

Article

Comparing Groups of Decision-Making Units in Efficiency Based on Semiparametric Regression

Hohsuk Noh ¹ and Seong J. Yang ^{2,*}

¹ Department of Statistics, Sookmyung Women's University, Seoul 04310, Korea; hsnoh@sookmyung.ac.kr

² Department of Statistics (Institute of Applied Statistics), Jeonbuk National University, Jeollabuk-do 54896, Korea

* Correspondence: sjyang@jbnu.ac.kr

Received: 26 December 2019; Accepted: 7 February 2020; Published: 11 February 2020



Abstract: We consider a stochastic frontier model in which a deviation of output from the production frontier consists of two components, a one-sided technical inefficiency and a two-sided random noise. In such a situation, we develop a semiparametric regression-based test and compare the technical efficiencies of the different decision-making unit groups, assuming that the production frontier function is the same for all the groups. Our test performs better than the previously proposed ones for the same purpose in numerical studies, and also has the theoretical advantage of working under more general assumptions. To illustrate our method, we apply the proposed test to Program for International Student Assessment (PISA) 2015 data and investigate whether an efficiency difference exists between male and female student groups at a specific age in terms of learning time and achievement in mathematics.

Keywords: data envelopment analysis; stochastic frontier model; semiparametric regression; group efficiency comparison

1. Introduction

Efficiency comparison between groups is currently used in various fields such as banking, insurance, sports, and R&D investment evaluation. Numerous empirical studies frequently analyze group efficiency using so-called Data Envelopment Analysis (DEA). DEA is a body of techniques for measuring relative efficiency by comparing it with the possible frontiers of decision-making units (DMUs) with multiple inputs and outputs. Here, the term DMU is used to collectively refer to all the units in which the production activity takes place. In the DEA framework, the DMU efficiency scores of each group can be obtained after specifying some assumptions appropriate to the situation, and then the comparison of the efficiency distributions of the groups is made on the basis of their obtained scores. For example, Golany and Storberg [1] and Lee et al. [2] applied non-parametric tests, such as the Mann–Whitney (MW) and Kruskal–Wallis tests, to the efficiency scores. Cummins et al. [3] introduced a dummy variable to indicate the groups, and then regressed the efficiency scores on the dummy variable. Simar and Zelenyuk [4] adapted the test developed in Li [5] to the DEA context and applied it to the obtained scores, to test the equality of efficiency distributions. O'Donnell et al. [6] used the concept of a meta-frontier to compare the technical efficiencies of firms that may be classified into different groups.

However, this stream of research under the DEA framework has a limitation in that it does not consider the noise factor in the production process. DEA typically assumes that the inefficiency of the DMU is the only cause of its production not reaching its maximum output, but obviously there are many uncontrolled factors which need to be considered as the cause. From this recognition, Aigner and Chu [7] and Meeusen and van den Broeck [8] first proposed the stochastic frontier model (SFM),

which allows for both unobserved variation in output: the technical inefficiency of the production unit and the noise which represents the effect of innumerable uncontrollable factors. For illustrative comparison between DEA and SFM frameworks, see Figure 1.

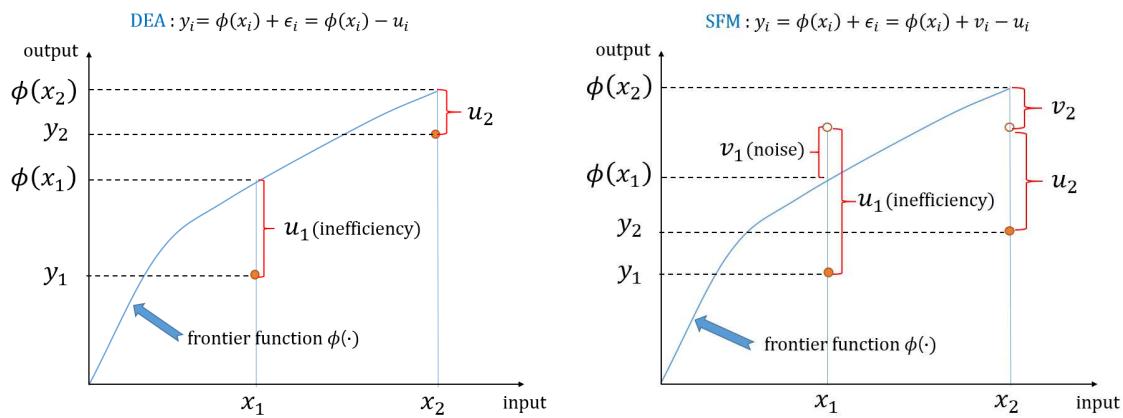


Figure 1. Comparison between Data Environment Analysis (DEA) (left panel) and SFM (right panel) frameworks (y_i : output, x_i : input, $\phi(x_i)$: the maximum output which can be obtained from the input x_i , ϵ_i : deviation from the production frontier function $\phi(x_i)$, u_i : technical inefficiency, v_i : noise). Note that the technical inefficiency u_i is a nonnegative random variable with unknown distribution.

Nowadays, the stochastic frontier model is used in a large literature of studies of production. Hence, we feel the need to develop a method and compare the efficiency difference between groups under SFM framework. One pioneering work in this direction is Banker et al. [9]. They developed five DEA-based hypothesis tests to compare the efficiency of groups under SFM. Although the paper referred above is an important development toward group efficiency comparison under SFM, their tests need to improve further.

First, their rather strong assumptions might limit the applicability of the proposed methods. For their parametric tests, they assumed the equality of both noise variance and inefficiency variance across groups. Second, their theoretical justification of the proposed methods needs to be checked. As regards their ordinary least squares (OLS) test of the mean difference in inefficiency, they provided its asymptotic normality as theoretical basis, but to our knowledge, such asymptotic normality is difficult to obtain because of the slow convergence rate of the DEA estimator when the number of input variables is greater than or equal to 2. The same comment is given in Section 3.2 of Simar and Wilson [10] on a similar type of asymptotic normality result as proof of Proposition 1 in Banker and Natarajan [11]. Finally, because they used the DEA methods for SFM, the tests they developed were based not directly on inefficiency itself, but on the inefficiency contaminated by positive measurement error due to noise. This indirect approach can lower the performance of their tests.

This observation has motivated us to develop a theoretically sound tool for comparison of group inefficiencies in the presence of noise. We develop such a methodology using a semi-parametric regression technique instead of DEA methods. The newly developed test performs better than the tests of Banker et al. [9] in numerical studies. It also has the theoretical advantage of working under more general assumptions compared to Banker et al. [9].

The rest of this paper is organized as follows. Section 2 describes our proposed test for group inefficiency comparison. We then perform some simulation studies and compare our test with the tests proposed by Banker et al. [9] in Section 3. We illustrate our method by applying the proposed test to Program for International Student Assessment (PISA) 2015 data and investigate whether an efficiency difference exists between male and female student groups at a specific age in terms of learning time and achievement in mathematics in Section 4. Section 5 provides some discussion and future research topics.

2. Group Efficiency Comparison under SFM

Assume that we have observations on n DMUs, where each observation consists of a vector of p inputs $\mathbf{X}_i = (X_{1,i}, \dots, X_{p,i})^\top$ and the corresponding output Y_i . We consider the case where n DMUs can be divided into two distinct groups with n_1 observations ($n = n_1 + n_2$). We assume the following stochastic frontier model for two groups of DMUs:

$$\begin{aligned} \text{(The first group)} \quad Y_i &= \phi(\mathbf{X}_i) + \varepsilon_{1,i}, \quad i \in \{1, \dots, n_1\}; \\ \text{(The second group)} \quad Y_i &= \phi(\mathbf{X}_i) + \varepsilon_{2,i}, \quad i \in \{n_1 + 1, \dots, n_1 + n_2\}, \end{aligned} \quad (1)$$

where $\varepsilon_{l,i} = V_{l,i} - U_{l,i}$, $V_{l,i}$ is a random noise term of the l th group with $E(V_{l,i}|\mathbf{X}_i) = 0$, and $U_{l,i}$ is an inefficiency term of the l th group with $U_{l,i} \geq 0$ for $l = 1, 2$. We assume that the same production technology is applied to both DMU groups. Hence, the production frontier function $\phi(\cdot)$ is the same throughout the groups, as in Banker et al. [9]. Under this model, we need to estimate the difference $E(U_1) - E(U_2)$ and test the hypothesis

$$H_0 : E(U_1) - E(U_2) = 0 \text{ vs. } H_1 : E(U_1) - E(U_2) > 0 (< 0) \quad (2)$$

to know which DMU group is more efficient. A novelty of our approach in developing the test is to implement the test without imposing any parametric assumption on the frontier function $\phi(\cdot)$, and with minimal assumptions on inefficiency and random noise. Banker et al. [9] also implemented the test without any parametric assumption on $\phi(\cdot)$, but with additional restrictive parametric assumptions on noise and inefficiency. In the following sections, we first review the work of Banker et al. [9] and then explain the development of our semiparametric regression-based test.

2.1. The Previous Work

To apply the DEA methods to SFM, Banker et al. [9] assumed that the random noise variables $V_{1,i}$, $V_{2,i}$ are bounded above by V^{max} , that is, $V_{1,i}$, $V_{2,i} \leq V^{max}$. Under this assumption, they transformed model (1) as

$$Y_i = (\phi(\mathbf{X}_i) + V^{max}) - (V^{max} - V_{l,i} + U_{l,i}) \equiv \tilde{\phi}(\mathbf{X}_i) - \tilde{U}_{l,i}, \quad l = 1, 2. \quad (3)$$

Since $\tilde{U}_{l,i} = (V^{max} - V_{l,i}) + U_{l,i} \geq 0$, they considered the translated production function $\tilde{\phi}(\cdot) = \phi(\cdot) + V^{max}$ as a new production function, and $\tilde{U}_{l,i}$ as the inefficiency of the DEA framework. The new inefficiency \tilde{U}_i^l can be estimated as $\hat{\phi}(\mathbf{X}_i) - Y_i$ after $\tilde{\phi}(\cdot)$ is estimated using the conventional DEA methods. After estimating $\tilde{U}_{l,i}$ using DEA methods, they used it for group efficiency comparison. This approach is advantageous in that we use the strength of the existing well-developed DEA techniques. However, the approach has one disadvantage in that the tests developed are based not on inefficiency ($U_{l,i}$) itself, but on the inefficiency contaminated by the positive measurement error ($V^{max} - V_{l,i}$) due to random noise. Additionally, the distributional property of the inefficiency estimated using DEA methods is generally hard to derive or quite complicated, making it very difficult to develop a statistical test theory based on estimated inefficiency (estimate of $\tilde{U}_{l,i}$). Hence, we are motivated to develop a test for (2) directly based on inefficiency $U_{l,i}$. We will explain this in the following section.

2.2. The Proposed Test

This section introduces our approach to testing the hypothesis in (2). Unlike Banker et al. [9], we do not require that neither the noise variance nor inefficiency should be equal across groups. Moreover, we allow for distributional difference in the composite error ε and input vector \mathbf{X}_i from the production environmental factors of each group. Specifically, the variance of V_l and mean of inefficiency U_l can differ by the group as well as conditional distribution of \mathbf{X}_i , given group l .

First, model (1) can be written as two nonparametric mean regression models as follows:

$$\begin{aligned} Y_i &= [\phi(\mathbf{X}_i) - E(U_1)] + [V_{1,i} - (U_{1,i} - E(U_1))] \\ &\equiv \phi^*(\mathbf{X}_i) + \varepsilon_{1,i}^*, \quad i \in \{1, \dots, n_1\}; \end{aligned} \quad (4)$$

$$\begin{aligned} Y_i &= [E(U_1) - E(U_2)] + [\phi(\mathbf{X}_i) - E(U_1)] + [V_{2,i} - (U_{2,i} - E(U_2))] \\ &\equiv \beta_0 + \phi^*(\mathbf{X}_i) + \varepsilon_{2,i}^*, \quad i \in \{n_1 + 1, \dots, n_1 + n_2\}, \end{aligned} \quad (5)$$

where $E(\varepsilon_{1,i}^*) = E(\varepsilon_{2,i}^*) = 0$, and $\beta_0 = E(U_1) - E(U_2)$. If a dummy variable is defined for groups letting $T_i = 0$ for $i \in \{1, \dots, n_1\}$ and $T_i = 1$ for $i \in \{n_1 + 1, \dots, n_1 + n_2\}$, the two models (4) and (5) can be integrated into a single partial linear semiparametric regression model as follows:

$$Y_i = \beta_0 T_i + \phi^*(\mathbf{X}_i) + \varepsilon_i^*, \quad i \in \{1, \dots, n\}, \quad (6)$$

where $\varepsilon_i^* = (1 - I(T_i = 1))\varepsilon_{1,i}^* + I(T_i = 1)\varepsilon_{2,i}^*$ and $E(\varepsilon_i^*|T_i, \mathbf{X}_i) = 0$. Using this model (6), we can test hypothesis (2) by testing hypothesis

$$H_0 : \beta_0 = 0 \text{ vs. } H_1 : \beta_0 > 0 (< 0). \quad (7)$$

Note that $Var(\varepsilon_i^*|T_i, \mathbf{X}_i) = (1 - I(T_i = 1))Var(\varepsilon_{1,i}^*) + I(T_i = 1)Var(\varepsilon_{2,i}^*)$. Thus, model (6) is a heteroscedastic partial linear model. Liang [12] and Ma et al. [13] studied model (6) when \mathbf{X}_i is univariate. By extending the theory from there to the case where \mathbf{X}_i is multivariate, we can test hypothesis (7). In Appendix A, we prove the asymptotic normality of the kernel-based profile estimator of β_0 based on a local linear model smoother when \mathbf{X}_i is multivariate, and provide the necessary assumptions for it. As with the estimator in Liang [12], the kernel-based profile estimator of β_0 when \mathbf{X}_i is multivariate is given as

$$\hat{\beta}_0 = (\mathbf{T}^\top(\mathbf{I} - \mathbf{S})^\top(\mathbf{I} - \mathbf{S})\mathbf{T})^{-1}\mathbf{T}^\top(\mathbf{I} - \mathbf{S})^\top(\mathbf{I} - \mathbf{S})\mathbf{Y} \equiv \mathbf{H}\mathbf{Y}, \quad (8)$$

where $\mathbf{T} = (T_1, \dots, T_n)^\top$, $\mathbf{Y} = (Y_1, \dots, Y_n)^\top$, and \mathbf{S} the smoother matrix for estimating the vector $(E(\cdot|\mathbf{X}_1), \dots, E(\cdot|\mathbf{X}_n))^\top$. If we choose local linear regression as the smoothing method, the smoothing matrix $\mathbf{S} = [\mathbf{s}_{\mathbf{X}_1} \cdots \mathbf{s}_{\mathbf{X}_n}]^\top$ will be a collection of row vectors, each of which is the smoother vector

$$\mathbf{s}_{\mathbf{x}}^\top = \mathbf{e}_1^\top(\mathbf{X}_x^\top \mathbf{W}_x \mathbf{X}_x)^{-1} \mathbf{X}_x^\top \mathbf{W}_x, \quad (9)$$

where $\mathbf{e}_1^\top = (1, 0, \dots, 0)$ is a $(p+1) \times 1$ vector; $\mathbf{W}_x = \text{diag}\{K_h(\mathbf{X}_1 - \mathbf{x}), \dots, K_h(\mathbf{X}_n - \mathbf{x})\}$ for some kernel function K and bandwidth vector $\mathbf{h} = (h_1, \dots, h_p)^\top$; and

$$\mathbf{X}_x = \begin{bmatrix} 1 & (\mathbf{X}_1 - \mathbf{x})^\top \\ \vdots & \vdots \\ 1 & (\mathbf{X}_n - \mathbf{x})^\top \end{bmatrix}. \quad (10)$$

Here, $K_h(\mathbf{X}_i - \mathbf{x}) = \prod_{j=1}^p h_j^{-1} K((X_{j,i} - x_j)/h_j)$. From the theorem in Appendix A, under some regularity conditions, $\sqrt{n}(\hat{\beta}_0 - \beta_0)$ is asymptotically normal with mean zero and variance $\sigma^2 = E(\tilde{T}^2)^{-2}E(\varepsilon^*\tilde{T})^2$, where $\tilde{T} = T - E(T|\mathbf{X})$. Using a consistent estimator of σ^2 , we can test (7) with significance level α by rejecting H_0 if $Z = \hat{\beta}_0 / (\hat{\sigma}/\sqrt{n}) \geq z_\alpha$ (or $\leq -z_\alpha$), where z_α is the $(1 - \alpha)$ -quantile of the standard normal distribution.

As regards the estimation of σ^2 , we can first directly estimate variance σ^2 using the estimates $\hat{\varepsilon}_i^*$ and $\hat{E}(T|\mathbf{X}_i)$, where $\hat{\varepsilon}_i^* = Y_i - \hat{\beta}_0 T_i - \hat{\phi}^*(\mathbf{X}_i)$ and $\hat{\phi}^*(\cdot)$ is the local linear estimator of $\phi^*(\cdot)$ based on $Y_i - \hat{\beta}_0 T_i$, $i = 1, \dots, n$. We can also estimate it using the sandwich covariance estimate based on (8),

$$\widehat{Var}(\hat{\beta}_0|\mathbf{T}, \mathbf{X}_1, \dots, \mathbf{X}_n) = \mathbf{H} \widehat{Var}(\mathbf{Y}|\mathbf{T}, \mathbf{X}_1, \dots, \mathbf{X}_n) \mathbf{H}^\top. \quad (11)$$

Matrix $\widehat{Var}(\mathbf{Y}|\mathbf{T}, \mathbf{X}_1, \dots, \mathbf{X}_n)$ is diagonal, with the i th diagonal element equal to

$$\widehat{E}(\varepsilon^{*2}|T_i, \mathbf{X}_i) = (1 - I(T_i = 1))\widehat{Var}(\varepsilon_1^*) + I(T_i = 1)\widehat{Var}(\varepsilon_2^*), \quad (12)$$

where

$$\widehat{Var}(\varepsilon_1^*) = n_1^{-1} \sum_{i: T_i=0} (\varepsilon_i^*)^2 - \left(n_1^{-1} \sum_{i: T_i=0} \hat{\varepsilon}_i^* \right)^2, \quad (13)$$

$$\widehat{Var}(\varepsilon_2^*) = n_2^{-1} \sum_{i: T_i=1} (\varepsilon_i^*)^2 - \left(n_2^{-1} \sum_{i: T_i=1} \hat{\varepsilon}_i^* \right)^2. \quad (14)$$

Since the frontier function is generally (coordinatewise) non-decreasing with respect to the input variables, one might consider it necessary to impose such a monotonicity on $\phi^*(\cdot)$. However, from Theorem 2.1 in Huang [14], such imposition will not decrease the asymptotic variance of $\hat{\beta}_0$; that is, it shows no theoretical improvement in performance. We therefore choose to develop the test without the monotonicity assumption for simplicity.

Note that our test directly estimates the mean difference in inefficiency $E(U_1) - E(U_2)$ using the semiparametric regression technique. Thus, the proposed test can work under assumptions that are more general than those in Banker et al. [9]. Additionally, we do not have to assume that noise has a finite upper support bound (V_{max}). However, the tests in Banker et al. [9] need such assumptions because they estimate $\tilde{U}_{l,i} = (V_{max} - V_{l,i}) + U_{l,i}$ and use it as a surrogate estimate of $U_{l,i}$. However, $V_{max} - V_{l,i}$ may hamper the tests and degrade their performance.

3. Numerical Studies

In this section, we compare the performance of our test with those of Banker et al. [9]. We consider single and multiple input cases and use sandwich formulas to estimate the variance in estimators.

3.1. Single Input Case

We first consider a single input case using the following model:

$$\begin{aligned} Y_i &= \phi(X_i) + V_{1,i} - U_{1,i}, \quad i \in \{1, \dots, n_1\} \\ Y_i &= \phi(X_i) + V_{2,i} - U_{2,i}, \quad i \in \{n_1 + 1, \dots, n_1 + n_2\}, \end{aligned}$$

where $\phi(x) = 30x - 9x^2$, $X \sim U(0, 1)$, $U_{l,i} \sim N_+(0, \sigma_{l,u}^2)$, and $V_{l,i}$ follow the truncated normal distribution with mean 0 and variance $\sigma_{l,v}^2$, which lies within $(-6\sigma_{l,v}, 6\sigma_{l,v})$, $l = 1, 2$. Here, N_+ stands for a normal distribution limited to the domain $[0, \infty)$. As for $V_{l,i}$, we try two cases to reflect both the equal and unequal error variances between groups. We set $\sigma_{1,v} = \sigma_{2,v} = 1$ for the equal error variance case and $\sigma_{1,v} = \sqrt{2}$, $\sigma_{2,v} = 1$ for the unequal variance case. To evaluate the type I error rate and power, we again consider two cases based on whether a mean difference (β_0) exists or does not exist between group inefficiencies: $\sigma_{1,u} = \sigma_{2,u} = 1$ ($\beta_0 = 0$) and $\sigma_{1,u} = \sqrt{2}$, $\sigma_{2,u} = 1$ ($\beta_0 = 0.3305$). Here, the type I error rate implies the rate of supporting group difference in mean inefficiency when there is no difference and the power means the rate of supporting group difference in mean inefficiency when there are really inefficiency differences between groups.

We consider three sample sizes, $n = 100, 200$, and 400 ; the proportion of each group is approximately 50% and number of replications is 1000. For a comparison, we report the type I error and power of the following five tests with significance level $\alpha = 0.05$: our proposed test (PT), the OLS test, the T-test, the Mann-Whitney (MW) test, the Kolmogorov-Smirnov (KS) test, and the F-test. The last five tests are from Banker et al. [9]. We used a plug-in principle (see Ruppert et al. [15]) to find the bandwidth for our PT.

The test results are depicted in Table 1. Four of these tests, that is, except the KS test and the F-test, seem to respect the significance level in both the equal and unequal variance cases. However, the KS test obviously shows a larger type I error rate than expected for unequal error variances and the F-test seems to be a conservative test, which gives much smaller type I error probabilities than expected. As regards the power, our PT performs best, with the largest power among all the tests. In unequal variance cases where $n = 200, 400$, the KS test has larger power than our PT. However, the KS test is not reliable since it tends to reject the null hypothesis too easily in those cases. Finally, all tests tend to show higher power with larger sample sizes.

Table 1. Type I error and power of the single input case with equal and unequal error variances.

Variances	<i>n</i>	Type I Error (Rejection Rate When $\beta_0 = 0$)						Power (Rejection Rate When $\beta_0 = 0.3305$)					
		PT	OLS	T	MW	KS	F	PT	OLS	T	MW	KS	F
Equal $(\sigma_{1,v} = \sigma_{2,v})$	100	0.052	0.050	0.050	0.050	0.037	0.013	0.377	0.336	0.320	0.283	0.200	0.152
	200	0.062	0.058	0.056	0.062	0.047	0.006	0.595	0.555	0.547	0.483	0.397	0.300
	400	0.063	0.064	0.062	0.061	0.052	0.003	0.848	0.822	0.818	0.761	0.665	0.533
Unequal $(\sigma_{1,v} \neq \sigma_{2,v})$	100	0.047	0.065	0.060	0.050	0.062	0.037	0.337	0.337	0.328	0.276	0.277	0.279
	200	0.053	0.064	0.063	0.052	0.099	0.029	0.505	0.468	0.463	0.388	0.521	0.401
	400	0.044	0.051	0.050	0.048	0.179	0.026	0.738	0.722	0.712	0.645	0.836	0.650

3.2. Multiple Input Case

We next consider a multiple input case with $p = 3$. All the components in the simulation, except for the frontier function ϕ , are the same as in the single input case. As regards the frontier function, we consider two scenarios: the production function has an additive form, and the production function does not have an additive form. The additive assumption on the production function is used in Ferrara and Vidoli [16], but it may not be satisfied in some cases. However, in case of multiple covariates, the practical applicability of our proposed method may become worse since it requires multivariate smoothing and therefore suffers from the well-known “curse of dimensionality” problem as the dimension of the covariates becomes higher. In this case, additive modeling can be a meaningful alternative. For this, we try to estimate the difference in group efficiencies and test whether it is zero with an alternative estimating strategy, where we employ a backfitting procedure, which is a well-known estimating approach under the additive assumption. See Appendix B for details of the alternative method. Considering these two scenarios (additive and non-additive production functions), we examine how our PT(n) and its alternative based on the additive assumption, PT(a), behave depending on the validity of the assumption. The model considered here is

$$\begin{aligned} Y_i &= \phi(\mathbf{X}_i) + V_{1,i} - U_{1,i}, \quad i \in \{1, \dots, n_1\} \\ Y_i &= \phi(\mathbf{X}_i) + V_{2,i} - U_{2,i}, \quad i \in \{n_1 + 1, \dots, n_1 + n_2\}, \end{aligned}$$

where $X_{j,i}$, $j = 1, 2, 3$, are generated from $U(0, 1)$ independently. For the first scenario, we set $\phi(\mathbf{x}) = (30x_1 - 9x_1^2) + (5 + 2 \arctan(10(x_2 - 0.5))) + (4\sqrt{x_3})$; this has an additive structure. For the second scenario, we consider $\phi(\mathbf{x}) = (4\sqrt{x_1} + 7\sqrt{x_2} + 5\sqrt{x_3} + 8\sqrt{x_1x_2} + 10\sqrt{x_2x_3} + 9\sqrt{x_1x_3})^{1.1}$. Note that both these production functions are concave. For PT(n), we select bandwidths by a generalized cross validation method (see Hastie and Tibshirani [17]), and for PT(a), we adopt a plug-in principle, as in the single input case.

The performances of PT(n) and PT(a) as well as the other five tests are reported in Tables 2 and 3. From Table 2, our PT(n) outperforms the competitors overall, especially in the unequal variances case. Note that its type I error rates do not deviate much from 0.05, which means that the type I error rate is under control as desired; those of other competitors such as the OLS, T and KS tests tend to be a bit smaller than this level in case of equal variances, and considerably larger in case of unequal variances. The MW test seems to respect the level like PT(n) but PT(n) turns to be more powerful

than MW. Note that PT(a) shows good results in terms of type I error rate, with comparable power to the MW test. Its power is lower than that of PT(n), but this is natural since the true production function is not additive. The F-test seems to be anticonservative leading to too high of a type I error probability, especially in the unequal variances case. However, in the equal variance case, the F-test shows very good performance in terms of type I error rate and power in large samples, as reported in Banker et al. [9]. From Table 3, our proposed two tests outperform their competitors when the additive assumption is true. Note that under the additive assumption, both PT(a) and PT(n) correctly specify the model. From our simulation, PT(a) slightly outperforms PT(n), since their type I error rates are close to 0.05 and the power of PT(a) is larger than that of PT(n). In case of equal variances, the type I error rates of the five competitors tend to be below 0.05, but when the variances are unequal, their power becomes much lower than our proposed tests although overall they show satisfactory type I errors. The only exception is the F-test. It shows the largest power in the unequal variances case but such merit is dimmed by considerably larger type I errors than other tests.

Table 2. Type I error and power of the multiple input case with equal and unequal error variances when the true production function is not additive.

Variances	n	Type I Error (Rejection Rate When $\beta_0 = 0$)							Power (Rejection Rate When $\beta_0 = 0.3305$)						
		PT(a)	PT(n)	OLS	T	MW	KS	F	PT(a)	PT(n)	OLS	T	MW	KS	F
Equal $(\sigma_{1,v} = \sigma_{2,v})$	100	0.066	0.053	0.044	0.043	0.054	0.040	0.141	0.220	0.294	0.234	0.235	0.170	0.133	0.492
	200	0.059	0.077	0.049	0.050	0.044	0.035	0.113	0.328	0.488	0.473	0.471	0.356	0.268	0.644
	400	0.048	0.056	0.039	0.039	0.037	0.031	0.064	0.509	0.765	0.705	0.706	0.580	0.517	0.815
Unequal $(\sigma_{1,v} \neq \sigma_{2,v})$	100	0.061	0.063	0.101	0.099	0.054	0.046	0.316	0.199	0.258	0.346	0.340	0.173	0.143	0.649
	200	0.058	0.074	0.139	0.137	0.062	0.053	0.381	0.302	0.433	0.583	0.580	0.347	0.331	0.849
	400	0.048	0.060	0.128	0.130	0.046	0.093	0.385	0.468	0.682	0.793	0.792	0.530	0.650	0.947

Table 3. Type I error and power of the multiple input case with equal and unequal error variances when the true production function is additive.

Variances	n	Type I Error (Rejection Rate When $\beta_0 = 0$)							Power (Rejection Rate When $\beta_0 = 0.3305$)						
		PT(a)	PT(n)	OLS	T	MW	KS	F	PT(a)	PT(n)	OLS	T	MW	KS	F
Equal $(\sigma_{1,v} = \sigma_{2,v})$	100	0.054	0.072	0.046	0.045	0.041	0.034	0.065	0.364	0.328	0.213	0.208	0.170	0.132	0.279
	200	0.066	0.060	0.063	0.061	0.056	0.045	0.045	0.578	0.551	0.387	0.384	0.318	0.258	0.403
	400	0.054	0.055	0.034	0.034	0.033	0.033	0.018	0.820	0.801	0.572	0.571	0.503	0.419	0.517
Unequal $(\sigma_{1,v} \neq \sigma_{2,v})$	100	0.057	0.070	0.066	0.062	0.044	0.048	0.144	0.301	0.295	0.236	0.228	0.153	0.121	0.381
	200	0.068	0.060	0.089	0.089	0.054	0.060	0.137	0.501	0.481	0.415	0.414	0.281	0.290	0.566
	400	0.057	0.054	0.058	0.058	0.032	0.056	0.096	0.719	0.695	0.588	0.586	0.446	0.535	0.720

4. Application to PISA 2015 Data

In this section, we applied our PT to PISA 2015 data and test the efficiency difference between male and female student groups at a specific age in terms of learning time (X_i) and achievement (Y_i) in mathematics. The data can be downloaded from <http://www.oecd.org/pisa/data/>. PISA is a worldwide study to evaluate educational systems by measuring the scholastic performance of 15-year-old school students in mathematics, science, and reading. We considered the regional averages of the students' learning time and achievement in mathematics based on test results of the 2015 version as production data. Out of the 103 regions in the data, two regions, Nova Scotia in Canada and Chile, were excluded from our analysis in view of their outlier characteristics in efficiency analysis.

Usually, international large-scale assessments data include measurement errors at the individual as well as group level. Therefore, we considered the following stochastic frontier model for such data:

$$\begin{aligned} \text{(Male Students)} Y_i &= \phi(X_i) + V_i^{male} - U_i^{male}, i \in \{1, \dots, 101\} \\ \text{(Female Students)} Y_i &= \phi(X_i) + V_i^{female} - U_i^{female}, i \in \{102, \dots, 202\}. \end{aligned}$$

In this model, we assumed that there would be no gender difference in learning ability from a biological point of view and use the same production frontier for both gender group. It means that all the socio-economic characteristics of differentiation between the gender groups were in the random error terms and not introduced in the frontier function.

Table 4 shows summary statistics of each student group data. We applied the six tests in Section 3 to the data and calculated the p-values for the following hypothesis testing:

$$H_0 : E(U^{male}) = E(U^{female}) \text{ vs. } H_1 : E(U^{male}) < E(U^{female})$$

Table 4. Summary statistics of our PISA 2015 data.

		min	Q ₁	median	mean	Q ₃	max
male	X_i	27.89	39.52	41.72	43.10	47.81	56.70
	Y_i	338.5	470.6	499.7	483.9	513.8	565.6
female	X_i	25.23	38.99	41.49	41.98	45.26	56.67
	Y_i	339.0	456.9	487.9	474.8	501.9	565.0

From Samuelsson and Samuelsson [18], it is known that male students are often more involved in mathematics classes than female students. Additionally, since women are more involved in domestic chores than men and for men time is often made free by their families and relatives for the learning activity, male students are likely to be in an environment where they can focus more on studying than female students. Hence, we expected the results of the test to indicate that the effectiveness of male students was greater than that of female students in average.

Table 5 gives the test results. At a significance level of around $\alpha = 0.05$, our PT, the MW test, and the KS test (with p-value slightly higher than $\alpha = 0.05$) supported the hypothesis that on average male students are more efficient in mathematics than female students. However, the OLS test, the T-test and the F-test reported no significant difference in learning efficiency between the two groups. The reason for this could be the somewhat restrictive assumptions for test validity. Thus, the three tests seem to face the risk of unreasonable results if the assumptions are not satisfied in practice, but our PT does not seem to suffer from this problem.

Table 5. P-values of the six tests to detect efficiency difference in groups of male and female students in terms of learning time and achievement in mathematics.

test	PT	OLS	T	MW	KS	F
p-value	0.049	0.150	0.150	0.044	0.057	0.322

5. Discussion and Conclusions

In this study, we developed tests with sound statistical theory for group efficiency comparison under SFM with considerably better performance than the previous tests proposed in numerical simulations. However, there is still room for improvement in our methods.

First, since we perform full nonparametric modeling for the frontier function $\phi(\cdot)$, which can be multivariate, our test might suffer from the “curse of dimensionality” and require high-order kernels for implementation with four or more input variables. In such a situation, we can consider an alternative test with spirit as in our test when the frontier function $\phi(\mathbf{X})$ has an additive structure, that is, $\phi(\mathbf{X}) = \sum_{j=1}^p \phi_j(X_j)$, or could be well-approximated by it.

Second, we only deal with one output case, which limits practical applications. Our methods should be extended to cover the case of multi-output production frontiers, which DEA methods cover.

Third, we assume the same production frontier for both group, which is a clear limitation in practice since such situation is not frequently observed. If it is important to assume separate production frontier functions for different groups, one can use the meta-frontier approach. O’Donnell et al. [6] proposed a meta-frontier approach to compare the group technical efficiencies under stochastic frontier framework. The proposed method has the advantage that it can be used without assuming a common frontier function. However, the use of their method sometimes can be restricted by their assumption that the frontier production function is log-linear.

Finally, if one is interested in estimating the mean inefficiency of each group, we refer to Noh and Van Keilgom [19], which is a recent work along that direction.

Author Contributions: Conceptualization, H. Noh; methodology, H. Noh and S. J. Yang; software, H. Noh and S. J. Yang; formal analysis, H. Noh and S. J. Yang; investigation, H. Noh and S. J. Yang; writing—original draft preparation, H. Noh and S. J. Yang; writing—review and editing, H. Noh and S. J. Yang. All authors have read and agree to the published version of the manuscript.

Funding: H. Noh was supported by the Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Education (NRF-2017R1D1A1A09000804), and S. J. Yang was supported by research funds for newly appointed professors of Jeonbuk National University in 2018.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

In this appendix, we provide details of the asymptotic normality of the proposed estimator $\hat{\beta}_0$ in Section 2.2. For this, we first list the relevant assumptions.

Assumption

1. The kernel function K is symmetric, and Lipschitz continuous in $[-1, 1]$.
2. ϕ is twice partially continuously differentiable.
3. The density functions of X_j ($j \in \{1, \dots, p\}$) are continuous, and bounded away from zero and infinity on their supports \mathcal{C}_j , which are bounded.
4. V_1, V_2, U_1 and U_2 have finite second moments.
5. For $j \in \{1, \dots, p\}$, h_j are asymptotic to n^{-a} for $a > 0$ such that $n(\prod_{j=1}^p h_j)^2 / \log n \rightarrow \infty$ and $nh_j^8 \rightarrow 0$ as n goes to infinity.

Theorem A1. Under the above assumptions,

$$\sqrt{n}(\hat{\beta}_0 - \beta_0) \longrightarrow N(0, \sigma^2)$$

where

$$\begin{aligned}\sigma^2 &= E(\tilde{T}^2)^{-2}E(\varepsilon^{*2}\tilde{T}^2) \\ \tilde{T} &= T - E(T|\mathbf{X})\end{aligned}$$

Proof. We write

$$\begin{aligned}\sqrt{n}(\hat{\beta}_0 - \beta_0) &= \left[n^{-1}\mathbf{T}^\top(\mathbf{I} - \mathbf{S})^\top(\mathbf{I} - \mathbf{S})\mathbf{T} \right]^{-1} \\ &\quad \times n^{-1/2} \left[\mathbf{T}^\top(\mathbf{I} - \mathbf{S})^\top(\varepsilon^* + \boldsymbol{\phi}^* - \mathbf{S}(\mathbf{Y} - \beta_0\mathbf{T})) \right]\end{aligned}$$

where $\boldsymbol{\phi}^* = (\phi^*(X_1), \dots, \phi^*(X_n))^\top$ and $\varepsilon^* = (\varepsilon_1^*, \dots, \varepsilon_n^*)^\top$. It suffices to show that

$$n^{-1}\mathbf{T}^\top(\mathbf{I} - \mathbf{S})^\top(\mathbf{I} - \mathbf{S})\mathbf{T} \xrightarrow{p} E(\tilde{T}^2), \text{ and} \quad (\text{A1})$$

$$n^{-1/2} \left[\mathbf{T}^\top(\mathbf{I} - \mathbf{S})^\top(\varepsilon^* + \boldsymbol{\phi}^* - \mathbf{S}(\mathbf{Y} - \beta_0\mathbf{T})) \right] \xrightarrow{d} N(0, E(\varepsilon^*)\tilde{T}^2). \quad (\text{A2})$$

To prove these, we first give the following fact.

$$\sup_{\mathbf{x} \in \mathcal{C}_1 \times \dots \times \mathcal{C}_p} |\hat{\xi}(\mathbf{x}) - \xi(\mathbf{x})| = O_p(n^{-2a} + n^{-(1-ap)/2} \log n), \quad (\text{A3})$$

where $\xi(\mathbf{x}) = E(R|\mathbf{X} = \mathbf{x})$ and $\hat{\xi}(\mathbf{x})$ is its local linear estimator. That is, $\hat{\xi}(\mathbf{x}) = \mathbf{s}_x^\top \mathbf{R}$ with $\mathbf{R} = (R_1, \dots, R_n)^\top$ when R_j is a response variable. (A3) can be shown from the standard theory of kernel smoothing. Note that

$$(\mathbf{I} - \mathbf{S})\mathbf{T} = (\mathbf{T} - E(\mathbf{T}|\mathbf{X})) + (E(\mathbf{T}|\mathbf{X}) - \mathbf{S}\mathbf{T}).$$

The second term of the right-hand side of the above equation is $o_p(1)$ from (A3). This proves (A1). Next, we write

$$\begin{aligned}&n^{-1/2} \left[\mathbf{T}^\top(\mathbf{I} - \mathbf{S})^\top(\varepsilon^* + \boldsymbol{\phi}^* - \mathbf{S}(\mathbf{Y} - \beta_0\mathbf{T})) \right] \\ &= n^{-1/2}(\mathbf{T} - E(\mathbf{T}|\mathbf{X}))^\top \varepsilon^* + n^{-1/2}(\mathbf{T} - E(\mathbf{T}|\mathbf{X}))^\top(\boldsymbol{\phi}^* - \mathbf{S}(\mathbf{Y} - \beta_0\mathbf{T})) \\ &\quad + n^{-1/2}(E(\mathbf{T}|\mathbf{X}) - \mathbf{S}\mathbf{T})^\top \varepsilon^* + n^{-1/2}(E(\mathbf{T}|\mathbf{X}) - \mathbf{S}\mathbf{T})^\top(\boldsymbol{\phi}^* - \mathbf{S}(\mathbf{Y} - \beta_0\mathbf{T})) \\ &\equiv n^{-1/2}(\mathbf{T} - E(\mathbf{T}|\mathbf{X}))^\top \varepsilon^* + \mathbf{A}_{1,n} + \mathbf{A}_{2,n} + \mathbf{A}_{3,n}.\end{aligned}$$

Since $n^{-1/2}(\mathbf{T} - E(\mathbf{T}|\mathbf{X}))^\top \varepsilon^* = N(0, E(\varepsilon^*)\tilde{T}^2) + o_p(1)$, it is enough to show that $\mathbf{A}_{j,n} = o_p(1)$, $j = 1, 2, 3$, to claim (A2).

To treat $A_{1,n}$, we note that

$$\sup_{x_j \in \mathcal{C}_j} \left| \frac{\partial}{\partial x_j} \hat{\xi}(\mathbf{x}) - \frac{\partial}{\partial x_j} \xi(\mathbf{x}) \right| = O_p(n^{-2a} + n^{-(1-ap-2a)/2} \log n), \quad j \in \{1, \dots, p\} \quad (\text{A4})$$

and denote $\hat{\boldsymbol{\phi}}^* = \mathbf{S}(\mathbf{Y} - \beta_0\mathbf{T})$. Then,

$$\mathbf{A}_{1,n} = n^{-1/2} \sum_{i=1}^n (T_i - E(T_i|\mathbf{X}_i))(\phi^*(\mathbf{X}_i) - \hat{\phi}^*(\mathbf{X}_i)).$$

Let \mathcal{G} denote a class of functions satisfying $|g(\mathbf{x}) - g(\mathbf{y})| \leq \|\mathbf{x} - \mathbf{y}\|$ for $\mathbf{x}, \mathbf{y} \in [0, 1]^p$. Then, $n^{a_0}(\hat{\phi}^*(\cdot) - \phi^*(\cdot))$ belongs to \mathbf{G} with probability tending to 1 from (A3) and (A4), where $a_0 < \max\{2a, (1 - ap - 2a)/2\}$. We can show that the δ -entropy of \mathbf{G} for the supremum norm satisfies

$$H_\infty(\delta, \mathcal{G}) \leq K \left(\log \frac{1}{\delta} + \frac{1}{\delta^p} \right)$$

for some constant K . Here, we consider an empirical process

$$n^{-1/2} \sum_{i=1}^n (T_i - E(T_i | \mathbf{X}_i))g(\mathbf{X}_i), \quad g \in \mathcal{G}$$

indexed by \mathcal{G} . Then, $E[(T_i - E(T_i | \mathbf{X}_i))g(\mathbf{X}_i)] = 0$, and by the Corollary 8.8 of van de Geer [20], we conclude that $\sup_{g \in \mathcal{G}} \left| n^{-1/2} \sum_{i=1}^n (T_i - E(T_i | \mathbf{X}_i))g(\mathbf{X}_i) \right| = O_p(1)$, to result in $\mathbf{A}_{1,n} = o_p(1)$. Note that the exponential tail condition, required to apply the empirical process technique, is automatically satisfied in our case since T is a binary variable.

As for $\mathbf{A}_{2,n}$, we first note that $E(\mathbf{A}_{2,n} | (T_1, \mathbf{X}_1), \dots, (T_n, \mathbf{X}_n)) = 0$. Moreover,

$$\begin{aligned} & E(\mathbf{A}_{2,n}^2 | (T_1, \mathbf{X}_1), \dots, (T_n, \mathbf{X}_n)) \\ & \leq n^{-1} [var(\varepsilon_1^{1,*}) + var(\varepsilon_1^{2,*})] (E(\mathbf{T} | \mathbf{X}) - \mathbf{ST})^\top (E(\mathbf{T} | \mathbf{X}) - \mathbf{ST}) \\ & = O_p(n^{-4a} + n^{-(1-ap)} \log n) \end{aligned}$$

from (A3). This establishes $\mathbf{A}_{2,n} = o_p(1)$. Finally, $\mathbf{A}_{3,n} = O_p(n^{1/2-4a} + n^{-1/2+ap/2} \log n) = o_p(1)$ from (A3), to complete the proof. \square

Appendix B

In this appendix, we describe an alternative test for (2) when the frontier function $\phi(\mathbf{X})$ has an additive structure, that is, $\phi(\mathbf{X}) = \sum_{j=1}^p \phi_j(X_j)$. Under the additive structure assumption, model (1) can be written as two nonparametric mean regression models:

$$\begin{aligned} Y_i &= \left[\sum_{j=1}^p E\phi_j(X_{j,i}) - E(U_1) \right] + \left[\sum_{j=1}^p (\phi_j(X_{j,i}) - E\phi_j(X_j)) \right] + [V_{1,i} - (U_{1,i} - E(U_1))] \\ &\equiv \mu + \sum_{j=1}^p \phi_j^*(X_{j,i}) + \varepsilon_{1,i}^*, \quad i \in \{1, \dots, n_1\}; \end{aligned} \tag{A5}$$

$$\begin{aligned} Y_i &= \left[\sum_{j=1}^p E\phi_j(X_{j,i}) - E(U_1) \right] + [E(U_1) - E(U_2)] + \left[\sum_{j=1}^p (\phi_j(X_{j,i}) - E\phi_j(X_j)) \right] \\ &\quad + [V_{2,i} - (U_{2,i} - E(U_2))] \\ &\equiv \mu + \beta_0 + \sum_{j=1}^p \phi_j^*(X_{j,i}) + \varepsilon_{2,i}^*, \quad i \in \{n_1 + 1, \dots, n_1 + n_2\}, \end{aligned} \tag{A6}$$

where $\phi_j^*(X_j) = \phi_j^*(X_j) - E\phi_j(X_j)$ and $E(\phi_j^*(X_j)) = 0$ for $j \in \{1, \dots, p\}$. If we introduce the same dummy variable T_i as in the single input case, models (A5) and (A6) can be integrated into one single semiparametric regression model, which would be a (heteroscedastic) partial linear additive model:

$$Y_i = \mu + \beta_0 T_i + \sum_{j=1}^p \phi_j^*(X_{j,i}) + \varepsilon_i^*, \quad i \in \{1, \dots, n\}, \tag{A7}$$

where $\varepsilon_i^* = (1 - I(T_i = 1))\varepsilon_{1,i}^* + I(T_i = 1)\varepsilon_{2,i}^*$ and $E(\varepsilon_i^* | T_i, \mathbf{X}_i) = 0$. Partial linear additive models have been studied by several authors; for example, Fan et al. [21], Fan and Li [22], Li [23], and Wei and

Liu [24]. For the test, we use the profile least square estimator of β_0 in Wei and Liu [24]. However, Wei and Liu [24] only showed the asymptotic distribution of the estimator of the parametric component vector (in our case, the estimator of $\beta = (\mu, \beta_0)^\top$) under the homoscedasticity assumption of the error (Theorem 2.1 of their paper), and so we extended their result to the heteroscedasticity case.

To introduce the profile least square estimator of β_0 using the method of Wei and Liu [24], we define some notations. Let

$$\mathbf{X}_{des} = \begin{bmatrix} 1 & X_1 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix} = [\mathbf{1}_n \ \mathbf{X}], \mathbf{S}_T = \begin{bmatrix} \mathbf{I}_n & \mathbf{S}_1^* & \cdots & \mathbf{S}_1^* \\ \mathbf{S}_2^* & \mathbf{I}_n & \cdots & \mathbf{S}_2^* \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{S}_p^* & \mathbf{S}_1^* & \cdots & \mathbf{I}_n \end{bmatrix}, \mathbf{C} = \begin{bmatrix} \mathbf{S}_1^* \\ \mathbf{S}_2^* \\ \vdots \\ \mathbf{S}_p^* \end{bmatrix}, \quad (A8)$$

where \mathbf{S}_k is the smoothing matrix for local linear regression with respect to the j th ($j \in \{1, \dots, p\}$) covariate vector $\mathbf{X}_j = (X_{j,1}, \dots, X_{j,n})^\top$ with kernel function $K(\cdot)$ and bandwidth h_j , $\mathbf{S}_j^* = (\mathbf{I}_n - \mathbf{1}_n \mathbf{1}_n^\top / n) \mathbf{S}_j$, and $\mathbf{1}_n = (1, \dots, 1)^\top$ with length n . Additionally, we define the additive smoother matrix \mathbf{W}_j as $\mathbf{W}_j = \mathbf{E}_j \mathbf{S}_T^{-1} \mathbf{C}$, where \mathbf{E}_j is a partitioned matrix of dimension $n \times np$ with $n \times n$ identity matrix as the j th “block” and zeros elsewhere. Then, the profile least squares estimator of $\beta = (\mu, \beta_0)^\top$ is obtained as the estimator of the coefficient vector β of a synthetic linear regression model

$$Y_i - \tilde{Y}_i = (\mathbf{T}_{des,i} - \tilde{\mathbf{T}}_{des,i})^\top \beta + \varepsilon_i, \quad (A9)$$

where $\mathbf{W}_M = \sum_{j=1}^p \mathbf{W}_j$, $\tilde{Y} = (\tilde{Y}_1, \dots, \tilde{Y}_n)^\top = \mathbf{W}_M \mathbf{Y}$ and $\tilde{\mathbf{T}}_{des} = (\tilde{\mathbf{T}}_{des,1}, \dots, \tilde{\mathbf{T}}_{des,n})^\top = \mathbf{W}_M \mathbf{T}_{des}$. Additionally, since $\mathbf{W}_M \mathbf{1}_n = (0, \dots, 0)^\top$, we know that the linear model (A9) becomes

$$Y_i - \tilde{Y}_i = \mu + (T_i - \tilde{T}_i) \beta_0 + \varepsilon_i^*, \quad (A10)$$

where $\tilde{\mathbf{T}} = (\tilde{T}_1, \dots, \tilde{T}_n)^\top = \mathbf{W}_M \mathbf{T}$. Hence, after a standard calculation in linear model theory, we obtain the profile least squares estimator of β_0 as

$$\hat{\beta}_0 = \left[\mathbf{T}^\top (\mathbf{I}_n - \mathbf{W}_M)^\top (\mathbf{I}_n - \mathbf{J}) (\mathbf{I}_n - \mathbf{W}_M) \mathbf{T} \right]^{-1} \mathbf{T}^\top (\mathbf{I}_n - \mathbf{W}_M)^\top (\mathbf{I}_n - \mathbf{J}) (\mathbf{I}_n - \mathbf{W}_M) \mathbf{Y}, \quad (A11)$$

where $\mathbf{J} = \mathbf{1}_n \mathbf{1}_n^\top / n$. Using the results to prove Theorem 2.1 in Wei and Liu [24], we show below that under some regularity conditions, estimator $\hat{\beta}_0$ is asymptotically normal with mean zero and variance

$$\sigma_{add}^2 = E(\tilde{T}^2)^{-2} E(\varepsilon^{*2} \tilde{T}^2), \quad (A12)$$

where $\tilde{T} = T - E(T) - \sum_{j=1}^p [E(T|X_j) - E(T)]$. Once we have a consistent estimate of σ_{add}^2 , we can test (7) for a given significance level α . As with the case of single input variable, we can directly estimate the variance $\sigma_{\beta_0}^2$ using estimates $\hat{\varepsilon}_i^*$ and $\hat{E}(T|X_{j,i})$. Here, $\hat{E}(T|X_{j,i})$ can be obtained as the i th element of $\mathbf{S}_j(T_1, \dots, T_n)^\top$. Alternatively, we can estimate the variance via the sandwich formula estimate based on (A11) following similar steps in Section 2.2.

Now, we can show the asymptotic property of the profile least square estimator of β_0 . For this, we first list the relevant assumptions.

Assumption

1. The kernel function K is symmetric, and Lipschitz continuous in $[-1, 1]$.
2. ϕ_j ($j \in \{1, \dots, p\}$) are twice continuously differentiable.
3. The density functions of X_j ($j \in \{1, \dots, p\}$) are continuous, and bounded away from zero and infinity on their supports, which are bounded.
4. V_1, V_2, U_1 and U_2 have finite second moments.
5. For $j \in \{1, \dots, p\}$, $h_j \rightarrow 0$, $nh_j / \log n \rightarrow \infty$ and $nh_j^8 \rightarrow 0$ as n goes to infinity.

Theorem A2. Under the above assumptions,

$$\sqrt{n}(\hat{\beta}_0 - \beta_0) \longrightarrow N(0, \sigma_{add}^2)$$

where

$$\begin{aligned}\sigma_{add}^2 &= E(\tilde{T}^2)^{-2}E(\varepsilon^{*2}\tilde{X}^2) \\ \tilde{T} &= T - E(T) - \sum_{j=1}^p [E(T|X_j) - E(T)]\end{aligned}$$

Proof. $\hat{\beta}_0$ can be expressed as follows:

$$\hat{\beta}_0 = \left[\mathbf{T}^\top (\mathbf{I}_n - \mathbf{W}_M)^\top (\mathbf{I}_n - \mathbf{J})(\mathbf{I}_n - \mathbf{W}_M)\mathbf{T} \right]^{-1} \mathbf{T}^\top (\mathbf{I}_n - \mathbf{W}_M)^\top (\mathbf{I}_n - \mathbf{J})(\mathbf{I}_n - \mathbf{W}_M)\mathbf{Y}$$

where $\mathbf{T} = (T_1, \dots, T_n)^\top$ and $\mathbf{J} = \mathbf{1}_n \mathbf{1}_n^\top / n$. Then,

$$\begin{aligned}&\sqrt{n}(\hat{\beta}_0 - \beta_0) \\ &= \left[n^{-1} \mathbf{T}^\top (\mathbf{I}_n - \mathbf{W}_M)^\top (\mathbf{I}_n - \mathbf{J})(\mathbf{I}_n - \mathbf{W}_M)\mathbf{T} \right]^{-1} \\ &\quad \times n^{-1/2} \mathbf{T}^\top (\mathbf{I}_n - \mathbf{W}_M)^\top (\mathbf{I}_n - \mathbf{J})(\mathbf{I}_n - \mathbf{W}_M)(\boldsymbol{\phi}^* + \boldsymbol{\varepsilon}^*),\end{aligned}$$

where $\boldsymbol{\phi}^* = (\sum_{j=1}^p \phi^*(X_{j,1}), \dots, \sum_{j=1}^p \phi^*(X_{j,n}))^\top$ and $\boldsymbol{\varepsilon}^* = (\varepsilon_1^*, \dots, \varepsilon_n^*)^\top$. Here, the term associated with the intercept μ vanishes because $\mathbf{W}_M \mathbf{1}_n = (0, \dots, 0)^\top$. To prove the theorem, it suffices to show that

$$\begin{aligned}&n^{-1} \mathbf{T}^\top (\mathbf{I}_n - \mathbf{W}_M)^\top (\mathbf{I}_n - \mathbf{J})(\mathbf{I}_n - \mathbf{W}_M)\mathbf{T} \\ &= n^{-1} \sum_{i=1}^n \left[T_i - E(T_i) - \sum_{j=1}^p [E(T_i|X_{j,i}) - E(T_i)] \right]^2 + o_p(1)\end{aligned}\tag{A13}$$

and

$$\begin{aligned}&n^{-1/2} \mathbf{T}^\top (\mathbf{I}_n - \mathbf{W}_M)^\top (\mathbf{I}_n - \mathbf{J})(\mathbf{I}_n - \mathbf{W}_M)(\boldsymbol{\phi}^* + \boldsymbol{\varepsilon}^*) \\ &= n^{-1/2} \sum_{i=1}^n \left[T_i - E(T_i) - \sum_{j=1}^p [E(T_i|X_{j,i}) - E(T_i)] \right] \varepsilon_i^* + o_p(1).\end{aligned}\tag{A14}$$

Note that $(\mathbf{I}_n - \mathbf{J})(\mathbf{I}_n - \mathbf{W}_M)\mathbf{X} = (\mathbf{I}_n - \mathbf{W}_M)\mathbf{X} - \mathbf{1}_n \bar{X}$, where $\bar{X} = n^{-1} \sum_{i=1}^n X_i$ because fact $\mathbf{1}_n^\top \mathbf{W}_M = (0, \dots, 0)$. Then, one can easily see that

$$\begin{aligned}&n^{-1} \mathbf{X}^\top (\mathbf{I}_n - \mathbf{W}_M)^\top (\mathbf{I}_n - \mathbf{J})(\mathbf{I}_n - \mathbf{W}_M)\mathbf{T} \\ &= n^{-1} (\mathbf{T} - \mathbf{1}_n \mu_T)^\top (\mathbf{I}_n - \mathbf{W}_M)^\top (\mathbf{I}_n - \mathbf{W}_M)(\mathbf{T} - \mathbf{1}_n \mu_T) + O_p(n^{-1/2})\end{aligned}$$

for $\mu_T = E(T_1)$. Therefore, Equation (A13) can be verified as in the proof of Lemma 6.2 in Wei and Liu (2012). For Equation (A14), note that

$$\begin{aligned}&n^{-1/2} \mathbf{T}^\top (\mathbf{I}_n - \mathbf{W}_M)^\top (\mathbf{I}_n - \mathbf{J})(\mathbf{I}_n - \mathbf{W}_M)\boldsymbol{\phi}^* \\ &= n^{-1/2} (\mathbf{T} - \mathbf{1}_n \mu_T)^\top (\mathbf{I}_n - \mathbf{W}_M)^\top (\mathbf{I}_n - \mathbf{W}_M)(\boldsymbol{\phi}^* - \mathbf{1}_n \mu_{\boldsymbol{\phi}^*}) + O_p(n^{-1/2})\end{aligned}\tag{A15}$$

and

$$\begin{aligned}&n^{-1/2} \mathbf{T}^\top (\mathbf{I}_n - \mathbf{W}_M)^\top (\mathbf{I}_n - \mathbf{J})(\mathbf{I}_n - \mathbf{W}_M)\boldsymbol{\varepsilon}^* \\ &= n^{-1/2} (\mathbf{T} - \mathbf{1}_n \mu_X)^\top (\mathbf{I}_n - \mathbf{W}_M)^\top (\mathbf{I}_n - \mathbf{W}_M)\boldsymbol{\varepsilon}^* + O_p(n^{-1/2}),\end{aligned}\tag{A16}$$

where $\mu_{\phi^*} = E(\sum_{j=1}^p \phi^*(X_{j,1}))$. Then, we have

$$(\mathbf{I}_n - \mathbf{W}_M)\boldsymbol{\varepsilon}^* = \boldsymbol{\varepsilon}^* - \sum_{j=1}^p \mathbf{S}_j \boldsymbol{\varepsilon}^* + O_p \left(n \sum_{j=1}^p h_j^4 \right)$$

from Lemma B.6 in [25]. Note that this is true as long as the conditional variance of the $\boldsymbol{\varepsilon}^*$ given covariates exists. This is guaranteed by assumption 4. Wei and Liu (2012) used a similar fact under the homoscedastic error assumption. Then, with a derivation similar to that in Lemma 6.3 of Wei and Liu (2012), we can show that (A15) converges to zero in probability and (A16) can be written as:

$$n^{-1/2} \sum_{i=1}^n \left[T_i - E(T_i) - \sum_{j=1}^p [E(T_i|X_{j,i}) - E(T_i)] \right] \boldsymbol{\varepsilon}_i^* + o_p(1),$$

to complete the proof. \square

References

1. Golany, B.; Storberg, J. A data envelopment analysis of the operational efficiencies of bank branches. *Interfaces* **1999**, *29*, 14–26.
2. Lee, H.; Park, Y.; Choi, H. Comparative evaluation of performance of national R&D programs with heterogeneous objectives: A DEA approach. *Eur. J. Oper. Res.* **2009**, *196*, 847–855.
3. Cummins, J.D.; Weiss, M.A.; Zi, H. Organizational form and efficiency: The coexistence of stock and mutual property-liability insurers. *Manag. Sci.* **1999**, *45*, 1254–1269.
4. Simar, L.; Zelenyuk, V. On testing equality of distributions of technical efficiency scores. *Econom. Rev.* **2006**, *25*, 497–522.
5. Li, Q. Nonparametric testing of closeness between two unknown distribution functions. *Econom. Rev.* **1996**, *15*, 261–274.
6. O'Donnell, C.J.; Rao, D.S.P.; Battese, G.E. Metafrontier frameworks for the study of firm-level efficiencies and technology ratios. *Empir. Econ.* **2008**, *34*, 231–255.
7. Aigner, D.; Chu, S. On estimating the industry production function. *Am. Econ. Rev.* **1968**, *58*, 826–839.
8. Meeusen, W.; van den Broeck, J. Efficiency estimation from Cobb-Douglas production functions with composed error. *Int. Econ. Rev.* **1977**, *18*, 435–444.
9. Banker, R.D.; Zheng, Z.; Natarajan, R. DEA-based hypothesis tests for comparing two groups of decision making units. *Eur. J. Oper. Res.* **2010**, *206*, 231–238.
10. Simar, L.; Wilson, P.W. Two-stage DEA: Caveat emptor. *J. Product. Anal.* **2011**, *36*, 205–218.
11. Banker, R.D.; Natarajan, R. Evaluating contextual variables affecting productivity using data envelopment analysis. *Oper. Res.* **2008**, *56*, 48–58.
12. Liang, H. Estimation in partially linear models and numerical comparisons. *Comput. Stat. Data Anal.* **2006**, *50*, 675–687.
13. Ma, Y.; Chiou, J.M.; Wang, N. Efficient semiparametric estimator for heteroscedastic partially linear models. *Biometrika* **2006**, *93*, 75–84.
14. Huang, J. A note on estimating a partly linear model under monotonicity constraints. *J. Stat. Plan. Inference* **2002**, *107*, 343–351.
15. Ruppert, D.; Sheather, S.J.; Wand, M.P. An effective bandwidth selector for local least squares regression. *J. Am. Stat. Assoc.* **1995**, *90*, 1257–1270.
16. Ferrara, G.; Vidoli, F. Semiparametric stochastic frontier models: A generalized additive model approach. *Eur. J. Oper. Res.* **2017**, *258*, 761–777.
17. Hastie, T.J.; Tibshirani, R.J. *Generalized Additive Models*; Chapman and Hall/CRC: New York, NY, USA, 1990.
18. Samuelsson, M.; Samuelsson, J. Gender differences in boys' and girls' perception of teaching and learning mathematics. *Open Rev. Educ. Res.* **2016**, *3*, 18–34.
19. Noh, H.; Van Keilgom, I. On relaxing the distributional assumption of stochastic frontier models. *J. Korean Stat. Soc.* **2020**, in press.

20. Van de Geer, S. *Empirical Processes in M-Estimation*; Cambridge University Press, 2000.
21. Fan, J.; Hardle, W.; Mammen, E. Direct estimation of low-dimensional components in additive models. *Ann. Stat.* **1998**, *26*, 943–971. doi:10.1214/aos/1024691083.
22. Fan, Y.; Li, Q. A kernel-based method for estimating additive partially linear models. *Stat. Sin.* **2003**, *13*, 739–762.
23. Li, Q. Efficient estimation of additive partially linear models. *Int. Econ. Rev.* **2000**, *41*, 1073–1092.
24. Wei, C.H.; Liu, C. Statistical inference on semi-parametric partial linear additive models. *J. Nonparametric Stat.* **2012**, *24*, 809–823.
25. Fan, J.; Jiang, J. Nonparametric inferences for additive models. *J. Am. Stat. Assoc.* **2005**, *100*, 890–907.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).