Abstract

Quantiles and expected shortfalls are commonly used risk measures in financial risk management. The two measurements are correlated while having distinguished features. In this project, our primary goal is to develop a stable and practical inference method for the conditional expected shortfall. We consider the joint modelling of conditional quantile and expected shortfall to facilitate the statistical inference procedure. While the regression coefficients can be estimated jointly by minimizing a class of strictly consistent joint loss functions, the computation is challenging, especially when the dimension of parameters is large since the loss functions are neither differentiable nor convex. We propose a two-step estimation procedure to reduce the computational effort by first estimating the quantile regression parameters with standard quantile regression. We show that the two-step estimator has the same asymptotic properties as the joint estimator, but the former is numerically more efficient. We develop a score-type inference method for hypothesis testing and confidence interval construction. Compared to the Wald-type method, the score method is robust against heterogeneity and is superior in finite samples, especially for cases with many confounding factors. The advantages of our proposed method over existing approaches are demonstrated by simulations and empirical studies based on income and college education data.

KEYWORDS
expected shortfall, quantile, score test, two-step estimation

1 | INTRODUCTION

A tail quantile of the profit-and-loss distribution, known as value at risk (VaR), measures the risk of loss for investments. It has been widely applied for capital allocation and risk management (McNeil et al., 2015). Although an intuitive measure, VaR has been criticized since it fails to capture tail risks beyond itself. Expected shortfall (ES), defined as the average above or below a specific quantile, fulfils such a deficiency as it better characterizes the tail behaviour by consolidating information from the entire tail region. In addition, ES has the desired property of subadditivity which VaR lacks in general (Artzner, 1997; Artzner et al., 1999). Due to these appealing features, ES has attracted increasing attention in recent years and has become a more widely used risk management tool. In financial risk management, the Basel Committee (2013) recommended shifting the quantitative risk metrics system from VaR to ES.

In many applications, risk measures might depend on exogenous covariates. For instance, market risks often change across investment conditions, such as macroeconomic, financial and political environments. Most clinical studies usually associate patient outcomes with demographic and therapeutic information. It is thus of interest to focus on the conditional risk measures adjusting for specific covariates. In this project, we consider the problem of inference for the conditional expected shortfall (CES).

He et al. (2010) introduced a COVariate-adjusted Expected Shortfall (COVES) test to detect treatment effects through CES, which is motivated by a clinical study with a balanced design. However, as shown in Sections 3 and 4, the statistical power of the COVES test may be
affected when there are unbalanced covariates. We consider inference based on a regression framework to evaluate CES beyond the scope of treatment differences. As pointed out by Gneiting (2011), CES is not ‘elicitable’ in the sense that it cannot be represented as the minimizer of an expected loss, and hence, the stand-alone regression for CES is infeasible. To overcome the problem of ‘elicitability’, Leorato et al. (2012) and Peracchi and Tanase (2008) suggested approximating CES by fitting an entire quantile process, which imposes both computational and theoretical challenges. Alternative methods such as those proposed by Cai and Wang (2008), Kato (2012) and Xiao (2014) rely on kernel-smoothing estimation for the conditional distribution function, which is subject to the ‘curse-of-dimensionality’ and practically feasible only for data with a few covariates.

For inference on CES, we consider an alternative approach through the joint modelling of conditional quantile and CES. Fissler and Ziegel (2016) recently showed that VaR and ES are jointly ‘elicitable’, and they provided a class of strictly consistent joint loss functions for the pairs of quantile and ES at the same probability level. Dimitriadis and Bayer (2019) utilized the joint loss functions in a regression setup for quantile and ES. The computation of the joint estimator is challenging especially when the dimension of parameters is large since the joint loss function is neither differentiable nor convex. To reduce the computational effort, we propose a two-step estimation procedure. We first estimate the quantile parameters with standard quantile regression (Koenker, 2005) and then estimate the ES regression coefficients by minimizing the simplified objective function with the quantile estimators plugged in. We show that the two-step estimator has the same asymptotic properties as the joint estimator, but the former is numerically more efficient. In addition, the CES estimation in the second step is locally robust to the quantile estimation in the first step, which implies that the local misspecification of the quantile parameters has no effect on the asymptotic distribution of the ES estimator; see Chernozhukov et al. (2016) for an elaboration on local robustness.

The Wald-type inference method can be conducted based on the asymptotic distribution of the parameter estimator. However, it has been shown in quantile regression literature that the Wald-type test is generally unstable for small sample sizes, partly due to the uncertainty from estimating nuisance parameters involved in the asymptotic variance, such as the conditional densities of the response (Chen & Wei, 2005; Kocherginsky et al., 2005). We develop a score-type inference method for hypothesis testing and confidence interval construction. Numerical studies suggest that the proposed score-type method is superior to the Wald-type method in finite samples, especially when the data are heterogeneous and involve a large number of confounding factors. Furthermore, the method provides more accurate results than the COVES approach for unbalanced design.

In Section 2, we first present the two-step estimation procedure for the joint-regression model and the large sample properties of the resulting estimators and then develop the score-type inference method for the ES regression parameters. We assess the finite sample performance of the proposed inference procedure with simulation studies in Section 3. The merit of the proposed method is illustrated by analysing two real data sets in Section 4. Some concluding remarks are provided in Section 5. All proofs are contained in the Supporting Information.

2 | PROPOSED METHOD

2.1 | Joint-regression models

Consider a continuous response $Y$ and a $p$-dimensional design vector $X$. Suppose that $\{(Y_i, X_i), i = 1, ..., n\}$ is an independent and identically distributed (i.i.d.) sample of $(Y, X)$. At a given probability level $\tau \in (0, 1)$, the conditional quantile of $Y$ given $X$ is defined as $Q_{\tau}(Y|X) = \inf\{y \in \mathbb{R} : F_{Y|X}^{-1}(y) \geq \tau\}$, where $F_{Y|X}$ is the conditional distribution function of $Y$ given $X$. The corresponding (left tail) $\tau$th CES is defined as

$$ES_{\tau}(Y|X) = \tau^{-1} \int_0^\tau F_{Y|X}^{-1}(u)du.$$ 

Artzner et al. (1999) proposed a set of four desirable properties for risk measures, including monotonicity, translation invariance, homogeneity and subadditivity. As ES satisfies these four conditions, it is referred to as a ‘coherent’ risk measurement, whereas the conditional quantile does not always meet the fourth condition. Our objective in this paper is to develop inference methods for the CES of $Y$ given $X$ at a certain probability level under a regression framework. As pointed out by Gneiting (2011), the stand-alone regression for CES is infeasible since the CES cannot be represented as the minimizer of an expected loss. To overcome this problem, we adopt the idea in Fissler and Ziegel (2016) and employ a joint-regression framework that simultaneously models the conditional quantile and CES.

In order to simplify presentation, let $X$ denote the design vector for both quantile and ES regression models. However, each model can have its own design vector. At a given probability level $\tau \in (0, 1)$, we jointly model the conditional quantile and CES of $Y$ given $X$ as

$$Q_{\tau}(Y|X) = X'\theta^q_\tau \text{ and } ES_{\tau}(Y|X) = X'\theta^e_\tau.$$ 

(1)
where the parameter vector \( \theta_0 = (\theta_0^*, \theta_0^j)^t \) is \( \tau \)-specific. Denote \( u^q = Y - Q_{0j}(X) \) and \( u^\tau = Y - ES_j(X) \). We assume \( Q_j(u^q|X) = ES_j(u^\tau|X) = 0 \) for identifiability purpose.

To estimate the regression coefficients, we utilize the class of strictly consistent joint loss functions for the pair of quantile and ES (Fissler & Ziegel, 2016),

\[
\rho_i(Y, X, \theta) = \{I(Y \leq X^\theta^q) - \tau\} G_1(X^\theta^q) - I(Y \leq X^\theta^\tau) G_1(Y) \\
+ G_2(X^\theta^q) \left\{ X^\theta^q - X^\theta^\tau + \frac{(X^\theta^q - Y)I(Y \leq X^\theta^q)}{\tau} \right\} \\
- G_2(X^\theta^q) + a(Y),
\]

(2)

where \( G_1 \) is an increasing and twice continuously differentiable function, \( G_2 \) is a three-times continuously differentiable function, \( G_2^{(1)} = G_2 \), \( G_1 \) and \( G_2^{(1)} \) are strictly positive and \( G_1 \) and \( a \) are integrable functions. The first line in (2) is a strictly consistent loss function for the quantile (Gneiting, 2011) and only depends on the quantile, but the second line involves both ES and quantile and cannot be split into two separate parts. Fissler and Ziegel (2016) showed that, under some regularity conditions, there exist no strictly consistent loss functions outside the class of functions given above, which implies that (2) is the most general class of objective functions that can be applied for the joint-regression model. Given data \( \{Y_i, X_i\} \), the corresponding joint estimators \( \hat{\theta} = (\hat{\theta}^*, \hat{\theta}^j)^t \) can be obtained by

\[
\hat{\theta} = \arg\min_{\theta} \sum_{i=1}^n \rho_i(Y_i, X_i, \theta).
\]

(3)

Dimitriadis and Bayer (2019) utilized the joint loss functions in a regression setup and proposed to estimate the quantile and ES parameters jointly by (3). However, the one-step procedure is computationally challenging, especially when the dimension of parameters is large since the loss functions are neither differentiable nor convex. To reduce the computational effort, we propose a two-step estimation procedure, which first estimates the quantile parameters using standard quantile regression and then estimates the CES parameters by minimizing the joint loss function with the quantile estimators plugged into it.

2.2 Two-step estimation

The first part of the joint loss function (2) depends only on the quantile. If the quantile parameters are known by ‘oracle’, we can plug the quantile parameters into the joint loss function and use only the second part to estimate the CES parameters, thus effectively reducing the computational complexity. Let \( \hat{\theta}^q \) be a consistent estimator of \( \theta_0^q \); then, the first part in (2) is fixed given the quantile estimate. Also note that the function \( a(Y) \) depends only on \( Y \). Therefore, we can estimate the CES parameters by minimizing the following simpler plug-in objective function,

\[
\rho_i(Y, X, \hat{\theta}^q, \theta^j) = G_2(X^\hat{\theta}^q) \left\{ X^\hat{\theta}^q - X^\theta^j + \frac{(X^\hat{\theta}^q - Y)I(Y \leq X^\hat{\theta}^q)}{\tau} \right\} \\
- G_2(X^\theta^q),
\]

(4)

leading to the CES estimator

\[
\hat{\theta}^j = \arg\min_{\theta^j} \sum_{i=1}^n \rho_i(Y_i, X_i, \hat{\theta}^q, \theta^j).
\]

(5)

In our implementation, we obtain the first-step estimator \( \hat{\theta}^q \) by standard quantile regression using the quantreg package in R. Compared to the joint estimator \( \hat{\theta} \), the two-step estimator \( \hat{\theta}^q \) is computationally more efficient because the minimization is performed over a smaller dimension. Moreover, we can show that the two estimators are asymptotically equivalent under the following regularity assumptions.

A1. The matrix \( E(XX') \) is positive definite.
A2. The conditional distribution of \( Y \) given \( X \), \( F_{Y|X}() \), has finite second moment and is absolutely continuous with a continuous density \( f_{Y|X} \), which is strictly positive, continuous and bounded in a neighbourhood of the \( r \)th conditional quantile of \( Y \).
A3. The class of strictly consistent joint loss functions is given by (2), where \( G_1 \) is an increasing and twice continuously differentiable function, \( G_2 \) is a three-times continuously differentiable function, \( G_2^{(1)} = G_2 \), \( G_1 \) and \( a \) are integrable functions.
A4. \( \hat{\theta}^q \) is a \( \sqrt{n} \)-consistent estimator of \( \theta_0^q \).
Theorem 1. Under Assumptions A1–A4 and the moment conditions \((\mathcal{M}-1)\) in the Supporting Information, we have

\[
\sqrt{n} \left( \hat{\theta} - \theta_0 \right) \overset{d}{\to} N(0, \Lambda^{-1} \Omega^{-1}),
\]

where \(\theta_0 = (\theta^0_q, \theta^0_p)^\top\) is the true parameter vector and

\[
\Lambda = E \left\{ \{XX^\top \} G_2^{(2)}(X \theta^0_p) \right\},
\]

\[
\Omega = E \left\{ \{XX^\top \} \left( \frac{\partial}{\partial \theta^0_q} \right)^2 \} \times \left\{ \frac{1}{\tau} \psi(u^\tau) + \frac{1-\tau}{\tau} \psi^2 (X \theta^0_p, \theta^0_q) \right\} \right\},
\]

\[
\psi(u^\tau) = \text{Var}(u^\tau | u^\tau \leq 0, X) = \text{Var}(Y - \mathbf{X} \theta^0_q | Y \leq \mathbf{X} \theta^0_q, \mathbf{X}),
\]

\[
\phi(X \theta^0_q, \theta^0_p) = \mathbf{X} \theta^0_q - \mathbf{X} \theta^0_p.
\]

Assumption A1 is required to exclude the multicollinearity of the stochastic explanatory variables. Assumption A2 includes standard assumptions in mean and quantile regression, and the finite conditional moment of \(Y\) given \(X\) is assumed since ES is a truncated mean of quantiles. The conditions on functions \(G_1\) and \(G_2\) are required for the strictly consistent joint loss functions. Assumption A4 can be relaxed to \(\sqrt{n} \| \hat{\theta}^\tau - \theta^0_q \|^2 = o_p(1)\): see discussions in the proof of Lemma 3 in the Supporting Information. Under the linear quantile regression model (1) and A1–A2, the quantile estimator \(\hat{\theta}^\tau\) from the standard quantile regression was shown to be \(\sqrt{n}\)-consistent; see Koenker (2005). The efficiency of the estimator depends on the form of specification functions \(G_1\) and \(G_2\). Dimitriadis and Bayer (2019) discusses several feasible choices, and their simulation analysis suggested that \(G_1(z) = z\) and \(G_2(z) = - \log(-z)\) provide the most consistent estimation results under all scenarios considered. We will employ these two functions in our implementation.

Remark 1. Dimitriadis and Bayer (2019) established the asymptotic normality of the joint estimators \((\hat{\theta}^\tau, \hat{\theta}^\delta)\) and showed that the two estimators are asymptotically independent. Theorem 1 implies that the proposed two-step estimator \(\hat{\theta}^\delta\) is asymptotically equivalent to the joint estimator \(\hat{\theta}^\tau\), but the former is obtained with two steps and is numerically more efficient. In addition, the error involved in the quantile estimation \(\hat{\theta}^\delta\) in the first step does not affect the asymptotic distribution of \(\hat{\theta}^\tau\). This asymptotic independence result follows due to the Neyman orthogonality (Foster & Syrgkanis, 2019; Neyman, 1959, 1979), that is,

\[
\bar{\Omega}^2 E \left( \frac{\partial^2 E(p(Y, X \theta^0_q, \theta^0_p) | X) \big| \theta^0_q \neq \theta^0_q}{(\theta^0_q - \theta^0_q) \cdot \theta^0_q} \right) = (XX^\top) G_2^{(1)}(X \theta^0_q) F_{\rho \times (X \theta^0_q)} - \tau \frac{\theta^0_q}{\theta^0_q - \theta^0_q} = 0.
\]

Therefore, when \(\sqrt{n} \| \hat{\theta}^\tau - \theta^0_q \|^2 = o_p(1)\), \(\hat{\theta}^\tau\) is locally robust to the prior quantile estimation or its local misspecification; see Chernozhukov et al. (2016) for an elaboration on local robustness. Furthermore, even though we assume both linear models for quantile and ES regression, the local robustness property enables us to consider more general models for the quantile estimation in the first step. For instance, the conditional quantile in the first step can be obtained by nonparametric regression, and this will not affect the asymptotic property of the two-step estimator \(\hat{\theta}^\delta\) as long as the ES regression model is correctly specified and the conditional quantile estimation is consistent with a certain rate.

Remark 2. The major computational challenge of the joint estimation in Dimitriadis and Bayer (2019) is that the method requires minimizing a joint loss function that is neither differentiable nor convex. On the other hand, the joint loss function is sufficiently smooth, so the first-step quantile estimator provides a good initialization for optimizing over the ES parameter \(\theta^\delta\). Therefore, our proposed procedure overcomes this computational hurdle by estimating in two steps and reducing parameter space in the more complex optimization problem in the second step. Barendse (2020) introduced an alternative two-step estimator that employs quantile regression in the first step and weighted least-squares for estimating the ES parameters in the second step. It is known that estimators based on minimizing \(L_2\)-type losses are often susceptible to outliers and heavy-tailed errors. In contrast, the second step of our proposed method is based on minimizing the plug-in objective function (4) that involves no squared term; thus, the estimation procedure is more tail-robust than the two-step method in Barendse (2020).

2.3 Estimation of the asymptotic variance of \(\hat{\theta}^\delta\)

Based on the asymptotic normality of the two-step estimator \(\hat{\theta}^\delta\), we can construct a Wald-type test on \(\theta^0_q\) by directly estimating the asymptotic covariance matrix. We can easily estimate the matrix \(\Lambda\) by \(\hat{\Lambda} = \frac{n}{n-1} \sum_{i=1}^n (XX_i) \times G_2^{(1)}(X \hat{\theta}^\delta_i)\). However, the \(\Omega\) matrix involves a nuisance function,
the conditional variance of the quantile residuals $\psi(u^q)$, for which accurate estimation is challenging. First of all, for tail quantile levels, for example, $\tau$ close to 0, corresponding to the left tail, there exist very few (about $\eta t$) observations after conditional on $u^q \leq 0$. Also, taking into account the dependence on the covariates $X$ further complicates the estimation in finite samples.

We adopt two approaches from Dimitriadis and Bayer (2019) to estimate the nuisance quantity $\psi(u^q) = \text{Var}(u^q|u^q \leq 0, X)$, which are based on simplifications of the dependence of $u^q$ on $X$. Denote \{ $u^q_i = Y_i - X_i\beta^q, i = 1, \ldots, n$ \} as the estimated quantile residuals. The first approach assumes homogeneous errors such that the distribution of $u^q$ does not depend on $X$. Then, we can estimate $\psi(u^q)$ simply by the sample variance of the negative quantile residuals, and we refer to the corresponding estimator $\hat{\psi}$ as the iid estimator. The second approach assumes a location-scale dependence structure:

$$u^q = X\alpha + (X\nu)\epsilon,$$

where $\alpha$ and $\nu$ are $p$-dimensional parameter vectors and $\epsilon \sim F_\tau(0,1)$ with zero-mean and unit variance. The conditional variance $\psi(u^q)$ can be estimated by quasi generalized pseudo maximum likelihood (Gourieroux et al., 1984) based on the scaling formula

$$\text{Var}(u^q|u^q \leq 0, X) = \int_{-\infty}^{0} z^2 h(z|X) dz - \left( \int_{-\infty}^{0} z h(z|X) dz \right)^2,$$

where $h(z|X)$ is the truncated conditional density of $u^q$ given $u^q \leq 0$ and $X$. In our numerical implementation, we first obtain initial estimations of the conditional mean and variance of $u^q$ given $X$ (i.e., $\alpha$ and $\nu$) by maximum likelihood estimator (MLE) under a working normality assumption and then employ kernel density estimation to estimate the unknown distribution $F_\tau$ nonparametrically. The resulting estimator of $\psi$ is referred to as the iid estimator (denoted by $\hat{\psi}_n$).

Define

$$\hat{\Omega}(\hat{\psi}) = n^{-1} \sum_{i=1}^{n} (X_i\hat{\psi}) \left\{ G_2^{(1)}(X_i\hat{\theta}) \right\}^2 \left\{ \frac{1}{\tau} \psi(u_i^q) + \frac{1-\tau}{\tau} \phi^2(X_i\hat{\theta}, \hat{\theta}) \right\},$$

and denote $\hat{\Omega}(\hat{\psi}_1)$ and $\hat{\Omega}(\hat{\psi}_n)$ as the estimated covariance matrices with $\psi(u^q)$ estimated by $\hat{\psi}_1$ and $\hat{\psi}_n$, respectively. The following theorem establishes the consistency of $\hat{\Lambda}$ and $\hat{\Omega}(\hat{\psi}_1)$ and $\hat{\Omega}(\hat{\psi}_n)$.

**Theorem 2.** Under assumptions of Theorem 1, $\hat{\Lambda} - \Lambda = o_p(1)$, and if model (7) holds, we have $\hat{\Omega}(\hat{\psi}_n) - \Omega = o_p(1)$. Furthermore, if $u^q$ is independent with $X$, then $\Omega(\hat{\psi}_1) - \Omega = o_p(1)$.

**2.4 Proposed score test**

Theorem 2 shows that the covariance matrix of $\hat{\theta}^\phi$ can be estimated consistently. However, in the quantile regression literature, it has been shown that Wald-type test based on direct estimation of the asymptotic covariance matrix is often unstable for small sample sizes. One reason is due to the sensitivity of Wald-type test to the smoothing parameter involved in estimating the unknown conditional density function. The score test has been shown to have more stable performance than Wald test for quantile regression in finite samples (Chen & Wei, 2005; Kocherginsky et al., 2005). Due to the connection between quantile and ES, we propose an alternative score-type test for the inference on CES parameters.

We partition $\theta^\phi$ into two parts $\theta^\phi_1 \in \mathbb{R}^{p_1}$ and $\theta^\phi_2 \in \mathbb{R}^{p_2}$ with $p_1 + p_2 = p$, and let $W$ and $Z$ be the design vectors corresponding to $\theta^\phi_1$ and $\theta^\phi_2$, respectively. Suppose we want to test the hypotheses $H_0 : \theta^\phi_2 = 0$ against $H_1 : \theta^\phi_2 \neq 0$ in the joint-regression model

$$Q_i(Y|X) = X\theta^\phi_1 + W\theta^\phi_2 + Z\theta^\phi_1.$$

Denote

$$\Pi_W = (W_1, \ldots, W_n), \quad \Pi_Z = (Z_1, \ldots, Z_n),$$

$$G = \text{diag} \left\{ G_2^{(1)}(X_i\hat{\theta}^\phi), \ldots, G_2^{(1)}(X_i\hat{\theta}^\phi) \right\},$$

$$P = \Pi_W G \Pi_W^{-1} \Pi_W G, \quad \Pi_Z = (I - P)\Pi_Z,$$
and let $Z_i^*$ be the rows of $\Pi Z$ corresponding to the $i$th subject. We consider the orthogonal transformation on $Z$ to adjust for the dependence of $Z$ and $W$. This transformation is needed to cancel out the first-order bias involved in $\hat{\theta}_l$ to prove Lemma 2 in the Supporting Information. The weighted projection in the orthogonal transformation through $G$ is needed to account for the heteroscedasticity; see some related discussion under quantile regression in Koenker and Machado (1999). Our proposed score test statistic is defined as

$$T_n(\hat{\psi}) = S_n(\hat{\psi})^{-1} S_n,$$

where

$$S_n = n^{-1/2} \sum_{i=1}^n Z_i^* \{ X_i(\theta^0 - \hat{\theta})^T - \tau^{-1} u_i^0 (\hat{\theta}^0 \leq 0) \},$$

$$u_i^0 = Y_i - X_i \hat{\theta}^0,$$

$$\Sigma_n(\hat{\psi}) = n^{-1} \sum_{i=1}^n \{ Z_i^* Z_i^* \} \{ G_i^2(X_i \theta^0) \}^2 \left\{ \frac{1}{\tau} \psi(u_i^0) + \frac{1-\tau}{\tau} \phi^2(X_i, \theta^0, \hat{\theta}) \right\}.$$

Here, $\hat{\theta}_l$ is the two-step estimator of the CES regression parameter $\theta^0$ under $H_0$, and $\hat{\theta}$ is the two-step estimator of $\theta^0 = (\theta_0^0, \theta_2^0)^T$ under the unrestricted model. To obtain $Z_i^*$, the weight matrix $G$ given in (8) can be estimated with $\hat{\theta}$.

Before presenting the asymptotic distribution of $T_n$, we define

$$\Sigma = E \left[ \left\{ Z^* Z^* \right\} \left\{ G^2(X \theta_0^0) \right\}^2 \left\{ \frac{1}{\tau} \psi(u^0) + \frac{1-\tau}{\tau} \phi^2(X, \theta_0^0, \theta_0) \right\} \right].$$

We also introduce an additional assumption A5.

**A5.** The minimum eigenvalue of $\Sigma$ is bounded away from zero.

**Theorem 3.** Suppose that assumptions in Theorem 1 and A5 hold, we have the following:

(i) Under $H_0$, $T_n(\hat{\psi}_n) \overset{d}{\longrightarrow} \chi^2_2$ as $n \to \infty$. Furthermore, if $u^0$ is independent with $X$, $T_n(\hat{\psi}_l) \overset{d}{\longrightarrow} \chi^2_2$ as $n \to \infty$.

(ii) Under the local alternative hypothesis $H_1: \theta_2^0 = \theta_2^0/\sqrt{n}$ with $\theta_2^0$ is some non-zero parameter vector corresponding to $Z$, $T_n(\hat{\psi}_n)$ asymptotically follows a non-central $\chi^2_2$ distribution with the noncentrality parameter

$$\zeta = E \left\{ Z^* G^2(X \theta_0^0) (Z \theta_2^0) \right\} \Sigma^{-1} E \left\{ Z^* G^2(X \theta_0^0) (Z \theta_2^0) \right\}.$$

Furthermore, if $u^0$ is independent with $X$, then $T_n(\hat{\psi}_l) \overset{d}{\longrightarrow} \chi^2_2$ with the noncentrality parameter $\zeta$.

**Remark 3.** The matrix $\Sigma_n(\hat{\psi})$ in the score test statistic involves the estimation of the truncated conditional variance of $u^0$ given $u^0 \leq 0$ and $X$. Similar to the Wald-type test, we consider both iid and iid estimators for the nuisance parameter $\psi$ to accommodate different scenarios.

**Remark 4.** The proposed inference methods can be directly applied for analysing the upper tail CES at the probability level $\tau$, which is defined as

$$ES_u(Y|X) = (1-\tau)^{-1} \int_{\tau}^1 F_{Y\mid X}(u)du = (1-\tau)^{-1} \int_{\tau}^1 \Phi_u(Y|X)du.$$  

Assuming the upper tail ES regression model: $ES_u(Y|X) = X \theta^*$. That is, $\theta^*$ and $\theta^0$ are the upper and lower tail ES regression parameters, respectively. Note that $F_{Y \mid X}(u) = -F_{Y \mid X}(1-u)$. In practice, if we are interested in the inference on $\theta^* \ast$, we can (1) change $Y$ to $Y^* = -Y$ and let $\tau^* = 1-\tau$ and (2) apply the proposed two-step estimation and inference methods on the lower tail ES of $Y^*$ conditional on $X$ with probability level $\tau^*$. Then, we have $\hat{\theta}^* = -\hat{\theta}$ and $\text{Var}(\hat{\theta}^*) = \text{Var}(\hat{\theta})$. 


3 | SIMULATION STUDY

In this section, we investigate the finite sample performance of the proposed inference method through Monte Carlo simulation studies. For comparison, we include the results from the Wald-type method in Section 2.3 and the bootstrap method, which are based on the joint estimation implemented in the R package esreg (Dimitriadis & Bayer, 2019). For the bootstrap method, we first generate $B = 1000$ bootstrap samples by randomly selecting the $n$ pairs of $(Y_i, X_i)$ with replacement and then obtain the bootstrap variance by the sample covariance of the CES coefficient estimators obtained by analysing the bootstrap samples. For both Wald and score methods, we consider $W$-IID and $S$-IID approaches, where the asymptotic variance is estimated under the homogeneous error assumption; and $W$-NID and $S$-NID method, where the conditional variance $ψ(u^2)$ is estimated by $ψ_k$. In addition, we also report the results of the COVES test introduced by He et al. (2010). Since the COVES method focuses on the treatment difference at the right tail of the response distribution, for the simulation study in Section 3 and the real data analysis in Section 4, we will focus on the inference for the upper tail CES at the probability level $r$, as defined in (10).

3.1 | Simulation design

The first two models we consider have simple set-ups:

- **Scenario 1**: $Y = 5 + ηD + x_1 + ε$;
- **Scenario 2**: $Y = 5 − ηD + C + (1 + 0.25D + 2C)ε$.

0.52) (truncated normal distribution) in the treatment group and $C ∼ TN(\min = −0.5, μ = 0, σ^2 = 0.5^2)$ in control group.

Scenario 1 is a homogeneous model where the regression error does not interact with any covariates. For Scenario 2, the error depends on both the treatment variable and the unbalanced covariate $C$. At the probability level $r$, the marginal impact on the CES due to treatment effect is

$$η(τ) = −η + 0.25ES_τ(ε),$$

which is the ES coefficient associated with the treatment variable $D$. In contrast, the expectation of the COVES test statistic is given by

$$T_τ = η(τ) + 2E(C|D = 1) − E(C|D = 0) \{E(ε) − Q_τ(ε)\}. \tag{11}$$

where $Q_τ(ε)$ and $ES_τ(ε)$ are the marginal $τ$th quantile and upper tail ES of $ε$. Due to the imbalance of covariate $C$, COVES may inflate or deflate the treatment difference $η(τ)$, which consequently makes the test either too liberal or too conservative.

**Remark 5.** The difference between $T_τ$ and the treatment difference $η(τ)$ is

$$T_τ − η(τ) = 2E(C|D = 1) − E(C|D = 0) \{E(ε) − Q_τ(ε)\}. \tag{11}$$

Suppose that $C$ is an unbalanced covariate such that the mean of $C$ differs for the two treatment groups and the error $ε$ depends on $C$. Then, it is possible to have $T_τ − η(τ) ≠ 0$, and this may affect the power of the COVES test. Specifically, if the second term on the right-hand side (RHS) of (11) cancels out with $η(τ)$, COVES may fail to detect the treatment difference. On the other hand, if the treatment has no impact on CES such that $η(τ) = 0$, but the second term on the RHS of (11) is non-zero, then COVES may over-reject the null hypothesis and thus lead to higher false positive rate.

There’s only one confounding variable in the first two models, and both error terms follow normal distributions. To examine the robustness of the proposed method, we further consider another two scenarios where more regression covariates are included (Scenario 3) and the error has a heavy-tailed distribution (Scenario 4). The data are generated from

$$Y = 5 + ηD + \sum_{i=2}^{7} x_i + (1 + γD)ε,$$

where $x_2$ is Ber(0.4); $x_3$ and $x_4$ have standard log-normal distribution; $(x_5, x_6)$ is bivariate normal with mean (2,2), variance (1,1) and correlation 0.8; and $x_7$ is chi-square distributed with one degree of freedom. Except for $x_5$ and $x_6$, all other covariates are independently generated.
- **Scenario 3**: We take $\gamma = 0$, and the error term $\epsilon \sim N(0,1)$.
- **Scenario 4**: We take $\gamma = 0.2$, and the error term $\epsilon \sim t_3/2$.

We consider two sample sizes $n = 50$ and $n = 100$ for each treatment group, and we focus on $\tau = 0.8$ and $\tau = 0.9$ in this study. The simulation is repeated 600 times for each scenario with a given value of $\eta$.

### 3.2 Statistical power for testing the treatment effect

Table 1 shows that Wald-type and score-type approaches both maintain the significance level reasonably well and the corresponding type I errors stay close to the nominal level of 0.05. However, the bootstrap method and the COVES test yield inflated false positive rates for most cases, especially when sample size $n = 50$ and $\tau = 0.9$. Under Scenario 1, with only one covariate besides the treatment indicator, all methods perform quite similarly. However, as we add more confounding factors, the score-type testing methods show higher statistical power than the Wald-type approaches; see Figure 1 for some typical examples in Scenarios 3 and 4 at $\tau = 0.8$ and $n = 100$. Power curves in Scenario 2 confirm that COVES test gives biased estimation of the treatment difference due to the unbalanced covariate.

### 3.3 Confidence intervals and point estimation for the treatment coefficient

To access the performance of different methods for confidence interval construction, we fix $\eta = 1.35, 2.25$ and 3.5 in Scenarios 1–4, respectively. Tables 2 and 3 summarize the coverage percentages and average lengths of 95% confidence intervals for the treatment difference $\eta(\tau)$. Under Scenario 1, the Wald-type and score methods show similar accuracy. For Scenarios 3 and 4 with more confounding factors, the score-type methods provide shorter confidence intervals with relatively higher coverage, which agrees with the results of the power analysis in Section 3.2. Furthermore, when errors are i.i.d, W-IID and W-NID approaches perform similarly. For Scenarios 2 and 4 when the errors are heterogeneous, W-NID method shows better performance in the sense that it provides confidence intervals with coverage closer to the nominal level and shorter length than the W-IID method. On the other hand, the score methods are less sensitive to the violation of homogeneity assumption, as S-IID and S-NID give similar results across all scenarios considered. Overall, the COVES method gives lowest coverage percentage, much below the nominal level 95%, especially under Scenario 2 where the error term depends on an unbalanced covariate.

### Table 1: Type I error (percentage) for testing $H_0: \eta(\tau) = 0$ with nominal level of 5%.

| Scenario | $n$ | $\tau$ | W-IID | W-NID | S-IID | S-NID | BOOT | COVES |
|----------|-----|--------|-------|-------|-------|-------|------|-------|
| 1        | 50  | 0.8    | 7.3   | 6.7   | 6.2   | 6.3   | 6.7  | 8.5   |
|          |     | 0.9    | 9.3   | 8.5   | 9.7   | 8.0   | 8.0  | 13.2  |
|          | 100 | 0.8    | 6.5   | 6.0   | 6.2   | 6.2   | 6.3  | 6.7   |
|          |     | 0.9    | 6.7   | 5.5   | 6.2   | 5.5   | 6.7  | 7.7   |
| 2        | 50  | 0.8    | 2.3   | 3.8   | 2.7   | 4.2   | 6.8  | 13.0  |
|          |     | 0.9    | 4.7   | 6.2   | 5.8   | 7.0   | 7.0  | 12.5  |
|          | 100 | 0.8    | 1.2   | 2.5   | 1.7   | 3.0   | 5.5  | 18.3  |
|          |     | 0.9    | 2.0   | 3.0   | 2.3   | 3.2   | 5.3  | 11.8  |
| 3        | 50  | 0.8    | 5.7   | 6.0   | 4.5   | 4.8   | 8.2  | 12.5  |
|          |     | 0.9    | 8.2   | 8.2   | 8.5   | 8.2   | 10.0 | 21.8  |
|          | 100 | 0.8    | 3.5   | 4.3   | 4.0   | 4.3   | 5.2  | 6.7   |
|          |     | 0.9    | 4.0   | 4.0   | 5.2   | 4.8   | 6.2  | 9.8   |
| 4        | 50  | 0.8    | 2.5   | 3.0   | 2.0   | 3.0   | 6.0  | 9.5   |
|          |     | 0.9    | 3.0   | 3.5   | 4.7   | 5.0   | 11.0 | 17.8  |
|          | 100 | 0.8    | 3.2   | 3.7   | 1.8   | 2.7   | 7.7  | 5.5   |
|          |     | 0.9    | 2.7   | 3.3   | 3.2   | 4.0   | 9.5  | 10.7  |

Abbreviations: BOOT, bootstrap method based on the joint estimation; COVES, method in He et al. (2010); S-IID (S-NID), score methods with $\psi(u^\tau)$ estimated by $\hat{\psi}_I(\hat{u}^\tau)$; W-IID (W-NID), Wald-type methods with $\psi(u^\tau)$ estimated by $\hat{\psi}_N(\hat{u}^\tau)$.
We also compared the point estimation of the joint and two-step estimators on treatment effect \( \eta(\tau) \), denoted by \( \sim \eta(\tau) \) and \( \hat{\eta}(\tau) \), respectively. Results under Scenarios 3 and 4 are summarized in Table 4. Overall, the two-step estimator shows superiority over the joint estimator in terms of both average bias and MSE, especially when sample size is small \( (n = 50) \).

4 | REAL DATA ANALYSES

4.1 | Application to 2018 Current Population Survey (CPS) income data

We illustrate the merit of the proposed score-type inference approach by analysing a data set from the CPS database, which can be assessed at https://www.census.gov/programs-surveys/cps.html. The CPS is a monthly survey of about 60,000 US households conducted by the United States Census Bureau for the Bureau of Labor Statistics. Information collected in the survey includes employment status, income from work and a number of demographic characteristics. From the 2018 CPS Annual Social and Economic Supplement (ASEC) Bridge Files, we compile...
a data set, which contains 870 individuals (401 male and 469 female). To find out if there exists a pay gap between women and men, we use hourly wage in US dollars as the response and consider Gender, Age, Age$^2$, Education level (ordinal variable with three levels) and Work status (full-time vs. part-time) as explanatory variables in the joint-regression model.

| Scenario | $\tau$ | W-IID | W-NID | S-IID | S-NID | BOOT | COVES |
|----------|--------|-------|-------|-------|-------|-------|-------|
| 1        | 0.8    | 92.2  | 92.8  | 94.0  | 94.0  | 93.0  | 91.8  |
|          |        | (120) | (122) | (120) | (120) | (124) | (113) |
|          | 0.9    | 90.5  | 91.5  | 90.7  | 90.7  | 91.3  | 88.3  |
|          |        | (146) | (149) | (146) | (146) | (144) | (133) |
| 2        | 0.8    | 98.0  | 96.0  | 97.3  | 97.3  | 93.5  | 85.8  |
|          |        | (268) | (240) | (265) | (265) | (203) | (232) |
|          | 0.9    | 95.5  | 94.5  | 94.2  | 94.2  | 93.3  | 84.7  |
|          |        | (327) | (297) | (320) | (320) | (253) | (268) |
| 3        | 0.8    | 92.7  | 92.3  | 95.5  | 95.5  | 94.7  | 87.7  |
|          |        | (184) | (182) | (157) | (157) | (187) | (110) |
|          | 0.9    | 90.8  | 91.5  | 91.5  | 91.5  | 92.5  | 77.3  |
|          |        | (210) | (210) | (175) | (175) | (192) | (119) |
| 4        | 0.8    | 97.0  | 96.3  | 98.0  | 98.0  | 96.5  | 90.7  |
|          |        | (234) | (227) | (200) | (200) | (197) | (143) |
|          | 0.9    | 96.3  | 96.2  | 95.5  | 95.5  | 94.3  | 83.3  |
|          |        | (343) | (333) | (284) | (284) | (238) | (201) |

Note: All values are in percentages.

Abbreviations: BOOT, bootstrap method based on the joint estimation; COVES, method in He et al. (2010); S-IID (S-NID), score methods with $\psi(u_q)$ estimated by $\hat{\psi}_I(\hat{\psi}_N)$; W-IID (W-NID), Wald-type methods with $\psi(u_q)$ estimated by $\hat{\psi}_I(\hat{\psi}_N)$.

| Scenario | $\tau$ | W-IID | W-NID | S-IID | S-NID | BOOT | COVES |
|----------|--------|-------|-------|-------|-------|-------|-------|
| 1        | 0.8    | 93.3  | 93.8  | 93.8  | 93.8  | 93.7  | 92.3  |
|          |        | (87.1)| (88.4)| (87.2)| (87.2)| (87.8)| (83.8)|
|          | 0.9    | 93    | 93.2  | 93.5  | 93.5  | 92.5  | 91.5  |
|          |        | (109) | (110) | (108) | (108) | (107) | (103) |
| 2        | 0.8    | 98.7  | 97.7  | 98.3  | 98.3  | 94.7  | 82.7  |
|          |        | (194) | (173) | (192) | (192) | (143) | (174) |
|          | 0.9    | 98.2  | 97.5  | 97.7  | 97.7  | 94.7  | 86.8  |
|          |        | (242) | (215) | (238) | (238) | (178) | (213) |
| 3        | 0.8    | 94.7  | 94.5  | 96.2  | 96.2  | 95.7  | 93.7  |
|          |        | (129) | (129) | (108) | (108) | (128) | (81.4)|
|          | 0.9    | 94.8  | 94.0  | 95.2  | 95.2  | 95.0  | 89.5  |
|          |        | (150) | (149) | (130) | (130) | (137) | (96.3)|
| 4        | 0.8    | 96.2  | 95.7  | 98.2  | 98.2  | 96.8  | 94.3  |
|          |        | (170) | (161) | (152) | (152) | (142) | (113) |
|          | 0.9    | 96.2  | 95.3  | 96.8  | 96.8  | 93.7  | 89.3  |
|          |        | (260) | (247) | (233) | (233) | (188) | (176)|

Note: All values are in percentages.

Abbreviations: BOOT, bootstrap method based on the joint estimation; COVES, method in He et al. (2010); S-IID (S-NID), score methods with $\psi(u_q)$ estimated by $\hat{\psi}_I(\hat{\psi}_N)$; W-IID (W-NID), Wald-type methods with $\psi(u_q)$ estimated by $\hat{\psi}_I(\hat{\psi}_N)$.
To check whether the quantile error depends on the predictors, we conduct a heterogeneity analysis by analysing the residual patterns. At a given quantile level $\tau$, we fit a linear quantile regression model and compare the variance of the quantile residuals for different covariate groups. Table 5 summarizes quantile residual variances associated with different education levels at $\tau = 0.7, 0.75$ and $0.8$. The result shows that the quantile residuals appear to depend on the Education level. Therefore, for Wald and score inference approaches, we focus on $W$-NID and $S$-NID methods. Moreover, the Education level is an unbalanced covariate since the female group has a higher proportion of subjects with Education Level 3.

Let $\theta_{eg}$ denote the upper tail CES difference of the hourly wage between female and male groups. At the probability levels 0.7, 0.75 and 0.8, we apply the proposed inference method to test $H_0: \theta_{eg} = 0$ against $H_1: \theta_{eg} \neq 0$. Except for the COVES method, all the other approaches suggest a significant pay gap between the two gender groups. We further calculate 95% confidence intervals of $\theta_{eg}$ using different methods and the results are summarized in Table 6. Under the significance level of 5%, results of $W$-NID, $S$-NID and the bootstrap approaches suggest that the female employee is substantially under-paid compared to the male employee, when other characteristics are kept the same. In contrast, the COVES test may be negatively affected by the unbalanced covariate Education level and thus fails to capture the tail difference.

### 4.2 Power analysis for the Opportunity Knocks (OK) data

To further assess the finite sample performance of the proposed score test, we conduct a power analysis based on the OK experiment (Angrist et al., 2014), which was designed to explore the effects of academic achievement awards for first-year and second-year college students. For our analysis, we consider a subset with all second-year students, which consists of 183 treated subjects and 337 untreated subjects. Treated students can receive bonus awards and have the opportunity to interact with randomly assigned peer advisors who can provide advice about study strategies, time management and university bureaucracy.
The award scheme offered cash incentives to students with course grades above 70. Therefore, the academic performance of students can be measured by the amount they earned in the OK experiment. To determine how the academic performance of students are motivated by the merit award, we define the response variable as the earning of students (in 1000 US dollars) from the OK programme. Besides the treatment indicator \( D \), we consider six additional covariates, including gender \( (X_1) \), high school grade \( (X_2) \), an indicator for English mother tongue \( (X_3) \), whether the student answers the scholarship formula question correctly \( (X_4) \), yes vs. no and mother’s and father’s education levels \( (X_5 \) and \( X_6) \), defined as above college degree or not. We apply the proposed inference method to test the treatment effect \( \delta_D^75 \) (upper tail ES regression coefficient associated with treatment variable \( D \)) at \( \tau = 0.70 \) and 0.8, and the results are summarized in Table 7. The data indicate that there’s a tendency that the merit award has a positive effect on the academic performance. However, none of the methods show that the treatment is statistically significant on the CES based on the original data, which might be due to the small sample size. To determine the sample size needed for different methods to capture the treatment difference, we conduct a power analysis.

In order to mimic the response distribution based on the linearity model assumption, we fit a linear quantile regression model using the original data at \( \tau = 0.75 \),

\[
\hat{Q}_i(\tau | D, X_1, ..., X_6) = \hat{\beta}_0(\tau) + \hat{\beta}_1(\tau)D + \sum_{j=1}^6 \hat{\alpha}_j(\tau)X_j .
\]

We then obtain quantile residuals, defined as \( \hat{e}_i(\tau) = Y_i - \hat{Q}_i(\tau | D, X_1, ..., X_6) \), and perform a heterogeneity analysis similar to the one in Section 4.1. The results indicate that the residual term depends on \( X_4 \). Therefore, for power analysis, we focus on S-NID and W-NID approaches, and the response is generated by

\[
Y_{\text{sim}} = \hat{\beta}_0(\tau) + \hat{\beta}_1(\tau)D + \sum_{j=1}^6 \hat{\alpha}_j(\tau)X_j^d + \xi(D = 1) + \epsilon,
\]

where \( \hat{\beta}_0(\tau), \hat{\beta}_1(\tau) \) and \( \hat{\alpha}_j(\tau) \) are regression coefficient estimators in (12); \( X_j^d \) follows the empirical distribution of covariate \( X_j \) in group \( D = d \); and \( \epsilon \) is randomly sampled from the quantile residuals \( \{\hat{e}_i(\tau)\} \) stratified by different grouped values of \( X_4 \). Since the estimated treatment effect from the observed sample is not statistically significant, we add an additional signal \( \xi \) in (13) to increase the treatment difference, and let the sample sizes of the two treatment groups be the same. We apply the proposed S-NID method to the synthetic data. Table 8 summarizes the sample size needed for different methods to reach a power of 0.9 at \( \tau = 0.75 \). The results show that the score test is clearly outperforming the Wald test and the bootstrap method, and the latter two require a trial with more subjects.

**Table 7** Point estimation and the 95% confidence intervals (within the parentheses) of \( \delta_D^75 \) given by different approaches based on the original OK data, where \( \delta_D^75 \) is the ES regression coefficient associated with treatment variable \( D \) and measures the treatment effect of the merit award on the upper tail CES.

| \( \tau \)  | S-NID       | W-NID       | BOOT       |
|------------|-------------|-------------|------------|
| 0.70       | 0.298       | 0.375       | 0.375      |
|            | (-0.086,0.651) | (-0.001,0.751) | (-0.327,1.078) |
| 0.75       | 0.286       | 0.304       | 0.304      |
|            | (-0.126,0.678) | (-0.116,0.724) | (-0.379,0.987) |
| 0.80       | 0.278       | 0.204       | 0.204      |
|            | (-0.190,0.726) | (-0.229,0.637) | (-0.469,0.878) |

**Table 8** Sample size needed for each treatment group in the synthetic OK data to reach power 0.9 at \( \tau = 0.75 \).

| \( \xi \)  | S-NID | W-NID | BOOT |
|------------|-------|-------|------|
| 0.2        | 436   | 484   | 494  |
| 0.3        | 282   | 292   | 314  |
| 0.4        | 210   | 215   | 218  |
| 0.5        | 134   | 155   | 168  |
| 0.6        | 107   | 120   | 129  |
| 0.7        | 82    | 94    | 101  |
5 | CONCLUSION

In this paper, we considered the joint modelling of conditional quantile and ES. A two-step estimation procedure is proposed to reduce the computational effort. We showed that the resulting two-step estimator is asymptotically equivalent to the joint estimator, but the former is numerically more efficient. In addition, the two-step estimator is locally robust to the perturbation of the quantile estimation in the first step. We further developed a score-type inference method for hypothesis testing and confidence interval construction. The proposed score method is robust in performance, especially for cases with a large number of confounding factors and heterogeneous errors.

We chose parametric linear models for the joint-regression framework due to its computational efficiency and model interpretability. This framework can be further extended by considering more general models. Wang et al. (2018), Taylor (2019) and Patton et al. (2019) consider dynamic models for ES with autoregressive features. To model CES based on exogenous covariates, another feasible alternative is to employ some nonparametric or semiparametric models, for example, varying coefficient models (T. Hastie & Tibshirani, 1993) and generalized additive models (T. J. Hastie & Tibshirani, 1990). The proposed two-step estimation procedure and inference methods can be adapted accordingly, but further theoretical and practical investigations are needed.

AUTHOR CONTRIBUTIONS

Both authors conceived of the presented idea and contributed to the final manuscript. Dr. Xiang Peng performed the numerical simulations and conducted the real data analysis, and the project is under the supervision of Dr. Huixia Judy Wang.

ACKNOWLEDGEMENTS

This work was partly supported by the IR/D programme and Grant DMS-1712760 from the National Science Foundation. Any opinion, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available in Current Population Survey (CPS) database at https://www.census.gov/programs-surveys/cps.html.

ORCID

Xiang Peng https://orcid.org/0009-0007-0534-2121
Huixia Judy Wang https://orcid.org/0000-0002-5195-8564

REFERENCES

Angrist, J., Oreopoulos, P., & Williams, T. (2014). When opportunity knocks, who answers? New evidence on college achievement awards. Journal of Human Resources, 49(3), 572–610.
Artzner, P. (1997). Thinking coherently. Risk, 10, 68–71.
Artzner, P., Delbaen, F., Eber, J.-M., & Heath, D. (1999). Coherent measures of risk. Mathematical Finance, 9(3), 203–228.
Barendse, S. (2020). Efficiently weighted estimation of tail and interquantile expectations. Tinbergen Institute Discussion Paper 2017-034/III, Available at SSRN: https://ssrn.com/abstract=2937665 or https://doi.org/10.2139/ssrn.2937665
Basel Committee. (2013). Basel Committee on Banking Supervision. Consultative Document. Fundamental review of the trading book: A revised market risk framework. Available at: https://www.bis.org/publ/bcbs265.pdf
Cai, Z., & Wang, X. (2008). Nonparametric estimation of conditional VaR and expected shortfall. Journal of Econometrics, 147(1), 120–130.
Chen, T., & Wei, Y. (2005). Computational issues for quantile regression. Sankhyā: The Indian Journal of Statistics, 67(2), 399–417.
Chernozhukov, V., Escanciano, J. C., Ichimura, H., Newey, W. K., & Robins, J. M. (2016). Locally robust semiparametric estimation. arXiv preprint arXiv: 1608.00033.
Dimitriadis, T., & Bayer, S. (2019). A joint quantile and expected shortfall regression framework. Electronic Journal of Statistics, 13(1), 1823–1871.
Fissler, T., & Ziegel, J. F. (2016). Higher order elicitation and Osband’s principle. The Annals of Statistics, 44(4), 1680–1707.
Foster, D. J., & Syrgkanis, V. (2019). Orthogonal statistical learning. arXiv. Retrieved from https://arxiv.org/abs/1901.09036
Gneiting, T. (2011). Making and evaluating point forecasts. Journal of the American Statistical Association, 106(494), 746–762.
Gourieroux, C., Monfort, A., & Trognon, A. (1984). Pseudo maximum likelihood methods: Theory. Econometrica: Journal of the Econometric Society, 52(3), 681–700.
Hastie, T., & Tibshirani, R. (1993). Varying-coefficient models. Journal of the Royal Statistical Society: Series B (Methodological), 55(4), 757–779.
Hastie, T. J., & Tibshirani, R. J. (1990). Generalized Additive Models, Vol. 43: CRC Press.
He, X., Hsu, Y.-H., & Hu, M. (2010). Detection of treatment effects by covariate-adjusted expected shortfall. The Annals of Applied Statistics, 4(4), 2114–2125.
Kato, K. (2012). Weighted Nadaraya–Watson estimation of conditional expected shortfall. Journal of Financial Econometrics, 10(2), 265–291.
Kocherginsky, M., He, X., & Mu, Y. (2005). Practical confidence intervals for regression quantiles. Journal of Computational and Graphical Statistics, 14(1), 41–55.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available in Current Population Survey (CPS) database at https://www.census.gov/programs-surveys/cps.html.

ORCID

Xiang Peng https://orcid.org/0009-0007-0534-2121
Huixia Judy Wang https://orcid.org/0000-0002-5195-8564

REFERENCES

Angrist, J., Oreopoulos, P., & Williams, T. (2014). When opportunity knocks, who answers? New evidence on college achievement awards. Journal of Human Resources, 49(3), 572–610.
Artzner, P. (1997). Thinking coherently. Risk, 10, 68–71.
Artzner, P., Delbaen, F., Eber, J.-M., & Heath, D. (1999). Coherent measures of risk. Mathematical Finance, 9(3), 203–228.
Barendse, S. (2020). Efficiently weighted estimation of tail and interquantile expectations. Tinbergen Institute Discussion Paper 2017-034/III, Available at SSRN: https://ssrn.com/abstract=2937665 or https://doi.org/10.2139/ssrn.2937665
Basel Committee. (2013). Basel Committee on Banking Supervision. Consultative Document. Fundamental review of the trading book: A revised market risk framework. Available at: https://www.bis.org/publ/bcbs265.pdf
Cai, Z., & Wang, X. (2008). Nonparametric estimation of conditional VaR and expected shortfall. Journal of Econometrics, 147(1), 120–130.
Chen, T., & Wei, Y. (2005). Computational issues for quantile regression. Sankhyā: The Indian Journal of Statistics, 67(2), 399–417.
Chernozhukov, V., Escanciano, J. C., Ichimura, H., Newey, W. K., & Robins, J. M. (2016). Locally robust semiparametric estimation. arXiv preprint arXiv: 1608.00033.
Dimitriadis, T., & Bayer, S. (2019). A joint quantile and expected shortfall regression framework. Electronic Journal of Statistics, 13(1), 1823–1871.
Fissler, T., & Ziegel, J. F. (2016). Higher order elicitation and Osband’s principle. The Annals of Statistics, 44(4), 1680–1707.
Foster, D. J., & Syrgkanis, V. (2019). Orthogonal statistical learning. arXiv. Retrieved from https://arxiv.org/abs/1901.09036
Gneiting, T. (2011). Making and evaluating point forecasts. Journal of the American Statistical Association, 106(494), 746–762.
Gourieroux, C., Monfort, A., & Trognon, A. (1984). Pseudo maximum likelihood methods: Theory. Econometrica: Journal of the Econometric Society, 52(3), 681–700.
Hastie, T., & Tibshirani, R. (1993). Varying-coefficient models. Journal of the Royal Statistical Society: Series B (Methodological), 55(4), 757–779.
Hastie, T. J., & Tibshirani, R. J. (1990). Generalized Additive Models, Vol. 43: CRC Press.
He, X., Hsu, Y.-H., & Hu, M. (2010). Detection of treatment effects by covariate-adjusted expected shortfall. The Annals of Applied Statistics, 4(4), 2114–2125.
Kato, K. (2012). Weighted Nadaraya–Watson estimation of conditional expected shortfall. Journal of Financial Econometrics, 10(2), 265–291.
Kocherginsky, M., He, X., & Mu, Y. (2005). Practical confidence intervals for regression quantiles. Journal of Computational and Graphical Statistics, 14(1), 41–55.
Koenker, R. (2005). *Quantile Regression*. Cambridge University Press.

Koenker, R., & Machado, J. A. F. (1999). Goodness of fit and related inference processes for quantile regression. *Journal of the American Statistical Association*, 94(448), 1296–1310.

Leorato, S., Peracchi, F., & Tanase, A. V. (2012). Asymptotically efficient estimation of the conditional expected shortfall. *Computational Statistics & Data Analysis*, 56(4), 768–784.

McNeil, A. J., Frey, R., & Embrechts, P. (2015). *Quantitative Risk Management: Concepts, Techniques and Tools-revised edition*. Princeton University Press.

Neyman, J. (1959). Optimal asymptotic tests of composite statistical hypotheses. In Grenander, V. (Ed.), *Probability and Statistics*: John Wiley & Sons, pp. 13–34.

Neyman, J. (1979). C(α) tests and their use. *Sankhyā The Indian Journal of Statistics, Series A* (1961-2002), 41(1/2), 1–21. Retrieved 2022-09-27, from http://www.jstor.org/stable/25050174

Patton, A. J., Ziegel, J. F., & Chen, R. (2019). Dynamic semiparametric models for expected shortfall (and value-at-risk). *Journal of Econometrics*, 211(2), 388–413.

Peracchi, F., & Tanase, A. V. (2008). On estimating the conditional expected shortfall. *Applied Stochastic Models in Business and Industry*, 24(5), 471–493.

Taylor, J. W. (2019). Forecasting value at risk and expected shortfall using a semiparametric approach based on the asymmetric Laplace distribution. *Journal of Business & Economic Statistics*, 37(1), 121–133.

Wang, C., Gerlach, R., & Chen, Q. (2018). A semi-parametric realized joint value-at-risk and expected shortfall regression framework. arXiv preprint arXiv:1807.02422.

Xiao, Z. (2014). Right-tail information in financial markets. *Econometric Theory*, 30, 94–126.

**AUTHOR BIOGRAPHIES**

**Dr. Xiang Peng** received his PhD degree from the George Washington University; his dissertation is about subgroup analysis and expected shortfall regression. He is currently working as a biostatistician in Merck & Co.

**Dr. Huixia Judy Wang** received her PhD from the University of Illinois in 2006. She was a faculty member in the Department of Statistics at North Carolina State University from 2006 to 2014. Currently, she is a professor and chair in the Department of Statistics at The George Washington University.

**SUPPORTING INFORMATION**

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article**: Peng, X., & Judy Wang, H. (2023). Inference for joint quantile and expected shortfall regression. *Stat*, 12(1), e619. https://doi.org/10.1002/sta4.619