

Inference for Support Vector Regression under ℓ_1 Regularization

Yuehao Bai

Department of Economics
University of Chicago
ybai@uchicago.edu

Hung Ho

University of Chicago
hqdhftw@uchicago.edu

Guillaume A. Pouliot

Harris School of Public Policy
University of Chicago
guillaumepouliot@uchicago.edu

Joshua K. C. Shea

Department of Economics
University of Chicago
jkshea@uchicago.edu

October 8, 2019

Abstract

We show that support vector regression (SVR) consistently estimates linear median regression functions and we develop a large sample inference method based on the inversion of a novel test statistic in order to produce error bars for SVR with ℓ_1 -norm regularization. Under a homoskedasticity assumption commonly imposed in the quantile regression literature, the procedure does not involve estimation of densities. It is thus unique amongst large sample inference methods for SVR in that it circumvents the need to select a bandwidth parameter. Simulation studies suggest that our procedure produces narrower error bars than does the standard inference method in quantile regression.

1 Introduction

This paper studies inference for support vector regression (SVR) with ℓ_1 -norm regularization (ℓ_1 -SVR). SVR is the extension of the support vector machine (SVM) classification method (Vapnik, 1998) to the regression problem (Smola and Schölkopf, 2004; Drucker et al., 1997; Basak et al., 2007; Cherkassky and Ma, 2004). SVR is designed to reproduce the good out-of-sample performance of SVM classification in the regression setting. It has been frequently used in regression analysis across fields such as geophysical sciences (Ghorbani et al., 2016; Li et al., 2011), environmental science (Sánchez et al., 2011), engineering (Xi et al., 2007; Pelosof et al., 2004; Li et al., 2012), and image compression (Jiao et al., 2005).

Some methodology is currently available for producing error bars for support vector machines regression. For instance, Gao et al. (2002) produce a Gaussian process formulation which delivers exact, small sample inference. Law and Kwok (2001) cast support vector machines regression in a Bayesian framework, and likewise produce exact, small sample inference. The existing inference methods share the shortcoming that they require distributional assumptions which are rarely satisfied in practice.

In many standard inference problems, such distributional assumptions are typically circumvented by resorting to either asymptotic approximations of the distribution of the estimated coefficients, or by inverting asymptotically valid test procedures.

Meanwhile, theory for asymptotic and exact inference with quantile regression is well developed (Koenker, 2005; Bai et al., 2019), and analogy with quantile regression (see Figure 1) suggests that similar ideas and methods may work for support vector machines regression with ℓ_1 -norm regularization.

Indeed, it has been shown that calculations akin to those used to derive asymptotic distributions of quantile regression coefficient estimates may be used to produce asymptotic approximations of conditional (on features) probabilities of classification for SVM (Pouliot, 2018). These derivations produce asymptotic distributions for the coefficients of the SVM classification problem, and are readily extended to produce asymptotic distributions for the regression coefficients of SVR.

However, regression estimators with support vectors¹ share the unfortunate property that the asymptotic variance of the estimated regression coefficients depends on the density of the regression errors (Koenker and Machado, 1999). This is intuitive since the stability of the fitted hyperplane resting on support vectors ought to depend on the stability of the support vectors. It is indeed an unfortunate property. Any estimate of the asymptotic variance, a key ingredient for producing the error bars, will rely on a nonparametric density estimate. Density estimation itself requires the somewhat arbitrary choice of a bandwidth parameter, which will induce an arbitrarily scaling of the width of error bars.

Here again, the analogy with quantile regression is fruitful. Although the asymptotic distribution of the regression coefficient inherently depends on the density of the regression errors, there are powerful statistical tests whose own asymptotic distribution does not depend on the density of the regression errors. By the duality between testing and inference, these tests may be inverted to produce confidence intervals

¹There is no general, agreed upon definition of support vectors. For both quantile regression and SVM regression with ℓ_1 -norm, they may be rigorously defined as the active basis in the dual program, when written as a linear program in standard form (Bertsimas and Tsitsiklis, 1997) with box constraints (Koenker, 2005).

for regression coefficients which do not require plugging in an estimate of a nonparametric density estimate or selecting a bandwidth parameter.

Contribution Our main contribution is to deliver what is, to the best of our knowledge, the first derivation of error bars for SVR which requires neither distributional assumptions nor the estimation of a nonparametric density—and thus the choice of a bandwidth parameter. Our results apply to support vector machines regression with ℓ_1 -penalty. We establish the asymptotic validity of the procedure, and in doing so, provide what is to the best of our knowledge the first rigorous results on the asymptotic distribution of SVR.

Outline The remainder of the paper is organized as follows. Section 2 introduces the setup and notation. Section 3 introduces the inference procedure. Section 4 presents a simulation study comparing the properties of the median regression rankscore test and those of the ℓ_1 -SVR rankscore test. Section 5 concludes. Technical proofs are deferred to a supplementary appendix.

2 Setup and Notation

Let $W_i = (Y_i, X_i, Z_i) \in \mathbb{R} \times \mathbb{R}^{d_x} \times \mathbb{R}^{d_z}$, $1 \leq i \leq n$ be i.i.d. random vectors. We assume the first element of X_i is 1. Let P denote the distribution of W_i . For a random variable (vector) A , define the vector (matrix) $\mathbf{A}_n = (A_1, \dots, A_n)'$. Let $Q_Y(x, z)$ denote the conditional median of Y given $X = x, Z = z$. We assume that this regression function is linear,

$$Q_Y(x, z) = x'\beta(P) + z'\gamma(P), \quad (1)$$

where $\beta(P) \in \mathbf{R}^{d_x}$ and $\gamma(P) \in \mathbf{R}^{d_z}$ are unknown parameters. We omit the dependence of β and γ on P whenever it is clear from the context. Define $F_Y(y|x, z)$ as the conditional distribution at $Y = y$ given $X = x$ and $Z = z$ and $f_Y(y|x, z)$ as the corresponding conditional density. We impose the following conditions on the distribution P .

Assumption 2.1. The distribution P is such that

- (a) $0 < E \left[\begin{pmatrix} X_i X_i' & X_i Z_i' \\ Z_i X_i' & Z_i Z_i' \end{pmatrix} f_Y(X_i' \beta + Z_i' \gamma - \epsilon | X_i, Z_i) \right] < \infty$.
- (b) $f_Y(y|x, z)$ exists for all $(y, x, z) \in \mathbb{R} \times \mathbb{R}^{d_x} \times \mathbb{R}^{d_z}$.
- (c) $F_Y(\cdot|x, z)$ is symmetric around $x'\beta + z'\gamma$ for all $(x, z) \in \mathbb{R}^{d_x} \times \mathbb{R}^{d_z}$.
- (d) $f_Y(x'\beta + z'\gamma - \epsilon|x, z) > 0$ for all $(x, z) \in \mathbb{R}^{d_x} \times \mathbb{R}^{d_z}$.
- (e) Define

$$\Gamma = \{ (x, z) \in \mathbb{R}^{d_x} \times \mathbb{R}^{d_z} : \\ y \in [x'\beta + z'\gamma - c, x'\beta + z'\gamma + c] \}.$$

There exists $c > 0$ such that

$$\sup_{(x,z) \in \Gamma} \frac{|f_Y(y|x,z) - f_Y(x'\beta + z'\gamma|x,z)|}{|y - x'\beta - z'\gamma|} < \infty .$$

Assumption 2.1(a), (b) and (e) are commonly imposed in the quantile regression literature (Koenker, 2005; Bai et al., 2019) in order to establish the asymptotic distributions of estimators. Assumption 2.1(c)–(d) are imposed so that the estimators from ℓ_1 -SVR are consistent for the coefficients of linear conditional medians. See Theorem 9.8 of Steinwart and Christmann (2008) for more details.

At times we will require the following homoskedasticity assumption on P . This strong but powerful assumption delivers the pivotal procedure.

Assumption 2.2. The distribution P is such that

- (a) $f_Y(x'\beta + z'\gamma - \epsilon|x,z) = f_Y(x'\beta + z'\gamma - \epsilon)$ for all $(x,z) \in \mathbb{R}^{d_x} \times \mathbb{R}^{d_z}$.
- (b) $F_Y(x'\beta + z'\gamma - \epsilon|x,z) \equiv p_\epsilon$ across all $(x,z) \in \mathbb{R}^{d_x} \times \mathbb{R}^{d_z}$, where p_ϵ is a constant which may depend on ϵ .

Conditions similar to Assumption 2.2(a) are imposed, often implicitly, in the study of regression rankscore tests in quantile regression, so that the density terms cancel in the expression of the limiting variances of test statistics. As Bai et al. (2019) point out, the assumptions are tacitly imposed in the framework of fixed regressors, such as in Gutenbrunner and Jurecková (1992) and Gutenbrunner et al. (1993). Assumption 2.2(b) is imposed so that the test statistic is simpler, but is not required in general. See Remark 3.3 for more details.

Consider the following ℓ_1 -SVR:

$$\min_{(b,r) \in \mathbb{R}^{d_x} \times \mathbb{R}^{d_z}} n^{-1} \sum_{1 \leq i \leq n} \max\{0, |Y_i - X_i' b - Z_i' r| - \epsilon\} + \lambda_n (\|b\|_1 + \|r\|_1) , \tag{2}$$

where

$$\|b\|_1 = \sum_{1 \leq j \leq d_x} |b_j|$$

and similarly for $\|r\|_1$.

As can be seen in Figure 1, the loss functions for median regression and support vector regression are similar, which suggests a close analogy between the methods and means for inference. In support vector regression, the errors are also penalized linearly, but only if they are bigger than ϵ in absolute value. We impose the following condition on the tuning parameter λ_n . It is satisfied when $\lambda_n = \lambda/n$, where λ is a constant.

Assumption 2.3. $\lambda_n \rightarrow 0$ as $n \rightarrow \infty$.

Linear programming duality is at the core of the design of our proposed test statistic. Let $\mathbf{1}_d$ denote the $d \times 1$ vector of 1's. Note that the ℓ_1 -SVR problem (2) has the following primal linear programming

formulation.

$$\max \mathbf{1}'_n \sigma + \lambda_n (\mathbf{1}'_{d_x} b^+ + \mathbf{1}'_{d_x} b^- + \mathbf{1}'_{d_z} r^+ + \mathbf{1}'_{d_z} r^-) \quad (3)$$

$$\begin{aligned} \text{subject to } u - v &= \mathbf{Y}_n - \mathbf{Z}_n r - \mathbf{X}_n b \\ \sigma - s &= u + v - \epsilon \mathbf{1}_n \\ b^+, b^-, r^+, r^-, u, v, \sigma, s &\geq 0, \end{aligned}$$

where the optimization is over $b^+, b^-, r^+, r^-, u, v, \sigma, s$.

It follows from standard duality arguments that the dual of (3) is:

$$\begin{aligned} \max_{a^+, a^-} \mathbf{Y}'_n a^+ + \epsilon \mathbf{1}'_n a^- \quad (4) \\ \text{subject to } -\lambda_n \mathbf{1}_{d_x} &\leq \mathbf{X}'_n a^+ \leq \lambda_n \mathbf{1}_{d_x} \\ -\lambda_n \mathbf{1}_{d_z} &\leq \mathbf{Z}'_n a^+ \leq \lambda_n \mathbf{1}_{d_z} \\ a^- &\leq a^+ \leq -a^- \\ a^- &\in [-1, 0]^n. \end{aligned}$$

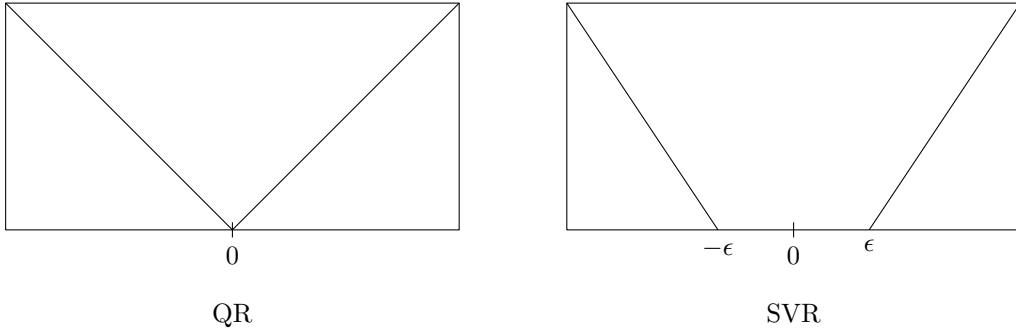


Figure 1: Loss Functions for Median Regression (QR) and for Support Vector Regression (SVR)

3 Inference

Our construction of confidence regions builds on the duality between hypothesis testing and inference. The duality relies on the tautological statement that the confidence region made of all γ_0 's for which the null $H_0 : \gamma = \gamma_0$ fails to be rejected at critical level α has coverage $1 - \alpha$.

Our error bars are obtained by test inversion as follows. Suppose, for simplicity of exposition, that $d_z = 1$. With a test statistic T_n and its null distribution in hand, we may find a critical value $c_{1-\alpha}$ such that

$$\lim_{n \rightarrow \infty} P\{T_n(\mathbf{W}_n, \gamma_0) \leq c_{1-\alpha}\} = 1 - \alpha$$

$\forall P$ such that $\gamma(P) = \gamma_0$ and $\forall \gamma_0 \in \mathbb{R}$. Define the confidence region as

$$C_n = \{\gamma_0 \in \mathbb{R} : T_n(\mathbf{W}_n, \gamma_0) \leq c_{1-\alpha}\} .$$

Theorem 3.2 below ensures that the confidence region C_n is indeed an interval. This guarantees that the formula for the error bars for γ is given by

$$[\underline{\gamma}, \bar{\gamma}] := \left[\min_{\gamma_0 \in C_n} \gamma_0, \max_{\gamma_0 \in C_n} \gamma_0 \right] .$$

For instance, Theorem 3.1 below justifies using $c_{0.95} = 1.96$ for error bars with 95% coverage, with a specific choice of $T_n(\mathbf{W}_n, \gamma_0)$.

We now introduce the test statistic for general values of d_z . For a prespecified $\gamma_0 \in \mathbb{R}^{d_z}$, we are interested in inverting tests of

$$H_0 : \gamma(P) = \gamma_0 \text{ versus } H_1 : \gamma(P) \neq \gamma_0 \quad (5)$$

at level $\alpha \in (0, 1)$.

For that purpose, consider the following “short” ℓ_1 -SVR problem constructed by replacing \mathbf{Y}_n with $\mathbf{Y}_n - \mathbf{Z}_n \gamma_0$ and omitting \mathbf{Z}_n from the regressors.

$$\begin{aligned} \max_{b^+, b^-, u, v, \sigma, s} \quad & \mathbf{1}'_n \sigma + \lambda_n (\mathbf{1}'_{d_x} b^+ + \mathbf{1}'_{d_x} b^-) \\ \text{subject to} \quad & u - v = \mathbf{Y}_n - \mathbf{Z}_n \gamma_0 - \mathbf{X}_n (b^+ - b^-) \\ & \sigma - s = u + v - \epsilon \mathbf{1}_n \\ & b^+, b^-, u, v, \sigma, s \geq 0 . \end{aligned} \quad (6)$$

Define $\hat{\beta}_n$ as $b^+ - b^-$ where b^+ and b^- are part of the solution to (6).

It follows from standard duality arguments that the dual of (6) is

$$\begin{aligned} \max_{a^+, a^-} \quad & (\mathbf{Y}_n - \mathbf{Z}_n \gamma_0)' a^+ + \epsilon \mathbf{1}'_n a^- \\ \text{subject to} \quad & -\lambda_n \mathbf{1}_{d_x} \leq \mathbf{X}'_n a^+ \leq \lambda_n \mathbf{1}_{d_x} \\ & a^- \leq a^+ \leq -a^- \\ & a^- \in [-1, 0]^n . \end{aligned} \quad (7)$$

Denote the solution to (7) by \hat{a}^+ and \hat{a}^- .

We collect a characterization of the complementary slackness condition for the ℓ_1 -SVR problem which will be key in the design of our proposed test statistic.

Lemma 3.1. *The solution to (7) satisfies:*

(a) *If $Y_i - Z'_i \gamma_0 - X'_i \hat{\beta}_n > \epsilon$, then $\hat{a}_i^- = -1$ and $\hat{a}_i^+ = 1$.*

(b) If $Y_i - Z_i'\gamma_0 - X_i'\hat{\beta}_n < -\epsilon$, then $\hat{a}_i^- = -1$ and $\hat{a}_i^+ = -1$.

(c) If $0 < |Y_i - Z_i'\gamma_0 - X_i'\hat{\beta}_n| < \epsilon$, then $\hat{a}_i^+ = \hat{a}_i^- = 0$.

PROOF OF LEMMA 3.1. By complementary slackness,

$$\begin{aligned} |Y_i - Z_i'\gamma_0 - X_i'\hat{\beta}_n| < \epsilon &\Leftrightarrow s_i > 0 \Rightarrow \hat{a}_i^- = 0 \\ |Y_i - Z_i'\gamma_0 - X_i'\hat{\beta}_n| > \epsilon &\Leftrightarrow \sigma_i > 0 \Rightarrow \hat{a}_i^- = -1 \\ Y_i - Z_i'\gamma_0 - X_i'\hat{\beta}_n > 0 &\Leftrightarrow u_i > 0 \Rightarrow \hat{a}_i^+ + \hat{a}_i^- = 0 \\ Y_i - Z_i'\gamma_0 - X_i'\hat{\beta}_n < 0 &\Leftrightarrow v_i > 0 \Rightarrow \hat{a}_i^- - \hat{a}_i^+ = 0, \end{aligned}$$

and the result follows. ■

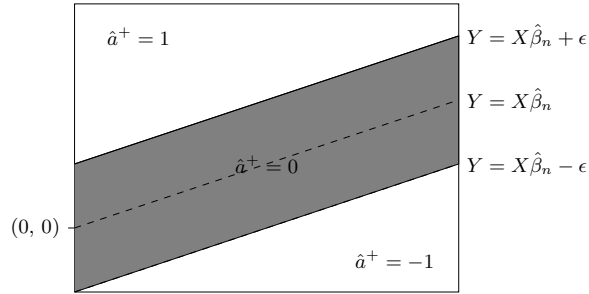


Figure 2: \hat{a}^+ in Different Regions of Regression Residuals $Y - X\hat{\beta}_n$.

Following Bai et al. (2019), we construct the ℓ_1 -SVR rankscore test statistic as

$$T_n(\mathbf{W}_n, \gamma_0) = \frac{n^{-1/2} \mathbf{Z}'_n \hat{a}^+}{\sqrt{n^{-1} \mathbf{Z}'_n \mathbf{M}_n \mathbf{Z}_n \hat{p}_n}}, \quad (8)$$

$$\mathbf{M}_n = I - \mathbf{X}_n (\mathbf{X}'_n \mathbf{X}_n)^{-1} \mathbf{X}'_n \quad (9)$$

and

$$\hat{p}_n = \frac{1}{n} \sum_{1 \leq i \leq n} I\{|Y_i - X_i'\hat{\beta}_n - Z_i'\gamma_0| \geq \epsilon\}. \quad (10)$$

We define the ℓ_1 -SVR rankscore test as

$$\phi_n(\mathbf{W}_n, \gamma_0) = I\{|T_n(\mathbf{W}_n, \gamma_0)| > z_{1-\frac{\alpha}{2}}\}, \quad (11)$$

where $z_{1-\frac{\alpha}{2}}$ is the $(1 - \frac{\alpha}{2})$ -th quantile of the standard normal distribution.

We give two intuitive interpretations for $T_n(\mathbf{W}_n, \gamma_0)$ in (8). Since $\lambda_n \rightarrow 0$, we intuitively view it as 0. First, as shown in Lemma 3.1 and Figure 2, \hat{a}_i^+ is a monotonic transformation of the residuals $\hat{u}_i = Y_i - Z_i'\gamma_0 - X_i'\hat{\beta}_n$. If H_0 holds, i.e., $\gamma(P) = \gamma_0$, then we expect a low correlation between Z and \hat{u} , and any monotonic transformation of \hat{u} . Therefore $\mathbf{Z}'_n \hat{a}^+$ should be small under the null, but larger the more $\gamma(P)$ differs from γ_0 . Second, consider running the SVR of $Y - Z'\gamma_0$ on X , with \hat{a}^+ as the solution to the dual

problem. If H_0 holds, i.e., $\gamma(P) = \gamma_0$, then running a regression of $Y - Z'\gamma_0$ on X and Z should result in an estimated coefficient “close” to 0 on Z . Hence, including Z or not in the regression should have “close” to zero effect on the primal or dual results. Equivalently, adding the constraint $\mathbf{Z}'_n \hat{a}^+ = 0$ to (7) should not change the solution very much, so that $\mathbf{Z}'_n \hat{a}^+ = 0$ holds approximately when the null hypothesis holds, but may be large otherwise.

Remark 3.1. The complementary slackness conditions in (6) and (7) are the key ingredients in the construction of the test statistic in (8). These conditions are summarized in Lemma 3.1. Figure 2 displays the regions defined by regression residuals and the corresponding values of \hat{a}^+ . For simplicity, we assume $d_x = 2$, the estimated intercept is 0, and $\gamma_0 = 0$. Note that unit i contributes to $T_n(\mathbf{W}_n, \gamma_0)$ only when $|Y_i - X_i \hat{\beta}_n| > \epsilon$. The graph is different from that under the median regression, where the shaded region in $\hat{a}_i^+ = 0$ collapses to a single line. ■

Remark 3.2. The ℓ_1 -SVR rankscore test is equivalent to the median regression rankscore test, when $\epsilon = \lambda = 0$ and \hat{p}_n is set to $\frac{1}{2}$. In Section 4, we compare the finite sample performances of the two tests. ■

The following theorem is our main result. It establishes the asymptotic distribution of the test statistic defined in (8) under the null, and as a result, the asymptotic exactness of the test defined in (11), in the sense that the limiting rejection probability under the null is equal to the nominal level.

Theorem 3.1. *Suppose P satisfies Assumption 2.1, λ_n satisfies Assumption 2.3, and P additionally satisfies the null hypothesis, i.e., $\gamma(P) = \gamma_0$. Then,*

$$n^{-1/2} \mathbf{Z}'_n \hat{a}^+ \xrightarrow{d} N(0, 2E[\tilde{Z}_i \tilde{Z}'_i F_Y(X'_i \beta + Z'_i \gamma_0 - \epsilon | X_i, Z_i)]) , \quad (12)$$

where

$$\begin{aligned} \tilde{Z}_i &= Z_i - E[Z_i X'_i f_Y(X'_i \beta + Z'_i \gamma_0 - \epsilon | X_i, Z_i)] \times \\ &\quad E[X_i X'_i f_Y(X'_i \beta + Z'_i \gamma_0 - \epsilon | X_i, Z_i)]^{-1} X_i . \end{aligned}$$

If P additionally satisfies Assumption 2.2, then

$$T_n(\mathbf{W}_n, \gamma_0) \xrightarrow{d} N(0, 1) ,$$

and therefore, for the problem of testing (5) at level $\alpha \in (0, 1)$, $\phi_n(\mathbf{W}_n)$ defined in (11) satisfies

$$\lim_{n \rightarrow \infty} E[\phi_n(\mathbf{W}_n, \gamma_0)] = \alpha .$$

Corollary 3.1. *Suppose P satisfies Assumptions 2.1–2.2 and the null hypothesis, and λ_n satisfies Assumption 2.3. Then the asymptotic variance in (12) can be consistently estimated without density estimation by*

$$\frac{1}{n} \mathbf{Z}'_n \mathbf{M}_n \mathbf{Z}_n \hat{p}_n$$

where \mathbf{M}_n and \hat{p}_n are defined in (9) and (10), respectively.

Remark 3.3. If P violates Assumption 2.2(b) but satisfies Assumption 2.2(a), it is straightforward to construct a consistent estimator of the asymptotic variance in (12) by using the law of iterated expectations, and the statistic studentized by the consistent estimator will be asymptotically exact. The estimator does not involve any bandwidth because it involves only conditional distributions rather than densities. We do not pursue this generalization here. If P violates Assumption 2.2(a), it is also possible to construct consistent estimators of the asymptotic variance in (12), and the statistic studentized by any consistent estimator will be asymptotically exact. See, for example, Powell (1991). The studentization, however, involves density estimation, which is why Assumption 2.2(a) is important for our purposes. ■

According to Theorem 3.1, we could construct confidence regions by inverting the test $\phi_n(\mathbf{W}_n, \gamma_0)$ in (11). The following corollary shows that the limiting coverage probability of the confidence region is indeed correct.

Corollary 3.2. Let $\phi_n(\mathbf{W}_n, \gamma_0)$ denote the test in (11) with level α . Define

$$C_n = \{\gamma_0 \in \mathbb{R} : \phi_n(\mathbf{W}_n, \gamma_0) = 0\} . \quad (13)$$

Suppose P satisfied Assumptions 2.1 and 2.2, and λ_n satisfies Assumption 2.3. Then,

$$\lim_{n \rightarrow \infty} P\{\gamma \in C_n\} = 1 - \alpha .$$

Monotonicity of the test statistic is essential to the tractability of the inversion procedure, as it limits the procedure to a search for the two points where the test statistic, as a function of the posited null parameter, crosses the critical value. Note that without monotonicity, we cannot construct the confidence region as an interval. When $d_z = 1$, the following theorem shows that the test statistic is indeed monotonic, so that C_n is an interval.

Theorem 3.2. If $d_z = 1$, then $T_n(\mathbf{W}_n, \gamma_0)$ in (8) is monotonically decreasing in γ_0 .

PROOF OF THEOREM 3.2. First note that the denominator of $T_n(\mathbf{W}_n, \gamma_0)$ is monotonically increasing in γ_0 because \hat{p}_n in (10) is so. Next, we show the numerator is monotonically decreasing in γ_0 . Denote by $\hat{a}^+(r)$ the solution to (7) with $\gamma_0 = r$. Given $r_1 > r_2$, by definition of the optimization problem,

$$(\mathbf{Y}_n - r_1 \mathbf{Z}_n)' \hat{a}^+(r_1) - (\mathbf{Y}_n - r_1 \mathbf{Z}_n)' \hat{a}^+(r_2) > 0 ,$$

while

$$(\mathbf{Y}_n - r_2 \mathbf{Z}_n)' \hat{a}^+(r_1) - (\mathbf{Y}_n - r_2 \mathbf{Z}_n)' \hat{a}^+(r_2) < 0 .$$

The two observations indicate that

$$r_1 \mathbf{Z}_n' (\hat{a}^+(r_2) - \hat{a}^+(r_1)) > 0 > r_2 \mathbf{Z}_n' (\hat{a}^+(r_2) - \hat{a}^+(r_1)) ,$$

so that

$$(r_1 - r_2)\mathbf{Z}'_n(\hat{a}^+(r_2) - \hat{a}^+(r_1)) > 0,$$

and the result follows since $r_1 > r_2$. ■

4 Simulation

In this section, we compare in a simulation study the finite-sample performance of the median regression rankscore test introduced in Remark 3.2 with that of the ℓ_1 -SVR rankscore test. The regression rankscore test is the default inference procedure for median regression (Koenker et al., 2019), and as such makes for the natural comparison method. In both models, the data is generated according to the following equation,

$$Y = -0.8 + 2X + \gamma Z + G^{-1}(\tau),$$

where

$$(X, Z) \sim N\left(\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 10 & 4 \\ 4 & 8 \end{pmatrix}\right),$$

G^{-1} is the inverse-CDF for the error, τ is uniformly distributed over a subset of the $[0, 1]$ interval, and $(X, Z) \perp\!\!\!\perp \tau$.

We consider five distributions for the error terms: Gaussian, Laplace, a symmetric mixture of Gaussian distributions, Student's t , and χ^2 . The latter three distributions permit us to measure the performance of test when the error distribution exhibits either multiple modes, fat tails, or asymmetry, respectively. For each distribution, we also consider the following two models which differ in the support restrictions on the error terms.

Unrestricted support : $\tau \sim \text{Unif}[0, 1]$.

Restricted support : $\tau \sim \text{Unif}([0, 0.4] \cup [0.6, 1])$.

The restricted support model may be thought of as an extreme version of the data generating processes which SVR is meant for when conceived as a regression extension of SVM classification.

Throughout the simulation, we set $\lambda_n = 0$. The bandwidth ϵ is adjusted according to the distribution of the error so that $G(\epsilon) - G(-\epsilon) = 0.2$. For symmetric distributions of the error term, the bandwidth thus excludes the quantiles in the interval $(0.4, 0.6)$. We set the true parameter $\gamma = 0$ under the null to study size properties, and $\gamma = 0.5$ under the alternative to study power properties when $\alpha = 0.05$. We compare these properties for the SVR rankscore test against its natural competitor, the median regression rankscore test.

Columns 1–4 of Table 1 present the simulations in which the errors are homoskedastic. In all the cases of unrestricted support for the errors, the distributions are centered and scaled to be mean zero with a

Table 1: Rejection Probabilities for Different Distributions of the Errors, Under Homoc/Heteroskedasticity

Distribution	Homoskedastic				Heteroskedastic				Homo. stat, hetero. data			
	SVR		QR		SVR		QR		SVR		QR	
	$\gamma = 0$	$\gamma = 0.5$	$\gamma = 0$	$\gamma = 0.5$	$\gamma = 0$	$\gamma = 0.5$	$\gamma = 0$	$\gamma = 0.5$	$\gamma = 0$	$\gamma = 0.5$	$\gamma = 0$	$\gamma = 0.5$
Gaussian	5.7	34.4	5.4	31.3	4.5	34.9	4.7	30.3	5.5	37.0	5.5	32.4
Gaussian, restricted	5.6	15.7	5.7	13.1	4.7	15.5	4.1	10.9	5.5	17.8	5.4	12.6
Laplace	5.6	64.8	5.3	65.4	4.2	63.6	4.7	67.4	5.5	67.9	5.7	70.4
Laplace, restricted	5.7	23.6	5.6	17.8	4.2	25.2	3.9	17.1	5.3	30.1	5.1	20.3
Mixture	4.6	37.9	4.3	30.6	4.5	37.7	4.4	32.1	4.7	38.8	4.7	32.7
Mixture, restricted	5.2	16.4	5.0	14.5	4.8	18.4	4.6	11.7	5.1	19.1	4.8	13.5
Student's t	5.7	40.1	5.4	35.9	4.4	37.0	4.6	31.7	5.7	40.2	5.3	35.9
Student's t , restricted	5.9	17.1	5.7	14.0	4.8	14.1	4.7	11.7	5.4	16.8	5.6	14.0
χ^2	5.1	39.2	5.4	39.6	4.7	16.3	3.7	15.4	4.7	17.0	4.4	17.3
χ^2 , restricted	5.4	14.3	5.4	14.8	4.8	16.1	4.0	15.6	5.5	16.6	4.7	16.4

standard deviation of 15.² Once restricting the support, the standard deviation may no longer be 15, albeit the errors continue to be mean zero. Columns 1 and 3 indicate that the size properties of the SVR and median regression rankscore tests are about equal under homoskedasticity. However, columns 2 and 4 suggest the SVR rankscore tests exhibits better power properties than median regression, the former outperforming the latter in 7 of the 10 settings.

Columns 5–8 of Table 1 present the simulations in which the errors are heteroskedastic, their variance being determined by covariate X .³ To account for the heteroskedastic errors, we obtain a consistent estimate of the variance expression in (12) using the methodology of Powell (1991). This requires that we estimate the density of the errors. The tuning parameters used in the density estimation are shown in Table 2. The test statistic is then studentized using the variance estimate. We keep the bandwidth ϵ the same as in the homoskedastic simulations. As before, columns 5–8 of Table 1 suggest that the size properties of the two tests are roughly equal under heteroskedasticity, whereas the SVR rankscore test exhibits greater power than the median regression rankscore test in 9 of the 10 cases.

Table 2: Tuning Parameters for Heteroskedastic Test Statistic

Distribution	Unrestricted		Restricted	
	h	κ	h	κ
Gaussian	1	1.75	2.5	2
Laplace	1	0.75	1.5	1.75
Mixture	2	2	2.75	2.5
Student's t	1	1	2	2.75
χ^2	1.5	1.5	2.75	1.5

To gauge the robustness of the SVR rankscore test, columns 9–12 present the simulations in which the homoskedastic test statistic is used with heteroskedastic errors. Similar to the earlier results, columns 9 and 11 suggest both tests have similar size properties. However, columns 10 and 12 reveal that the SVR rankscore test demonstrates greater statistical power than the median regression rankscore test, the former having higher rejection rates in 8 of the 10 simulations.

In each iteration of the simulations in Table 1, we can construct the confidence interval by inverting (11). From the duality between hypothesis testing and inference, the results in Table 1 suggest the SVR rankscore test has greater power against the alternatives considered when inverting the test, as compared to the median regression rankscore test. This would imply tighter confidence intervals under the former test procedure. Indeed, we find this to be the case. Figures 3–5 present the average confidence interval for each simulation where $\gamma = 0.5$. In all three figures, the results closely align with those of Table 1. That is, for

² The mixture consists of evenly weighted Gaussian distributions centered around 10 and -10 , with the same scale parameter. Errors under the Student's t distribution are drawn from the counterpart with 10 degrees of freedom, and then rescaled. Errors under the χ^2 distribution are drawn from the counterpart with 3 degrees of freedom, and then rescaled and recentered.

³The scale parameters of the distributions are set equal to a normalized value of X . For the Gaussian, Laplace, and mixture distributions, X is normalized to have a standard deviation of 1 and is then recentered so its mean is equal to the scale parameter required for the error distribution to have a standard deviation of 15. For the Student's t and χ^2 distributions, X is instead recentered around the degrees of freedom stated in Footnote 2. In the rare event that the normalized X falls below 0, its absolute value is taken. While the distributions of parameters determining the standard deviation of the errors are centered around the value that would correspond to a standard deviation of 15, the standard deviations of the actual error terms need not be 15.

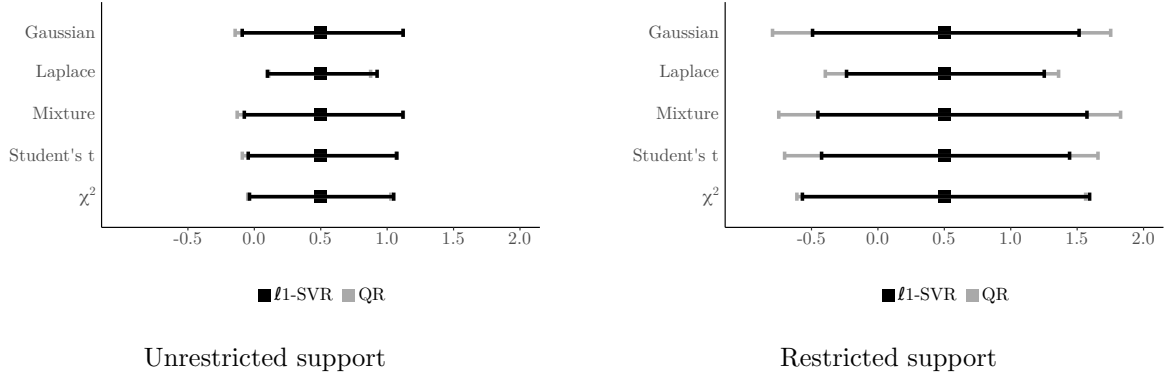


Figure 3: Confidence Intervals Under Homoskedasticity

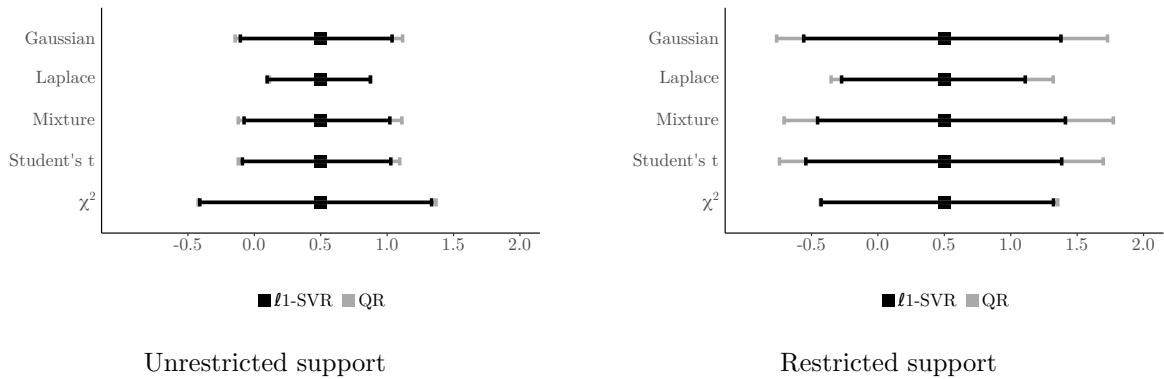


Figure 4: Confidence Intervals Under Heteroskedasticity

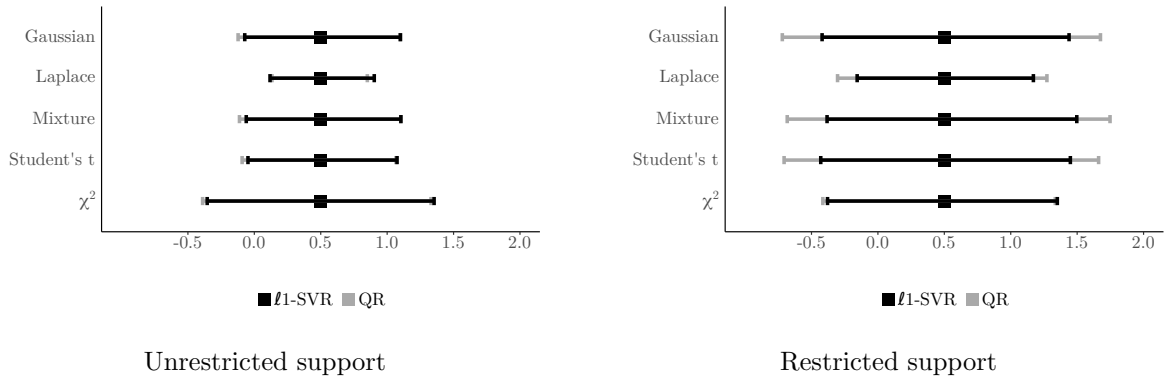


Figure 5: Confidence Intervals Under Heteroskedasticity Using Homoskedastic Test Statistic

the data generating process where the SVR rankscore test has greater power than the median regression rankscore test, the confidence interval of the former is narrower than that of the latter. The reduction in the error bars becomes rather substantial when we restrict the support of the error terms. This is to be expected, as the SVR loss function is able to account for the restricted support of the error term, whereas the median regression loss function cannot (see Figure 1).

It is rather remarkable that modifications to the quantile regression procedure intended for robustness

deliver greater inference accuracy. This naturally suggests using the SVR regression rankscore test for inference in standard quantile regression analysis, even if the point estimate is obtained using quantile regression. Similar robust behavior in heteroskedastic environments of regression rankscore tests constructed under homoskedasticity assumptions are documented using real and simulated data in [Bai et al. \(2019\)](#) and [Pouliot \(2019\)](#).

5 Conclusion

While SVR is largely used in practice, the lack of a large sample inference method which does not require the rather arbitrary choice of a bandwidth parameter was limiting its practicality. In this article, we developed theory and delivered methodology to produce asymptotically valid error bars while circumventing the need to select a bandwidth parameter. The ℓ_1 -SVR rankscore test is the natural analog of the default procedure for inference in quantile regression. The asymptotic theory developed to establish the validity of the error bars is novel for SVR, and may be of independent interest. Unexpectedly, simulation evidence suggests that the regression rankscore test with our proposed test statistic may outperform the standard median regression rankscore test in inference for the regression parameters of the linear median regression function.

Appendix A Proofs

Since all results are derived under the null that $\gamma(P) = \gamma_0$, we assume without loss of generality that $\gamma(P) = \gamma_0 = 0$. We use $a \lesssim b$ to denote that there exists $l > 0$ such that $a \leq lb$. We use $\|\cdot\|$ to denote the Euclidean norm.

A.1 Proof of Theorem 3.1

Follows immediately from Lemma A.5, Lemma A.6, and Lemma A.7. ■

A.2 Proof of Corollary 3.2

Follows immediately from Theorem 3.1 and the duality between hypotheses tests and confidence regions. ■

A.3 Auxiliary Lemmas

Lemma A.1. *Suppose U_i , $1 \leq i \leq n$ are i.i.d. random variables where $E|U_i|^r < \infty$. Then*

$$n^{-1/r} \max_{1 \leq i \leq n} |U_i| \xrightarrow{P} 0 .$$

PROOF OF LEMMA A.1. Note that for all $\eta > 0$,

$$\begin{aligned} P \left\{ n^{-1/r} \max_{1 \leq i \leq n} |U_i| > \eta \right\} &\leq P \left\{ \max_{1 \leq i \leq n} |U_i|^r > n\eta^r \right\} \\ &\leq nP\{|U_i|^r > n\eta^r\} \leq \frac{n}{n\eta^r} E[|U_i|^r I\{|U_i|^r > n\eta^r\}] = \frac{1}{\eta^r} E[|U_i|^r I\{|U_i|^r > n\eta^r\}] \rightarrow 0 , \end{aligned}$$

where the convergence follows from the dominated convergence theorem and $E|U_i|^r \rightarrow 0$. ■

Lemma A.2. *Suppose P satisfies Assumption 2.1(b)-(d). Then*

$$S(b) = E[\max\{|Y - X'b - \epsilon|, 0\}]$$

is uniquely minimized at $b = \beta$.

PROOF OF LEMMA A.2. Follows immediately upon noting that Theorem 9.8 of [Steinwart and Christmann \(2008\)](#) holds under Assumption 2.1(b)-(d). ■

Lemma A.3. *Suppose P satisfies Assumption 2.1(b)-(d) and λ_n satisfies Assumption 2.3. Then,*

$$\hat{\beta}_n \xrightarrow{P} \beta .$$

PROOF OF LEMMA A.3. Define

$$S_n(b) = n^{-1} \sum_{1 \leq i \leq n} \max\{0, |Y_i - X_i'b| - \epsilon\} + \lambda_n \|b\|_1 .$$

To begin with, note that $S_n(b)$ is convex in b . Without any loss of generality suppose $\beta = 0$. For any $\delta > 0$, let B_δ denote the closed δ -ball around 0. By definition,

$$S_n(\hat{\beta}_n) \leq S_n(0) . \quad (14)$$

For all $b \in \mathbb{R}^{d_x} \setminus B_\delta$, we have by convexity that

$$S_n(b_\delta) \leq \frac{\delta}{|b|} S_n(b) + \frac{|b| - \delta}{|b|} S_n(0) ,$$

where

$$b_\delta = \frac{\delta}{|b|} b .$$

Therefore

$$S_n(b) \geq \frac{|b|}{\delta} S_n(b_\delta) - \frac{|b| - \delta}{\delta} S_n(0) . \quad (15)$$

Since $\{b : |b| = 1\}$ is compact and $S(b)$ is continuous in b , we have by Lemma A.2 that there exists $\eta > 0$ such that

$$\min_{b:|b|=1} S(b) \geq S(0) + \eta . \quad (16)$$

By Lemma 2.6.18 of [Van Der Vaart and Wellner \(1996\)](#), $\{b \rightarrow |y - x'b| - \epsilon + \lambda_n \|b\|_1 : |b| = 1\}$ is a VC class, thus Donsker, and thus Glivenko-Cantelli, i.e.,

$$\sup_{b:|b|=1} |S_n(b) - S(b)| = o_P(1) . \quad (17)$$

Combining (15), (16), and (17), and that $S_n(0) \xrightarrow{P} S(0)$, we have that

$$\{|\hat{\beta}_n - \beta| > \delta\} \Rightarrow \{S_n(\hat{\beta}_n) \geq S_n(0) + \eta + o_P(1)\} ,$$

which has probability approaching zero because of (14). ■

Lemma A.4. *Suppose P satisfies Assumption 2.1(a)-(e). Then,*

$$n^{1/2}(\hat{\beta}_n - \beta) = \frac{1}{2} E[X_i X_i' f_Y(X_i' \beta - \epsilon | X_i, Z_i)]^{-1} n^{-1/2} \sum_{1 \leq i \leq n} X_i (I\{Y_i - X_i' \beta > \epsilon\} - I\{Y_i - X_i' \beta < -\epsilon\}) + o_P(1) .$$

PROOF OF LEMMA A.4. Define

$$\hat{L}_n = n^{-1/2} \sum_{1 \leq i \leq n} X_i (I\{Y_i - X_i' \hat{\beta}_n > \epsilon\} - I\{Y_i - X_i' \hat{\beta}_n < -\epsilon\}) \quad (18)$$

$$L_n = n^{-1/2} \sum_{1 \leq i \leq n} X_i (I\{Y_i - X_i' \beta > \epsilon\} - I\{Y_i - X_i' \beta < -\epsilon\}) . \quad (19)$$

To begin with, note that

$$\left| n^{-1/2} \sum_{1 \leq i \leq n} X_i \hat{a}_i^+ - \hat{L}_n \right| \lesssim n^{-1/2} \max_{1 \leq i \leq n} |X_i| \sum_{1 \leq i \leq n} (I\{Y_i = X_i' \hat{\beta}_n\} + I\{Y_i - X_i' \hat{\beta}_n = \epsilon\}) = o_P(1) , \quad (20)$$

because of by Lemma 3.1, Lemma A.1, Assumption 2.1(a), and that the number of support vectors are bounded. By

similar arguments,

$$\left| n^{-1/2} \sum_{1 \leq i \leq n} X_i \hat{\alpha}_i^+ \right| \leq \left| n^{-1/2} \lambda_n \mathbf{1}_{d_x} \right| + o_P(1) = o_P(1), \quad (21)$$

where the second equality follows from (7) and the last follows from Lemma A.1. Next, we write

$$\begin{aligned} \hat{L}_n - L_n &= n^{-1/2} \sum_{1 \leq i \leq n} X_i (I\{Y_i \leq X_i' \beta + \epsilon\} - I\{Y_i \leq X_i' \hat{\beta}_n + \epsilon\}) \\ &\quad + n^{-1/2} \sum_{1 \leq i \leq n} X_i (I\{Y_i \leq X_i' \beta - \epsilon\} - I\{Y_i \leq X_i' \hat{\beta}_n - \epsilon\}) \\ &= R_{1,n} + R_{2,n} + R_{1,n}^- + R_{2,n}^-, \end{aligned} \quad (22)$$

where

$$\begin{aligned} R_{1,n} &= n^{-1/2} \sum_{1 \leq i \leq n} X_i I\{Y_i \leq X_i' \beta + \epsilon\} - E[X_i I\{Y_i \leq X_i' \beta + \epsilon\}] \\ &\quad - (X_i I\{Y_i \leq X_i' \hat{\beta}_n + X_i' t + \epsilon\} - E[X_i I\{Y_i \leq X_i' \beta + X_i' t + \epsilon\}])|_{t=\hat{\beta}_n - \beta} \\ R_{2,n} &= n^{1/2} E[X_i (I\{Y_i \leq X_i' \beta + \epsilon\} - I\{Y_i \leq X_i' \beta + n^{-1/2} X_i' t + \epsilon\})]_{t=n^{1/2}(\hat{\beta}_n - \beta)}, \end{aligned}$$

and similarly for $R_{1,n}^-$ and $R_{2,n}^-$.

Since $\hat{\beta}_n - \beta = o_P(1)$ by Lemma A.3, $R_{1,n} \xrightarrow{P} 0$ by the similar arguments as those used in the last part of the proof of Lemma A.1 of Bai et al. (2019). For $R_{2,n}$, note that

$$\begin{aligned} R_{2,n} &= n^{1/2} E[X_i (F_Y(X_i' \beta + \epsilon | X_i, Z_i) - F_Y(X_i' \beta + n^{-1/2} X_i' t + \epsilon | X_i, Z_i))]_{t=n^{1/2}(\hat{\beta}_n - \beta)} \\ &= -n^{1/2} E[X_i n^{-1/2} X_i' t f_Y(X_i' \beta + \epsilon + s_i n^{-1/2} X_i' t | X_i, Z_i)]_{t=n^{1/2}(\hat{\beta}_n - \beta)} \\ &= -n^{1/2} (\hat{\beta}_n - \beta) E[X_i X_i' f_Y(X_i' \beta + \epsilon + s_i n^{-1/2} X_i' t | X_i, Z_i)]_{t=n^{1/2}(\hat{\beta}_n - \beta)} \end{aligned} \quad (23)$$

where $s_i \in [0, 1]$ is a random variable. The first equality above holds by the law of iterated expectation and the second holds by the mean-value theorem. A similar decomposition holds for $R_{1,n}^-$ and $R_{2,n}^-$.

We then argue that $L_n = O_P(1)$. Indeed, by Assumption 2.1(c), the conditional distributions are symmetric so that the individual terms of L_n are i.i.d. mean zero and therefore $L_n = O_P(1)$ by the central limit theorem.

By applying similar arguments as those used to establish Lemma A.2 of Bai et al. (2019), where assumptions are satisfied under Assumption 2.1(a)-(e), and noting that $L_n = O_P(1)$, it follows from (18), (19), (20), (21), and (23) that

$$n^{1/2}(\hat{\beta}_n - \beta) = O_P(1). \quad (24)$$

and

$$n^{1/2}(\hat{\beta}_n - \beta) (E[X_i X_i' (f_Y(X_i' \beta + \epsilon | X_i, Z_i) + f_Y(X_i' \beta - \epsilon | X_i, Z_i))] + o_P(1)) = -L_n + o_P(1). \quad (25)$$

The proof is finished by plugging (24) in (25), and noting that

$$f_Y(X_i' \beta + \epsilon | X_i, Z_i) = f_Y(X_i' \beta - \epsilon | X_i, Z_i)$$

by Assumption 2.1(c). ■

Lemma A.5. *Suppose P satisfies Assumption 2.1(a)-(d) and ϵ and λ_n satisfies Assumption 2.3. Then,*

$$n^{-1/2} \mathbf{Z}'_n \hat{a}^+ \xrightarrow{d} N(0, 2E[\tilde{Z}_i \tilde{Z}'_i F_Y(X'_i \beta - \epsilon | X_i, Z_i)]) ,$$

where

$$\tilde{Z}_i = Z_i - E[Z_i X'_i f_Y(X'_i \beta - \epsilon | X_i, Z_i)] E[X_i X'_i f_Y(X'_i \beta - \epsilon | X_i, Z_i)]^{-1} X_i$$

PROOF OF LEMMA A.5. It follows from Lemma A.3, Lemma A.4, and similar arguments used to establish Lemma A.1 of Bai et al. (2019) that

$$n^{-1/2} \mathbf{Z}'_n \hat{a}^+ = n^{-1/2} \sum_{1 \leq i \leq n} \tilde{Z}_i (I\{Y_i - X'_i \beta > \epsilon\} - I\{Y_i - X'_i \beta < -\epsilon\}) + o_P(1) .$$

Note that that

$$\begin{aligned} & E[(I\{Y_i - X'_i \beta > \epsilon\} - I\{Y_i - X'_i \beta < -\epsilon\})^2 | X_i, Z_i] \\ &= E[I\{Y_i \geq X'_i \beta + \epsilon\} + I\{Y_i \leq X'_i \beta - \epsilon\} | X_i, Z_i] \\ &= 2E[F_Y(X'_i \beta - \epsilon | X_i, Z_i)] , \end{aligned}$$

where the last equality follows from Assumption 2.1(c). The lemma now follows from the Central Limit Theorem and Assumption 2.1(a). ■

Lemma A.6. *Suppose P satisfies Assumption 2.1(a) and Assumption 2.2(a). Then,*

$$n^{-1} \mathbf{Z}'_n \mathbf{M}_n \mathbf{Z}_n \xrightarrow{P} E[Z_i Z'_i] - E[Z_i X'_i] E[X_i X'_i]^{-1} E[X_i Z'_i] .$$

PROOF OF LEMMA A.6. Follows from Assumption 2.1(a), Assumption 2.2(a), and an application of the weak law of large numbers. ■

Lemma A.7. *Suppose P satisfies Assumptions 2.1(a)-(d) and 2.2, and λ_n satisfies Assumption 2.3. Then,*

$$\hat{p}_n \xrightarrow{P} 2p_\epsilon .$$

PROOF OF LEMMA A.7. We consider

$$\frac{1}{n} \sum_{1 \leq i \leq n} I\{Y_i \leq X'_i \beta + Z'_i \gamma_0 - \epsilon + X'_i (\hat{\beta}_n - \beta)\} ,$$

and the other half follows similarly. By Lemma A.3, since Assumptions 2.1(a)-(d) and 2.3 hold, $\hat{\beta}_n \xrightarrow{P} \beta$. Fix $\eta > 0$. For any $\delta > 0$, consider the empirical process indexed by the class of functions

$$\{t \rightarrow I\{y \leq x' \beta + z' \gamma_0 - \epsilon + x' t\} : \|t\| \in [0, \delta]\} .$$

It is easy to see the class of functions is VC by Lemma 9.12 of Kosorok (2008), so that is Donsker hence Glivenko-Cantelli by Theorem 2.6.7 of Van Der Vaart and Wellner (1996), i.e.,

$$\sup_{t \in [0, \delta]} \left| \frac{1}{n} \sum_{1 \leq i \leq n} I\{Y_i \leq X'_i \beta + Z'_i \gamma_0 - \epsilon + X'_i t\} - P\{Y_i \leq X'_i \beta + Z'_i \gamma_0 - \epsilon + X'_i t\} \right| \xrightarrow{P} 0 .$$

Next,

$$P\{Y_i \leq X_i'\beta + Z_i'\gamma_0 - \epsilon + X_i't\}$$

is continuous at $t = 0$. Since $\hat{\beta}_n - \beta = o_P(1)$, with probability approaching 1, $\|\hat{\beta}_n - \beta\| \leq \delta$, and therefore

$$\left| \frac{1}{n} \sum_{1 \leq i \leq n} I\{Y_i \leq X_i'\beta + Z_i'\gamma_0 - \epsilon + X_i't\} - P\{Y_i \leq X_i'\beta + Z_i'\gamma_0 - \epsilon\} \right| \leq \eta + \eta_\delta,$$

where $\eta_\delta \rightarrow 0$ as $\delta \rightarrow 0$. Let $\delta \rightarrow 0$ to finish the proof. ■

References

- Bai, Y., G. A. Pouliot, and A. M. Shaikh (2019). On regression rankscore inference. Working paper.
- Basak, D., S. Pal, and D. C. Patranabis (2007). Support vector regression. *Neural Information Processing-Letters and Reviews* 11(10), 203–224.
- Bertsimas, D. and J. N. Tsitsiklis (1997). *Introduction to linear optimization*, Volume 6. Athena Scientific Belmont, MA.
- Cherkassky, V. and Y. Ma (2004). Practical selection of svm parameters and noise estimation for svm regression. *Neural networks* 17(1), 113–126.
- Drucker, H., C. J. Burges, L. Kaufman, A. J. Smola, and V. Vapnik (1997). Support vector regression machines. In *Advances in neural information processing systems*, pp. 155–161.
- Gao, J. B., S. R. Gunn, C. J. Harris, and M. Brown (2002). A probabilistic framework for svm regression and error bar estimation. *Machine Learning* 46(1-3), 71–89.
- Ghorbani, M., G. Zargar, and H. Jazayeri-Rad (2016). Prediction of asphaltene precipitation using support vector regression tuned with genetic algorithms. *Petroleum* 2(3), 301–306.
- Gutenbrunner, C. and J. Jurečková (1992). Regression rank scores and regression quantiles. *The Annals of Statistics*, 305–330.
- Gutenbrunner, C., J. Jurečková, R. Koenker, and S. Portnoy (1993). Tests of linear hypotheses based on regression rank scores. *Nonparametric Statistics* 2(4), 307–331.
- Jiao, R., Y. Li, Q. Wang, and B. Li (2005). Svm regression and its application to image compression. In *International Conference on Intelligent Computing*, pp. 747–756. Springer.
- Koenker, R. (2005). *Quantile Regression*. Econometric Society Monographs. Cambridge University Press.
- Koenker, R. and J. A. Machado (1999). Goodness of fit and related inference processes for quantile regression. *Journal of the american statistical association* 94(448), 1296–1310.
- Koenker, R., S. Portnoy, P. T. Ng, A. Zeileis, P. Grosjean, and B. D. Ripley (2019). R package `quantreg`.
- Kosorok, M. R. (2008). *Introduction to empirical processes and semiparametric inference*. Springer.
- Law, M. H. and J. T.-Y. Kwok (2001). Bayesian support vector regression. In *AISTATS*. Citeseer.
- Li, J., S. West, and G. Platt (2012). Power decomposition based on svm regression. In *2012 Proceedings of International Conference on Modelling, Identification and Control*, pp. 1195–1199. IEEE.
- Li, P., Z. Tan, L. Yan, and K. Deng (2011). Time series prediction of mining subsidence based on a svm. *Mining Science and Technology (China)* 21(4), 557–562.

- Peloso, R., A. Miller, P. Allen, and T. Jebara (2004). An svm learning approach to robotic grasping. In *IEEE International Conference on Robotics and Automation, 2004. Proceedings. ICRA'04. 2004*, Volume 4, pp. 3512–3518. IEEE.
- Pouliot, G. (2018). Equivalence of multicategory svm and simplex cone svm: Fast computations and statistical theory. In *International Conference on Machine Learning*, pp. 4130–4137.
- Pouliot, G. A. (2019). Instrumental variables quantile regression with multivariate endogenous variable. Working paper.
- Powell, J. L. (1991). Estimation of monotonic regression models under quantile restrictions. *Nonparametric and semiparametric methods in Econometrics*, 357–384.
- Sánchez, A. S., P. G. Nieto, P. R. Fernández, J. del Coz Díaz, and F. J. Iglesias-Rodríguez (2011). Application of an svm-based regression model to the air quality study at local scale in the avilés urban area (spain). *Mathematical and Computer Modelling* 54(5-6), 1453–1466.
- Smola, A. J. and B. Schölkopf (2004). A tutorial on support vector regression. *Statistics and computing* 14(3), 199–222.
- Steinwart, I. and A. Christmann (2008). *Support vector machines*. Springer Science & Business Media.
- Van Der Vaart, A. W. and J. A. Wellner (1996). Weak convergence. In *Weak convergence and empirical processes*, pp. 16–28. Springer.
- Vapnik, V. (1998). *Statistical Learning Theory*. John Wiley & Sons Inc., New York.
- Xi, X.-C., A.-N. Poo, and S.-K. Chou (2007). Support vector regression model predictive control on a hvac plant. *Control Engineering Practice* 15(8), 897–908.