**-0.6056**	**3.2483**	**0.1171**	**0.7111**	**0.2481**	**0.7243**	**0.9533**

Thus, as **Table 7** shows, we have the lowest correlation coefficient of 0.899 between the doubly heteroscedastic exponential (Xuv) and gamma (Guv) models. We conclude that the estimated efficiency scores are moderately consistent, although the correlation coefficient between unrestricted models is fairly high: 0.963.

Table 7: Correlations for Estimated Efficiencies from Doubly Heteroscedastic Frontier Models[a]

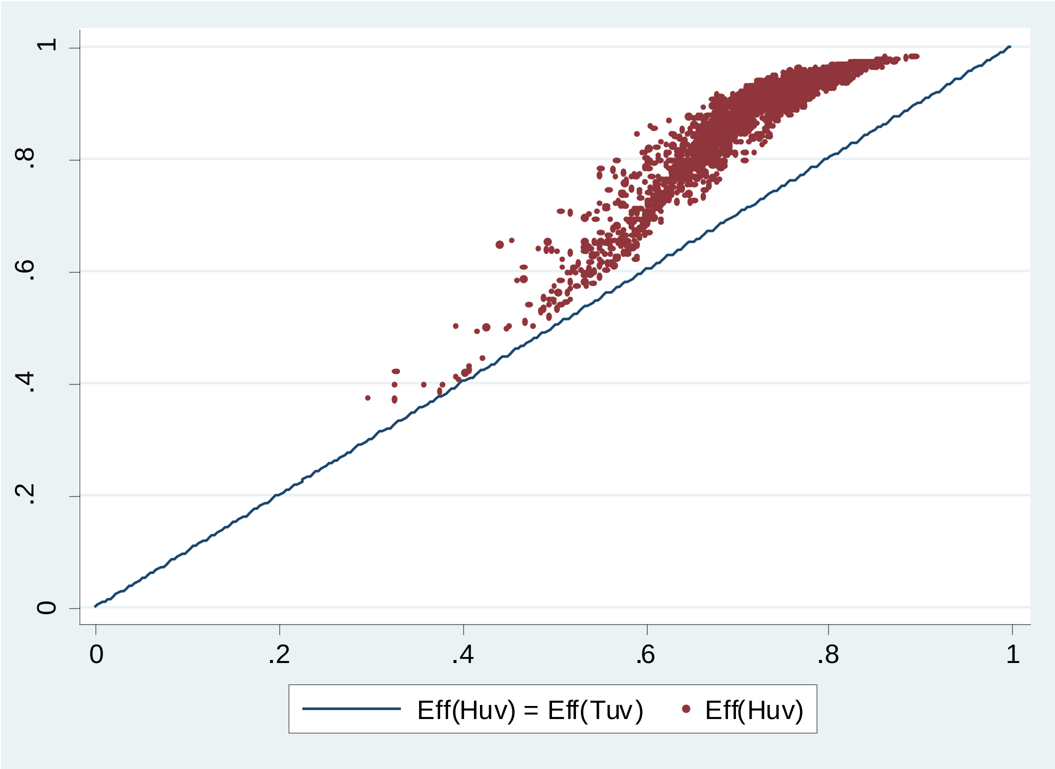
Model	Huv	Tuv	eXuv	Guv
Huv	1	0.9425	0.9878	0.9272
Tuv	0.9506	1	0.9136	0.9630
Xuv	0.9898	0.9170	1	0.8991
Guv	0.9444	0.9537	0.9215	1

a) Spearman rank correlations below diagonal and Pearson correlations above diagonal.

In the context of efficiency rankings, the highest correlation is 0.990 between Huv and Xuv, and the lowest correlation is 0.917 between Tuv and Xuv. Thus we can still maintain a conclusion from the above homoscedastic models; efficiency rankings are consistent among these four types of models.

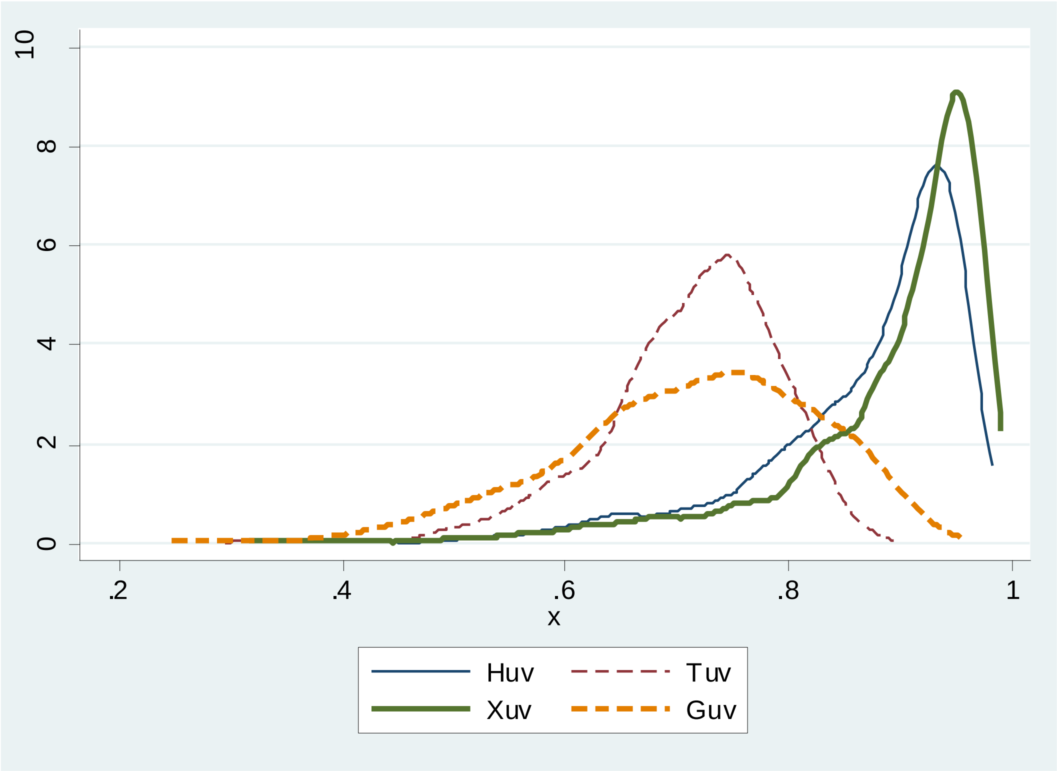
A slight decrease in these correlation coefficients indicates that correcting heteroscedasticity is (to some extent) sensitive to the distributional assumptions. For example, **Figure 3** suggests that the estimated efficiency distribution from the doubly heteroscedastic half normal (Huv) model is now concave rather than convex to that from the doubly heteroscedastic truncated normal (Tuv) model. Interestingly, another restricted Xuv model also takes a similar concave form to another unrestricted Guv model. These results explain why the correlation coefficients between a restricted model and an unrestricted model are relatively lower.

Figure 3: Estimated Efficiencies: Huv against Tuv



Comparing **Figure 4** with **Figure 2**, we can observe that estimated efficiency distributions from both unrestricted models move to the left and become flatter. On the other hand, the efficiency distribution for the half normal model moves to the right and becomes more peaked. Thus, the unrestricted model is now apt to be able to distinguish more efficient utilities in a more precise way, and the restricted models share the opposite pattern.

Figure 4: Estimated Efficiency Distributions from Four Doubly Heteroscedastic Models



A Doubly Heteroscedastic Variable Mean Model and the Nested Models

We further examine the above sensitivity to heteroscedasticity corrections by introducing a doubly heteroscedastic variable mean model. Whereas the half normal and truncated normal models assume $\boldsymbol{\mu_i = 0}$ and $\boldsymbol{\mu_i = \mu_0}$ in (3) respectively, our truncated normal variable mean model has a more flexible functional form:

$$\mu_i = \eta_0 + Z_i^u \eta \tag{7}$$

where Z_i^u is the above defied efficiency-related environmental variables in (6), and $\boldsymbol{\eta}$ captures the corresponding unknown parameters. Thus, these three models are nested.

We also can examine a doubly heteroscedastic Variable Mean (Muv) model by combining (5), (6) and (7): then three models (Huv, Tuv and Muv) are also nested. That is, the models use the same kinds of assumptions, but the extent of the restrictions is different: the Half-normal model is a special case of the Truncated-normal model, and the Truncated-normal model is a special case of the Variable Mean model. In addition, in order for a more comprehensive heteroscedastic correction, we also introduce more arguments, Z_i^{u2}, which is achieved by adding the following efficiency-related environmental variables to Z_i^u:

rsubp: subsidy ratio on profit and loss account defined by the sum of subsidies on profit and loss account per water supply revenue,

aveope: average operation rate defined by average delivered water volume per delivered water capacity,

cusden: customer density defined by the number of customers per the length of all pipes.

Then we can estimate the half normal, truncated normal and variable mean doubly heteroscedastic models when the number of arguments for the one-sided error component increases for a more comprehensive heteroscedastic correction.

Table 8 presents estimates of the frontier parameters as well as the estimates of feasible general least squares (**FGLS**) by using the same arguments: Z_i^{u2} and Z_i^v: **Hsuv**, **Tsuv**, and **Msuv** denote doubly heteroscedastic (uv) models with more explanatory variables, yielding a stronger heteroscedastic correction for half-normal (H) and truncated normal (T) distributions, and a variable mean (M) model, respectively.

Table 8: Doubly Heteroscedastic Production-Frontier Models with s stronger correction

	FGLS	Hsuv	Tsuv	Msuv	Muv
Constant	**2.3188*** **	**2.2129*** **	**2.2089*** **	**2.6134*** **	**3.2664*** **
	(0.1820)	**(0.1406)**	**(0.1408)**	**(0.1484)**	**(0.1608)**
Log(K)	**0.4019*** **	**0.2468** **	**0.2446** **	**0.3003*** **	**0.2199***
	(0.1014)	**(0. 0783)**	**(0.0786)**	**(0.0774)**	**(0.0859)**
Log(L)	**0.1572***	**0. 1516***	**0.1515**	**0.1771** **	**0.3812*** **
	(0.0796)	**(0.0628)**	**(0.0638)**	**(0.0615)**	**(0.0656)**
Log(O)	**0.2159*** **	**0.2969*** **	**0.2977*** **	**0.2653*** **	**0.2250*** **
	(0.0195)	**(0.0155)**	**(0.0155)**	**(0.0138)**	**(0.0167)**
Log(P)	**0.1669*** **	**0.2638*** **	**0.2649*** **	**0.2304*** **	**0.2043*** **
	(0.0141)	**(0.0131)**	**(0.0131)**	**(0.0112)**	**(0.0133)**
L(K)L(K)	**-0.0764** **	**-0.0342**	**-0.0343**	**-0.0338**	**-0.0179**
	(0.0311)	**(0.0251)**	**(0.0253)**	**(0.0226)**	**(0.0262)**
L(L)L(L)	**-0.0477**	**-0.0285**	**-0.0290**	**-0.0328***	**-0.0092**
	(0.0224)	**(0.0172)**	**(0. 0178)**	**(0.0156)**	**(0.0180)**
L(O)L(O)	**0.0558*** **	**0.0402*** **	**0.0401*** **	**0.0279*** **	**0.0238*** **
	(0.0026)	**(0.0018)**	**(0.0018)**	**(0.0015)**	**(0.0020)**
L(P)L(P)	**0.0567*** **	**0.0451*** **	**0.0450*** **	**0.0344*** **	**0.0373*** **
	(0.0021)	**(0.0014)**	**(0.0014)**	**(0.0012)**	**(0.0014)**
L(K)L(L)	**-0.0002*** **	**-0.0347**	**-0. 0344**	**-0.0186**	**-0.0369**
	(0.0246)	**(0.0195)**	**(0.0198)**	**(0.0178)**	**(0.0203)**
L(K)L(O)	**0.0068*** **	**0.0169*** **	**0.0170*** **	**0.0135*** **	**0.0168*** **
	(0.0039)	**(0.0030)**	**(0.0030)**	**(0.0027)**	**(0.0033)**
L(K)L(P)	**0.0165**	**0.0188*** **	**0.0189*** **	**0.0185*** **	**0.0120*** **
	(0.0028)	**(0.0024)**	**(0.0024)**	**(0.0021)**	**(0.0025)**
L(L)L(O)	**0.0109** **	**0.0228*** **	**0.0228*** **	**0.0214*** **	**0.0104** **
	(0.0037)	**(0.0032)**	**(0.0032)**	**(0.0030)**	**(0.0036)**
L(L)L(P)	**0.0131*** **	**0.0250*** **	**0.0250*** **	**0.0185*** **	**0.0132**
	(0.0027)	**(0.0024)**	**(0.0023)**	**(0.0021)**	**(0.0024)**
L(O)L(P)	**-0.0542*** **	**-0.0665*** **	**-0.0666*** **	**-0.0559*** **	**-0.0476**
	(0.0019)	**(0.0017)**	**(0.0017)**	**(0.0015)**	**(0.0017)**
R^2 / LL	**0.9715**	**1087.7965**	**1095.2768**	**1408.0258**	**906.2368**

Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

These models are nested; LR tests strongly reject the restriction of the zero mean and homoscedastic mean. We also add a doubly heteroscedastic variable mean model (**Muv**) based on (7) by using the arguments of only Z_i^u and Z_i^v, which can be compared with heteroscedastic models in **Table 5**. We cannot compare these models with exponential or gamma models because the assumptions are fundamentally different.

Again, the agreement between Hsuv and Tsuv is striking; the frontier parameters are almost identical. On the other hand, estimated parameters from Msuv are not close to those estimated by the other models. Note that the estimated parameters from Muv are not close to those of Huv and Tuv in **Table 5**. Thus, it appears that these differences are mainly caused from the heteroscedastic mean assumption rather than the number of arguments utilized for the one-sided variance function.

In sum, however, we conclude that the estimates of the frontier parameters are not as consistent when we include a more appropriate variable mean statistical model. On the other hand, we can say that an increase in the one-sided error arguments produces more consistent estimates of the frontier parameters.

Table 9 shows that the sample mean efficiencies become much closer by the stronger heteroscedastic correction.

Table 9: Estimated Efficiency Distributions from Doubly Heteroscedastic Frontier Models

Model	Skewness	Kurotsis	S.D.	Mean	Min	Median	Max
Hsuv	-1.3686	4.2516	0.1177	0.8663	0.4698	0.9064	0.9966
Tsuv	-2.0837	8.4576	0.1028	0.8951	0.2959	0.9278	0.9992
Msuv	-1.4309	5.7861	0.0838	0.9011	0.4019	0.9206	0.9999
Muv	-0.2074	3.1323	0.1004	0.5511	0.1827	0.5561	0.8658

Tsuv and Msuv produce especially higher values of efficiencies than Tuv and Muv. **Figure 5** and **Figure 6** capture these movements and indicate the important role of adopting an appropriate heteroscedastic correction. Then, as **Table 10** shows, the correlation coefficients for efficiency scores and their rankings between Hsuv and Tsuv are 0.945 and 0.966, both of which are higher than those between Huv and Tuv. A proper and sufficient heteroscedasticity correction produces increases in the consistency of the efficiency scores and their rankings, as well as consistency in the estimates of the frontier parameters.

Figure 5: Efficiency Distributions for Doubly Heteroscedastic Models (Weaker Correction)

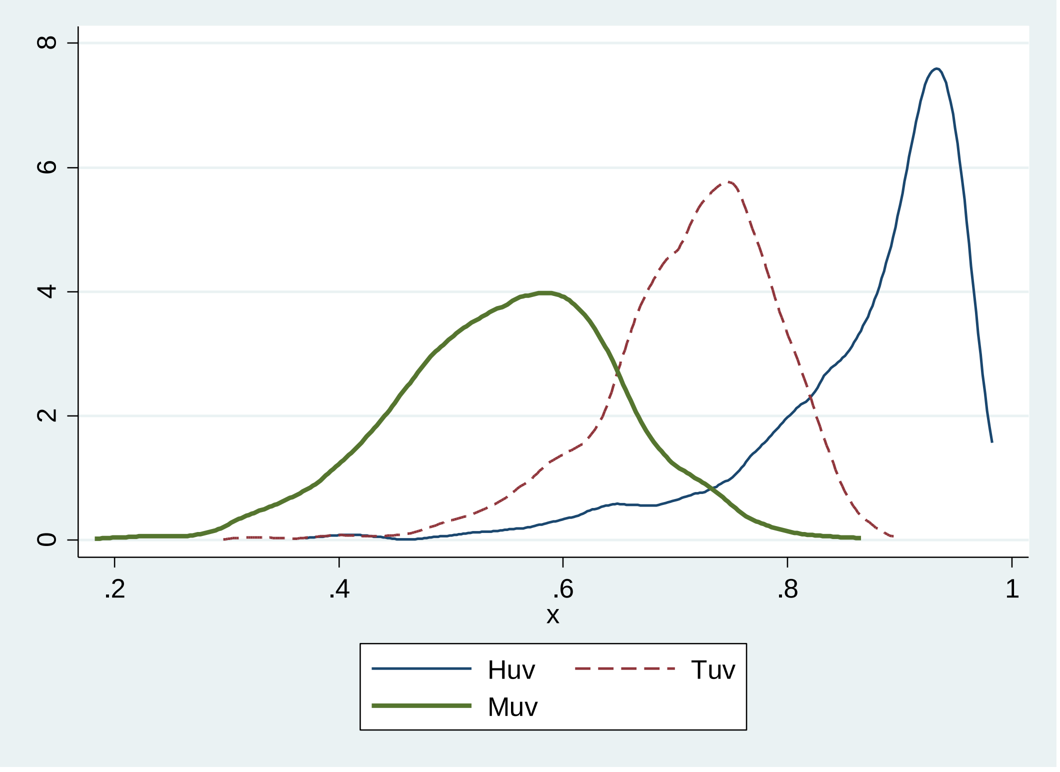
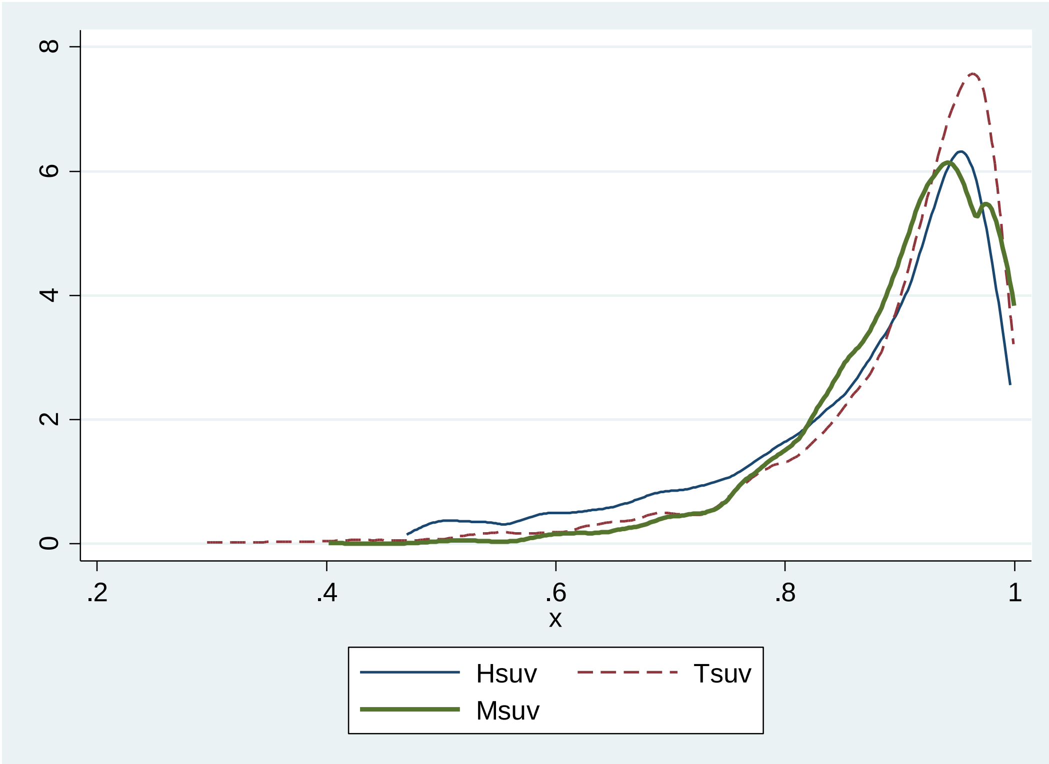


Figure 6: Efficiency Distributions for Doubly Heteroscedastic Models (Stronger Correction)



However, when we include a variable mean model, the lowest correlation coefficient is 0.799 and the lowest rank correlation coefficient is 0.838: between Hsuv and Msuv. Note that these relatively low correlation coefficients are not caused from the heteroscedastic mean assumption itself because estimated efficiencies from Muv are highly correlated with those of Huv, as shown in **Table 10**. The differences are due to the fact that estimated efficiencies from the Variable Mean model are quite sensitive to a stronger heteroscedasticity correction, which is statistically favored among our nested models. Therefore, we can conclude that the estimated efficiency scores and their rankings are only moderately consistent.

Table 10: Correlations for Estimated Efficiencies from Doubly Heteroscedastic Frontier Models[a]

Stronger Correction

Model	Msuv	Tsuv	Hsuv	Trunc
Msuv	1	0.8400	0.7985	0.6603
Tsuv	0.873	1	0.9454	0.8286
Hsuv	0.8376	0.9660	1	0.8952
Trunc	0.6322	0.7395	0.8545	1

Weaker Correction

Model	Muv	Tuv	Huv	Trunc
Muv	1	0.9577	0.9107	0.7727
Tuv	0.9623	1	0.9425	0.8831
Huv	0.9770	0.9506	1	0.8611
Trunc	0.7985	0.8589	0.8040	1

4. Implications

We estimate homoscedastic and doubly heteroscedastic stochastic production-frontier models of the Japanese water industry under four distributional assumptions: half-normal, truncated normal, exponential and gamma distributions. The results for the homoscedastic frontier models support that the view that both efficiency scores and their rankings are consistent among these four types of models; this result is similar that obtained by Greene (2008, p.183).

The four types of doubly heteroscedastic frontier models produce modest improvements: efficiency rankings are still consistent but the efficiency scores themselves are somewhat consistent. These results are in line with conclusions by Kumbhakar and Lovell (2000, p.90), although their observations are based on only a homoscedastic frontier model. We can explain a slight decrease in these correlation coefficients by the different sensitivity of different distributional assumptions used to correct for heteroscedasticity. In particular, unrestricted models produce lower efficiencies than restricted models, and the shifted distributions result in relatively low correlations.

We further examine this sensitivity problem by introducing a doubly heteroscedastic Variable Mean model, increasing the number of statistically significant arguments for the one-sided error component. The half normal, truncated normal and variable mean doubly heteroscedastic models are nested. The likelihood ratio tests reject the restriction of the zero mean and homoscedastic mean. The stronger correction for heteroscedasticity brings greater consistency of estimates for parameters, efficiencies and their rankings between half normal and truncated normal models, whereas it reduces their correlation coefficients with the doubly heteroscedastic variable mean model.

These empirical results suggest three possibilities regarding the sensitivity of efficiency ranking to distributional assumptions. When we apply the four types of distributional assumptions to a homoscedastic stochastic frontier model, an efficiency ranking will be clearly consistent. When we apply them to a doubly heteroscedastic stochastic frontier model, we were able to make an efficiency ranking consistent whenever we can find proper and sufficient arguments for the variance functions. When a more general model, like a variable mean model, is statistically supported, the efficiency ranking is quite sensitive to heteroscedasticity correction schemes.

From the policy-standpoint, the results underscore the point that individual efficiency scores are not necessarily robust with respect to different error specifications, let alone different specifications of the model itself, treatment of outliers, or other elements that can influence the coefficients that determine “expected output” relative to actual output—for given inputs and exogenous conditions. Rather, this analysis of Japanese water utilities reminds us that the decision-relevance of technical benchmarking studies depends on sensible use of the scores (Berg, 2010, p. 115). A regulator setting price caps would have to establish catch-up times for utilities which seem to be lagging in performance—that decision requires judgment and awareness that groupings of firms makes better sense than using individual scores. Similarly, a government ministry determining whether support subsidies are being wasted or used wisely by utilities would want to group firms (say, in quartiles or deciles) so that incentives could be applied in a manner that can be supported by performance patterns (and not individual scores). These observations are not meant to detract from efforts to refine and improve benchmarking—just to remind analysts that humility is called for when so many factors remain beyond managerial control (and outside analytical models).

Endnotes

[1] De White and Marques (2009a, b). See also Davis and Garces (2009, Chapter 3), Coelli and Perelman (2003) and Haney and Pollitt (2009) for practical applications of yardstick comparisons.

[2] In our sample, 203 observations (8.3%) have zero value of O and 1209 observations (49.5%) have zero value of P. Thus we adopt a standard practice, and calculate the log values of O and P by adding one to these original values.

[3] We used LIMDEP (NLOGIT v.4.3) to estimate all of stochastic production-frontier models in this paper.

[4] The results are also similar in that a truncated normal model results in a large variance for the inefficiency error component.

[5] In our homoscedastic case, however, we should recall that the LR test cannot reject the restriction of the exponential model. In addition, a half normal model rather than a gamma model exhibits a different efficiency distribution. Thus, some of Greene's observations on a gamma distribution apply to a heteroscedastic model as well as to a homoscedastic model in our case.

[6] See Caudill et al. (1995, p.107) for a discussion of the advantages of this functional form.

References

Alvarez, A. et al., 2006. "Interpreting and Testing the Scaling Property in Models where Inefficiency Depends on Firm Characteristics." *Journal of Productivity Analysis*, 25(3), 201-212.

Berg, Sanford V. 2010. *Water Utility Benchmarking: Measurement, Methodologies, and Performance Incentives*, International Water Association, xii-170.

Caudill, S.B., Jon M. Ford and Gropper, D.M., 1995. "Frontier Estimation and Firm-Specific Inefficiency Measures in the Presence of Heteroscedasticity." *Journal of Business & Economic Statistics*, 13(1), 105-111.

Coelli, T., Estache, A. and Perelman, S., 2003. *A Primer for Efficiency Measurement for Utilities and Transport Regulators*, World Bank Publications.

Coelli, T and S., Walding, 2007. "Performance Measurement in the Australian Water Supply Industry: A Preliminary Analysis." pp. 29-62 in E.Coelli, T. & Lawrence, D., 2007. *Performance Measurement and Regulation of Network Utilities*, Edward Elgar Publishing.

Davis, P. and Garces, E., 2009. *Quantitative Techniques for Competition and Antitrust Analysis*, Princeton University Press.

De Witte, K. and Marques, R., 2009a. Designing performance incentives, an international benchmark study in the water sector. *Central European Journal of Operations Research*.

De Witte, K. and Marques, R., 2009b. Capturing the environment, a metafrontier approach to the drinking water sector. *International Transactions in Operational Research*, 16(2), pp.257-71.

Haney, A. B. and M. Pollitt, 2009. Efficiency analysis of energy networks: An international survey of regulators. Energy Policy, In Press.

Greene, W.H., 1990. A Gamma-distributed stochastic frontier model. *Journal of Econometrics*, 46(1-2), 141-163.

Greene, W., 2004. Distinguishing between heterogeneity and inefficiency: stochastic frontier analysis of the World Health Organization's panel data on national health care systems. *Health Economics*, 13(10), 959-980.

Greene, W., 2005a. Reconsidering Heterogeneity in Panel Data Estimators of the Stochastic Frontier Model. *Journal of Econometrics*, 126, 269-303.

Greene, W., 2005b. Fixed and Random Effects in Stochastic Frontier Models. *Journal of Productivity Analysis*, 23(1), 7-32.

Greene, W.H., 2008. The Economic Approach of Efficiency Analysis, pp. 92-250 in Fried, H.O., Lovell, C.A.K. & Schmidt, S.S.. *The Measurement of Productive Efficiency and Productivity Growth*, illustrated edition., Oxford University Press, USA.

Hadri, K., 1999. "Estimation of a Doubly Heteroscedastic Stochastic Frontier Cost Function." *Journal of Business & Economic Statistics*, 17(3), 359-363.

Hadri, K., Guermat, C. and Whittaker, J., 2003. "Estimating Farm Efficiency in the Presence of Double Heteroscedasticity using Panel Data," *Journal of Applied Economics*, 6(2), 255-268.

Kumbhakar, S.C. & Lovell, C.A.K., 2000. *Stochastic Frontier Analysis*, Cambridge University Press.

Wang, H., and Schmidt, P., 2002. One-Step and Two-Step Estimation of the Effects of Exogenous Variables on Technical Efficiency Levels. *Journal of Productivity Analysis*, 18, 129-144.

Wang, W.S. & Schmidt, P., 2009. On the distribution of estimated technical efficiency in stochastic frontier models. *Journal of Econometrics,* 148(1), 36-45.