行政院國家科學委員會專題研究計畫　成果報告

核密度函數估計量之改良式交叉有效法

計畫類別：個別型計畫
計畫編號：NSC93-2118-M-032-013-
執行期間：93 年 08 月 01 日至 94 年 07 月 31 日
執行單位：淡江大學統計學系

計畫主持人：鄧文舜

報告類型：精簡報告

處理方式：本計畫可公開查詢

中　華　民　國 94 年 10 月 24 日

A Feasible Method for Fitting Partially Linear Panel Data Models with Fixed Effect

Deng Wen-Shuenn *
Department of Statistics
Tamkang University

Huang Tai-Hsin
Department of Money and Banking
National Chengchi University

October 21, 2005

Abstract

There are more data sets being available in panel form, making the estimation techniques for the regression model with panel data very popular in applied econometrics. In this presentation, we consider a partially linear model with fixed-effect, under the framework of panel data. Based on the ideas of Stock (1989) and Robinson (1988), we proposed a double residual technique to estimate the fixed-effect and slope coefficients. A simulation study is carried out to assess the finite-sample performances of the proposed estimators. The results reveal that the estimators can have mild and reasonable percentage biases. In addition, they tend to be efficient relative to those for correctly specified parametric models, in terms of their simulated standard error ratios. An empirical study employing 96 Taiwan's electronic firms' data is also conducted. Evidence is found that the partially linear models considered here are preferable to the parametric ones, due to possible misspecification of the latter.

中文摘要

在實際應用上, 觀測常資料常是以追蹤資料 (panel data) 型式出現的, 因此計量經濟學上追縱資料之迴歸模式的估計技巧, 是一個重要且應用性高的課題。在本文中, 我們考慮一個追縱資料下, 具固定效果 (fixed effect) 的部分線性 (partially linear) 迴歸模型。根據 Stock(1989) 及 Robinson(1988) 的雙重殘差 (double residual) 概念, 我們提出一種估計固定效果及迴歸係數的方法。模擬研究顯示, 此估計方法所產生的估計量可以具有合理的百分比偏量 (percentage bias)。此外, 當觀察期數增加或觀察之合併資料之數目 (number of panel) 增加時, 本文所提出的估計量之估計效率, 會趨近於設定正確的參數回歸模式下, 最小平方法估計量的估計效率。亦即, 兩者的模擬標準誤 (simulated standard error) 之比值趨近於一。此外本文針對一組實際追縱資料資料, 建立具固定效果 (fixed effect) 的部分線性 (partially linear) 迴歸模型, 並利用拔靴法 (boorastrpping) 來針對固定效果及迴歸係數進行假設檢定。由於部分線性迴歸模型, 可以減低模型設定錯誤的可能性, 因此本文的所考慮的模型及估計方法, 較之於參數方法更能適用於追縱資料下具固定效果之迴歸模型建立。

JEL Classification: C14 C15 C23

Keywords: partially linear model, kernel estimators, fixed effect parameters

*This work was supported in part by the National Science Council grant NSC-93-2118-M-032-013.

1 Introduction

In the past few decades, there are two main approaches to the studies of efficiency and productivity of producers for a particular industry. One of them is the mathematical programming or data envelopment analysis (DEA). The other is referred to as the econometric approach. The DEA approach, dating back to Farrell (1957), is known as function-free such that the potential specification error on the functional form avoids. However, this advantage is not without expense. Due to its incapability of disentangling the impact of random disturbances on efficiency, the estimated efficiency scores may be contaminated with noise and thus unable to reflect the true performance of a representative firm. On the contrary, the statistical noise is explicitly incorporated in a regression model as a two-part error component. Thus, the estimated production efficiency by the econometric approach does not suffer from the same weakness as the DEA does. But, this approach is frequently criticized as requiring the imposition of a specific presumption on the functional form of a production technology and/or an optimal cost frontier, as well as on the distributions of the composed error term.

As noted by Schmidt and Sickles (1984) and Berger (1993), the availablility of a panel data set enables ones to relax the imposition of strong distributional assumptions on the error components. However, the likelihood of functional form misspecification remains to be solved in the econometric approach. Recently, Fan et al. (1996) extended the linear stochastic frontier model, first proposed by Aigner et al. (1977) and Meeusen and van den Broeck (1977), to a semiparametric frontier model where the production function takes an unknown form and the composite error term follows the convention[1]. Li et al. (2002) further developed a semiparametric smooth coefficient model for studying a general regression association with varying coefficients, while ignoring the possibility of production inefficiency. The above two works appear to be suitable only in the context of either cross-sectional or time series data.

Given that the purely parametric and purely nonparametric models have their own strengths and weaknesses, the current paper proposes an intermediate strategy employing a semiparametric method – the partial linear model – to estimate either production, or cost, or profit frontiers, as well as to estimate technical efficiency, under the framework of the fixed-effects model. Specifically, let's consider the semiparametric production regression model for the ith firm ($i = 1, \cdots, n$) at period t ($t = 1, \cdots, T$):

$$y_{it} = \beta' z_{it} + f(x_i t) + u_i + v_{it}, \tag{1-1}$$

where y_{it} is the (log) output level of the ith firm at time t, $z_{it} = (z_{1,it}, z_{2,it}, , z_{p,it})$ and $x_{it} = (x_{1,it}, x_{2,it}, , x_{q,it})$ are, respectively, observations on the continuous independent variable of p and q-dimensional (log) inputs z and x. Notations $\beta = (\beta_1, \beta_2, \cdots, \beta_p)$ and f are respectively, the unknown slope coefficient vector and function of x. Notation u_i represents the unknown and time-invariant group (firm) specific inefficiency, and v_{it} denotes the random disturbance with mean zero and constant variance. The u_i's are treated as fixed effects by this exercise. Incidentally, the use of panel data provides rich source of information about, e.g., the behaviors of firms in an industry.

Three attractive features of this article are worth mentioning. First, the adoption of the fixed-effects model allows the inefficiency term to be correlated with the regressors and with the statistical noise, more compliant with the reality. Second, the employment of the partial linear structure (1-1) is one of the three types of restrictions, which considerably reduces approximation error derived by the pure nonparametric case[2]. Robinson (1988) has shown under regularity conditions that the estimators generalizing the ordinary least squares (OLS) estimators $\widehat{\beta}$ are $\sqrt{N}$-consistent for and asymptotically

[1] In fact, they assumed that the one-sided error term is distributed as a half-normal random variable truncated at zero from below and the two-sided statistical noise is an independently and identically distributed normal variable with mean zero and constant variance.

[2] The other two empirically useful restrictions are an additive separable function on $f(x_{it})$, where x_{it} is a vector, and smoothness assumptions. See Yatchew (1998) for details

normal; a consistent estimator of its limiting covariance matrix was also given. Third, our fixed-effects treatment can be regarded as an extension of Stock (1989) from the setting of cross sections to that of panel data and from a nonparametric estimation to a semiparametric counterpart. Thus, the critical element of the "curse of dimensionality", presented in nonparametric methods, is less of an issue. This is because the dimension of x_{it} in a semi-parametric model is lowered. It is in turn that the asymptotic results, obtained by Stock (1989), may be applied in this article.

This paper is organized as follows. Section 2 states the model and proposes the semiparametric estimation procedures under a panel data setting. Section 3 performs the Monte Carlo simulations and presents the results of the experiments. An empirical study, using the data of 96 firms in the industry of the electronics and information in Taiwan over the period 1996-2001, is addressed in Section 4, while Section 5 concludes the paper.

2 The Econometric Model and the Proposed Estimation Method

Based on equation (1-1), our proposed estimation method follows the vein of Robinson(1988) and Stock(1989) with the generalizations to panel data and to a semiparametric model. It is assumed that (z_{it}, x_{it}, y_{it}) are independently and identically distributed (i.i.d.) as the $(R^p \times R^q \times R)$-valued random variable (x, y, z), ϵ_{it} are assumed to be i.i.d. random variable with mean zero and constant variance σ^2, $0 < \sigma^2 < \infty$. The regressors (z_{it}, x_{it}) are assumed to be independent of the error term. Notation f is an unknown function continuous in variable x. The purpose of the regression analysis is to use the observations (z_{it}, x_{it}, y_{it}) to estimate the production function:

$$y = \beta' z + f(x) + u \tag{2-1}$$

Either the time series or the cross-sectional version of population regression function (2-1) has been examined by, for example, Stock (1989), Engle et al. (1986), and Robinson (1988), among others, for various objectives. Our own ends are accurate estimation of β, suitably measuring technical efficiency, and precisely predicting y. The elements β of are the parameters of interest that contain useful information on production technology when a production or a cost function is under consideration. Some statistical testing hypotheses may be expressible solely in terms of β. In the area of productivity and efficiency analysis, how to correctly assess the technical efficiency of a firm is a core issue, concerned not only by academic researchers, but also by policy makers. Finally, although economic theory might relate a valid set of independent variables to a dependent variable, it often fails to offer guidance about the exact functional form to estimate[3]. A misspecified parametric model will lead to inconsistent parameter estimates and subsequent statistical inferences. This paper proposes a procedure to predict the dependent variable when the exact functional form is unknown.

Define $Y = (y_{11}, \cdots, y_{1T}, y_{21}, \cdots, y_{2T}, \cdots, \cdots, y_{n1}, \cdots, y_{nT})'$; and let v, $Z_d (d = 1, \cdots, p)$ and f_x be similarly defined respectively by $v_{it}, z_{d,it} (d = 1, \cdots, p)$ and $f(x_{it})$, for $i = 1, \cdots, n$ and $t = 1, \cdots, T$. Let $N = n \times T$ and $J_c (c = 1, \cdots, n)$ be an $N \times 1$ vector having 1's in the entries from $(c-1) \times T$ to $c \times T$, and 0 otherwise. For example, if $T = 3$, then $J_2 = (0, 0, 0, 1, 1, 1, 0, 0, 0, \cdots, 000)'$. Using these notations, equation (1-1) can be rewritten as

$$\begin{aligned} Y &= \sum_{d=1}^{p} Z_d \beta_d + f_x + \sum_{d=1}^{p} J_c u_c + v \\ &= Z\beta + f_x + JU + v \end{aligned} \tag{2-2}$$

[3] The commonly used functional forms include the Cobb-Douglas, translog, constant elasticity of substitution, the Leontief, and so on. These are all approximation to the true but unknown functions.

where $Z = (Z_1 Z_2 \cdots Z_p)$, $J_{N\times n} = (J_1 J_2 \cdots J_n)$, and $U = (u_1, u_2, \cdots, u_n)'$.

Based on the idea of Robinson(1989), Stock(1989) and (2-2), our estimating strategy is simple. Before addressing the estimation strategy, we need to introduce some notations. Let $(\delta_{11}, \cdots, \delta_{1T}, \delta_{21}, \cdots, \delta_{2T}, \cdots, \delta_{n1}, \cdots, \delta_{nT})$ be any panel observation of a random variable δ, and w the observation of the q-dimensional random input x. Given a kernel function K as a symmetric probability density function supported on R^q, and a bandwidth h, the kernel estimator $E[\delta|w]$ is formulated by

$$E[\delta|w] = \frac{\sum_{i=1}^{n}\sum_{t=1}^{T} K(\frac{x_{it}-w}{h})\delta_{it}}{\sum_{i=1}^{n}\sum_{t=1}^{T} K(\frac{x_{it}-w}{h})} \tag{2-3}$$

Note that local polynomial and other nonparametric smoothers can also serve to formulate $E[\delta|w]$, however, simulation studies in Section 3 demonstrate that $\widehat{u}_i$ and $\widehat{\beta}_d$ based on local linear type estimators suffer larger simulated mean square error. Therefore, a local constant type of estimator (2-3) will be used throughout this work.

We now propose our method of estimating U and β. Note that (2-1) and (2-2) imply that, based on the "double residual" idea proposed by Robinson(1988), the residuals from estimating $E[y|x], E[J|x]$,and $E[z|x]$ can form the following parametric model

$$y - E[y|x] = u - E[u|x] + (z - E[z|x])\beta \tag{2-4}$$

One can use (2-4) and to estimate U and β, simultaneously via ordinary least squares procedure (OLS). This suggests, for each $i = 1, \cdots, n$ and $t = 1, \cdots, T$, producing nonparametric estimates of $\widehat{E}[y|x_{it}]$, $\widehat{E}[z|x_{it}]$, $i = 1, \cdots, n$, and $\widehat{E}[Z_d|x_{it}]$, $d = 1, \cdots, p$, to construct matrices $\widehat{E}[y|x]$, $\widehat{E}[J|x]$, and $\widehat{E}[Z|x]$. Their elements come from the respective kernel estimates. Based on (2-4), one can write

$$y - \widehat{E}[y|x] \cong (J - \widehat{E}[J|x])u + (Z - \widehat{E}[z|x])\beta + v \tag{2-5}$$

Let $\Lambda = (Z\ X)$ and $\widehat{E}[\Lambda|x] = (\widehat{E}[J|x]\ \widehat{E}[Z|x])$. The simple OLS estimator can be expressed as

$$\begin{pmatrix} u_i \\ b_{n+d} \end{pmatrix} = \begin{pmatrix} e_i \\ e_{n+d} \end{pmatrix} \Omega,\ i = 1, \cdots, n \text{ and } d = 1, \cdots, p$$

where $e_k, i = 1, \cdots, n+p$, is the $1 \times (1+p)$ vector having 1 at the kth entry and all other entries 0, and $\Omega = \{(\Lambda - \widehat{E}[\Lambda|x])'(\Lambda - \widehat{E}[\Lambda|x])\}^{-1}(\Lambda - \widehat{E}[\Lambda|x])'(y - \widehat{E}[y|x])$

Variance as σ^2 well as the covariance matrix $\sum_{u_i,b_d}$of all u_i's and β_d's can be estimated respectively by (analogous to Robinson 1988)

$$\widehat{\sigma}^2 = \frac{1}{N}\{(y - \widehat{E}[y|x] - (\Lambda - \widehat{E}[\Lambda|x]))(\begin{matrix}\widehat{U}\\ \widehat{\beta}\end{matrix})\}\{(y - \widehat{E}[y|x] - (\Lambda - \widehat{E}[\Lambda|x]))(\begin{matrix}\widehat{U}\\ \widehat{\beta}\end{matrix})\}'$$

and

$$\widehat{\Sigma}_{u_i,b_d} = \widehat{\sigma}^2\{(\Lambda - \widehat{E}[\Lambda|x])(\Lambda - \widehat{E}[\Lambda|x])'\}^{-1}$$

where $\widehat{U} = (\widehat{u}_1, \widehat{u}_2, \cdots \widehat{u}_n)'$ and $\widehat{\beta} = (\widehat{\beta}_1, \widehat{\beta}_2, \cdots \widehat{\beta}_n)'$.

To estimate $f(x_{it})$, one can simply perform a kernel regression of $y - (J - \widehat{E}[J|x])U + (Z - \widehat{E}[z|x])\beta$ on x_{it}. That is, $f(x_{it})$ is estimated by the kernel estimate $\widehat{f}(x_{it}) = E(\,y - (J - \widehat{E}[J|x])\widehat{U} + (Z - \widehat{E}[z|x])\widehat{\beta} \mid x)$ as suggested by Yatchew (2003). Finally, the predicted values of y_{it} and v_{it} are thus given by $\widehat{y}_{it} = \widehat{u}_i + \widehat{\beta}' z_{it} + f(x_{it})$ and $\widehat{v}_{it} = y_{it} - \widehat{y}_{it}$

It is worth mentioning that in Section 5 a bootstrap procedure will be adopted to compute the bootstrapped p-values, where residuals are used to generate the bootstrapped samples repeatedly.

3 Monte Carlo Experiments

In this section, Monte Carlo simulations are carried out to give some insights into the finite-sample performance of our proposed estimators. Subsection 3.1 assumes that the nonparametric function $f(.)$ depends on a single variable, while Subsection 3.2 extends to two variables.

3.1 Single Variable in $f(.)$

Regression model (1-1) is arbitrarily chosen as , $z_{it} = (z_{1,it}, z_{2,it}, z_{3,it})'$, and $\beta = (15, 10, 5)\prime$. In the model settings, and $f(x_{it}) = 10 + 2\sqrt{x_{it}} + x_{it}/20$ are independently drawn from a uniform distribution $U(0, 25)$ and normal distribution $N(5, 5^2)$, respectively, while and , whose marginal distributions are respectively $N(150, 30^2)$ and $N(50, 10^2)$, are drawn from a bivariate normal distribution with a correlation coefficient 1/10. The regression error are i.i.d. $N(0, \sigma^2)$ variables, where for all i and t. We consider five different (n, T) combinations: $(10, 50), (10, 15), (10, 6), (30, 6)$ and $(100, 6)$. The fixed effect coefficients $u_i, i = 1, \cdots, n$ are specified by a random sample drawn from $U(5, 25)$.

The kernel to be employed throughout this current simulation is $K(r) = 0.75(1 - r^2)1_{[-1,1]}(r)$ with bandwidth $h = cn^{-1/5}\text{sd}(x)$, where $\text{sd}(x)$ is the sample standard deviation of regressor x, and constant c adjusted through the simulation. Note here that, the nonparametric component in the model contains a constant 10, which serves as the intercept when the parametric OLS estimation method is applied. We attempt to estimate coefficients β as well as the normalized fixed effect coefficients $\alpha_i = u_i - u_n$, $i = 1, \cdots, n$. The normalization technique is pivotal if one wants to sidestep multicollinearity. For each of the above five (n, T) combinations, 1000 replications of regression data $(y_{it}, z_{1it}, z_{2it}, z_{3it}, x_{it})$, for $i = 1, \ldots, n$ and $t = 1, \ldots, T$, are generated and thus applied to both the proposed semiparametric and the parametric OLS estimation methods, assuming the true functional form to be known a priori.

Table 1 provides the simulation results for coefficients estimates $\widehat{\beta}_1, \widehat{\beta}_2, \widehat{\beta}_3$, and (columns 5 to 7) and the selected minimal, median, and maximal fixed-effect coefficient estimates $\widehat{\alpha}_1, \widehat{\alpha}_2, \widehat{\alpha}_3$ (columns 2 to 4), yielded by the 1000 replications. Recall that the parameters marked with "ˆ" represent the semiparametric estimates, while the parameters with "˜" are the usual OLS estimates. The overheaded "–" denotes the average of the resulting 1000 coefficient estimates. We report their respective simulated percentage biases as well as the relative efficiencies in columns 2 to 7. The latter are defined as either the ratio of simulated standard error $\text{se}(\widetilde{\alpha})$ to $\text{se}(\widehat{\alpha})$ or $\text{se}(\widetilde{\beta})$ to $\text{se}(\widehat{\beta})$ and are given in the parentheses. Column 8 provides the minimal and maximal (in the brackets) relative efficiency selected among all $\widehat{\alpha}_i$'s, which are measured by the ratio of the root mean squared error (RMSE) of $\widetilde{\alpha}_i$ to that of $\widehat{\alpha}_i$.

The simulation results indicate that, for a fixed n, the biases and relative efficiency performances improve as the time period T increases. On the other hand, for $T = 6$, the estimation biases appear to be irrelevant to the group size n. It is evident that the semiparametric estimators tend to be as efficient as their parametric counterparts

[Insert Table 1 here]

For the three different time spans $T = 6$, 15, and 50, given $n = 10$, Figure 1a draws the relative efficiencies of all to , based on the same 1000 replications of data. Figure 1b draws those for $n = 10, 30$, and 100, given $T = 6$. It is clear from the two figures that as T increases, the relative efficiencies tends to 1, indicating that our semiparametric estimators tend to be as efficient as the parametric ones. On the other hand, when n is getting larger, both the semiparametric and the OLS estimators become more volatile and suffer from larger simulated variances. However, our proposed semiparametric estimators exhibit less variability as n increases (detailed not reported) than the OLS estimators do. Hence the improvements in efficiencies are still found.

The foregoing simulation studies suggest the following conclusions. First, the mean square errors of the coefficient estimates deteriorate substantially when a local linear smoothing kernel is employed. Therefore, we suggest instead using local constant type estimators (based on higher order kernel or boundary Jackknifing carpentry when bias reduction is necessary) to establish the proposed estimator. Second, when n is small, say 10, the percentage biases of the slope estimators are all within about 2%, and the bias vanishes as T grows. In addition, the fixed-effects estimators perform slightly worse, with the percentage bias less than at most 6%; the bias decreases quickly as T increases. On the other hand, given a small T, say 6, similar results are found for the slope estimators when n grows. It can be seen that the pattern of change in the percentage bias of the fixed-effects parameters is ambiguous, but its magnitude is at most 8.7%, as n increases to 100. Third, even for a small T, the variances (standard deviations) of our proposed estimators are still quite close to those of the correctly specified OLS estimators. As n increases, both semiparametric and OLS estimators become less efficient because their variances enlarge. It is seen from the last column of Table 1 that the semiparametric estimators are nearly as efficient as the parametric OLS method, implying that the loss of efficiencies is negligible. The partial linear regression model proposed by this paper is clearly preferable to the parametric ones due to its efficiency property and failure to imposing a specific, most likely incorrect, functional form.

[Insert Figures 1a and 1b here]

3.2 Two Variables in $f(.)$

To understand the dimensional effect of the nonparametric function $f(.)$ on the performance of estimators $\widehat{\alpha}_i(i = 1, \cdots, n)$ and $\widehat{\beta}_d(d = 1, \cdots, p)$, a second simulation is conducted, in which variable $z_{1,it}$ is moved into $f(.)$, leaving $z_{2,it}$ and $z_{3,it}$ unchanged. We use the same 1000 dataset generated in the previous subsection, but assumed now that the nonparametric component $f(.)$ comprises both z_1 and x. Therefore, regression model (1-1) is now specified as $z_{it} = (z_{2,it}, z_{3,it})$, and $f(z_{1,,it}, x_{it}) = 15z_{1,it}+10+2\sqrt{x_{it}}+x_{it}/20$. Note here that the OLS estimates under this context all remains intact due to the linearity of model (1-1). Note here that a product kernel function for $q = 2$ has to be applied to (2-3), now let w_1 and w_2 denote any panel observations of z_1 and x, and define $r_1 = (z_{1,it} - w)/h_1, r_2 = (x_{it} - w)/h_2$, where h_1 and h_2 are the bandwidths. The product kernel used here is thus formulated as $K(r_1, r_2) = 0.75^2(1 - r_1^2)(1 - r_2^2)1_{[-1,1]\times[-1,1]}(r_1, r_2)$,

Under the above model setting, we attempt to quantify whether the bias and efficiency loss are getting worse when the nonparametric function $f(.)$ has two variables.

[Insert Table 2 here]

Table 2 summarizes the second simulation results and gives almost the same implications as Table 1. Given $n = 10$, the estimators are increasingly biased toward 0 as T grows. Although the bias of the estimator grows slightly as the number of firm increases, when T is fixed at 6, it is nevertheless acceptable. Generally speaking, the increase in the number of regressor contained by $f(.)$ has insignificant impact on the bias.

As an additional comparison, we compute the ratio of simulated standard error se($\widetilde{\alpha}$) to se($\widehat{\alpha}$) or se($\widetilde{\beta}$) to se($\widehat{\beta}$). The outcomes are shown in the parentheses. Except for the case of small n and T, i.e., $n = 10$ and $T = 6$, all these relative efficiency measures are above 75% – in some they exceed 90% – and increase quickly with either n or T grows. For the instance of $n = 10$ and $T = 6$, the ratio exceeds 57%.

As most empirical studies that use panel data contain much more units than 10, the employment of our semiparametric model appears to be feasible.

It is seen that most of the percentage bias figures in Table 2 are less than those in Table 1. However, all the relative efficiency measures of Table 2 fall below those of Table 1. The trade-off between the bias and efficiency is obvious. Evidence is found that the higher is the dimensionality of $f(.)$, the lower the bias will be, but the less efficient the estimator is in expense.

Again, Table 2 presents the ratio of RMSE of to the RMSE of in the last column. In all cases this ratio is in excess of 55% and in some it exceeds 85%. The comparison is most favorable to when either T or n increases. Thus for this model the disadvantage of using the semiparametric estimator is likely to be moderate. This is in sharp contrast to Table 1, where the same ratio is quite close to unity.

In summary, as far as the bias is concerned, both of the one- and two-variable cases are close to each other with the former case slightly worse than the latter. Furthermore, the loss of efficiency rises with the dimension in $f(.)$, but remains modest and, hence, acceptable at least for the two-variable model. The bias is inclined to be negatively correlated with the loss of efficiency.

4 An Empirical Application

In this section, we exemplify our semiparametric model by estimating a widely used translog production frontier, which allows for the expenditure on research and development (R&D) entering in an unknown form. The conventional translog production frontier is also estimated for the sake of comparison. Subsection 4.1 describes the data and Subsection 4.2 presents the empirical results.

4.1 The Data Source

(omitted)

[Insert Table 3 Here]

4.2 Empirical Results

The translog part of (1-1) excluding intercept term is specified as

$$y_{it} = \sum_{d=1}^{3} \beta_d z_{d,it} + \frac{1}{2}\sum_{d=1}^{3}\sum_{k=1}^{3} \beta_{d,k} z_{d,it} z_{k,it} + \beta_\tau t + \frac{1}{2}\beta_{\tau\tau} t^2 + \sum_{d=1}^{3} \beta_{\tau d} z_{d,it} t \tag{4-1}$$

where the symmetrical constraints, $\beta_{jk} = \beta_{kj} (\forall j \neq k)$, have to be imposed and linear trend t and quadratic trend are introduced to account for the possible productivity change over time. It is noteworthy that fixed parameters are estimated as the coefficients of the firm specific dummies. The last dummy variable is arbitrarily selected to be omitted in such a way as to avoid the problem of linear dependence between the dummy variables and nonparametric function . This appears to be understandable, because the functional from of is completely free such that the intercept term is likely to be included in.

[Insert Table 4 Here]

Table 4 shows the parameter estimates of (4-1) for the partially linear model and the conventional linear model. The estimated fixed effects are presented in appendix. In addition, the parametric model incorporates R&D as an extra term of the explanatory variables.

It is deserved a specific mention that the asymptotic distributions of $\widehat{\beta}_d$ and $\widehat{\alpha}_j$ from the semiparametric model are yet to be derived. To test the point null hypothesis $H_0 : \beta_d = 0$ (or $H_0 : \alpha_j = 0$) against the alternative $H_1 : \beta_d \neq 0$ (or $H_1 : \alpha_j \neq 0$), we propose using the bootstrap method to examine whether or not the estimates $\widehat{\beta}_d$ and $\widehat{\alpha}_j$ from our sample of data are compatible with the null hypotheses. The bootstrap method is a common practice to get the small-sample distribution of the chosen estimator or test statistic. Table 4 lists the bootstrapped single tail p-values for the semiparametric case, while the two-tail p-values for the parametric case (not shown) are directly adapted from the t distribution. Readers are suggested to refer to Yatchew (2002) for the exact resampling procedure adopted by this exercise.

Under a two-tail test, there are 4 parameter estimates are significant at the 10% level of significance for the both models. Interesting enough, all the slope parameter estimates of the both models have the same signs and the magnitudes of each pair of these estimates are close to some extent. Similar inference can be drawn from the fixed parameter estimates as shown in the appendix. There are only three out of 95 pairs of the fixed parameters having distinct signs. Since the coefficient estimates of R&D for the parametric model is insignificant, one is led to conclude that either the variable is an irrelevant independent variable to the regression model, or the functional form of the regression model is subject to specification errors. However, as the semiparametric model that introduces the same variable – R&D – is found to have equal number of significant slope parameter estimates and all these estimates do not deviate very far from those of the parametric model, variable R&D seems to be an important factor in the description of the production process for the electronics industry. Hence, the functional form misspecification may be the primary cause.

5 Concluding Remarks

This paper has proposed a feasible estimation procedure particularly suitable for a fixed-effect panel data model under the framework of partially linear model settings. For the past few decades, such data banks having several years of data on a number of firms, households, or geographical areas become more and more popular. Although there exist several econometric approaches amenable to panel data, it is nevertheless devoid of valid semiparametric approaches that are able to utilize the enriched sample data and to yield unbiased and almost the same efficient estimators as the ones obtained by the parametric approach. The estimation procedure suggested by this paper attempts to help shed some light on filling up this gap existent in the literature.

According to the results of the Monte Carlo simulations, our procedure appears to be able to result in desirable estimators in terms of the percentage bias and the relative efficiencies. For the case of single variable, the slope estimators perform much better than the fixed parameter estimators. However, the performance of the fixed effects estimators of the semiparametric model is close to those of the correctly specified parametric model, on the basis of the relative RMSE of to . As far as the instance of two variables is concerned, the efficiency of the semiparametric estimator deteriorates moderately, while its bias tends to be lessened somewhat.

To illustrate our estimation procedure, an empirical study on the production function of Taiwan's electronic industry is carried out, in the context of both semiparametric and parametric settings. When the true functional form of the production frontier is not known, a priori, which is usually the case, the semiparametric model should be more competent than the parametric model. The use of the semiparametic model appears to enhance our empirical capacity for understanding a firm's production technology

Figure 1:

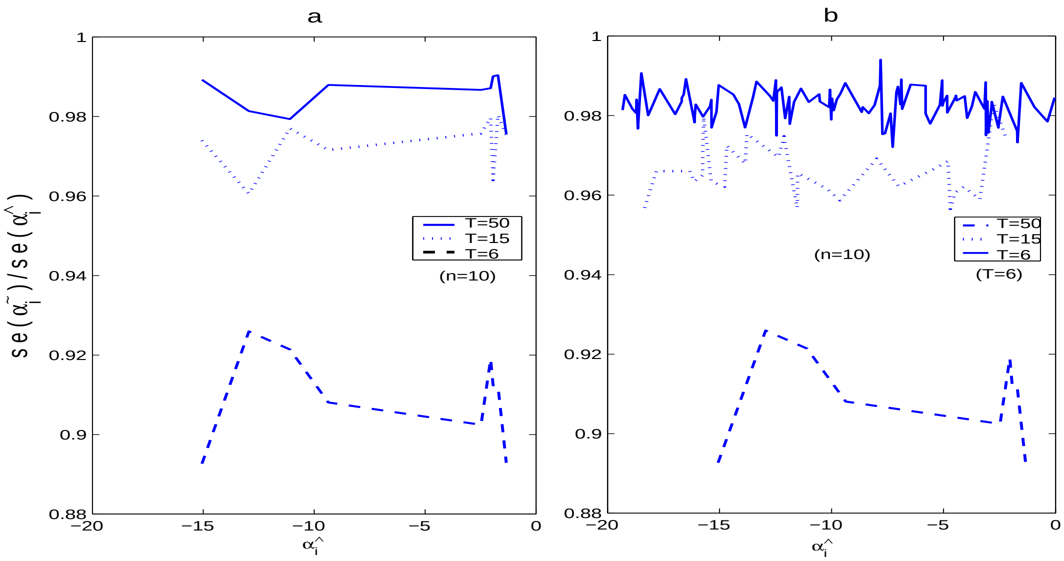

TABLE 3.1

(n,T)	$\frac{\bar{\bar{\alpha}}_1-\alpha_1}{\alpha_1}$	$\frac{\bar{\bar{\alpha}}_{\frac{n}{2}}-\alpha_{\frac{n}{2}}}{\alpha_{\frac{n}{2}}}$	$\frac{\bar{\bar{\alpha}}_n-\alpha_n}{\alpha_n}$	$\frac{\bar{\bar{\beta}}_1-\beta_1}{\beta_1}$	$\frac{\bar{\bar{\beta}}_2-\beta_2}{\beta_2}$	$\frac{\bar{\bar{\beta}}_3-\beta_3}{\beta_3}$	$\frac{RMSE(\widehat{\alpha}_i)}{RMSE(\widetilde{\alpha}_i)}$
(10,30)	-2.46	0.25	-0.42	0.10	-0.98	0.33	97.65
	(97.40)	(97.58)	(97.54)	(99.45)	(98.38)	(99.01)	(99.05)
(10,15)	5.70	0.84	0.32	0.09	2.03	-0.30	95.06
	(96.17)	(96.16)	(97.02)	(95.28)	(96.53)	(95.05)	(97.67)
(10,6)	6.10	-3.10	-2.30	-0.14	1.20	0.40	88.26
	(92.97)	(92.47)	(93.51)	(92.63)	(91.51)	(90.19)	(93.49)
(30,6)	-8.70	1.89	1.35	-0.01	-1.81	0.37	95.06
	(95.68)	(97.36)	(97.41)	(96.16)	(96.48)	(97.12)	(98.30)
(100,6)	2.14	6.50	0.96	0.02	-0.15	-0.11	97.20
	(98.14)	(98.55)	(98.45)	(99.36)	(98.54)	(98.57)	(99.38)

TABLE 3.2

(n,T)	$\frac{\bar{\bar{\alpha}}_1-\alpha_1}{\alpha_1}$	$\frac{\bar{\bar{\alpha}}_{\frac{n}{2}}-\alpha_{\frac{n}{2}}}{\alpha_{\frac{n}{2}}}$	$\frac{\bar{\bar{\alpha}}_n-\alpha_n}{\alpha_n}$	$\frac{\bar{\bar{\beta}}_2-\beta_2}{\beta_2}$	$\frac{\bar{\bar{\beta}}_3-\beta_3}{\beta_3}$	$\frac{RMSE(\widehat{\alpha}_i)}{RMSE(\widetilde{\alpha}_i)}$
(10,50)	-0.61	0.09	-0.99	0.10	-0.20	(98.00)
	(91.29)	(91.15)	(90.02)	(91.24)	(92.71)	(92.69)
(10,15)	1.93	1.56	-0.03	2.66	0.32	77.12
	(78.09)	(78.63)	(82.39)	(78.06)	(77.11)	(82.39)
(10,6)	-1.65	-5.42	-0.06	2.82	-0.14	57.40
	(60.46)	(59.90)	(57.40)	(59.44)	(57.86)	(61.65)
(30,6)	-4.56	-1.19	-0.76	-0.76	-2.42	77.79
	(79.62)	(82.3)	(79.94)	(79.04)	(79.54)	(82.38)
(100,6)	0.92	4.68	0.27	-0.32	-0.09	87.37
	(88.89)	(92.39)	(89.00)	(92.10)	(89.63)	(92.40)