

Statistica Sinica Preprint No: SS-2016-0399.R2

Title	Composite Estimation: An Asymptotically Weighted Least Squares Approach
Manuscript ID	SS-2016-0399.R2
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202016.0399
Complete List of Authors	Lu Lin Feng Li Kangning Wang and Lixing Zhu
Corresponding Author	Lixing Zhu
E-mail	lzhu@hkbu.edu.hk
Notice: Accepted version subject to English editing.	

COMPOSITE ESTIMATION: AN ASYMPTOTICALLY WEIGHTED LEAST SQUARES APPROACH

Lu Lin^{1,5}, Feng Li², Kangning Wang³ and Lixing Zhu^{4,6}

¹*Shandong University, Jinan, China*

²*Zhengzhou University, Zhengzhou, China*

³*Shandong Technology and Business University, Yantai, China*

⁴*Hong Kong Baptist University, Hong Kong, China*

⁵*Qufu Normal University, Qufu, China*

⁶*Beijing Normal University, Beijing, China*

Abstract: The purpose of this paper is three-fold. First, based on the asymptotic presentation of initial estimators, and model-independent parameters either hidden in the model or combined with the initial estimators, a pro forma linear regression between the initial estimators and the parameters is defined in an asymptotic sense. Then a weighted least squares estimation is constructed within this framework. Second, systematic studies are conducted to examine when both variance and bias reductions can be achieved simultaneously and when only variance can be reduced. Third, a generic rule of constructing composite estimation and unified theoretical properties are introduced. Some important examples such as quantile regression, nonparametric kernel estimation, blockwise empirical likelihood estimation are investigated in detail to explain the methodology and theory. Simulations are conducted to examine its performance in finite sample situations and a real dataset is analysed for illustration. The comparison with existing competitors is also made.

Key words and phrases: Asymptotic representation, model-independent parameter, weighted least squares, composite quantile regression, nonparametric regression.

1. Introduction

1.1 Motivation and existing methodologies

In the area of point estimation, how to promote estimation efficiency in various models remains an important issue. Recently, composition methodologies have received much attention in the literature. The main goal of these methodologies focuses on estimation variance reduction. Zou and Yuan (2008) proposed a composite quantile linear regression to reduce asymptotic variance. Kai, Li and Zou (2010) extended it to construct a variance-reduced nonparametric regression estimation. For further developments of this methodology in semiparametric settings, see Kai, Li and Zou (2011). To achieve variance-reduction as well as robustness, Bradic, Fan and Wang (2011) introduced a penalized composite quasi-likelihood for ultrahigh dimensional variable selection by combining several convex loss functions, together with a weighted L_1 -penalty. As a common purpose of these methodologies is to reduce estimation variance, we call them *the variance-reduction methodologies*.

Two common approaches to construct composite estimator can be reviewed as follows. One is to directly define a weighted sum of initial estimators as a composite estimator:

$$\tilde{\theta} = \sum_{k=1}^m w_k \hat{\theta}_k, \quad (1.1)$$

if a set of initial estimators $\hat{\theta}_k$ of the parameter of interest θ can be defined. We call the estimation method of (1.1) *the direct composition*. The estimation efficiency could be achieved by properly selecting weights under a criterion such as minimization of the estimation variance, see, e.g. Koenker

(1984) and Kai, Li and Zou (2010). More generally, minimizing user-chosen risk such as mean squared error can be adopted for this purpose, see, e.g. Lavancier and Rochet (2016) and the references therein. Another similar methodology, called the aggregation estimation, mimics estimation weighted averaging. The resulting composite estimator is approximately at least as good as the best linear or convex combination of initial estimators, see for instance Juditsky and Nemirovski (2000) and Rigollet and Tsybakov (2007), and the references therein.

Note that when the risk is chosen to be mean squared error, the corresponding methodology could reduce either estimation bias or estimation variance or both in a balanced manner. But when the biases of initial estimators are of the same magnitude, this methodology often fails to reduce bias unless the weights are chosen to be negative (see Rigollet and Tsybakov 2007) or some strong constraints are assumed on initial estimators (see Sun, Gai and Lin, 2013).

The second method defines a composite estimator via minimizing a weighted sum of objective functions, namely, the estimator can be expressed as

$$\tilde{\theta} = \arg \min_{\theta \in \Theta} \sum_{k=1}^m w_k g_k(Z, \theta), \quad (1.2)$$

provided that the predetermined objective functions $g_k(Z, \theta)$, $k = 1, \dots, m$, contain the same parameter θ . We call it *the objective function composition*. For example, Zou and Yuan (2008) suggested this method for linear quantile regression. In their method, different objective functions g_k are related to different quantiles τ_k , but the parameter θ of interest is free of τ_k . Compared with the estimator of θ obtained via a single quantile τ , the com-

posite estimator can reduce the estimation variance when the weights w_k are properly selected. However, this method cannot be extended to handle many other problems. As an example, for nonparametric quantile regression, we have no chance to get a weighted sum of the objective functions in (1.2) so that the parameter of interest is free of the quantiles τ . Actually, for different α_k (the $100\tau_k\%$ quantile of the model error), the parameters in the objective functions $g_k(Z, \theta_k)$ are $\theta_k = r(x) + \alpha_k$. Although the nonparametric regression function $r(x)$ is what we want to estimate, it is not easy to separate $r(x)$ and α_k (see Kai, Li and Zou, 2010). Sun, Gai and Lin (2013) showed that the weights in the above composition asymptotically play no role in estimation efficiency enhancement, and the bias cannot be reduced to have a faster convergence rate to zero.

1.2 The contributions of the new method

To explore the new methodology, we observe a common feature in several cases. That is, a model-independent parameter, say τ , plays a crucial role in the procedure of constructing a set of initial estimators. This parameter is not of interest for us to estimate, whereas with different values τ_k of the parameter τ , several initial estimators $\hat{\theta}_{\tau_k}$ for the parameter θ of interest can be defined. Then, the first question we need to answer is how to find a model-independent parameter for this purpose. In some scenarios, it is hidden in the model such as the quantile in quantile regression. But in some other scenarios, particularly in semiparametric and nonparametric setups, such a parameter does not exist in the model. But we will find it from estimation procedure. We now list a few examples: the quantile in the parametric and nonparametric quantile regression estimators (Zou

and Yuan, 2008; Kai, Li and Zou, 2010); the bandwidth in the Nadaraya-Watson kernel estimator (N-W estimator) and the local linear estimator for nonparametric regression (Fan and Gijbels, 1996); the size of block in the blockwise likelihood (Kitamura, 1997; Lin and Zhang, 2001). For more details see the examples given in Section 3.

In this paper, we establish a unified relationship between the estimation and the model-independent parameter under a generic framework. The basic idea for establishing the relationship is to use the asymptotic representation of the initial estimator. Specifically, we use (or define) the model-independent parameter and the corresponding initial estimators to build a pro forma linear regression model.

Then we construct a composite estimator that is a weighted least squares estimator through the linear regression model. We call the method *the asymptotically weighted least squares* (AWLS) and the resultant estimation *the AWLS estimation*. The details will be presented in Sections 2 and 3.

From the above description, We will see in the later development that this method has some desirable features.

1. (*Generality*) The AWLS can be constructed as long as an estimator has an asymptotically linear representation with a known function of model-independent parameter.
2. (*Variance reduction*) By selecting proper weights, the AWLS estimation could be asymptotically more efficient than those obtained by existing composite methods such as the composite maximum likelihood and the composite least squares.
3. (*Bias reduction*) The AWLS can, in some cases, reduce estimation

bias to accelerate the convergence rate. Nonparametric estimation is an example.

4. (*Generic rule*) More importantly, the results explain how the composition depends on the structure of asymptotic representation. This has never been explored in the literature before. From the construction, we can know in which cases the AWLS can achieve both bias and variance reduction and in which cases, it can only reduce variance.

The remainder of this paper is organized as follows. In Section 2, a unified framework of the AWLS is introduced, and a generic rule of the AWLS and its theoretical property are investigated. In Section 3, two typical models, the linear quantile regression and nonparametric regression, are used as examples to illustrate the study described in the previous section, and blockwise empirical likelihood is also briefly discussed. Numerical studies including simulation study and real data analysis are given in Section 4 and the proofs of the theorems are postponed to the Supplement.

2. A generic framework for AWLS

In this section, we first introduce a generic framework for the construction of the AWLS estimation and then investigate its theoretical properties in different scenarios. The verifiable examples will be given in the next section.

2.1 Models and estimations

Suppose that for given m values $\tau_k, k = 1, \dots, m$, of a model-independent parameter τ , m initial estimators $\hat{\theta}_{\tau_k}$ of the parameter θ of interest depend on τ_k respectively, and have the following asymptotic representation:

$$\hat{\theta}_{\tau_k} = \theta + b_n \xi_n(\tau_k) + \epsilon_n(\tau_k), \quad k = 1, \dots, m. \quad (2.1)$$

Here n is the sample size, the random variable $\xi_n(\tau)$ is a known function of τ satisfying $\xi_n(\tau) = O_p(1)$, b_n is independent of τ and is an infinitesimal of lower order than the order of $\epsilon_n(\tau)$ in probability. The convergence rate of $\hat{\theta}_{\tau_k} - \theta$ is then of the order $O_p(b_n)$ for all $k = 1, \dots, m$. The framework in (2.1) sets a pro forma linear model with response variables $\hat{\theta}_{\tau_k}$, covariates τ_k (or $\xi_n(\tau_k)$), intercept θ and model error $\epsilon_n(\tau_k)$. The intercept θ is actually the parameter of interest.

This formula has four possible combinations: b_n is either known or unknown, and $\xi_n(\tau)$ is either free of θ or dependent on θ . When this artificial covariate ξ_n in (2.1) is related to θ , we write it as $\xi_n = \xi_n(\tau, \theta)$ for clarity. An initial estimator $\hat{\theta}$ is then required to replace θ . In this case, denote $\hat{\xi}_n(\tau) = \xi_n(\tau, \hat{\theta})$. We will see that the different combinations will lead to different asymptotic properties for the corresponding AWLS estimator. In the following, we will separately consider two different cases when b_n is known or unknown as the corresponding AWLS estimators have different expressions. But for ξ_n , we only give the estimators for the case with ξ_n depending on θ . When ξ_n is free of θ , the AWLS estimators have the same forms if $\hat{\xi}_n$ is replaced by ξ_n .

Case 1. (b_n is unknown). An AWLS estimator $\tilde{\theta}$ of θ can be constructed as the first component of the following minimizers:

$$\begin{pmatrix} \tilde{\theta} \\ \tilde{b}_n \end{pmatrix} = \arg \min_{\theta, b_n} \frac{1}{m} \sum_{k=1}^m w_k (\hat{\theta}_{\tau_k} - \theta - b_n \hat{\xi}_n(\tau_k))^2, \quad (2.2)$$

where $w_k, k = 1, \dots, m$, are weights satisfying $\sum_{k=1}^m w_k = 1$. The estimator

has the following closed form:

$$\tilde{\theta} = \sum_{k=1}^m w_k \hat{\theta}_{\tau_k} - \hat{b}_n \bar{\xi}_n, \quad (2.3)$$

where $\bar{\xi}_n = \sum_{k=1}^m w_k \hat{\xi}_n(\tau_k)$ and $\hat{b}_n = \frac{\sum_{k=1}^m w_k \hat{\theta}_{\tau_k} (\hat{\xi}_n(\tau_k) - \bar{\xi}_n)}{\sum_{k=1}^m w_k (\hat{\xi}_n(\tau_k) - \bar{\xi}_n)^2}$.

Case 2 (b_n is known). By the weighted least squares, the AWLS estimator can be simply expressed as

$$\tilde{\theta} = \sum_{k=1}^m w_k \left(\hat{\theta}_{\tau_k} - b_n \hat{\xi}_n(\tau_k) \right), \quad (2.4)$$

where $w_k, k = 1, \dots, m$, are weights satisfying $\sum_{k=1}^m w_k = 1$.

2.2 Properties

We now investigate the asymptotic properties of the AWLS estimators defined in (2.3) and (2.4).

2.2.1 Convergence rate

First, consider the case where $\xi_n(\tau)$ is free of θ . We define regenerated weights as

$$\tilde{w}_k = w_k - \bar{\xi}_n \frac{w_k (\xi_n(\tau_k) - \bar{\xi}_n)}{\sum_{k=1}^m w_k (\xi_n(\tau_k) - \bar{\xi}_n)^2}, \quad k = 1, \dots, m. \quad (2.5)$$

They are free of the initial estimators and still satisfy $\sum_{k=1}^m \tilde{w}_k = 1$, but not necessary to be positive. We have the following theorem.

THEOREM 2.1. When $\xi_n(\tau)$ is free of θ , the AWLS estimators $\tilde{\theta}$ defined in (2.3) and (2.4) satisfy

$$\tilde{\theta} - \theta = \sum_{k=1}^m \tilde{w}_k \epsilon_n(\tau_k),$$

where $\epsilon_n(\tau_k)$ are the error terms in the asymptotic representation defined in (2.1).

REMARK 2.1. The theorem gives an important conclusion: when ξ_n is free of the parameter of interest, the convergence rate of the AWLS estimator can be accelerated. More precisely, $\tilde{\theta} - \theta$ has the same convergence rate as that of the error term $\epsilon_n(\tau)$.

Now consider the case where $\xi_n(\tau)$ depends on θ . We need the following condition:

(C1) There are constants $c_1 > 0$ and $c_2 > 0$, such that when n is large enough, $c_1 \leq |b_n \xi'_n(\tau, \theta)| \leq c_2$ and $|b_n \xi''_n(\tau, \theta)| \leq c_2$ in probability, where $\xi'_n(\tau, \theta)$ and $\xi''_n(\tau, \theta)$ are respectively the first- and second-order partial derivatives of $\xi_n(\tau, \theta)$ with respect to θ .

Condition (C1) is usually mild. For example, when the asymptotic representation (2.1) is obtained by Bahadur representation or asymptotic linear estimation (van der Vaart (1998) and Bickel (1998)), this condition holds under some regularity conditions.

THEOREM 2.2. When $\xi_n(\tau)$ depends on θ , and condition (C1) holds, then, both the AWLS estimators $\tilde{\theta}$ in (2.3) and (2.4) have the same convergence rate in probability as that of the initial estimator $\hat{\theta}_\tau$.

REMARK 2.2. Theorem 2.1, Theorem 2.2 and Theorem 2.3 given below show that the asymptotic representation can determine whether an AWLS estimator would have both bias and variance reduction. We may choose the one with ξ_n being free of the parameter of interest if possible.

2.2.2 Variance reduction

Now we consider the variance reduction issue. When $\xi_n(\tau)$ is free of θ , Theorem 2.1 shows that the AWLS estimator has a faster convergence rate

than the initial estimator and thus is of course variance reduced, asymptotically. Consider the case when $\xi_n(\tau)$ depends on θ . The following condition is assumed:

(C2) There is a function $g(\tau)$ such that $g(\tau) \neq 0$ and $b_n \xi'_n(\tau) = g(\tau) + O_p(b_n)$.

From model (2.1) we can see that this condition is mild as well. Under parametric situation, for instance, $b_n = 1/\sqrt{n}$ and $g(\tau)$ is the expectation of $b_n \xi'_n(\tau)$.

Let $\mathbf{w}_g = (w_1 g(\tau_1), \dots, w_m g(\tau_m))^T$ and $\mathbf{1}$ be an m -dimensional column vector with all components 1. We have the following theorem.

THEOREM 2.3. If $\xi_n(\tau)$ depends on θ , and (C1) and (C2) hold, then, the AWLS estimators $\tilde{\theta}$ defined in (2.3) and (2.4) satisfy

$$\tilde{\theta} = - \sum_{k=1}^m w_k g(\tau_k) (\hat{\theta}_{\tau_k} - \theta) + b_n O_p(\hat{\theta}_{\tau_k} - \theta) + o_p(\hat{\theta}_{\tau_k} - \theta) + \epsilon_n(\tau_k),$$

where $b_n O_p(\hat{\theta}_{\tau_k} - \theta) + o_p(\hat{\theta}_{\tau_k} - \theta) + \epsilon_n(\tau_k)$ is an infinitesimal of higher order than the first term. Particularly, if θ is a scale parameter of interest, the asymptotic variance of $\sqrt{n} \tilde{\theta}$ defined in (2.3) and (2.4) can be expressed as

$$\lim_{n \rightarrow \infty} n \text{Var}(\tilde{\theta}) = \mathbf{w}_g^T \lim \Sigma_{\hat{\theta}} \mathbf{w}_g,$$

where $\lim \Sigma_{\hat{\theta}}$ is the asymptotic covariance matrix of $\sqrt{n}(\hat{\theta}_{\tau_1}, \dots, \hat{\theta}_{\tau_m})^T$. Moreover, the optimal weight vector (written as \mathbf{w}^*) has the form: $\mathbf{w}^* = (\mathbf{1}^T (\lim \Sigma_{\hat{\theta}})^{-1} \mathbf{1})^{-1} (\lim \Sigma_{\hat{\theta}})^{-1} \mathbf{1}$, and then, $\lim_{n \rightarrow \infty} \text{Var}(\tilde{\theta}) \leq \lim_{n \rightarrow \infty} \text{Var}(\hat{\theta}_{\tau_k})$ for $k = 1, \dots, m$.

For this theorem, we have the following remark.

REMARK 2.3. (a) Optimal weights. In Theorem 2.3, the optimal weight vector $\mathbf{w}^* = (w_1^*, \dots, w_m^*)^T = (\mathbf{1}^T (\lim \Sigma_{\hat{\theta}})^{-1} \mathbf{1})^{-1} (\lim \Sigma_{\hat{\theta}})^{-1} \mathbf{1}$ is related to the unknown covariance matrix $\lim \Sigma_{\hat{\theta}}$ and vector $\mathbf{w}_g = (w_1 g(\tau_1), \dots, w_m g(\tau_m))^T$. They can be consistently estimated by classical methods such as jackknife (see, e.g., Shao and Wu (1989)).

(b) Weight selection under multivariate θ case. For scalar θ , a closed representation for the optimal weight vector \mathbf{w}^* has been derived in Theorem 2.3. When θ is a vector, we can see that

$$\lim_{n \rightarrow \infty} Cov(\tilde{\theta}) = \sum_{j=1}^m \sum_{k=1}^m w_j g(\tau_j) w_k g(\tau_k) \lim_{n \rightarrow \infty} Cov(\hat{\theta}_{\tau_j}, \hat{\theta}_{\tau_k}).$$

In general, a closed solution for the optimal weight may not be attained, unless numerical approximation is adopted. However, if the initial estimators satisfy

$$HCov(\hat{\theta}_{\tau_k}, \hat{\theta}_{\tau_j})H \rightarrow a_{kj}D, \quad (2.6)$$

where $H = \text{diag}(n^{\delta_1}, \dots, n^{\delta_p})$ with $0 < \delta_j \leq 1/4$, a_{jk} are constants and D is a positive definite matrix, and both are given or estimable, we can get a closed solution. For example, the asymptotic covariances of the quantile regression estimators satisfy this, see the results in Section 3. Under this situation, by the same argument as used above, the closed representation for the optimal weight vector is $\mathbf{w}^* = (\mathbf{1}^T D^{-1} \mathbf{1})^{-1} D^{-1} \mathbf{1}$. When (2.6) does not hold, the following suboptimal weights can be considered. Note that $\lim_{n \rightarrow \infty} tr(Cov(\tilde{\theta})) = \sum_{j=1}^m \sum_{k=1}^m w_j g(\tau_j) w_k g(\tau_k) \lim_{n \rightarrow \infty} tr(Cov(\hat{\theta}_{\tau_j}, \hat{\theta}_{\tau_k}))$. A suboptimal weight vector can be obtained as $\mathbf{w}_S^* = (\mathbf{1}^T A_S^{-1} \mathbf{1})^{-1} A_S^{-1} \mathbf{1}$, where

$$A_S = \left(\lim_{n \rightarrow \infty} tr(Cov(\hat{\theta}_{\tau_j}, \hat{\theta}_{\tau_k})) \right)_{j,k=1}^p.$$

2.3 Choices of m and values of τ

For practical use, we must choose the number m of initial estimators to be combined and the values of model-independent parameter τ . m can be regarded as a tuning parameter, as its choice influences the performance of the AWLS estimator. It seems challenging to have a criterion to select an optimal m and values of τ , as they appear model-dependent. Thus, the choices in the present paper are empirical. As was shown by Zou and Yuan (2008), for the composite quantile regression, a number around 19 of m would be practically useful. Thus, the equally spaced quantiles $\tau_k = k/(m+1)$ amounts to using the 5%, 10%, \dots , 95% quantiles. In the simulations, we see that the AWLS estimator for the composite quantile regression is not sensitive to the choice of m and τ_k . If τ is the bandwidth h in the kernel estimation we will discuss in the next section or the number of knots in the B-spline estimation, the AWLS estimator is reasonably affected by the choice of h because the bandwidth often affects the performance of nonparametric estimation. However, we show that the AWLS is not very sensitive to the number m when h_k 's are around the optimal bandwidth. In practical use, we may determine a data-driven bandwidth first and then take values τ_k so that h_k 's are around it. In the simulation studies in Section 4, we will discuss this issue in more detail.

On the other hand, although we have a generic framework for the use of composition, the model-independent parameter selection is in general a challenge because it relies on the asymptotic presentation of initial estimator and the relationship between such a parameter and initial estimators. It deserves a further study to see whether there is a general way to select this parameter even when the user has little knowledge about the asymptotics.

3. Some examples

In this section, we use the linear quantile, nonparametric regressions and blockwise empirical likelihood as examples to verify the methods and theory proposed in the previous section. Some conclusions given below can be thought of as direct corollaries of those proposed in Section 2. But for validation and further discussion, we still give the detailed conclusions and proofs. Moreover, for these specific models, some special results are obtained.

3.1 AWLS for linear quantile regression

Consider the following linear regression model:

$$Y = \beta^T X + e.$$

Suppose that the conditional 100 τ % quantile of $Y|X$ can be expressed as the following linear regression form: $\beta^T X + \alpha_\tau$, where α_τ is the 100 τ % quantile of $Y - \beta^T X$. See Koenker (2005) for the details. The quantile regression estimator of $(\alpha_\tau, \beta^T)^T$ can be obtained as

$$\begin{pmatrix} \hat{\alpha}_\tau \\ \hat{\beta}_\tau \end{pmatrix} = \arg \min_{\alpha_\tau, \beta} \sum_{i=1}^n \rho_\tau(Y_i - \alpha_\tau - \beta^T X_i),$$

where $\rho_\tau(t) = \tau t_+ + (1 - \tau)t_-$ is the so-called check function with $+$ and $-$ standing for positive and negative parts, respectively. Denote $F_i(y) = F(y|X_i) = P(Y_i < y|X_i)$ and suppose that $F_i(y), i = 1, \dots, n$, are i.i.d. with a common density function $f(y) > 0$ for all y . Under some regularity conditions (see, e.g., Bahadur 1966; Kiefer 1967; Koenker 2005), for different quantile positions $\tau = \tau_k, k = 1, \dots, m$, we have the following Bahadur

representation:

$$\begin{aligned}\hat{\beta}_{\tau_k} &= \beta + \frac{1}{f_e(\alpha_{\tau_k})n} D_n^{-1} \sum_{i=1}^n X_i(\tau_k - I(Y_i \leq \alpha_{\tau_k} + \beta^T X_i)) + \epsilon_n(\tau_k) \\ &=: \beta + b_n \xi_n(\tau_k) + \epsilon_n(\tau_k), \quad k = 1, \dots, m,\end{aligned}\quad (3.1)$$

where $D_n = \frac{1}{n} \sum_{i=1}^n X_i X_i^T$, $f_e(\cdot)$ is the density function of error $e = Y - \beta^T X$, $\epsilon_n(\tau_k)$, $k = 1, \dots, m$, are of order $O_p(n^{-3/4})$, and b_n and $\xi_n(\tau_k)$ are defined respectively as $b_n = 1/\sqrt{n}$ and

$$\xi_n(\tau_k) = \frac{1}{f_e(\alpha_{\tau_k})\sqrt{n}} D_n^{-1} \sum_{i=1}^n X_i(\tau_k - I(Y_i \leq \alpha_{\tau_k} + \beta^T X_i)).$$

It can be seen that $\xi_n(\tau_k) = O_p(1)$, and b_n is of order $n^{-1/2}$, an infinitesimal of lower order than that of $\epsilon_n(\tau_k)$. We first suppose $f_e(\cdot)$ is a given function. Then, the asymptotical representation (3.1) can be included into the framework of (2.1). When α_{τ_k} and β in $\xi_n(\tau_k)$ are replaced by their consistent estimators $\hat{\alpha}_{\tau_k}$ and $\hat{\beta}_{\tau_k}$ respectively, from (2.4), the AWLS estimator of β has the following form:

$$\tilde{\beta} = \sum_{k=1}^m w_k \left\{ \hat{\beta}_{\tau_k} - \frac{1}{f_e(\hat{\alpha}_{\tau_k})n} D_n^{-1} \sum_{i=1}^n X_i(\tau_k - I(Y_i \leq \hat{\alpha}_{\tau_k} + \hat{\beta}_{\tau_k}^T X_i)) \right\}. \quad (3.2)$$

By comparing the above with the Bahadur representation (3.1), we see that besides the initial estimators $\hat{\beta}_{\tau_k}$, the main term $b_n \xi_n(\tau_k)$ plays a key role in constructing the AWLS estimator in (3.2). This term is mainly related to the directional derivative of the objective function. This method can be extended to the case when the density function $f_e(\cdot)$ is unknown, but can be consistently estimated. For ease of exposition, we only present the result with a given $f_e(\cdot)$ because by the Slutsky theorem, the asymptotic distribution of $\tilde{\beta}$ is changeless when a consistent estimator of $f_e(\cdot)$ is used.

We now investigate the properties of the above AWLS estimator. To this end, we assume the following conditions:

- (C3) $\max_{1 \leq i \leq n} \|X_i\| \leq cn^\nu$ for some constants $c > 0$ and $0 \leq \nu < 1/2$, where $\|\cdot\|$ is Euclidean norm, and there exists a positive definite matrix D such that $D = \lim_{n \rightarrow \infty} D_n$.
- (C4) The density function $f_e(\cdot)$ of the error e is continuously differentiable and positive at α_{τ_k} for $k = 1, \dots, m$.

The following theorem states the asymptotic properties of the AWLS estimator.

THEOREM 3.1. Under the conditions (C3) and (C4), the AWLS estimator (3.2) has the following asymptotic representation:

$$\tilde{\beta} - \beta = D^{-1} \frac{1}{n} \sum_{i=1}^n X_i \sum_{k=1}^m \frac{w_k}{f_e(\alpha_{\tau_k})} (\tau_k - I(Y_i \leq \alpha_{\tau_k} + \beta^T X_i)) + O_p(n^{-3/4}).$$

Consequently,

$$\sqrt{n}(\tilde{\beta} - \beta) \xrightarrow{\mathcal{D}} N(0, \mathbf{w}^T A_0 \mathbf{w} D^{-1}),$$

where $\mathbf{w} = (w_1, \dots, w_m)^T$ and $A_0 = \left(\frac{\min(\tau_k, \tau_j)(1 - \max(\tau_k, \tau_j))}{f_e(\alpha_{\tau_k})f_e(\alpha_{\tau_j})} \right)_{k,j=1}^m$.

Actually, this theorem can be thought of as a corollary of Theorem 2.3 and Remark 2.3(b) because the initial estimator $\hat{\beta}_\tau$ satisfies (2.6) (see Koenker, 2005). Moreover, from the theorem, we have the following findings:

REMARK 3.1. (1) When w_k are particularly chosen as $w_k = \frac{f_e(\alpha_{\tau_k})}{\sum_{k=1}^m f_e(\alpha_{\tau_k})}$, the limiting variance of the AWLS estimator (2.2) is identical to that in Zou and Yuan (2008). In other words, we can have smaller limiting variance by

choosing proper weights. The optimal weight vector is $\mathbf{w}^* = \min_{\mathbf{1}^T \mathbf{w} = 1} \mathbf{w}^T A_0 \mathbf{w}$. Lagrange multipliers lead to the optimal weight vector and the optimal limiting variance respectively in the following closed forms:

$$\mathbf{w}^* = (\mathbf{1}^T A_0^{-1} \mathbf{1})^{-1} A_0^{-1} \mathbf{1}, \quad \mathbf{w}^{*T} A_0 \mathbf{w}^* D^{-1} = (\mathbf{1}^T A_0^{-1} \mathbf{1})^{-1} D^{-1}.$$

This is the same as the optimal weight in Remark 2.3 (b). For univariate linear regression, Koenker (1984) obtained the above estimation efficiency by the direct composition. However, in their method, the computation is not easy to implement as the optimal weights are the solutions of m nonlinear equations.

(2) When the density function f_e is unknown, the matrix A_0 can be estimated using a plug-in estimator \hat{f}_e of f_e . For example, as shown by Sun and Lin (2013), $f_e(\cdot)$ can be consistently estimated by the kernel estimator as $\hat{f}_e(t) = \frac{1}{n} \sum_{i=1}^n K_h(\hat{e}_i - t)$, where $\hat{e}_i = Y_i - \hat{\beta}^T X_i$ with $\hat{\beta}$ being a root- n consistent estimator, $K_h(t) = \frac{1}{h} K(t/h)$, $K(\cdot)$ is a kernel function and h is a bandwidth. With the plug-in estimator \hat{f}_e , the property of weight vector \mathbf{w}^* is not discussed here.

3.2 AWLS estimation for nonparametric regression

Consider the following nonparametric regression:

$$Y = r(X) + e,$$

where $r(x)$ is a smooth nonparametric regression function for $x \in [0, 1]$, and the error term satisfies $E(e|X) = 0$ and $Var(e|X) = \sigma^2$. We now consider the AWLS kernel estimator of $r(x)$ for $x \in (0, 1)$. As is known, $x \in (0, 1)$ is not a necessary constraint; that is, we use it only for simplicity of presentation. In this section, we give two types of composite estimators

such that we can explore how the estimation efficiency depends on the structure of asymptotical representation.

Type-1: Expectation-based estimator. It is well known that under certain regularity conditions such as the second-order continuous and bounded derivatives, a commonly used kernel estimator $\hat{r}_\tau(x)$ (e.g., N-W estimator) of the regression function $r(x)$ has the mean value:

$$E(\hat{r}_\tau(x)) = r(x) + \frac{1}{2} \left[r''(x) + 2 \frac{r'(x)f'_X(x)}{f_X(x)} \right] \mu_2(K)h^2 + o(h^2), \quad x \in (0, 1) \quad (3.3)$$

where $f_X(x)$ is the density function of X , $\mu_2(K) = \int u^2 K(u) du$, $K(x)$ is a kernel function and h is a bandwidth satisfying $h = \tau n^{-\eta}$ for constants $\tau > 0$ and $0 < \eta < 1$. Then, for different values of $\tau = \tau_k, k = 1, \dots, m$, we have the following asymptotic representation: for $x \in (0, 1)$ and $k = 1, \dots, m$,

$$\begin{aligned} \hat{r}_{\tau_k}(x) &= r(x) + \left[\frac{1}{2} \left\{ r''(x) + 2 \frac{r'(x)f'_X(x)}{f_X(x)} \right\} \mu_2(K) n^{-2\eta} \right] \tau_k^2 + \epsilon_n(\tau_k) \\ &=: r(x) + b_n \xi_n(\tau_k) + \epsilon_n(\tau_k), \end{aligned} \quad (3.4)$$

where $b_n = \frac{1}{2} \left\{ r''(x) + 2 \frac{r'(x)f'_X(x)}{f_X(x)} \right\} \mu_2(K) n^{-2\eta}$ and $\xi_n(\tau_k) = \tau_k^2$. Under the regularity condition in (C5) specified later, $\epsilon_n = \hat{r}_\tau(x) - E(\hat{r}_\tau(x)) + o(n^{-2\eta})$. It has a mean of order $o(n^{-2\eta})$ and a variance of order $O(n^{-(1-\eta)})$ and thus is of order $o_p(n^{-2\eta})$, provided that $0 < \eta < 1/5$. The asymptotic representation (3.4) has the same framework as in (2.1). From (2.3), the resulting AWLS estimator of $a = r(x)$ is

$$\tilde{r}_1(x) = \sum_{k=1}^m w_k \hat{r}_{\tau_k}(x) - \tilde{b}_n(x) \overline{\tau^2}, \quad (3.5)$$

where τ_k are chosen for forming bandwidths $h_k = \tau_k n^{-\eta}$, $k = 1, \dots, m$, $\tilde{b}_n(x) = \frac{\sum_{k=1}^m w_k \hat{r}_{\tau_k}(x) (\tau_k^2 - \overline{\tau^2})}{\sum_{k=1}^m w_k (\tau_k^2 - \overline{\tau^2})^2}$ and $\overline{\tau^2} = \frac{\sum_{k=1}^m w_k \tau_k^2}{\sum_{k=1}^m w_k}$. Unlike the quantile linear re-

gression, we use the expectation representation (3.4) to construct the AWL-S estimator (3.5) for nonparametric regression function. The construction procedure is relatively simple.

Type-2: Bahadur representation-based estimator. We can also use the Bahadur representation (see, e.g., Bhattacharya and Gangopadhyay 1990; Chaudhuri 1991; Hong 2003) to construct a composite estimator. Under certain regularity conditions, for different values of $\tau = \tau_k, k = 1, \dots, m$, the N-W estimators $\hat{r}_{\tau_k}(x)$ have the following Bahadur representation:

$$\begin{aligned}\hat{r}_{\tau_k}(x) &= r(x) + \frac{1}{v_{\tau_k}(x)n} \sum_{i=1}^n K_{\tau_k}(X_i - x)(Y_i - r(x)) + \epsilon_n(\tau) \\ &=: r(x) + b_n \xi_n(\tau_k) + \epsilon_n(\tau_k), \quad x \in (0, 1),\end{aligned}\quad (3.6)$$

where $b_n = 1/\sqrt{n}$, and $\xi_n(\tau_k) = \frac{1}{v_{\tau_k}(x)\sqrt{n}} \sum_{i=1}^n K_{\tau_k}(X_i - x)(Y_i - r(x))$, $v_{\tau}(x) = \int K_{\tau}(u)f_X(x + hu)du$ with $K_{\tau}(x) = h^{-1}K(x/h)$ and $h = \tau n^{-\eta}$. In this presentation, b_n is a constant and the covariates $\xi_n(\tau_k)$ are related to the function of interest $r(x)$, and again $\epsilon_n(\tau)$ is of order $O_p(n^{-3(1-\eta)/4})$. Then, the asymptotic representation (3.6) has the same framework as in (2.1). If the density $f_X(\cdot)$ is known to lead to a given v_{τ} , according to the corresponding estimator (2.4), the resulting AWLS estimator can be expressed as

$$\tilde{r}_2(x) = \sum_{k=1}^m w_k \left(\hat{r}_{\tau_k}(x) - \frac{1}{v_{\tau_k}(x)n} \sum_{i=1}^n K_{\tau_k}(X_i - x)(Y_i - \hat{r}_{\tau_k}(x)) \right), \quad x \in (0, 1).\quad (3.7)$$

If $f_X(\cdot)$ is unknown, we can use its estimator instead and then get an estimator of v_{τ} .

We now investigate the asymptotic properties of the estimators in (3.5) and (3.7). Consider respectively the following two regularity conditions:

(C5) Kernel function $K(u)$ is symmetric with respect to $u = 0$, and satisfies $\int K(u)du = 1$, $\int u^2 K(u)du < \infty$ and $\int u^2 K^2(u)du < \infty$. Regression function $r(x)$ defined above and density function $f_X(x)$ of X have the second-order continuous and bounded derivatives and $f_X(x) > 0$ for all x .

(C6) Kernel function $K(u)$ is symmetric with respect to $u = 0$, and satisfies $\int K(u)du = 1$, $\int u^4 K(u)du < \infty$ and $\int u^2 K^2(u)du < \infty$. Functions $r(x)$ and $f_X(x)$ have the fourth-order continuous and bounded derivatives and $f_X(x) > 0$ for all x .

Denote $s_k(\mathbf{w}) = 1 - \frac{\bar{\tau}^2(\tau_k^2 - \bar{\tau}^2)}{\sum_{k=1}^m w_k(\tau_k^2 - \bar{\tau}^2)^2}$, $g_k = w_k - \bar{\tau}^2 \frac{w_k(\tau_k^2 - \bar{\tau}^2)}{\sum_{k=1}^m w_k(\tau_k^2 - \bar{\tau}^2)^2}$, $A_1(\mathbf{w}) = \left(\frac{s_k(\mathbf{w})s_j(\mathbf{w})}{\tau_k\tau_j} \int K\left(\frac{u}{\tau_k}\right)K\left(\frac{u}{\tau_j}\right)du \right)_{k,j=1}^m$, $A_2 = \left(\frac{1}{\tau_k\tau_j} \int K\left(\frac{u}{\tau_k}\right)K\left(\frac{u}{\tau_j}\right)du \right)_{k,j=1}^m$.

The following theorem states some interesting results.

THEOREM 3.2. Suppose $h_k = \tau_k n^{-\eta}$, $k = 1, \dots, m$.

(1) Under Condition (C5) or (C6), if $0 < \eta < 1/5$, then, accordingly there is an $c_n(x) = o(n^{-2\eta})$ or $c_n(x) = n^{-4\eta}c(x) \sum_{k=1}^m g_k \tau_k^4$ with $c(x)$ being a known function, the AWLS estimator $\tilde{r}_1(x)$ in (3.5) achieves the following asymptotic normality:

$$\sqrt{n^{1-\eta}} \left(\tilde{r}_1(x) - r(x) - c_n(x) \right) \xrightarrow{\mathcal{D}} N \left(0, \mathbf{w}^T A_1(\mathbf{w}) \mathbf{w} \frac{\sigma^2}{f_X(x)} \right), \quad x \in (0, 1).$$

(2) For the AWLS estimator $\tilde{r}_2(x)$ in (3.7), under Condition (C6), if $1/5 \leq \eta < 1$, then

$$\sqrt{n^{1-\eta}} \left(\tilde{r}_2(x) - r(x) - n^{-2\eta}d(x) \sum_{k=1}^m w_k \tau_k^2 \right) \xrightarrow{\mathcal{D}} N \left(0, \mathbf{w}^T A_2 \mathbf{w} \frac{\sigma^2}{f_X(x)} \right), \quad x \in (0, 1),$$

where $d(x)$ is a given function.

By the theorem, we present the following conclusions.

REMARK 3.2. (a) *Rate-accelerated convergence.* Note that under Condition (C5) $\tilde{r}_1(x)$ achieves a rate-accelerated bias $o(n^{-2\eta})$ rather than the classical optimal rate $O(n^{-2\eta})$ the N-W estimator achieves. Under Condition (C5), when the optimal bandwidth $h = O(n^{-1/9})$ is used, $\tilde{r}_1(x)$ has the convergence rate of $O(n^{-4/9})$ without higher-order smoothness conditions on the regression and density functions, and more importantly, without higher-order kernel. But the classical N-W estimator requires these to reach the convergence rate of order $O(n^{-4/9})$. This illustrates the conclusion about the convergence rate acceleration in Theorem 2.1. We also show later that by choosing proper weight, the AWLS estimator $\tilde{r}_1(x)$ can have smaller variance as well. In contrast, $\tilde{r}_2(x)$ cannot have a faster convergence rate while the estimation variance can be reduced.

(b) *Weight selection.* Invoking the same argument as in Remark 3.1, the optimal weight vector for the second estimator $\tilde{r}_2(x)$ has a closed form:

$$\mathbf{w}_2^* = (\mathbf{1}^T A_2^{-1} \mathbf{1})^{-1} A_2^{-1} \mathbf{1}.$$

It is easy to compute when τ_k and kernel function $K(\cdot)$ are given. However, the definition right before Theorem 3.2 tells that $A_1(\mathbf{w})$ of the expectation-based estimator $\tilde{r}_1(x)$ depends on the weight vector \mathbf{w} as well. Thus, the corresponding optimal weight vector for $\tilde{r}_1(x)$ has no closed form. To handle the problem, we approximate $A_1(\mathbf{w})$ by

$$A_1 = \left(\frac{s_k s_j}{\tau_k \tau_j} \int K\left(\frac{u}{\tau_k}\right) K\left(\frac{u}{\tau_j}\right) du \right)_{k,j=1}^m,$$

where $s_k = 1 - \frac{\tau^2(\tau_k^2 - \tau^2)}{\sum_{k=1}^m (\tau_k^2 - \tau^2)^2}$ is free of the weight vector \mathbf{w} . A “suboptimal” weight vector for $\tilde{r}_1(x)$ is then

$$\mathbf{w}_1^* = (\mathbf{1}^T A_1^{-1} \mathbf{1})^{-1} A_1^{-1} \mathbf{1}.$$

This “suboptimal weight”, \mathbf{w}_1^* , can be easily computed. With the weights \mathbf{w}_1^* and \mathbf{w}_2^* , $\sqrt{n^{1-\eta}} \tilde{r}_1(x)$ and $\sqrt{n^{1-\eta}} \tilde{r}_2(x)$ have the limiting variances as

$$(\mathbf{1}^T A_1^{-1} \mathbf{1})^{-1} \frac{\sigma^2}{f_X(x)} \quad \text{and} \quad (\mathbf{1}^T A_2^{-1} \mathbf{1})^{-1} \frac{\sigma^2}{f_X(x)}, \quad (3.8)$$

respectively. The two limiting variances could be smaller than those of the classical kernel estimators in certain scenarios. For example, when kernel function is chosen as $K(u) = e^{-\frac{u^2}{2}} / \sqrt{2\pi}$, then

$$A_1 = \left(\frac{s_k s_j}{(2\pi)^{1/2} \sqrt{\tau_k^2 + \tau_j^2}} \right)_{k,j=1}^m, \quad A_2 = \left(\frac{1}{(2\pi)^{1/2} \sqrt{\tau_k^2 + \tau_j^2}} \right)_{k,j=1}^m.$$

It is known that with this kernel function, the limiting variance of the N-W estimator is $\frac{\sigma^2}{2\sqrt{\pi}f_X(x)}$, which is just a special case of the variances in (3.8) with $m = 1$ and $\tau_1 = 1$. Thus, when $\min\{\tau_k; k = 1, \dots, m\} < 1 < \max\{\tau_k; k = 1, \dots, m\}$ and the above weights are used, the limiting variances of the AWLS estimators are smaller.

(c) *Kernel selection.* As mentioned above, the AWLS estimators can have either accelerated convergence rate or smaller limiting variance, or both. From the technical proof, we see that the estimators still have the kernel estimation types. A natural concern is whether the classical N-W estimator, or an adaption of the N-W estimator, could also enjoy this rate-acceleration property through a delicate selection of kernel function. The details of the proof tell that this is not possible and there is no such a kernel function for any single N-W estimator. This is because the AWLS estimators, particularly the expectation-based estimator $\tilde{r}_1(x)$ is not simply a weighted sum of the initial estimator with positive weights summing to one.

3.3 AWLS estimation for blockwise empirical likelihood

The values of the model-independent parameters, quantile τ and bandwidth h in the two examples above, can be continuously selected. In this subsection, we use an example to see that the value of the model-independent parameter can be discrete.

Blockwise likelihood (see, e.g., Varin, Reid and Firth 2011) is typically used in models with dependent data. To reduce the data dependency, the blockwise versions of the data are considered. Let Y_1, \dots, Y_n be dependent observations from an unknown d -variate distribution $f(y; \theta)$, where the parameter vector $\theta \in \Theta \subset R^p$. The information about θ and $f(y; \theta)$ is available in the form of an unbiased estimating function $u(y; \theta)$, i.e. $E(u(Y; \theta^0)) = 0$, where θ^0 is the true value of θ and $u(y; \theta)$ is a given function vector: $R^d \times \Theta \rightarrow R^r$ with $r \geq p$. Let τ and l be integers satisfying $\tau = \lfloor n^{1-c_1} \rfloor$ and $l = \lfloor c_2 n^{1-c_1} \rfloor$ for some constants $0 < c_1 \leq 1$ and $0 < c_2 \leq 1$, where $\lfloor x \rfloor$ stands for the integer part of x . Denote $B_i = (Y_{(i-1)l+1}, \dots, Y_{(i-1)l+\tau})^T$, $i = 1, \dots, q$, where $q = \lfloor (n - \tau)/l \rfloor + 1$. It can be verified that $q = O(n^{c_1})$. We can see that B_i are blocks of observations, τ is the window-width, and l is the separation between the block starting points. The observation blocks B_i are used to construct the following estimating function:

$$U_i(\theta, \tau) = \frac{1}{\tau} \sum_{k=1}^{\tau} u(Y_{(i-1)l+k}; \theta).$$

Then, the blockwise empirical Euclidean log-likelihood ratio for dependent data is defined as

$$l_{\tau}(\theta) = \sup \left\{ -\frac{1}{2} \sum_{i=1}^q (qp_i - 1)^2 \left| \sum_{i=1}^q p_i = 1, p_i \geq 0, \sum_{i=1}^q p_i U_i(\theta, \tau) = 0 \right. \right\},$$

and the empirical Euclidean likelihood estimator of θ is defined as

$$\hat{\theta}_{\tau} = \arg \sup_{\theta \in \Theta} l_{\tau}(\theta).$$

Here, we only consider the case of $p = r = 1$. It follows from the asymptotic representation given in the proof of Theorem 2 of Lin and Zhang (2001) that under certain regularity conditions, the following asymptotic representation holds:

$$\hat{\theta}_{\tau_k} = \theta + b_n \xi_n(\tau_k) + o_p\left(\frac{1}{\sqrt{n}}\right), \quad k = 1, \dots, m, \quad (3.9)$$

where $b_n = \frac{1}{\sqrt{n}\Delta(\theta)}$, $\xi_n(\tau_k) = \sqrt{n}\bar{U}(\theta, \tau_k)$, $\bar{U}(\theta, \tau) = \frac{1}{q} \sum_{i=1}^q U_i(\theta, \tau)$ and $\Delta(\theta) = E(u'(Y; \theta))$ with $u'(y; \theta)$ being the derivative of $u(y; \theta)$ with respect to θ . Clearly, the above is also within the framework of (2.1) with the unknown b_n and a parameter-dependent $\xi_n(\tau_k)$.

In this example, the positive integer τ is the model-independent parameter. Such a parameter determines the size of blocks of data points, and has discrete values. From the asymptotic representation of the empirical likelihood (3.9), we see that the blockwise empirical likelihood AWLS estimator has the form as in (2.3), i.e.,

$$\tilde{\theta} = \sum_{k=1}^m w_k \hat{\theta}_{\tau_k} - \hat{b}_n \bar{\xi}_n, \quad (3.10)$$

where

$$\hat{b}_n = \frac{\sum_{k=1}^m w_k \hat{\theta}_{\tau_k} \left(\sqrt{n} \bar{U}(\hat{\theta}, \tau_k) - \bar{\xi}_n \right)}{\sum_{k=1}^m w_k \left(\sqrt{n} \bar{U}(\hat{\theta}, \tau_k) - \bar{\xi}_n \right)^2}, \quad \bar{\xi}_n = \sum_{k=1}^m w_k \sqrt{n} \bar{U}(\hat{\theta}, \tau_k)$$

with $\hat{\theta}$ being an initial estimator of θ .

The theoretical property and the optimal choice of weights can be determined by Theorem 2.3 and Remark 2.3. The details are omitted here.

4. Numerical studies

4.1 Simulations

In this subsection, we examine the finite sample behavior of the newly proposed estimators using simulation studies. To obtain thorough comparisons, we comprehensively compare it with several competitors that are based on the objective function composition, the direct composition and the aggregation for some linear and nonparametric models. Mean squared error (for parametric model) and mean integrated squared error (for nonparametric model) are used to evaluate the performances of the involved estimators. We also report the simulation results for estimation bias. Moreover, we consider the asymptotic relative efficiency (RE) defined as $RE(\hat{\beta}, \tilde{\beta}) = \text{Var}(\hat{\beta})/\text{Var}(\tilde{\beta})$, where $\tilde{\beta}$ is the proposed AWLS estimator and $\hat{\beta}$ is a competitor. Here $RE > 1$ indicates better performance of the AWLS estimator.

Experiment 1. Consider the linear regression of the form

$$Y = X^T \beta + \epsilon,$$

where $\beta = (3, 2, -1, -2)^T$, the covariate vector $X = (X_1, X_2, X_3, X_4)^T$ follows a multivariate normal distribution $N(0, \Sigma)$ with $\Sigma_{ij} = 0.7^{|i-j|}$ for $1 \leq i, j \leq 4$, and the error term ϵ follows centralized $Gamma(2, 2)$ so that its expectation is zero.

We choose $\tau = 0.3$ for the asymmetry distribution of the error term to construct the common quantile regression (QR) estimator $\hat{\beta}_\tau$ defined in Subsection 3, and select $\tau_k = \frac{k}{10}$ for $k = 1, 2, \dots, 9$ ($m = 9$) to construct the AWLS estimator $\tilde{\beta}$ defined in (2.2). According to Zou and Yuan (2008), the CQR estimator $\hat{\beta}$ is defined by minimizing the following composite objective function:

$$(\hat{\beta}_{CQ}^T, \hat{\alpha}_{\tau_1}, \dots, \hat{\alpha}_{\tau_m})^T = \arg \min_{\beta, \alpha_{\tau_1}, \dots, \alpha_{\tau_m}} \sum_{i=1}^n \sum_{k=1}^m \rho_{\tau_k}(Y_i - \alpha_{\tau_k} - \beta^T X_i). \quad (4.1)$$

According to Bradic, Fan and Wang (2011), the WCQR estimator $\hat{\beta}_{WCQ}$ is determined by minimizing the composite objective function (4.1) with weight w_k to each $\rho_{\tau_k}(\cdot)$.

To obtain a consistent estimator of the density function $f_e(\alpha_{\tau_k})$, we first use the OLS method to estimate a preliminary estimator $\hat{\beta}_{OLS}$, and then compute the residuals as $\hat{\epsilon}_i = Y_i - X_i^T \hat{\beta}_{OLS}$. $f_e(\alpha_{\tau_k})$ is then estimated by the nonparametric kernel density estimator through $\hat{\epsilon}_i, i = 1, 2, \dots, n$. Consequently, the optimal weights in the AWLS estimator $\tilde{\beta}$ defined in (2.2) can be attained.

For the sample sizes $n = 100, 200$ and 400 , the empirical bias, relative efficiency (RE) and mean squared error (MSE) of the four estimators and ordinary least square estimator (OLS) over 500 replications are reported in Table 1. The boxplots with sample size $n = 200$ for the five estimators are depicted in Figure 1. For the different sample sizes of n , the boxplot trends are similar. Further, to check the influence of m on the AWLS estimator, the quantile levels τ 's are valued from 0.1 to 0.9 with three step lengths 0.2, 0.1 and 0.05. In these cases, the compositions are based respectively on 5, 9, and 17 initial estimators. The boxplots of the AWLS estimators with different choices of m and the same sample size $n = 200$ are presented in Figure 2. We also did simulations for the cases of $n = 100$ and $n = 400$. As the results do not have significant difference, we do not report them here.

Table 1 and Figures 1 and 2 about here

Table 1, Figures 1 and 2 obviously suggest the following conclusions. (1) The AWLS estimator $\tilde{\beta}$ and the WCQR estimator of Bradic, Fan and Wang (2011) behave comparably better than the other competitors in the

sense that the MSEs are significantly reduced, the boxplots are observably thinned, and nearly all the relative efficiencies are greater than 1. (2) Without composition, the QR estimator is better than the OLS estimator due to the skewness of the Gamma distribution. (3) In each subfigure of Figure 2, the boxplots are almost identical, showing that the AWLS estimator for the linear quantile regression model is robust to the choice of m .

It is worth pointing out that the simulation result depends the assumption on the distribution of the error term. As shown by a referee, if the error is Gaussian, the OLS is by far the best method in this setting because the basic quantile regression estimators are much worse than OLS in this case. In fact our AWLS aims at combining several quantile regression estimators, and then it can be guaranteed that the AWLS estimator is better than any single quantile regression estimator.

Experiment 2. Consider the dependent data Y_1, Y_2, \dots, Y_n generated from the model

$$Y_i = X_i\theta + \varepsilon_i,$$

where $X_i \sim N(0, 1)$, $\theta = 5$, $\varepsilon_1 = \epsilon_1$, $\varepsilon_i = 0.7\varepsilon_{i-1} + \epsilon_i$ for $i = 2, 3, \dots, n$, and $\epsilon_i, i = 1, \dots, n$, are independent and identically distributed as $N(0, 1)$. We compare the finite sample behaviors of the blockwise composite likelihood estimator and the AWLS estimator. To get blockwise data and composite likelihood estimator, we take $c = 1/3$, $\tau_k = (k + 1)/10, k = 1, 2, \dots, 8$. The simulation results of bias, MSE and RE obtained by different sample sizes and 500 repetitions are listed in Table 2. We can conclude that the proposed AWLS estimator makes progresses in bias, MSE compared with the original blockwise likelihood estimator.

Experiment 3. For the nonparametric regression

$$Y_i = \sin(2\pi X_i) + 2 \exp(X_i^2) + \epsilon_i, \quad i = 1, \dots, n,$$

where $X_i \sim U(0, 1)$, the errors are chosen as $\epsilon_i \sim N(0, 0.5^2)$, and the sample sizes are designed as $n = 100, 200$ and 400 , respectively. The common local constant (LC) estimator (kernel estimator) is defined as

$$\hat{r}_h(x) = \frac{\sum_{i=1}^n Y_i K\left(\frac{X_i - x}{h}\right)}{\sum_{i=1}^n K\left(\frac{X_i - x}{h}\right)}. \quad (4.2)$$

As a comparison, we here define a composite estimator using the composite objective function method: for $h_k = \tau_k n^{-\eta}, k = 1, \dots, m$, the composite local constant (CLC) estimator is the minimizer of the form:

$$\hat{r}(x) = \arg \min_a \sum_{i=1}^n \sum_{k=1}^m (Y_i - a)^2 K\left(\frac{X_i - x}{h_k}\right).$$

This estimator has a closed representation:

$$\hat{r}(x) = \frac{\sum_{i=1}^n \sum_{k=1}^m Y_i K\left(\frac{X_i - x}{h_k}\right)}{\sum_{i=1}^n \sum_{k=1}^m K\left(\frac{X_i - x}{h_k}\right)}, \quad (4.3)$$

which can be regarded as an indirect composition of the LC estimators (4.2) with different bandwidths. Besides, an aggregation (AGG) estimator (Bunea et al. (2004)) is also considered as a competitor, which has the form

$$\hat{r}^*(x) = \sum_{k=1}^m w_k \hat{r}_{h_k}(x),$$

where w_k satisfies $\sum_{k=1}^m w_k = 1$. The optimal weights are obtained by L_1 -type penalized least squares defined by the equation (2.1) in Bunea et al. (2004). To compute the optimal weights w_k 's, the sample is randomly split into two independent subsamples with equal sample size, one (training sample) is used to construct estimators \hat{r}_{h_k} and the other (validation sample) is

used to aggregate them. As the weights rely on the split, 10 random splits of the sample are run and then the aggregation estimator is obtained by an average through the equation (4.1) in Rigollet and Tsybakov (2007).

In this experiment, the Epanechnikov kernel $K(u) = 0.75(1-u^2)\mathbf{1}_{|u|\leq 1}$ is employed, and to facilitate computation of the optimal weights for AWLS estimator, the integral in $A_1(\mathbf{w})$ is approximated using 40 grid intervals. For the three estimation procedures, we use two-fold cross-validation to select a basic bandwidth h_{op} . In the LC estimation procedure, h_{op} is used to define the LC estimator. For the CLC, AGG and the AWLS with the bandwidths of the form $h_k = \tau_k h_{op}$, m values of τ_k s are chosen in the range $[0.5, 1.5]$ with the step length $0.5/l$. We consider the case: $l = 6$ and thus, $m = 13$ and the resulting bandwidths $h_k = \frac{0.5(l+k)}{l}h_{op}$ for $k = 0, \dots, 12$. The simulation results are reported in Table 3, in which the MISE is the empirical mean integrated squared errors through 500 repetitions. The quantile curves of the LC, CLC, AGG and AWLS estimators for $r(x)$ are also presented. Because the results are similar for different sample sizes of n , only the quantile curves with $n = 200$ are depicted in Figure 3 to save space. Each subfigure contains 0.05, 0.5 and 0.95 quantile curves of the nonparametric estimator and the true curve of $r(x)$. To evaluate the influence of m , the 0.05, 0.5 and 0.95 quantile curves of the AWLS estimator with $n = 200$ and $l = 3, 6$ and 9 (i.e., $m = 7, 13$ and 19), and the true curve of $r(x)$ are presented in Figure 4. We can see that the MISEs of the AWLS estimators are all about 0.0158.

Table 2 and Figures 3 and 4 about here

By comparing the MISEs and the quantile curves of the four estimators

in Table 3 and Figures 3 and 4, we have the following findings: (1) the AGG works well with small MISE compared with the LC and CLC, but the AWLS is the best one among the four estimators; (2) the CLC estimator $\hat{r}(x)$ given in (4.3) is the worst one among these estimators, implying that the composite objective function is not always efficient and (3) the AWLS is robust to the choice of m .

In summary, the AWLS estimation usually works well and is not very sensitive to the choice of the number m of initial estimators. Based on the limited simulations, a value m between 10 and 15 is recommended.

4.2. Real data analysis

In this subsection, the cholestyramine dataset in Efron and Tibshirani (1993) is analysed by the LC and AWLS for illustration. The dataset contains 164 individuals who took part in an experiment to see if the drug cholestyramine can lower blood cholesterol levels. The men were supposed to take six packets of cholestyramine per day, but many of them actually took much less. The covariate denoted by X measures ‘Compliance’ as a percentage of the intended dose actually taken. The response denoted by Y is ‘Improvement’ and makes a decrease in total blood plasma cholesterol level from the beginning to the end of the experiment.

The scatter plot of Y against X in Figure 5 shows that the men who took more cholestyramine tend to exhibit bigger improvements in their cholesterol levels, but the model structure seems complex. Thus, a nonparametric regression model $Y = r(X) + \epsilon$ is modelled for the relationship between ‘Improvement’ and ‘Compliance’ (see Efron and Tibshirani (1993)). To estimate the function $r(\cdot)$, the local constant (LC) estimator defined in

(4.2) and the AWLS estimator are employed. In the estimation procedures, we use the Epanechnikov kernel $K(u) = 0.75(1 - u^2)\mathbf{1}_{|u|\leq 1}$ to construct nonparametric estimators and use the equal weights to build the AWLS estimator for simplicity. As did in Experiment 2, the two-fold cross-validation is used to determine the basic bandwidth h_{op} . In the LC estimation procedure, the resulting bandwidth is $h = 8$. Then, as in Experiment 2, $m = 13$ is chosen. Figure 5 depicts the scatter plot of Compliance and Improvement and the curves of the LC estimator and the AWLS estimator of $r(x)$.

Figure 5 and Figure 6 about here

We have three observations: the drug cholestyramine can lower blood cholesterol levels in general, when Compliance varies in the intervals $[20, 50]$ and $[70, 100]$, the blood cholesterol levels rapidly improve, and the curves of the LC estimator and the AWLS estimator are close to each other.

Finally, we use the R^2 values of the LC estimator and the AWLS estimator to further confirm the advantage of the new method, where $R^2 = 1 - \sum_{j=1}^n (Y_j - \hat{Y}_j)^2 / \sum_{j=1}^n (Y_j - \bar{Y})^2$, \hat{Y}_j is the predicted value of Y_j and \bar{Y} is the sample mean of Y_j s. We first use two-fold cross-validation to generate the optimal bandwidth h_{op} and then use the method suggested in Experiment 2 to produce the bandwidths h_k for composite estimation construction. The 500 values of R^2 for the two estimators are computed when such a procedure is repeated 500 times. The boxplots of R^2 values are given in Figure 6. It shows that R^2 values of the the AWLS estimators are larger than those of the LC estimators in general. And the AWLS estimator is more stable than LC estimator due to smaller variation. Thus, the AWLS fits the data better.

Supplementary Materials

Contain the brief description of the online supplementary materials.

Acknowledgements

Lu Lin was supported by NNSF projects (11571204, 11231005, 11526205 and 11626247) of China. Feng Li was supported by NNSF project (U1404104) and a foundation of Zhengzhou University. Kangning Wang was supported by NSF project (ZR2017BA002) of Shandong Province of China. Lixing Zhu was supported by a grant from the University Grants Council of Hong Kong, Hong Kong, China and a NNSF grant (NSFC11671042). The authors thank for the editor, associate editor and referees for their constructive comments that led to a great improvement of the early manuscript.

References

- Bahadur, R. R. (1966). A note on quantiles in Large Samples. *Ann. Math. Statist.*, **37**, 577-580.
- Bhattacharya, P. K. and Gangopadhyay, A. (1990). Kernel and nearest neighbor estimation of a conditional quantile. *Ann. Statist.*, **18**, 1400-1415.
- Bickel, P. J., Klaassen, C. A. J., Ritov, Y. and Wellner, J. A. (1998). *Efficient and adaptive estimation for semiparametric models*. Springer-Verlag New York, Inc.
- Bradic, J., Fan, J. and Wang, W. (2011). Penalized composite quasi-likelihood for ultrahigh dimensional variable selection. *J. R. Statist. Soc., B*, **73**, 325-349.
- Bunea, F., Tsybakov, A. and Wegkamp, M. (2004). Aggregation for regression learning. Available at <https://arxiv.org/abs/math/0410214>.
- Chaudhuri, P. (1991). Nonparametric estimates of regression quantiles and their local Bahadur representation. *Ann. Statist.*, **19**, 760-777.
- Efron, B. and Tibshirani, R. (1993). *An Introduction to the Bootstrap*. Chapman and Hall, New

REFERENCES³²

- York, London.
- Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*. Chapman and Hall, London.
- Hong, S. Y. (2003). Bahadur representation and its applications for local polynomial estimation in nonparametric M -regression. *Nonparametric Statistics*, **15**, 237-251.
- Juditsky, A. and Nemirovski, A. (2000). Functional aggregation for nonparametric regression. *Ann. Statist.*, **28**, 681-712.
- Kai, B, Li, R. and Zou, H. (2010). Local composite quantile regression smoothing: an efficient and safe alternative to local polynomial regression. *J. R. Statist. Soc. B* **72**, 49-69.
- (2011). New efficient estimation and variable selection methods for semiparametric varying-coefficient partially linear models. *Ann. Statist.*, **39**, 305-332.
- Kiefer, J. (1967). On Bahadur's representation of sample quantiles. *Ann. Math. Statist.*, **38**, 1323-1342.
- Kitamura, Y. (1997). Empirical likelihood methods with weakly dependent processes. *Ann. Statist.*, **25**, 2084 - 2102.
- Koenker, R. (1984). A note on L-estimates for linear models. *Statistics and Probability Letters*, **2**, 323-325.
- (2005). *Quantile Regression*. Cambridge Univ. Press.
- Lavancier, F. and Rochet, P. (2016). A general procedure to combine estimators. *Computational Statistics & Data Analysis*, **94**, 175-192.
- Lin, L. and Zhang, R. C. (2001). Blockwise empirical Euclidean likelihood for weakly dependent processes. *Statistics and Probability Letters*, **53**, 143-52.
- Rigollet, P. and Tsybakov, A. (2007). Linear and convex aggregation of density estimators. *Mathematical Methods of Statistics*, **16**(3), 260-280.
- Shao, J. and Wu, C. F. J. (1989). A general theory for jackknife variance estimation. *Ann. Stat.*, **17**, 1176-1197.
- Sun, J., Gai Y. J. and Lin, L. (2013). Weighted local linear composite quantile estimation for the case of general error distributions. *Journal of Statistical Planning and Inference*, **143**,

1049-1063.

Varin, C., Reid, N. and Firth, D. (2011). An overview of composite likelihood methods. *Statist. Sinica*, **21**, 5-42.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.

Zou, H. and Yuan, M. (2008). Composite quantile regression and the oracle model selection theory. *Ann. Statist.*, **36**, 1108-1126.

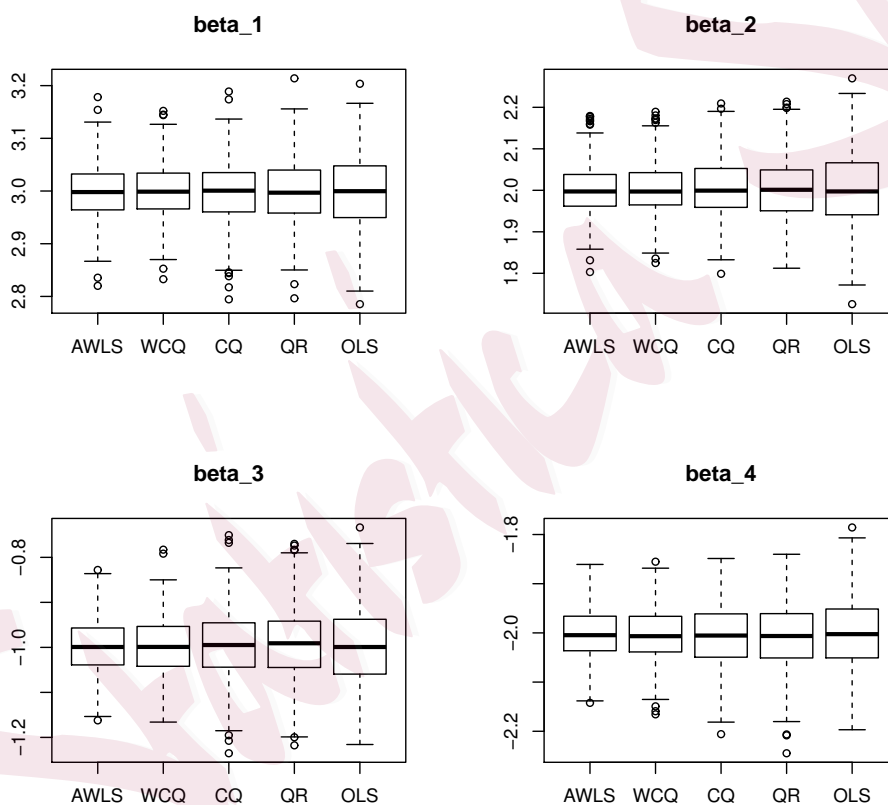


Figure 1. The boxplots of the estimators for $\beta_1, \beta_2, \beta_3$ and β_4 in Experiment 1.

Table 1: Simulation results in Experiment 1

n			$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
100	AWLS	Bias	-0.0016	0.0042	-0.0042	0.0015
		MSE	0.0055	0.0085	0.0092	0.0058
	WCQR	Bias	-0.0007	0.0017	-0.0009	-0.0015
		MSE	0.0053	0.0080	0.0090	0.0063
		RE	0.9668	0.9457	0.9800	1.0751
	CQR	Bias	-0.0026	0.0021	-0.0010	-0.0008
		MSE	0.0077	0.0108	0.0117	0.0080
		RE	1.4032	1.2708	1.2756	1.3739
	QR	Bias	-0.0011	0.0036	0.0004	-0.0029
		MSE	0.0077	0.0116	0.0129	0.0082
		RE	1.4160	1.3636	1.4003	1.3922
	OLS	Bias	-0.0015	0.0037	-0.0037	0.0008
		MSE	0.0098	0.0146	0.0153	0.0103
		RE	1.7938	1.7138	1.6640	1.7635
	200	AWLS	Bias	-0.0008	-0.0002	0.0025
MSE			0.0026	0.0036	0.0037	0.0027
WCQR		Bias	-0.0010	0.0010	0.0026	-0.0031
		MSE	0.0026	0.0037	0.0039	0.0027
		RE	1.0006	0.9955	1.0604	1.0281
CQR		Bias	-0.0017	0.0024	0.0035	-0.0039
		MSE	0.0036	0.0053	0.0053	0.0039
		RE	1.4126	1.4591	1.4472	1.4630
QR		Bias	-0.0010	-0.0001	0.0075	-0.0056
		MSE	0.0037	0.0057	0.0060	0.0043
		RE	1.4430	1.5693	1.6230	1.6188
OLS		Bias	-0.0015	0.0043	0.0009	-0.0000
		MSE	0.0049	0.0075	0.0073	0.0050
		RE	1.9276	2.0504	1.9909	1.8754
400		AWLS	Bias	-0.0023	0.0024	0.0000
	MSE		0.0011	0.0016	0.0017	0.0012
	WCQR	Bias	-0.0014	0.0021	-0.0008	-0.0006
		MSE	0.0011	0.0017	0.0016	0.0012
		RE	1.0008	1.0129	0.9748	0.9976
	CQR	Bias	-0.0009	0.0033	-0.0021	-0.0002
		MSE	0.0019	0.0029	0.0027	0.0018
		RE	1.5535	1.5467	1.4763	1.5712
	QR	Bias	-0.0011	0.0030	-0.0007	-0.0012
		MSE	0.0019	0.0029	0.0027	0.0018
		RE	1.7319	1.7638	1.6553	1.5533
	OLS	Bias	0.0005	0.0021	-0.0013	-0.0009
		MSE	0.0026	0.0035	0.0034	0.0026
		RE	2.3196	2.1202	2.0401	2.2269

Table 2: Simulation results in Experiment 2

n		Bias	MSE	RE
100	AWLS	-0.0004	0.0228	-
	BCEL	0.0006	0.0254	1.1157
200	AWLS	-0.0012	0.0091	-
	BCEL	0.0011	0.0098	1.0764
400	AWLS	-0.0007	0.0057	-
	BCEL	-0.0021	0.0060	1.0564

Table 3: MISE for nonparametric estimators in Experiment 2

	n=100	n=200	n=400
LC	0.0397	0.0251	0.0114
CLC	0.0555	0.0475	0.0453
AGG	0.0287	0.0192	0.0121
AWLS	0.0264	0.0157	0.0109

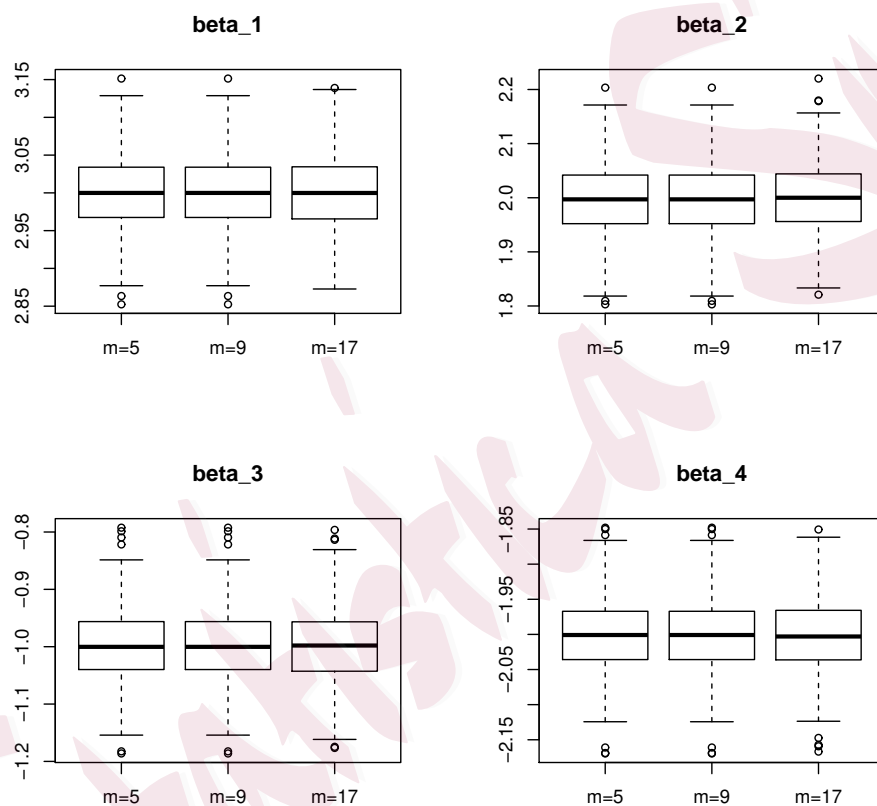


Figure 2. The boxplots of the AWLS estimators for $\beta_1, \beta_2, \beta_3$ and β_4 with different m in Experiment 1.

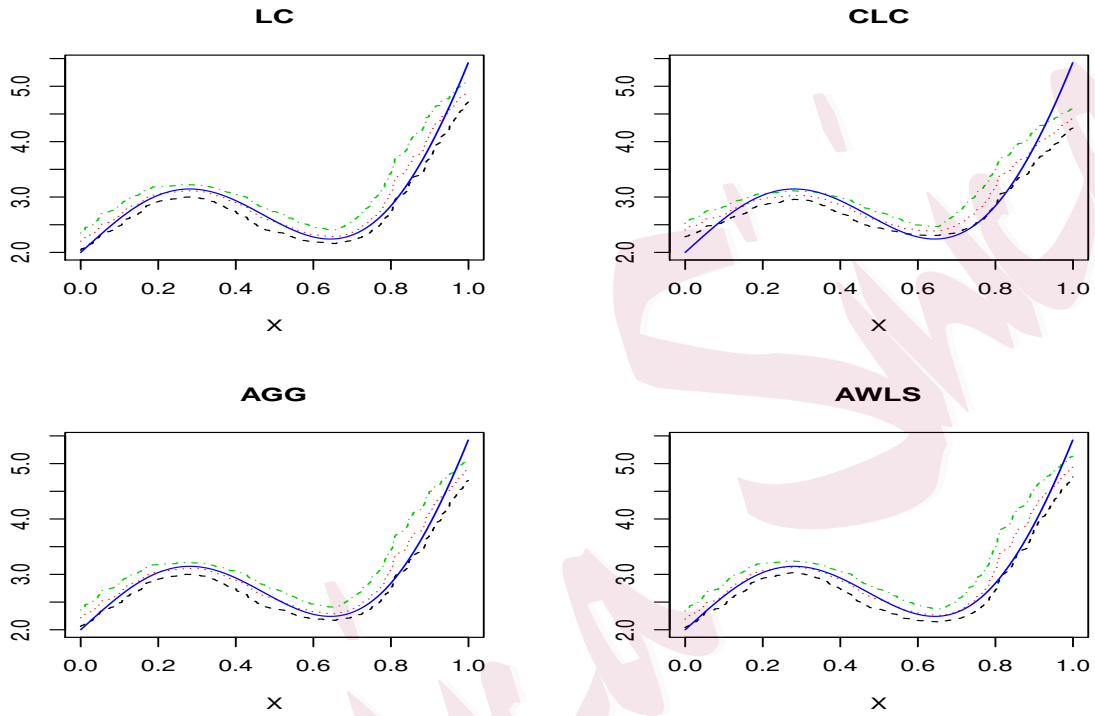


Figure 3. Quantiles curves for the LC, CLC , AGG and AWLS estimators in Experiment 2.

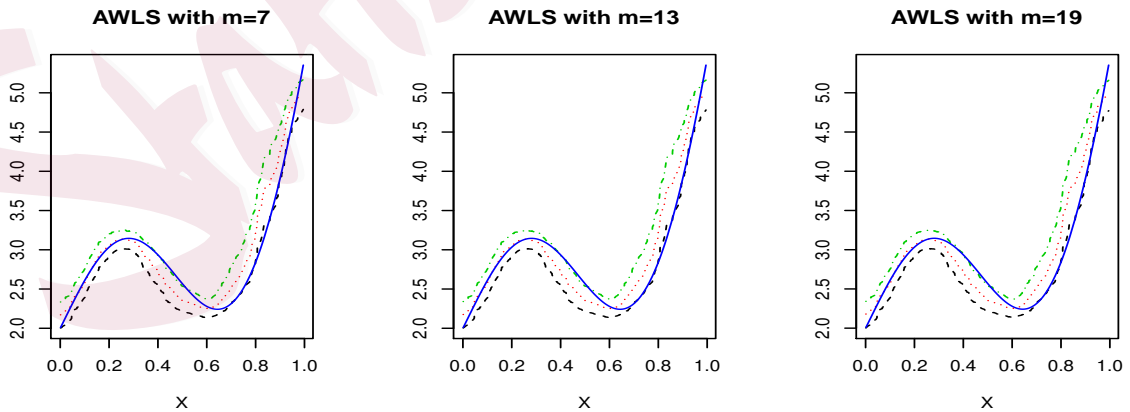


Figure 4. Quantiles curves for the AWLS estimators with different values of m in Experiment 2.

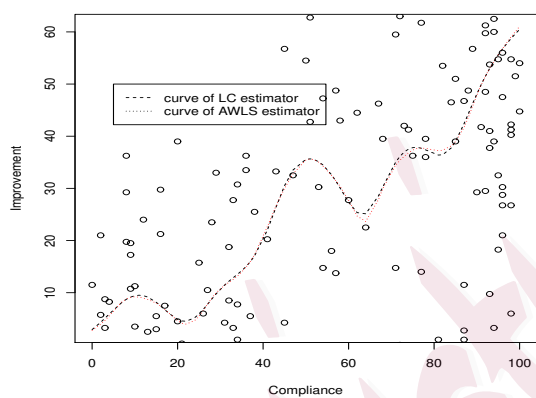


Figure 5. The curves of the LC estimator and the AWLS estimator.

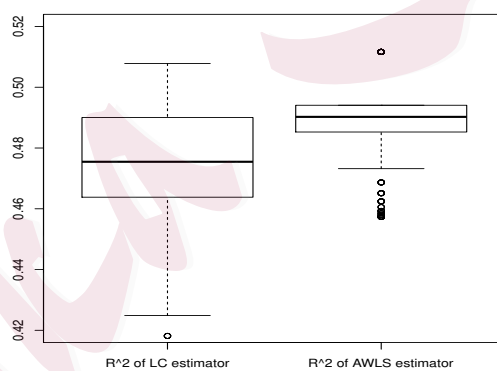


Figure 6. The boxplots of R^2 for LC estimator and AWLS estimator.

REFERENCES39

Lu Lin

Zhongtai Securities Institute for Financial Studies, Shandong University, Jinan, China

School of Statistics, Qufu Normal University, Qufu, China

E-mail: (linlu@sdu.edu.cn)

Feng Li

School of Mathematics and Statistics, Zhengzhou University, Zhengzhou, China

E-mail: (lifengstat@zzu.edu.cn)

Kangning Wang

School of Statistics, Shandong Technology and Business University, Yantai, China

E-mail: (wkn1986@126.com)

Lixing Zhu

Department of mathematics, Hong Kong Baptist University, Hong Kong, China

School of Statistics, Beijing Normal University, Beijing, China

E-mail: (lzhu@hkbu.edu.hk)