Empirical likelihood for heteroscedastic partially linear models

Xuewen Lu

Department of Mathematics, Hunan Normal University, Changsha, Hunan, 410081, China
Department of Mathematics and Statistics, University of Calgary, Calgary, Alberta, T2N 1N4, Canada

A R T I C L E   I N F O

Article history:
Received 7 August 2007
Available online 16 May 2008

AMS 2000 subject classifications:
62F35
62G20

Keywords:
Double robustness
Empirical likelihood
Heteroscedasticity
Kernel estimation
Partially linear model
Semiparametric efficiency

A B S T R A C T

We make empirical-likelihood-based inference for the parameters in heteroscedastic partially linear models. Unlike the existing empirical likelihood procedures for heteroscedastic partially linear models, the proposed empirical likelihood is constructed using components of a semiparametric efficient score. We show that it retains the double robustness feature of the semiparametric efficient estimator for the parameters and shares the desirable properties of the empirical likelihood for linear models. Compared with the normal approximation method and the existing empirical likelihood methods, the empirical likelihood method based on the semiparametric efficient score is more attractive not only theoretically but empirically. Simulation studies demonstrate that the proposed empirical likelihood provides smaller confidence regions than that based on semiparametric inefficient estimating equations subject to the same coverage probabilities. Hence, the proposed empirical likelihood is preferred to the normal approximation method as well as the empirical likelihood method based on semiparametric inefficient estimating equations, and it should be useful in practice.

© 2008 Elsevier Inc. All rights reserved.

1. Introduction

Consider a partially linear regression model,

\[ Y_i = X_i^T \beta + g(Z_i) + \varepsilon_i, \quad (i = 1, \ldots, n), \]

where \((X_i, Z_i, Y_i)\)'s are i.i.d. copies from a population \((X, Z, Y)\), \(Y\) is a response variable, \(X\) is a random \(k\)-vector, \(Z\) is a random scalar, \(\beta\) is an unknown parameter vector, \(g(\cdot)\) is an unknown function, \(\varepsilon_i\)'s are independent random errors with \(E(\varepsilon_i|X_i, Z_i) = 0\) and \(\text{Var}(\varepsilon_i|X_i, Z_i) = v(X_i, Z_i) > 0\), \(v(X, Z)\) is a function of \((X, Z)\) representing possible heteroscedasticity. We assume that \(v(X_i, Z_i)\) is an unknown function of one-dimensional estimable variable \(\xi(X_i, Z_i)\) and \(A^T\) is the transpose of a column vector or matrix \(A\). Since its introduction by Engle et al. [3], this model has been broadly and deeply studied by many authors in various disciplines. Härdle et al. [5] wrote a monograph on this model. Recently, Ma et al. [12] summarized a variety of approaches and considered a family of consistent estimators of \(\beta\). They showed that the optimal semiparametric efficiency bound for \(\beta\) can be reached by a semiparametric kernel estimator in the family. Theoretical illustration and numerical simulations have shown the advantages of the semiparametric efficient estimator.

To make inference about \(\beta\) in model (1) using either semiparametric efficient estimators or semiparametric inefficient estimators, many authors have considered the normal approximation method. The empirical likelihood method introduced in Owen [13,14] might be useful for the purpose of making semiparametric inference for this model. To the best of our
knowledge, there is not much research in the literature on this model using the empirical likelihood method, although it has been successfully applied to various models, to name a few, linear models [15], generalized linear models [8], quantile estimation [1], generalized estimating equations [17], dependent process models [7], errors-in-covariables models [19,10], Cox regression models [16], additive risk models [11,20], and nonparametric regression models [2]. When models are semiparametric type like model (1), plug-in estimates of nuisance parameters in estimating equations are often needed. Hjort et al. [6] have provided a general investigation on this issue, but their method is not directly applicable to confidence regions in our case.

What we know about semiparametric inference for partially linear models with random designs based on the empirical likelihood method is the work by [18,21]. Wang and Jing [18] studied the empirical likelihood for partially linear models, but they assumed that the random errors were homoscedastic, that is, \( e_i \)'s are independent of \((X_i, Z_i)\)'s. This is not realistic in many applications. Zhu and Xue [21] investigated likelihood confidence regions in a partially linear single-index model and treated the partially linear model as a special case of that model. They considered heteroscedastic errors, but their empirical likelihood was constructed from the components of a semiparametric inefficient estimating equation. We think that it might be more informative if we use components of a semiparametric efficient score to construct the empirical likelihood. Motivated by the research of these authors and Ma et al. [12], in this paper, we propose a new approach to the empirical likelihood inference about \( \beta \) based on the semiparametric efficient score given by Ma et al. [12]. We will show that the empirical log-likelihood ratio is asymptotically a standard chi-square random variable. We will conduct simulation studies to compare the proposed method with the normal approximation method and the existing empirical likelihood methods based on semiparametric inefficient estimating equations.

The plan of the paper is as follows. In Section 2, we introduce the empirical likelihood method for the inference of partially linear models and present our main results. In Section 3, we report Monte Carlo simulation results. All the technical conditions and proofs are presented in the Appendix.

2. Methodology and main results

First, we introduce the estimation method of Ma et al. [12] for the parameter \( \beta \) in model (1) and propose the empirical likelihood method for \( \beta \).

To estimate \( \beta \), Ma et al. [12] reviewed the following set of estimating equations:

\[
0 = \Psi(Y, X, Z, w, \beta, \hat{g}) = n^{-1} \sum_{i=1}^{n} w_i \{ Y_i - X_i^T \beta - \hat{g}(Z_i, \beta) \} \{ X_i - \hat{E}(X|Z_i) \},
\]

(2)

where \( \hat{g}(z, \beta) \) could be a given or estimated function of \( z \), it may or may not contain \( \beta, \hat{E}(\cdot|z) \) denotes a consistent estimator of \( E(\cdot|z) \), and \( w_i = w(X_i, Z_i) \) denotes the weights. The common choice of \( w \) is \( w \equiv 1 \) [18,9] or \( w = \text{Var}(e|X, Z)^{-1} \) [5, eq. 2.1.4]. They pointed out that if one cannot obtain a consistent estimator of \( g \), including a nonconstant weight \( w \) in (2) results in an inconsistent estimator of \( \beta \). Given that weights are commonly needed in practice and \( g \) could simply be a nuisance parameter, it is desirable to develop estimators which are robust to a not well estimated or misspecified \( g \) but still accommodate the inclusion of weights, i.e. estimators are double robust. To achieve this goal, Ma et al. [12] proposed a new family of weighted estimating equations:

\[
0 = n^{-1} \sum_{i=1}^{n} w_i \{ Y_i - X_i^T \beta - \hat{g}(Z_i, \beta) \} \left[ X_i - \frac{\hat{E}(w(X_i, Z_i)X_i|Z_i)}{\hat{E}(w(X_i, Z_i)|Z_i)} \right].
\]

(3)

They showed that when \( w \) is taken to be \( w = w(X, Z) = E(e^2|X, Z)^{-1} \), (3) mimics the semiparametric efficient score for \( \beta \) (Proposition 1, [12]). To estimate \( \beta \), functions \( g, w, E(w|Z) \) and \( E(wX|Z) \) need to be nonparametrically estimated. Suppose their corresponding estimators are \( \hat{g}, \hat{w}, \hat{E}(\hat{w}(X, Z)|Z) \) and \( \hat{E}(\hat{w}(X, Z)X|Z) \), respectively. The estimator of \( \beta \) is the solution to

\[
0 = n^{-1} \sum_{i=1}^{n} \{ Y_i - X_i^T \beta - \hat{g}(Z_i, \beta) \} \hat{w}(X_i, Z_i) \left[ X_i - \frac{\hat{E}(\hat{w}(X_i, Z_i)X_i|Z_i)}{\hat{E}(\hat{w}(X_i, Z_i)|Z_i)} \right],
\]

(4)

where \( \hat{g}(Z_i, \beta) \) is an estimator of \( g(Z_i) \) for given \( \beta \), one option for it is

\[
\hat{E}(\hat{w}(Y_i, Z_i))/\hat{E}(\hat{w}(|Z_i)) - (\hat{E}(\hat{w}X_i^T|Z_i)/\hat{E}(\hat{w}|Z_i)) \beta,
\]

(5)

where \( \hat{E}(\hat{w}(Y_i, Z_i)) \) is a nonparametric estimator of \( E(wY_i|Z_i) \).

The following steps outline the iterative algorithm in calculating the estimates:

Step 1. Estimate \( \hat{E}(Y|Z) \) and \( \hat{E}(X|Z) \) nonparametrically and obtain the initial estimates \( \hat{\beta}_1 = [(X - \hat{E}(X|Z))^T(X - \hat{E}(X|Z))]^{-1}[(X - \hat{E}(X|Z))^TY - \hat{E}(Y|Z)] \), and

\[
\hat{g}(Z) = \hat{E}(Y_i|Z) - \hat{E}(X_i|Z)^T \hat{\beta}_1,
\]

where \( X \) is the design matrix, \( Y \) and \( Z \) are the vectors composed of \( Y_i \) and \( Z_i \), respectively, \( \hat{E}(X|Z) \) is the matrix composed of \( \hat{E}(X_i|Z) \) and \( \hat{E}(Y|Z) \) is the vector composed of \( \hat{E}(Y_i|Z) \), respectively.
Step 2. Calculate \( \hat{\xi}_i = Y_i - X_i^T \hat{\beta} - \hat{g}(Z_i) \).

Step 3. Obtain nonparametrically estimated \( \hat{w}_i = \hat{E}(\hat{\xi}_i^2 | X_i, Z_i)^{-1} \), for \( i = 1, \ldots, n \).

Step 4. For the set of \( \{\hat{w}_i\} \), obtain \( \hat{E}(\hat{w}_i | Z_i) \), \( \hat{E}(\hat{w}_i X_i | Z_i) \) and \( \hat{E}(\hat{w}_i Y_i | Z_i) \), for \( i = 1, \ldots, n \), in the form of local-linear estimators.

Step 5. Let \( \tilde{X}_i = X_i - \hat{E}(\hat{w}_i X_i | Z_i)/\hat{E}(\hat{w}_i | Z_i) \) and \( \tilde{Y}_i = Y_i - \hat{E}(\hat{w}_i Y_i | Z_i)/\hat{E}(\hat{w}_i | Z_i) \), and let \( W \) be a diagonal matrix with \( \hat{w}_i \) being the \( i \)th diagonal element. Calculate

\[
\hat{\beta} = (\tilde{X}^T W \tilde{X})^{-1} \tilde{X}^T W \tilde{Y},
\]

and \( \hat{g} \) as in (5) with \( \beta = \hat{\beta} \), where \( \tilde{X} \) is the matrix and \( \tilde{Y} \) is the vector composed of \( \tilde{X}_i \) and \( \tilde{Y}_i \), respectively.

Step 6. Iterate Steps 2–5 until convergence and use the sandwich covariance estimate based on (6) to estimate the asymptotic variance matrix of \( \beta \), with the estimated variance matrix of \( Y \) being a diagonal matrix with \( i \)th diagonal element equal to \( \hat{E}(\hat{\xi}_i^2 | X_i, Z_i) \). Specifically, the estimated asymptotic variance matrix of \( \beta \) is given by \( n^{-1} \hat{V} \) with

\[
\hat{V} = \left( \frac{\tilde{X}^T W \tilde{X}}{n} \right)^{-1},
\]

where all the estimated terms in \( \hat{V} \) are taken from the final step of the above algorithm.

In fact, the initial estimator obtained in Step 1 is already root-\( n \) consistent, a one step Newton–Raphson update from it gives an estimator that has the same asymptotic distribution as the solution of (4). To simplify computations, we will just perform the above iterative algorithm one time in the implementation of simulation studies.

Under appropriate conditions, \( \hat{\beta} \) is an asymptotic normal estimator of the true parameter value \( \beta \), stated as follows.

**Theorem 1** (Proposition 2, Ma et al. [12]). Assume \( \hat{\beta} \) solves (4). Then, under the regularity conditions given in the Appendix, when \( n \to \infty \),

\[
\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{D} N(0, V),
\]

where \( \xrightarrow{D} \) stands for convergence in distribution,

\[
V = \left[ E(S_{\text{eff}} S_{\text{eff}}^T) \right]^{-1} = \left[ E \left( w X X^T - \frac{E(w X | Z) E(w X | Z)^T}{E(w | Z)} \right) \right]^{-1},
\]

with the semiparametric efficient score

\[
S_{\text{eff}} = w E \left( X - \frac{E(w X | Z)}{E(w | Z)} \right).
\]

Therefore, a large sample \((1 - \alpha)\) 100% confidence region for \( \beta \) based on the above normal approximation is given by

\[
R_{\text{NA, } \alpha} = \{ \beta : n(\hat{\beta} - \beta)^T \hat{V}^{-1} (\hat{\beta} - \beta) \leq \chi^2_{\alpha}(k) \},
\]

where \( \hat{V} \) is a consistent estimator of \( V \) and is given in (7), and \( \chi^2_{\alpha}(k) \) is the \((1 - \alpha)\)th upper quantile of the chi-square distribution \( \chi^2(k) \) with \( k \) degrees of freedom.

We now introduce an auxiliary random vector using the semiparametric efficient score \( S_{\text{eff}} \). Let

\[
\xi_i(\beta) = S_{\text{eff}}(X_i, Z_i, Y_i) = w(X_i, Z_i)(Y_i - X_i^T \beta - g(Z_i)) \left\{ X_i - \frac{E(w X_i | Z_i)}{E(w_i | Z_i)} \right\}.
\]

Note that \( E[\xi_i(\beta)] = 0, i = 1, \ldots, n \), if \( \beta \) is the true parameter. Using this fact, we apply Owen’s empirical likelihood method [13, 14] to make inference about \( \beta \). Let \( p = (p_1, \ldots, p_n) \) be a probability vector satisfying \( \sum_{i=1}^{n} p_i = 1 \) and \( p_i \geq 0 \) for \( i = 1, \ldots, n \). Let \( F_p \) be a distribution function which assigns probability \( p_i \) at point \( \xi_i(\beta) \). Therefore,

\[
\delta(F_p, \beta) = \sum_{i=1}^{n} p_i \xi_i(\beta).
\]

Then, the empirical likelihood, evaluated at the true parameter value \( \beta \), is defined by

\[
L(\beta) = \sup \left\{ \prod_{i=1}^{n} p_i : \sum_{i=1}^{n} p_i = 1, \delta(F_p, \beta) = 0 \right\}.
\]
Since \( \delta(F_p, \beta) \) depends on the unknown functions \( g(Z_i), E(w|Z_i) \) and \( E(wX_i|Z_i) \), we replace them by the local-linear estimators from the final step of the iterative algorithm: \( \hat{g}(Z_i, \beta) = \hat{E}(\hat{w}_i|Y_i|Z_i)/\hat{E}(\hat{w}_i|Z_i) - \{ \hat{E}(\hat{w}_i X_i^T|Z_i)/\hat{E}(\hat{w}_i|Z_i) \} \beta, \hat{E}(\hat{w}_i|Z_i), \hat{E}(\hat{w}_i X_i|Z_i) \) and \( \hat{E}(\hat{w}_i Y_i|Z_i) \). Hence, an estimated \( \hat{\xi}_i(\beta) \) is given by

\[
W_{ni} \equiv W_{ni}(\beta) = \hat{w}(X_i, Z_i)[\hat{Y}_i - \hat{X}_i^T \beta] \hat{X}_i,
\]

where \( \hat{Y}_i \) and \( \hat{X}_i \) are defined in Step 5 of the iterative algorithm. Consequently, an estimated empirical likelihood by the plug-in method, evaluated at the true value \( \beta \), is defined by

\[
\tilde{L}(\beta) = \sup \left\{ \prod_{i=1}^n p_i : \sum_{i=1}^n p_i = 1, \, \hat{\delta}(F_p, \beta) = 0 \right\},
\]

where \( \hat{\delta}(F_p, \beta) = \sum_{i=1}^n p_i W_{ni} \). Then, by the method of Lagrange multipliers, we can easily get

\[
p_i = \frac{1}{n} \{ 1 + \lambda^T W_{ni} \}^{-1}, \quad i = 1, \ldots, n,
\]

where \( \lambda = (\lambda_1, \ldots, \lambda_k)^T \) is the solution of

\[
\frac{1}{n} \sum_{i=1}^n \frac{W_{ni}}{1 + \lambda^T W_{ni}} = 0. \tag{9}
\]

Note that \( \prod_{i=1}^n p_i, \text{subject to} \sum_{i=1}^n p_i = 1 \), attains its maximum \( n^{-n} \) at \( p_i = n^{-1} \). So we define the empirical likelihood ratio at \( \beta \) by

\[
R(\beta) = \prod_{i=1}^n (np_i) = \prod_{i=1}^n \{ 1 + \lambda^T W_{ni} \}^{-1},
\]

and the corresponding empirical log-likelihood ratio is defined as

\[
\mathcal{L}(\beta) = -2 \log R(\beta) = 2 \sum_{i=1}^n \log \{ 1 + \lambda^T W_{ni} \}. \tag{10}
\]

The following theorem gives the asymptotic distribution of \( \mathcal{L}(\beta) \).

**Theorem 2** (Wilks' Theorem). Assume the conditions in Theorem 1 hold. If \( \beta \) is the true parameter value, then

\[
\mathcal{L}(\beta) \xrightarrow{D} \chi^2(k).
\]

**Theorem 2** can be used to construct empirical likelihood confidence regions for \( \beta \). A \( (1 - \alpha) \) 100% confidence region for \( \beta \) is given by

\[
R_{EL, \alpha} = \{ \beta : \mathcal{L}(\beta) \leq \chi^2_{\alpha}(k) \}. \tag{11}
\]

**Corollary 1.** Under the conditions of Theorem 2,

\[
P(\beta \in R_{EL, \alpha}) \rightarrow 1 - \alpha, \quad \text{as} \ n \rightarrow \infty.
\]

In (4), if we take \( w \equiv 1 \), we obtain an unweighted estimating equation for \( \beta \). The result in **Theorem 2** still holds. This corresponds to the empirical likelihood method considered by Wang and Jing [18] and Zhu and Xue [21] for partially linear models, where the empirical likelihood is constructed from the components of the unweighted estimating equation, which is not semiparametric efficient for heteroscedastic partially linear models. In our simulation studies for heteroscedastic designs, we will investigate the finite sample performance of the proposed empirical likelihood method and their methods along with the normal approximation method using the semiparametric efficient estimator of Ma et al. [12].

**3. Monte Carlo simulation studies**

We conduct some Monte Carlo experiments to compare the finite sample performance of the three methods described in Section 2: semiparametric inefficient (or unweighted) empirical likelihood (SIEL), semiparametric efficient normal approximation (SENA), and semiparametric efficient empirical likelihood (SEEL). We examine two models, one with a lower-dimensional \( X (k = 1) \) and one with a higher-dimensional \( X (k = 3) \).
Table 1
Model 1: Comparisons of three methods (SIEL, SENA, SEEL) for semiparametric inference of $\beta = 1$

<table>
<thead>
<tr>
<th>n</th>
<th>SIEL</th>
<th>SENA</th>
<th>SEEL</th>
<th>SIEL</th>
<th>SENA</th>
<th>SEEL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>m = 1</td>
<td>m = 2</td>
<td>m = 3</td>
<td>m = 4</td>
<td>m = 3</td>
<td>m = 4</td>
</tr>
<tr>
<td>100</td>
<td>(a) [0.255, 1.733]</td>
<td>[0.450, 1.541]</td>
<td>[0.458, 1.534]</td>
<td>[0.250, 1.730]</td>
<td>[0.461, 1.570]</td>
<td>[0.462, 1.570]</td>
</tr>
<tr>
<td></td>
<td>(b) 1.478</td>
<td>1.091</td>
<td>1.076</td>
<td>1.480</td>
<td>1.108</td>
<td>1.107</td>
</tr>
<tr>
<td></td>
<td>(c) 93.1</td>
<td>91.6</td>
<td>90.5</td>
<td>93.5</td>
<td>93.4</td>
<td>93.8</td>
</tr>
<tr>
<td></td>
<td>(0.149)</td>
<td>(0.109)</td>
<td>(0.146)</td>
<td>(0.094)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>(a) [0.469, 1.532]</td>
<td>[0.603, 1.391]</td>
<td>[0.605, 1.389]</td>
<td>[0.468, 1.532]</td>
<td>[0.611, 1.407]</td>
<td>[0.611, 1.406]</td>
</tr>
<tr>
<td></td>
<td>(b) 1.063</td>
<td>0.789</td>
<td>0.784</td>
<td>1.064</td>
<td>0.797</td>
<td>0.795</td>
</tr>
<tr>
<td></td>
<td>(c) 94.3</td>
<td>92.2</td>
<td>92.2</td>
<td>94.7</td>
<td>94.1</td>
<td>94.0</td>
</tr>
<tr>
<td></td>
<td>(0.077)</td>
<td>(0.047)</td>
<td>(0.074)</td>
<td>(0.046)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>300</td>
<td>(a) [0.560, 1.431]</td>
<td>[0.678, 1.325]</td>
<td>[0.679, 1.324]</td>
<td>[0.558, 1.430]</td>
<td>[0.686, 1.338]</td>
<td>[0.686, 1.338]</td>
</tr>
<tr>
<td></td>
<td>(b) 0.872</td>
<td>0.648</td>
<td>0.645</td>
<td>0.872</td>
<td>0.653</td>
<td>0.653</td>
</tr>
<tr>
<td></td>
<td>(c) 93.2</td>
<td>91.9</td>
<td>92.0</td>
<td>93.6</td>
<td>93.4</td>
<td>93.4</td>
</tr>
<tr>
<td></td>
<td>(0.053)</td>
<td>(0.034)</td>
<td>(0.052)</td>
<td>(0.033)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(a) average 95% confidence intervals. (b) average lengths of 95% confidence intervals. (c) empirical coverage probabilities (%). The numbers in parentheses are the mean squared errors of the SIE and SE estimators, $m$ is the bandwidth tuning parameter, $n$ is the sample size.

Model 1: Consider the following heteroscedastic partially linear model with a one-dimensional $X$,

$$Y = X\beta + g(Z) + \varepsilon,$$

where $\beta = 1$, $g(Z) = \exp(Z - 2)$, $Z$ is from a uniform distribution $U(2, 4)$, $X$ is from a uniform distribution $U(Z/2, Z)$, and $Y$ is from a normal distribution with mean $E(Y|X, Z) = X\beta + g(Z)$ and variance $\text{Var}(\varepsilon|X, Z) = [E^2(Y|X, Z) + 1]/16$. Since $\text{Var}(\varepsilon|X, Z)$ is a function of $\xi(X, Z) = E(Y|X, Z)$, $\hat{\xi} = X\hat{\beta} + \hat{g}(Z)$ is obtained from the initial step of the iterative algorithm and is used to estimate $\text{Var}(\varepsilon|X, Z)$. The initial estimator $\hat{\beta}$ is updated once using the estimated weights and the semiparametric efficient score to obtain the semiparametric efficient estimator $\hat{\beta}$. In nonparametric regressions, we use the Epanechnikov kernel for smoothing.

We calculate average length and coverage probability of confidence intervals given by three methods at the nominal level 95%. We report the results over 1000 replicates in Table 1. To assess the sensitivity of the estimates and the coverage probabilities to bandwidth and sample size simulations are conducted for different bandwidth values and sample sizes. The sample sizes used are $n = 100, 200, 300$. In the estimation of required nonparametric functions of $Z$ and $\xi$, the bandwidth is taken to be $m \times \text{sd}(Z) \times n^{-1/5}$ for $Z$ and $m \times 1.2\text{sd}(\hat{\xi}) \times n^{-1/5}$ for $\xi$, where $\text{sd}(Z)$ and $\text{sd}(\hat{\xi})$ represent the sample standard deviations of the regressors $Z$ and $\hat{\xi}$, respectively. $m$ is a bandwidth tuning parameter taking values 1, 2, 3 and 4. These chosen bandwidths satisfy the bandwidth conditions stated in the Appendix.

Model 2: Consider the following heteroscedastic partially linear model with a three-dimensional covariate vector $X$,

$$Y = X^T\beta + g(Z) + \varepsilon,$$

where $\beta = (\beta_1, \beta_2, \beta_3)^T = (1, -1, 2)^T$, $g(Z) = 6\sin(2Z)$, $Z$ is from a uniform distribution $U(2, 4)$, $X = (X_1, X_2, X_3)^T$. $X_1$ is from a uniform distribution $U(Z/2, Z)$, $X_2$ is from a uniform distribution $U(-Z, Z)$, $X_3$ is from a triangle distribution on $[-2, 2]$, independent of $Z$, $X_1$ and $X_2$, and $Y$ is from a normal distribution with mean $E(Y|X, Z) = X\beta + g(Z)$ and variance $\text{Var}(\varepsilon|X, Z) = [E^2(Y|X, Z) + 1]/16$. The estimation procedure is the same as that of Model 1.

Model 2 has a higher-dimensional $X$ than Model 1. The bandwidth and sample size are set in the same way as in Model 1. We obtain confidence regions instead of confidence intervals from the SIE and SE methods. In Table 2, we report the coverage probabilities of these confidence regions based on 1000 replicates, as well as the mean squared errors of the SIE estimator and relative efficiency of the SE estimator related to the SE estimator.

From these simulation results, we draw the following conclusions. In general, regardless of the dimensionality of $X$, SENA and SEEL work equally well and both outperform SIEL in terms of the average length of the confidence intervals and
Table 2
Model 2: Comparisons of three methods (SIEL, SENA, SEEL) for semiparametric inference of \( \beta = (1, -1, 2)^T \)

<table>
<thead>
<tr>
<th>n</th>
<th>SIEL</th>
<th>SENA</th>
<th>SEEL</th>
<th>SIEL</th>
<th>SENA</th>
<th>SEEL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( m = 1 )</td>
<td>( m = 2 )</td>
<td></td>
<td>( m = 3 )</td>
<td>( m = 4 )</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>CP</td>
<td>86.8</td>
<td>93.4</td>
<td>87.7</td>
<td>88.8</td>
<td>96.1</td>
</tr>
<tr>
<td></td>
<td>( \beta )</td>
<td>RE</td>
<td>MSE</td>
<td>( \beta )</td>
<td>RE</td>
<td>MSE</td>
</tr>
<tr>
<td></td>
<td>( \hat{\beta}_1 )</td>
<td>0.313</td>
<td>0.149</td>
<td>( \hat{\beta}_1 )</td>
<td>0.294</td>
<td>0.146</td>
</tr>
<tr>
<td></td>
<td>( \hat{\beta}_2 )</td>
<td>0.009</td>
<td>0.010</td>
<td>( \hat{\beta}_2 )</td>
<td>0.584</td>
<td>0.009</td>
</tr>
<tr>
<td></td>
<td>( \hat{\beta}_3 )</td>
<td>0.425</td>
<td>0.038</td>
<td>( \hat{\beta}_3 )</td>
<td>0.518</td>
<td>0.037</td>
</tr>
<tr>
<td>200</td>
<td>CP</td>
<td>92.6</td>
<td>94.2</td>
<td>91.8</td>
<td>93.3</td>
<td>97.3</td>
</tr>
<tr>
<td></td>
<td>( \beta )</td>
<td>RE</td>
<td>MSE</td>
<td>( \beta )</td>
<td>RE</td>
<td>MSE</td>
</tr>
<tr>
<td></td>
<td>( \hat{\beta}_1 )</td>
<td>0.269</td>
<td>0.070</td>
<td>( \hat{\beta}_1 )</td>
<td>0.343</td>
<td>0.067</td>
</tr>
<tr>
<td></td>
<td>( \hat{\beta}_2 )</td>
<td>0.005</td>
<td>0.017</td>
<td>( \hat{\beta}_2 )</td>
<td>0.514</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>( \hat{\beta}_3 )</td>
<td>0.379</td>
<td>0.017</td>
<td>( \hat{\beta}_3 )</td>
<td>0.486</td>
<td>0.017</td>
</tr>
<tr>
<td>300</td>
<td>CP</td>
<td>94.2</td>
<td>95.4</td>
<td>92.3</td>
<td>94.5</td>
<td>96.8</td>
</tr>
<tr>
<td></td>
<td>( \beta )</td>
<td>RE</td>
<td>MSE</td>
<td>( \beta )</td>
<td>RE</td>
<td>MSE</td>
</tr>
<tr>
<td></td>
<td>( \hat{\beta}_1 )</td>
<td>0.223</td>
<td>0.045</td>
<td>( \hat{\beta}_1 )</td>
<td>0.291</td>
<td>0.044</td>
</tr>
<tr>
<td></td>
<td>( \hat{\beta}_2 )</td>
<td>0.003</td>
<td>0.012</td>
<td>( \hat{\beta}_2 )</td>
<td>0.527</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>( \hat{\beta}_3 )</td>
<td>0.346</td>
<td>0.012</td>
<td>( \hat{\beta}_3 )</td>
<td>0.471</td>
<td>0.012</td>
</tr>
</tbody>
</table>

CP, empirical coverage probability of the 95% confidence region (%); RE, mean squared error of the SE estimator over that of the SI estimator \((\text{MSE}(\hat{\beta}_{SE})/\text{MSE}(\hat{\beta}_{SI}))\); MSE: mean squared error of the SI estimator; \( m \), the bandwidth tuning parameter; \( n \), sample size.

the efficiency of the estimates, subject to the same coverage probability. A trivial impact of the dimensionality of \( X \) is expected since we have assumed that the conditional variance is a function of the conditional mean to avoid the curse of dimensionality. SIEL gives much wider confidence intervals or regions and lower efficiency though the coverage probabilities are closer to the nominal level than SENA and SEEL. SENA and SEEL tend to have narrower confidence intervals or regions than SIEL for general bandwidth. SIEL seems to be more robust against bandwidth selection. We believe this is because SENA and SEEL need to estimate conditional variance. However, it is apparent that the SE estimator is more efficient than the SI estimator for a wide range of bandwidth values where the SI estimator is best estimated with minimum mean squared errors. When the sample size is small, the coverage probabilities of all three methods deviate from the nominal level, the coverage accuracy is somewhat sensitive to the bandwidth. If the bandwidth is appropriately chosen (taking \( m \) between 2 and 3) so that the mean squared errors of both SE and SI estimators are minimum, and the sample size is moderate or large \(( n \geq 200)\), all the coverage probabilities are in agreement with the nominal level, the average length of the confidence intervals becomes shorter, but SIEL still systematically has wider confidence intervals or regions than SENA and SEEL. We have used some empirical bandwidths. A data driven bandwidth selection method may be used in nonparametric estimation, for example, least squares cross-validation (LSCV) or generalized cross-validation (GCV), as suggested in [21]. But they require more computational efforts for implementation. In summary, for inference of heteroscedastic partially linear models, both SENA and SEEL have better overall performance than SIEL. Given the advantages that SEEL has: double robustness, semiparametric efficiency and all the features of the empirical likelihood method (see [4]), it is preferred to SENA and SIEL.
Acknowledgments

I wish to thank Professor Yanyuan Ma for allowing me to read the technical report of her paper and, the Associate Editor and two referees for their constructive comments. This research was partly supported by a grant from the Natural Sciences and Engineering Research Council of Canada.

Appendix. Conditions and proofs

A.1. Conditions

The following regularity conditions are taken from Ma et al. [12] with some modifications.

**Condition 1.** The errors $e_i (i = 1, \ldots, n)$ are independent and $0 < \text{Var}(e_i^2) < \infty$.

**Condition 2.** There exists a variance function, $v(\cdot)$, $\xi_i = \xi(X_i, Z_i)$ and a positive constant $\gamma$, such that $E(e_i^2 | X_i, Z_i) = v(\xi_i)$, with $v(\cdot) > \gamma > 0$.

**Condition 3.** The functions $g(z), E(X|x), E(Y|z), v(\xi), E(w|z), E(wX|z)$ and $E(wY|z)$ are twice continuously differentiable with finite derivatives. As a function of $(x, z), \xi$ is three times differentiable with finite derivatives.

**Condition 4.** There exists a function $\mu_4(\cdot)$ such that $E(e_i^4) = E[\mu_4(\xi_i)]$. The function $\mu_4(\cdot)$ is continuous; furthermore, there exists an $s > 2$, such that $\max_{1 \leq i \leq n} E(e_i^{2s}) < c < \infty$ for some $c > 0$. This condition is necessary for obtaining uniform consistency of the estimator for the variance function $v(\cdot)$.

**Condition 5.** Assume that the random variables $X_i$ have a density, $f_X$, and that the support of $f_X$ is a compact interval. This condition ensures that $X_i^2 \beta$ is bounded.

**Condition 6.** Assume that the random variables $\xi_i$ and $Z_i$ have densities, $f_\xi$ and $f_Z$, respectively, that the supports of $f_\xi$ and $f_Z$ are compact intervals and that $f_\xi$ and $f_Z$ are twice continuously differentiable, satisfying $0 < \inf f_\xi(\cdot) \leq \sup f_\xi(\cdot) < \infty$ and $0 < \inf f_Z(\cdot) \leq \sup f_Z(\cdot) < \infty$.

**Condition 7.** The kernel function $K$ is symmetric and continuously differentiable with compact support $[-1, 1]$.

**Condition 8.** The bandwidth $h$ used in the kernel estimators satisfies $h \rightarrow 0$, $nh^3 \rightarrow \infty$ and $nh^8 \rightarrow 0$ as $n \rightarrow \infty$, and

$$\lim_{n \rightarrow \infty} \inf (nh/ \log n)^{1/2} n^{-2/r} > 0$$

for a constant $r$ ($0 < r < s$), where $s$ is given in **Condition 4**.

**Condition 9.** The estimator of the variance function $v(\cdot)$ is such that $\hat{v}(\cdot)$ is truncated below by a sequence $\zeta_n > 0$, where $\zeta_n \rightarrow 0$. This sequence satisfies $h/\zeta_n \rightarrow 0$, $nh^2 \zeta_n^2 \rightarrow \infty$ and $nh \zeta_n^2 / \log n \rightarrow \infty$.

A.2. Proof of Theorem 2

To prove **Theorem 2**, it suffices to prove the following lemmas. Throughout the proof, $| \cdot |$ is understood to be $\| \cdot \|$ when the argument is a vector.

**Lemma 1.** Under the conditions of **Theorem 2**, we have

(a) $\max_{1 \leq i \leq n} |W_{ni}| = o_p(n^{1/2})$.

(b) $| \lambda | = o_p(n^{-1/2})$.

**Lemma 2.** Under the conditions of **Theorem 2**, we have

(a) $\mathcal{L}(\beta) = (n^{-1/2} \sum_{i=1}^{n} W_{ni})^T (n^{-1} \sum_{i=1}^{n} W_{ni} W_{ni}^T)^{-1} (n^{-1/2} \sum_{i=1}^{n} W_{ni}) + o_p(1)$.

(b) $n^{-1/2} \sum_{i=1}^{n} W_{ni} \xrightarrow{d} N(0, V^{-1})$.

(c) $n^{-1} \sum_{i=1}^{n} W_{ni} W_{ni}^T \xrightarrow{p} V^{-1}$, where $\xrightarrow{p}$ stands for convergence in probability.
First, suppose that Lemmas 1 and 2 hold, we prove Theorem 2. It follows from Lemma 2

\[ L(\beta) = \left( n^{-1/2} \sum_{i=1}^{n} W_{ni} \right)^{T} V \left( n^{-1/2} \sum_{i=1}^{n} W_{ni} \right) + o_{p}(1) \]

\[ = \left( V^{1/2} n^{-1/2} \sum_{i=1}^{n} W_{ni} \right)^{T} \left( V^{1/2} n^{-1/2} \sum_{i=1}^{n} W_{ni} \right) + o_{p}(1) \]

and \( V^{1/2} n^{-1/2} \sum_{i=1}^{n} W_{ni} \xrightarrow{D} N(0, I_{p}) \), here \( I_{p} \) denotes an identity matrix. Therefore, Theorem 2 is proved.

Next, we provide proofs of Lemmas 1 and 2.

Proof of Lemma 1. Part (b) can be proved using the similar arguments as those used in the proof of (2.14) of [14]. Now, we prove (a).

The proof of part (a) utilizes the proof of Proposition 2 in [12]. We will adopt some results given by their proof and outline only the required steps. Ma et al. in their technical report have shown that if \( \| \hat{\beta} - \beta \| = O_{p}(1/\sqrt{n}) \), then, for \( \xi \in I, I \) is the support of the density function of \( \xi_i \),

\[ \sup_{\xi \in I} |\hat{\nu}_{n}(\xi) - v(\xi)| = O_{p} \left\{ \left( \frac{\log n}{nh} \right)^{1/2} + h^{1/2} \right\} \tag{12} \]

where \( \hat{\nu}_{n}(\xi) \) is the nonparametric variance estimation of \( v(\xi) \) using the estimated sequences \( \hat{\xi}_{i} \) and \( \hat{\xi}_{i} \). Finally, \( \hat{w}(X, Z) = \hat{v}_{n}(\xi)^{-1} \) estimates \( w \), the inverse of the true variance function \( v \). It is guaranteed that \( \hat{w} = w + O_{p}(\log n/(nh))^{1/2} + h^{1/2} = \hat{v}_{n}(\xi) \) always holds. Using the usual kernel regression with the same bandwidth \( h \) in all the regressions, we can obtain

\[ \hat{E}(X|Z) = E(X|Z) \left( 1 + a_{x} h^{2} + b_{x} \frac{1}{\sqrt{nh}} \right) \]

\[ \hat{E}(Y|Z) = E(Y|Z) \left( 1 + a_{y} h^{2} + b_{y} \frac{1}{\sqrt{nh}} \right) \]

\[ \hat{E}(\hat{w}|Z) = E(\hat{w}|Z) \left( 1 + a_{\hat{w}} h^{2} + b_{\hat{w}} \frac{1}{\sqrt{nh}} \right) \]

\[ \hat{E}(\hat{w}X|Z) = E(\hat{w}X|Z) \left( 1 + a_{\hat{w}x} h^{2} + b_{\hat{w}x} \frac{1}{\sqrt{nh}} \right) \]

\[ \hat{E}(\hat{w}Y|Z) = E(\hat{w}Y|Z) \left( 1 + a_{\hat{w}y} h^{2} + b_{\hat{w}y} \frac{1}{\sqrt{nh}} \right) \]

where \( a_{x}, a_{y}, a_{\hat{w}x}, a_{\hat{w}y}, b_{x}, b_{y}, b_{\hat{w}x}, b_{\hat{w}y} \) are functions of \( Z \) with constant order. Let

\[ b_{i}^{x} = -E(\hat{w}_{i}X_{i}|Z) \frac{(a_{\hat{w}x} - a_{w}) h^{2} + (b_{\hat{w}x} - b_{w}) h}{1 + a_{w} h^{2} + b_{w} h^{1/2}} \]

and

\[ b_{i}^{y} = -E(\hat{w}_{i}Y_{i}|Z) \frac{(a_{\hat{w}y} - a_{w}) h^{2} + (b_{\hat{w}y} - b_{w}) h}{1 + a_{w} h^{2} + b_{w} h^{1/2}} \]

Then, we have

\[ W_{ni} = \left\{ Y_{i} - \frac{\hat{E}(\hat{w}_{i}Y_{i}|Z)}{\hat{E}(\hat{w}_{i}|Z)} \right\} - \left\{ X_{i} - \frac{\hat{E}(\hat{w}_{i}X_{i}|Z)}{\hat{E}(\hat{w}_{i}|Z)} \right\}^{T} \beta \]

\[ = \left\{ Y_{i} - \frac{E(\hat{w}_{i}Y_{i}|Z)}{E(\hat{w}_{i}|Z)} \right\} - \left\{ X_{i} - \frac{E(\hat{w}_{i}X_{i}|Z)}{E(\hat{w}_{i}|Z)} \right\}^{T} \beta + (b_{i}^{x} - b_{i}^{y} \beta) \]

\[ = \left\{ Y_{i} - \frac{E(w_{i}Y_{i}|Z)}{E(w_{i}|Z)} \right\} - \left\{ X_{i} - \frac{E(w_{i}X_{i}|Z)}{E(w_{i}|Z)} \right\}^{T} \beta \cdot w_{i} \left\{ X_{i} - \frac{E(w_{i}X_{i}|Z)}{E(w_{i}|Z)} \right\} + r(X_{i}, Z_{i}, Y_{i}) o_{p}(1) \]

\[ = o_{p}(1), \]
where \( r(X_i, Z_i, Y_i) \) is some function of \((X_i, Z_i, Y_i)\) with finite second moment and \( o_p(1) \) is independent of \((X_i, Z_i, Y_i)\). Therefore,

\[
\max_{1 \leq i \leq n} |W_{ni}| \leq \max_{1 \leq i \leq n} \left| \varepsilon_i w_i \left\{ X_i - \frac{E(w_i X_i | Z_i)}{E(w_i | Z_i)} \right\} \right| + \max_{1 \leq i \leq n} |r(X_i, Z_i, Y_i)| o_p(1) \\
= o_p(n^{1/2}) + o_p(n^{1/2}) o_p(1) \\
= o_p(n^{1/2}).
\]

Lemma 1(a) is proved.

**Proof of Lemma 2.** In fact, Lemma 2(b) and (c) can be proved using the same arguments in the proof of Proposition 2 of [12]. We give the proof of Lemma 2(a) only. Taylor’s expansion of \( \mathcal{L}(\beta) \) in (10) with respect to \( \lambda^T W_{ni} \) gives

\[
\mathcal{L}(\beta) = 2 \sum_{i=1}^{n} \left( \lambda^T W_{ni} - (1/2) (\lambda^T W_{ni})^2 \right) + R_n,
\]

where \( R_n \), in probability, satisfies the following inequality in light of Lemma 1 and Lemma 2(c) for some constant \( C > 0 \),

\[
|R_n| \leq C \sum_{i=1}^{n} (\lambda^T W_{ni})^3 \leq C |\lambda|^3 \max_{1 \leq i \leq n} |W_{ni}| \sum_{i=1}^{n} |W_{ni}|^2 = o_p(1).
\]

By Lemma 1, Lemma 2(c) and similar arguments as above, we obtain

\[
\sum_{i=1}^{n} \frac{(\lambda^T W_{ni})^3}{1 + \lambda^T W_{ni}} = o_p(1).
\]

By (9), we obtain

\[
0 = \sum_{i=1}^{n} \frac{\lambda^T W_{ni}}{1 + \lambda^T W_{ni}} = \sum_{i=1}^{n} (\lambda^T W_{ni}) - \sum_{i=1}^{n} (\lambda^T W_{ni})^2 + \sum_{i=1}^{n} \frac{(\lambda^T W_{ni})^3}{1 + \lambda^T W_{ni}}.
\]

By (14) and (15), we obtain

\[
\sum_{i=1}^{n} (\lambda^T W_{ni}) = \sum_{i=1}^{n} (\lambda^T W_{ni})^2 + o_p(1).
\]

Again by (9), we obtain

\[
0 = \sum_{i=1}^{n} \frac{W_{ni}}{1 + \lambda^T W_{ni}} = \sum_{i=1}^{n} \left[ 1 - \lambda^T W_{ni} + \frac{(\lambda^T W_{ni})^2}{1 + \lambda^T W_{ni}} \right] \\
= \sum_{i=1}^{n} W_{ni} - \sum_{i=1}^{n} \left( W_{ni} W_{ni}^T \right) \lambda + \sum_{i=1}^{n} \frac{W_{ni}(\lambda^T W_{ni})^2}{1 + \lambda^T W_{ni}}.
\]

By Lemma 1 and Lemma 2(c), we obtain

\[
n^{-1} \sum_{i=1}^{n} \frac{|W_{ni}(\lambda^T W_{ni})^2|}{1 + \lambda^T W_{ni}} \leq C |\lambda|^2 \max_{1 \leq i \leq n} |W_{ni}| n^{-1} \sum_{i=1}^{n} |W_{ni}|^2 = o_p(n^{-1/2}).
\]

Hence, we have

\[
\lambda = \left( \sum_{i=1}^{n} W_{ni} W_{ni}^T \right)^{-1} \sum_{i=1}^{n} W_{ni} + \left( n^{-1} \sum_{i=1}^{n} W_{ni} W_{ni}^T \right)^{-1} \left( n^{-1} \sum_{i=1}^{n} W_{ni}(\lambda^T W_{ni})^2 \right) \\
= \left( \sum_{i=1}^{n} W_{ni} W_{ni}^T \right)^{-1} \sum_{i=1}^{n} W_{ni} + o_p(n^{-1/2}).
\]

By (13), (16) and (17), finally we obtain

\[
\mathcal{L}(\beta) = \sum_{i=1}^{n} \lambda^T W_{ni} W_{ni}^T \lambda + o_p(1) \\
= \left( n^{-1/2} \sum_{i=1}^{n} W_{ni} \right)^T \left( n^{-1} \sum_{i=1}^{n} W_{ni} W_{ni}^T \right)^{-1} \left( n^{-1/2} \sum_{i=1}^{n} W_{ni} \right) + o_p(1). \quad \square
\]
References