Statistica Sinica **15**(2005), 117-133

BOOTSTRAP OF A SEMIPARAMETRIC PARTIALLY LINEAR MODEL WITH AUTOREGRESSIVE ERRORS

Jinhong You and Xian Zhou

The Hong Kong Polytechnic University

Abstract: This paper is concerned with a semiparametric partially linear regression model with unknown regression coefficients, an unknown nonparametric function for the non-linear component, and unobservable serially correlated random errors. The random errors are modeled by an autoregressive time series. We show that the distributions of the feasible semiparametric generalized least squares estimator of the parametric component, and the estimator of the autoregressive coefficients of the error process, admit bootstrap approximation. Simulation results show that the bootstrap substantially outperforms the normal approximation not only for small to medium sample sizes, but also for highly correlated random errors. A data example is provided to illustrate the method.

Key words and phrases: Asymptotic property, autoregressive error, bootstrap, partially linear regression model, semiparametric least squares estimator.

1. Introduction

Over the last two decades, a great deal of effort has been devoted to the theory and methods of nonparametric regression analysis. When multiple predictor variables are included in the regression equation, however, nonparametric regression faces the so called *curse of dimensionality.* As a result, parametric regression remains popular due to simplicity in computation and interpretation, but a wrong model for the regression function can lead to excessive modeling biases and misleading conclusions. Semiparametric regression models become natural alternatives in such situations, as they can reduce the risk of misspecifying a parametric model while avoiding some drawbacks of fully nonparametric methods. An important semiparametric model is the partially linear regression model introduced by Engle, Granger, Rice and Weiss (1986) to study the effect of weather on electricity demand. It can be written as

$$y_i = x_i'\beta + g(t_i) + \varepsilon_i, \quad i = 1, \ldots, n, \tag{1.1}$$

where y_i's are responses, $x_i = (x_{i1}, \ldots, x_{ip})'$ are design points for linear regression, $\beta = (\beta_1, \ldots, \beta_p)'$ is a vector of unknown parameters to be estimated, $t_i \in [0, 1]$

are additional design points, $g(\cdot)$ is an unknown bounded real-valued function defined on $[0, 1]$, and the ε_i's are unobservable random errors.

The model in (1.1) has been extensively studied. When the errors ε_i's are i.i.d. random variables, various estimation methods have been developed by Heckman (1986), Rice (1986), Chen (1988), Speckman (1988), Robinson (1988), Chen and Shiau (1991, 1994), Donald (1994), Eubank and Speckman (1990) and Hamilton and Truong (1997), among others. These works have used the kernel method, the spline method, series estimation, local linear estimation, two-stage estimation, and so on. They also discussed the asymptotic properties of the estimators. The independence assumption for the errors, however, is not always appropriate in applications, especially for sequentially collected economic data. For example, in the process of fitting the relationship between temperature and electricity usage, Engle et al. (1986) took the data to be autoregressive with order one. Schick (1994) presented an estimator of the autocorrelation coefficient for (1.1) with AR(1) errors. Schick (1996, 1998) went further to construct efficient estimators of β and the autocorrelation coefficient. Gao (1995a) studied the estimation problem for (1.1) with serially correlated errors. A semiparametric least squares estimator (SLSE) for the parametric component β was proposed and its asymptotic properties were discussed. You and Chen (2002a) constructed a semiparametric generalized least squares estimator (SGLSE) for the parametric component of (1.1) with autoregressive errors. However, due to the nonparametric component $g(\cdot)$, the SGLSE is biased even if the error structure is known. The bias prevents this estimator from attaining the optimal Berry-Essen rate $n^{-1/2}$. Actually, the best normal approximation rate is only $n^{-1/5}$ (cf. Hong (2002)) under Assumptions 2.1 and 2.2 of Section 2 below. In this situation, the bootstrap becomes an attractive alternative. As a matter of fact, the application of the bootstrap to (1.1) has recently attracted attention in the literature. For example, Hong and Cheng (1993) considered bootstrap approximation of the estimators for the parameters in model (1.1). In their case $\{x_i', t_i, \varepsilon_i, i = 1, \ldots, n\}$ are i.i.d. random variables and $g(\cdot)$ is estimated by a kernel smoother. Liang, Härdle and Sommerfeld (2000) explained the advantage of the bootstrap method and constructed bootstrap statistics for parameters β and $\sigma^2 = \mathrm{Var}(\varepsilon_1)$, and studied their asymptotic normality when the (x_i', t_i) are known design points, ε_i are i.i.d. random variables and $g(\cdot)$ is estimated by a general nonparametric fitting. You and Chen (2002b) investigated a wild bootstrap approximation and showed it robust against heteroscedasticity. However, the bootstrap schemes mentioned above did not consider possible serial correlation in the errors.

In this paper, by fitting the error structure and generating time series replicates we devise an alternative bootstrap method that accounts for autoregressive errors, and apply it to (1.1). We show that the distributions of the feasible SGLSE of β and the estimator of the autoregressive coefficient of the error process admit this bootstrap approximation. The same problem was studied by Stute (1995) and Vilar-Fernández and Vilar-Fernández (2002) in the context of traditional linear regression models, but they did not consider the bootstrap approximation for the autoregressive coefficients. We do this here.

To evaluate the performance of our method, a small simulation is conducted. The results show that the bootstrap outperforms the normal approximation, substantially, when the random errors are heavily correlated. Furthermore, application to a data set is provided to illustrate the method.

The rest of this paper is organized as follows. Section 2 presents the feasible SGLSE of the regression coefficients β. The bootstrap methodology, along with its asymptotic properties, are discussed in Section 3. Section 4 reports the results of a small simulation and an application. Proofs of the main results are given in Section 5, followed by conclusions in Section 6.

2. Feasible SGLSE of the Parametric Component

Throughout this paper we assume that the design points x_i and t_i are fixed, $i = 1, \ldots, n$. In addition, suppose that the vector $(1, \ldots, 1)'$ is not in the space spanned by the column vectors of $X = (x_1, \ldots, x_n)'$, which ensures the identifiability of the model in (1.1) according to Chen (1988). For convenience we assume that the errors arise from a stationary autoregressive sequence with order one, namely,

$$\varepsilon_i = \rho\varepsilon_{i-1} + e_i, \quad i = 1, \ldots, n, \tag{2.1}$$

where ρ is unknown (with $|\rho| < 1$) and the e_i's are i.i.d. following distribution $F(\cdot)$ with zero mean and finite variance σ_e^2. Extension of our results to a more general autoregressive model is conceptually straightforward.

There are several methods to construct a feasible SGLSE for β, including the partial spline method, the partial kernel method, the series approximation and so on. Here we adopt the partial kernel method proposed by Speckman (1988) because, in comparison with the partial spline method, it does not need to underestimate the nonparametric component in order to obtain a $\sqrt{n}$-consistent estimator of the parametric component. Details of the partial kernel method are summarized below.

Assume (1.1). If β is known to be the true parameter then, as $E(\varepsilon_i) = 0$, we have $g(t_i) = E(y_i - x_i'\beta)$ for $i = 1, \ldots, n$. Hence, a natural nonparametric estimator of $g(\cdot)$ given β is $\tilde{g}(t, \beta) = \sum_{i=1}^n W_{ni}(t)(y_i - x_i'\beta)$, where $W_{ni}(\cdot)$ are the weight functions satisfying Assumption 2.1 below. Substituting $\tilde{g}(t,\beta)$ into (1.1) gives $y_i = x_i'\beta + \tilde{g}(t,\beta) + \varepsilon_i$, which can be written as

$$\hat{y}_i = \hat{x}_i'\beta + \tilde{\varepsilon}_i, \quad i = 1, \dots, n, \tag{2.2}$$

where $\hat{y}_i = y_i - \sum_{j=1}^n W_{nj}(t_i)y_j$, $\hat{x}_i = x_i - \sum_{j=1}^n W_{nj}(t_i)x_j$ and $\tilde{\varepsilon}_i = g(t_i) - \sum_{j=1}^n W_{nj}(t_i)g(t_j) + \varepsilon_i - \sum_{j=1}^n W_{nj}(t_i)\varepsilon_j$. Write (2.2) in matrix form as

$$\hat{Y} = \hat{X}\beta + \tilde{\varepsilon}, \tag{2.3}$$

where $\hat{Y} = (\hat{y}_1, \dots, \hat{y}_n)'$, $\hat{X} = (\hat{x}_1, \dots, \hat{x}_n)'$ and $\tilde{\varepsilon} = (\tilde{\varepsilon}_1, \dots, \tilde{\varepsilon}_n)'$. Regard $\tilde{\varepsilon}_i$ as new random errors. Then by (2.3) a semiparametric least square estimator (SLSE) of β is given by $\hat{\beta}_n = (\hat{X}'\hat{X})^{-1}\hat{X}'\hat{Y}$. Since the errors are serially correlated in (1.1), the SLSE $\hat{\beta}_n$ is not asymptotically efficient. To overcome this problem, we consider weighted estimation. First we fit the error structure. Based on $\hat{\beta}_n$, the estimated residuals can be obtained as

$$\hat{\varepsilon}_i = \hat{y}_i - \hat{x}_i'\hat{\beta}_n, \quad i = 1, \dots, n. \tag{2.4}$$

Thus we estimate the autoregressive coefficient ρ by $\hat{\rho}_n = \left(\sum_{i=1}^n \hat{\varepsilon}_i^2\right)^{-1} \sum_{i=1}^{n-1} \hat{\varepsilon}_{i+1}\hat{\varepsilon}_i$.

By (2.1) we have

$$E(\varepsilon\varepsilon') = \sigma^2 \begin{bmatrix} 1 & -\rho & 0 & 0 & \cdots & 0 \\ -\rho & 1+\rho^2 & -\rho & 0 & \cdots & 0 \\ 0 & -\rho & 1+\rho^2 & -\rho & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & -\rho & 1+\rho^2 & -\rho \\ 0 & 0 & \cdots & 0 & -\rho & 1 \end{bmatrix}^{-1} \hat{=} \sigma^2\Omega, \tag{2.5}$$

where $\sigma^2 = E(\varepsilon_1^2)$. According to Lemmas 5.1 and 5.2 in Section 5, $\tilde{\varepsilon}_i = \varepsilon_i + o(1)$ a.s. Therefore, a feasible SGLSE is given by $\hat{\beta}_n^w = (\hat{X}'\hat{\Omega}^{-1}\hat{X})^{-1}\hat{X}'\hat{\Omega}^{-1}\hat{Y}$, where $\hat{\Omega}$ is similar to Ω in (2.5) with $\hat{\rho}_n$ in place of ρ.

We now state some assumptions required to obtain the asymptotic property of $\hat{\beta}_n^w$. These assumptions, while a bit lengthy, are actually quite mild and can be easily satisfied. First suppose, as is common in the setting of partially linear regression model, that $\{x_i\}$ and $\{t_i\}$ are related via $x_{is} = h_s(t_i) + u_{is}$, $i = 1, \dots, n$; $s = 1, \dots, p$. Justification for this can be found in Speckman (1988). Consider the following assumptions.

Assumption 2.1. The probability weight functions $W_{ni}(\cdot)$ satisfy

(i) $\max_{1\le i\le n} \sum_{j=1}^n W_{ni}(t_j) = O(1)$,

(ii) $\max_{1\le i,j\le n} W_{ni}(t_j) = O(b_n)$,

(iii) $\max_{1\le j\le n} \sum_{i=1}^{n} W_{ni}(t_j) I(|t_j - t_i| > c_n) = O(d_n)$, where $b_n = O(n^{-2/3})$, $\limsup_{n\to\infty} nc_n^3 < \infty$, $\limsup_{n\to\infty} nd_n^3 < \infty$, and $I(A)$ is the indicator function of a set A.

Assumption 2.2. $\max_{1\le i\le n} \|\sum_{j=1}^{n} W_{nj}(t_i)u_j\| = o(n^{-1/6})$, $\|u_j\| \le c$ and $\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} u_i \Omega^{-1} u_i' = \Sigma > 0$, where $\|\cdot\|$ denotes the Euclidean norm, and c is a constant.

Assumption 2.3. The functions $g(\cdot)$ and $h_1(\cdot), \ldots, h_p(\cdot)$ satisfy a Lipschitz condition of order 1 on $[0,1]$.

Remark 2.1. Partition the interval $[0,1]$ by $0 = t_0 \le t_1 \le \cdots \le t_n \le t_{n+1} = 1$ such that $\max_{1\le i\le n+1} |t_i - t_{i-1}| = O(n^{-1})$. Take $W_{nj}(\cdot)$ to be the k_n-nearest neighbor type weight functions, namely $W_{nj}(t) = k_n^{-1}$ if t_j belongs to the k_n-nearest neighbor of t and $W_{nj}(t) = 0$ otherwise, where $k_n = n^{2/3}$. Then $W_{nj}(\cdot)$ satisfies Assumption 2.1.

Remark 2.2. Note that $\sum_{j=1}^{n} W_{nj}(t_i)u_j$ is a weighted average of the locally centered quantities $\{u_j\}_{j=1}^{n}$. Hence $\max_{1\le i\le n} \|\sum_{j=1}^{n} W_{nj}(t_i)u_j\| = o(n^{-1/6})$ is a mild condition.

Remark 2.3. Assumption 2.3 is mild and holds for most commonly used functions, polynomial and trigonometric functions, for example.

Under Assumptions 2.1–2.3 and $E(e_1^4) < \infty$, You and Chen (2002a) proved that

$$\sqrt{n}(\hat{\beta}_n^w - \beta) \to_D N(0, \sigma^2\Sigma^{-1}) \quad \text{and} \quad \sqrt{n}(\hat{\rho}_n - \rho) \to_D N(0, 1-\rho^2), \tag{2.6}$$

where "$\to_D$" denotes convergence in distribution. In addition, they also showed that $\hat{\beta}_n^w$ is asymptotically efficient. In the following sections we show that the distributions of $\hat{\beta}_n^w$ and $\hat{\rho}_n$ can be approximated by a bootstrap procedure.

3. Bootstrap Methodology

We begin with describing our bootstrap.

Step 1. Given the initial sample (x_i', t_i, y_i), $i = 1, \ldots, n$, we construct a SGLSE $\hat{\beta}_n^w$ by the method described in Section 2.

Step 2. The estimated residuals $\hat{\varepsilon}_i^w = \hat{y}_i - \hat{x}_i'\hat{\beta}_n^w$ are evaluated and the noise of the AR(1) model is obtained as $\hat{e}_i^w = \hat{\varepsilon}_i^w - \hat{\rho}_n \hat{\varepsilon}_{i-1}^w$ for $i = 2, \ldots, n$.

Step 3. Depending on the assumptions imposed on the error distribution F, and for a sufficiently large number N, a random sample of $e_i^\star$ for $-N \le$

$i \leq n$ is drawn from an estimator of F. For instance, if F is assumed to belong to a finite dimensional parametric family of distributions, i.e., $F \in \{F(\cdot;\alpha)\}$ with parameter α, then $\{e_i^\star\}$ can be obtained as an i.i.d. sequence from $F(\cdot;\hat{\alpha})$, where $\hat{\alpha}$ is a consistent estimator of α. Alternatively, in the absence of any a priori assumptions on the distribution of the true errors, the series $e_i^\star$ can be obtained as a random sample from the empirical distribution function $\hat{F}_n$ of F which puts mass $(n-p)^{-1}$ on each of the centered residual vectors $\hat{e}_i$.

Step 4. Using this noise series, bootstrap replicates $\{\varepsilon_i^\star, i = 1, \ldots, n\}$ are obtained as $\varepsilon_i^\star = \sum_{j=0}^{\infty} \hat{\rho}_n^j e_{i-j}^\star$. In practice, an initial value $\varepsilon_i^\star = \sum_{j=0}^{\infty} \hat{\rho}_n^j e_{i-j}^\star$ is computed and then the equation $\varepsilon_i^\star = \hat{\rho}_n \varepsilon_{i-1}^\star + e_i^\star$ is iteratively applied to obtain the values $\varepsilon_i^\star$ for $i = 1, \ldots, n$. In calculations, we can approximate $\varepsilon_i^\star$ by $\sum_{j=0}^{N} \hat{\rho}_n^j e_{i-j}^\star$ for some large N.

Step 5. The bootstrap sample $y_i^\star$ is obtained by means of $y_i^\star = x_i'\hat{\beta}_n^w + \hat{g}_n^w(t_i) + \varepsilon_i^\star$ for $i = 1, \ldots, n$, where $\hat{g}_n^w(t_i) = \sum_{j=1}^{n} W_{nj}(t_i)(y_j - x_j'\hat{\beta}_n^w)$.

Step 6. The SGLSE is computed with the bootstrap sample. Thus $\hat{\beta}_n^{\star w} = (\hat{X}'\hat{\Omega}^{\star-1}\hat{X})^{-1}\ \hat{X}'\hat{\Omega}^{\star-1}\hat{Y}^\star$, where $\hat{Y}^\star$ and $\hat{\Omega}^{\star-1}$ are analogous to $\hat{Y}$ and $\hat{\Omega}$, with $\hat{Y}^\star$ from the bootstrap sample $\{y_i^\star,\ i = 1, \ldots, n\}$ and $\hat{\Omega}^{\star-1}$ using $\rho_n^\star = \sum_{i=1}^{n-1} \varepsilon_i^\star \varepsilon_{i+1}^\star \,/ \sum_{i=1}^{n} \varepsilon_i^{\star 2}$ as the estimator of ρ.

Step 7. For a large value of M, Step 3-6 are repeated M times to obtain the bootstrap replications of the estimators $\{\rho_{n,1}^\star, \ldots, \rho_{n,M}^\star\}$ and $\{\hat{\beta}_{n,1}^{\star w}, \ldots, \hat{\beta}_{n,M}^{\star w}\}$.

Remark 3.1. The same bootstrap method was used in Stute (1995) and Vilar-Fernández and Vilar-Fernández (2002). However, they deal only with traditional linear regression models.

For $\hat{\beta}_n^{\star w}$ and $\hat{\rho}_n^\star$ we have the following asymptotic results.

Theorem 3.1. *Suppose that Assumptions* 2.1 *to* 2.3 *hold and* $E(e_1^4) < \infty$. *Then* $\sqrt{n}(\hat{\beta}_n^{\star w} - \hat{\beta}_n^w) \to_{D^\star} N(0, \sigma^2\Sigma^{-1})$ *as* $n \to \infty$, *where* $\to_{D^\star}$ *denotes the convergence in distribution underlying the bootstrap samples.*

Theorem 3.2. *Under the assumptions of Theorem* 3.1, $\sqrt{n}(\hat{\rho}_n^\star - \hat{\rho}_n) \to_{D^\star} N(0, 1-\rho^2)$ *as* $n \to \infty$.

Combining Theorems 3.1 and 3.2 with the asymptotic normality of $\hat{\beta}_n^w$ and $\hat{\rho}_n$ (see (2.6)), leads to the following corollary.

Corollary 3.1. *Under the assumptions of Theorem* 3.1,

(i) $\sup_{x \in \mathcal{R}^p} \left| P^\star(\sqrt{n}(\hat{\beta}_n^{\star w} - \hat{\beta}_n) \leq x) - P(\sqrt{n}(\hat{\beta}_n^w - \beta) \leq x) \right| \to_p 0$ *as* $n \to \infty$, *where $P^\star$ denotes the probability distribution under the resampling and $\to_p$ stands for convergence in probability;*

(ii) $\sup_{x\in\mathcal{R}} |P^\star(\sqrt{n}(\hat\rho_n^\star - \hat\rho_n) \le x) - P(\sqrt{n}(\hat\rho_n - \rho) \le x)| \to_p 0$ *as* $n \to \infty$.

The above results can be used to construct large sample confidence intervals for β or ρ. For instance, a $100(1-\alpha)\%$ two-sided confidence interval for $a'\beta$ is

$$\left[a'\hat\beta_n^w - \frac{1}{\sqrt{n}} z(1-\alpha/2), a'\hat\beta_n^w - \frac{1}{\sqrt{n}} z(\alpha/2)\right],$$

where a is a nonzero constant p-vector and $P^\star(\sqrt{n}(a'\hat\beta_n^{\star w} - a'\hat\beta_n^w) \le z(\alpha)) = \alpha$.

It should be noted that although our results are established under the assumption that the nonparametric regressor is one-dimensional, it is not difficult to extend them to the multi-dimensional nonparametric regressor case using product kernels.

4. Simulation Studies and an Application

In this section we carry out some simulation studies to compare the bootstrap approach with a normal approximation. We also illustrate this method via its application to a set of spirit consumption data.

4.1. The finite sample performance

We investigated the model $y_i = x_i\beta + g(t_i) + \varepsilon_i$ with $\varepsilon_i = \rho\varepsilon_{i-1} + e_i$, $i = 1, \ldots, n$, where $g(t_i) = \sin(2\pi t_i)$, $\beta = 5$. Two forms of the error distribution F were tested: the standard normal distribution $N(0,1)$ and the uniform distribution $U[-1,1]$. For the autoregressive coefficient ρ, we considered three cases: $\rho = 0.3$, $\rho = 0.5$ and $\rho = 0.8$, to reflect different levels of serial correlation in the errors. The independent variables x_i and t_i are generated from the $U[0,1]$ distribution.

For comparison, we calculated the confidence intervals of β and ρ based on simulated samples using both normal approximation and the bootstrap. We then compare the coverage percentages of the confidence intervals, with sample sizes ranging from 50 to 500.

The confidence intervals are constructed as follows. According to (2.6),

$$D_n = (\hat\beta_n^w - \beta)\hat\sigma_n^{-1}(\hat X'\hat\Omega^{-1}\hat X)^{1/2} \to_D N(0,1) \quad \text{as } n \to \infty, \tag{4.1}$$

where $\hat\sigma_n^2$ is the sample variance computed from the residuals $\hat\varepsilon_i$ given in (2.4). Therefore, (4.1) may be utilized to construct confidence intervals for β. Likewise, Theorem 3.1 and the consistency results in Section 5 imply that the actual distribution of D_n may be approximated by that of $D_n^\star = (\hat\beta_n^{\star w} - \hat\beta_n^w)\hat\sigma_n^{\star -1}(\hat X'\hat\Omega^{\star -1}$

$\hat{X})^{1/2}$. Thus, the confidence intervals for β based on normal approximation and the bootstrap are, respectively,

$$I_n^N(\beta) = \left[\hat{\beta}_n^w - \hat{\sigma}_n(\hat{X}'\hat{\Omega}^{-1}\hat{X})^{-1/2}u_N,\ \hat{\beta}_n^w - \hat{\sigma}_n(\hat{X}'\hat{\Omega}^{-1}\hat{X})^{-1/2}l_N\right],$$
$$I_n^B(\beta) = \left[\hat{\beta}_n^w - \hat{\sigma}_n(\hat{X}'\hat{\Omega}^{-1}\hat{X})^{-1/2}u_B(\beta),\ \hat{\beta}_n^w - \hat{\sigma}_n(\hat{X}'\hat{\Omega}^{-1}\hat{X})^{-1/2}l_B(\beta)\right],$$

where u_N, l_N and $u_B(\beta), l_B(\beta)$ denote the $1-\alpha/2$ and $\alpha/2$ quantiles computed from $N(0,1)$ and the bootstrap distribution, respectively.

Correspondingly, the confidence intervals for ρ based on normal approximation and bootstrap are, respectively,

$$I_n^N(\rho) = \left[\hat{\rho}_n - n^{-1/2}\sqrt{1-\hat{\rho}_n^2}u_N, \hat{\rho}_n - n^{-1/2}\sqrt{1-\hat{\rho}_n^2}l_N\right],$$
$$I_n^B(\rho) = \left[\hat{\rho}_n - n^{-1/2}\sqrt{1-\hat{\rho}_n^2}u_B(\rho), \hat{\rho}_n - n^{-1/2}\sqrt{1-\hat{\rho}_n^2}l_B(\rho)\right],$$

where u_N, l_N and $u_B(\rho), l_B(\rho)$ have the same definitions as those for β.

Samples of sizes $n = 50, 100$, 200 and 500 were drawn repeatedly. In each case the number of simulated realizations was $10,000$, as was the number of bootstrap replicates (the x_i values are generated once for each n). For each simulated realization, the bootstrap distributions are computed individually. For the weight function $W_{ni}(t_j)$, we use the Priestley and Chao's weight with a Gaussian kernel:

$$W_{ni}(t_j) = \frac{1}{nh}K\left(\frac{t_i - t_j}{h}\right) = \frac{1}{nh}\frac{1}{\sqrt{2\pi}}e^{-\frac{(t_i-t_j)^2}{2h^2}}.$$

The bandwidth is selected by Cross-Validation (CV). The coverage percentages by 90% confidence intervals from simulation are listed in Tables 1 and 2 below.

From Tables 1 and 2 we can see that the bootstrap has higher coverage percentage than the normal approximation in almost every scenario considered. The difference is more significant for smaller n and larger ρ. This shows that the bootstrap outperforms the normal approximation not only for small to medium sample sizes, which is usually a main reason to use bootstrap, but also for high serial correlation and hence is advantageous when the data exhibit serial dependence.

Table 1. The actual coverage percentages obatined for β and ρ by normal approximation (N) and bootstrap (B) when F is $N(0,1)$.

	$\rho = 0.3$				$\rho = 0.5$				$\rho = 0.9$			
	β		ρ		β		ρ		β		ρ	
	N	B	N	B	N	B	N	B	N	B	N	B
$n = 50$	84%	88%	63%	76%	80%	86%	64%	82%	72%	83%	59%	82%
$n = 100$	83%	92%	75%	80%	82%	89%	79%	87%	74%	86%	70%	86%
$n = 200$	86%	90%	84%	91%	86%	91%	82%	88%	80%	90%	80%	87%
$n = 500$	88%	91%	89%	93%	87%	91%	86%	90%	85%	93%	87%	90%

Table 2. The actual coverage percentages obatined for β and ρ by normal approximation (N) and bootstrap (B) when F is $U[-1,1]$.

	$\rho=0.3$				$\rho=0.5$				$\rho=0.9$			
	β		ρ		β		ρ		β		ρ	
	N	B	N	B	N	B	N	B	N	B	N	B
$n=50$	86%	89%	76%	85%	85%	86%	73%	85%	82%	88%	69%	84%
$n=100$	86%	93%	79%	86%	84%	88%	77%	88%	81%	89%	75%	88%
$n=200$	88%	91%	84%	88%	87%	90%	85%	89%	86%	93%	82%	86%
$n=500$	91%	90%	89%	92%	89%	93%	90%	91%	91%	90%	88%	91%

4.2. Application to spirit consumption data

We now illustrate the methodology via its application to the spirit consumption data in the United Kingdom from 1870 to 1938. The data set can be found on page 427 of Fuller (1976). In this data set, the dependent variable y_i is the annual per capita consumption of spirits in the United Kingdom. The explanatory variables x_{i1} and x_{i2} are per capita income and price of spirits, respectively, both deflated by a general price index. All data are in logarithms. Fuller (1976) used the following linear regression model to fit this data set, where 1,869 is the origin for t and ε_i is assumed to be a stationary time series:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 t_i + \beta_4 (t_i - 35)^2 + \varepsilon_i. \tag{4.2}$$

Least squares (LS) regression is $\hat{y}_i = 2.1373 + 0.6808 x_{i1} - 0.6333 x_{i2} - 0.0095 t_i - 0.00011(t_i - 35)^2$. The residual mean square is 9.2204×10^{-4}. By the Durbin-Watson d test, Fuller (1976) took the errors to be autocorrelated with order one with autocoefficient 0.7633. Then, based on this estimated autocoefficient he applied the weighted least squares (WLS) regression to fit model (4.2) to obtain $\hat{y}_i = 2.3579 + 0.7091 x_{i1} - 0.7836 x_{i2} - 0.0081 t_i - 0.00012(t_i - 35)^2$.

We now relax (4.2) to a semiparametric partially linear regression model

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + g(t_i) + \varepsilon_i, \tag{4.3}$$

where $g(\cdot)$ is an unknown function of t. The semiparametric least squares (SLS) regression gives $\hat{y}_i = 0.6463 x_{i1} - 0.9535 x_{i2} - \hat{g}(t_i)$, where $\hat{g}(t_i)$ is shown in Figure 1. The residual mean square is 2.1914×10^{-4}, only about one quarter of the residual mean square from the traditional linear regression model. Moreover, from Figure 2 we can see that at almost every point the fitted residual of the former is less than that of the latter. This illustrates that (4.3) is a more suitable model here than is (4.2). The autocoefficient for (4.3) is 0.2074. Figure 3 shows the residual plot after fitting the AR(1) process to $\{\varepsilon_i\}$ in (4.3), which is not significantly different from white noise, validating the choice of AR(1) for (4.3). The semiparametric generalized least squares (SGLS) regression gives $\hat{y}_i = 0.6821x_{i1} - 0.9394x_{i2} - \hat{g}(t_i)$, and the bootstrap estimation gives $\hat{y}_i = 0.6402x_{i1} - 0.9549x_{i2} - \hat{g}(t_i)$.

More details of the data analysis can be found in Table 3 below, where the LSE and WLSE correspond to (4.2), while the SLSE, SGLSE and bootstrap estimates are from (4.3).

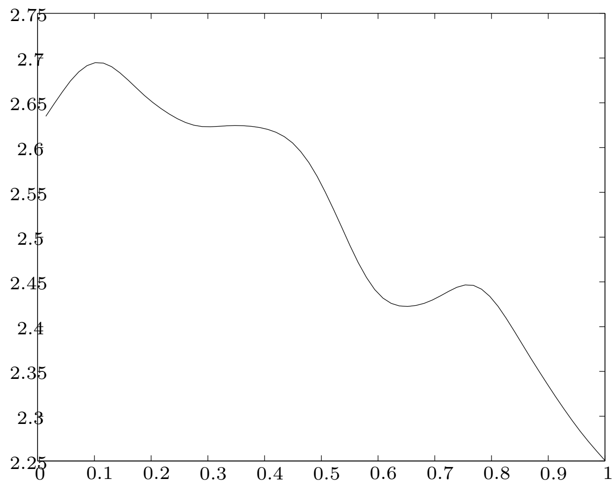

Figure 1. The estimated nonparametric component $g(\cdot)$ in partially linear regression model (4.3) with the spirit consumption data.

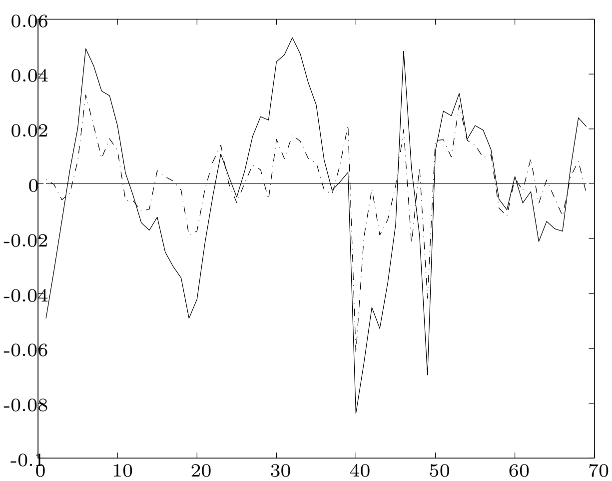

Figure 2. The fitted residuals by linear regression model (4.2) (——) and partially linear regression model (4.3) ($- \cdot - \cdot -$) for the spirit consumption data.

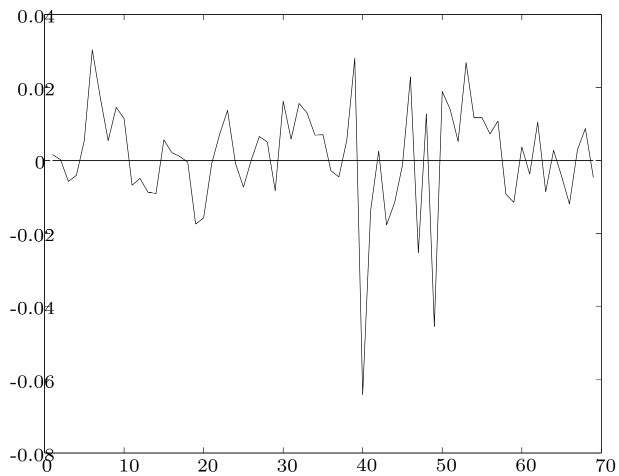

Figure 3. Residuals after fitting AR(1) to the errors in partially linear regression model (4.3) with the spirit consumption data.

Table 3. The fitting of the spirit consumption data.

	Mean	Standard Error	95% Confidence Interval
LSE(β_1)	0.6808	0.2387	(0.6245, 0.7371)
LSE(β_2)	-0.6333	0.1025	(-0.6575, -0.6091)
LSE(ρ)	0.7153	0.0841	(0.6955, 0.7351)
WLSE(β_1)	0.7091	0.0742	(0.6916, 0.7266)
WLSE(β_2)	-0.7836	0.0283	(-0.7769, -0.7903)
SLSE(β_1)	0.6463	0.1245	(0.6169, 0.6757)
SLSE(β_2)	-0.9535	0.0678	(-0.9695, -0.9375)
SLSE(ρ)	0.2074	0.1072	(0.1821, 0.2327)
SGLSE(β_1)	0.6821	0.1122	(0.6556, 0.7086)
SGLSE(β_2)	-0.9394	0.0592	(-0.9534, -0.9254)
Bootstrap(β_1)	0.6402	0.1232	(0.4939, 0.7795)
Bootstrap(β_2)	-0.9549	0.0670	(-1.0841, -0.8105)
Bootstrap(ρ)	0.2072	0.1136	(0.0170, 0.3274)

5. Proofs of Main Results

To prove the main results, we first introduce several lemmas. The first lemma can be found in Gao (1995).

Lemma 5.1. (i) *Suppose that Assumptions* 2.1 (iii) *and* 2.3 *hold. Then as* $n \to \infty$, $\max_{0\le s\le p} \max_{1\le i\le n} |G_s(t_i) - \sum_{j=1}^n W_{nj}(t_i)G_s(t_j)| = O(c_n) + O(d_n)$, *where* $G_0(\cdot) = g(\cdot)$ *and* $G_s(\cdot) = h_s(\cdot)$, $s = 1, \ldots, p$;
(ii) *Under Assumptions* 2.1 *to* 2.3, *as* $n \to \infty$, $\max_{1\le s\le p} \max_{1\le i\le n} |\sum_{j=1}^n W_{nj}(t_i) x_{js} - h_s(t_i)| = O(c_n) + O(d_n) + o(n^{-1/6})$.

Lemma 5.2. *Suppose that $\{\varepsilon_i\}$ is a linear process so $\varepsilon_i = \sum_{j=0}^{\infty} \psi_j e_{i-j}$ with $\sup_n n \sum_{j=n}^{\infty} |\psi_j| = O(1)$, where e_i's are i.i.d. with mean 0 and $E(e_1^4) < \infty$, and suppose that Assumption 2.1 holds. If the spectral density $f(\omega)$ of $\{\varepsilon_i\}$ satisfies $c_1 \leq f(\omega) \leq c_2$ for all $\omega \in (-\pi, \pi]$, where c_1 and c_2 are positive constants, then $\max_{1\leq i\leq n}|\sum_{j=1}^{n} W_{nj}(t_i)\varepsilon_j| = O(n^{-1/3}\log n)$ a.s.*

Proof. By separating the MA(∞) error process into two parts, a procedure widely applied in time series, similar to the proof of Lemma A.3 in Härdle, *et al.* (2000), we can prove Lemma 5.2. The details can be found in You (2002).

Lemma 5.3. *Suppose that Assumptions 2.1 to 2.3 hold and $E(e_1^4) < \infty$. Then as $n \to \infty$, $|\hat{\beta}_{ni}^w - \beta_i| = O[(\log\log n/n)^{1/2}]$ a.s. and $|\hat{\rho}_n - \rho| = O[(\log\log n/n)^{1/2}]$ a.s., where $\hat{\beta}_{ni}^w$ and β_i denote the i-th components of $\hat{\beta}_n^w$ and β respectively.*

The proof of Lemma 5.3 can be found in You and Chen (2002a).

Now we define a d_2 metric (also called Mallow's metric or Wasserstein distance) for probability measures P and Q with $\int |x|^2 dP < \infty$ and $\int |x|^2 dQ < \infty$, $d_2(P,Q) = \inf(E|X-Y|^2)^{1/2}$, where the inf is taken over pairs (X,Y) of random variables, with X and Y distributed according to P and Q respectively. The following lemma shows that the empirical distribution function $\hat{F}_n$, obtained in Step 3 of Section 3, converges to the distribution F of the noise process $\{e_i\}$ in terms of the Mallow's metric.

Lemma 5.4. *Under the assumptions of Lemma 5.3, $d_2(\hat{F}_n, F) \to 0$ a.s. as $n \to \infty$.*

Proof. Denote by $F_n^\star$ the empirical distribution function based on the error series $\{\tilde{e}_i\}_{i=2}^n$, which is obtained in the same way as $\hat{e}_i^w$ (Steps 2 and 3 in Section 3), but from the unobservable residuals $\{\varepsilon_i\}_{i=1}^n$. Then we have $d_2(\hat{F}_n, F) \leq d_2(\hat{F}_n, F_n^\star) + d_2(F_n^\star, F)$. By Theorem 3.1 in Kreiss and Franke (1992), $d_2(F_n^\star, F) \to 0$ as $n \to \infty$ a.s. Hence it suffices to show that $d_2(\hat{F}_n, F_n^\star) \to 0$ a.s., as we do below.

$$
\begin{aligned}
d_2^2(\hat{F}_n, F_n^\star) &\leq \frac{1}{n-1}\sum_{i=2}^{n}(\hat{e}_i^w - e_i^\star)^2 = \frac{1}{n-1}\sum_{i=2}^{n}\left(\hat{\varepsilon}_i^w - \hat{\rho}_n\hat{\varepsilon}_{i-1}^w - \varepsilon_i + \rho\varepsilon_{i-1}\right)^2 \\
&\leq \frac{2}{n-1}\sum_{i=2}^{n}(\hat{\varepsilon}_i^w - \varepsilon_i)^2 + \frac{4}{n-1}\sum_{i=2}^{n}(\rho - \hat{\rho}_n)^2\varepsilon_{i-1}^2 \\
&\quad + \frac{4}{n-1}\sum_{i=2}^{n}\hat{\rho}_n^2(\hat{\varepsilon}_{i-1}^w - \varepsilon_{i-1})^2 = I_1 + I_2 + I_3 \quad \text{say.}
\end{aligned}
$$

According to Lemmas 5.1, 5.2 and 5.3, and the fact that

$$
\begin{aligned}
\sum_{i=1}^{n}\hat{x}_i\hat{x}_i' &= \sum_{i=1}^{n}\Big(u_i - \sum_{j=1}^{n}W_{nj}(t_i)u_j + \tilde{H}(t_i)\Big)\sum_{i=1}^{n}\Big(u_i - \sum_{j=1}^{n}W_{nj}(t_i)u_j + \tilde{H}(t_i)\Big)' \\
&= O(n),
\end{aligned}
$$

where $\tilde{H}(t_i) = (\tilde{h}_1(t_i), \ldots, \tilde{h}_p(t_i))$ with $\tilde{h}_s(t_i) = h_s(t_i) - \sum_{j=1}^n W_{nj}(t_i)h_s(t_j)$, we have

$$I_1 \le \frac{4}{n-1}\sum_{i=2}^n (\hat{\beta}_n^w - \beta)'\hat{x}_i\hat{x}_i'(\hat{\beta}_n^w - \beta) + \frac{4}{n-1}\sum_{i=2}^n \tilde{g}^2(t_i) + \frac{4}{n-1}\sum_{i=2}^n \tilde{\varepsilon}_i^2 = o(1) \text{ a.s.},$$

where $\tilde{g}(t_i) = g(t_i) - \sum_{j=1}^n W_{nj}(t_i)g(t_j)$ and $\tilde{\varepsilon}_i = \sum_{j=1}^n W_{nj}(t_i)\varepsilon_j$. Moreover, it is easy to see from Lemma 5.3 that $I_2 = o(1)$ a.s. Along the same lines we can show that

$$I_3 \le \frac{4}{n-1}\sum_{i=2}^n (\hat{\rho}_n^2 - \rho^2)(\hat{\varepsilon}_{i-1}^w - \varepsilon_{i-1})^2 + \frac{4}{n-1}\sum_{i=2}^n \rho^2(\hat{\varepsilon}_{i-1}^w - \varepsilon_{i-1})^2 = o(1) \text{ a.s..}$$

The proof is thus complete.

Lemma 5.5. *Under the assumptions of Lemma* 5.3, $\hat{\rho}_n^\star - \hat{\rho}_n \to 0$ *a.s. as* $n \to \infty$, *where* $\hat{\rho}_n^\star$ *is defined in Section* 3.

Proof. The proof is similar to that of Lemma 4.2 in Vilar-Fernández and Vilar-Fernández (2002).

Lemma 5.6. *Under the assumptions of Lemma* 5.3, $n^{-1/2}\hat{X}'\varepsilon^\star = O_p(1)$, $n^{-1}\hat{X}'(\hat{\Omega}^{\star-1} - \hat{\Omega}^{-1})\hat{X} \to_p 0$ *and* $n^{-1}\hat{X}'(\hat{\Omega}^{\star-1} - \hat{\Omega}^{-1})\varepsilon^\star \to_p 0$.

Proof. The proof of Lemma 5.6 is similar to that of Theorem 3.1 in You and Chen (2002a), but this time for the bootstrap variables. We omit the details.

Lemma 5.7. *Under the assumptions of Lemma* 5.3, $n^{-1}\sum_{i=1}^{n-1}(\hat{\varepsilon}_i - \varepsilon_i)\varepsilon_{i+1} = o(n^{-1/2})$ *and* $n^{-1}\sum_{i=1}^{n-1}(\hat{\varepsilon}_i - \varepsilon_i)(\hat{\varepsilon}_{i+1} - \varepsilon_{i+1}) = o(n^{-1/2})$ *a.s., where* $\hat{\varepsilon}_i = \hat{y}_i - \hat{x}_i'\hat{\beta}_n$, $i = 1, \ldots, n$.

Proof. The proof is similar to those of Lemmas 5.4 and 5.5 in You and Chen (2002a).

Lemma 5.8. *Under the assumptions of Lemma* 5.3, $|E_\star(\varepsilon_i^{\star 2}) - E(\varepsilon_1^2)| \to_p 0$ *and* $|n^{-1}\sum_{i=1}^n \varepsilon_i^{\star 2} - E(\varepsilon_1^2)| \to_p 0$ *as* $n \to \infty$, *where* $E_\star$ *denotes the expectation under the resampling.*

Proof. Combining Lemmas 5.1 to 5.3 we obtain

$$\begin{aligned}
|E_\star(e_i^{\star 2}) - \sigma_e^2| &= \left| n^{-1}\sum_{i=1}^n \hat{e}_i^{w2} - \sigma_e^2 \right| \le n^{-1}\sum_{i=1}^n |\hat{e}_i^{w2} - e_i^2| + \left| n^{-1}\sum_{i=1}^n e_i^2 - \sigma_e^2 \right| \\
&\le n^{-1}\sum_{i=1}^n |\hat{e}_i^w - e_i||\hat{e}_i^w| + n^{-1}\sum_{i=1}^n |\hat{e}_i^w - e_i||e_i| + O_p\left(n^{-1/2}\right) \\
&\le n^{-1}\sum_{i=1}^n |\varepsilon_i - \hat{\rho}_n\varepsilon_{i-1} - e_i + \tilde{g}(t_i) + \hat{x}_i'(\beta - \hat{\beta}_n^w) - \hat{\rho}_n[\tilde{g}(t_{i-1}) + \hat{x}_{i-1}'(\beta - \hat{\beta}_n^w)]| \\
&\qquad \cdot|\varepsilon_i - \hat{\rho}_n\varepsilon_{i-1} + \tilde{g}(t_i) + \hat{x}_i'(\beta - \hat{\beta}_n^w) - \hat{\rho}_n[\tilde{g}(t_{i-1}) + \hat{x}_{i-1}'(\beta - \hat{\beta}_n^w)]|
\end{aligned}$$

$$+n^{-1}\sum_{i=1}^{n}|\varepsilon_i-\hat\rho_n\varepsilon_{i-1}-e_i+\tilde g(t_i)+\hat x_i'(\beta-\hat\beta_n^w)-\hat\rho_n[\tilde g(t_{i-1})+\hat x_{i-1}'(\beta-\hat\beta_n^w)]||e_i|$$
$$+O_p\left(n^{-1/2}\right)$$
$$=n^{-1}\sum_{i=1}^{n}\Big|(\rho-\hat\rho_n)\varepsilon_{i-1}+O(n^{-1/3}\log n)\Big|\Big|\varepsilon_i-\hat\rho_n\varepsilon_{i-1}+O(n^{-1/3}\log n)\Big|$$
$$+n^{-1}\sum_{i=1}^{n}\Big|(\rho-\hat\rho_n)\varepsilon_{i-1}+O(n^{-1/3}\log n)\Big||e_i|+o_p\left(n^{-1/2}\right)=o_p(1).$$

Therefore, $|E_\star(\varepsilon_i^{\star 2})-E(\varepsilon_1^2)|\le\sum_{j=0}^{\infty}[|\hat\rho_n^j-\rho^j||\rho^j|E_\star(e_i^{\star 2})+|\rho^j||\hat\rho_n^j||E_\star(e_i^{\star 2})-\sigma_e^2|+|\hat\rho_n^j-\rho^j||\rho^j|\sigma_e^2]=o_p(1)$. By the same argument we can prove the second limit.

Proof of Theorem 3.1. Let $\tilde\beta_n^{\star w}=(\hat X'\hat\Omega^{-1}\hat X)^{-1}\hat X'\hat\Omega^{-1}\hat Y^\star$. Then $n^{1/2}(\hat\beta_n^{\star w}-\hat\beta_n^w)=n^{1/2}(\hat\beta_n^{\star w}-\tilde\beta_n^w)+n^{1/2}(\tilde\beta_n^{\star w}-\hat\beta_n^w)=I_1+I_2$, say. Combining Lemmas 5.7 and 5.8, similar to the proof of Theorem 3.1 in Vilar-Fernández and Vilar-Fernández (2002), we can show $I_1=o_p(1)$. Furthermore, $I_2=n^{1/2}(\hat X'\hat\Omega^{-1}\hat X)^{-1}\hat X'\hat\Omega^{-1}\varepsilon^\star$. Put $\alpha^\star=n^{-1/2}\hat X'\hat\Omega^{-1}\varepsilon^\star$ and $\alpha=n^{-1/2}\hat X'\hat\Omega^{-1}\varepsilon$. Applying Lemma 5.3 it is not difficult to show that $d_2(\Phi^\star,\Phi)$ converges to zero a.s., where $\Phi^\star$ and Φ are the distribution functions of $\alpha^\star$ and α respectively. Therefore, Theorem 3.1 follows from (2.6).

Proof of Theorem 3.2. It is easy to see that, in order to complete the proof of this theorem, it suffices to show the following equations:

$$\sqrt{n}(\hat\rho_n-\rho)=\frac{1}{\sqrt{n}E(\varepsilon_1^2)}\sum_{i=1}^{n-1}e_{i+1}\varepsilon_i+o_p(1),\tag{5.1}$$

$$\sqrt{n}(\hat\rho_n^\star-\hat\rho_n)=\frac{1}{\sqrt{n}E_\star(\varepsilon_1^{\star 2})}\sum_{i=1}^{n-1}e_{i+1}^\star\varepsilon_i^\star+o_p(1),\tag{5.2}$$

$$d_2(\Phi_3,\Phi_4)\to_p 0 \text{ as } n\to\infty,\tag{5.3}$$

where Φ_3 and Φ_4 are respectively the distribution functions of $n^{-1/2}\sum_{i=1}^{n-1}e_{i+1}\varepsilon_i/E(\varepsilon_1^2)$ and $n^{-1/2}\sum_{i=1}^{n-1}e_{i+1}^\star\varepsilon_i^\star/E_\star(\varepsilon_1^{\star 2})$. Since

$$\frac1n\sum_{i=1}^{n-1}\hat\varepsilon_i\hat\varepsilon_{i+1}=\frac1n\sum_{i=1}^{n-1}(\hat\varepsilon_i-\varepsilon_i)(\hat\varepsilon_{i+1}-\varepsilon_{i+1})+\frac1n\sum_{i=1}^{n-1}(\hat\varepsilon_i-\varepsilon_i)\varepsilon_{i+1}+\frac1n\sum_{i=1}^{n-1}(\hat\varepsilon_{i+1}-\varepsilon_{i+1})\varepsilon_i,$$

it follows from Lemma 5.7 that $n^{-1}\sum_{i=1}^{n-1}\hat\varepsilon_i\hat\varepsilon_{i=1}=n^{-1}\sum_{i=1}^{n-1}\varepsilon_i\varepsilon_{i-1}+o(n^{-1/2})$ a.s. Similarly, we can get $n^{-1}\sum_{i=1}^{n}\hat\varepsilon_i^2=n^{-1}\sum_{i=1}^{n}\varepsilon_i^2+o(n^{-1/2})$ a.s. Therefore, $\sqrt{n}(\hat\rho_n-\rho)=n^{-1/2}\sum_{i=1}^{n-1}e_{i+1}\varepsilon_i/\sum_{i=1}^{n}\varepsilon_i^2+o_p(1)=n^{-1/2}\sum_{i=1}^{n-1}e_{i+1}\varepsilon_i/E(\varepsilon_1^2)+o_p(1)$ by the fact that $n^{-1}\sum_{i=1}^{n}\varepsilon_i^2\to E(\varepsilon_1^2)$ as $n\to\infty$ and $\sum_{i=1}^{n-1}e_{i+1}\varepsilon_i/\sqrt{n}=O_p(1)$. This implies (5.1).

Next, as $\varepsilon_i^\star = \sum_{j=0}^{\infty} \hat{\rho}_n^j e_{i-j}^\star$ we can write $\sqrt{n}(\hat{\rho}_n^\star - \hat{\rho}_n) = n^{-1/2}\sum_{i=1}^{n-1} e_{i+1}^\star \varepsilon_i^\star / \sum_{i=1}^n \varepsilon_i^{\star 2}$. On the other hand, according to Lemma 5.3 and the proof of Lemma 5.7, $n^{-1}\sum_{i=1}^n \hat{e}_i^{w2} = n^{-1}\sum_{i=1}^n(\hat{\varepsilon}_i^w - \hat{\rho}_n \hat{\varepsilon}_{i-1}^w)^2 \le 2n^{-1}(1+\hat{\rho}_n^2)\sum_{i=1}^n \hat{\varepsilon}_i^{w2} = O_p(1)$. Therefore,

$$E_\star\Big(\sum_{i=1}^{n-1} e_{i+1}^\star \varepsilon_i^\star\Big)^2 = \sum_{i=1}^{n-1} E_\star(e_{i+1}^{\star 2})E(\varepsilon_i^{\star 2}) = O(n)\cdot \frac{1}{n}\sum_{i=1}^n \hat{e}_i^{w2}\cdot \frac{1}{n}\sum_{i=1}^n \hat{\varepsilon}_i^{w2} = O_p(n).$$

Further, by Lemma 5.8,

$$\Big|\frac{1}{n}\sum_{i=1}^n (\varepsilon_i^{\star 2}) - E_\star(\varepsilon_i^{\star 2})\Big| \le \Big|\frac{1}{n}\sum_{i=1}^n \varepsilon_i^{\star 2} - E(\varepsilon_1^2)\Big| + |E_\star(\varepsilon_i^{\star 2}) - E(\varepsilon_1^2)| \to_p 0 \text{ as } n\to\infty.$$

So, (5.2) holds. The proof of (5.3) is similar to that of Theorem 3.2 in Paparoditis (1996). We omit details. The proof is thus complete.

6. Concluding Remarks

It should be noted that the results of this paper are limited to the case of fixed designs. Sometimes in practice, especially in econometrics, it may be necessary to allow random regressors as well (cf. Robinson (1988)). In such a case different approaches, such as the paired bootstrap, may be more appropriate. We will investigate this topic in a separate paper.

Another important issue is to improve our bootstrap performance by Edgeworth correction. This is a more difficult issue due to the nonparametric component in (1.1) (cf., Linton (1995)), and further investigations are called for.

Other interesting topics for future studies include extension of our results to an ARMA error structure, and possibly to nonlinear time series error structures such as the ARCH and GARCH models.

References

Chen, H. (1988). Convergence rates for parametric components in a partially linear model. *Ann. Statist.* **16**, 136-146.

Chen, H. and Shiau, J. (1991). A two-stage spline smoothing method for partially linear models. *J. Statist. Plann. Inference* **27**, 187-202.

Chen, H. and Shiau, J. (1994). Data-driven efficient estimation for a partially linear model. *Ann. Statist.* **22**, 211-237.

Chen, Z. G. and Ni, J. Y. (1989). Subset regression time series and its modeling procedures. *J. Multivariate Anal.* **32**, 55-69.

Donald, G. and Newey, K. (1994). Series estimation of semilinear models. *J. Multivariate Anal.* **50**, 30-40.

Engle, R. F., Granger, W. J., Rice, J. and Weiss, A. (1986). Semiparametric estimates of the relation between weather and electricity sales. *J. Amer. Statist. Assoc.* **80**, 310-319.

Eubank, R. and Speckman, P. (1990). Trigonometric series regression estimators with an application to partially linear models. *J. Multivariate Anal.* **32**, 70-85.

Fuller, W. A. (1976). *Introduction to Statistical Time Series*. Wiley, New York.

Gao, J. T. (1995a). Asymptotic theory for partially linear models. *Commun. Statist. Theory Method* **22**, 3327-3354.

Gao, J. T. (1995b). The laws of the iterated logarithm of some estimates in a partially linear models. *Statist. Probab. Lett.* **25**, 153-162.

Hamilton, A. and Truong, K. (1997). Local linear estimation in partially linear models. *J. Multivariate Anal.* **60**, 1-19.

Härdle, W., Liang, H. and Gao, J. (2000). *Partially Linear Models*. Physica-Verlag, Heidelberg.

Heckman, N. (1986). Spline smoothing in a partially linear model. *J. Roy. Statist. Soc. Ser. B* **48**, 244-248.

Hong, S. Y. (2002). Normal approximation rate and bias reduction for data-driven kernel smoothing estimator in a semiparametric regression model. *J. Multivariate Anal.* **80**, 1-20.

Hong, S. Y. and Cheng, P. (1993). Bootstrap approximation of estimation for parameter in a semiparametric regression model. *Sci. China Ser. A* **23**, 239-251.

Kreiss, J. P. and Frabke, J. (1992). Bootstrap stationary autoregressive moving-average models. *J. Time Ser. Anal.* **13**, 296-317.

Liang, H., Härdle, W. and Sommerfeld, V. (2000). Bootstrap approximation in a partially linear regression model. *J. Statist. Plann. Inference* **91**, 413-426.

Linton, O. (1995). Second order approximation in the partially linear regression model. *Econometrica* **63**, 1079-1112.

Paparoditis, E. (1996). Bootstrapping autoregressive and moving average parameter estimates of infinite order vector autoregressive processes. *J. Multivariate Anal.* **57**, 277-296.

Rice, J. (1986). Convergence rates for partially splined models. *Statist. Probab. Lett.* **4**, 203-208.

Robinson, P. (1988). Root-N-consistent semiparametric regression. *Econometrica* **56**, 931-954.

Schick, A. (1994). Estimation of the autocorrelation coefficient in the presence of a regression trend. *Statist. Probab. Lett.* **21**, 371-380.

Schick, A. (1996). Efficient estimation in a semiparametric additive regression model with autoregressive errors. *Stochastic Process. Appl.* **61**, 339-361.

Schick, A. (1998). An adaptive estimator of the autocorrelation coefficient in regression models with autoregressive errors. *J. Time Ser. Anal.* **19**, 575-589.

Speckman, P. (1988). Kernel smoothing in partial linear models. *J. Royal Statist. Soc. B* **50**, 413-436.

Stute, W. (1995). Bootstrap of a linear model with AR-error structure. *Metrika* **18**, 395-410.

Sun, X. Q., You, J. H., Chen, G. M and Zhou, X. (2002). Convergence rates of estimators in partially linear regression models with MA(∞) error process. *Comm. Statist. Theory Method* **31**, 2251-2273.

Vilar-Fernández, J. M. and Vilar-Fernández, J. A. (2002). Bootstrap of minimum distance estimators in regression with correlated disturbances. *J. Statist. Plann. Inference* **108**, 283-299.

You, J. (2002). Semiparametric regression models with serially correlated and/or heteroscedastic errors. Unpublished thesis, University of Regina.

You, J. and Chen, G. (2002a). Semiparametric generalized least squares estimation in partially linear regression models with correlated errors. Submitted to *J. Statist. Plann. Inference*

You, J. and Chen, G. (2002b). Wild bootstrap estimation in partially linear regression models with heteroscedasticity. Submitted to *Statist. Probab. Lett.*

Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong.

E-mail: jyou@bios.unc.edu

Department of Applied Mathematics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong.

E-mail: maxzhou@polyu.edu.hk

(Received march 2003; accepted February 2004)