EFFICIENT ESTIMATION OF A PARTIAL LINEAR MODEL UNDER HETEROSKEDASTICITY WITH UNKNOWN FORM

ERIC S. LIN[1]

Abstract

In this paper we consider the series estimator for the partial linear regression model proposed in Li (2000) to allow for heteroskedastictiy with unknown form. We propose an alternative estimator and prove that it achieves Chamberlain's (1992) semi-parametric efficiency bound. The proposed estimator shares the same first-order asymptotic properties as Li (2000). The Monte Carlo experiment shows that our estimator behaves in a way that is quite similar to Li (2000) estimator. To overcome the problem of picking smoothing parameters in series estimation, we propose minimizing the bootstrapping approximate mean square error to choose the smoothing parameters. By using the true mean square error as the benchmark, the bootstrap method works well and provides us with the criteria to choose two smoothing parameters simultaneously.

Keywords: Partial linear model, Semiparametric efficiency bound, Heteroskedasticity

JEL Classification: C01, C13, C14

1. Introduction

Nonparametric methods have become quite popular in economics in recent decades. While nonparametric regression is flexible in recovering the true shape of the regression curve without specifying a parametric family for the data, it has some disadvantages. The most fundamental problem is the well known "curse of dimensionality". To overcome this problem, a useful approach is to keep certain variables nonparametric but to adopt a parametric form for the variables of interest. A popular method for doing this is to specify the regression model as:

$$y_i = x_i' \beta + g(z_i) + u_i, \qquad \ldots (1)$$

where $g(\cdot)$ is an unknown nonparametric function and is usually highly dimensional. This model is referred to as a partial linear or semilinear regression model. Engle, Granger, Rice and Weiss (1986) apply this model to study the effect of weather on electricity demand. The partial linear specification also appears in various sample selection models such as Newey, Powell and Walker (1990), and Lee, Rosenzweig and Pitt (1992).

Previous studies on the estimation of partial linear models include Engle, Granger, Rice and Weiss (1986), Heckman (1986), Rice (1986), Chen (1988), Speckman (1988), Robinson

---

[1] Assistant Professor, Department of Economics, National Tsing Hua University, Hsin-Chu, Taiwan 30013. Email: slin@mx.nthu.edu.tw. I would like to thank Stephen Donald and Aman Ullah for helpful comments. Of course, all errors are mine.

(1988), Linton (1995), Donald and Newey (1994), Hong and Cheng (1999), Li (2000), and many others. Engle, Granger, Rice and Weiss (1986) propose the partial spline smoothing approach. This method was further studied by Rice (1986) and Heckman (1986). Rice (1986) obtains the asymptotic bias of a partial spline smoothing estimator of β and shows that this approach cannot attain the Berry-Esseen rate $\sqrt{n}$ for the estimator of β unless x and z are uncorrelated or the unknown nonparametric component $g(\cdot)$ is undersmoothed.[2] Chen (1988) proposes a kind of piecewise polynomial approximation to $g(\cdot)$, and the convergence rate of $\hat{\beta}$ is shown to be $\sqrt{n}$ consistent with the smallest possible variance even when x and z are dependent. Speckman (1988) considers kernel smoothing and proves that the parametric rate of $\hat{\beta}$ is attainable for the usual "optimal" bandwidth choice under the optimal nonparametric convergence rate for the estimation of $g(\cdot)$. Robinson (1988) constructs a feasible least squares estimator of β using Nadaraya-Watson kernel estimators of $E[y \mid z]$ and $E[x \mid z]$. He proves that $\hat{\beta}$ is $\sqrt{n}$ consistent and asymptotically normal. Linton (1995) proposes the local polynomial regression method to estimate $E[y \mid z]$ and $E[x \mid z]$.[3] He establishes the $\sqrt{n}$ consistent estimator of β and finds that it is second-order optimal using a second-order approximation of $\sqrt{n}(\hat{\beta}-\beta)$. Donald and Newey (1994) use a series approximation for the unknown function $g(\cdot)$. They show that the estimator is a $\sqrt{n}$ consistent estimator and asymptotically normal under weak conditions.[4] Hong and Cheng (1999) revisit the kernel smoothing method and show that the normal approximation rate of β is achieved only when bandwidth h is at a rate $n^{-1/4}$ instead of the usual "optimal" bandwidth rate $n^{-1/5}$. Li (2000) considers the additive partial linear model using the series estimation method and proves that the estimator of finite dimensional parameter β reaches the semiparametric efficiency bound under homoskedasticity.[5] Another approach to the partial linear model is to avoid the nonparametric estimation procedure. Yatchew (1997) proposes a differencing estimator to remove the effect of the unknown function $g(\cdot)$. The differencing estimator is in general not efficient but Yatchew also illustrates the generalized method of differencing to achieve the same asymptotic efficiency bound obtained by Robinson (1988).

---

2 This means that the $\sqrt{n}$ parametric rate for the estimation of β and the optimal nonparametric rate for the estimation of $g(\cdot)$ could not be attained simultaneously in the partial spline smoothing approach.

3 Linton (1995) adopts a local polynomial regression estimator instead of Robinson's (1988) Nadaraya-Watson kernel estimator due to the nice properties of the local polynomial regression estimator which is design adaptive and is able to correct the boundary bias problem.

4 The condition is weaker than in previous studies in that the modulus of continuity of $g(z)$ and $E[x \mid z]$ is higher than 1/4 of the dimension of z and the number of terms is chosen appropriately. In addition, the covariates z may not only be multidimensional but also discrete.

5 Li's result is based on homoskedastic errors. However, Chamberlain's (1992) semiparametric efficiency bound can allow for conditional heteroskedasticity.

In this paper we will focus on Li's (2000) estimator under heteroskedasticity with unknown form and show that it attains Chamberlain's (1992) semiparametric efficiency bound. In addition, we contribute to the literature on partially linear models by proposing an alternative estimator which has the same first-order asymptotics as Li's estimator. Both Li's estimator and our estimator involve dealing with not only the unknown function $g(\cdot)$ but also an unknown variance function which is allowed to depend on all of the regressors. Since picking two smoothing parameters may be difficult in practice, we propose bootstrapping the approximate mean square errors to choose the two smoothing parameters.

The remainder of this paper is organized as follows. In Section 2 we describe the model and estimation techniques adopted in this paper. The first-order asymptotic results for Li's estimator and our estimator are provided in Section 3. In Section 4 we conduct a small-scale Monte Carlo experiment. Section 5 concludes this paper.

2. The Model

Consider a partial linear regression model as in (1.1):

$$y_i = x_i'\beta + g(z_i) + u_i, \qquad \text{... (2.1)}$$

where the covariates x_i and z_i are of dimension r and q, respectively, $\beta = (\beta_1,...,\beta_r)'$ is a $r \times 1$ vector of an unknown parameter, and $g(\cdot)$ is an unknown function. Of course, we can extend (2.1) to the additive partially linear regression model by setting:

$$g(z_i) = g_1(z_{1i}) + g_2(z_{2i}) + ... + g_L(z_{Li}),$$

where $g_l(z_{li})$ is a scalar, and z_{li} is of dimension q_l ($q_l \geq 1, l = 1,2,...,L$). For simplicity, we assume $L = 1$, and $q_l = q$ in this paper. In matrix form, we can write (2.1) as

$$y = x\beta + g(z) + u. \qquad \text{... (2.2)}$$

The identification condition for β in (2.2) is stated below.

Assumption 2.1. (Identification) To identify the partial linear regression model in (2.2), we need $E[(x - E[x \mid z])'(x - E[x \mid z])]$ to be positive definite.

Put literally, we need the random variable x to not be fully contained in z. To understand the identification condition, taking the expectation of (2.2) conditional on z gives:

$$E[y \mid z] = E[x \mid z]\beta + g(z) + E[u \mid z]. \qquad \text{... (2.3)}$$

Substracting (2.3) from (2.2) gives:

$$y - E[y \mid z] = [x - E[x \mid z]]\beta + u - E[u \mid z]. \qquad \text{... (2.4)}$$

From (2.4), it is obvious that the identification of β requires the full rank of $x - E[x \mid z]$. In the context of a sample selection model where z would represent the variables that affect selection, we can have a situation where z is a linear function of some variable that appears in x provided that there is also a variable that predicts selection but does not appear in x (see Newey, Powell and Walker (1990)).

We now describe the estimation methods that are to be used in this paper. The estimation strategy of model (2.2) recommended in Robinson (1988) involves estimating $E[y \mid z]$ and $E[x \mid z]$ nonparametrically (e.g., using the Nadaraya-Watson or local linear kernel smoothing method) and regressing $y - E[y \mid z]$ on $x - E[x \mid z]$ to yield the estimate of β. Alternatively, we can use the sieve method (e.g., power series or spline) to estimate the conditional mean functions of $E[y \mid z]$ and $E[x \mid z]$.[6] We define the series basis functions as follows:

$$p_K(z) = (p_{1K}(z), p_{2K}(z), \ldots, p_{KK}(z))',$$

Here p_K denotes the $n \times K$ matrix with the ith row $p_{Ki} = p_K(z_i)$. The projection matrix is defined by $Q = p_K\left(p_K' p_K\right)^{-1} p_K'$. Then the estimator of β using the partialled out series based method, as first suggested in Donald and Newey (1994), is given by:

$$\begin{aligned}\hat{\beta} &= \left[(x - Qx)'(x - Qx)\right]^{-1}(x - Qx)(y - Qy) \\ &= \left[x'(I - Q)x\right]^{-1} x'(I - Q)y.\end{aligned}$$

Now, the unknown function $g(\cdot)$ can be estimated by $\hat{g} = p_K(z)\hat{\pi}$, where $\hat{\pi}$ is given by:

$$\hat{\pi} = \left(p_K' p_K\right)^{-1} p_K'(y - x\hat{\beta}).$$

Li (2000) verifies that, under the homoskedasticity assumption, $\hat{\beta}$ will be semiparametrically efficient in the sense that the inverse of the asymptotic variance of $\sqrt{n}(\hat{\beta} - \beta)$ achieves Chamberlain's semiparametric efficiency bound. He also establishes $\sqrt{n}$-consistency of $\hat{\beta}$ under conditional homoskedasticity. However, if the disturbances are heteroskedastic, $\hat{\beta}$ will in general not be semiparametrically efficient.[7] Therefore, Li (2000) suggests using a GLS-type

6 Li (2000) discusses the advantages of using the series approach to estimate the partial linear model as compared with the kernel based approach.

7 According to Li (2000), the estimator $\hat{\beta}$ is said to be "local efficient" since its efficiency is attained when some restrictions are satisfied. Here, it means that the assumption of homoskedasticity is satisfied.

estimator by regressing y_i/σ_i on x_i/σ_i and $p_K(z_i)/\sigma_i$, where $\sigma_i^2 = E(u_i^2 \mid x_i, z_i)$. We let $\hat{\beta}_{GLS}$ denote the corresponding estimator of β and note that it has the form:

$$\hat{\beta}_{GLS} = \left[x^{*\prime}\left(I - Q^*\right)x^*\right]^{-1}\left[x^{*\prime}\left(I - Q^*\right)y^*\right], \qquad \text{... (2.5)}$$

where $x^* = \left(x_1/\sigma_1, ..., x_n/\sigma_n\right)'$, $Q^* = p_K^*\left(p_K^{*\prime}p_K^*\right)^{-1}p_K^{*\prime}$, and $p_K^* = (p_{K1}/\sigma_1, ..., p_{Kn}/\sigma_n)'$. Without providing the proof, Li (2000) asserts that the method should produce a semiparametric efficient estimator of β. We will prove this fact in Theorem 3.1. To implement this estimator, we use an estimate of the variance function σ_i^2 which can then be plugged into the projection matrix Q^*. This task can be performed by using a preliminary consistent estimator of the model, such as $\hat{\beta}_{OLS}$, and then regressing the squared OLS residuals ($\hat{u}_i = y_i - x_i'\hat{\beta}_{OLS} - \hat{g}(z_i)$) on x and z using some nonparametric regression methods. Carrol (1982) proposes the kernel estimation method. Robinson (1987) suggests using the k-nearest neighbor method to estimate σ_i^2. Alternatively, we can utilize a series-based method. To sum up, the GLS approach proposed by Li (2000) is essentially to obtain the weighted least square in the first stage, and then to partial out the unknown g function in the second stage to obtain the efficient estimator $\hat{\beta}_{GLS}$.[8]

An alternative estimator proposed in this paper differs from Li (2000) in that we implement a weighted least square by regressing $y - E(y \mid x)$ on $x - E(x \mid z)$ using weights that are the inverse of the variance. All conditional expectations are estimated via a series regression so that if the variances were known we would have the following GLS estimator:

$$\tilde{\beta}_{GLS} = \left[x'(I-Q)\Sigma^{-1}(I-Q)x\right]^{-1}\left[x'(I-Q)\Sigma^{-1}(I-Q)y\right] \qquad \text{... (2.6)}$$

Although the estimators in (2.5) and (2.6) look different, we can prove that our estimator shares the same first-order asymptotic result as the Li's estimator. The reason why weighting after removing the mean works is that essentially we are estimating the model:

$$y_i - E[y_i \mid z_i] = [x_i - E[x_i \mid z_i]]\beta + u_i \qquad \text{... (2.7)}$$

by weighted least squares with weights that are the inverse of the variances. The regression in (2.7) is then equivalent to performing the following regression:

$$\frac{y_i - E[y_i \mid z_i]}{\sigma_i} = \frac{[x_i - E[x_i \mid z_i]]}{\sigma_i}\beta + \frac{u_i}{\sigma_i}. \qquad \text{... (2.8)}$$

The error term in (2.8) is clearly conditionally homoskedastic. This means that our estimator inherits the semiparametric efficiency.

---

8 Of course, we need to use a preliminary estimator of β to estimate the variance function before implementing the weighted least square.

A difficulty with the feasible GLS approach for Li (2000) and the alternative estimators is that it is necessary to know how the estimator depends on the smoothing parameters and in this case there will be two. One smoothing parameter relates to the number of functions used to approximate $g(\cdot)$ (say, K) as well as the number of functions used to approximate the variance function (say, H). We will discuss the issue of picking the smoothing parameters in Section 3.

3. First-Order Asymptotics

The following assumptions are needed to establish our main results.

Assumption 3.1. (i) $(y_i, x_i, z_i), i = 1,...,n$ are i.i.d. (independent and identically distributed); the support of (x,z) is a compact subset of R^{q+r}; (ii) $E[u_i \mid x_i, z_i] = 0$, $E[u_i^2 \mid x_i, z_i] = \sigma_i^2$ and u_i has bounded fourth moments; (iii) Let $x_i = E(x_i \mid z_i) + \varepsilon_i = h(z_i) + \varepsilon_i$, $E(\varepsilon_i \mid z_i) = 0$, and $E(\varepsilon_i^2 \mid z_i)$ is bounded away from ∞; (iv) All of $h(z_i)$ and σ_i^2 are bounded functions on the support of (x,z).

Assumption 3.1 (i) is quite standard in a regression model. Assumption 3.1 (ii) allows for conditional heteroskedasticity. Assumption 3.1 (iii) assumes that x_i is a function of z_i plus a random element that has a finite variance. These conditions plus Assumptions 3.2 and 3.3 discussed below will make it possible to estimate the various unknown functions.

Assumption 3.2. For every K there is a nonsingular constant matrix B such that for $P^K(z) = Bp^K(z)$: (i) the smallest eigenvalue of $E\left[p^K(z_i)p^K(z_i)'\right]$ is bounded away from zero uniformly in K and; (ii) there is a sequence of constants $\zeta(K)$ satisfying $\sup_{z\in Z}\left\|p^K(z)\right\| \le \zeta(K)$ and $K = K(n)$ such that $\zeta(K)^2K/n \to 0$ as $n \to \infty$, where Z is the support of z.

Assumption 3.2 is usually imposed on series estimators. See Newey (1997) for a further discussion. This assumption normalizes the approximating function. Part (i) bounds the second moment matrix away from singularity. Part (ii) controls the convergence rate of the series estimator.

Assumption 3.3. (i) For $f = g$ or $f = h$, there exist some π_f and α_f (>0) such that $\sup_{z\in Z}\left|f(z) - P^K(z)'\pi_f\right| = O\left(K^{-\alpha_f}\right)$ as $K \to \infty$; also, $\sqrt{n}K^{-\alpha_f} \to 0$ as $n \to \infty$. (ii) For σ^2, there exist some π_σ and α_σ (>0) such that $\sup_{(x,z)\in X\times Z}\left|\sigma^2(x,z) - P^H(x,z)'\pi_\sigma\right| = O\left(H^{-\alpha_\sigma}\right)$ as $H \to \infty$; also, $\sqrt{n}H^{-\alpha_\sigma} \to 0$ as $n \to \infty$.

Assumption 3.3 specifies the bound of the approximation error when we approximate unknown functions g or h as well as the variance function as will be required in order to implement a feasible GLS estimator of the partially linear model. Note that there are two smoothing parameters K and H that are required for estimation. Assumptions 2.1-3.3 make it possible to approximate the unknown functions and in turn estimate the parameter of interest, β. The following theorem gives the first-order asymptotic distribution of $\hat{\beta}_{GLS}$, which is the GLS estimator of Li in (2.5). All the proofs in this section are included in the Appendix.

Theorem 3.1. Define $x_i = h(z_i) + \varepsilon_i$ and assume that $E[\varepsilon_i \varepsilon_i'/\sigma_i^2]$ is positive definite, then under Assumptions 2.1-3.3, we have

$$\sqrt{n}(\hat{\beta}_{GLS} - \beta) \xrightarrow{d} N(0, J_0^{-1}),$$

where $J_0 = E\{[x_i - h(z_i)][var(u_i \mid x_i, z_i)][x_i - h(z_i)]\} = E[\varepsilon_i \varepsilon_i'/\sigma_i^2]$ is Chamberlain's semi-parametric efficiency bound.

Proof. The proof is given in the Appendix.

The next theorem states that our estimator $\tilde{\beta}_{GLS}$ in (2.6) is semiparametric efficient.

Theorem 3.2. Define $x_i = h(z_i) + \varepsilon_i$ and assume that $E[\varepsilon_i \varepsilon_i'/\sigma_i^2]$ is positive definite, then under Assumptions 2.1-3.3, we have

$$\sqrt{n}(\tilde{\beta}_{GLS} - \beta) \xrightarrow{d} N(0, J_0^{-1}),$$

where $J_0 = E\{[x_i - h(z_i)][var(u_i \mid x_i, z_i)][x_i - h(z_i)]\} = E[\varepsilon_i \varepsilon_i'/\sigma_i^2]$ is Chamberlain's semi-parametric efficiency bound.

Proof. The proof is given in the Appendix.

From Theorems 3.1 and 3.2, we can see that the Li (2000) and the proposed estimators under heteroskedasticity are the same in terms of the first-order asymptotics. They are both asymptotically normally distributed with the same asymptotic variance covariance matrix, which achieves the semi-parametric efficiency bound. However, it would be interesting to compare the finite sample properties of the two efficient estimators. One way to compare the two estimators is through the higher-order stochastic expansion. For more details see Ullah (2004). We now write the feasible version of the Li (2000) estimators as follows:[9]

$$\sqrt{n}(\hat{\beta}_{FGLS} - \beta) = \left[\frac{1}{n}\sum_{i=1}^{n}\frac{[x_i - \tilde{x}_i^*]^2}{\hat{\sigma}_i^2}\right]^{-1}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{[x_i - \tilde{x}_i^*][g_i - \tilde{g}_i^* + u_i - \tilde{u}_i^*]}{\hat{\sigma}_i^2}\right], \quad \ldots (3.1)$$

where the "$*$" represents a variable that is multiplied by a projection matrix which is formed using the weighted version of the approximating functions. As we expand equation (3.1), we have to expand not only the term $\hat{\sigma}_i^2$ in the general denominator, but also the implicit $\hat{\sigma}_i^2$ in the

---

9 For illustration purposes, we assume that the variables are scalars.

weighted approximating function (e.g., $\tilde{x}_i^*$ and $\tilde{g}_i^*$). This will complicate the expansions. As for the proposed estimator:

$$\sqrt{n}(\tilde{\beta}_{FGLS}-\beta)=\left[\frac{1}{n}\sum_{i=1}^{n}\frac{[x_i-\tilde{x}_i]^2}{\hat{\sigma}_i^2}\right]^{-1}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{[x_i-\tilde{x}_i][g_i-\tilde{g}_i+u_i-\tilde{u}_i]}{\hat{\sigma}_i^2}\right]. \quad \text{... (3.2)}$$

The projection matrix in forming $\tilde{x}$ does not involve the weighted random factor, $\hat{\sigma}_i^2$, and should be more amenable to higher-order expansion.

4. Monte Carlo Experiment

4.1. Simulation Design

We design the data generating process for the Monte Carlo experiment in this section as follows:

$$y_i=\beta\cdot x_i+\exp(z_i)+u_i\cdot\sigma^2(x_i,z_i) \quad \text{... (4.1)}$$

$$x_i=d\cdot z_i+v_i,\ z_i=\frac{i}{n},\ i=1,\dots,n$$

$$u_i\sim N(0,1),\ v_i\sim N(0,1)$$

$$d=10,\ \beta=1$$

The setting for the skedastic function σ_i^2 is given by:.

$$\sigma^2(x_i,z_i)=0\cdot x_i^2+0.3\cdot z_i^2.$$

We consider 1,000 replications for sample sizes of $n=100$, 200 and 400. The mean absolute bias (BIAS) and mean square error (MSE) are computed for four possible estimators, which include the preliminary estimator[10], the Li (2000) estimator, our estimator, and the kernel estimator. The procedures for computing different estimators are described below.

4.2. Estimators of β

4.2.1 Li (2000) estimator

1. Regress y_i on $p_K(z_i)$ and x_i on $p_K(z_i)$ and obtain the residuals $y_i-\tilde{y}_i$ and $x_i-\tilde{x}_i$, where

[10] The preliminary estimator represents the series estimator of the partial linear model without taking heteroskedasticity into consideration.

$$y_i - \tilde{y}_i = y_i - p_K(z_i)\left(p_K(z)' p_K(z)\right)^{-1} p_K(z)' y$$

$$x_i - \tilde{x}_i = x_i - p_K(z_i)\left(p_K(z)' p_K(z)\right)^{-1} p_K(z)' x.$$

2. Regress $y_i - \tilde{y}_i$ on $x_i - \tilde{x}_i$ and obtain the preliminary estimator of β, say b_o.

3. Estimate g by $\hat{g}(z_i) = p_K(z_i)\left(p_K(z)' p_K(z)\right)^{-1} p_K(z)'[y - xb_o]$.

4. Estimate u_i by $\hat{u}_i = y_i - x_i b_o - \hat{g}(z_i)$.

5. Estimate σ_i^2 by $\hat{\sigma}_i^2 = p_H(z_i)\left(p_H(z)' p_H(z)\right)^{-1} p_H(z)' \widehat{u2}$, where $\widehat{u2}$ is a column vector of $\hat{u}_i^2$.

6. Regress $y_i/\hat{\sigma}_i$ on $x_i/\hat{\sigma}_i$, and $p_K(z_i)/\hat{\sigma}_i$ to yield the Li (2000) semiparametric efficient estimator of β.

4.2.2 Alternative estimator

1. Regress y_i on $p_K(z_i)$ and x_i on $p_K(z_i)$ and obtain the residuals $y_i - \tilde{y}_i$ and $x_i - \tilde{x}_i$, where

$$y_i - \tilde{y}_i = y_i - p_K(z_i)\left(p_K(z)' p_K(z)\right)^{-1} p_K(z)' y$$

$$x_i - \tilde{x}_i = x_i - p_K(z_i)\left(p_K(z)' p_K(z)\right)^{-1} p_K(z)' x.$$

2. Regress $y_i - \tilde{y}_i$ on $x_i - \tilde{x}_i$ and obtain the preliminary estimator of β, say b_o.

3. Estimate g by $\hat{g}(z_i) = p_K(z_i)\left(p_K(z)' p_K(z)\right)^{-1} p_K(z)'[y - xb_o]$.

4. Estimate u_i by $\hat{u}_i = y_i - x_i b_o - \hat{g}(z_i)$.

5. Estimate σ_i^2 by $\hat{\sigma}_i^2 = p_H(z_i)\left(p_H(z)' p_H(z)\right)^{-1} p_H(z)' \widehat{u2}$, where $\widehat{u2}$ is a column vector of $\hat{u}_i^2$.

6. Regress $(y_i - \tilde{y}_i)/\hat{\sigma}_i$ on $(x_i - \tilde{x}_i)/\hat{\sigma}_i$ to obtain the proposed semiparametric efficient estimator of β.

It can be seen that our estimator differs from the Li (2000) estimator in the final stage of the estimation procedure, which is detailed in Section 2. To compare the series approximation of the unknown function with the kernel smoothing, we also consider the Nadaraya-Watson kernel type method in estimating g and σ_i^2.

4.2.3 Kernel estimator

1. Compute the kernel estimators of $E(y \mid z)$ and $E(x \mid z)$ by

$$\hat{y}_i = \frac{\overline{y}_i}{\hat{f}_i} = \frac{\frac{1}{nh}\sum_{j=1}^{n} y_j K\left(\frac{z_i - z_j}{h}\right)}{\frac{1}{nh}\sum_{j=1}^{n} K\left(\frac{z_i - z_j}{h}\right)}$$

$$\hat{x}_i = \frac{\overline{x}_i}{\hat{f}_i} = \frac{\frac{1}{nh}\sum_{j=1}^{n} x_j K\left(\frac{z_i - z_j}{h}\right)}{\frac{1}{nh}\sum_{j=1}^{n} K\left(\frac{z_i - z_j}{h}\right)},$$

where $K(\cdot)$ is the kernel function and h is the bandwidth.[11]

2. Obtain initial $b_o = \left[(x-\hat{x})'(x-\hat{x})\right]^{-1}\left[(x-\hat{x})(y-\hat{y})\right]$.

3. Obtain the residuals by $\hat{u} = (y-\hat{y}) - (x-\hat{x})b_o$.

4. Form the kernel regression (or local linear regression) of $\hat{\sigma}_i^2$ using $\hat{u}_i^2$.[12]

5. Regress $(y_i - \hat{y}_i)/\hat{\sigma}_i$ on $(x_i - \hat{x}_i)/\hat{\sigma}_i$ to obtain an efficient estimator of β.

4.3. Simulation Results

We know that the series estimator for the variance function is not necessarily positive. To guarantee the positive variance estimate, trimming will be needed. The choice of the trimming parameters may affect the performance of the various estimators. We do not explore this issue in this paper. Instead, we arbitrarily set three possible trimming points (TP), $TP = .1$, $.01$ and $.001$.

The estimators for comparison include the proposed estimator (b_{Lin}), the Li (2000) estimator (b_{Li}), the preliminary estimator (b_o), and the kernel estimator (b_K). The simulation results are summarized in Table 1. We expect that b_o, which ignores the heteroskedasticity, will perform the worst. Table 1 confirms the expectation in terms of the bias and mean square error. The Li (2000) estimator has the minimum bias and MSE in almost all cases. Our estimator only dominates in the case where $n = 400$. However, it is worth noting that the alternative estimator performs in a way that is pretty much similar to the Li (2000) estimator. As the sample size increases to 400, the behavior of the two estimators is just about equivalent. This finding also confirms the equivalent first-order asymptotics for the two estimators as derived in Section 3. As for the kernel estimator, it is dominated by both the Li (2000) estimator and our estimator.

---

[11] Here we utilize the Gaussian kernel and pick the bandwidth by

[12] Here we still utilize the Gaussian kernel and pick the bandwidth in the same way as in Step 1.

To see the impact of a different setup of the unknown g function on the simulation results, we then change the setting by letting $g(z) = (1+z)^3$. The results are presented in Table 2. One can see that in this setting the Li (2000) estimator still dominates although our estimator is quite close to Li's. It should be noted that the kernel estimator performs badly and even worse than the preliminary estimator in this particular setting.

Note that throughout the simulation we arbitrarily pick the approximating functions for $g(z)$ and $\sigma^2(z)$ as $(1, z, z^2)$ and $(1, z, z^2)$, respectively. The issue of how to pick the optimal smoothing parameters will be discussed in the next subsection.

4.4. Picking Smoothing Parameters

Even though the series estimator we propose in this paper is quite easy to implement, we still need to pick the number of approximating functions. We need to pick a smoothing parameter K for the approximation of g and H for the approximation of the variance function σ_i^2. The method we consider here is the use of the bootstrap to approximate the MSE and to pick smoothing parameters to minimize the estimated MSE. The use of the bootstrap-based procedure for selecting the moment condition has been discussed in Inoue (2006).

Of course, the bootstrapping method we suggest in this section can be easily applied to Li's estimator as well. In this experiment, we consider three possible sets which serve as the functions for approximating $g(z)$ and $\sigma^2(z)$. The DGP follows the same setup in (4.1). The numbers of Monte Carlo and bootstrapping replications are set to 1,000 and 399 for all cases. The resampling scheme involves bootstrapping the $\{x_i, z_i, y_i\}$ triple. The potential instrument sets for approximating $g(z)$ and $\sigma^2(z)$ are:

$$z^1 = (1, z)$$
$$z^2 = (1, z, z^2)$$
$$z^3 = (1, z, z^2, z^3).$$

Conducting the series estimation allows for 9 combinations of the instruments. We use the following notation to record each combination:

$$K_{11} = (z^1, z^1),\quad K_{21} = (z^2, z^1),\quad K_{31} = (z^3, z^1),$$
$$K_{12} = (z^1, z^2),\quad K_{22} = (z^2, z^2),\quad K_{32} = (z^3, z^2),$$
$$K_{13} = (z^1, z^3),\quad K_{23} = (z^2, z^3),\quad K_{33} = (z^3, z^3),$$

where K_{32} stands for using $(1, z, z^2, z^3)$ and $(1, z, z^2)$ as the instruments for approximating $g(z)$ and $\sigma^2(z)$, respectively. That is to say, we employ 4 and 3 instruments respectively, in forming the approximating functions. Note that we restrict our attention to the case of $TP = .001$.

The results are shown in Table 3. It can be observed that it is not the best strategy to choose as many functions such as picking K_{33}. Most of the situations (such as K_{21} or K_{31}) tend to choose more (say, three or four) instruments for $g(z)$ and just two instruments for $\sigma^2(z)$.[13] Because we know the DGP, the true MSE can be precisely calculated for different combinations of instruments. We compare the true MSE with the bootstrapping method and find that the optimal smoothing parameters chosen by the two methods are quite similar. For instance, as $n = 100$, the true MSE and bootstrapping MSE pick K_{21} and K_{31}, respectively.

5. Concluding Remarks

In this paper we explicitly prove that the series estimator considered in Li (2000) is semiparametric efficient under heteroskedasticity with unknown form. In addition, we propose an alternative estimator for the partial linear regression model and show that it shares the same first-order asymptotics as Li's (2000) estimator.

In the Monte Carlo experiment, we compare our estimator with the Li (2000) and kernel estimators in terms of the mean absolute bias and mean squared error. The simulation results show that our estimator behaves in a way that is quite similar to Li's (2000) estimator. In addition, the performance of the series type esitmators seems more robust to the setting of the unknown g function than the kernel estimator. It is necessary to determine two smoothing parameters in estimating the unknown g and σ_i^2. The usual way of doing this is to derive the approximate MSE through the higher-order stochastic expansion. The alternative estimator seems easier to apply when performing a higher-order expansion than the Li's estimator. This research direction is still ongoing. In this paper, we propose bootstrapping the approximate mean square error in order to choose the smoothing parameters. By using the true MSE as the benchmark, the bootstrapping method works very well and provides us with the criteria for choosing the two smoothing parameters simultaneously.

References

Carrol, R. J. (1982), "Adapting for Heteroskedasticity in Linear Model," *Annals of Statistics*, 10, 1224-1233.

Chamberlain, G. (1992), "Efficiency Bounds for Semiparametric Regression," *Econometrica*, 60, 567-596.

Chen, H. (1988), "Convergence Rates for Parametric Components in a Partly Linear Model," *Annals of Statistics*, 16, 136-146.

Donald, S. G. and W. K. Newey (1994), "Series Estimation of Semilinear Models," *Journal of Multivariate Analysis*, 50, 30-40.

[13] This may reflect the design of our DGP, i.e., we need more terms in the power series to approximate the exponential g function than the squared skedastic function.

Engle, R. F., C. W. J. Granger, J. Rice, and A. Weiss (1986), "Semiparametric Estimates of the Relation between Weather and Electricity Demand," *Journal of the American Statistical Association*, 76, 817-823.

Heckman, N. E. (1986), "Spline Smoothing in a Partially Linear Model," *Journal of the Royal Statistical Society, Series B*, 48, 244-248.

Hong, S. Y. and P. Cheng (1999), "Normal Approximation Rate of the Kernel Smoothing Estimator in a Partial Linear Model," *Journal of Multivariate Analysis*, 70, 207-220.

Inoue, A. (2006), "A Bootstrap Approach to Moment Selection," *Econometric Journal*, 9, 48-75.

Lee, L. F., M. R. Rosenzweig, and M. Pitt (1997), "The Effects of Improved Nutrition, Sanitation and Water Purity on Child Health in High-Mortality Populations," *Journal of Econometrics*, 77, 209-235.

Li, Q. (2000), "Efficient Estimation of Additive Partially Linear Models," *International Economic Review*, 41(4), 1073-1092.

Linton, O. (1995), "Second Order Approximation in the Partially Linear Regression Model," *Econometrica*, 63, 1079-1112.

Newey, W. K. (1997), "Convergence Rates and Asymptotic Normality for Series Estimator," *Journal of Econometrics*, 79, 147-168.

Newey, W. K., J. L. Powell, and J. R. Walker (1990), "Semiparametric Estimation of Selection Models: Some Empirical Results," *American Economic Review*, 80(2), 324-328.

Rice, J. (1986), "Convergence Rates for Partially Splined Models," *Statistics & Probability Letters*, 4, 203-208.

Robinson, P. M. (1987), "Asymptotically Efficient Estimation in the Presence of Heteroskedasticity," *Econometrica*, 55, 875-891.

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

Ullha (2004), *Finite Sample Econometrics*, Oxford University Press, New York.

Speckman, P. (1988), "Kernel Smoothing in Partial Linear Models," *Journal of the Royal Statistical Society, Series B*, 50, 413.

Yatchew, A. (1997), "An Elementary Estimator of the Partial Linear Model," *Economics Letters* 57, 135-143.

Appendix A: Proofs

Let C denote the generic constants throughout this Appendix. The Euclidean norm $\|\cdot\|$ for a matrix A is defined as $\|A\| = [tr(A'A)]^{1/2}$. Let C.S. denote the Cauchy-Schwartz inequality. According to the notation by Robinson (1988), for the scalar or column vector sequences A_i and B_i, we define $S_{A,B} = n^{-1}\sum_{i=1}^{n} A_i B_i'$ and $S_A = S_{A,A}$. The following lemma of Li (2000) are useful in the proof of our theorems. The proofs are referred to in Li (2000: p.1089-1090).[14]

Lemma 1. $\hat{Q} - I = O_p\left(\zeta(K)\sqrt{K}/\sqrt{n}\right)$, where $\hat{Q} = (P'P/n)$.

Lemma 2. $\|\tilde{\pi}_f - \pi_f\| = O_p\left(K^{-\alpha}\right)$, where $\tilde{\pi}_f = (P'P)^{-1}P'f$, and $f = g$ or $f = h$.

Lemma 3. $(Q'\eta/n) = O_p\left(\zeta(K)/\sqrt{n}\right) = o_p(1)$.

Lemma 4. $S_{f-\tilde{f}} = O_p\left(K^{-2\alpha}\right) = o_p\left(n^{-1/2}\right)$, where $f = g$ or $f = h$.

Lemma 5. (i) $S_{\tilde{v}} = O_p(K/n)$, (ii) $S_{\tilde{u}} = O_p(K/n)$, (iii) $S_{\tilde{\eta}} = o_p(1)$.

Corollary 1. If we replace the approximating function p_K by the normalized version (say $p_K^* = (p_{K1}/\sigma_1, p_{K2}/\sigma_2, \ldots, p_{Kn}/\sigma_n)'$), Lemmas 1-5 still hold.

Corollary 2. If we replace the random variables by the normalized version (e.g., $f = (f_1/\sigma_1, f_2/\sigma_2, \ldots, f_n/\sigma_n)'$), Lemmas 1-5 still hold.

Proof of Theorem 3.1

We can write $\sqrt{n}(\hat{\beta}_{GLS} - \beta)$ as

$$\sqrt{n}\left(\hat{\beta}_{GLS} - \beta\right) = \left[\frac{x^{*\prime}\left(I - Q^*\right)x^*}{n}\right]^{-1} \sqrt{n}\left[\frac{x^*\left(I - Q^*\right)\left(g^* + u^*\right)}{n}\right]$$

$$= S_{x^*-\tilde{x}^{**}}^{-1} \sqrt{n} S_{x^*-\tilde{x}^{**}, g^*-\tilde{g}^{**}+u^*-\tilde{u}^{**}}. \qquad \ldots(A-1)$$

What we want to do is to prove that the first term in (A-1) converges in probability based on the Law of Large Number and the second term converges in distribution by the Lindberg-Levi Central Limit Theorem. We use the following propositions to prove the results:

Proposition 1. $x^{*\prime}\left(I - Q^*\right)x^*/n = S_{x^*-\tilde{x}^{**}} = E[\varepsilon_i \varepsilon_i'/\sigma_i^2] + o_p(1)$.

Proof. Let $\tilde{x}_i^* = p_{Ki}^*\left(p_K^{*\prime} p_K^*\right)^{-1} p_K^{*\prime} x$. Using the definition of x_i and $\tilde{x}_i^*$ gives

---

[14] Note that we adopt the notation from Li (2000). In our paper, since we do not consider the additive structure of the partial linear model, there is no need to decompose ε as $v + \eta$. However, the order related to $\tilde{\varepsilon}$ in our proof is similar to the order of $\tilde{v}$ in Li's paper. For instance, it is trivial to see that $S_{\tilde{\varepsilon}} = O_p(K/n)$.

$$\frac{1}{n}x^{*\prime}\left(I-Q^{*}\right)x^{*}$$

$$=\frac{1}{n}\sum_{i=1}^{n}\frac{\varepsilon_i\varepsilon_i'}{\sigma_i^2}+\frac{1}{n}\sum_{i=1}^{n}\frac{\left[\left(h_i-\tilde{h}_i^{*}\right)-\tilde{\varepsilon}_i^{*}\right]\left[\left(h_i-\tilde{h}_i^{*}\right)-\tilde{\varepsilon}_i^{*}\right]'}{\sigma_i^2}$$

$$+\frac{2}{n}\sum_{i=1}^{n}\frac{\varepsilon_i\left[\left(h_i-\tilde{h}_i^{*}\right)-\tilde{\varepsilon}_i^{*}\right]'}{\sigma_i^2}$$

$$=S_{\varepsilon^{*}}+S_{(h^{*}-\tilde{h}^{**})-\tilde{\varepsilon}^{**}}+S_{\tilde{\varepsilon}^{**},(h^{*}-\tilde{h}^{**})-\tilde{\varepsilon}^{**}}. \qquad \text{...(A-2)}$$

Note that the variables with a single "*" represent normalization by σ_i and the variables with a double "**" stand for normalized variables which are premultiplied by the normalized projection matrix Q^{*}. By LLN, the first term in (A-2) will converge to $E\left[\varepsilon_i\varepsilon_i'/\sigma_i^2\right]+o_p(1)$. We also have the inequality $S_{(h^{*}-\tilde{h}^{**})-\tilde{\varepsilon}^{**}}\leq 2\left[S_{h^{*}-\tilde{h}^{**}}+S_{\tilde{\varepsilon}^{**}}\right]=o_p(1)$ by Corollary 1, Corollary 2, Lemma 4, and Lemma 5 (i) and (iii). Applying CS to the last term in (A-2) gives

$$S_{\tilde{\varepsilon}^{**},(h^{*}-\tilde{h}^{**})-\tilde{\varepsilon}^{**}}\leq\left(S_{\tilde{\varepsilon}^{**}}S_{(h^{*}-\tilde{h}^{**})-\tilde{\varepsilon}^{**}}\right)^{1/2}=\left(O_p(1)o_p(1)\right)^{1/2}=o_p(1). \qquad \square$$

Proposition 2. $S_{x^{*}-\tilde{x}^{**},g^{*}-\tilde{g}^{**}}=o_p\left(n^{-1/2}\right)$

Proof. Using definition of x^{*} and $\tilde{x}^{**}$ gives

$$S_{x^{*}-\tilde{x}^{**},g^{*}-\tilde{g}^{**}}=S_{\varepsilon^{*},g^{*}-\tilde{g}^{**}}+S_{h^{*}-\tilde{h}^{**},g^{*}-\tilde{g}^{**}}-S_{\tilde{\varepsilon}^{**},g^{*}-\tilde{g}^{**}}.$$

1. $S_{\varepsilon^{*},g^{*}-\tilde{g}^{**}}\leq\left(S_{\varepsilon^{*}}S_{g^{*}-\tilde{g}^{**}}\right)^{1/2}=O_p\left(K^{-\alpha}\right)$ by C.S., Proposition 1 and Lemma 4.
2. $S_{h^{*}-\tilde{h}^{**},g^{*}-\tilde{g}^{**}}\leq\left(S_{h^{*}-\tilde{h}^{**}}S_{g^{*}-\tilde{g}^{**}}\right)^{1/2}=O_p\left(K^{-2\alpha}\right)$ by C.S., and Lemma 4.
3. $S_{\tilde{\varepsilon}^{**},g^{*}-\tilde{g}^{**}}\leq\left(S_{\tilde{\varepsilon}^{**}}S_{g^{*}-\tilde{g}^{**}}\right)^{1/2}=o_p(1)O_p\left(K^{-\alpha}\right)$ by C.S., Lemma 5 (i) and Lemma 4. □

Proposition 3. $S_{x^{*}-\tilde{x}^{**},\tilde{u}^{**}}=o_p\left(n^{-1/2}\right)$.

Proof. Using the definitions of x^{*} and $\tilde{x}^{**}$ gives

$$S_{x^{*}-\tilde{x}^{**},g^{*}-\tilde{g}^{**}}=S_{\varepsilon^{*},\tilde{u}^{**}}+S_{h^{*}-\tilde{h}^{**},\tilde{u}^{**}}-S_{\tilde{\varepsilon}^{**},\tilde{u}^{**}}.$$

1. $E\left[\left\|S_{\varepsilon^{*},\tilde{u}^{**}}\right\|^2 \mid Z\right]=n^{-2}tr\left[Q^{*}\varepsilon^{*}\varepsilon^{*\prime}Q^{*}E\left[u^{*}u^{*\prime}\mid Z\right]\right]\leq Cn^{-2}tr\left[\tilde{\varepsilon}^{**}\tilde{\varepsilon}^{**\prime}\right]$

$=Cn^{-1}tr\left(S_{\tilde{\varepsilon}^{**}}\right)=O_p\left(K/n^2\right)$ by C.S. and Lemma 5 (i).

2. $S_{h^{*}-\tilde{h}^{**},\tilde{u}^{**}}\leq\left(S_{h^{*}-\tilde{h}^{**}}S_{\tilde{u}^{**}}\right)^{1/2}=O_p\left(K^{-\alpha}\right)O_p\left(\sqrt{K}/\sqrt{n}\right)$ by C.S., Lemma 4 and Lemma 5 (ii).

3. $S_{\bar{\varepsilon}^{**},\bar{u}^{**}} \le \left(S_{\bar{\varepsilon}^{**}} S_{\bar{u}^{**}}\right)^{1/2} = O_p(K/n)$ by C.S., Lemma 5 (i) and Lemma 5 (ii). □

Proposition 4. $\sqrt{n}S_{x^*-\tilde{x}^{**},u^*} \xrightarrow{d} N(0,\ E[\varepsilon_i\varepsilon_i'/\sigma_i^2])$.

Proof. Using the definitions of x^* and $\tilde{x}^{**}$ gives

$$S_{x^*-\tilde{x}^{**},u^*} = S_{\varepsilon^*,u^*} + S_{h^*-\tilde{h}^{**},u^*} - S_{\tilde{\varepsilon}^{**},u^*}.$$

1. $\sqrt{n}S_{\varepsilon^*,u^*} = \sum_{i=1}^{n}\left[\varepsilon_i u_i/\sigma_i^2\right]/\sqrt{n} \xrightarrow{d} N\left(0, E\left[\varepsilon_i\varepsilon_i'/\sigma_i^2\right]\right)$ by Lindberg-Levi Central Limit Theorem.

2. $E\left[\left\|S_{h^*-\tilde{h}^{**},u^*}\right\|^2 \mid Z\right] = n^{-2}tr\left[\left(h^*-\tilde{h}^{**}\right)\left(h^*-\tilde{h}^{**}\right)' E\left(u^*u^{*\prime} \mid Z\right)\right] \le Cn^{-1}tr\left[S_{h^*-\tilde{h}^{**}}\right] = o_p\left(n^{-1}\right)$ by C.S. and Lemma 4.

3. $E\left[\left\|S_{\tilde{\varepsilon}^{**},u^*}\right\|^2 \mid Z\right] = n^{-2}tr\left[Q^*\varepsilon^*\varepsilon^{*\prime}Q^*E\left(u^*u^{*\prime} \mid Z\right)\right] \le Cn^{-1}tr\left[S_{\tilde{\varepsilon}^{**}}\right] = o_p\left(n^{-1}\right)$ by C.S. and Lemma 5 (i). □

Combining Propositions 1-4 proves Theorem 3.1.

Proof of Theorem 3.2

Our new estimator could be written as

$$\sqrt{n}\left(\tilde{\beta}_{GLS}-\beta\right) = \left[\frac{x(I-Q)\Sigma^{-1}(I-Q)x}{n}\right]^{-1}\sqrt{n}\left[\frac{x(I-Q)\Sigma^{-1}(I-Q)(g+u)}{n}\right]$$

$$= S_{x^*-\tilde{x}^*}^{-1}\sqrt{n}S_{x^*-\tilde{x}^*,g^*-\tilde{g}^*+u^*-\tilde{u}^*}. \qquad \text{...(A-3)}$$

What we want is to prove that the first term in (A-3) coverges in probability by Law of Large Number and the second term converges in distribution by Central Limit Theorem. We use the following propositions to prove the results.

Proposition 5. $x^{*\prime}(I-Q)x^*/n = S_{x^*-\tilde{x}^*} = E[\varepsilon_i\varepsilon_i'/\sigma_i^2] + o_p(1)$.

Proof. Let $\tilde{x}_i = p_{Ki}\left(p_K'p_K\right)^{-1}p_K'x$. Using the definitions of x_i and $\tilde{x}_i$ gives

$$\frac{1}{n}x^{*\prime}(I-Q)x^*$$

$$= \frac{1}{n}\sum_{i=1}^{n}\frac{\varepsilon_i\varepsilon_i'}{\sigma_i^2} + \frac{1}{n}\sum_{i=1}^{n}\frac{\left[\left(h_i-\tilde{h}_i\right)-\tilde{\varepsilon}_i\right]\left[\left(h_i-\tilde{h}_i\right)-\tilde{\varepsilon}_i\right]'}{\sigma_i^2}$$

$$+\frac{2}{n}\sum_{i=1}^{n}\frac{\varepsilon_i\left[\left(h_i-\tilde{h}_i\right)-\tilde{\varepsilon}_i\right]'}{\sigma_i^2}$$

$$= S_{\varepsilon^*} + S_{(h^*-\tilde{h}^*)-\tilde{\varepsilon}^*} + S_{\varepsilon^*,(h^*-\tilde{h}^*)-\tilde{\varepsilon}^*}. \qquad \ldots(A-4)$$

Note that here we only have the variables with single "*" representing normalization by σ_i. And the projection matrix Q is not normalized by σ_i. By LLN, the first term in (A-4) will converge to $E[\varepsilon_i\varepsilon_i'/\sigma_i^2] + o_p(1)$. We also have the inequality $S_{(h^*-\tilde{h}^*)-\tilde{\varepsilon}^*} \le 2\left[S_{h^*-\tilde{h}^*} + S_{\tilde{\varepsilon}^*}\right] = o_p(1)$ based on Corollary 2, Lemma 4, and Lemma 5 (i) and (iii). Applying CS to the last term in (A-4) gives

$$S_{\varepsilon^*,(h^*-\tilde{h}^*)-\tilde{\varepsilon}^*} \le \left(S_{\varepsilon^*} S_{(h^*-\tilde{h}^*)-\tilde{\varepsilon}^*}\right)^{1/2} = \left(O_p(1)o_p(1)\right)^{1/2} = o_p(1).$$

Proposition 6. $S_{x^*-\tilde{x}^*,g^*-\tilde{g}^*} = o_p\left(n^{-1/2}\right)$

Proof. Using the definitions of x^* and $\tilde{x}^*$ gives

$$S_{x^*-\tilde{x}^*,g^*-\tilde{g}^*} = S_{\varepsilon^*,g^*-\tilde{g}^*} + S_{h^*-\tilde{h}^*,g^*-\tilde{g}^*} - S_{\tilde{\varepsilon}^*,g^*-\tilde{g}^*}.$$

1. $S_{\varepsilon^*,g^*-\tilde{g}^*} \le \left(S_{\varepsilon^*} S_{g^*-\tilde{g}^*}\right)^{1/2} = O_p\left(K^{-\alpha}\right)$ by C.S., Proposition 1 and Lemma 4.

2. $S_{h^*-\tilde{h}^*,g^*-\tilde{g}^*} \le \left(S_{h^*-\tilde{h}^*} S_{g^*-\tilde{g}^*}\right)^{1/2} = O_p\left(K^{-2\alpha}\right)$ by C.S., and Lemma 4.

3. $S_{\tilde{\varepsilon}^*,g^*-\tilde{g}^*} \le \left(S_{\tilde{\varepsilon}^*} S_{g^*-\tilde{g}^*}\right)^{1/2} = o_p(1)O_p\left(K^{-\alpha}\right)$ by C.S., Lemma 5 (i) and Lemma 4.

Proposition 7. $S_{x^*-\tilde{x}^*,\tilde{u}^*} = o_p\left(n^{-1/2}\right)$.

Proof. Using the definitions of x^* and $\tilde{x}^*$ gives

$$S_{x^*-\tilde{x}^*,g^*-\tilde{g}^*} = S_{\varepsilon^*,\tilde{u}^*} + S_{h-\tilde{h}^*,\tilde{u}^*} - S_{\tilde{\varepsilon}^*,\tilde{u}^*}.$$

1. $E\left[\left\|S_{\varepsilon,\tilde{u}^*}\right\|^2 \mid Z\right] = n^{-2}\mathrm{tr}\left[Q\varepsilon\varepsilon'QE\left[u^*u^{*\prime} \mid Z\right]\right] \le Cn^{-2}\mathrm{tr}[\tilde{\varepsilon}\tilde{\varepsilon}'] = Cn^{-1}\mathrm{tr}(S_{\tilde{\varepsilon}}) \quad = O_p\left(K/n^2\right)$ by C.S. and Lemma 5 (i).

2. $S_{h-\tilde{h}^*,\tilde{u}^*} \le \left(S_{h-\tilde{h}^*} S_{\tilde{u}^*}\right)^{1/2} = O_p\left(K^{-\alpha}\right)O_p\left(\sqrt{K}/\sqrt{n}\right)$ by C.S., Lemma 4 and Lemma 5 (ii).

3. $S_{\tilde{\varepsilon}^*,\tilde{u}^*} \le \left(S_{\tilde{\varepsilon}^*} S_{\tilde{u}^*}\right)^{1/2} = O_p(K/n)$ by C.S., Lemma 5 (i) and Lemma 5 (ii).

Proposition 8. $\sqrt{n}S_{x-\tilde{x}^{*},u} \xrightarrow{d} N(0,\ E[\varepsilon_i\varepsilon_i'/\sigma_i^2])$.

Proof. Using the definitions of x and $\tilde{x}^{*}$ gives

$$S_{x^{*}-\tilde{x}^{*},u^{*}} = S_{\varepsilon^{*},u^{*}} + S_{h^{*}-\tilde{h}^{*},u^{*}} - S_{\tilde{\varepsilon}^{*},u^{*}}.$$

1. $\sqrt{n}S_{\varepsilon^{*},u^{*}} = \sum_{i=1}^{n}\left[\varepsilon_i u_i/\sigma_i^2\right]/\sqrt{n} \xrightarrow{d} N\left(0, E\left[\varepsilon_i\varepsilon_i'/\sigma_i^2\right]\right)$ by Lindberg-Levi Central Limit Theorem.

2. $E\left[\left\|S_{h^{*}-\tilde{h}^{*},u^{*}}\right\|^2 \mid Z\right] = n^{-2}tr\left[\left(h^{*}-\tilde{h}^{*}\right)\left(h^{*}-\tilde{h}^{*}\right)' E\left(u^{*}u^{*}{}' \mid Z\right)\right] \le Cn^{-1}tr\left[S_{h^{*}-\tilde{h}^{*}}\right] = o_p\left(n^{-1}\right)$ by C.S. and Lemma 4.

3. $E\left[\left\|S_{\tilde{\varepsilon}^{*},u^{*}}\right\|^2 \mid Z\right] = n^{-2}tr\left[Q\varepsilon^{*}\varepsilon^{*}{}'QE\left(u^{*}u^{*}{}' \mid Z\right)\right] \le Cn^{-1}tr\left[S_{\tilde{\varepsilon}^{*}}\right] = o_p\left(n^{-1}\right)$ by C.S. and Lemma 5 (i).

Combining Propositions 5-8 proves Theorem 3.2.

Appendix B: Tables

Table 1. Simulation Results of the Partial Linear Regression Model Assuming $g(z) = \exp(z)$

	TP = .1		TP = .01		TP = .001	
Estimators [a]	*BIAS* [b]	*MSE*	*BIAS*	*MSE*	*BIAS*	*MSE*
n = 100						
b_{Lin}	0.2410	**0.0921**	0.1714	0.0471	0.1922	0.0601
b_{Li}	**0.2409**	**0.0921**	**0.1713**	**0.0469**	**0.1899**	**0.0584**
b_0	0.2704	0.1180	0.2704	0.1180	0.2704	0.1180
b_K	0.2584	0.1057	0.2259	0.0754	0.2360	0.0814
n = 200						
b_{Lin}	0.1683	0.0443	0.1124	**0.0201**	0.1244	0.0253
b_{Li}	**0.1682**	**0.0442**	**0.1123**	**0.0201**	**0.1240**	**0.0251**
b_0	0.1902	0.0573	0.1902	0.0573	0.1902	0.0573
b_K	0.1796	0.0503	0.1554	0.0351	0.1703	0.0407
n = 400						
b_{Lin}	**0.1202**	**0.0223**	**0.0824**	**0.0109**	**0.0929**	0.0135
b_{Li}	**0.1202**	**0.0223**	0.0825	**0.0109**	**0.0929**	**0.0134**
b_0	0.1369	0.0286	0.1369	0.0286	0.1369	0.0286
b_K	0.1306	0.0256	0.1080	0.0174	0.1245	0.0221

[a] b_{Lin}, b_{Li}, b_0 and b_K represent my estimator, Li's (2000) estimator, the preliminary estimator and the kernel estimator, respectively.

[b] Note that the numbers in boldface represent the minimum of the corresponding BIAS or MSE.

Table 2. Simulation Results of the Partial Linear Regression Model Assuming $g(z) = (1+z)^3$

	TP = .1		TP = .01		TP = .001	
Estimators [a]	*BIAS* [b]	*MSE*	*BIAS*	*MSE*	*BIAS*	*MSE*
$n = 100$						
b_{Lin}	0.2416	0.0923	0.1731	0.0479	0.1944	0.0615
b_{Li}	0.2415	0.0922	0.1719	0.0472	0.1889	0.0578
b_0	0.2712	0.1182	0.2712	0.1182	0.2712	0.1182
b_K	0.4668	0.2866	0.4753	0.2836	0.4753	0.2836
$n = 200$						
b_{Lin}	0.1687	0.0444	0.1135	0.0205	0.1277	0.0267
b_{Li}	0.1686	0.0444	0.1128	0.0203	0.1234	0.0248
b_0	0.1909	0.0575	0.1909	0.0575	0.1909	0.0575
b_K	0.3080	0.1287	0.3141	0.1226	0.3141	0.1226
$n = 400$						
b_{Lin}	0.1203	0.0223	0.0829	0.0110	0.0934	0.0138
b_{Li}	0.1203	0.0223	0.0827	0.0109	0.0921	0.0133
b_0	0.1371	0.0285	0.1371	0.0285	0.1371	0.0285
b_K	0.2142	0.0617	0.2140	0.0570	0.2140	0.0570

[a] b_{Lin}, b_{Li}, b_0 and b_K represent my estimator, Li's (2000) estimator, the preliminary estimator and the kernel estimator, respectively.

[b] Note that the numbers in boldface represent the minimum of the corresponding BIAS or MSE.

Table 3. Choosing smoothing parameters assuming $g(z) = \exp(z)$

Estimators	n = 100		n = 200		n = 400	
	True MSE[a]	*Bootstrap MSE*	*True MSE*	*Bootstrap MSE*	*True MSE*	*Bootstrap MSE*
K_{11}	0.1062	0.1439	0.0519	0.0624	0.0227	0.0265
K_{21}	0.0559	0.1354	0.0297	0.0569	0.0140	0.0244
K_{31}	0.0569	0.1326	0.0296	0.0539	0.0141	0.0238
K_{12}	0.0949	0.1617	0.0413	0.0712	0.0172	0.0316
K_{22}	0.0560	0.1578	0.0288	0.0675	0.0121	0.0288
K_{32}	0.0561	0.1515	0.0284	0.0637	0.0123	0.0278
K_{13}	0.1043	0.2009	0.0481	0.0874	0.0208	0.0392
K_{23}	0.0639	0.1778	0.0313	0.0745	0.0142	0.0319
K_{33}	0.0643	0.1761	0.0313	0.0715	0.0143	0.0314

[a] Note that the numbers in boldface represent the minimum of the corresponding MSE.

◆◆◆