SCIENCE CHINA
Mathematics

• ARTICLES •

September 2017 Vol. 60 No. 9: 1705–1716
doi: 10.1007/s11425-015-0158-0

A weighted Wilcoxon estimate for the covariate-specific ROC curve

ZHANG QingZhao[1], DUAN XiaoGang[2] & ZHOU XiaoHua[3,*]

[1]*School of Economics and Wang Yanan Institute for Studies in Economics, Xiamen University, Xiamen* 361005*, China;*
[2]*Department of Statistics, Beijing Normal University, Beijing* 100875*, China;*
[3]*Department of Biostatistics, University of Washington, Seattle, WA* 98195*, USA*

Email: qzzhang@xmu.edu.cn, xgduan@bnu.edu.cn, azhou@uw.edu

Received March 7, 2015; accepted September 24, 2016; published online February 23, 2017

Abstract The covariate-specific receiver operating characteristic (ROC) curve is an important tool for evaluating the classification accuracy of a diagnostic test when it is associated with certain covariates. In this paper, a weighted Wilcoxon estimator is constructed for estimating this curve under the framework of location-scale model for the test result. The asymptotic normality is established, both for the regression parameter estimator and the estimator for the covariate-specific ROC curve at a fixed false positive point. Simulation results show that the Wilcoxon estimator compares favorably to its main competitors in terms of the standard error, especially when outliers exist in the covariates. As an illustration, the new procedure is applied to the dementia data from the national Alzheimer's coordinating center.

Keywords covariates, location-scale model, receiver operating characteristic curve, Wilcoxon method

MSC(2010) 62E20, 62F35

Citation: Zhang Q Z, Duan X G, Zhou X H. A weighted Wilcoxon estimate for the covariate-specific ROC curve. Sci China Math, 2017, 60: 1705–1716, doi: 10.1007/s11425-015-0158-0

1 Introduction

The receiver operating characteristic (ROC) analysis, which was developed in the statistical decision theory, has been employed successfully in a variety of fields, such as experimental psychology, medical diagnosis and medical imaging. The ROC curve plots a test's sensitivity versus its 1-specificity as one varies the decision threshold for test positivity, and is a useful tool to evaluate the classification ability of a medical diagnostic test.

In practice, the ROC analysis is often complicated due to the existence of covariates. In particular, the discrimination ability of a diagnostic test is often varied for different subpopulations defined by some or all involved covariates. For example, for the uniform data example analyzed in Section 4, the test under evaluation is the mini-mental state examination (MMSE), a 30-point questionnaire used to screen for cognitive impairment. This test includes questions and problems in a number of areas, such as repeating lists of words, arithmetic such as the serial sevens, language use and comprehension. It is likely that this test has different discrimination ability for patients from different background. For example, controlling

* Corresponding author

other possible confounding factors, the discrimination accuracy of the MMSE test may be different for patients with different education levels. Moreover, some exploratory analysis shows that outliers may exist in some covariates, such as age. Section 4 contains detailed discussion about this data example.

The covariate-specific ROC curve is widely used to evaluate the classification accuracy of the test result for a specific subpopulation defined by the involved covariates. Like the ROC curve, a covariate-specific ROC curve also depicts the position of each quantile point of the nondiseased population in the diseased population. However, a covariate-specific ROC curve compares two conditional distributions. This usually complicates the analysis and modeling process. In the literature, one uses either direct or indirect approach to analyze the covariate-specific ROC curve (see [20, Chapter 5]). At the heart of the direct approach is the specification of a model, either parametrically or semiparametrically, to connect the underlying covariates directly to the target curve. This facilitates the interpretation of how covariates affect the curve, but meanwhile causes estimation inconvenience since the curve itself is unobservable. Quite differently, a key step for the indirect approach is the specification of a regression model for the test result. Based on the specific model, one can straightforwardly write out the expression for the target curve. Compared with the direct approach, the indirect modeling is easy to implement. Moreover, many developed tools can be borrowed to make inference about the target curve. Due to these benefits, the indirect approach is usually a first choice for modeling the covariate-specific ROC curve. A location-scale model is usually employed for the indirect modeling (see [2–4, 8, 11–13, 19]), which is also the regression model we focused on in the current paper.

Recently, Duan and Zhou [1] proposed to estimate the covariate-specific ROC curve through the composite quantile regression method (see [21]) under a location-scale model. To facilitate the estimation, they reformulate the classic location-scale model and estimate the regression parameters under the reformulated framework. One drawback of the procedure is that they have to introduce extra nuisance parameters when estimating the regression parameters in the location-scale model. This increases the computational burden for using their procedure, particularly when the number of quantile points is large. Moreover, their procedure performs bad when outliers exist in the covariates. However, as will be shown below, the weighted Wilcoxon procedure avoids all the shortcomings mentioned above, and meanwhile it retains the efficiency and robustness advantages of the composite quantile regression method.

The Wilcoxon method is an important alternative to the least squares and quantile type methods for estimating the regression parameters. By appropriately choosing the weight function, a weighted Wilcoxon estimator is robust to outliers both in the response and covariates, which is frequently encountered in biomedical and many other application fields (see [7, 17]). The survey paper by McKean [10] contains many useful references for the Wilcoxon method. In this paper, we propose to use the weighted Wilcoxon method to estimate the regression parameters in a reformulated location-scale model, which is further used to estimate the covariate-specific ROC curve. The simulation results suggest that the weighted Wilcoxon procedure is very competitive to existing ones in terms of standard error, particularly when outliers exist in the covariates.

The rest of the paper is organized as follows. In Section 2, we introduce the model and present the main estimation procedure. Following this, we display the asymptotic results of the proposed estimators, both for the regression parameters and the ROC curve at a fixed false positive point. We also present the result for a linear location-scale model to gain insights into the new procedure. In Section 3, numerical comparisons are conducted through simulation studies. In Section 4, the dementia data from the national Alzheimer's coordinating center (NACC) is analyzed for illustration. Section 5 contains a brief discussion, and all the technical details are displayed in Appendix.

2 Methodology

2.1 Model

Let Y, D and X be the continuous test result, the true disease status and the covariates available for a subject, respectively. Let $D = 1$ denote a diseased subject and 0 a healthy subject. We assume that

larger values of Y are more indicative of the disease. Consider the location-scale model

$$Y = \tilde{\mu}(X, D, \alpha) + D\sigma_1\epsilon_1 + (1 - D)\sigma_0\epsilon_0, \tag{2.1}$$

where $\tilde{\mu}$ is a known function, α is a p-dimensional parameter vector, σ_1 and σ_0 are disease-specific scale parameters, and ϵ_0 and ϵ_1 are disease-specific errors, both with mean 0 and variance 1, and independent of X and D. Under (2.1), the derived ROC curve for a given $X = x$ is

$$R_x(t) = s_1\left\{\frac{\sigma_0}{\sigma_1}s_0^{-1}(t) + \frac{\tilde{\mu}(x, 0, \alpha) - \tilde{\mu}(x, 1, \alpha)}{\sigma_1}\right\},$$

where s_0 and s_1 denote the survival functions of ϵ_0 and ϵ_1, respectively.

To facilitate comparisons among different estimation procedures, Duan and Zhou [1] proposed to reorganize (2.1) as

$$Y = \mu(X, D, \beta) + D\varepsilon_1 + (1 - D)\varepsilon_0, \tag{2.2}$$

where μ is a known function, β is a q-dimensional parameter vector with $q \leqslant p$, ε_0 and ε_1 are disease-specific errors with the respective survival functions S_0 and S_1. Besides the usually simpler functional form compared to (2.1), a key feature of (2.2) is that no specific moment restrictions are imposed on both ε_1 and ε_0, except their being independent of X and D. Under (2.2), the covariate-specific ROC curve is

$$R_x(t) = S_1\{S_0^{-1}(t) + \mu(x, 0, \beta) - \mu(x, 1, \beta)\}. \tag{2.3}$$

2.2 The proposed method

Let $\{(Y_i, X_i, D_i) : i = 1, \ldots, n\}$ be an independent and identically distributed sample from (2.2). The weighted Wilcoxon estimator of β, denoted by $\hat{\beta}_n$ throughout, is the minimizer of

$$Q_n(\beta) = \sum_{j=1}^{n}\sum_{i<j} b_{ij}\{D_iD_j + (1 - D_i)(1 - D_j)\}|e_i(\beta) - e_j(\beta)|, \tag{2.4}$$

where $e_i(\beta) = Y_i - \mu(X_i, D_i, \beta)$, and b_{ij} is weight attached to the comparison of $e_i(\beta)$ and $e_j(\beta)$. In the simulation studies and real data analysis, we adopt the generalized rank (GR) weights (see [16]), i.e., $b_{ij} = h(X_i)h(X_j)$, where

$$h(X_i) = \min\left\{1, \frac{b}{(X_i - \psi)^{\mathrm{T}}W^{-1}(X_i - \psi)}\right\},$$

with (ψ, W) being the robust minimum volume ellipsoid estimator of the location and scatter, and b being the 95th percentile of $\chi^2(r)$; here r is the dimension of X. As noted in [15], use of between-group comparisons can bias the estimates if the error distributions are different. Hence we introduce the term $D_iD_j + (1 - D_i)(1 - D_j)$ in (2.4), to eliminate comparisons of the residuals between the diseased and healthy group.

Compared with [1], the proposed method has two advantages. First, we need not to introduce additional parameters for estimating β. In contrast, [1] defines $2K$ quantile points of the error terms to estimate β, where K is the number of quantile points in the composite quantile regression procedure. In this sense, the proposed procedure is a better partner of the reformulated model framework. Second, the weighted Wilcoxon method is robust to outliers both in the response and covariates, while the procedure in [1] can be seriously destroyed by outliers in the covariates.

After obtaining $\hat{\beta}_n$, one can estimate S_0 and S_1 naturally through

$$\hat{S}_0(\varepsilon) = \frac{\sum_{i=1}^{n}(1 - D_i)I(\hat{\varepsilon}_i > \varepsilon)}{\sum_{i=1}^{n}(1 - D_i)} \quad \text{and} \quad \hat{S}_1(\varepsilon) = \frac{\sum_{i=1}^{n}D_iI(\hat{\varepsilon}_i > \varepsilon)}{\sum_{i=1}^{n}D_i},$$

respectively, where $\hat{\varepsilon}_i = D_i\{Y_i - \mu(X_i, 1, \hat{\beta}_n)\} + (1 - D_i)\{Y_i - \mu(X_i, 0, \hat{\beta}_n)\}$. Define $\hat{S}_0^{-1}(t) = \inf\{y : \hat{S}_0(y) < t\}$. Plugging $\hat{\beta}_n$, $\hat{S}_0$ and $\hat{S}_1$ into (2.3), the proposed estimator for $R_x(t)$ is

$$\hat{R}_x(t) = \hat{S}_1\{\hat{S}_0^{-1}(t) + \mu(x, 0, \hat{\beta}_n) - \mu(x, 1, \hat{\beta}_n)\}. \tag{2.5}$$

2.3 Asymptotic normality

We first introduce some notation. Write $n_1 = \sum_{i=1}^n D_i$ and $n_0 = n - n_1$. Define $\mu_1 = \mathrm{E}\{D\dot{\mu}(X, 1, \beta^*)\}$ and $\mu_0 = \mathrm{E}\{(1-D)\dot{\mu}(X, 0, \beta^*)\}$, where β^* denotes the true value of β, and $\dot{\mu}$ represents the partial derivative of μ with respect to β. Write $\dot{\boldsymbol{\mu}}_n = \{\dot{\mu}(X_1, D_1, \beta^*), \ldots, \dot{\mu}(X_n, D_n, \beta^*)\}^{\mathrm{T}}$. Let $\mathcal{B}$ be the parameter space for β, and $a^{\otimes 2} = aa^{\mathrm{T}}$ for a vector or matrix a. Let $W_n = (w_{ij})$ be an $n \times n$ matrix with

$$w_{ij} = \begin{cases} -n_0^{-1} b_{ij} I\{D_i = 0, D_j = 0\} - n_1^{-1} b_{ij} I\{D_i = 1, D_j = 1\}, & i \neq j, \\ n_0^{-1} I\{D_i = 0\} \sum_{k \neq i}^n b_{ik} + n_1^{-1} I\{D_i = 1\} \sum_{k \neq i}^n b_{ik}, & i = j. \end{cases}$$

We assume the following conditions hold:

A1. Write $\pi = \mathrm{pr}(D = 1)$. Then $0 < \pi < 1$.

A2. The parameter space $\mathcal{B}$ is compact, and the true value β^* is an interior point of $\mathcal{B}$.

A3. For $j = 0, 1$, the distribution function F_j of ε_j is absolutely continuous, with continuous density f_j.

A4. (i) The function $\mu(x, d, \beta)$ is continuous in (x, d) for each β, and is differentiable at β^* for each (x, d), with derivative $\dot{\mu}(x, d, \beta)$ such that $\mathrm{E}\{\dot{\mu}(X, D, \beta)^{\otimes 2}\}$ is positive definite.

(ii) There exists a measurable function $U(x, d)$ with $\mathrm{E}\{U(X, D)^2\} < \infty$, such that $|\mu(x, d, \beta) - \mu(x, d, \tilde{\beta})| \leqslant U(x, d)\|\beta - \tilde{\beta}\|$, for each x, d, β and $\tilde{\beta}$.

(iii) There exists positive definite matrices V and C such that $n^{-1}\dot{\boldsymbol{\mu}}_n^{\mathrm{T}} W_n^2 \dot{\boldsymbol{\mu}}_n$ converges in probability to V and $n^{-1}\{\tau_1 \dot{\boldsymbol{\mu}}_n^{\mathrm{T}} \mathscr{D} W_n \mathscr{D} \dot{\boldsymbol{\mu}}_n + \tau_0 \dot{\boldsymbol{\mu}}_n^{\mathrm{T}} (I_n - \mathscr{D}) W_n (I_n - \mathscr{D}) \dot{\boldsymbol{\mu}}_n\}$ converges in probability to C, where $\tau_0 = \int f_0^2(t)dt$, $\tau_1 = \int f_1^2(t)dt$, $\mathscr{D} = \mathrm{diag}\{D_1, \ldots, D_n\}$, and I_n denotes the $n \times n$ identity matrix.

(iv) The class of functions $(y, x, d) \mapsto I\{y - \mu(x, d, \beta) > \omega\}$, for β and ω in some neighborhood of the associated true values, is Donsker.

Assumption A1 is trivial in ROC-related literatures. Assumptions A2 and A3 are regular conditions which are required for proving the consistency of $\hat{\beta}_n$. Conditions A4(i)–A4(iii) are modified versions of those used for investigating rank method under a linear model. Condition A4(iv) has been used in [1] to prove the asymptotic normality of $\hat{R}_x(t)$. Our main results are as follows.

Theorem 2.1. *Suppose that Conditions* A1–A4 *hold. Then the sequence* $n^{1/2}(\hat{\beta}_n - \beta^*)$ *is asymptotically normal with mean zero and covariance matrix* $12^{-1}C^{-1}VC^{-1}$.

Theorem 2.2. *Suppose that Conditions* A1–A4 *hold. Then, for each fixed* $t \in (0, 1)$, *and a fixed covariate value* x, *the sequence* $n^{1/2}\{\hat{R}_x(t) - R_x(t)\}$ *is asymptotically normal with mean zero and variance*

$$\pi^{-1} S_1(w^*)\{1 - S_1(w^*)\} + \frac{t(1-t) f_1^2(w^*)}{(1-\pi) f_0^2\{S_0^{-1}(t)\}} + \frac{1}{12} f_1^2(w^*) \nabla_x^{\mathrm{T}} C^{-1} V C^{-1} \nabla_x, \tag{2.6}$$

where $w^* = S_0^{-1}(t) + \mu(x, 0, \beta^*) - \mu(x, 1, \beta^*)$, *and* $\nabla_x = \{\dot{\mu}(x, 1, \beta^*) - \dot{\mu}(x, 0, \beta^*)\} - \{\pi^{-1}\mu_1 - (1-\pi)^{-1}\mu_0\}$.

The proof of Theorem 2.2 is similar to that for proving [1, Theorem 2], and thus is omitted here. We see that the asymptotic variance of $n^{1/2}\{\hat{R}_x(t) - R_x(t)\}$ is affected by the specific estimation procedures only through the third term in (2.6), for which the core part is the asymptotic covariance matrix of $n^{1/2}(\hat{\beta}_n - \beta^*)$. This provides us a unified framework for comparing different procedures chosen for estimating the target ROC curve. Given the first two terms in (2.6), a procedure with smaller asymptotic covariance for estimating β is more appealing for estimating the target ROC curve.

2.4 A special case: Linear location-scale model

To gain insights into the proposed estimation procedure, we consider a linear location-scale model $Y = \alpha_0 + \alpha_1 D + \alpha_2^{\mathrm{T}} X + D\alpha_3^{\mathrm{T}} X + D\sigma_1\epsilon_1 + (1 - D)\sigma_0\epsilon_0$, which can be reorganized as $Y = \beta_1^{\mathrm{T}} X + D\beta_2^{\mathrm{T}} X + D\varepsilon_1 + (1 - D)\varepsilon_0$, where $\beta_1 = \alpha_2$, $\beta_2 = \alpha_3$, and ε_0 and ε_1 are newly defined disease-specific errors. Write $\beta = (\beta_1^{\mathrm{T}}, \beta_2^{\mathrm{T}})^{\mathrm{T}}$, then we can estimate it by minimizing

$$\bar{Q}_n(\beta) = \sum_{j=1}^n \sum_{i<j} b_{ij} I(D_i = 1, D_j = 1)|Y_i - Y_j - (\beta_1 + \beta_2)^{\mathrm{T}}(X_i - X_j)|$$

$$+\sum_{j=1}^{n}\sum_{i<j} b_{ij}I(D_i=0,D_j=0)|Y_i-Y_j-\beta_1^{\mathrm{T}}(X_i-X_j)|,$$

which is a special case of (2.4). It follows that we can estimate β_1 only based on the second term of $\bar{Q}_n(\beta)$, and estimate β_2 via $\hat{\theta}-\hat{\beta}_1$, where $\theta=\beta_1+\beta_2$ is estimated based on the first term of $\bar{Q}_n(\beta)$. We have the following result.

Corollary 2.3. *Suppose that Conditions* A1–A4 *hold. Let* $b_{ij}\equiv 1$, $\tau_0=\int f_0^2(t)dt$, *and* $\tau_1=\int f_1^2(t)dt$. *Furthermore, let* $C_1=\mathrm{E}[D\{\dot{\mu}_\beta(X,1,\beta^*)\}^{\otimes 2}]$ *and* $C_0=\mathrm{E}[(1-D)\{\dot{\mu}(X,0,\beta^*)\}^{\otimes 2}]$. *Then* $n^{1/2}(\hat{\beta}_n-\beta^*)$ *is asymptotically normal with mean zero and covariance matrix* $12^{-1}\Gamma^{-1}\Delta\Gamma^{-1}$, *where* $\Gamma=\tau_1(C_1-\pi^{-1}\mu_1^{\otimes 2})+\tau_0\{C_0-(1-\pi)^{-1}\mu_0^{\otimes 2}\}$ *and* $\Delta=(C_1-\pi^{-1}\mu_1^{\otimes 2})+\{C_0-(1-\pi)^{-1}\mu_0^{\otimes 2}\}$.

Under the linear location-scale model with $b_{ij}\equiv 1$, there is some interesting findings about the relationship between the Wilcoxon procedure and the composite quantile regression procedure for estimating β, which further implies comparison between the two methods for estimating the target ROC curve. First, if ε_1 and ε_0 are identically distributed, the composite quantile regression estimator has the same asymptotic distribution as the Wilcoxon estimator when the number of the selected quantiles tends to infinity. Second, it can be shown that the asymptotic relative efficiency of the rank with respect to least squares method is never below 0.864, which sharpens the bound 0.703 given in [21]; see [7] for more discussions. Combined this fact to the discussions below Theorem 2.2, it appears that the newly proposed weighted Wilcoxon procedure is more appealing for estimating the ROC curve than most of its competitors, including the one introduced in [1].

3 Simulation

Simulation studies were carried out to assess the finite-sample performance of the proposed method for estimating $R_x(t)$. Our simulations were designed to demonstrate the robustness and efficiency of the proposed weighted Wilcoxon (WW) estimator, compared with the Wilcoxon (W) estimator, the composite quantile regression (CQR) estimator in [1], and the quasilikelihood (QL) estimator in [12]. All the results were based on 1,000 replications. We totally considered two model setups.

Model 1. We used the model considered in [1], i.e.,

$$Y=1+X+D+XD+De_1+(1-D)e_0,$$

where $X\sim N(0,1)$, $D\sim B(1,0.5)$, e_1 and e_0 were both standard normal errors, and the four variables X, D, e_1 and e_0 were generated independently with each other. To investigate the effect of outliers on the estimation of $R_x(t)$, we considered three cases:

Case 1. A contamination of the response was done by replacing a 5% of Y randomly with $Y+\delta$, where $\delta\sim N(0,5^2)$.

Case 2. A contamination of the covariates was done by replacing a 5% of X randomly with $X+\delta$, where $\delta\sim N(0,5^2)$.

Case 3. Independently replaced 5% of Y and 5% of X with $Y+\delta_1$ and $X+\delta_2$, where δ_1 and δ_2 were independent and both followed $N(0,5^2)$.

Table 1 summarizes the bias (Bias), the standard error (SE) and the root mean squared error (RMSE) for the estimated $R_x(t)$ with false positive points fixed at 0.1, 0.3, 0.5 and 0.7, and covariate at $x=0.5$, respectively. We observe that the WW, W and CQR procedures are more robust than the QL method when outliers exist in the Y direction. Furthermore, the WW procedure is robust to outliers in the X direction, in contrast to the adversely affected performance of the W, CQR and QL methods.

Model 2. We considered the same model as previously except that $X\sim U(0,1)$. We displayed the results for three configurations of e_1 and e_0. (1) both from $N(0,1)$; (2) both from a contaminated normal distribution $0.95N(0,1)+0.05N(0,50^2)$; and (3) both from the Cauchy distribution with probability density function proportional to $(1+u^2)^{-1}$.

Table 1 Bias, SE and RMSE for estimating $R_{0.5}(t)$. All values are multiplied by 100

	false positive point							
	0.1		0.3		0.5		0.7	
method	Bias (SE)	RMSE	Bias (SE)	RMSE	Bias (SE)	RMSE	Bias (SE)	RMSE
	Case 1 (outliers in only Y)							
W	−3.7 (10.0)	10.7	−2.0 (5.6)	5.9	−1.7 (3.4)	3.8	−1.6 (2.0)	2.6
WW	−3.7 (10.0)	10.7	−2.0 (5.6)	6.0	−1.7 (3.4)	3.8	−1.6 (2.0)	2.6
CQR	−3.9 (10.1)	10.8	−2.2 (5.6)	6.0	−1.8 (3.5)	3.9	−1.6 (2.0)	2.5
QL	−4.2 (10.2)	11.1	−2.3 (5.8)	6.2	−1.8 (3.5)	4.0	−1.7 (2.1)	2.7
	Case 2 (outliers in only X)							
W	−7.3 (9.7)	12.2	−6.3 (8.8)	10.8	−5.1 (7.0)	8.7	−3.9 (5.0)	6.3
WW	−3.6 (9.3)	10.0	−2.2 (5.6)	6.0	−2.0 (3.5)	4.0	−2.0 (2.1)	2.9
CQR	−8.6 (9.8)	13.0	−7.9 (9.3)	12.2	−6.3 (7.7)	9.9	−4.6 (5.6)	7.3
QL	−13.7 (8.9)	16.3	−15.5 (9.8)	18.4	−12.9 (8.9)	15.7	−9.3 (7.0)	11.6
	Case 3 (outliers in both X and Y)							
W	−11.3 (10.2)	15.3	−8.8 (9.2)	12.7	−7.0 (7.3)	10.1	−5.5 (5.1)	7.5
WW	−8.1 (10.3)	13.2	−4.2 (5.9)	7.2	−3.7 (3.9)	5.4	−3.5 (2.6)	4.4
CQR	−12.7 (10.1)	16.2	−10.4 (9.5)	14.0	−8.3 (7.9)	11.5	−6.3 (5.6)	8.4
QL	−16.8 (9.1)	19.1	−17.1 (10.0)	19.8	−14.3 (8.9)	16.8	−10.4 (6.8)	12.4

Table 2 Bias, SE and RMSE for estimating $R_{0.5}(t)$. All values are multiplied by 100

	false positive point							
	0.1		0.3		0.5		0.7	
method	Bias (SE)	RMSE	Bias (SE)	RMSE	Bias (SE)	RMSE	Bias (SE)	RMSE
	Normal							
W	0.9 (8.4)	8.4	0.3 (5.1)	5.1	0.0 (3.0)	3.0	0.0 (1.6)	1.6
WW	0.8 (8.4)	8.4	0.3 (5.1)	5.1	0.0 (3.0)	3.0	0.0 (1.6)	1.6
CQR	0.8 (8.4)	8.4	0.2 (5.1)	5.1	0.0 (3.0)	3.0	0.0 (1.6)	1.6
QL	0.8 (8.3)	8.4	0.2 (5.0)	5.1	0.1 (5.1)	5.1	0.0 (1.6)	1.6
	Contanimated normal							
W	−0.1 (9.8)	9.8	0.2 (5.2)	5.2	0.1 (3.3)	3.3	0.0 (2.2)	2.2
WW	−0.1 (9.8)	9.8	0.2 (5.2)	5.2	0.1 (3.3)	3.3	0.0 (2.2)	2.2
CQR	−0.2 (9.8)	9.8	0.1 (5.2)	5.2	0.0 (3.4)	3.4	0.0 (2.2)	2.2
QL	−14.0 (16.9)	21.9	−13.8 (14.2)	19.8	−7.1 (9.3)	11.7	−3.8 (7.3)	8.3
	Cauchy							
W	2.7 (10.6)	10.9	−1.1 (6.8)	6.9	−0.2 (4.1)	4.1	−0.1 (3.5)	3.5
WW	2.7 (10.6)	10.9	−1.1 (6.8)	6.9	−0.2 (4.1)	4.1	−0.1 (3.5)	3.5
CQR	2.7 (10.6)	10.9	−1.2 (6.9)	7.0	−0.2 (4.1)	4.1	−0.1 (3.5)	3.5
QL	4.2 (14.6)	15.2	−20.2 (19.8)	28.3	−9.9 (14.9)	17.9	−4.3 (12.5)	13.2

Table 2 summarizes the corresponding simulation results. For the contaminated normal and Cauchy distributions, the proposed method greatly outperforms the QL method in most cases, and is similar to CQR. For the normal error distribution, the four procedures behave very similarly to each other.

The simulation findings can be summarized as follows. First, the newly proposed WW procedure behaves very similar to the CQR procedure for normal setups without outliers existing in the covariates, and both of them perform better than the remaining two procedures. Second, the WW procedure performs the best among all the considered procedures when outliers exist in the covariates, and the efficiency gain of the WW over the CQR method can be substantial. Due to its competitive numerical performance, the WW procedure appears to be a promising choice in the covariate-specific ROC analysis.

4 Example

A real data set collected by the NACC is analyzed to illustrate the proposed method. The test under evaluation is the MMSE, a 30-point questionnaire used to screen for cognitive impairment, with lower values indicating more severe impairment. We used 30 minus the original test score as our response T, so that larger response values were more indicative of the disease status. Our interest is how well the test can predict progression to dementia among subjects who have at least four follow-ups. The total sample size was $n = 1{,}124$.

The dementia status D was defined through the clinical diagnosis result at a patient's fourth visit to one of the participating research centers. We considered six covariates, i.e.,

$$X = (X_1, \ldots, X_6)^{\mathrm{T}},$$

which suggested in turn the patient's age, gender (1 for male, and 0 for female), years of education, the presence or not (1 for presence, and 0 otherwise) of depression, Parkinson's disease, and stroke.

To model the test score, we used a linear location model, including the main effects of D and X as well as their first order interactions. The scale model was assumed to depend only on the disease status. As revealed by Figures 1 and 2, the distribution of T is highly skewed, and X_1 contains some obvious outliers. One may suggest a log-transformation of the response variable. However, our preliminary analysis indicates that the log-transformed T is still unsatisfactory. Since the transformation may not remove the outliers and meanwhile brings additional issues such as the interpretation, we chose to analyze the variables on their original scale.

Table 3 summarizes the point estimates of the model parameters and their bootstrap SE based on 500 bootstrap resampling. Figure 3 displays the estimated $R_x(t)$ at two education levels, i.e., 3 years and 23 years, for a white man aged 75 with no stroke history, no depression and no Parkinson's disease, by the four different procedures as adopted in the simulation studies. We observe that, compared to other procedures, the ROC curve produced by the WW method shows much more difference for patients with different education levels we investigated, with the MMSE becoming less accurate for predicting the dementia status with the increase of the education levels. This is also demonstrated by the corresponding areas under the curve as summarized in Figure 4.

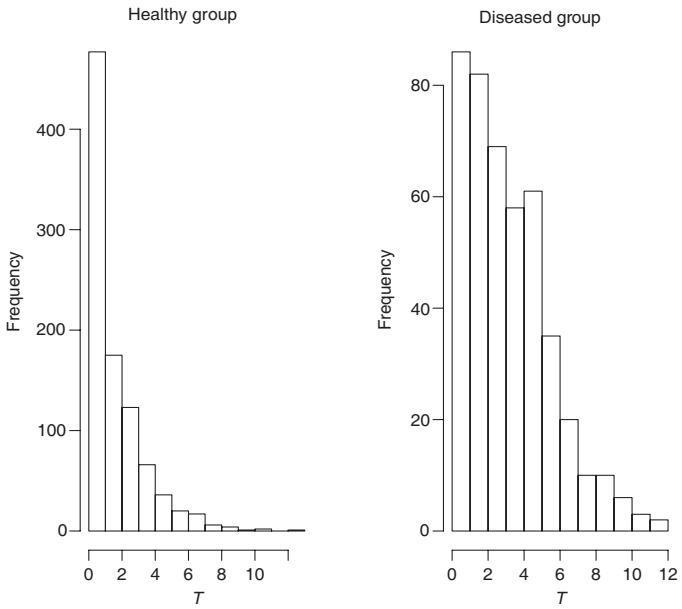

Figure 1 Histograms of T (i.e., 30 minus the original MMSE scores) for the healthy ($D = 0$) and diseased ($D = 1$) groups

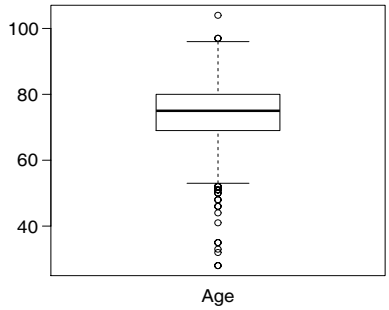

Figure 2 Boxplots of the age variable

Table 3 Estimated parameters and their bootstrap SE (in parentheses) for the dementia data

	W	WW	CQR	QL
$\hat{\sigma}_1$	$*$	$*$	$*$	2.290 (0.100)
$\hat{\sigma}_0$	$*$	$*$	$*$	1.720 (0.070)
Intercept	$*$	$*$	$*$	2.220 (0.720)
X_1	0.021 (0.007)	0.023 (0.007)	0.019 (0.007)	0.021 (0.008)
X_2	0.490 (0.120)	0.440 (0.140)	0.470 (0.110)	0.620 (0.140)
X_3	−0.119 (0.023)	−0.116 (0.025)	−0.130 (0.024)	−0.154 (0.028)
X_4	0.050 (0.140)	−0.000 (0.130)	0.090 (0.130)	0.110 (0.170)
X_5	−0.060 (0.250)	−0.090 (0.260)	−0.040 (0.260)	−0.170 (0.270)
X_6	0.310 (0.300)	0.370 (0.300)	0.350 (0.280)	0.270 (0.320)
D	$*$	$*$	$*$	2.890 (1.520)
$D \times X_1$	−0.009 (0.016)	−0.023 (0.015)	−0.008 (0.014)	−0.012 (0.016)
$D \times X_2$	0.100 (0.280)	0.220 (0.310)	0.080 (0.280)	−0.050 (0.300)
$D \times X_3$	−0.022 (0.052)	−0.050 (0.056)	−0.010 (0.052)	−0.010 (0.053)
$D \times X_4$	−0.390 (0.310)	−0.330 (0.300)	−0.460 (0.290)	−0.560 (0.320)
$D \times X_5$	−0.160 (0.580)	−0.240 (0.600)	−0.130 (0.580)	−0.150 (0.560)
$D \times X_6$	−0.930 (0.910)	−1.040 (0.830)	−1.020 (0.900)	−0.480 (0.930)

Note. X_1: age; X_2: gender; X_3: education; X_4: depression; X_5: Parkinson; X_6: stroke; D: true dementia status.

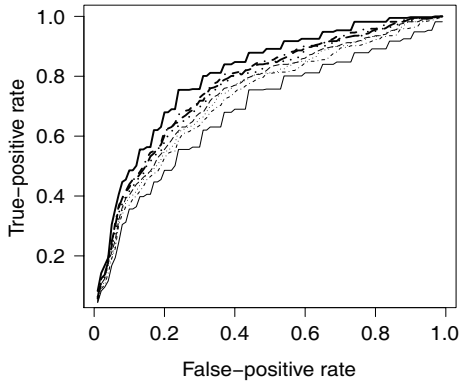

Figure 3 Estimated $R_x(t)$ by the W method (dot dashed lines), the WW method (solid lines), the CQR method (dotted lines) and the QL method (long dashed lines), for 3 years of education (thick lines) and 23 years of education (thin lines)

5 Discussion

In this paper, we have proposed a new method for estimating the covariate-specific ROC curve based on the Wilcoxon rank method, which is an important alternative to the least squares and the quantile method

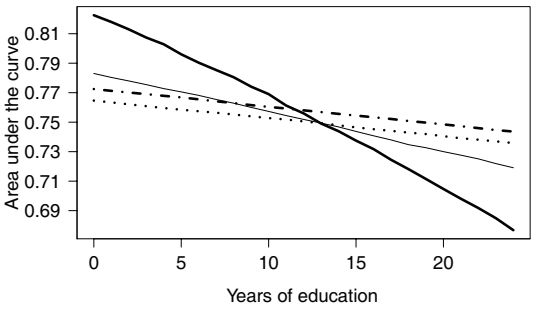

Figure 4 Estimated area under the curve by the W method (thin solid lines), the WW method (thick solid lines), the CQR method (dotted lines) and the QL method (dot dashed lines), with changing education levels

in the regression problem. Our results suggest that the weighted Wilcoxon method is very competing for estimating the covariate-specific ROC curve, particularly when outliers exist in the covariates. To conclude, we point out some possible research avenues in the future. When there are many covariates, the variable selection issue arises for model interpretation and estimation efficiency. Ma and Huang [9] and Wang et al. [18] studied the variable selection for the area under the curve, the most commonly used summary measure of the ROC curve. To our knowledge, there is few variable selection studies focusing on the indirect modeling approach for the ROC curve. It is challenging but interesting to develop in the future flexible selection approaches in the context of the covariate-specific ROC curve.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant Nos. 11401561 and 11301031). The authors thank two referees for their valuable suggestions, and the National Alzheimer's Coordinating Center for providing the data for analysis (the interpretation and reporting of the data are the sole responsibility of the authors).

References

1 Duan X G, Zhou X H. Composite quantile regression for the receiver operating characteristic curve. Biometrika, 2013, 100: 889–900
2 Faraggi D. Adjusting receiver operating characteristic curves and related indices for covariates. J Roy Statist Soc Ser D, 2003, 52: 179–192
3 González-Manteiga W, Pardo-Fernández J C, Van Keilegom I. ROC curves in nonparametric location-scale regression models. Scand J Statist, 2011, 38: 169–184
4 Heagerty P J, Pepe M S. Semiparametric estimation of regression quantiles with application to standardizing weight for height and age in US children. J Roy Statist Soc Ser C, 1999, 48: 533–551
5 Hettmansperger T P, Mckean J W. Robust Nonparametric Statistical Methods. New York: Wiley, 1998
6 Knight K. Limiting distributions for l_1 regression estimators under general conditions. Ann Statist, 1998, 26: 755–770
7 Leng C. Varible selection and coefficient estimation via regularized rank regression. Statist Sinica, 2010, 20: 167–181
8 Liu D P, Zhou X H. Nonparametric estimation of the covariate-specific ROC curve in presence of ignorable verification bias. Biometrics, 2011, 67: 906–916
9 Ma S G, Huang J. Regularized ROC method for disease classification and biomarker selection with microarray data. Bioinformatics, 2005, 21: 4356–4362
10 McKean J W. Robust analysis of linear models. Statist Sci, 2004, 19: 562–570
11 Pepe M S. A regression modelling framework for receiver operating characteristic curves in medical diagnostic testing. Biometrika, 1997, 84: 595–608
12 Pepe M S. Three approaches to regression analysis of receiver operating characteristic curves for continuous test results. Biometrics, 1998, 54: 124–135
13 Pepe M S. Statistical Evaluation of Medical Tests for Classification and Prediction. Oxford: Oxford University Press, 2003
14 Sherman R. The limiting distribution of the maximum rank correlation estimator. Econometrica, 1993, 61: 123–137
15 Sievers G. A weighted dispersion function for estimation in linear models. Comm Statist Theory Methods, 1983, 12: 1161–1179
16 Terpstra J T, McKean J W. Rank-based analyses of linear models using R. J Statist Softw, 2005, 14: 1–26

17 Wang L, Li R Z. Weighted Wilcoxon-type smoothly clipped absolute deviation method. J Amer Statist Assoc, 2009, 104: 1631–1645
18 Wang Y J, Chen H H, Li R Z, et al. Prediction-based structured variable selection through the receiver operating characteristic curves. Biometrics, 2011, 67: 896–905
19 Zheng Y, Heagerty P J. Semiparametric estimation of time-dependent ROC curves for longitudinal marker data. Biostatistics, 2004, 5: 615–632
20 Zhou X H, Obuchowski N A, McClish D K. Statistical Methods in Diagnostic Medicine, 2nd ed. New York: Wiley, 2011
21 Zou H, Yuan M. Composite quantile regression and the oracle model selection theory. Ann Statist, 2008, 36: 1108–1126

Appendix A

Proof of Theorem 2.1. Let $Y_{ij} = Y_i - Y_j$ and $\dot{\mu}_{ij}(s,\beta^*) = \dot{\mu}(X_i,s,\beta^*) - \dot{\mu}(X_j,s,\beta^*)$. For $Q_n(\beta)$ in (2.4), it is equivalent to consider the following objective function:

$$\begin{aligned} L_n(\beta) &= \frac{1}{n(n-1)} \sum_{s\in\{0,1\}} \sum_{i\neq j} b_{ij} 1\{D_i = s, D_j = s\} |Y_{ij} - \{\mu(X_i,s,\beta) - \mu(X_j,s,\beta)\}| \\ &= \frac{1}{n(n-1)} \sum_{s\in\{0,1\}} \sum_{i\neq j} b_{ij} I\{D_i = s, D_j = s\} |e^*_{ij} - \Delta^s_{ij}|, \end{aligned}$$

where $e^*_{ij} = e_i(\beta^*) - e_j(\beta^*)$ and $\Delta^s_{ij} = \{\mu(X_i,s,\beta) - \mu(X_j,s,\beta)\} - \{\mu(X_i,s,\beta^*) - \mu(X_j,s,\beta^*)\}$. By the identity in [6], i.e., $(|r-t| - |r|)/2 = t\{I(r<0) - 1/2\} + \int_0^t \{I(r \leqslant w) - I(r \leqslant 0)\} dw$, we have

$$\begin{aligned} \Gamma_n(\beta) &= L_n(\beta) - L_n(\beta^*) \\ &= \frac{1}{n(n-1)} \sum_{s\in\{0,1\}} \sum_{i\neq j} b_{ij} I(D_i = s, D_j = s) \{|e^*_{ij} - \Delta^s_{ij}| - |e^*_{ij}|\} \\ &= \frac{1}{n(n-1)} \sum_{s\in\{0,1\}} \sum_{i\neq j} b_{ij} I(D_i = s, D_j = s) \Delta^s_{ij} \{2I(e^*_{ij} < 0) - 1\} \\ &\quad + \frac{2}{n(n-1)} \sum_{s\in\{0,1\}} \sum_{i\neq j} b_{ij} I(D_i = s, D_j = s) \int_0^{\Delta^s_{ij}} \{I(e^*_{ij} < t) - I(e^*_{ij} < 0)\} dt \\ &=: Z_{n1} + Z_{n2}. \end{aligned} \tag{A.1}$$

Before analyzing $\Gamma_n(\beta)$, we first define

$$\phi_n = n^{-3/2} \sum_{s\in\{0,1\}} \sum_{i,j} b_{ij} I(D_i = s, D_j = s) \dot{\mu}_{ij}(s,\beta^*) \{2I(e^*_{ij} < 0) - 1\}.$$

Recall that $\dot{\mu}_{ij}(s,\beta^*) = \dot{\mu}(X_i,s,\beta^*) - \dot{\mu}(X_j,s,\beta^*)$. Following [5, Lemma 5.2.3], we have

$$\phi_n = n^{-3/2} \sum_{s\in\{0,1\}} \sum_{i,j} b_{ij} I(D_i = D_j = s) \dot{\mu}_{ij}(s,\beta^*) \{2F_s(e^*_j) - 1\} + o_p(1). \tag{A.2}$$

Denote the first term on the right-hand side of (A.2) by $\phi^\dagger_n$. It is easy to find that

$$\phi^\dagger_n = n^{-1/2} \sum_{s\in\{0,1\}} \sum_{j=1}^n I(D_j = s) \Big\{ \frac{1}{n} \sum_{j=1}^n b_{ij} I(D_i = s) \dot{\mu}_{ij}(s,\beta^*) \Big\} \{2F_s(e^*_j) - 1\}. \tag{A.3}$$

Conditional on (X,D), $\phi^\dagger_n$ is a sum of independent but not identically distributed random variables. Obviously, by Condition A3 and the fact $F_s(e^*_j) \sim U(0,1)$ when $D_j = s$, we have $\mathrm{E}(\phi^\dagger_n) = 0$. In addition, recall the definition of W_n and $\dot{\boldsymbol{\mu}}_n$ in Subsection 2.3, we have

$$\mathrm{Var}(\phi^\dagger_n | X, D) = n^{-3} \sum_{s\in\{0,1\}} \sum_{j=1}^n I(D_j = s) \Big\{ \frac{1}{n} \sum_{j=1}^n b_{ij} I(D_i = s) \dot{\mu}_{ij}(s,\beta^*) \Big\}^2 \mathrm{E}\{2F_s(e^*_j) - 1\}^2$$

$$= (3n)^{-1} \sum_{s\in\{0,1\}} \sum_{i,j,k} I(D_i = s, D_j = s, D_k = s) b_{ij} b_{kj} \dot{\mu}_{ij}(s, \beta^*) \dot{\mu}_{kj}(s, \beta^*)$$
$$= (3n)^{-1} \dot{\boldsymbol{\mu}}_n^{\mathrm{T}} W_n^2 \dot{\boldsymbol{\mu}}_n, \tag{A.4}$$

which converges to $V/3$ in probability. The conditional normality can be established via the Lindeberg-Feller central limit theorem. By Slutsky's lemma, together with (A.2)–(A.4), we have

$$\phi_n \xrightarrow{D} \phi, \quad \text{with} \quad \phi \sim N_q(0, V/3). \tag{A.5}$$

Now we will show that

$$\Gamma_n(\beta) = (\beta - \beta^*)^{\mathrm{T}} C(\beta - \beta^*) + n^{-1/2} \bar{\phi}_n^{\mathrm{T}} (\beta - \beta^*) + o_p(\|\beta - \beta^*\|^2) + o_p(n^{-1}) \tag{A.6}$$

holds uniformly in an $o_p(1)$ neighborhood of β^*, where $\bar{\phi}_n$ converges in distribution to $N_q(0, V/3)$.

To verify (A.6), we first show that

$$Z_{n1} = \frac{1}{\sqrt{n}} \bar{\phi}_n^{\mathrm{T}} (\beta - \beta^*) + o_p\left(\frac{\|\beta - \beta^*\|}{\sqrt{n}} \right) \tag{A.7}$$

holds uniformly in an $o_p(1)$ neighborhood of β^*. By Taylor's expansion for Δ_{ij}^s, we have

$$\begin{aligned} Z_{n1} &= \frac{1}{n(n-1)} \sum_{s\in\{0,1\}} \sum_{i\neq j} b_{ij} I(D_i = s, D_j = s) \{2I(e_{ij}^* < 0) - 1\} \dot{\mu}_{ij}(s, \beta^*)^{\mathrm{T}} (\beta - \beta^*) \\ &\quad + \frac{1}{n(n-1)} \sum_{s\in\{0,1\}} \sum_{i\neq j} b_{ij} I(D_i = s, D_j = s) \{2I(e_{ij}^* < 0) - 1\} \\ &\quad \times \{\mu_{ij}(s, \beta) - \mu_{ij}(s, \beta^*) - \dot{\mu}_{ij}(s, \beta^*)^{\mathrm{T}} (\beta - \beta^*)\} \\ &= \frac{1}{\sqrt{n}} \frac{n^2}{n(n-1)} \phi_n^{\mathrm{T}} (\beta - \beta^*) + R_{n1}. \end{aligned} \tag{A.8}$$

Let $\bar{\phi}_n = [n^2/\{n(n-1)\}]\phi_n$. By (A.5) and Slutsky's lemma, $\bar{\phi}_n$ converges in distribution to $N_q(0, V/3)$. Under Condition A4 and some calculus, $\mathrm{Var}(R_{n1}) = o(n^{-1}\|\beta - \beta^*\|^2)$. Together with $\mathrm{E}(R_{n1}) = 0$, we conclude that $R_{n1} = o_p(n^{-1/2}\|\beta - \beta^*\|)$. Based on the above discussions, (A.7) is proved.

Write $Z_{n2} = \mathrm{E}(Z_{n2} \mid X, D) + Z_{n2} - \mathrm{E}(Z_{n2} \mid X, D)$. Note that e_{ij}^* is defined within the same group. For the healthy group, we denote the distribution function and density function of e_{ij}^* as G_0 and g_0, respectively. Similarly, we can define G_1 and g_1 for the diseased group. A simple calculation yields $g_0(0) = \tau_0$ and $g_1(0) = \tau_1$. Then we have

$$\begin{aligned} &\mathrm{E}(Z_{n2} \mid X, D) \\ &= \frac{2}{n(n-1)} \sum_{s\in\{0,1\}} \sum_{i\neq j} b_{ij} I(D_i = s, D_j = s) \int_0^{\Delta_{ij}^s} \mathrm{E}\{I(e_{ij}^* < t) - I(e_{ij}^* < 0)\} dt \\ &= \frac{2}{n(n-1)} \sum_{s\in\{0,1\}} \sum_{i\neq j} b_{ij} I(D_i = s, D_j = s) \int_0^{\Delta_{ij}^s} \{G_s(t) - G_s(0)\} dt \\ &= \frac{2}{n(n-1)} \sum_{s\in\{0,1\}} \sum_{i\neq j} b_{ij} I(D_i = s, D_j = s) \int_0^{\Delta_{ij}^s} t\{1 + o(1)\} dt \\ &= \frac{1}{n(n-1)} \sum_{s\in\{0,1\}} g_s(0) \sum_{i\neq j} b_{ij} I(D_i = s, D_j = s) (\Delta_{ij}^s)^2 \{1 + o(1)\} \\ &= \frac{(\beta - \beta^*)^{\mathrm{T}}}{n(n-1)} \sum_{s\in\{0,1\}} g_s(0) \sum_{i\neq j} b_{ij} I(D_i = s, D_j = s) \dot{\mu}_{ij}(s, \beta^*) \dot{\mu}_{ij}(s, \beta^*)^{\mathrm{T}} (\beta - \beta^*) + o_p(\|\beta - \beta^*\|^2) \\ &= (\beta - \beta^*)^{\mathrm{T}} n^{-1} \{g_1(0) \dot{\boldsymbol{\mu}}_n^{\mathrm{T}} \mathscr{D} W_n \mathscr{D} \dot{\boldsymbol{\mu}}_n + g_0(0) \dot{\boldsymbol{\mu}}_n^{\mathrm{T}} (I_n - \mathscr{D}) W_n (I_n - \mathscr{D}) \dot{\boldsymbol{\mu}}_n\} (\beta - \beta^*) + o_p(\|\beta - \beta^*\|^2) \end{aligned}$$

$$= (\beta - \beta^*)^{\mathrm{T}} C(\beta - \beta^*) + o_p(\|\beta - \beta^*\|^2). \tag{A.9}$$

Similarly, one can verify that $\mathrm{Var}(Z_{n2} \mid X, Z) = o_p(n^{-2})$. Therefore,

$$Z_{n2} - \mathrm{E}(Z_{n2} \mid X, D) = o_p(n^{-1}). \tag{A.10}$$

Combining (A.9) and (A.10), we show that

$$Z_{n2} = (\beta - \beta^*)^{\mathrm{T}} C(\beta - \beta^*) + o_p(\|\beta - \beta^*\|^2) + o_p(n^{-1}) \tag{A.11}$$

uniformly in $o_p(1)$ neighborhoods of β^*.

Since $\|\beta - \beta^*\|^2 + n^{-1} \geqslant 2n^{-1/2}\|\beta - \beta^*\|$, (A.6) follows by combining (A.1), (A.7) and (A.11). The remaining results follow from [14, Theorems 1 and 2]. This completes the proof. □