Semiparametric Smooth-coefficient Stochastic Frontier Model

Subal C. Kumbhakar Kai Sun

Abstract

This paper proposes a semiparametric smooth-coefficient stochastic production frontier model where all the coefficients are expressed as some unknown functions of environmental factors. The inefficiency term is multiplicatively decomposed into a scaling function of the environmental factors and a standard truncated normal random variable. A testing procedure is suggested for the relevance of the environmental factors. Monte Carlo study shows plausible finite sample behavior of our proposed estimation and inference procedure. An empirical example is given, where both the semiparametric and standard parametric models are estimated and results are compared.

Key Words: Semiparametric smooth-coefficient Model, Stochastic Frontier Model, Environmental Factor

JEL Codes: C13, C14, D24

1 Introduction

Following the seminal work of Aigner, Lovell and Schmidt (1977) and Meeusen and van den Broeck (1977), there was an abundant literature on the estimation of technical inefficiency in stochastic frontier framework (see Kumbhakar and Lovell (2000) for references). More recently, attention was paid to the modeling of environmental factors (hereafter, Z variables) affecting inefficiency. They are the exogenous factors, such as time, R&D, etc., in addition to traditional input(s) and output(s) in frontier models. For example, Kumbhakar (1990) proposed a multiplicative decomposition of technical inefficiency into a time-varying part and a noise term (see also Battese and Coelli (1992)). Alvarez, Amsler, Orea and Schmidt (2006) named this decomposition the scaling property of technical inefficiency. The idea of this property was proposed earlier by Reifschneider and Stevenson (1991), Simar, Lovell and van den Eeckaut (1994) and Caudill, Ford and Gropper (1995), and further studied by Wang and Schmidt (2002), where they assumed that the technical inefficiency is distributed as truncated normal (from the left at zero) with mean zero and variance as a function of the Z variables. This is equivalent to that the technical inefficiency is distributed as a standard truncated normal (with mean zero and unit variance) scaled by a function of Z. Alvarez et al. (2006) interpreted the standard truncated normal random variable as "the firms' base efficiency level which captures things like the manager's natural skills", but "how well these natural skills are exploited to manage the firm efficiently depends on ... measures of the environment in which the firm operates." Alternatively, Kumbhakar, Ghosh and McGuckin (1991) and Battese and Coelli (1995) proposed an additive decomposition

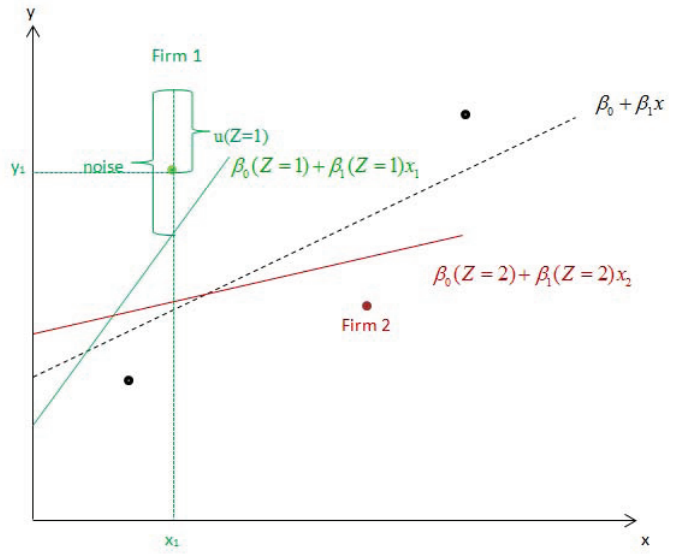
of technical inefficiency into a function of Z and a noise term (see also Huang and Liu (1994) and Simar and Wilson (2007)).

The novelty of this paper lies in the fact that we not only consider the impact of Z variables on the technical inefficiency part, but we also introduce the Z variables into the frontier part in a semiparametric fashion, following Zhang, Sun, Delgado and Kumbhakar (2012), who ignored the inefficiency part. Specifically, in a production framework, we express the intercept and slope coefficients as unknown functions of the Z variables. This allows the environmental factors to shift the frontier non-neutrally. In contrast, Saal, Parker and Weyman-Jones (2007) treated these factors as traditional inputs which can only affect the frontier neutrally. More formally, we propose the estimation of a stochastic frontier model such as

$$output = \beta_0(R\&D) + \beta_1(R\&D)labor + \beta_2(R\&D)capital + v - u(R\&D) \tag{1}$$

where β's are the coefficients, v is noise term, and u is technical inefficiency. This model works for either cross-sectional or panel data, and is able to yield individual-specific estimates. For example, different technologies can be estimated for different firms during different time periods depending on the levels of R&D. Figure 1 gives a simple graphical illustration of the proposed semiparametric stochastic frontier model with one input and one output. The black dashed line is the standard frontier that *all* firms share. It, however, does not capture the heterogeneity of firm's technology. The semiparametric model is able to estimate a particular frontier for a particular firm. For example, firm 1 (represented by the green line) can have a different intercept *and* slope from that of firm 2 (represented by the red line), depending on the level of a Z variable such as R&D. Meanwhile, Z can also affect technical inefficiency. This allows one to compare the technologies, including technical inefficiencies, for different firms which are linked by the Z variable, say, R&D. In this regard, our proposed model has several advantages over Battese, Prasada Rao and O'Donnell (2004) and O'Donnell, Prasada Rao and Battese (2008) (hereafter, B&O) who suggested a metafrontier framework for the comparison of firms under different technologies: (1) B&O's model is more liable to sample misclassification due to potentially different grouping criteria whereas grouping is not required in our model; (2) B&O's model only yields group-specific estimates while ours is individual-specific; (3) our model yields comparable estimates linked by the Z variables and there is no need to estimate a common metafrontier. To give more credibility of the inclusion of the Z variables into the model, a residual-based wild bootstrap testing procedure, borrowed from Li and Racine (2010), for the relevance of the environmental factors is proposed. We show that the model under the null of irrelevance of Z is the same as a standard parametric stochastic frontier model without environmental factors. We then apply our proposed methodology in the Norwegian forestry, with a cross-section of 3249 active forest owners. Both standard and semiparametric

Figure 1: **A Simple Illustration of the Semiparametric Stochastic Frontier Model**



frontier models are estimated and results are compared.

The rest of the paper is organized as follows. Section 2 presents the estimation procedure of a semiparametric stochastic production frontier model with environmental factors. Section 3 proposes a test for the relevance of the environmental factors. Section 4 performs a Monte Carlo study to show the finite sample behavior of our proposed estimator. Section 5 applies the method to the Norwegian forestry. Section 6 concludes.

2 Estimation of Technical Inefficiency

Consider a stochastic production frontier model with the following specification:

$$y_i = \alpha(Z_i) + X_i'\beta(Z_i) + v_i - u_i, \tag{2}$$

where y_i is the log of output, $X_i' = [x_{1i}, \ldots, x_{ki}]$ is a vector of the log of k-inputs,[1] Z_i is a p-vector of environmental factors (e.g., time, R&D, among others), $\alpha(\cdot)$ is the intercept and $\beta(\cdot)$ is a $k \times 1$ parameter vector. Both of them are expressed as unknown functions of Z_i. $v_i \sim iidN(0, \sigma_v^2)$ is the noise term, and $u_i = u(Z_i; \delta)$ is the positive technical inefficiency term as some function of the same set of environmental factors, $\delta' = [\delta_0,\ \delta_1']$ is a parameter vector where δ_0 is a scalar and δ_1 is a p-vector. Following Simar et al. (1994) and Caudill et al. (1995), this functional disturbance can be identified through heteroscedasticity by assuming $u_i = \sigma_u(Z_i)\eta_i$, where $\sigma_u(Z_i) = \exp(\delta_0 + \delta_1' Z_i)$, and $\eta_i \sim iidN^+(0,1)$; or equivalently, $u_i \sim iidN^+(0, \sigma_u^2(Z_i))$,

[1]For the translog specification (Christensen, Jorgenson and Lau 1971), X_i will include higher-order terms and interactions.

where $\sigma_u^2(Z_i) = \exp[2(\delta_0 + \delta_1' Z_i)]$. The functional form and distributional assumptions are required to guarantee the positivity of u_i. These assumptions indicate $E(u_i) = \sqrt{2/\pi}\sigma_u(Z_i) = \sqrt{2/\pi}\exp(\delta_0 + \delta_1' Z_i)$.

The production frontier in (2) is not the conditional expectation of y_i, because the composite error term $v_i - u_i$ does not have a zero mean. To solve for this problem, we can rewrite (2) as:

$$y_i = \alpha(Z_i) + X_i'\beta(Z_i) + v_i - (u_i - E(u_i)) - E(u_i), \tag{3}$$

or equivalently,

$$y_i = \theta(Z_i) + X_i'\beta(Z_i) + \varepsilon_i, \tag{4}$$

where $\theta(Z_i) = \alpha(Z_i) - E(u_i)$, and $\varepsilon_i = v_i - (u_i - E(u_i))$. (4) can be consistently estimated as a semiparametric smooth coefficient model (Li, Huang, Li and Fu 2002). Define $\hat{\rho}(Z_i) = [\hat{\theta}(Z_i),\ \hat{\beta}'(Z_i)]$, and $W_i' = [1,\ X_i']$, the smooth coefficient estimator can be written as:

$$\hat{\rho}(Z_i) = \left[\sum_{j=1}^{n} W_j W_j' K\left(\frac{Z_j - Z_i}{h}\right)\right]^{-1} \sum_{j=1}^{n} W_j y_j K\left(\frac{Z_j - Z_i}{h}\right), \tag{5}$$

where n is sample size, $K(\cdot)$ is product kernel function (Li and Racine 2007), and h is a p-vector of bandwidth, which can be selected via least-squares cross-validation method (Li and Racine 2010).

Residuals can be obtained from the estimated equation, viz., $\hat{\varepsilon}_i = y_i - \hat{\theta}(Z_i) - X_i'\hat{\beta}(Z_i)$. This is the first step of the estimation. Recall that we previously defined $\varepsilon_i = v_i - u_i + E(u_i)$, where $u_i = \sigma_u(Z_i)\eta_i$, $E(u_i) = \sqrt{2/\pi}\sigma_u(Z_i)$, and therefore, the estimating equation for the second step of the estimation is:

$$\begin{aligned} R_i &= \sqrt{2/\pi}\sigma_u(Z_i) + v_i - \sigma_u(Z_i)\eta_i \\ &= \sqrt{2/\pi}\exp(\delta_0 + \delta_1' Z_i) + v_i - \exp(\delta_0 + \delta_1' Z_i)\eta_i \end{aligned} \tag{6}$$

where $R_i = \hat{\varepsilon}_i$. A parametric stochastic frontier estimation technique can be applied in this step, using maximum likelihood estimation method. Define $\varepsilon_i^* = v_i - \exp(\delta_0 + \delta_1' Z_i)\eta_i$, the log-likelihood function can be written as:

$$\ln L = Constant - \frac{1}{2}\sum_i \ln\left[\sigma_u^2(Z_i) + \sigma_v^2\right] + \sum_i \ln\Phi\left(-\frac{\varepsilon_i^*\lambda_i}{\sigma_i}\right) - \frac{1}{2}\sum_i \frac{\varepsilon_i^{*2}}{\sigma_i^2}, \tag{7}$$

where $\sigma_u^2(Z_i) = \exp[2(\delta_0 + \delta_1' Z_i)]$, $\sigma_i^2 = \sigma_v^2 + \sigma_u^2(Z_i) = \sigma_v^2 + \exp[2(\delta_0 + \delta_1' Z_i)]$, and $\lambda_i = \sigma_u(Z_i)/\sigma_v = \exp(\delta_0 + \delta_1' Z_i)/\sigma_v$. δ and σ_v^2 can be first estimated by maximizing the log-likelihood function. λ_i and $\sigma_u^2(Z_i)$ can then be estimated. $E(u_i)$ is estimated as the scaled $\sigma_u(Z_i)$, which is used to identify the intercept in

(2) as $\alpha(Z_i) = \theta(Z_i) + E(u_i)$.

3 Testing for the Relevance of Environmental Factors

The environmental factors, Z_i, shift the production frontier (both intercept and slopes) as well as technical inefficiency. One naturally wants to test if Z_i matters, that is, to test if (2) can be estimated as a standard stochastic frontier model:

$$y_i = \alpha + X_i'\beta + \nu_i - \mu_i, \tag{8}$$

where ν_i is the normal noise term, and μ_i is the half-normal technical inefficiency term. In this model, neither the coefficients nor the technical inefficiency vary with Z_i. This is the same as testing if the parameters in (4) are constants, viz.,

$$y_i = \theta + X_i'\beta + \epsilon_i = W_i'\rho + \epsilon_i, \tag{9}$$

where $\rho' = [\theta,\ \beta']$, and $\epsilon_i = \nu_i - (\mu_i - E(\mu_i))$. The null hypothesis can be stated as $H_0 : \rho(Z_i) = \rho$.[2] Following Li and Racine (2010), the consistent model specification test statistic is constructed as:

$$\hat{I}_n = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}^{n} W_i'W_j\hat{\epsilon}_i\hat{\epsilon}_j K\left(\frac{Z_i - Z_j}{h}\right) \tag{10}$$

where $K(\cdot)$ is the product kernel function, $\hat{\epsilon}_i = y_i - \hat{\theta} - X_i'\hat{\beta}$ is obtained from the parametric model (9). We follow Li and Racine's (2010) residual-based wild bootstrap method to determine whether to reject the null hypothesis or not:

Step 1: Estimate (9), obtain $\hat{\rho}$ and $\hat{\epsilon}_i$, and generate wild bootstrap disturbance $\epsilon_i^\star$;

Step 2: From $\epsilon_i^\star$, generate $y_i^\star = W_i'\hat{\rho} + \epsilon_i^\star$;

Step 3: Use $\{y_i^\star, W_i\}_{i=1}^n$ to estimate the parametric model (9), and obtain $\hat{\rho}^\star$, and $\hat{\epsilon}_i^\star = y_i^\star - W_i'\hat{\rho}^\star$;

Step 4: The bootstrap statistic $\hat{I}_n^\star$ is obtained from (10), replacing $\hat{\epsilon}_i\hat{\epsilon}_j$ by $\hat{\epsilon}_i^\star\hat{\epsilon}_j^\star$.

Step 5: Repeat Steps 1-4 a large number of times, say $B = 399$ times, and calculate the p-value: $p = \frac{1}{B}\sum_{b=1}^{B}\mathbf{I}(I_n^* > I_n)$, where $\mathbf{I}(\cdot)$ is the indicator function with a value of 1 if the statement in the parenthesis is true.

Note that $y_i^\star$ is generated under the null hypothesis, and therefore, the p-value is the size of the test. The null hypothesis can be rejected if the p-value is less than the level of significance, say 0.05.

[2] Constant ρ implies constant θ and β, and constant θ implies constant α and $E(\mu_i)$. The log-likelihood function under this scenario is:

$$\ln L = Constant - \frac{1}{2}\sum_i \ln(\sigma_\mu^2 + \sigma_\nu^2) + \sum_i \ln\Phi\left(-\frac{\epsilon_i^*\lambda}{\sigma}\right) - \frac{1}{2}\sum_i \frac{\epsilon_i^{*2}}{\sigma^2},$$

where $\epsilon_i^* = \nu_i - \mu_i$, $\lambda = \sigma_\mu/\sigma_\nu$, and $\sigma^2 = \sigma_\mu^2 + \sigma_\nu^2$.

4 Monte Carlo Study

4.1 Estimation Procedure

To study the finite-sample behavior of the proposed semiparametric smooth-coefficient stochastic frontier estimation method, we conduct some Monte Carlo experiments using the following data generating process (DGP):

$$y_i = \alpha(Z_i) + \beta(Z_i)X_i + v_i - u_i,$$

where Z_i is generated on a equally-spaced grid between -2 and 2, $\alpha(Z_i) = \exp(Z_i)$, $\beta(Z_i) = \cos(Z_i)$, $X_i \sim N(0,1)$, $v_i \sim N(0, \sigma_v^2)$, $u_i = \exp(\delta_0 + \delta_1 Z_i)\eta_i$, where $\eta_i \sim N^+(0,1)$.

We draw $M = 1000$ Monte Carlo replications from this DGP, and consider sample size of $n = 100, 200, 400,$and 800. Table 1 reports the bias, variance, and mean squared error (MSE) for each of the parameters in (7). It can be seen that both variance and MSE decreases as the sample size increases. Since other parameters of interest are functions of Z_i - $\sigma_u^2(Z_i) = \exp[2(\delta_0 + \delta_1 Z_i)]$, $\sigma_i^2 = \sigma_v^2 + \sigma_u^2(Z_i)$, and $\lambda_i = \sigma_u(Z_i)/\sigma_v$, Figure 2-5 reports the trajectory of the true and estimated functional parameters along with their 95% confidence bands. It can be seen that the confidence bands of a particular parameter shrink as the sample size increases. Therefore, the Monte Carlo study shows evidence of the consistency of our proposed estimator.

Table 1: Performance of constant parameters

	$\delta_0 = 1$			$\delta_1 = 1$			$\sigma_v^2 = 1$		
n	Bias	Var	MSE	Bias	Var	MSE	Bias	Var	MSE
100	0.0118	0.0715	0.0717	-0.0857	0.0387	0.0460	-0.0753	0.1218	0.1275
200	0.0197	0.0262	0.0265	-0.0544	0.0185	0.0214	-0.0459	0.0506	0.0527
400	0.0166	0.0128	0.0130	-0.0364	0.0089	0.0102	-0.0285	0.0272	0.0280
800	0.0149	0.0068	0.0070	-0.0244	0.0050	0.0056	-0.0200	0.0135	0.0139

4.2 Testing Procedure

We now report simulations to examine the finite-sample performance of the bootstrapped-based test for the statistic $\hat{I}_n$. The model under the null hypothesis is a standard parametric stochastic frontier model of the form:

$$y_i = 1 + 0.5X_i + \nu_i - \mu_i,$$

where $X_i \sim N(0,1)$, $\nu_i \sim N(0, \sigma_\nu^2)$, and $\mu_i \sim N^+(0, \sigma_\mu^2)$. For what follows, we let $\sigma_\nu^2 = 1$ and $\sigma_\mu^2 = 1$.

The model under the alternative is generated from

$$y_i = \exp(Z_i) + \cos(Z_i)X_i + v_i - \exp(\delta_0 + \delta_1 Z_i)\eta_i,$$

where Z_i is generated on a equally-spaced grid between -2 and 2, $X_i \sim N(0,1)$, $v_i \sim N(0,\sigma_v^2)$, $\eta_i \sim N^+(0,1)$. For what follows, we let $\delta_0 = 1$, $\delta_1 = 1$, and $\sigma_v^2 = 1$.

The empirical size of the test statistic can be assessed by simulating data under the null, and empirical power can be assessed under the alternative. We choose to use two different bandwidth selection criteria: (1) the normal reference bandwidth which is given by $Z_{sd}n^{-1/5}$, where Z_{sd} is the standard deviation of Z_i, $i = 1,\ldots,n$, and (2) the data-driven least-squares cross-validation bandwidth (Li and Racine 2010). We draw $M = 1000$ Monte Carlo replications. For each replication, we compute $\hat{I}_n$ using the bandwidth for that particular replication, and conduct $B = 399$ bootstrap replications and compute the empirical p-value. The empirical rejection frequencies for $\alpha = 0.01, 0.05$, and 0.1 are reported in Tables 2 and 3. Table 2 shows that the test is better sized under the normal-reference bandwidth. This is because the least-squares cross-validated bandwidth can automatically detect and remove irrelevant variables before estimating the model. Since Z_i is irrelevant under the null, this can potentially increase the size of the test. Table 3 shows that the power of the test increases with n, and converges to 1.

Table 2: Empirical size for the proposed test

	Normal-reference bandwidth			LSCV bandwidth		
n	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
100	0.011	0.056	0.106	0.017	0.086	0.149
200	0.014	0.060	0.105	0.023	0.094	0.167
400	0.007	0.048	0.100	0.014	0.086	0.158
800	0.006	0.048	0.101	0.022	0.081	0.153

Table 3: Empirical power for the proposed test

	Normal-reference bandwidth			LSCV bandwidth		
n	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.1$
100	0.643	0.883	0.941	0.703	0.923	0.968
200	0.954	0.995	0.999	0.967	0.998	1.000
400	1.000	1.000	1.000	0.999	1.000	1.000
800	1.000	1.000	1.000	1.000	1.000	1.000

5 An Empirical Application

In this section, we consider estimation of stochastic production frontier in Norwegian forestry. The data, compiled by *Statistics Norway*, were drawn from a cross-section of 3249 active forest owners. All data are for the year 2003. The output variable consists of annual timber sales from the forest, measured in cubic meters. The labor input variable is the sum of hours worked by contractors and hours worked by the owner, his family or hired labor in 2003. The material input variable measures forest area cut in hectares, which is the

area of various types of final fellings in 2003. The capital input variable is the value of the increment from the forest. The forest owner can choose to cut the increment for either current or future period. Our choices of the environmental factors are: (1) income from outfield-related productions (i.e., recreational services), (2) income from agriculture, (3) wage income, (4) a binary variable with a value of 1 indicating there is a management plan, and 0 otherwise, (5) a binary variable with a value of 1 indicating the forest owner has an education level of Bachelor or higher, and 0 otherwise, (6) a binary variable with a value of 1 indicating its properties are located in central municipalities, and 0 otherwise. Lien, Størdal and Baardsen (2007) used this data to assess technical inefficiency of these Norwegian forests. Table 4 presents summary statistics in the sample. Further details on the source of the data and definitions of the variables were provided in their study.

Table 4: **Summary Statistics of the Variables**

Symbol	Variable Name	Mean	Sd.	Min.	Max.	Bandwidth[1]
y	Log of output (Harvesting level)	5.6680	1.665462	0.6931	10.74	-
x_1	Log of labor (Working hours)	2.882	1.637169	-2.072	7.876	-
x_2	Log of material (Forest area cut)	0.6692	1.522922	-4.4240	5.434	-
x_3	Log of capital (Value of increment)	11.780	1.184578	7.297	16.6	-
Z_1	Income from outfield related productions (1000NOK)	70.98	467.0222	0.00	11810	27.94376593
Z_2	Income from agricullture (1000NOK)	54.21	125.9468	0.00	2488	9.94951468
Z_3	Wage income (1000NOK)	240.3	269.1531	0.00	2183	122.03785955
Z_4	Management plan (0/1)	0.6898	0.4626668	0.00	1.00	0.21638380
Z_5	Education, Bachelor or higher (0/1)	0.2416	0.4281267	0.00	1.00	0.45613206
Z_6	Centrality (0/1)	0.3764	0.4845628	0.00	1.00	0.01324032

1. The bandwidths are selected via least-squares cross-validation.

We consider two specifications for the stochastic production frontier: (1) the semiparametric smooth-coefficient stochastic frontier model as described in (2) (i.e., with environmental factors which enter the coefficients and inefficiency nonparametrically), and (2) the standard parametric stochastic production frontier model as described in (8) (i.e., without environmental factors). Technical inefficiencies are calculated from both models using the JLMS method (Jondrow, Lovell, Materov and Schmidt 1982), after estimating all essential parameters and information. Specifically, for the first model (i.e., semiparametric), technical efficiency (TE) is calculated as $TE_i = \exp\left(-M(u_i|\varepsilon_i^*)\right)$, where $M(u_i|\varepsilon_i^*) = -\varepsilon_i^*\sigma_u^2(Z_i)/\sigma_i^2$ if $\varepsilon_i^* \leq 0$, and 0 otherwise; TE can be estimated in a similar fashion for the second model (i.e., parametric).

The average estimated technical efficiency is 0.97 for the semiparametric model and 0.89 for the standard parametric model. These results are comparable to Lien et al. (2007) who found the average technical

efficiency to be 0.90 using a different model. The semiparametric (parametric) model shows that almost 4% (7%) of the forest owners have an efficiency estimate of less than 0.75. To get an overall picture, the kernel distributions of the estimated technical efficiencies and the composite error term for the two models are reported in Figure 6 and 7, respectively, and those of the estimated functional parameters are reported in Figure 8. Figure 6 shows that, most of the forests are fully technically efficient under the semiparametric model, with the mode of technical efficiency around one; however, under the standard parametric model, the distribution is bi-modal (with the modes occurring at 0.9 and 1.0), and much fewer forests are estimated to be fully efficient. This may have some indication of the impact of model specification on the estimated technical efficiency. This indication is further revealed by Figure 7, which shows that the estimated composite error term (i.e., noise minus inefficiency) centers around zero under the semiparametric model, suggesting that inefficiency barely exists, while that under the standard parametric model has a mode that is less than zero, suggesting that inefficiency is more likely to exist.

Although the semiparametric model yields less variation in terms of TE, it generates more variation in the parameters. The distributions of the functional parameters in Figure 8 show that the semiparametric model better captures parameter heterogeneity while its standard parametric counterpart only yields estimates that are degenerate. More specifically, the labor, material, and capital productivity (represented by $\hat{\beta}_1(Z_i)$, $\hat{\beta}_2(Z_i)$, and $\hat{\beta}_3(Z_i)$, respectively) estimates under the standard parametric model only approximate the means of those estimates under the semiparametric model. With a micro-level data set, it is generally more interesting and informative to investigate each forest owner as opposed to an average forest owner. The density of $\hat{\sigma}^2(Z_i)$ under the semiparametric model in Figure 8 resembles that of a chi-squared distribution, and it obviously deviates from the degenerate $\hat{\sigma}^2$ estimate under the standard parametric model.

With all these differences in results between the semiparametric and its parametric counterpart, one would naturally perform specification test of one model against the other. We test the two models by testing the relevance of the environmental factors using the testing procedure described in section 3, since the semiparametric model without the environmental factors becomes the standard parametric model. The zero bootstrapped p-value suggests that these factors are relevant; and therefore the semiparametric model is preferred. This testing result is not very surprising based on the estimation results.

6 Conclusion

This paper proposes a semiparametric smooth-coefficient stochastic production frontier model, where all the coefficients, including intercept and slopes, along with the inefficiency term, are expressed as functions of a set of environmental factors. Thus, these factors affect the production frontier non-neutrally, as opposed

to traditional inputs which only affect the frontier neutrally. Using micro-level data, this model can yield a particular set of production frontier estimates for a particular, say, firm. Therefore, the potential heterogeneity of technology can be captured by this model. Since the environmental factors enter most parameters in the model nonparametrically and the elimination of these factors reduces the semiparametric model to its parametric counterpart, a testing procedure for the relevance of these factors is proposed. Monte Carlo study shows plausible finite sample behavior of our proposed estimation and inference procedure. An empirical example using real data is made and the advantages of the semiparametric approach over standard parametric approach are further revealed. A possible extension of this paper could be to relax the exponential functional form of the variance of the inefficiency term. This means, however, more work should be done to impose positivity constraint on the variance estimates.

References

Aigner, D. J., Lovell, C. A. K. and Schmidt, P. (1977), 'Formulation and estimation of stochastic frontier production functions', *Journal of Econometrics* **6**(1), 21–37.

Alvarez, A., Amsler, C., Orea, L. and Schmidt, P. (2006), 'Interpreting and testing the scaling property in models where inefficiency depends on firm characteristics', *Journal of Productivity Analysis* **25**(3), 201–212.

Battese, G. E. and Coelli, T. J. (1992), 'Frontier production functions, technical efficiency and panel data: With applications to paddy farmers in India', *Journal of Productivity Analysis* **3**, 153–169.

Battese, G. E. and Coelli, T. J. (1995), 'A model for technical inefficiency effects in a stochastic frontier production function for panel data', *Empirical Economics* **20**, 325–32.

Battese, G. E., Prasada Rao, D. S. and O'Donnell, C. (2004), 'A metafrontier production function for estimation of technical efficiencies and technology gaps for firms operating under different technologies', *Journal of Productivity Analysis* **21**, 91–103.

Caudill, S. B., Ford, J. M. and Gropper, D. M. (1995), 'Frontier estimation and firm-specific inefficiency measures in the presence of heteroskedasticity', *Journal of Business and Economic Statistics* **13**(1), 105–11.

Christensen, L. R., Jorgenson, D. W. and Lau, L. J. (1971), 'Conjugate duality and the transcendental logarithm production function', *Econometrica* **39**, 255–56.

Huang, C. J. and Liu, J.-T. (1994), 'Estimation of a non-neutral stochastic frontier production function', *Journal of Productivity Analysis* **5**, 171–180.

Jondrow, J., Lovell, C. A. K., Materov, I. S. and Schmidt, P. (1982), 'On the estimation of technical inefficiency in the stochastic frontier production function model', *Journal of Econometrics* **19**, 233–38.

Kumbhakar, S. C. (1990), 'Production frontiers, panel data, and time-varying technical efficiency', *Journal of Econometrics* **46**, 201–12.

Kumbhakar, S. C., Ghosh, S. and McGuckin, J. T. (1991), 'A generalized production frontier approach for estimating determinants of inefficiency in US dairy farms', *Journal of Business and Economic Statistics* **9**, 279–86.

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

Li, Q., Huang, C., Li, D. and Fu, T. (2002), 'Semiparametric smooth coefficient models', *Journal of Business and Economic Statistics* **20**(3), 412–422.

Li, Q. and Racine, J. (2007), *Nonparametric Econometrics: Theory and Practice*, Princeton University Press.

Li, Q. and Racine, J. S. (2010), 'Smooth varying-coefficient estimation and inference for qualitative and quantitative data', *Econometric Theory* **26**, 1607–1637.

Lien, G., Størdal, S. and Baardsen, S. (2007), 'Technical efficiency in timber production and effects of other income sources', *Small-scale Forestry* **6**, 65–78.

Meeusen, W. and van den Broeck, J. (1977), 'Efficiency estimation from Cobb-Douglas production functios with composed error', *International Economic Review* **18**(2), 435–44.

O'Donnell, C., Prasada Rao, D. S. and Battese, G. E. (2008), 'Metafrontier frameworks for the study of firm-level efficiencies and technology ratios', *Empirical Economics* **34**, 231–55.

Reifschneider, D. and Stevenson, R. (1991), 'Systematic departures from the frontier: A framework for the analysis of firm inefficiency', *International Economic Review* **32**(3), 715–23.

Saal, D., Parker, D. and Weyman-Jones, T. (2007), 'Determining the contribution of technical change, efficiency change and scale change to produtivity growth in the privatized English and Welsh water and sewerage industry: 1985-2000', *Journal of Productivity Analysis* **28**, 127–39.

Simar, L., Lovell, C. A. K. and van den Eeckaut, P. (1994), Stochastic frontiers incorporating exogenous influences on efficiency. Discussion Paper No.9403, Institut de Statistique, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.

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

Wang, H.-J. 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**, 129–44.

Zhang, R., Sun, K., Delgado, M. S. and Kumbhakar, S. C. (2012), 'Productivity in China's high technology industry: Regional heterogeneity and R&D', *Technological Forecasting & Social Change* **79**, 127–141.

Figure 2: **Performance of functional parameters: n=100**

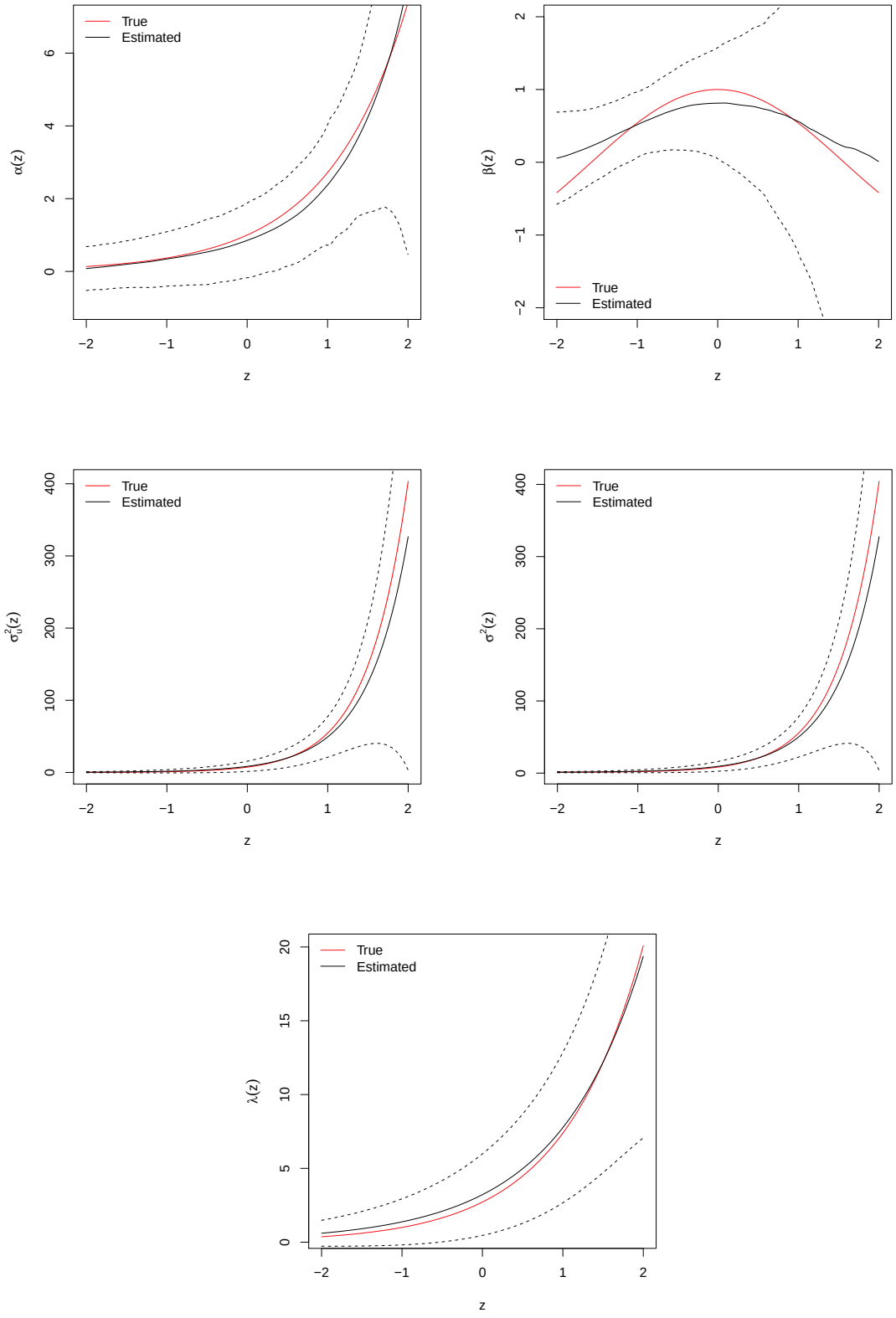


Figure 3: **Performance of functional parameters: n=200**

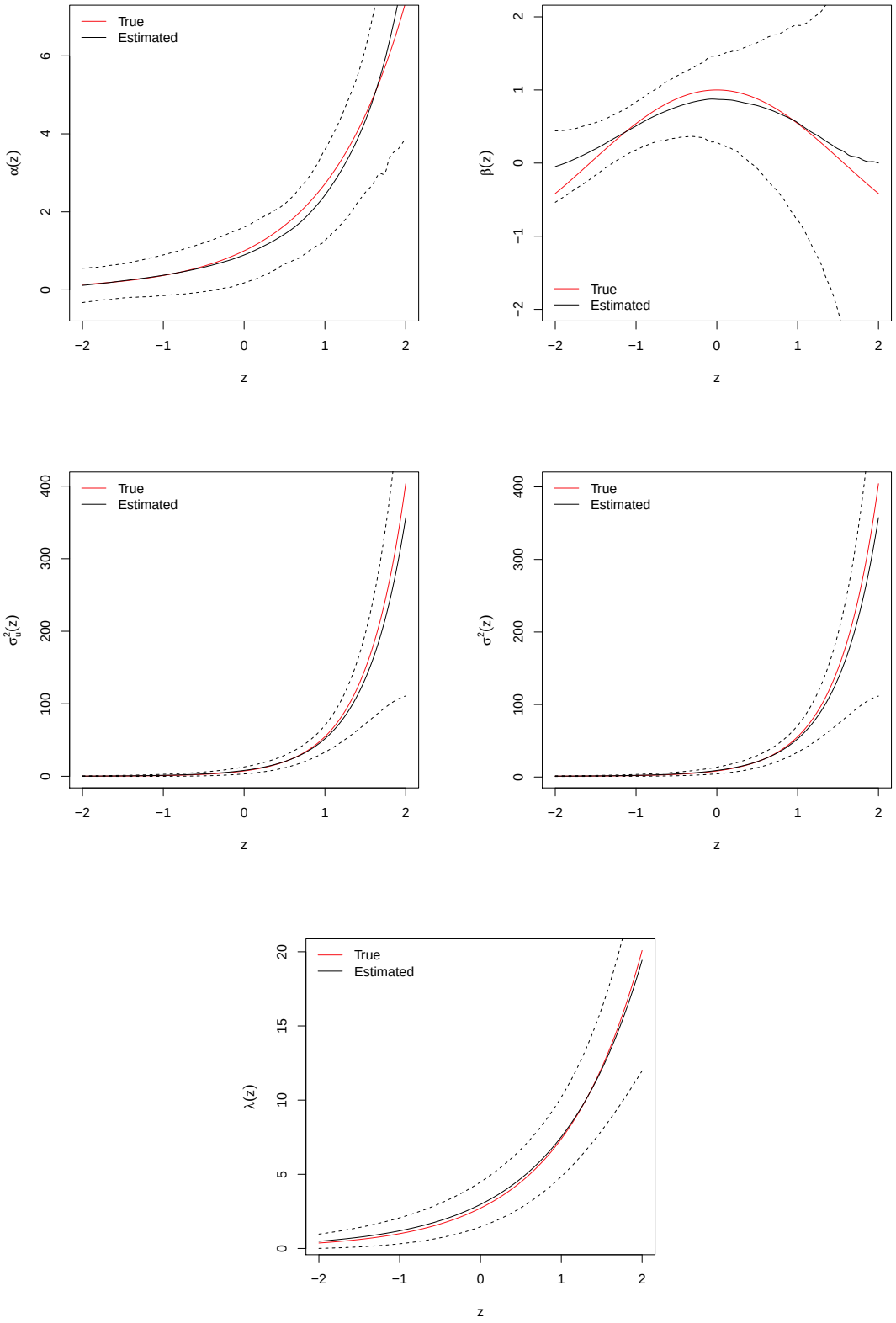


Figure 4: **Performance of functional parameters: n=400**

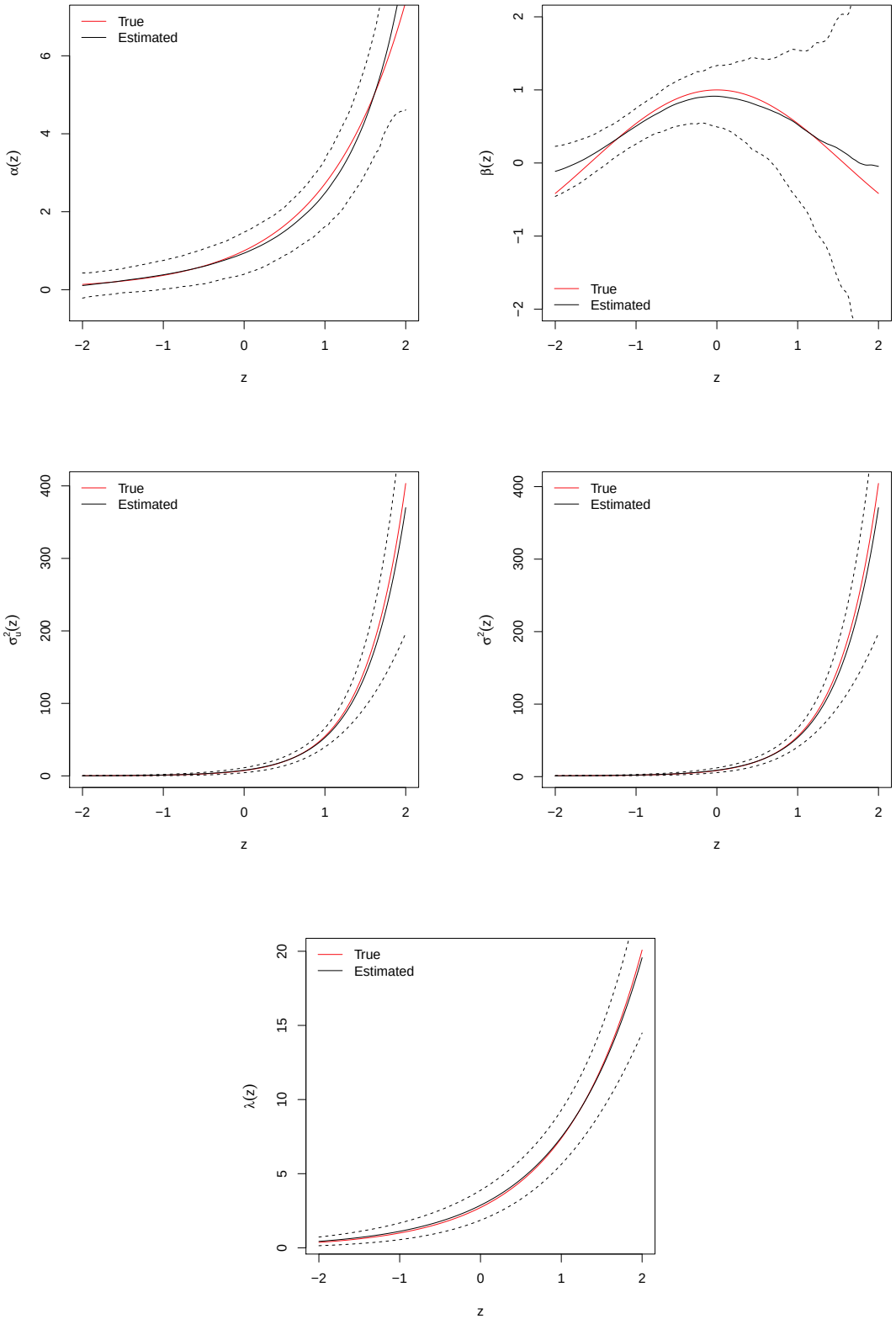


Figure 5: **Performance of functional parameters: n=800**

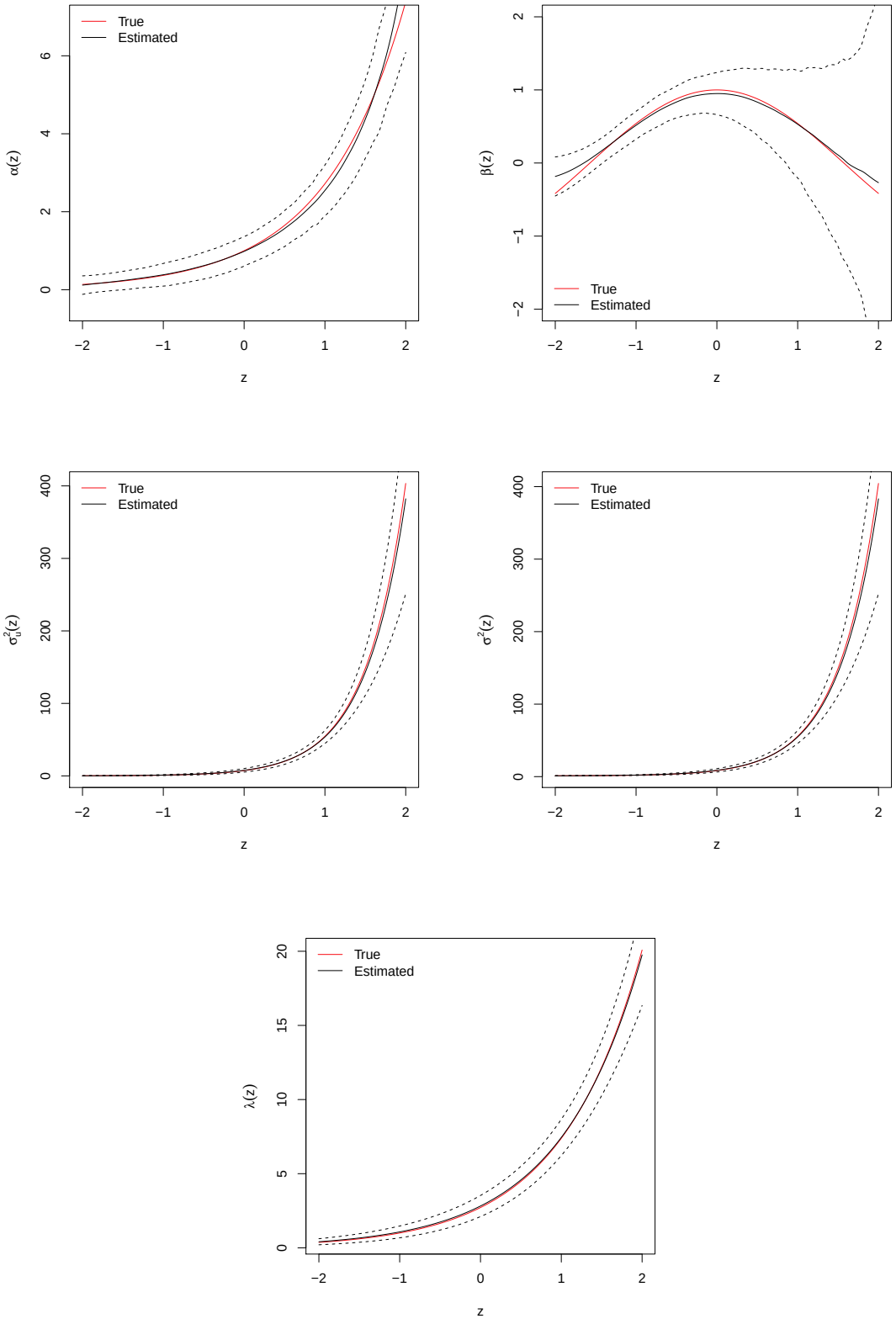


Figure 6: **Technical Efficiency**

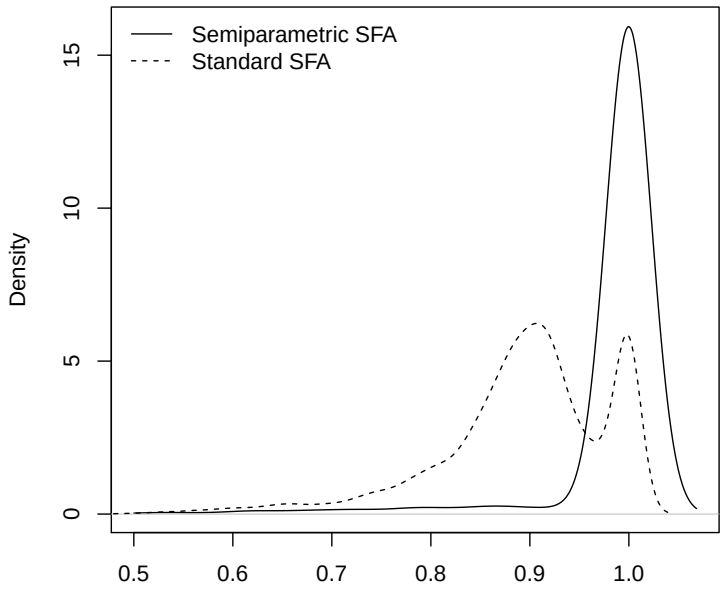


Figure 7: **Estimated Composite Error (noise minus inefficiency)**

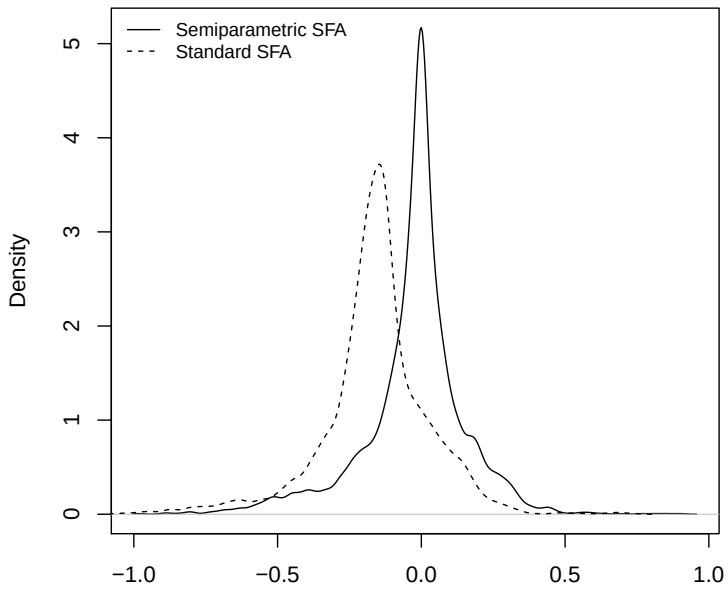


Figure 8: **Summary results**