Semiparametric Smooth Coefficient Models

Qi Li
Department of Economics, Texas A&M University, College Station, TX 77843 (*qi@econ.tamu.edu*)

Cliff J. Huang
College of Business, Feng Chia University, Seatwen, Taiwan

Dong Li
Department of Economics, Kansas State University, Manhattan, KS 66506 (*dongli@ksu.edu*)

Tsu-Tan Fu
Institute of Economics, Academia Sinica, Nankang, Taipei, Taiwan (*tfu@ieas.econ.sinica.edu.tw*)

In this article, we propose a semiparametric smooth coefficient model as a useful yet flexible specification for studying a general regression relationship with varying coefficients. The article proposes a local least squares method with a kernel weight function to estimate the smooth coefficient function. The consistency of the estimator and its asymptotic normality are established. A simple statistic for testing a parametric model versus the semiparametric smooth coefficient model is proposed. An empirical application of the proposed method is presented with an estimation of the production function of the nonmetal mineral industry in China. The empirical findings show that the intermediate production and management expense has played a vital role and is an unbalanced determinant of the labor and capital elasticities of output in production.

KEY WORDS: Semiparametric estimation; Smooth coefficient model; Specification test.

1. INTRODUCTION

Semiparametric and nonparametric estimation techniques have attracted much attention among econometricians and statisticians. One popular semiparametric specification is a partially linear model of the following form (e.g., Robinson 1988; Stock 1989):

$$y_i = \alpha(z_i) + x_i'\beta_0 + \epsilon_i, \tag{1}$$

where $x_i'\beta_0$ is the parametric component and $\alpha(z_i)$ is the nonparametric part of the model [the functional form of $\alpha(\cdot)$ is not specified].

In this article, we consider a more general semiparametric regression model: a semiparametric smooth coefficient model. A semiparametric smooth coefficient model nests a partially linear model as a special case and it is given by

$$y_i = \alpha(z_i) + x_i'\beta(z_i) + \epsilon_i, \tag{2}$$

where $\beta(z_i)$ is a vector of unspecified smooth functions of z_i. When $\beta(z_i) = \beta_0$, model (2) reduces to (1). The smooth coefficient model is an appropriate setting, for example, in the framework of a cross-sectional production function where the right-hand-side variables are labor, capital, and firm's R&D inputs. If we let $x_i = (labor_i, capital_i)$ and $z_i = \text{R\&D}_i$, then model (2) suggests that the labor and capital input coefficients may vary directly with the firm's R&D input. Thus, both the marginal productivity of labor and the capital depend on the firm's R&D values. As a result, the returns to scale may also be a function of R&D. The partially linear model (1) assumes the slope coefficients β_0 are invariant to R&D, and the R&D variable can only shift the level of the production frontier. In this case, the R&D variable is said to have "neutral" effects on the production frontier. In contrast to model (1), our smooth coefficient model (2) allows R&D to affect the stochastic frontier "nonneutrally." Section 4 provides an empirical example similar to this setting. There is a rich literature for using a one-sided error term to describe production efficiency. Aigner, Lovell, and Schmidt (1977) suggested using a one-sided normal error term to describe a firm's production inefficiency, which results in a parallel level shift of the firm's production function from an efficient production function. For using panel data to relax some of the restrictive assumptions made in Aigner et al. (1977), see Kumbhakar (1990) and Park, Sickles, and Simar (1998). See Fan and Li (1996) for nonparametric estimation of a production frontier function with an error structure as suggested by Aigner et al. (1977).

The time series smooth transition autoregressive (STAR) model is another example of the smooth coefficient model, $y_t = \alpha(y_{t-d}) + x_t'\beta(y_{t-d}) + \epsilon_t$, where $\alpha(y_{t-d})$ and $\beta(y_{t-d})$ are bounded. Other related models are considered in Chen and Tsay (1993) and Hastie and Tibshirani (1993), who considered the autoregressive model of the form $y_t = f_1(y_{t-d})y_{t-1} + f_2(y_{t-d})y_{t-2} + \cdots + f_p(y_{t-d})y_{t-p} + \epsilon_t$, where the functional form of $f_j(\cdot)$'s $(j = 1, \ldots, p)$ are not specified. When $d > p$, y_{t-d} and $(y_{t-1}, \ldots, y_{t-p})$ are nonnested, and their model is similar to our model (2). However, when $1 \le d \le p$, y_{t-d} is contained in $(y_{t-1}, \ldots, y_{t-p})$. Model (2) does not cover this case. Chen and Tsay (1993) and Hastie and Tibshirani (1993) discussed the identification of $f_j(\cdot)$ for the general case and suggested some recursive algorithms to estimate the unknown $f_j(\cdot)$ functions.

In this article, we only consider the case in which x_i and z_i are nonnested so that we do not need recursive or backfitting

Journal of Business & Economic Statistics
July 2002, Vol. 20, No. 3
DOI 10.1198/073500102288618531

algorithms as in Chen and Tsay (1993) or Hastie and Tibshirani (1993). As a result, the asymptotic distribution of our semiparametric estimator can be easily established. Also, we suggest a simple consistent test for testing a parametric model versus the semiparametric model (2).

The semiparametric model (2) has the advantage that it allows more flexibility in functional form than a parametric linear model or a semiparametric partially linear specification of (1). Further, the sample size required to obtain a reliable semiparametric estimation is not as large as that required for estimating a nonparametric model. Hence, a semiparametric smooth coefficient model provides a flexible specification and should be useful for applied researchers. It should be noted that when the dimension of z_i is greater than 1, our model (2) also suffers the curse of dimensionality problem, although to a lesser extent than a purely nonparametric model where both x_i and z_i enter the model nonparametrically.

The rest of the article is organized as follows. Section 2 presents a semiparametric smooth coefficient model and establishes the asymptotic properties of our proposed semiparametric estimator. Section 3 proposes a test for a parametric model against a semiparametric alternative. Section 4 uses the semiparametric smooth coefficient model to study the production function of the nonmetal mineral industry in China. Concluding remarks are given in Section 5.

2. SEMIPARAMETRIC SMOOTH COEFFICIENT MODEL

Model (2) can be expressed more compactly as

$$y_i = \alpha(z_i) + x_i'\beta(z_i) + \epsilon_i = (1, x_i')\begin{pmatrix}\alpha(z_i)\\ \beta(z_i)\end{pmatrix} + \epsilon_i$$
$$\equiv X_i'\delta(z_i) + \epsilon_i, \quad (3)$$

where $\delta(z_i) = (\alpha(z_i), (\beta(z_i))')'$. $\delta(z_i)$ is a vector of smooth, but unknown functions of z_i, x_i is a $p \times 1$ vector, and z_i is of dimension q. The model is, in fact, similar to the time-varying coefficient model considered by Robinson (1989):

$$y_t = x_t'\beta(t) + \epsilon_t, \qquad t = 1, \ldots, n, \quad (4)$$

where $\beta(t)$ is a smooth, unknown function of time t. Model (3) differs from model (4) in that the variable $z_i \in R^q$ is stochastic, whereas $t \in \{0, 1, \ldots, n\}$ is nonstochastic. Robinson proposed a local least squares method to estimate $\beta(t)$ using kernel weight function.

Following Robinson, we propose the following local least squares method to estimate $\delta(z)$:

$$\hat{\delta}(z) = \left[(nh^q)^{-1}\sum_{j=1}^{n} X_j X_j' K\left(\frac{z_j - z}{h}\right)\right]^{-1}$$
$$\times \left\{(nh^q)^{-1}\sum_{j=1}^{n} X_j y_j K\left(\frac{z_j - z}{h}\right)\right\}$$
$$\equiv [D_n(z)]^{-1} A_n(z), \quad (5)$$

where $D_n(z) = (nh^q)^{-1}\sum_j X_j X_j' K((z_j - z)/h)$, $A_n(z) = (nh^q)^{-1}\sum_j X_j y_j K((z_j - z)/h)$, $K(\cdot)$ is a kernel function, and $h = h_n$ is a smoothing parameter.

The intuition behind the preceding local least squares estimator is apparent. Let us assume that z is a scalar and $K(\cdot)$ is a uniform kernel. In this case, (5) becomes

$$\hat{\delta}(z) = \left[\sum_{|z_j - z| \le h} X_j X_j'\right]^{-1} \sum_{|z_j - z| \le h} X_j y_j. \quad (6)$$

$\hat{\delta}(z)$ is simply a least squares estimator obtained by regressing y_j on X_j using the observations of (X_j, y_j) where the corresponding z_j is close to $z(|z_j - z| \le h)$. Because $\delta(z)$ is a smooth function of z, $|\delta(z_j) - \delta(z)|$ is small when $|z_j - z|$ is small. The condition that nh is large ensures that we have sufficient observations within the interval $|z_j - z| \le h$ when $\delta(z_j)$ is close to $\delta(z)$. Therefore, under conditions such as $h \to 0$ and $nh \to \infty$ ($nh^q \to \infty$ if $z_i \in R^q$), one can show that the local least squares estimator $\hat{\delta}(z)$ provides a consistent estimate of $\delta(z)$.

The following theorem establishes the consistency and asymptotic normality of $\hat{\delta}(z)$.

Theorem 2.1. Under conditions (A.1) and (A.2) given in the Appendix, and for a fixed value of z with $f_z(z) > 0$ [$f_z(\cdot)$ is the marginal density function of z_i], we have (a) $\hat{\delta}(z) - \delta(z) \to 0$ in probability, and (b) if, in addition to (A.1) and (A.2), we also have $nh^{q+4} \to 0$ as $n \to \infty$, then $\sqrt{nh^q}(\hat{\delta}(z) - \delta(z)) \to \mathrm{N}(0, \Omega_z)$ in distribution, provided that $M_z \stackrel{\text{def}}{=} f_z(z)E[X_1X_1'|z_1 = z]$ is positive definite, where $\Omega_z = M_z^{-1}V_zM_z^{-1}$, $V_z = f_z(z)E[X_1X_1'\sigma_\epsilon^2(x_1, z_1)|z_1 = z][\int K^2(u)\,du]$, and $\sigma_\epsilon^2(x_1, z_1) = E(\epsilon_1^2|x_1, z_1)$.

Moreover, Ω_z can be consistently estimated by $\widehat{\Omega}_z = \widehat{M}_z^{-1}\widehat{V}_z\widehat{M}_z^{-1}$, $\widehat{M}_z = (nh^q)^{-1}\sum_i X_iX_i'K_{iz}$ with $K_{iz} \equiv K((z_i - z)/h)$, $\widehat{V}_z = (nh^q)^{-1}\sum_i X_iX_i'\hat{\epsilon}_i^2K_{iz}^2$, and $\hat{\epsilon}_i = y_i - X_i'\hat{\delta}(z_i)$.

The proof of Theorem 2.1 is presented in the Appendix.

3. TESTING PARAMETRIC VERSUS SEMIPARAMETRIC MODELS

Consider a parametric specification of model (2):

$$y_i = X_i'\delta_0(z_i) + \epsilon_i, \quad (7)$$

where $\delta_0(z_i)$ is a parametric function of z_i. For example, consider the simple case where z_i is a scalar and $\delta_0(z_i) = (\alpha_0 + z_i\gamma_0, \beta_0')'$. We have a standard linear regression model

$$y_i = \alpha_0 + z_i\gamma_0 + x_i'\beta_0 + \epsilon_i \equiv X_i'\delta_0(z_i) + \epsilon_i, \quad (8)$$

where $\delta_0(z_i) = (\alpha_0 + z_i\gamma_0, \beta_0')'$.

Even though model (3) is more general than model (7), one may want to estimate the parametric model (7) if it is, in fact, the true model. It is usually more efficient (in finite-sample applications) to estimate a correctly specified parametric model than to estimate a semiparametric model. However, if the semiparametric model (3) is a correct specification, but model (7) is not, the estimation results based on the misspecified parametric model (7) will usually lead to inconsistent estimation results. Therefore, in practice, it is of interest to test whether the parametric model (7) is an adequate description of the data. In the following, we propose a simple statistic for testing the parametric model (7) versus the semiparametric smooth coefficient model (3).

The null hypothesis that model (7) is a correct specification can be stated as follows: H_0: $\delta(z) - \delta_0(z) = 0$ almost everywhere. The alternative hypothesis is that model (3) is the correct specification, but not model (7): $H_1 : \delta(z) - \delta_0(z) \neq 0$ on a set with positive measure.

We will use integrated squared differences $I \stackrel{\text{def}}{=} \int [\delta(z) - \delta_0(z)]'[\delta(z) - \delta_0(z)]\, dz$ as the basis of our test. For using other methods to construct consistent model specification tests, see Bierens (1982), Wooldridge (1992), and Bierens and Ploberger (1997), among others. Notice that $I = 0$ under H_0 and $I > 0$ under H_1. Therefore, I serves as a proper measure for testing H_0 versus H_1. A feasible test statistic can be obtained by replacing $\delta(z)$ with the local least squares estimate $\hat{\delta}(z)$ given by (5) and by replacing $\delta_0(z)$ with its estimate $\hat{\delta}_0(z_i)$. In the case of a linear parametric function with a scalar z_i, $\hat{\delta}_0(z_i) = (\hat{\alpha}_0 + z\hat{\gamma}_0, z_i\hat{\beta}_0)'$, where $(\hat{\alpha}_0, \hat{\gamma}_0, \hat{\beta}_0)$ is the least squares estimate of $(\alpha_0, \gamma_0, \beta_0)$.

Because the random denominator $D_n(z)$ in (5) is not bounded away from 0, it is difficult to derive the asymptotic distribution of the test statistic based on the integrated squared differences I. To avoid this random denominator problem, we propose a test statistic based on a weighted version of I with $D_n(z)$ as the weight function,

$$I_n = \int \{D_n(z)[\hat{\delta}(z) - \hat{\delta}_0(z)]\}' \{D_n(z)[\hat{\delta}(z) - \hat{\delta}_0(z)]\} dz. \quad (9)$$

Using the identity that $D_n(z)\hat{\delta}(z) = A_n(z)$ [see (5)], we have

$$\begin{aligned} I_n &= \int [A_n(z) - D_n(z)\hat{\delta}_0(z)]'[A_n(z) - D_n(z)\hat{\delta}_0(z)]\, dz \\ &= (nh^q)^{-2} \sum_i \sum_j \int X_i'[y_i - X_i'\hat{\delta}_0(z)] X_j[y_j - X_j'\hat{\delta}_0(z)] \\ &\quad \times K\left(\frac{z_i - z}{h}\right) K\left(\frac{z_j - z}{h}\right) dz. \end{aligned} \quad (10)$$

The test statistic I_n given in (10) involves a q-dimensional (possibly numerical) integration and is not easy to compute in practice. It can also be shown that I_n involves a "nonzero center term" under H_0, which can cause substantial finite sample bias in testing H_0; see Li (1996) for some Monte Carlo evidence on this issue. In the following equation, we use some simple tricks to construct a new test statistic that does not involve numerical integration or a "nonzero center term."

Removing the $i = j$ term from (10) and replacing $\hat{\delta}_0(z)$ in the first set of brackets by $\hat{\delta}_0(z_i)$ and $\hat{\delta}_0(z)$ in the second set of brackets by $\hat{\delta}_0(z_j)$, we have a new test statistic,

$$\begin{aligned} \tilde{I}_n &= \frac{1}{n^2 h^{2q}} \sum_i \sum_{j \neq i} X_i'\big(y_i - X_i'\hat{\delta}_0(z_i)\big) X_j\big(y_j - X_j'\hat{\delta}_0(z_j)\big) \\ &\quad \times \int K\left(\frac{z_i - z}{h}\right) K\left(\frac{z_j - z}{h}\right) dz \\ &= \frac{1}{n^2 h^q} \sum_i \sum_{j \neq i} X_i'\big(y_i - X_i'\hat{\delta}_0(z_i)\big) \\ &\quad \times X_j\big(y_j - X_j'\hat{\delta}_0(z_j)\big) \bar{K}\left(\frac{z_i - z_j}{h}\right), \end{aligned} \quad (11)$$

where $\bar{K}(v) \stackrel{\text{def}}{=} \int K(u)K(u + v)\, du$ is the twofold convolution kernel derived from $K(\cdot)$.

Dropping the $i = j$ term in (10) removes a center term in I_n, and replacing $\delta(z)$ by $\delta(z_i)$ [or $\delta(z_j)$] gets rid of the integration in I_n. The reason that we can replace $\delta(z)$ by $\delta(z_i)$ [or $\delta(z_j)$] is that only z close to both z_i and z_j are important in the integration due to the kernel weight function $K((z_i - z)/h)K((z_j - z)/h)$.

In practice, one can use the product kernel $K(u) = \prod_{l=1}^q k(u_l)$ for computation. In this case, the convolution kernel is also a product kernel $\bar{K}(u) = \prod_{l=1}^q \bar{k}(u_l)$, where $k(\cdot)$ and $\bar{k}(\cdot)$ are univariate kernel functions. If $k(\cdot)$ is a standard normal kernel, that is, $k(x) = e^{-x^2/2}/\sqrt{2\pi}$, then the convolution kernel is $\bar{k}(x) = e^{-x^2/4}/\sqrt{4\pi}$ [an N(0, 2) density function]. This is because $\bar{k}(\cdot)$ is the density function obtained from the *sum* of two independent N(0, 1) random variables. Thus, in contrast to I_n in (10), the test statistic $\tilde{I}_n$ does not involve any integration.

In fact, one does not even have to use the convolution kernels. Simply replacing $\bar{K}((z_i - z_j)/h)$ by $K((z_i - z_j)/h)$ in (11) provides a simple consistent test. Therefore, our proposed test statistic is

$$\begin{aligned} \hat{I}_n &= \frac{1}{n^2 h^q} \sum_i \sum_{j \neq i} X_i'\big(y_i - X_i'\hat{\delta}_0(z_i)\big) \\ &\quad \times X_j\big(y_j - X_j'\hat{\delta}_0(z_j)\big) K\left(\frac{z_i - z_j}{h}\right) \\ &= \frac{1}{n^2 h^q} \sum_i \sum_{j \neq i} X_i' X_j \hat{\epsilon}_i \hat{\epsilon}_j K\left(\frac{z_i - z_j}{h}\right), \end{aligned} \quad (12)$$

where $K(\cdot)$ is a second-order kernel function [see condition (A.2) (i) in the Appendix]. $K(\cdot)$ could be the same kernel function as used in (5). The residual $\hat{\epsilon}_i = y_i - X_i\hat{\delta}_0(z_i)$ is obtained from the parametric model (7). Besides the advantages of not needing numerical integration or a nonzero center term, another advantage of using the $\hat{I}_n$ test of (12) over the I_n test of (10) is that the regularity conditions required for deriving the asymptotic normality of $\hat{I}_n$ (under H_0) are weaker than those required for I_n. As shown in the Appendix, for the conditions on smoothing parameter h, $h \to 0$ and $nh^q \to \infty$ (as $n \to \infty$) are sufficient for $nh^{q/2}\hat{I}_n$ to have an asymptotic normal distribution with a simple asymptotic variance under H_0, and $nh^{q/2}\hat{I}_n \to +\infty$ under H_1. However, it can be shown that the asymptotic variance of I_n is different, depending on whether the data are undersmoothed ($nh^{q+4} \to 0$), optimally smoothed ($nh^{q+4} \to$ a positive constant), or oversmoothed ($nh^{q+4} \to \infty$); see Hall (1984) and Fan (1994) for more details on how different amounts of smoothing can lead to different asymptotic variances for some nonparametric kernel-based tests.

Theorem 3.1. Assume that conditions (A.1) and (A.2) given in the Appendix hold. Then (a) under H_0, $J_n = nh^{q/2}\hat{I}_n/\hat{\sigma}_0 \to$ N(0, 1) in distribution, where

$\hat{\sigma}_0^2 = 2(n^2h^q)^{-1}\sum_i\sum_{j\neq i}\hat{\epsilon}_i^2\hat{\epsilon}_j^2(X_i'X_j)^2K^2((z_i-z_j)/h)$ is a consistent estimator of

$$\sigma_0^2 = 2f_z(z)E\big[(X_1'X_2)^2\sigma_\epsilon^4(x_1,z_1)|z_1=z\big]\Big[\int K^2(v)\,dv\Big];$$

(b) under H_1, $\text{Prob}[J_n > B_n] \to 1$ as $n\to\infty$, where B_n is any nonstochastic sequence with $B_n = o(nh^{q/2})$.

The proof of Theorem 3.1 is also presented in the Appendix.

Theorem 3.1 states that the test statistic $J_n = nh^{q/2}\widehat{I}_n/\hat{\sigma}_0$ is a consistent test for testing H_0 versus H_1. Note that, as $n\to\infty$, $J_n\to+\infty$ under H_1. Therefore, it is a one-sided test. In practice, H_0 is rejected if $J_n > c_\alpha$ at the significant level α, where c_α is the upper αth percentile from a standard normal distribution.

Here we would like to make some comments on the differences and similarities of the testing problem compared with those of Fan and Li (1996, 1999) and Zheng (1996) (hereafter FLZ). FLZ considered a conditional moment test of the form $E(\epsilon|z)=0$ almost surely, where ϵ is the error term from a regression model. They constructed their tests based on $E[\epsilon E(\epsilon|z)f(z)]$. After replacing $E(\epsilon|z)f(z)$ by a kernel estimator and replacing $E(\cdot)$ by the sample mean of $(\cdot)$, their test statistic had a simple form containing a double summation. The testing problem we consider here cannot be written as a conditional zero mean of the error term ϵ. Therefore, we choose to use integrated squared differences as the starting point for construction of our test. Unfortunately, the integrated squared differences–based test I_n involves numerical integration and is not convenient to compute. Its asymptotic distributions are complicated and depend on whether the data are undersmoothed, optimally smoothed, or oversmoothed. After using some simple manipulations, we are able to obtain a much simpler test that does not involve numerical integration. The final version of our test, the $\widehat{I}_n$ test, only has a double summation, which is very similar to the tests considered in FLZ. Consequently, the asymptotic distribution of $\widehat{I}_n$ is relatively simple and is similar to the asymptotic analysis of FLZ.

Finally, we would like to make some comments on how to choose the smoothing parameters. Both Theorem 2.1 and Theorem 3.1 require the smoothing parameter h to satisfy the usual conditions that $h\to0$ and $nh^q\to\infty$ as $n\to\infty$. In practice, one can use some rule-of-thumb formula like $h_l = z_{l,\text{sd}}n^{-1/(4+q)}$, where $z_{l,\text{sd}}$ is the sample standard deviation of $\{z_{l,i}\}_{i=1}^n$, $l=1,\ldots,q$. When $q=1$, z_i is normally distributed, and $K(\cdot)$ is a Gaussian kernel function, the so-called normal-reference rule is to choose $h=1.06\,z_{\text{sd}}n^{-1/5}$; see Silverman (1986) for a more detailed discussion of this (we thank a referee for pointing out the normal-reference rule to us). Alternatively, one may use some data-driven method such as the least squares cross-validation method to select h. Although we did consider this case in Theorems 2.1 and 3.1, it is natural to conjecture that the results of Theorems 2.1 and 3.1 remain valid with the least squares cross-validation selection of h.

4. PRODUCTION FUNCTION OF THE NONMETAL MINERAL MANUFACTURING INDUSTRY IN CHINA

In this section, we consider estimation of a production function in China's nonmetal mineral manufacturing industry to illustrate the application of the semiparametric smooth coefficient model. The data used in this article are drawn from the Third Industrial Census of China conducted by the National Statistical Bureau of China in 1995. The Third Industrial Census of China is currently the most comprehensive industrial survey in China. We use the data on all the firms in the nonmetal mineral manufacturing industry (code 31 in the survey) in this article. There are several reasons for choosing the nonmetal mineral manufacturing industry in our article: (1) To avoid heterogeneity across different industries, we have to include only those firms in the same industry. At the same time, we want to have a sufficiently large sample size to carry out the semiparametric estimation. This leaves us with seven industries from which to choose. (2) Due to the effort of the government since the early 1980s, there was significant foreign investment in most Chinese industries in 1995. It has been found that production performance is quite different across different ownership types (see, e.g., Murakami, Liu, and Otsuka 1994), which leads to different production functions across different ownership types, even within the same industry. For this reason, we want to study an industry in which there were very few firms owned or operated by foreign entities. The nonmetal mineral manufacturing industry is a suitable candidate according to this criterion. There were only 67 firms (wholly or partly) owned by foreign investors out of 1,473 firms in the nonmetal mineral industry in 1995. We removed the firms (wholly or partly) owned by foreign investors from our sample. This leaves 1,406 observations in the sample. The nonmetal mineral manufacturing industry includes cement manufacturing, asbestos manufacturing, glass manufacturing, and so forth.

Value added in thousand renminbi (hereafter RMB) (Y), value of capital assets in thousand RMB (K), average number of employees (L), and intermediate production and management expense in thousand RMB (Z) are the main variables used in the estimation of the production function. There were usually small fluctuations in the number of employees throughout the year. The reported number (L) is the average within the year reported. The yearly average number of employees of each firm includes all permanent, contracted, and temporary workers. The intermediate production and management expense includes those production expenses and management expenses that are not directly related to the production of output. For example, it includes research and development (R&D), upgrading of the existing equipment, and employee training. Note that R&D is part of the intermediate production and management expense, but R&D itself is not available in this dataset. It is hypothesized that firms with more intermediate production and management expense tend to produce more output (holding other inputs equal), and this may alter the marginal product of labor and capital.

In the production function, output is the value added (Y). The two inputs are the value of capital assets (K) and the average number of employees (L). Intermediate production and management expense (Z) is included in the production estimation as an argument of the unknown smooth coefficient function in the semiparametric frontier to examine the possible nonneutral efficiency effect in production. Table 1 shows the summary statistics for the logged variables in the dataset.

Table 1. Summary Statistics

Variable	Mean	Standard deviation	Minimum	Maximum
$\ln Y$	9.28	1.19	1.39	13.15
$\ln K$	11.11	.79	8.76	14.39
$\ln L$	6.81	.73	2.83	9.93
$\ln Z$	6.56	1.53	2.89	9.81

The semiparametric model we use allows the coefficients of $\ln K$ and $\ln L$ to vary with the variable $z = \ln Z$ without specifying the functional form:

$$\ln Y = \alpha(z) + \beta_K(z) \ln K + \beta_L(z) \ln L + \epsilon. \quad (13)$$

To estimate the semiparametric model (13), a standard normal kernel function $K(u) = e^{-u^2/2}/\sqrt{2\pi}$ is used to estimate $\delta(z) = (\alpha(z), \beta_K(z), \beta_L(z))$. The smoothing parameter is chosen via $h = z_{\text{sd}} n^{-1/5}$, where n is the sample size and z_{sd} is the sample standard deviation of z.

We will compare our semiparametric estimation results with two parametric models. The first one is a Cobb–Douglas production function:

$$\ln Y = \alpha_0 + \alpha_1 z + \alpha_2 z^2 + \beta_K \ln K + \beta_L \ln L + \epsilon. \quad (14)$$

In model (14), the intermediate production and management expense z is assumed to neutrally shift the production function, that is, $\alpha_0(z) = \alpha_0 + \alpha_1 z + \alpha_2 z^2$, and all other coefficients are constant, $\beta_j(z) = \beta_j$, $j = K, L$.

The second parametric model is a translog model that allows interaction terms:

$$\ln Y = \alpha_0 + \alpha_1 z + \alpha_2 z^2 + \beta_{1K} \ln K + \beta_{2K} \ln^2 K + \beta_{1L} \ln L + \beta_{2L} \ln^2 L + \beta_{Kz} z \ln K + \beta_{Lz} z \ln L + \beta_{KL} \ln K \ln L + \epsilon. \quad (15)$$

The results of the two parametric models (14) and (15) are given as follows:

$$\underset{(.03)}{\ln Y = -.016} + \underset{(.23)}{.027z} + \underset{(1.23)}{.011 z^2} + \underset{(14.85)}{.573 \ln K} + \underset{(8.59)}{.332 \ln L}$$

$$R^2 = .425, \qquad \text{Adjusted } R^2 = .423 \quad (16)$$

and

$$\ln Y = \underset{(1.08)}{3.721} + \underset{(.71)}{.197z} + \underset{(1.53)}{.017z^2} - \underset{(.48)}{.313 \ln K} + \underset{(.26)}{.010 \ln^2 K} + \underset{(1.08)}{.525 \ln L} + \underset{(.06)}{.002 \ln^2 L} + \underset{(1.68)}{.051 z \ln K} - \underset{(4.10)}{.0120 z \ln L} + \underset{(.84)}{.049 \ln K \ln L}$$

$$R^2 = .433, \qquad \text{Adjusted } R^2 = .430, \quad (17)$$

where the t statistics are given in parentheses.

In the simple production function model (16), the coefficients on both capital and labor are positive and significant. Even though the coefficients on z and z^2 are not significant, a joint test of these two coefficients yields a test statistic $F(2, 1{,}401) = 42.96$, which is highly significant [the 1% $F(2, 1{,}401)$ critical value is 4.61]. We choose a quadratic functional form for z in model (14) to allow for possible nonlinearity in z. We also estimated a simple model by removing the z^2 term from (14). The t statistic for the coefficient on z becomes 9.19. Therefore, the z variable has a significant positive effect on output. In fact, we also estimated a partially linear model where z enters the model nonparametrically. The estimation result is similar to that of the Cobb–Douglas model and is not reported here to save space.

From the translog model output [Eq. (17)], we observe that there are many insignificant coefficients. Our semiparametric model (13) does not nest model (15). However, if $\beta_{2K} = \beta_{2L} = \beta_{KL} = 0$, then model (15) becomes nested in model (13). A joint F test for testing $\beta_{2K} = \beta_{2L} = \beta_{KL} = 0$ gives a value of $F(3, 1{,}396) = .66$ with Prob $> F = .58$. Therefore, we do not reject the null hypothesis that $\beta_{2K} = \beta_{2L} = \beta_{KL} = 0$. The estimation results after removing $\ln^2 K, \ln^2 L$, and $\ln K \ln L$ are as follows:

$$\ln Y = \underset{(.13)}{.200} - \underset{(-.11)}{.024z} + \underset{(1.21)}{.013z^2} + \underset{(.93)}{.158 \ln K} + \underset{(6.15)}{.992 \ln L} + \underset{(2.60)}{.065 z \ln K} - \underset{(-4.22)}{.103 z \ln L}$$

$$R^2 = .433, \qquad \text{Adjusted } R^2 = .430. \quad (18)$$

Note that model (18) has the same R^2 value (up to the third decimal) as that of the more general model (15). This gives further evidence that $\ln^2 K$, $\ln^2 L$, and $\ln K \ln L$ are irrelevant regressors.

Table 2 tabulates the mean values and the 10th, 50th (median), and 90th percentiles of the estimates based on the semiparametric smoothing coefficient model. Because the semiparametric estimators of the smooth coefficients are functions of z, we plot the coefficients and their pointwise 90% and 95% bounds in Figures 1 and 2, respectively. For comparison, we also plot their counterparts from the Cobb–Douglas model (16). The coefficients $\hat{\beta}_j(z)$ and $\hat{\beta}_j$, $j = K, L$, are plotted against the intermediate production and management expense variable z. We also plot the intercept term and the returns to scale $\beta_K(z) + \beta_L(z)$. Except for the intercept, which is $\alpha_0(z) = \alpha_0 + \alpha_1 z + \alpha_2 z^2$, all other coefficient estimates in the parametric model (16) are constant over z. Comparing the intercept [Fig. 1(a)] in the parametric model with that in the semiparametric model, we see that the intercept in the semiparametric model is relatively flat, whereas it is increasing in z in the parametric model. The reason is that, in the Cobb–Douglas model, the coefficients on capital and labor are restricted to be constant and thus any (positive) effect of intermediate production and management expense on output is only reflected

Table 2. Semiparametric Estimation

Coefficient	Mean	10th percentile	Median	90th percentile
$\beta_K(z)$	.645	.615	.661	.701
$\beta_L(z)$	.336	.275	.336	.378
$\alpha(z)$	−.176	−.319	−.241	−.122
$R^2 = .452$				

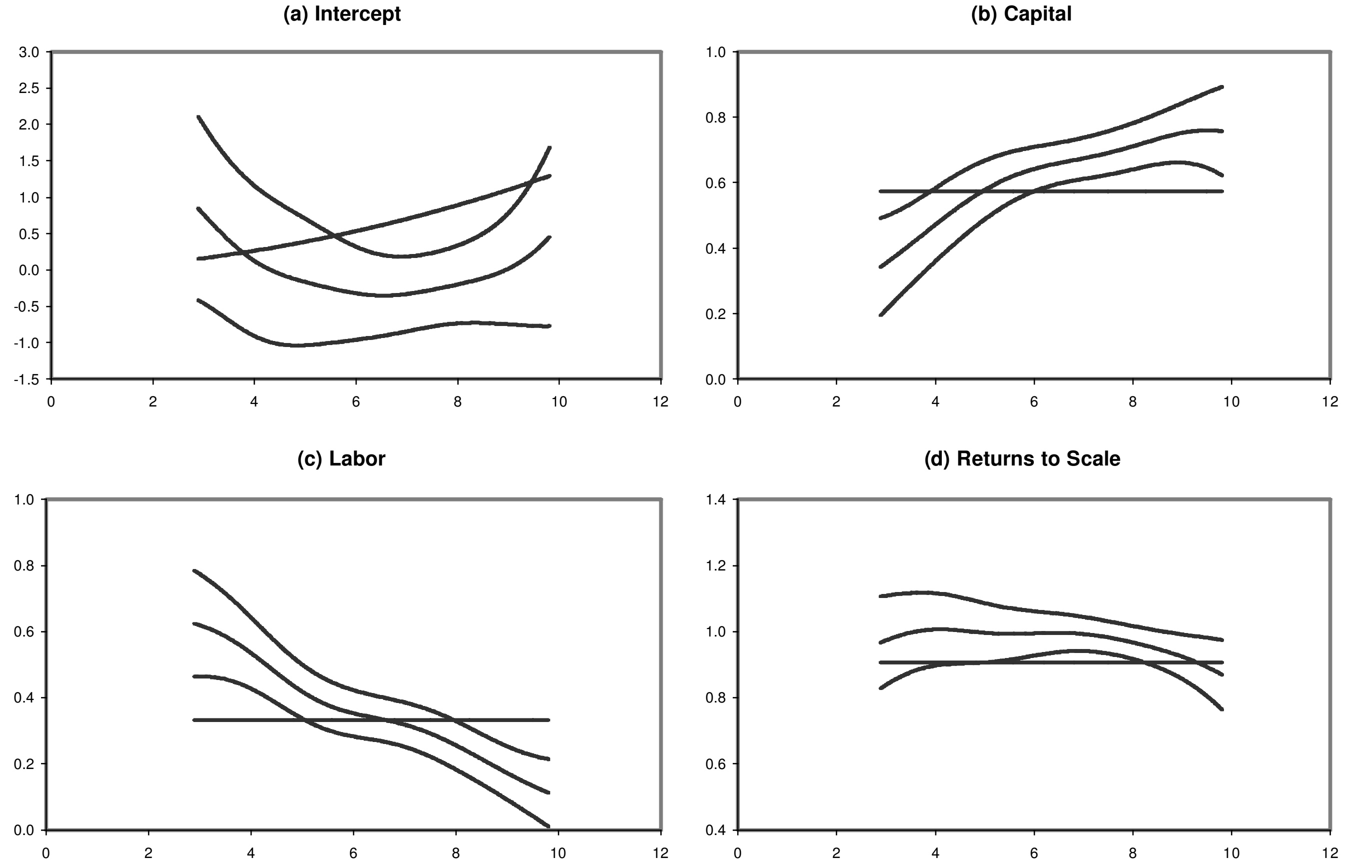

Figure 1. Cobb–Douglas Versus Semiparametric Production (90% bounds).

in the intercept term. Once we allow the coefficients on capital and labor to change with z, the "direct" (intercept) effect of the intermediate production and management expense on output is much smaller. If we look at the coefficient on capital and labor in the semiparametric model, we can see that the coefficient on capital is increasing in z [Fig. 1(b)], and the coefficient on labor is decreasing in z [Fig. 1(c)]. The coefficients from the benchmark parametric model are outside the 90% confidence bounds of the semiparametric model for a large proportion of z. Because our semiparametric model nests the parametric model, this may be viewed as an indication of misspecification of the parametric model. The fact that $\beta_L(z)$ is a decreasing function of z may look puzzling, because one may expect a higher value of intermediate expense to lead to increased labor elasticity of output. However, almost all the firms in our sample are state owned or quasi state owned. It is widely accepted that one common problem with these firms is that there is more labor in these firms than the efficient level. This is the so-called "concealed unemployment." The government did not allow these firms to lay off extra employees to avoid social unrest. There was a scarcity of capital and a surplus of labor in these firms. See Jefferson, Singh, Xing, and Zhang (1999) for a detailed discussion of this issue. The decreasing $\beta_L(z)$ partly reflects this fact about the particular dataset used in this article. Here we would like to mention that, in another study, we applied our smooth coefficient model to estimate the production function of Taiwan's electronic industry, where we found both of the coefficients $\beta_K(z)$ and $\beta_L(z)$ are increasing functions of z, where z is the firm's R&D expenditure. Taiwan's electronic industry is composed of private firms. R&D tends to increase both the capital elasticity and the labor elasticity.

The returns to scale (the summation of the coefficients on capital and labor) are plotted in Figure 1(d). Unlike other coefficients, the returns to scale are relatively flat in the semiparametric model and the returns to scale in the Cobb–Douglas model are within the 90% bounds for most z. The average return to scale, $n^{-1}\sum_{i=1}^{n}[\hat{\beta}_L(z_i)+\hat{\beta}_K(z_i)]$, is equal to .981 for the semiparametric model, which is larger than $\hat{\beta}_L+\hat{\beta}_K=.906$, the return to scale obtained from the parametric model estimation. The returns to scale are close to constant returns to scale in the semiparametric model, whereas returns are somewhat decreasing in the parametric model. The Cobb–Douglas model restricts the labor and capital elasticities of output to be constant, which leads to an increasing effect of z on the intercept term. The intermediate expense shifts the production function without any effect on the labor and capital elasticities of output. However, once we relax the restriction of constant coefficients, we can clearly see two very different effects, which would not be discovered under the Cobb–Douglas model. The semiparametric model reveals some important features of the production process in the dataset. The intermediate production and management expense tends to increase the capital elasticity of output and decrease the labor elasticity of output. Without more detailed accounting data, we cannot give a clear explanation of this

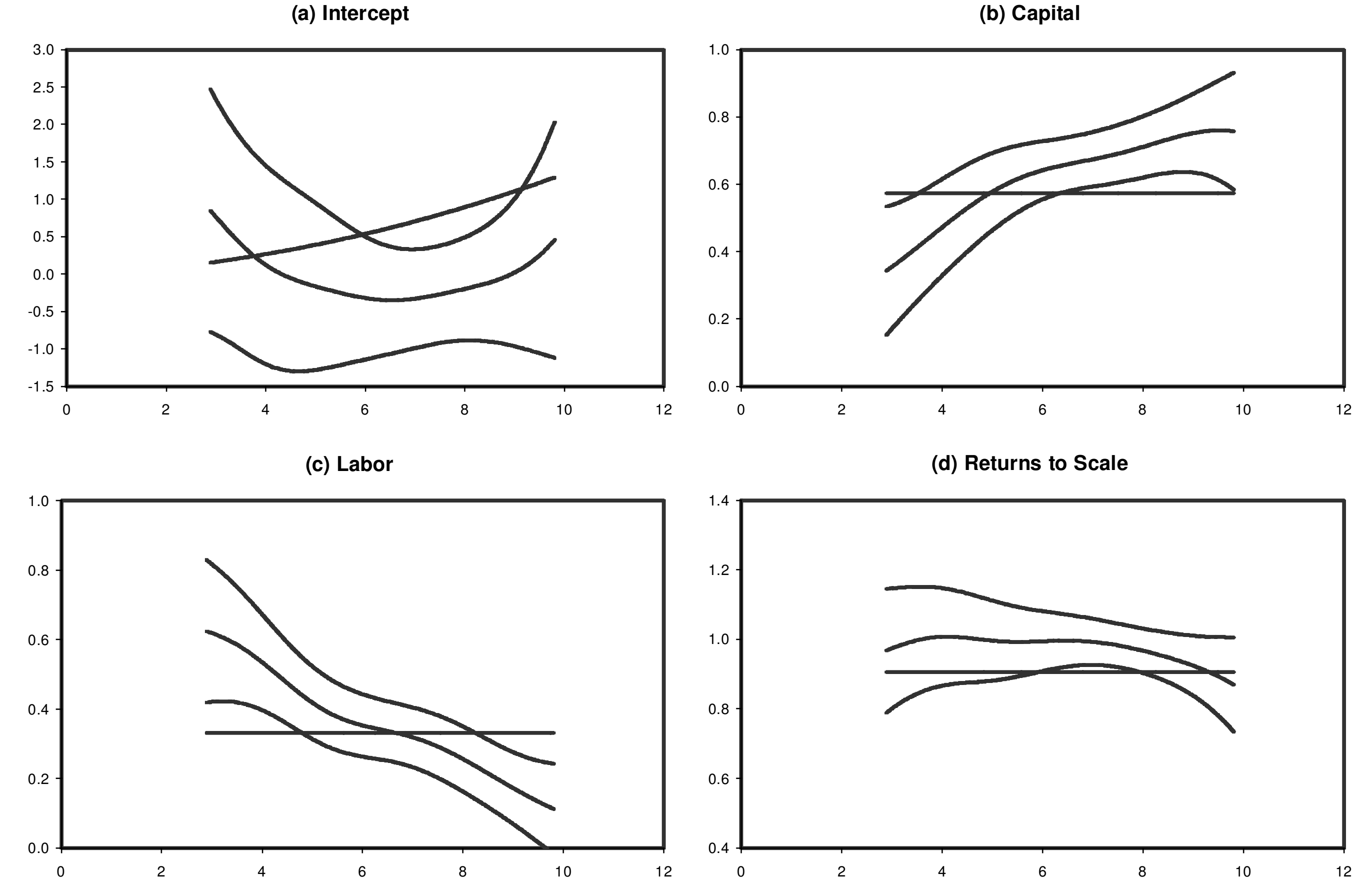

Figure 2. Cobb–Douglas Versus Semiparametric Production (95% bounds).

phenomenon. We speculate that most of the intermediate expense was spent on R&D and upgrading the existing equipment, which subsequently improves the capital elasticity of output.

From Figures 1 and 2, we observe some significant differences between the semiparametric and the parametric (Cobb–Douglas) estimation results. For testing the constant coefficient hypothesis of model (14), we use a standard normal kernel function and the smoothing parameter is chosen as $h = z_{\text{sd}} n^{-1/5}$. The test statistic $J_n = nh^{1/2}\widehat{I_n}/\hat{\sigma}_0$ is given in Theorem 3.1. We have done some simulations to examine the finite-sample performance of the J_n test, and find that it is significantly undersized for $n = 1{,}000$ if we use the asymptotic normal critical values. However, the wild bootstrap method gives very accurate estimated sizes. This is consistent with Li and Wang (1998), who found that their nonparametric test was significantly undersized and that the wild bootstrap test gave much better estimated sizes. Therefore, we will use the bootstrap critical value in testing the null hypothesis of a correctly specified Cobb–Douglas model. The computed J_n statistic yields a value of .464, with a p value of .032 (the 1,000 bootstrap statistics have mean $-.829$ and standard deviation .592). Therefore, we reject the Cobb–Douglas model at the 5% level based on the bootstrap critical value. This outcome confirms the observations from Figures 1 and 2 that the production frontier of the nonmetal mineral industry in China is of the variable coefficient type. The impact of the intermediate expense on output is nonneutral and is input specific.

The comparisons between the semiparametric estimation results and those from the simplified translog model (18) are given in Figures 3 and 4 with the 90% and 95% bounds for the semiparametric estimates. From Figure 3, we see that, compared with the semiparametric model (13), the parametric model (18) overestimates the intercept and underestimates the returns to scale. For the capital and labor coefficient estimates [$\beta_K(z)$ and $\beta_L(K)$], the parametric estimates lie inside the 90% bounds for about 2/3 of the range of z. Compared with the Cobb–Douglas model, the simplified translog model gives much closer estimation results than those of the semiparametric model. Figure 4 reports the same information as Figure 3 except that the 90% bounds are replaced by the 95% bounds. As the bounds become wider, we observe that the parametric estimates of $\beta_K(z)$ and $\beta_L(z)$ now lie inside the 95% bounds for most values of z. For the intercept and return to scale estimates, there are still some significant portions of z in which the parametric estimation differs from the semiparametric results. We also apply the J_n test for the correctness of model (18), and obtain a value of $J_n = .439$, with a p value of .043 (the 1,000 bootstrap statistics have mean $-.830$ and standard deviation .588). Therefore, we reject the simplified translog model (18) at the 5% level based on the bootstrap critical value.

Summarizing the preceding empirical results, we observe that the Cobb–Douglas model, as well as a partially linear model, does not provide an adequate description of the relationship between output and explanatory variables. This is due

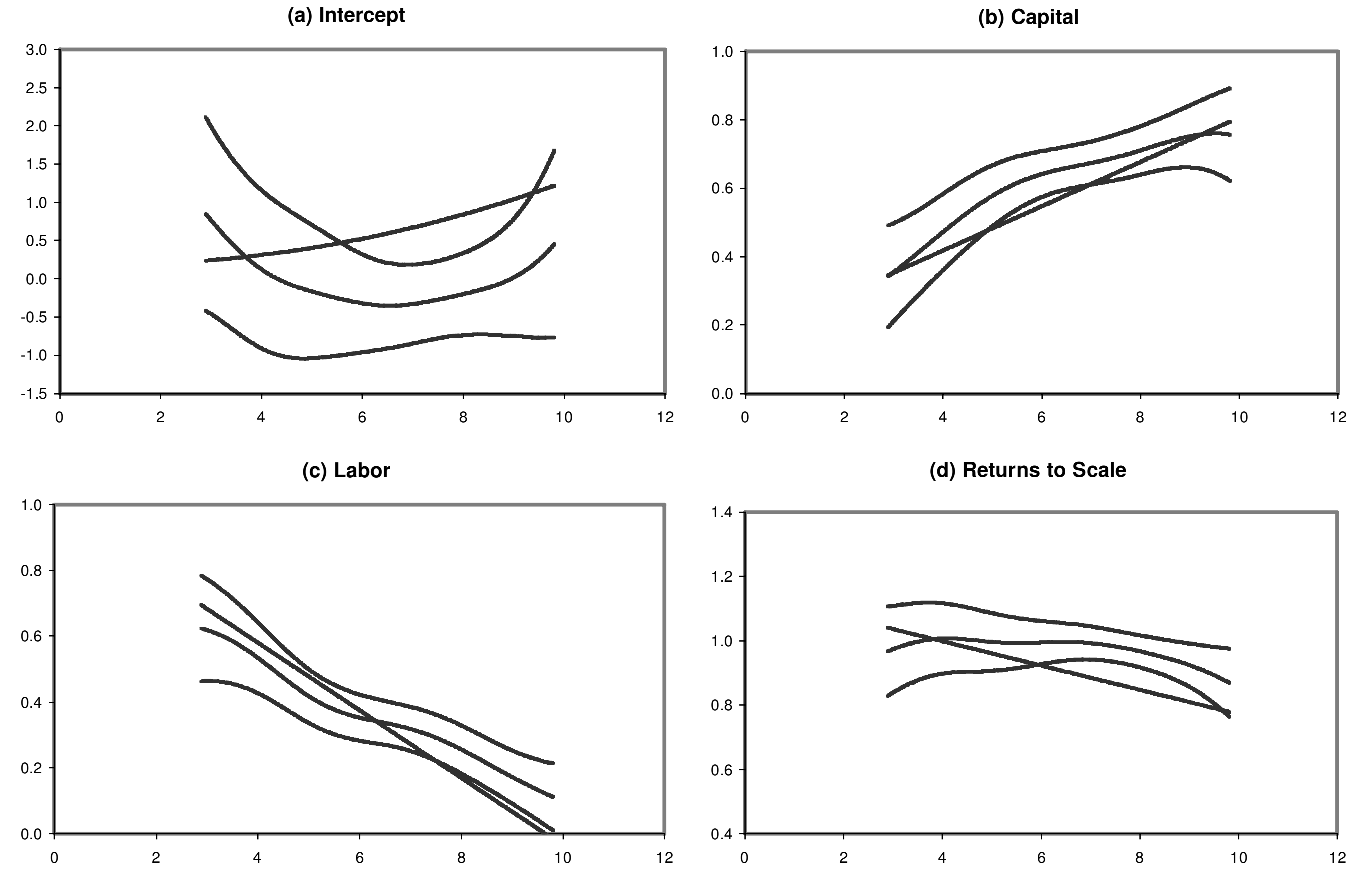

Figure 3. Translog Versus Semiparametric Production (90% bounds).

to the restriction that both models only allow the intermediate production and management expense variable to affect output neutrally. Our semiparametric model and a simplified translog model give much better estimation results. A model specification test and the goodness-of-fit R^2 suggest that our semiparametric model dominates the simplified translog model (18).

5. CONCLUSION

In this article, we propose that a semiparametric smooth coefficient model is a useful yet flexible specification for studying a general regression relationship with varying coefficients. Because the smooth coefficient functions are unspecified, the semiparametric model provides an alternative to a parametric model. Furthermore, the sample size required to obtain reliable semiparametric estimation is not as large as that required for estimating a purely nonparametric model. We suggest a local least squares method with a kernel weight function to estimate the smooth coefficient functions. The consistency of the estimator and its asymptotic normality are established. A simple test statistic for testing a parametric model versus the semiparametric smooth coefficient model is proposed based on the principle of integrated squared differences between the semiparametric and the parametric estimations of the $\delta(z)$ $[\delta_0(z)]$ coefficients. The proposed semiparametric estimator and the test statistic are simple to compute and should be useful to the applied researcher.

The results of an empirical application of a semiparametric smooth coefficient model to the nonmetal mineral industry in China are encouraging. The empirical findings show that the intermediate production and management expenditures have played a vital role and are unbalanced determinants of the labor and capital elasticities of output in production.

Although we only consider independent data in this article, all the results (Theorems 2.1 and 3.1) are still valid for time series data, provided x_t and z_t are nonnested (they can both contain lagged values of y_t) and the data are β-mixing with certain decay rates. For the estimation part (Theorem 2.1), the required conditions will be very similar to that of Robinson (1989), whereas for a consistent test (Theorem 3.1) with time series data, the regularity conditions are given in Fan and Li (1999).

ACKNOWLEDGMENTS

We thank three referees, an associate editor, and Jeff Wooldridge for their insightful comments that greatly improved the article. We also thank James Griffin for very helpful comments on an earlier version of the article. Huang's research is supported by the University Research Council of Vanderbilt University and the National Science Council, Republic of China. Q. Li's research is supported by the Social Sciences and Humanities Research Council of Canada, the Natural Sciences Engineering Research Council of Canada, Ontario Premier's Research Excellence Awards,

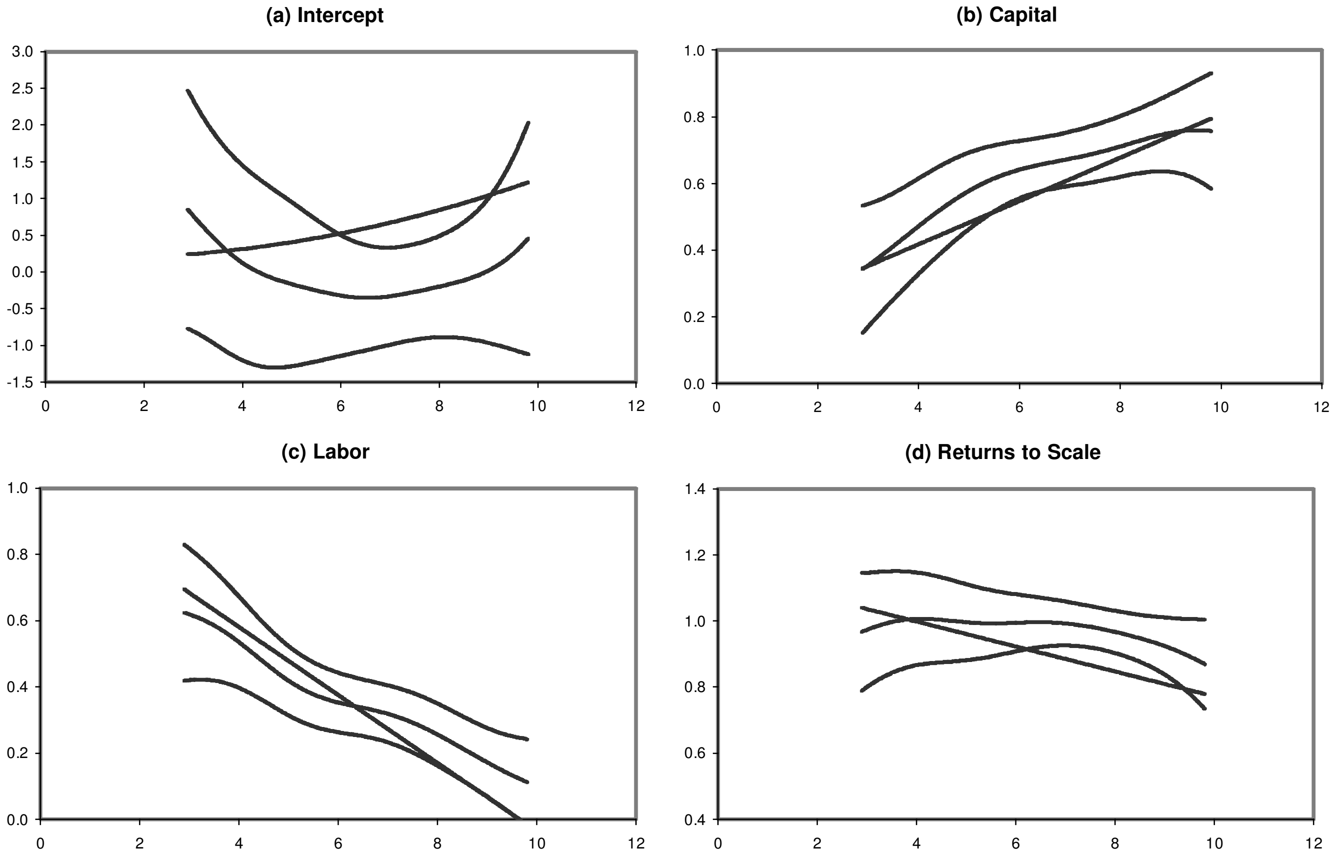

Figure 4. Translog Versus Semiparametric Production (95% bounds).

Bush Program in Economics and Public Policy, and Private Enterprise Research Center, Texas A&M University.

APPENDIX: PROOFS OF THEOREMS 2.1 AND 3.1

In this appendix, we provide proofs of Theorems 2.1 and 3.1. We shall use $\|\cdot\|$ to denote the Euclidean norm. First, we present some regularity conditions.

(A.1) (i) (y_i, x_i, z_i) is independent and identically distributed as $(\mathcal{Y}, \mathcal{X}, \mathcal{Z})$. $E[\mathcal{Y}|\mathcal{X} = x, \mathcal{Z} = z] = \alpha(z) + x'\beta(z)$ almost everywhere, and $\varepsilon = \mathcal{Y} - \mathcal{X}'\delta(\mathcal{Z})$ has finite fourth moments. (ii) Let $f(x,z)$ denote the joint density function of $(\mathcal{X}, \mathcal{Z})$ and let $f_z(z)$ be the marginal density function of $\mathcal{Z}$. $\alpha(z)$, $\beta(z)$, $f_z(z)$, $f(x,z)$, and $\sigma_\epsilon^2(x,z) = E[\varepsilon^2|\mathcal{X} = x, \mathcal{Z} = z]$ all satisfy some Lipschitz-type conditions (in z). $|m(u+v) - m(u)| \le D(u)\|v\|$, where $D(\cdot)$ has finite fourth moments; $m(\cdot)$ is $\alpha(\cdot), \beta(\cdot), f_z(\cdot), f(x,\cdot)$, or $\sigma_\epsilon^2(x,\cdot)$. (iii) $f(x,z)$ and $f_z(z)$ are bounded, and $\alpha(\mathcal{Z})$, $\beta(\mathcal{Z})$, and $(\mathcal{X}, \mathcal{Z})$ all have finite fourth moments.

(A.2) (i) $K(\cdot)$ is a bounded symmetric function with $\int K(u)\,du = 1$, and $\int K(u)uu'\,du = cI_q$, where $c > 0$ is a positive constant and I_q is the identity matrix of dimension q. (ii) As $n \to \infty$, $h \to 0$ and $nh^q \to \infty$.

Note that (A.1) (i) implies that $E(\epsilon_i|x_i, z_i) = 0$, and in (A.1) (ii) we allow conditional heteroscedastic error of unknown form. (A.2) (i) says that $K(\cdot)$ is a standard second-order kernel function. (A.2) (ii) are the usual conditions that ensure the bias and variance terms in the nonparametric estimation go to 0 as the sample size becomes large.

Throughout this appendix, we use $\sum_i$ to denote $\sum_{i=1}^n$ and $\sum_i \sum_{j\neq i}$ to denote $\sum_{i=1}^n \sum_{j=1, j\neq i}^n$.

Proof of Theorem 2.1(a).

$$
\begin{aligned}
\hat{\delta}(z) &= \left[\sum_j X_j X_j' K_{jz}\right]^{-1} \sum_j X_j y_j K_{jz} \\
&= \left[\sum_j X_j X_j' K_{jz}\right]^{-1} \sum_j X_j \big(X_j'\delta(z_j) + \epsilon_j\big) K_{jz} \\
&= \left[\sum_j X_j X_j' K_{jz}\right]^{-1} \\
&\quad \times \sum_j X_j \left[X_j'\delta(z) + X_j'(\delta(z_j) - \delta(z)) + \epsilon_j\right] K_{jz} \\
&= \delta(z) + \left[\sum_j X_j X_j' K_{jz}\right]^{-1} \\
&\quad \times \sum_j X_j \left[X_j'(\delta(z_j) - \delta(z)) + \epsilon_j\right] K_{jz} \\
&= \delta(z) + [D_n(z)]^{-1}\{A_{1n}(z) + A_{2n}(z)\},
\end{aligned}
$$

where

$$D_n(z) = (nh^q)^{-1} \sum_j X_j X_j' K_{jz},$$

$$A_{1n}(z) = (nh^q)^{-1} \sum_j X_j X_j' \big(\delta(z_j) - \delta(z)\big) K_{jz},$$

$$A_{2n}(z) = (nh^q)^{-1} \sum_j X_j \epsilon_j K_{jz}.$$

Obviously, $\hat{\delta}(z) - \delta(z) \xrightarrow{p} 0$ will be proved if we can show the following:

1. $D_n(z) = (nh^q)^{-1} \sum_j X_j X_j' K_{jz} \xrightarrow{p} M_z$ (M_z is positive definite).
2. $A_{1n}(z) = o_p(1)$.
3. $A_{2n}(z) = o_p(1)$.

These are proven next.

Case 1:

$$\begin{aligned}
E(D_n(z)) &= h^{-q} E[X_1 X_1' K_{1z}] \\
&= h^{-q} \int\int X_1 X_1' K\big((z_1 - z)/h\big) f(x_1, z_1)\, dx_1\, dz_1 \\
&= \int\int X_1 X_1' K(v) f(x_1, z + hv)\, dv\, dx_1 \\
&= \Big[\int X_1 X_1 f(x_1, z)\, dx_1\Big]\Big[\int K(v)\, dv + O(h)\Big] \\
&= f_z(z)\Big[\int X_1 X_1 f(x_1 | z_1 = z)\, dx_1\Big][1 + O(h)] \\
&= f_z(z) E[X_1 X_1' | z_1 = z] + o(1) = M_z + o(1).
\end{aligned}$$

Similarly, one can easily show that $\mathrm{var}(D_n(z)) = O((nh^q)^{-1}) = o(1)$. Therefore, $D_n(z) \xrightarrow{p} M_z$.

Case 2:

$$\begin{aligned}
&E[||A_{1n}(z)||^2] \\
&= E\big\{[A_{1n}(z)]' A_{1n}(z)\big\} \\
&= (n^2 h^{2q})^{-1} \sum_i \sum_j E[(\delta(z_i) - \delta(z))' K_{iz} X_i X_i' X_j X_j' \\
&\qquad \times (\delta(z_j) - \delta(z)) K_{jz}] \\
&= (n^2 h^{2q})^{-1} \sum_i E[(\delta(z_i) - \delta(z))' K_{iz} X_i X_i' X_i X_i' \\
&\qquad \times (\delta(z_i) - \delta(z)) K_{iz}] \\
&\quad + (n^2 h^{2q})^{-1} \sum_i \sum_{j \neq i} E[(\delta(z_i) - \delta(z))' K_{iz} X_i X_i' \\
&\qquad \times X_j X_j' (\delta(z_j) - \delta(z)) K_{jz}] \\
&= (nh^{2q})^{-1} E[(\delta(z_1) - \delta(z))' K_{1z} X_1 X_1' X_1 X_1' \\
&\qquad \times (\delta(z_1) - \delta(z)) K_{1z}] \\
&\quad + (h^{2q})^{-1} E[(\delta(z_1) - \delta(z))' K_{1z} X_1 X_1' \\
&\qquad \times X_2 X_2' (\delta(z_2) - \delta(z)) K_{2z}] \\
&= O(h^2 (nh^q)^{-1}) + O(h^4).
\end{aligned}$$

Therefore, $A_{1n}(z) = O_p(h(nh^q)^{-1/2} + h^2) = o_p(1)$.

Case 3:

$$\begin{aligned}
E\big[||A_{2n}(z)||^2\big] &= (nh^q)^{-2} \sum_i E\big[X_i' X_i \epsilon_i^2 K_{iz}^2\big] \\
&= (nh^{2q})^{-1} E\big[X_1' X_1 \sigma_\epsilon^2(x_1, z_1) K_{1z}^2\big] \\
&= (nh^{2q})^{-1} O(h^q) = O((nh^q)^{-1}) = o(1).
\end{aligned}$$

Therefore, $A_{2n}(z) = O_p((nh^q)^{-1/2}) = o_p(1)$.

Proof of Theorem 2.1(b). By the proof of cases 1 and 3, we know that $D_n(z) = M_z + o_p(1)$ and $A_{1n}(z) = O_p(h^2 + h(nh)^{-1/2})$. Also note that $nh^{q+4} = o(1)$. We have

$$\begin{aligned}
&\sqrt{nh^q}\big(\hat{\delta}(z) - \delta(z)\big) \\
&\quad = \big[D_n(z)\big]^{-1} \sqrt{nh^q}\big\{A_{1n}(z) + A_{2n}(z)\big\} \\
&\quad = \big[D_n(z)\big]^{-1} \sqrt{nh^q}\big\{O\big(h^2 + h(nh^q)^{-1/2}\big) + A_{2n}(z)\big\} \\
&\quad = \big[D_n(z)\big]^{-1} \big\{\sqrt{nh^q} A_{2n}(z)\big\} + o_p(1).
\end{aligned}$$

The term $\sqrt{nh^q} A_{2n}(z)$ has mean 0 and its variance is

$$\begin{aligned}
&nh^q \Big\{(nh^q)^{-2} \sum_i E\big[X_i X_i' \epsilon_i^2 K_{iz}^2\big]\Big\} \\
&\quad = h^{-q} \int f(x_1, z_1) \sigma_\epsilon^2(x_1, z_1) X_1 X_1' K_{1z}^2\, dx_1\, dz_1 \\
&\quad = \Big[\int X_1 X_1' \sigma_\epsilon^2(x_1, z) f(x_1, z)\, dx_1\Big]\Big[\int K^2(v)\, dv\Big] + O(h) \\
&\quad = f_z(z) E\big[X_1 X_1' \sigma_\epsilon^2(x_1, z_1) | z_1 = z\big]\Big[\int K^2(v)\, dv\Big] + o(1) \\
&\quad = V_z + o(1).
\end{aligned}$$

It is straightforward to check that the conditions for a triangular-array central limit theorem hold (e.g., Serfling 1980, p. 32). Thus, $\sqrt{nh^q} A_{2n}(z) \xrightarrow{p} N(0, V_z)$.

Therefore, $\sqrt{nh^q}(\hat{\delta}(z) - \delta(z)) \xrightarrow{p} M_z^{-1} N(0, V_z) = N(0, \Omega_z)$.

Proof of Theorem 3.1. The proof of Theorem 3.1 is similar to the proof of proposition 1 of Li and Wang (1998) and theorem 3.1 of Zheng (1996). We provide only a sketch proof for Theorem 3.1.

Proof of Theorem 3.1(a). Using the identity that, under H_0, $\hat{\epsilon}_i = y_i - X_i' \hat{\delta}(z_i) = \epsilon_i + X_i'(\delta_0(z_i) - \hat{\delta}_0(z_i))$ and the shorthand notation $K_{ij} = K((z_i - z_j)/h)$, we have

$$\begin{aligned}
\widehat{I}_n &= \frac{1}{n^2 h^{2q}} \sum_i \sum_{j \neq i} \hat{\epsilon}_i \hat{\epsilon}_j X_i' X_j K_{ij} \\
&= \frac{1}{n^2 h^q} \sum_i \sum_{j \neq i} \epsilon_i \epsilon_j X_i' X_j K_{ij} \\
&\quad + \frac{2}{n^2 h^{2q}} \sum_i \sum_j X_i' X_j \epsilon_i X_j' \big(\delta_0(z_j) - \hat{\delta}_0(z_j)\big) K_{ij} \\
&\quad + \frac{1}{n^2 h^{2q}} \sum_i \sum_{j \neq i} X_i' X_j X_i' \big(\delta_0(z_i) - \hat{\delta}_0(z_i)\big) \\
&\qquad \times X_j' (\delta_0(z_j) - \hat{\delta}_0(z_j)) X_j K_{ij} \\
&= \widehat{I}_{1n} + 2\widehat{I}_{2n} + \widehat{I}_{3n}.
\end{aligned}$$

The term $nh^{q/2}\widehat{I}_{1n} = nh^{q/2}\{(n^2h^q)^{-1}\sum_i\sum_{j\neq i}\epsilon_i\epsilon_j X_i'X_jK_{ij}\} \xrightarrow{d} \mathrm{N}(0,\sigma_0^2)$ follows a similar proof as in the proof of lemma 1 of Li and Wang (1998) or Theorem 3.1 of Zheng (1996). We will not repeat their proof here. Instead, we will only provide some intuitive arguments.

$\widehat{I}_{1n}$ can be written as a U statistic $\widehat{I}_{1n} = 2(n^2h^q)^{-1}\sum_{i=1}^n\sum_{j>i}^n \times H_n(w_i,w_j)$, where $w_i = (\epsilon_i, x_i, z_i)$ and $H_n(w_i,w_j) = \epsilon_i\epsilon_j X_i'X_jK_{ij}$. Note that $E[H_n(w_i,w_j)|w_i] = 0$. Therefore, $\widehat{I}_{1n}$ is a second-order degenerate U statistic. Obviously, $E(nh^{q/2}\widehat{I}_{1n}) = 0$ and it is easy to show that $\mathrm{var}(nh^{q/2}\widehat{I}_{1n}) = \sigma_0^2 + o(1)$. Therefore, one would expect $nh^{q/2}\widehat{I}_{1n} \xrightarrow{d} \mathrm{N}(0,\sigma_0^2)$. This is indeed true by Hall's (1984) central limit theorem for degenerate U statistics.

Using the facts that $(\hat{\alpha}_0 - \alpha_0) = O_p(n^{-1/2})$, $(\hat{\gamma}_0 - \gamma_0) = O_p(n^{-1/2})$, and $(\hat{\beta}_0 - \beta_0) = O_p(n^{-1})$ and using similar arguments as in Li and Wang (1998) and Zheng (1996), one can easily show that $\widehat{I}_{2n} = O_p(n^{-1})$, $\widehat{I}_{3n} = O_p(n^{-1})$, and $\hat{\sigma}_0^2 = \sigma_0^2 + o_p(1)$. Therefore, $nh^{q/2}\widehat{I}_n/\hat{\sigma}_0 = nh^{q/2}\widehat{I}_{1n}/\sigma_0 + o_p(1) \xrightarrow{d} \mathrm{N}(0, 1)$ under H_0.

Proof of Theorem 3.1(b). Again the detailed proof is very similar to that of Li and Wang (1998) and Zheng (1996). We give only the main steps here. Parallel to the proof of Theorem 3.1, it is easy to show that $\widehat{I}_n \xrightarrow{p} I \stackrel{\text{def}}{=} \int\{M_z[\delta(z) - \delta_0(z)]\}'\{M_z[\delta(z)-\delta_0(z)]\}\,dz (>0)$ under H_1. It is straightforward to show that $\hat{\sigma}_0 = C + o_p(1)$ under H_1, where C is a positive constant. Therefore, $J_n = nh^{q/2}\widehat{I}_n/\hat{\sigma}_0 = nh^{q/2}I/C + o_p(1)$, which leads to Theorem 3.1(b).

[Received July 1999. Revised February 2001.]

REFERENCES

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

Bierens, H. (1982), "Consistent Model Specification Tests," *Journal of Econometrics*, 10, 105–134.

Bierens, H., and Ploberger, W. (1997), "Asymptotic Theory of Integrated Conditional Moment Tests," *Econometrica*, 65, 1129–1151.

Chen, R., and Tsay, R. S. (1993), "Functional-Coefficient Autoregressive Models," *Journal of the American Statistical Association*, 88, 298–308.

Cornwell, C., Schmidt, P., and Sickles, R. C. (1990), "Production Frontiers With Cross-Sectional and Time-Series Variation in Efficiency Levels," *Journal of Econometrics*, 46, 185–200.

Fan, Y. (1994), "Testing the Goodness of Fit of a Parametric Density Function by Kernel Method," *Econometric Theory*, 10, 316–356.

Fan, Y., and Li, Q. (1996), "Consistent Model Specification Tests: Omitted Variables and Semiparametric Functional Forms," *Econometrica*, 65, 865–890.

——— (1999), "Central Limit Theorem for Degenerate U-Statistic of Absolutely Regular Process: With an Application to Model Specification Testing," *Journal of Nonparametric Statistics*, 10, 245–271.

Fan, Y., Li, Q., and Weersink, A. (1996), "Semiparametric Estimation of Stochastic Production Frontier Models," *Journal of Business & Economic Statistics*, 14, 460–468.

Hall, P. (1984), "Central Limit Theorem for Integrated Square Error of Multivariate Nonparametric Density Estimators," *Journal of Multivariate Analysis*, 14, 1–16.

Hastie, T., and Tibshirani, R. (1993), "Varying Coefficient Models," *Journal of the Royal Statistical Society*, Ser. B, 55, 757–796.

Jefferson, G. H., Singh, I., Xing, J., and Zhang, S. (1999), "China's Industrial Performance: A Review of Recent Findings," in *Enterprise Reform in China: Ownership, Transition, and Performance*, eds. G. W. Jefferson and I. Singh, Oxford, U.K.: Oxford University Press, pp. 127–152.

Kumbhakar, S. C. (1990), "Production Frontier, Panel Data, and Time-Varying Technical Inefficiency," *Journal of Econometrics*, 46, 201–211.

Li, Q. (1996), "Nonparametric Testing of Closeness Between Two Unknown Distributions," *Econometric Review*, 15, 261–274.

Li, Q., and Wang, S. (1998), "A Simple Consistent Bootstrap Test for a Parametric Regression Functional Form," *Journal of Econometrics*, 87, 145–165.

Murakami, N., Liu, D., and Otsuka, K. (1994), "Technical and Allocative Efficiency Among Socialist Enterprises: The Case of Garment Industry in China," *Journal of Comparative Economics*, 19 (3), 410–433.

Park, B., Sickles, R. C., and Simar, L. (1998), "Stochastic Panel Frontiers: A Semiparametric Approach," *Journal of Econometrics*, 84, 273–301.

Robinson, P. M. (1988), "Root-N-Consistent Semiparametric Regression," *Econometrica*, 56, 931–954.

——— (1989), "Nonparametric Estimation of Time-Varying Parameters," in *Statistical Analysis and Forecasting of Economic Structural Changes*, ed. P. Hackl, New York: Springer-Verlag.

Serfling, R. J. (1980), *Asymptotic Theorems of Mathematical Statistics*, New York: Wiley.

Silverman, W. (1986), *Density Estimation for Statistics and Data Analysis*, New York: Chapman & Hall.

Stock, J. H. (1989), "Nonparametric Policy Analysis," *Journal of the American Statistical Association*, 84, 567–575.

Wooldridge, J. (1992), "A Test for Functional Form Against Nonparametric Alternatives," *Econometric Theory*, 8, 452–475.

Zheng, J. X. (1996), "A Consistent Test of Functional Form via Nonparametric Estimation Technique," *Journal of Econometrics*, 75, 263–289.

This article has been cited by:

1. Qi Li , Jeffrey S. Racine , Jeffrey M. Wooldridge . 2009. Efficient Estimation of Average Treatment Effects with Mixed Categorical and Continuous DataEfficient Estimation of Average Treatment Effects with Mixed Categorical and Continuous Data. *Journal of Business and Economic Statistics* **27**:2, 206-223. [Abstract] [PDF] [PDF Plus]