Submitted to the Annals of Statistics

INFERENCE FOR CONDITIONAL VALUE-AT-RISK OF A PREDICTIVE REGRESSION

By Yi He[§], Yanxi Hou[‡,¶], Liang Peng[*,‖] and Haipeng Shen[†,**]

University of Amsterdam[§], *Fudan University*[¶], *Georgia State University*[‖] *and The University of Hong Kong*[**]

Conditional value-at-risk is a popular risk measure in risk management. We study the inference problem of conditional value-at-risk under a linear predictive regression model. We derive the asymptotic distribution of the least squares estimator for the conditional value-at-risk. Our results relax the model assumptions made in Chun et al. (2012) and correct their mistake in the asymptotic variance expression. We show that the asymptotic variance depends on the quantile density function of the unobserved error and whether the model has a predictor with infinite variance, which makes it challenging to actually quantify the uncertainty of the conditional risk measure. To make the inference feasible, we then propose a smooth empirical likelihood based method for constructing a confidence interval for the conditional value-at-risk based on either independent errors or GARCH errors. Our approach not only bypasses the challenge of directly estimating the asymptotic variance but also does not need to know whether there exists an infinite variance predictor in the predictive model. Furthermore, we apply the same idea to the quantile regression method, which allows infinite variance predictors and generalizes the parameter estimation in Whang (2006) to conditional value-at-risk in the supplementary material. We demonstrate the finite sample performance of the derived confidence intervals through numerical studies before applying them to real data.

1. Introduction.

One important aspect of risk management is to infer a risk measure with accurate statistical uncertainty quantification. There have been many well-known risk measures in the literature, among which the value-at-risk, VaR, is arguably one of the most widely used in finance and insurance. VaR is especially popular for summarizing distributional of

[*]Research is partially supported by the Simons Foundation.

[†]Research is partially supported by the Ministry of Science and Technology Major Project of China 2017YFC1310903, University of Hong Kong Stanley Ho Alumni Challenge Fund, and HKU University Research Committee Seed Funding Award 104004215.

[‡]Corresponding author. Research is partially supported by National Natural Science Foundation of China 71803026 and 71991471.

MSC 2010 subject classifications: 62G20, 62M10, 62P20

Keywords and phrases: conditional risk measures, empirical likelihood, least squares estimation, interval estimation, quantile regression, risk management, value-at-risk

tails of economic risk variables; see, e.g., Duffie and Pan (1997, 2001); Jorion (2007); Kou et al. (2013); Kou and Peng (2016). When an asset or a financial variable can be predicted by some market variables or risk factors, a conditional risk measure given the market situation is more meaningful than an unconditional risk measure, as is the case for conditional VaR. We consider the inference issue of deriving the confidence interval for a conditional VaR, which can be used to quantify estimation risk for regulatory purposes; see, e.g., Jorion (1996); Christoffersen and Gonçalves (2005); Gao and Song (2008).

A conditional VaR is defined as a conditional quantile of a return of interest given the market situation. One can directly model the conditional quantile as a parametric form of the market variables; hence, quantile regression techniques can be employed to infer the conditional VaR, which gives the standard rate of convergence $n^{-1/2}$; see, e.g., Engle and Manganelli (2004) and Fan et al. (2016). In this paper, we take this parametric approach and follow Chun et al. (2012) to estimate conditional VaR via a predictive regression model. Alternatively, a model-free estimator of the conditional VaR can be constructed by kernel smoothing techniques, which unfortunately results in a slower rate of convergence than the standard rate and a complicated asymptotic variance that is not easy to be estimated directly; see, e.g., Xu (2016).

To fix ideas, let Y denote the return of an asset or a financial variable, and $\boldsymbol{X} = (X_1, \dots, X_k)^T$ be the collection of predictors (market variables or risk factors). Here A^T denotes the transpose of a matrix or a vector A. Suppose that we have n observations (data points) $\{Y_t, \boldsymbol{X}_t = (X_{t,1}, \dots, X_{t,k})^T\}_{t=1}^n$ from the following linear predictive regression:

$$(1.1) \qquad Y_t = \beta_0 + \sum_{i=1}^{k} \beta_i X_{t,i} + \epsilon_t, \quad t = 1, \dots, n,$$

where $\{\epsilon_t\}$ is a sequence of independent and identically distributed random variables with zero mean and finite variance. Later we will study the case that $\{\epsilon_t\}$ follows a stationary generalized autoregressive conditional heteroscedasticity (GARCH) process [Bollerslev (1986)].

Under the model (1.1), the conditional VaR of Y_t at level $\alpha \in (0,1)$, given $\boldsymbol{X}_t = \boldsymbol{x} := (x_1, \dots, x_k)^T$, is proposed as the conditional quantile[Chun et al. (2012)]:

$$\mathrm{VaR}_{\boldsymbol{x}}(\alpha) := \inf\{q : P(Y_t \le q | \boldsymbol{X}_t = \boldsymbol{x}) \ge \alpha\} = F_\epsilon^{-1}(\alpha) + \boldsymbol{z}^T \boldsymbol{\beta},$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_k)^T$, $\boldsymbol{z} = (1, \boldsymbol{x}^T)^T$, F_ϵ denotes the distribution function of ϵ_t and F_ϵ^{-1} denotes the generalized inverse of F_ϵ. It then follows that a simple estimator for the above conditional VaR is

$$\widehat{\mathrm{VaR}}_{\boldsymbol{x}}(\alpha) = \widehat{\epsilon}_{n,[n\alpha]} + \boldsymbol{z}^T \widehat{\boldsymbol{\beta}}, \tag{1.2}$$

where $\widehat{\boldsymbol{\beta}}$ is a consistent estimator of $\boldsymbol{\beta}$, $\widehat{\epsilon}_t = Y_t - \widehat{\boldsymbol{\beta}}^T \boldsymbol{Z}_t$ with $\boldsymbol{Z}_t = (1, \boldsymbol{X}_t^T)^T$, $\widehat{\epsilon}_{n,1} \le \ldots \le \widehat{\epsilon}_{n,n}$ denote the order statistics of $\{\widehat{\epsilon}_1, \ldots, \widehat{\epsilon}_n\}$, and $[n\alpha]$ is the nearest integer of $n\alpha$.

Under Model (1.1), Chun et al. (2012) recently considered the least squares estimator:

$$\widehat{\boldsymbol{\beta}} = \left\{ \frac{1}{n} \sum_{t=1}^{n} \boldsymbol{Z}_t \boldsymbol{Z}_t^T \right\}^{-1} \frac{1}{n} \sum_{t=1}^{n} Y_t \boldsymbol{Z}_t, \tag{1.3}$$

and derived the asymptotic distribution of the corresponding $\widehat{\mathrm{VaR}}_{\boldsymbol{x}}(\alpha)$. They assume that the predictors $\boldsymbol{X}_t$'s are independent and identically distributed random vectors.

In this paper, we work under the same predictive model (1.1) and derive the asymptotic distribution of the conditional VaR (1.2) based on the least squares estimator (1.3). Our estimator of the conditional quantile is different from the one obtained via quantile regression [Koenker and Bassett (1978)]. In the next section we show that our least squares estimator is asymptotically more efficient than the quantile regression estimator in regular cases, which is consistent with the simulation results in Chun et al. (2012); see Remark 2 below. We extend the assumptions of Chun et al. (2012) by allowing the predictors in Model (1.1) to be a stationary sequence, instead of being just an independent sequence, and furthermore, incorporating scenarios where some predictors have infinite variance. In risk management, the phenomenon of infinite-variance variables has attracted great attention since the seminal work by Mandelbrot (1963); see also, e.g., Fama and Roll (1971), Granger and Orr (1972), Rachev and Mittnik (2000), Rachev et al. (2005), Resnick (2007) and many references therein.

Using the well-developed approximation results of the residual-based empirical processes, see Koul (2002), Mammen (1963), Müller et al. (2012) and references therein, we derive the asymptotic variance of $\widehat{\mathrm{VaR}}_{\boldsymbol{x}}(\alpha)$, and point out that the asymptotic variance actually depends on whether all predictors have a finite variance or some of them have an infinite variance. Our results also show that the variance formula derived by Chun et al. (2012) is actually wrong, even when $E(\boldsymbol{X}_t \boldsymbol{X}_t^T) < \infty$ is assumed; see Remark 1 in Section 2.

Even with our correction, the asymptotic variance is so complicated that statistical inference for the conditional VaR is not trivial at all. To make the inference feasible, we further propose a smooth empirical likelihood method to effectively construct a confidence interval for the conditional VaR, without explicitly calculating the asymptotic variance.

Our method does not depend on whether all predictors have a finite variance or some of them have an infinite variance. This not only avoids the challenging issue of testing whether the variance of a predictor is finite or infinite, but also relaxes the restrictive boundedness conditions on predictor variables in many existing smoothed inference procedures, such as the well-known bootstrap method proposed in Horowitz (1998) and the empirical likelihood method based on quantile regression in Whang (2006). Furthermore, we extend the study for independent ϵ_t's to a GARCH sequence with potentially infinite kurtosis using a self-weighting method motivated by Ling (2007). In the supplementary material [He et al. (2019)] we also apply the same idea to quantile regression method, which generalizes the interval estimation for $\boldsymbol{\beta}$ in Whang (2006) to conditional VaR by allowing some infinite-variance predictors. As such, our proposed confidence intervals for the conditional VaR, given the current market situation, are useful in monitoring the risk of an asset or a financial variable.

We focus on linear models (in mean and conditional variance) for simplicity, which are useful in economic and financial applications; see, e.g., Engle (2001, 2004) for GARCH applications in value-at-risk forecasting, Cochrane (2009) for the discussions on linear factor models in finance, Adrian and Brunnermeier (2016) and Adrian and Giannone (2019) for macroeconomic applications. In the most general case, it is important to consider the nonlinear effects, and we leave them as possible future works; see, e.g., Spokoiny (2009), Lai and Xing (2013) and many references therein.

We organize the rest of the paper as follows. In Section 2, we derive the asymptotic distribution of the (least-squares) conditional VaR estimator. In Section 3, we propose the empirical likelihood method for deriving the corresponding confidence intervals via least squares estimation; in the supplement[He et al. (2019)] we also extend the idea to the quantile regression method. Section 4 generalizes the study for independent errors to GARCH errors. We further demonstrate the performance of the proposed method through a simulation study in Section 5 and a real data application in Section 6, respectively. We conclude the paper in Section 7. Additional simulation and data analysis results and all the details of the proofs are provided in the supplementary material[He et al. (2019)]. Unless specified otherwise, our asymptotic results hold as the sample size $n \to \infty$.

2. Methodology Based On Least Squares Estimation. In this section, we study the asymptotic distribution of the conditional VaR (1.2) based on the least squares estimator (1.3). We consider the same predictive model (1.1) as in Chun et al. (2012) and generalize the model assumptions in two aspects that are important for financial applications: 1) we allow $\{\boldsymbol{X}_t\}$ to be a stationary sequence; 2) we consider cases where some predictors have an infinite variance. In this section, we only work with (strongly) exogenous predictors for simplicity, while the results may be extended for weakly exogenous predictors at the cost of greater complications. In our main results for the empirical likelihood method in the next section, we relax the exogeneity condition and allow general predictors, including autoregressors. We denote $\xrightarrow{P}$ as convergence in probability and $\xrightarrow{d}$ as convergence in distribution.

First, we consider the case that all the predicting variables in the regression model (1.1) have a finite variance. More specifically, we assume the following conditions:

A1 $\{\epsilon_t\}$ is a sequence of independent and identically distributed random variables with zero mean and finite variance σ^2; $\{\epsilon_t\}$ and $\{\boldsymbol{X}_t\}$ are independent.

A2 $\{\boldsymbol{X}_t\}$ is a stationary sequence with $E(\|\boldsymbol{X}_t\|^{2+\iota_0}) < \infty$ for some $\iota_0 > 0$, and an ergodic sequence such that

$$\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{X}_t \xrightarrow{P} E\boldsymbol{X}_1,\ \frac{1}{n}\sum_{t=1}^{n}\boldsymbol{X}_t\boldsymbol{X}_t^T \xrightarrow{P} E(\boldsymbol{X}_1\boldsymbol{X}_1^T).$$

A3 Let $\boldsymbol{Z}_1 = (1, \boldsymbol{X}_1^T)^T$ and $\Omega := E(\boldsymbol{Z}_1\boldsymbol{Z}_1^T)$. Assume that Ω is positive definite.

A4 $F'_\epsilon(y)$ is continuous and positive in a neighborhood of $y = F_\epsilon^{-1}(\alpha)$.

Using the approximation results for the residual-based empirical process, see, e.g., equation (2.3) of Mammen (1963), and applying the delta method, we can expand that

$$\sqrt{n}\left(\widehat{\mathrm{VaR}}_{\boldsymbol{x}}(\alpha) - \mathrm{VaR}_{\boldsymbol{x}}(\alpha)\right) = \sum_{t=1}^{n} D_{nt} + o_p(1),$$

where

$$D_{nt} = \frac{1}{\sqrt{n}}\left\{-\frac{I(\epsilon_t \le \gamma) - \alpha}{F'_\epsilon(\gamma)} + \epsilon_t\boldsymbol{Z}_t^T\Omega^{-1}\boldsymbol{z} - \epsilon_t\boldsymbol{Z}_t^T\Omega^{-1}E(\boldsymbol{Z}_1)\right\}.$$

Note that $\{D_{nt}\}$ is a martingale difference array adapted to the filtration $\{\mathcal{F}_t\}_{t\ge 0} = \{\sigma(\epsilon_1, \ldots, \epsilon_t, \boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_t, \boldsymbol{Z}_{t+1})\}_{t\ge 0}$, where $\sigma(\cdot)$ means the sigma-algebra generated by random variables. Applying a proper version of martingale central limit theorem, such as Theorem 3.2 in Hall and Heyde (1980), we can establish the asymptotic normality of our VaR estimator as follows.

THEOREM 1. *Under conditions* A1–A4*, for* $\alpha \in (0,1)$*,*

$$\sqrt{n}\{\widehat{\mathrm{VaR}}_{\boldsymbol{x}}(\alpha) - \mathrm{VaR}_{\boldsymbol{x}}(\alpha)\} \xrightarrow{d} N(0, \omega^2 + \sigma^2 \boldsymbol{z}^T \Omega^{-1} \boldsymbol{z} + \Delta),$$

where $\boldsymbol{z} = (1, \boldsymbol{x}^T)^T$, $\omega^2 = \frac{\alpha(1-\alpha)}{\{F'_\epsilon(F_\epsilon^{-1}(\alpha))\}^2}$, *and* $\Delta = \Delta_1 + \Delta_2$ *with*

$$\Delta_1 = \sigma^2 E(\boldsymbol{Z}_1^T)\Omega^{-1}E(\boldsymbol{Z}_1) + 2\frac{E(\epsilon_1 I(\epsilon_1 \le F_\epsilon^{-1}(\alpha)))}{F'_\epsilon(F_\epsilon^{-1}(\alpha))}E(\boldsymbol{Z}_1^T)\Omega^{-1}E(\boldsymbol{Z}_1) \quad \textit{and}$$

$$\Delta_2 = -2\sigma^2 E(\boldsymbol{Z}_1^T)\Omega^{-1}\boldsymbol{z} - 2\frac{E(\epsilon_1 I(\epsilon_1 \le F_\epsilon^{-1}(\alpha)))}{F'_\epsilon(F_\epsilon^{-1}(\alpha))}E(\boldsymbol{Z}_1^T)\Omega^{-1}\boldsymbol{z}.$$

REMARK 1. *The last term* Δ *in the above asymptotic variance is missing in the variance formula (33) of Chun et al. (2012). When* $\boldsymbol{X}_t$ *is demeaned, i.e.,* $E(\boldsymbol{X}_t) = 0$*, we have* $\Delta = -\sigma^2$ *and thus the confidence interval overcovers when the standard error of Chun et al. (2012) is used. The first component* Δ_1 *is due to the impact of the statistical uncertainty of the least squares estimator* $\widehat{\boldsymbol{\beta}}$ *on residual quantile estimation. The second component* Δ_2 *attributes to the interaction between* $\widehat{\boldsymbol{\beta}}$ *and the residual quantile estimation, and it is easy to verify that* $\Delta_2 = 0$ *if* ϵ_t *is normally distributed. Our simulation study in Section 5 demonstrates that overlooking the* Δ*-term has a non-negligible impact on the coverage performance of the resulting confidence intervals.*

REMARK 2. *An alternative way to estimate the conditional* VaR *for our linear model* (1.1) *is to use the so-called linear quantile regression method in econometrics; see, e.g., Koenker (2005). The simulation study in Chun et al. (2012), e.g., Tables 1 and 2 therein, clearly demonstrates that the least squares estimator outperforms the quantile-regression-based estimator for many error distributions when* $\alpha = 0.95$*. To understand this in large sample theory, let us consider a regular case that* $E(\boldsymbol{X_t}) = 0$ *and, without loss of generality,* $\sigma = 1$*. Under the conditions of the above theorem, our least squares estimator has an asymptotic variance* $\frac{1}{n}\left\{\omega^2 + \boldsymbol{x}^T\Omega_X^{-1}\boldsymbol{x}\right\}$ *with* $\Omega_X = E\left(\boldsymbol{X_1}\boldsymbol{X_1^T}\right)$*, which is smaller than the asymptotic variance of the linear quantile regression estimator (derived from (3.7) in Koenker (2005)) given as* $\frac{1}{n}\left\{\omega^2 + \omega^2 \cdot \boldsymbol{x}^T\Omega_X^{-1}\boldsymbol{x}\right\}$ *if* $\omega^2 = \frac{\alpha(1-\alpha)}{\{F'_\epsilon(F_\epsilon^{-1}(\alpha))\}^2} > 1$*. The condition* $\omega^2 = \omega^2(\alpha) > 1$ *is trivial for large* α *in* VaR *applications under weak conditions such as normally and student-t distributed errors.*

Next, we show that the infinite-variance predictor does not play a role in the asymptotic variance of VaR estimator. Without loss of generality, we only consider the case where the kth predictor has an infinite variance, i.e., $EX_{t,k}^2 = \infty$; the proof for the cases with multiple infinite variance variables are completely analogous. Our proposed empirical likelihood method in the next section allows arbitrarily many infinite variance predictors and some predictors to be the lags of Y_t because we assume that ϵ_t is independent of $\{\boldsymbol{X}_s : s \le t\}$ for all $t \ge 1$. We now assume the following conditions:

B1 $\{\epsilon_t\}$ is a sequence of independent and identically distributed random variables with zero mean and finite variance σ^2; $\{\epsilon_t\}$ is independent of $\{\boldsymbol{X}_t\}$.

B2 $\{\boldsymbol{X}_t\}$ is a stationary sequence with $\frac{1}{n}\sum_{t=1}^n \boldsymbol{X}_t \xrightarrow{P} E\boldsymbol{X}_1$, $EX_{t,i}^2 < \infty$ for $i = 1, \dots, k-1$, $EX_{t,k}^2 = \infty$.

B3 Let $\widetilde{\boldsymbol{X}}_t = (X_{t,1}, \dots, X_{t,k-1})^T$. Assume $E(\|\widetilde{\boldsymbol{X}}_t\|^{2+\iota_0}) < \infty$ for some $\iota_0 > 0$, and as $n \to \infty$,

$$\frac{1}{n}\sum_{t=1}^n \widetilde{\boldsymbol{X}}_t \widetilde{\boldsymbol{X}}_t^T \xrightarrow{P} E(\widetilde{\boldsymbol{X}}_1 \widetilde{\boldsymbol{X}}_1^T).$$

B4 Let $\widetilde{\boldsymbol{Z}}_1 = (1, X_{1,1}, \dots, X_{1,k-1})^T$ and $\widetilde{\Omega} := E(\widetilde{\boldsymbol{Z}}_1 \widetilde{\boldsymbol{Z}}_1^T)$. Assume $\widetilde{\Omega}$ is positive definite.

B5 $F_\epsilon'(y)$ is continuous and positive in a neighborhood of $y = F_\epsilon^{-1}(\alpha)$.

B6 The distribution function F_k of $X_{t,k}$ lies in the domain of attraction of a stable law with index $d \in (1, 2)$ (see Feller (1971) for details on stable law). Further assume $\sum_{t=1}^n (X_{t,k} - E(X_{t,k})) = O_p(n^{1/d} L(n))$, and $\sum_{t=1}^n X_{t,k}^2 = O_p(n^{2/d} L(n))$ for some slowly varying $L(n)$, i.e., $L(tx)/L(t) \to 1$ for any $x > 0$ as $t \to \infty$. See Davis and Hsing (1995) for detailed conditions to ensure the above rate.

THEOREM 2. *Under conditions* B1–B6, *for* $\alpha \in (0,1)$

$$\sqrt{n}\{\widehat{\mathrm{VaR}}_{\boldsymbol{x}}(\alpha) - \mathrm{VaR}_{\boldsymbol{x}}(\alpha)\} \xrightarrow{d} N(0, \omega^2 + \sigma^2 \tilde{\boldsymbol{z}}^T \widetilde{\Omega}^{-1} \tilde{\boldsymbol{z}} + \widetilde{\Delta}),$$

where $\tilde{\boldsymbol{z}} = (1, x_1, \dots, x_{k-1})^T$, $\omega^2 = \frac{\alpha(1-\alpha)}{\{F_\epsilon'(F_\epsilon^{-1}(\alpha))\}^2}$, *and* $\widetilde{\Delta} = \widetilde{\Delta}_1 + \widetilde{\Delta}_2$ *with*

$$\widetilde{\Delta}_1 = \sigma^2 E(\widetilde{\boldsymbol{Z}}_1^T)\widetilde{\Omega}^{-1}E(\widetilde{\boldsymbol{Z}}_1) + 2\frac{E(\epsilon_1 I(\epsilon_1 \le F_\epsilon^{-1}(\alpha)))}{F_\epsilon'(F_\epsilon^{-1}(\alpha))} E(\widetilde{\boldsymbol{Z}}_1^T)\widetilde{\Omega}^{-1}E(\widetilde{\boldsymbol{Z}}_1) \quad \textit{and}$$

$$\widetilde{\Delta}_2 = -2\sigma^2 E(\widetilde{\boldsymbol{Z}}_1^T)\widetilde{\Omega}^{-1}\tilde{\boldsymbol{z}} - 2\frac{E(\epsilon_1 I(\epsilon_1 \le F_\epsilon^{-1}(\alpha)))}{F_\epsilon'(F_\epsilon^{-1}(\alpha))} E(\widetilde{\boldsymbol{Z}}_1^T)\widetilde{\Omega}^{-1}\tilde{\boldsymbol{z}}.$$

The proof of Theorem 2 is similar to that of Theorem 1, except it is more involved to remove the effect of the infinite-variance predictor(s). Using the fact that $\widehat{\beta}_k - \beta_k$ has a faster rate of convergence than n^{-1/d^*} for some $d^* \in (d, 2)$ [see, e.g., Davis and Wu (1997), Section 4], we can establish a similar martingale approximation of the residual-based empirical process for the subsample with $|X_{t,k}| \le n^{1/d^*}$ using the techniques in Shorack and Wellner (1986), Chapter 4. The result then follows after showing that the truncation effect is asymptotically trivial. A direct calculation indicates the negligible contribution from the infinite variance predictor in the asymptotic variance by again recalling the faster convergence rate of $\widehat{\beta}_k - \beta_k$.

3. Empirical Likelihood Method with Independent Errors. It is known that quantifying the statistical uncertainty of a risk measure is important and challenging in risk management; see, e.g., McNeil et al. (2015). Theorems 1 and 2 in Section 2 have provided the theoretical foundation for measuring the uncertainty of the conditional VaR (1.2). However, the theorems show that the asymptotic variance depends on whether there exist some infinite-variance predictors, which are difficult to verify in practice. Furthermore, even if one can do so, the asymptotic variance is so complicated that its calculation is non-trivial at all. Moreover, the applicability of a bootstrap method remains unknown due to possibly infinite variance predictors.

To make it feasible to quantify the uncertainty of the conditional VaR, here we propose an empirical likelihood method that uses a combination of estimating equations in Qin and Lawless (1994) and the smoothing technique in Chen and Hall (1993). The empirical likelihood is a distribution-free statistical inference method based on a data-driven likelihood ratio function. We refer to Owen (2001) for an overview of the empirical likelihood method, which has been shown to be quite effective in interval estimation and hypothesis testing. To effectively construct interval estimation of a VaR for regulatory purposes, empirical likelihood method for a quantile can be employed; see Chen and Hall (1993) and Baysal and Staum (2008).

First, we consider the least-squares settings and present the relevant notations. Define $X_{t,0} := 1$ and $\boldsymbol{Z}_t = (1, X_{t,1}, \ldots, X_{t,k})^T$ for $t \ge 1$. Consider a smooth distribution function $K(\cdot)$ and a bandwidth parameter $h = h(n) > 0$, and define

$$\begin{aligned} \boldsymbol{W}_t(\boldsymbol{\beta}, \theta) &=: \frac{1}{\|\boldsymbol{Z}_t\|^2} \widetilde{\boldsymbol{W}}_t(\boldsymbol{\beta}, \theta) \\ &=: \frac{1}{\|\boldsymbol{Z}_t\|^2} \left(\widetilde{W}_{t,1}(\boldsymbol{\beta}, \theta),\ \widetilde{W}_{t,2}(\boldsymbol{\beta}), \ldots, \widetilde{W}_{t,k+2}(\boldsymbol{\beta}) \right)^T \end{aligned} \tag{3.1}$$

with

$$\widetilde{W}_{t,1}(\boldsymbol{\beta},\theta) = K\left(\frac{\theta - \boldsymbol{\beta}^T\boldsymbol{z} - (Y_t - \boldsymbol{\beta}^T\boldsymbol{Z}_t)}{h}\right) - \alpha,$$
$$\widetilde{W}_{t,i+2}(\boldsymbol{\beta}) = (Y_t - \boldsymbol{\beta}^T\boldsymbol{Z}_t)X_{t,i} \quad \text{for} \quad i = 0,1,\ldots,k,$$

where $\boldsymbol{z}$ is defined earlier in Theorem 1, and $\|\cdot\|$ denotes Euclidean norm.

The reason to employ the weight $\|\boldsymbol{Z}_t\|^{-2}$ is to remove the effect of infinite moments of $\boldsymbol{X}_t$, so that we can have a unified inference procedure. One can choose different weight functions such as $\|\boldsymbol{Z}_t\|_\infty^{-2} := \left(\max_{0\le i\le k}|X_{t,i}|\right)^{-2}$. The observations $\widetilde{\boldsymbol{W}}_t(\boldsymbol{\beta},\theta)$ can be weighted less (or even not weighted at all), if the first (and second) moment(s) of $\|\boldsymbol{Z}_t\|$ is (are) finite; see Remarks 3 and 4 below. Choosing an optimal weight function is beyond the scope of this paper.

We now can define the empirical likelihood function for $\boldsymbol{\beta}$ and θ as

$$L(\boldsymbol{\beta},\theta) = \sup\left\{\prod_{t=1}^n (np_t) : p_1 \ge 0,\ldots,p_n \ge 0, \sum_{t=1}^n p_t = 1, \sum_{t=1}^n p_t\boldsymbol{W}_t(\boldsymbol{\beta},\theta) = 0\right\}.$$

Since we are only interested in the conditional VaR, $\theta^0 = \mathrm{VaR}_{\boldsymbol{x}}(\alpha)$, we treat $\boldsymbol{\beta}$ as the nuisance parameter and consider the so-called profile empirical likelihood function: $L^P(\theta) = \max_{\boldsymbol{\beta}} L(\boldsymbol{\beta},\theta)$. We use the original empirical likelihood proposed by Owen (1990) for simplicity. It may not be well-defined over the entire parameter space as the zero vector may be outside the convex hull of $\{\boldsymbol{W}_t(\boldsymbol{\beta},\theta)\}$ for some $(\boldsymbol{\beta},\theta)$. This is strong evidence that these points are not the true value if the models are correctly misspecified. We then adopt the convention that assigns $L(\boldsymbol{\beta},\theta) = 0$, or $\log L(\boldsymbol{\beta},\theta) = -\infty$, at such points; see, e.g., Kitamura (2007), Section 8 for further discussions. To construct positive likelihood values everywhere, one may alternatively work with the adjusted empirical likelihood [Chen et al. (2008), Emerson and Owen (2009)] by adding some clever pseudo observation(s), the penalized empirical likelihood [Bartolucci (2007), Lahiri and Mukhopadhyay (2012)] that drops the convex hull constraint, or the extended empirical likelihood [Tsao and Wu (2013)] that maps the full parameter space to the natural domain of the empirical likelihood function. Our proofs imply that the empirical likelihood function is well-defined with probability tending to 1 in a neighborhood around the true value, and we restrict the domain of the nuisance parameters in this neighborhood if necessary.

To derive the confidence interval, we need to establish a Wilks type of result for the twice-negative (profile) empirical log-likelihood function. For that purpose, we assume the following:

C1 $\{\epsilon_t\}$ is a sequence of independent and identically distributed random variables with zero mean and $E(|\epsilon_t|^{2+\iota_0}) < \infty$ for some $\iota_0 > 0$. For all $t \geq 1$, ϵ_t is independent of $\{\boldsymbol{X}_s : s \leq t\}$. Furthermore, $F'_\epsilon(y)$ is Lipschitz continuous and positive in a neighborhood of $y = F_\epsilon^{-1}(\alpha)$.

C2 The stationary sequence $\{\boldsymbol{X}_t\}$ is ergodic such that

$$\frac{1}{n}\sum_{t=1}^{n}\frac{\boldsymbol{Z}_t\boldsymbol{Z}_t^T}{\|\boldsymbol{Z}_t\|^2} \xrightarrow{P} E\left(\frac{\boldsymbol{Z}_1\boldsymbol{Z}_1^T}{\|\boldsymbol{Z}_1\|^2}\right), \quad \frac{1}{n}\sum_{t=1}^{n}\frac{\boldsymbol{Z}_t\boldsymbol{Z}_t^T}{\|\boldsymbol{Z}_t\|^4} \xrightarrow{P} E\left(\frac{\boldsymbol{Z}_1\boldsymbol{Z}_1^T}{\|\boldsymbol{Z}_1\|^4}\right).$$

C3 Define

$$\Sigma_1 := E\left\{\begin{pmatrix}\frac{I(\epsilon_1\leq F_\epsilon^{-1}(\alpha))-\alpha}{\|\boldsymbol{Z}_1\|^2}\\ \epsilon_1\boldsymbol{Z}_1/\|\boldsymbol{Z}_1\|^2\end{pmatrix}\left(\frac{I(\epsilon_1\leq F_\epsilon^{-1}(\alpha))-\alpha}{\|\boldsymbol{Z}_1\|^2}, \quad \epsilon_1\boldsymbol{Z}_1^T/\|\boldsymbol{Z}_1\|^2\right)\right\},$$

$$\Sigma_2 := \begin{pmatrix}F'_\epsilon(F_\epsilon^{-1}(\alpha))E\left((\boldsymbol{Z}_1-\boldsymbol{z})^T/\|\boldsymbol{Z}_1\|^2\right)\\ -E\left(\boldsymbol{Z}_1\boldsymbol{Z}_1^T/\|\boldsymbol{Z}_1\|^2\right)\end{pmatrix}.$$

Assume Σ_1 is positive definite and Σ_2 has full (column) rank.

C4 Assume $K(x) = \int_{-1}^{x} g(s)\,ds$, where $g(x)$ is a symmetric density function on $[-1,1]$ with bounded derivative. Furthermore, assume $nh^4 \to 0$ and $nh^{r_0} \to \infty$ for some $r_0 \in (2,4)$.

THEOREM 3. *Under the above conditions* C1–C4, $-2\log L^P(\mathrm{VaR}_{\boldsymbol{x}}(\alpha))$ *converges in distribution to* $\chi^2(1)$, *a chi-squared limit with one degree of freedom as* $n \to \infty$.

We develop a new approach to prove the Wilks-type theorems throughout the paper, including Theorem 3 and more later on, using the so-called local asymptotic normality (LAN) property of the empirical likelihood ratio function; see, e.g., Chapter 7 in van der Vaart (2000). This is more flexible than the high-order expansion approach by Qin and Lawless (1994), which does not allow smoothing under heavy tails. Specifically, for $\boldsymbol{\beta}$ in a local neighborhood around the true value $\boldsymbol{\beta}^0$, we show that

$$\begin{aligned}(3.2)\qquad -2\log L(\boldsymbol{\beta},\theta^0) =& \boldsymbol{\nu}^T\Sigma_2^T\Sigma_1^{-1}\Sigma_2\boldsymbol{\nu} + 2\boldsymbol{\nu}^T\Sigma_2^T\Sigma_1^{-1}\mathbb{W}_n + \mathbb{W}_n^T\Sigma_1^{-1}\mathbb{W}_n\\ &+ o_P(1) + o_P(\|\boldsymbol{\nu}\|) + o_P(\|\boldsymbol{\nu}\|^2),\end{aligned}$$

where

$$\mathbb{W}_n := \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{W}_t(\boldsymbol{\beta}^0,\theta^0) \text{ and } \boldsymbol{\nu} := \sqrt{n}(\boldsymbol{\beta}-\boldsymbol{\beta}^0).$$

The asymptotic orders of the reminder are uniform. For this, we need to establish the uniform convergence of the smoothed sample (second) moments and gradients for the parameters, which involves many technical details given in the supplementary material [He et al. (2019)].

Now, with probability tending to one, there exists a maximum empirical likelihood estimator $\widetilde{\boldsymbol{\beta}}$ in this neighborhood such that

$$-2\log L^P(\theta^0) = -2\log L(\widetilde{\boldsymbol{\beta}}, \theta^0) \tag{3.3}$$

and

$$\sqrt{n}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^0) = -(\Sigma_2^T \Sigma_1^{-1} \Sigma_2)^{-1} \Sigma_2^T \Sigma_1^{-1} \mathbb{W}_n + o_P(1), \tag{3.4}$$

by combining some arguments similar to that for Lemma 1 in Qin and Lawless (1994) and that for Theorem 2 in Sherman (1993). Substituting (3.4) into (3.2) and (3.3) yields that

$$-2\log L^P(\theta^0) = \mathbb{W}_n^T \Sigma_1^{-1/2} D \Sigma_1^{-1/2} \mathbb{W}_n + o_P(1),$$

where

$$D := I_{k+2} - \Sigma_1^{-1/2} \Sigma_2 (\Sigma_2^T \Sigma_1^{-1} \Sigma_2)^{-1} \Sigma_2^T \Sigma_1^{-1/2}$$

is a symmetric, idempotent matrix with unit trace and I_{k+2} denotes the identity $(k+2) \times (k+2)$ matrix. The chi-squared limit follows from the orthogonal decomposition of quadratic forms and the central limit theorem that $\Sigma_1^{-1/2} \mathbb{W}_n \xrightarrow{d} N(0, I_{k+2})$.

Based on Theorem 3, an asymptotic confidence interval for $\mathrm{VaR}_{\boldsymbol{x}}(\alpha)$ with level $\xi \in (0,1)$ is

$$I_\xi = \{\theta : -2\log L^P(\theta) \le \chi^2_{1,\xi}\},$$

where $\chi^2_{1,\xi}$ denotes the ξ-quantile of $\chi^2(1)$. The above interval can be effectively determined by using a standard search algorithm and often has good finite-sample coverage accuracy. As noted above, we adopt the convention that assigns a zero empirical likelihood value to the points where the convex hull condition is violated. Therefore, we exclude these points from the confidence intervals. We will implement the procedure and demonstrate the good coverage property in Section 5.

REMARK 3 (Predictors with finite mean). *When $E(\|\boldsymbol{X}_t\|) < \infty$, Theorem 3 remains true if we replace the weight $\|\boldsymbol{Z}_t\|^{-2}$ in $\boldsymbol{W}_t$ (3.1) by $\|\boldsymbol{Z}_t\|^{-1}$, and in Conditions C1-C3 replace the weights $\|\boldsymbol{Z}_t\|^{-2}$ by $\|\boldsymbol{Z}_t\|^{-1}$, and $\|\boldsymbol{Z}_t\|^{-4}$ by $\|\boldsymbol{Z}_t\|^{-2}$.*

Remark 4 (Predictors with finite variance). *When $E(\|\boldsymbol{X}_t\|^2) < \infty$, Theorem 3 remains true, if we remove the weight $\|\boldsymbol{Z}_t\|^{-2}$ in $\boldsymbol{W}_t$ (3.1), as well as the weights $\|\boldsymbol{Z}_t\|^{-2}$ and $\|\boldsymbol{Z}_t\|^{-4}$ in Conditions C1-C3.*

Remark 5. *As in Qin and Lawless (1994), we can study the asymptotic limit of the maximum empirical likelihood estimator defined by $\widehat{\theta}_{MELE} = \arg\max_\theta L^P(\theta)$. It has a different asymptotic variance from those given in Theorems 1 and 2 since the proposed profile empirical likelihood method is based on a weighted least squares estimator instead of the ordinary least squares estimator.*

We can apply the same weighted idea to generalize the empirical likelihood method for quantile regression in Whang (2006) by allowing infinite variance predictors. We construct a confidence interval for the conditional VaR rather than a confidence region for the regression coefficients therein. Full details of the methodology and the proofs are provided in the supplementary material [He et al. (2019)].

4. Empirical Likelihood Method with GARCH Errors. In this section, we generalize the empirical likelihood method to allow GARCH errors in the regression model. That is, we assume the errors ϵ_t's in (1.1) follow a GARCH process defined by the equation

$$\epsilon_t = \eta_t \sqrt{h_t}, \quad h_t = \omega + \sum_{i=1}^{r} a_i \epsilon_{t-i}^2 + \sum_{j=1}^{s} b_j h_{t-j}. \tag{4.1}$$

Denote the parameters by $\boldsymbol{\zeta} = (\omega, a_1, \ldots, a_r, b_1, \ldots, b_s)^T$ and the parameter space by $\Theta_{\boldsymbol{\zeta}}$. The true value $\boldsymbol{\zeta}^0$ satisfies the following conditions:

- G1 $\Theta_{\boldsymbol{\zeta}}$ is a compact subset of $[0, \infty)^{r+s+1}$, and $\boldsymbol{\zeta}^0$ is an interior point.
- G2 $\omega > 0$, $a_i, b_j \geq 0$ for all $i = 1, \ldots, r$ and $j = 1, \ldots, s$, $\sum_{i=1}^r a_i + \sum_{j=1}^s b_j < 1$, and there exists no common root for equations $\sum_{i=1}^r a_i z^i = 0$ and $1 - \sum_{j=1}^s b_j z^j = 0$.
- G3 $\{\eta_t\}$ is a sequence of independent and identically distributed random variables with zero mean, variance 1 and distribution function F_η. For all $t \in \{0, \pm 1, \ldots\}$, η_t is independent of $\{\boldsymbol{X}_s : s \leq t\}$. Furthermore, $F_\eta'(y)$ is Lipschitz continuous and positive in a neighborhood of $y = F_\eta^{-1}(\alpha)$.
- G4 $E(|\eta_t|^{4(1+\iota_0)}) < \infty$, $E(|\epsilon_t|^{2+\iota_0}) < \infty$ and $E\,\|\boldsymbol{X}_t\|^{\iota_0} < \infty$ for some $\iota_0 > 0$.

Conditions G1–G3 are standard assumptions for the stationarity and identifiability of the GARCH model. The moment condition on ϵ_t in G4 is the same as that in condition C1 for the least-squares method. The moment condition on η_t is to ensure the asymptotic normality of the quasi maximum likelihood estimator for GARCH parameters; see, e.g., Hall and Yao (2003).

Combining models (1.1) and (4.1), we are interested in the conditional VaR given by

$$\begin{aligned}\mathrm{VaR}_{\boldsymbol{x},\sigma}(\alpha) =& \inf\left\{q : P\left(Y_t \le q | \boldsymbol{X}_t = \boldsymbol{x}, \sqrt{h_t} = \sigma\right) \ge \alpha\right\} \\ =& \sigma F_\eta^{-1}(\alpha) + \boldsymbol{z}^T\boldsymbol{\beta},\end{aligned} \tag{4.2}$$

where $\boldsymbol{z} = (1, \boldsymbol{x}^T)^T$ and σ are given values. Note that h_t depends on Y_s for $s < t$. Slightly extending the notation, in this section, we denote the parameter for $\mathrm{VaR}_{\boldsymbol{x},\sigma}(\alpha)$ by θ.

Given the observations $\{(Y_t, {\boldsymbol{X}_t}^T)^T : t = 1, \ldots, n\}$ and the initial values $\{(Y_t, {\boldsymbol{X}_t}^T)^T : t \le 0\}$ which are generated by models (1.1) and (4.1), we write the parametric model as

$$\begin{aligned}\epsilon_t(\boldsymbol{\beta}) =& Y_t - \boldsymbol{Z}_t^T\boldsymbol{\beta},\ \eta_t(\boldsymbol{\zeta}, \boldsymbol{\beta}) = \frac{\epsilon_t(\boldsymbol{\beta})}{\sqrt{h_t(\boldsymbol{\zeta}, \boldsymbol{\beta})}},\ \text{and} \\ h_t(\boldsymbol{\zeta}, \boldsymbol{\beta}) =& \omega + \sum_{i=1}^r a_i\epsilon_{t-i}^2(\boldsymbol{\beta}) + \sum_{i=1}^s b_i h_{t-i}(\boldsymbol{\zeta}, \boldsymbol{\beta}).\end{aligned}$$

Here, $\epsilon_t = \epsilon_t(\boldsymbol{\beta}^0)$, $h_t = h_t(\boldsymbol{\zeta}^0, \boldsymbol{\beta}^0)$ and $\eta_t = \eta_t(\boldsymbol{\zeta}^0, \boldsymbol{\beta}^0)$ where $\boldsymbol{\beta}^0$ and $\boldsymbol{\zeta}^0$ are true parameters. Throughout the partial derivative $\frac{\partial \epsilon_t}{\partial \boldsymbol{\beta}}$ means $\frac{\partial \epsilon_t(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}$, and we use similar notations for other partial derivatives for η_t and h_t. For estimating $\mathrm{VaR}_{\boldsymbol{x},\sigma}(\alpha)$ in (4.2), we first estimate $\boldsymbol{\beta}$ by minimizing $\sum_{t=1}^n \epsilon_t^2(\boldsymbol{\beta})$ and then estimate $\boldsymbol{\zeta}$ by maximizing the quasi log-likelihood function

$$\sum_{t=1}^n l_t(\boldsymbol{\zeta}, \boldsymbol{\beta}), \quad l_t(\boldsymbol{\zeta}, \boldsymbol{\beta}) = -\frac{1}{2}\log h_t(\boldsymbol{\zeta}, \boldsymbol{\beta}) - \frac{\epsilon_t^2(\boldsymbol{\beta})}{2h_t(\boldsymbol{\zeta}, \boldsymbol{\beta})}$$

with $\boldsymbol{\beta}$ being replaced by the obtained estimator in the first step, and then estimate $F_\eta^{-1}(\alpha)$ by the empirical quantile of $\{\eta_t(\boldsymbol{\zeta}, \boldsymbol{\beta})\}_{t=1}^n$ with $\boldsymbol{\beta}$ and $\boldsymbol{\zeta}$ replaced by the obtained estimators in the first and second step. Obviously such an estimator will have a complicated asymptotic variance. To construct a confidence interval for this conditional VaR by allowing infinite variance predictors, one may apply the empirical likelihood method to some weighted estimating equations derived from the above three-step estimation. Unlike the case of independent errors, we need to bound the partial derivatives $\frac{\partial^2}{\partial \boldsymbol{\beta}\partial \boldsymbol{\zeta}^T} l_t(\boldsymbol{\zeta}, \boldsymbol{\beta})$ as we have to profile $\boldsymbol{\beta}$ eventually.

Like in Ling (2007), there exist $\rho \in (0,1)$ and a neighborhood of $\boldsymbol{\upsilon}^0 := (\boldsymbol{\zeta}^0, \boldsymbol{\beta}^0)$, say $\Theta_{\boldsymbol{\upsilon}}$, such that

$$\sup_{\Theta_{\boldsymbol{\upsilon}}} \left\| \frac{\partial^2 l_t(\boldsymbol{\upsilon})}{\partial \boldsymbol{\zeta} \partial \boldsymbol{\beta}^T} \right\| \leq C \xi_{\rho,t}^3 (1+\eta_t^2)(1+\widetilde{\xi}_{\rho,t-1}^{\iota}),$$

where $\xi_{\rho,t} = 1 + \sum_{i=0}^{\infty} \rho^i \|\boldsymbol{X}_{t-i}\|$ and $\widetilde{\xi}_{\rho,t-1} = 1 + \sum_{i=0}^{\infty} \rho^i |\epsilon_{t-1-i}|$, ι is any sufficiently small positive value and C is some constant that does not depend on t and n. Since $E(\eta_t^2) < \infty$ and $E|\epsilon_t|^{\iota} < \infty$ for small ι, we only need to bound $\xi_{\rho,t}^3$.

For any given $\rho \in (0,1)$, there exists $N \geq 1$ such that $\rho^i = e^{i \log \rho} \leq e^{-\log^2(i+1)}$ for all $i \geq N$, i.e., $\sum_{i=N}^{\infty} \rho^i \|\boldsymbol{X}_{t-i}\| \leq \sum_{i=N}^{\infty} e^{-\log^2(i+1)} \|\boldsymbol{X}_{t-i}\|$. Obviously, $\sum_{i=0}^{N} \rho^i \|\boldsymbol{X}_{t-i}\| \leq C \sum_{i=0}^{N} e^{-\log^2(i+1)} \|\boldsymbol{X}_{t-i}\|$ for some constant $C > 0$. Therefore, $\sum_{i=0}^{\infty} \rho^i \|\boldsymbol{X}_{t-i}\|$ can be bounded up to a constant by $\sum_{i=0}^{\infty} e^{-\log^2(i+1)} \|\boldsymbol{X}_{t-i}\|$.

This motivates to maximize the following weighted quasi likelihood function

$$\sum_{t=1}^{n} \widetilde{w}_t l_t(\boldsymbol{\zeta}, \boldsymbol{\beta}) \quad \text{with}$$

$$\widetilde{w}_t = \left(\sum_{i=0}^{\infty} e^{-\log^2(i+1)} \left\{ \mathbb{1}[\|\boldsymbol{X}_{t-i}\| \leq C_0] + C_0^{-1} \|\boldsymbol{X}_{t-i}\| \mathbb{1}[\|\boldsymbol{X}_{t-i}\| > C_0] \right\} \right)^{-3}$$

for some constant $C_0 > 0$. In practice, we do not observe $\|\boldsymbol{X}_s\|$ for $s \leq 0$ and have to replace them with some (small) constants. In this way, we can drop the initial values in the weight $\widetilde{w}_t$ completely and rewrite it as

$$w_t = \left(\sum_{i=0}^{t-1} e^{-\log^2(i+1)} \left\{ \mathbb{1}[\|\boldsymbol{X}_{t-i}\| \leq C_0] + C_0^{-1} \|\boldsymbol{X}_{t-i}\| \mathbb{1}[\|\boldsymbol{X}_{t-i}\| > C_0] \right\} \right)^{-3}$$

and C_0 is chosen as the 95% sample quantile of $\|\boldsymbol{X}_t\|$'s in our simulation study and real data analysis. The initial values $\|\boldsymbol{X}_s\|$ for $s \leq 0$ are ignorable, since by using $E\|\boldsymbol{X}_t\|^{\iota_0} < \infty$, we can show that the reminder satisfies

$$\max_{t \geq n^{\delta}} \sum_{i=t}^{\infty} e^{-\log^2(i+1)} ||\boldsymbol{X}_{t-i}|| = o_P(1) \quad \text{for any} \quad \delta \in (0,1);$$

see Lemmas F.5 and F.6 in our supplementary material [He et al. (2019)].

By taking δ small enough, $\sum_{t=1}^{n^{\delta}} w_t l_t(\boldsymbol{\zeta}, \boldsymbol{\beta})$ will be a smaller order term of $\sum_{t=n^{\delta}}^{n} w_t l_t(\boldsymbol{\zeta}, \boldsymbol{\beta})$. That is, we simply consider $\sum_{t=1}^{n} w_t l_t(\boldsymbol{\zeta}, \boldsymbol{\beta})$, and construct the estimation function

(4.3)

$$\bar{\boldsymbol{W}}_t(\boldsymbol{\zeta},\boldsymbol{\beta},\theta) =: \left(w_t \bar{W}_{t,1}(\boldsymbol{\zeta},\boldsymbol{\beta},\theta), \frac{\widetilde{W}_{t,2}(\boldsymbol{\beta})}{\|\boldsymbol{Z}_t\|^2}, \ldots, \frac{\widetilde{W}_{t,k+2}(\boldsymbol{\beta})}{\|\boldsymbol{Z}_t\|^2}, w_t \frac{\partial l_t(\boldsymbol{\zeta},\boldsymbol{\beta})}{\partial \boldsymbol{\zeta}^T}\right)^T$$

where $\widetilde{W}_{t,2}(\boldsymbol{\beta}), \ldots, \widetilde{W}_{t,k+2}(\boldsymbol{\beta})$ are the same as in (3.1) and

$$\bar{W}_{t,1}(\boldsymbol{\zeta},\boldsymbol{\beta},\theta) = K\left(\frac{(\theta - \boldsymbol{\beta}^T\boldsymbol{z})/\sigma - \eta_t(\boldsymbol{\zeta},\boldsymbol{\beta})}{h}\right) - \alpha,$$

$$\frac{\partial l_t(\boldsymbol{\zeta},\boldsymbol{\beta})}{\partial \boldsymbol{\zeta}} = -\frac{1}{2h_t(\boldsymbol{\zeta},\boldsymbol{\beta})}\frac{\partial h_t(\boldsymbol{\zeta},\boldsymbol{\beta})}{\partial \boldsymbol{\zeta}} + \frac{\eta_t^2(\boldsymbol{\zeta},\boldsymbol{\beta})}{2h_t(\boldsymbol{\zeta},\boldsymbol{\beta})}\frac{\partial h_t(\boldsymbol{\zeta},\boldsymbol{\beta})}{\partial \boldsymbol{\zeta}}.$$

Our empirical likelihood function is given by

$$\bar{L}(\boldsymbol{\zeta},\boldsymbol{\beta},\theta) = \sup\left\{\prod_{t=1}^n (np_t) : p_1 \geq 0, \ldots, p_n \geq 0, \sum_{t=1}^n p_t = 1, \sum_{t=1}^n p_t \bar{\boldsymbol{W}}_t(\boldsymbol{\zeta},\boldsymbol{\beta},\theta) = 0\right\},$$

and the profile empirical likelihood function is

$$\bar{L}^P(\theta) = \max_{\boldsymbol{\zeta},\boldsymbol{\beta}} \bar{L}(\boldsymbol{\zeta},\boldsymbol{\beta},\theta).$$

To profile out $\boldsymbol{\beta}$ in addition to $\boldsymbol{\zeta}$, we have to extend the conditions C2 and C3 as follows:

G5 The sequence $\{\underline{\boldsymbol{Y}}_t = (\varepsilon_t, \boldsymbol{X}_t, \varepsilon_{t-1}, \boldsymbol{X}_{t-1}, \ldots)^T\}$ is strictly stationary and ergodic.

G6 Define

$$\Omega_1 := E\left\{\begin{pmatrix} \widetilde{w}_1\left(I(\eta_1 \leq F_\eta^{-1}(\alpha)) - \alpha\right) \\ \epsilon_1 \boldsymbol{Z}_1/\|\boldsymbol{Z}_1\|^2 \\ \frac{1}{2}\widetilde{w}_1\left(\eta_1^2 - 1\right)\frac{1}{h_1}\frac{\partial h_1}{\partial \boldsymbol{\zeta}} \end{pmatrix}\begin{pmatrix} \widetilde{w}_1\left(I(\eta_1 \leq F_\eta^{-1}(\alpha)) - \alpha\right) \\ \epsilon_1 \boldsymbol{Z}_1/\|\boldsymbol{Z}_1\|^2 \\ \frac{1}{2}\widetilde{w}_1\left(\eta_1^2 - 1\right)\frac{1}{h_1}\frac{\partial h_1}{\partial \boldsymbol{\zeta}} \end{pmatrix}^T\right\},$$

$$\Omega_2 := E\begin{pmatrix} \frac{1}{2}\widetilde{w}_1 F_\eta'(F_\eta^{-1}(\alpha))\left(F_\eta^{-1}(\alpha)\frac{1}{h_1}\frac{\partial h_1}{\partial \boldsymbol{\zeta}}, F_\eta^{-1}(\alpha)\frac{1}{h_1}\frac{\partial h_1}{\partial \boldsymbol{\beta}} + \frac{\boldsymbol{Z}_1}{\sqrt{h_1}} - \frac{\boldsymbol{z}}{\sigma}\right)^T \\ \boldsymbol{0}_{(k+1)\times(1+r+s)}, \ -\boldsymbol{Z}_1\boldsymbol{Z}_1^T/\|\boldsymbol{Z}_1\|^2 \\ -\frac{1}{2}\widetilde{w}_1\frac{1}{h_1^2}\frac{\partial h_1}{\partial \boldsymbol{\zeta}}\frac{\partial h_1}{\partial \boldsymbol{\zeta}^T}, \ -\frac{1}{2}\widetilde{w}_1\frac{1}{h_1^2}\frac{\partial h_1}{\partial \boldsymbol{\zeta}}\frac{\partial h_1}{\partial \boldsymbol{\beta}^T} \end{pmatrix}.$$

Assume Ω_1 is positive definite and Ω_2 has full (column) rank.

THEOREM 4. *Under conditions G1–G6 and C4,* $-2\log \bar{L}^P(\mathrm{VaR}_{\boldsymbol{x},\sigma}(\alpha))$ *converges in distribution to* $\chi^2(1)$*, a chi-squared limit with one degree of freedom as* $n \to \infty$*.*

REMARK 6. *Developing an empirical likelihood method based on quantile regression for GARCH errors requires reparameterizing the GARCH model due to the identification issue as in Lee and Noh (2013), which complicates the comparison with the above empirical likelihood method based on least squares estimation. On the other hand, the simulation study below for independent errors shows that the interval derived from the empirical likelihood method based on least squares estimation is more accurate than that based on quantile regression. Hence we skip the study of proposing an empirical likelihood method based on quantile regression for GARCH errors.*

Similar to Theorem 3, it suffices to establish the LAN property of the empirical likelihood function, that is, uniformly for the nuisance parameters $\boldsymbol{\upsilon} = (\boldsymbol{\zeta}^T, \boldsymbol{\beta}^T)^T$ in a neighborhood around the true value $\boldsymbol{\upsilon}^0 = (\boldsymbol{\zeta}^{0T}, \boldsymbol{\beta}^{0T})^T$,

$$\begin{aligned}(4.4)\qquad -2\log \bar{L}(\boldsymbol{\upsilon}, \theta^0) =& \bar{\boldsymbol{\nu}}^T \Omega_2^T \Omega_1^{-1} \Omega_2 \bar{\boldsymbol{\nu}} + 2\bar{\boldsymbol{\nu}}^T \Omega_2^T \Omega_1^{-1} \bar{\mathbb{W}}_n + \bar{\mathbb{W}}_n^T \Omega_1^{-1} \bar{\mathbb{W}}_n \\ &+ o_P(1) + o_P(\|\bar{\boldsymbol{\nu}}\|) + o_P(\|\bar{\boldsymbol{\nu}}\|^2),\end{aligned}$$

where

$$\bar{\mathbb{W}}_n := \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \bar{\boldsymbol{W}}_t(\boldsymbol{\upsilon}^0, \theta^0), \text{ and } \bar{\boldsymbol{\nu}} := \sqrt{n}(\boldsymbol{\upsilon} - \boldsymbol{\upsilon}^0).$$

For this, we need the non-asymptotic bounds for $h_t(\boldsymbol{\zeta}, \boldsymbol{\beta})$ and $\eta_t(\boldsymbol{\zeta}, \boldsymbol{\beta})$ and for their (second-order) gradients over this neighborhood. We generalize the bounds in Ling (2007) therein to allow more general predictors (not necessarily autoregressors) in our model. Using these bounds, we can show that the downweighting is sufficient for the asymptotic normality of the quasi-score functions and the uniform convergence of the quasi-information matrix.

5. Simulation Study. In this section, we carry out a simulation study to illustrate the performance of the proposed weighted least squares estimation, empirical likelihood methods based on least squares estimation, and quantile regression.

We consider the following predictive linear model:

$$Y_t = 1 + 2X_{t,1} + 2X_{t,2} + \epsilon_t,$$

with $\beta_0 = 1, \beta_1 = 2, \beta_2 = 2$, and choose various types of predictors and error distributions. More specifically, for the predictors $(X_{t,1}, X_{t,2})^T$, we assume that they are mutually independent, and distributed according to one of the following three settings:

- $\{X_{t,1}\} \sim$ Student-t with 1.5 degrees of freedom ($t(1.5)$), and

$\{X_{t,2}\} \sim$ stationary autoregressive process of order 1 (AR(1)):

$$X_{t,2} = 0.355X_{t-1,2} + \phi_t, \qquad \phi_t \sim \mathrm{N}(0,1).$$

Note that $\{X_{t,1}\}$ then has an infinite variance.

- $\{X_{t,1}\} \sim t(1.5)$, and
  $\{X_{t,2}\} \sim$ stationary generalized autoregressive conditional heteroscedasticity model of lag 1 (GARCH(1,1)):

  $$X_{t,2} = \sigma_t\eta_t, \quad \sigma_t^2 = 0.1 + 0.7X_{t-1,2}^2 + 0.1\sigma_{t-1}^2, \qquad \eta_t \sim \mathrm{N}(0,1).$$

- $\{X_{t,1}\} \sim$ AR(1):

  $$X_{t,1} = 0.355X_{t-1,1} + \phi_t, \qquad \phi_t \sim \mathrm{N}(0,1), \quad \text{and}$$

  $\{X_{t,2}\} \sim$ GARCH(1,1):

  $$X_{t,2} = \sigma_t\eta_t, \quad \sigma_t^2 = 0.1 + 0.7X_{t-1,2}^2 + 0.1\sigma_{t-1}^2, \qquad \eta_t \sim \mathrm{N}(0,1).$$

The error ϵ_t is generated from one of the following three settings: 1) i.i.d. standard normal distribution (N$(0,1)$); 2) i.i.d. centered (i.e. shifted to have mean zero) log normal distribution with location parameter 0 and scale parameter 1/16 (LN(0,1/16)); 3) GARCH model $\epsilon_t = \sigma_t e_t$, $\sigma_t^2 = 0.1 + 0.3\sigma_{t-1}^2 + 0.3\epsilon_{t-1}^2$ with e_t standard normal innovations. Therefore, we study 9 combinations of different regressors and errors.

We choose the sample size(s) $n = 2000$ or/and 5000 for the value-at-risk level $\alpha = 0.95, 0.99$, the biweight kernel function $g(u) = \frac{15}{16}(1-u^2)^2\mathbb{1}[|u| \leq 1]$, the bandwidth $h = 0.5 \times n^{-1/3}$ or $1 \times n^{-1/3}$ or $1.5 \times n^{-1/3}$, and consider the predictor values $\boldsymbol{x} = (0.1, 0.1)^T$. Note that the order $n^{-1/3}$ for the bandwidth is motivated by the optimal bandwidth in smooth distribution function estimation.

5.1. *Point estimation.* Before investigating the coverage performance, we compare the mean and root of mean squared error (RMSE) of the weighted least squares estimator (1.2) and the standard quantile regression estimator. We set the value-at-risk level $\alpha = 0.95, 0.99$ and the sample size $n = 2000$. We repeat 10000 times and report the mean and RMSE in Table 1. Note that the true conditional VaR θ_0 for the case of GARCH errors is calculated via the average of $100,000$ quantiles at level α by simulating the GARCH error with sample size $n = 5000$. The weighted least squares estimator is consistently better in terms of RMSE, while the RMSE increases as α becomes larger. In next subsections, we shall show that the least-squares confidence interval also has better coverage performance especially when α is large.

TABLE 1

Mean and RMSE based on 10000 replications for the weighted least squares estimator and the quantile regression estimator with $\alpha = 0.95$ and $\alpha = 0.99$ and $n = 2000$.

Error	Regressors	θ_0	LSE		QR	
			Mean	RMSE	Mean	RMSE
$\alpha = 0.95$						
Normal	t+AR	3.0449	3.0416	0.0470	3.0438	0.0474
	t+GARCH		3.0419	0.0475	3.0440	0.0479
	AR+GARCH		3.0421	0.0473	3.0442	0.0480
Log Normal	t+AR	1.8769	1.8758	0.0180	1.8768	0.0180
	t+GARCH		1.8756	0.0179	1.8766	0.0181
	AR+GARCH		1.8756	0.0177	1.8766	0.0180
GARCH	t+AR	2.2046	2.2041	0.0302	2.2056	0.0305
	t+GARCH		2.2044	0.0310	2.2057	0.0313
	AR+GARCH		2.2043	0.0303	2.2056	0.0307
$\alpha = 0.99$						
Normal	t+AR	3.7263	3.7146	0.0839	3.7221	0.0835
	t+GARCH		3.7167	0.0835	3.7242	0.0847
	AR+GARCH		3.7150	0.0823	3.7236	0.0842
Log Normal	t+AR	2.1571	2.1522	0.0375	2.1560	0.0379
	t+GARCH		2.1518	0.0369	2.1559	0.0374
	AR+GARCH		2.1526	0.0365	2.1569	0.0375
GARCH	t+AR	2.6039	2.6145	0.0442	2.6171	0.0456
	t+GARCH		2.6143	0.0445	2.6168	0.0460
	AR+GARCH		2.6146	0.0447	2.6174	0.0462

5.2. *Interval Estimation for Independent Errors.* By considering independent errors, this subsection uses the R package 'emplik' to implement the proposed empirical likelihood confidence interval with level γ based on the weighted least squares estimator (LSE) and that based the quantile regression estimator (QR) developed in our supplement (see the discussion at the end of Section 3), and compares with the interval based on the asymptotic variance in Chun et al. (2012) (CSU) which tends to over-cover in theory.

By drawing 10000 random samples from each setting, we calculate the coverage probabilities of these three intervals at γ= 95% confidence level for $\mathrm{VaR}_{\boldsymbol{x}}(0.95)$ and $\mathrm{VaR}_{\boldsymbol{x}}(0.99)$. We repeat the analysis for $\gamma = 90\%$ confidence intervals in the supplementary material [He et al. (2019)], and the conclusions are qualitatively the same.

From the simulation results in Tables 2 and 3, we observe the followings.

i) The proposed empirical likelihood confidence interval based on least squares estimation has an accurate coverage probability for both $\alpha = 0.95$ and $\alpha = 0.99$, its performance improves as the sample size increases, and it is robust against the considered three choices of the

TABLE 2
Empirical coverage probabilities for the 95% confidence intervals of $\mathrm{VaR}_{\boldsymbol{x}}(0.95)$ *and* $\mathrm{VaR}_{\boldsymbol{x}}(0.99)$ *for* N*(0,1) errors, with bandwidth* $h_i = 0.5i \times n^{-1/3}$.

Regressors	t+AR		t+GARCH		AR+GARCH	
n	2000	5000	2000	5000	2000	5000
95% CI of $\mathrm{VaR}_{\boldsymbol{x}}(0.95)$, $\boldsymbol{x} = (0.1, 0.1)^T$						
CSU	0.9646	0.9679	0.9689	0.9703	0.9675	0.9696
LSE, h_1	0.9492	0.9509	0.9497	0.9513	0.9513	0.9490
LSE, h_2	0.9498	0.9507	0.9499	0.9506	0.9500	0.9493
LSE, h_3	0.9491	0.9508	0.9498	0.9494	0.9508	0.9481
QR, h_1	0.9479	0.9525	0.9539	0.9555	0.9564	0.9538
QR, h_2	0.9477	0.9507	0.9513	0.9530	0.9542	0.9516
QR, h_3	0.9456	0.9498	0.9505	0.9522	0.9521	0.9502
95% CI of $\mathrm{VaR}_{\boldsymbol{x}}(0.99)$, $\boldsymbol{x} = (0.1, 0.1)^T$						
CSU	0.9338	0.9469	0.9389	0.9462	0.9322	0.9506
LSE, h_1	0.9715	0.9478	0.9705	0.9516	0.9654	0.9528
LSE, h_2	0.9708	0.9470	0.9699	0.9510	0.9643	0.9527
LSE, h_3	0.9692	0.9470	0.9693	0.9511	0.9648	0.9529
QR, h_1	0.9489	0.9418	0.9215	0.9301	0.9199	0.8907
QR, h_2	0.9447	0.9399	0.9218	0.9270	0.9187	0.8923
QR, h_3	0.9430	0.9385	0.9239	0.9285	0.9159	0.8962

TABLE 3
Empirical coverage probabilities for the 95% confidence intervals of $\mathrm{VaR}_{\boldsymbol{x}}(0.95)$ *and* $\mathrm{VaR}_{\boldsymbol{x}}(0.99)$ *for centered-LN(0,1/16) errors, with bandwidth* $h_i = 0.5i \times n^{-1/3}$.

Regressors	t+AR		t+GARCH		AR+GARCH	
n	2000	5000	2000	5000	2000	5000
95% CI of $\mathrm{VaR}_{\boldsymbol{x}}(0.95)$, $\boldsymbol{x} = (0.1, 0.1)^T$						
CSU	0.9978	1.0000	0.9984	1.0000	0.9984	1.0000
LSE, h_1	0.9490	0.9511	0.9482	0.9497	0.9497	0.9489
LSE, h_2	0.9489	0.9486	0.9473	0.9489	0.9465	0.9459
LSE, h_3	0.9400	0.9415	0.9346	0.9388	0.9320	0.9359
QR, h_1	0.9479	0.9496	0.9325	0.9468	0.9379	0.9433
QR, h_2	0.9478	0.9480	0.9296	0.9431	0.9303	0.9318
QR, h_3	0.9382	0.9399	0.9178	0.9297	0.9170	0.9247
95% CI of $\mathrm{VaR}_{\boldsymbol{x}}(0.99)$, $\boldsymbol{x} = (0.1, 0.1)^T$						
CSU	0.9630	0.9729	0.9650	0.9728	0.9621	0.9740
LSE, h_1	0.9469	0.9469	0.9506	0.9508	0.9469	0.9525
LSE, h_2	0.9439	0.9458	0.9511	0.9500	0.9448	0.9513
LSE, h_3	0.9403	0.9437	0.9481	0.9468	0.9418	0.9455
QR, h_1	0.9223	0.8896	0.8294	0.7033	0.7851	0.6900
QR, h_2	0.9143	0.8804	0.8482	0.7164	0.7923	0.6956
QR, h_3	0.9020	0.8766	0.8492	0.7388	0.8078	0.7003

bandwidth h_i, $i = 1, 2, 3$.

ii) The proposed empirical likelihood confidence interval based on quantile regression has an accurate coverage probability when $\alpha = 0.95$. However its performance is unsatisfying when $\alpha = 0.99$ and $\{X_{t,2}\}$ follows from a GARCH(1,1) model. This may be explained by the fact that the conventional inference for quantile regression does not apply sufficiently far in the tails; see, e.g., Chernozhukov (2005).

iii) The poor performance of Chun et al. (2012)'s confidence intervals, especially for the non-normal cases in Table 3, confirms that the asymptotic variance in Chun et al. (2012) is incorrect.

5.3. *Empirical Likelihood Method for GARCH Errors.* This subsection investigates the finite sample performance of the proposed empirical likelihood interval based on the least squares estimator for GARCH errors given in Section 4.

TABLE 4

Empirical coverage probabilities for the 95% confidence intervals of $\mathrm{VaR}_{\boldsymbol{x},\sigma}(0.95)$ *and* $\mathrm{VaR}_{\boldsymbol{x},\sigma}(0.99)$ *for GARCH(1,1) errors (*$\epsilon_t = \sigma_t e_t$, $\sigma_t^2 = 0.1 + 0.3\sigma_{t-1}^2 + 0.3\epsilon_{t-1}^2$, $e_t \sim N(0,1)$*), with bandwidth* $h_i = 0.5i \times n^{-1/3}$.

Regressors	t+AR		t+GARCH		AR+GARCH	
n	2000	5000	2000	5000	2000	5000
95% CI of $\mathrm{VaR}_{\boldsymbol{x},\sigma}(0.95)$, $\boldsymbol{x} = (0.1, 0.1)^T$ and $\sigma = \sqrt{0.25}$						
LSE, h_1	0.952	0.965	0.952	0.948	0.960	0.960
LSE, h_2	0.946	0.962	0.942	0.949	0.952	0.952
LSE, h_3	0.946	0.959	0.944	0.950	0.947	0.953
95% CI of $\mathrm{VaR}_{\boldsymbol{x},\sigma}(0.99)$, $\boldsymbol{x} = (0.1, 0.1)^T$ and $\sigma = \sqrt{0.25}$						
LSE, h_1	0.966	0.954	0.966	0.961	0.960	0.964
LSE, h_2	0.974	0.945	0.967	0.956	0.952	0.957
LSE, h_3	0.970	0.941	0.958	0.946	0.951	0.953

Due to the heavy computation, we draw 1000 instead of 10000 random samples from the three settings for predictors and report the coverage probabilities in Tables 4. Again we only report the results for 95% confidence intervals and postpone that for 90% confidence intervals in the supplementary material [He et al. (2019)]. The results prefer a larger bandwidth and are (slightly) less accurate than those for the independent errors given in the previous subsection.

6. Real Data Analysis. This section illustrates how our proposed method may be employed to monitor the conditional value-at-risk of an individual financial institution, given the lagged information of economic state variables. We consider four US banks: Citigroup, JPMorgan Chase, Wachovia, and Wells Fargo. We assume a linear regression similar to Model (11a) in Adrian and Brunnermeier (2016).

Particularly, for each institution indexed by j, we construct the weekly prediction interval(s) for its next-week conditional VaR at 95% level based on the following model:

$$X_t^{(j)} = \beta_0^{(j)} + \sum_{i=1}^{7} \beta_i^{(j)} M_{i,t-1} + \epsilon_t^{(j)}, \tag{6.1}$$

$$\epsilon_t^{(j)} = \eta_t^{(j)} \sqrt{h_t^{(j)}},\ h_t^{(j)} = w^{(j)} + a^{(j)} \{\epsilon_{t-1}^{(j)}\}^2 + b^{(j)} h_{t-1}^{(j)}, \tag{6.2}$$

where $X_t^{(j)}$ is the weekly equity loss of the financial institution j at week t, and the predictors $M_{1,t-1}, \ldots, M_{7,t-1}$ are the one-week lagged value of the following seven economic state variables:

- $M_{1,t-1}$ – the change in the three-month yield;
- $M_{2,t-1}$ – the change in the slope of the yield curve, measured by the spread between the composite long-term bond yield and the three-month bill rate;
- $M_{3,t-1}$ – a short-term TED spread, defined as the difference between the three-month LIBOR rate and the three-month secondary market Treasury bill rate;
- $M_{4,t-1}$ – the change in the credit spread between Moody's Baa-rated bonds and the ten-year Treasury rate;
- $M_{5,t-1}$ – the weekly market return on the S&P500 index;
- $M_{6,t-1}$ – the weekly real estate sector return in excess of the market financial sector return;
- $M_{7,t-1}$ – equity volatility, which is computed as the 22-day rolling standard deviation of the daily CRSP equity market return.

For more details and explanations, we refer to Adrian and Brunnermeier (2016), page 1719. All the data are downloadable from `https://www.aeaweb.org/aer/data/10607/20120555_data.zip`.

To show the importance of avoiding the moments conditions, we plot the Hill (1975) estimates of the tail index (that is, the largest order of finite moment) as a function of the number of upper order statistics used in estimation for three predictors: the change in the three-month yield (yeild3m), the change in the credit spread (credit) and the change in the slope of the yield curve (term). Figure 1 shows that they all have tail indices close to or below 2, suggesting the possibility of an infinite variance. These Hill plots suggest that it is necessary to take the infinite variance into account.

Before computing the proposed intervals based on either independent errors (i.e., $a^{(j)} = b^{(j)} = 0$ in (6.2)) or GARCH errors, we examine the autocorrelation functions of $\{\epsilon_t^{(j)}\}$ and $\{\eta_t^{(j)}\}$ above. First, we fit model (6.1) by

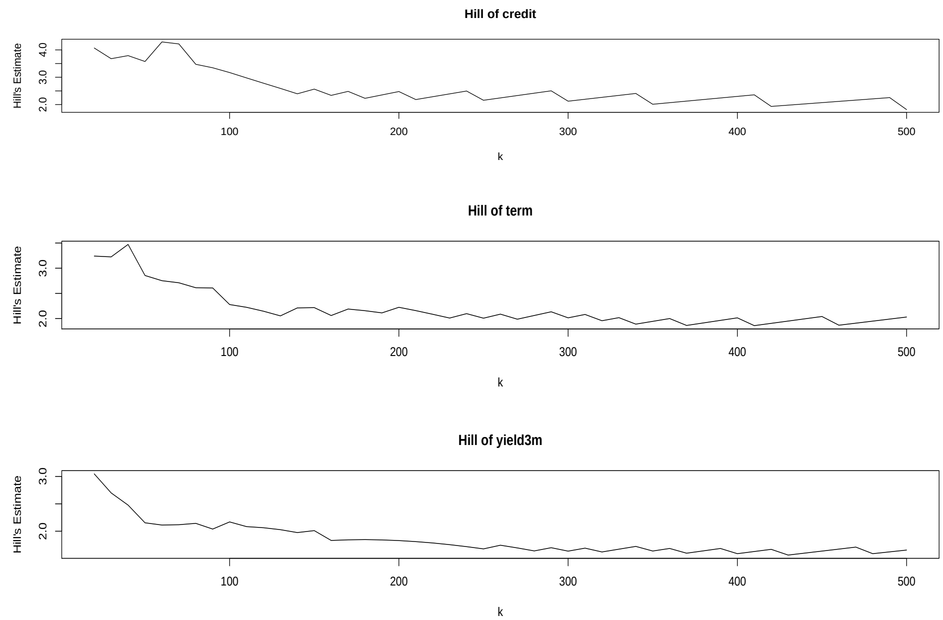


Fig 1: Hill estimates as a function of the number of upper order statistics for three weekly economic state variables that may have infinite variances. The names of the variables are from the original data set.

the least squares method and then plot the time series of (fitted) residuals $\{\epsilon_t^{(j)}\}$ and their autocorrelation functions. Second, we fit the GARCH model (6.2) to the least squares residuals by the quasi maximum likelihood method, and then plot the (fitted) GARCH residuals $\{\eta_t^{(j)}\}$ and their autocorrelation functions. Figure 2 shows the plots for our first bank, Citigroup, and both $\{\epsilon_t^{(j)}\}$ and $\{\eta_t^{(j)}\}$ suggest significant correlation at lag one. We carefully repeat the analysis for all other banks (JPMorgan Chase, Wachovia, and Wells Fargo), and observe the same pattern; see our supplementary materials [He et al. (2019)] for the other plots. These results suggest considering the following extended regression model instead of (6.1) in Adrian and Brunnermeier (2016):

$$X_t^{(j)} = \beta_0^{(j)} + \sum_{i=1}^{7} \beta_i^{(j)} M_{i,t-1} + \beta_8^{(j)} X_{t-1}^{(j)} + \beta_9^{(j)} X_{t-2}^{(j)} + \epsilon_t^{(j)}, \tag{6.3}$$

i.e., the considered model is a combination of equations (6.3) and (6.2), where the case of $a^{(j)} = b^{(j)} = 0$ leads to independent errors.

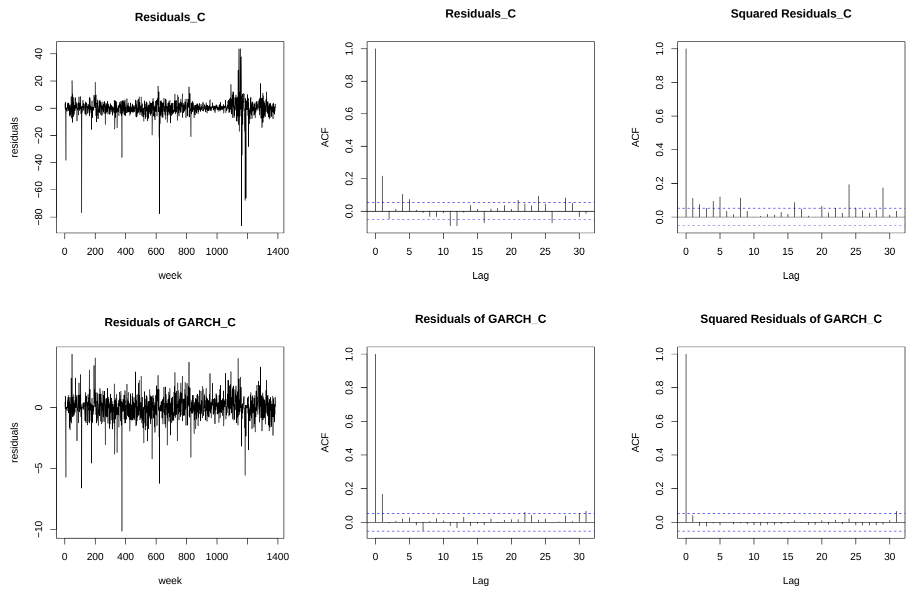


Fig 2: Time series and autocorrelation plots for Citigroup. The top plots are for the least-squares residuals $\{\epsilon_t^{(j)}\}$ from the model (6.1), and bottom plots are for the corresponding GARCH residual $\{\eta_t^{(j)}\}$ from the model (6.2). The middle plots are for their nominal levels, and right plots are for their squares.

Again, in Figure 3 we plot the time series, and the autocorrelations of the (squared) fitted least-squares residuals and GARCH residuals for Citigroup based on the extended regression model (6.3) and the GARCH model (6.2). The plots show a better fitness at least in terms of autocorrelations. We carefully check the same model specifications for all other banks and consistently observe good fitness, while the volatility clustering effects seem weaker for Wachovia and Wells Fargo. The results are comparable over all banks, and therefore we again leave the plots for other banks in the supplementary material[He et al. (2019)].

Hence, we apply the proposed empirical likelihood methods to construct intervals for the conditional Value-at-Risk, $\mathrm{VaR}_{\boldsymbol{x}}(0.95)$, based on (6.3) with independent errors, and $\mathrm{VaR}_{\boldsymbol{x},\sigma}(0.95)$ based on GARCH(1,1) errors (i.e., models (6.3) and (6.2)), where $\boldsymbol{x}$ is chosen as the latest one hundred observations of the predicting variables, i.e., $\{(M_{1,t-1},\dots,M_{7,t-1},X_{t-1}^{(j)},X_{t-2}^{(j)})^T : t=n,n-1,\dots,n-99\}$, and σ is chosen as the latest one hundred estimated σ_t in case of GARCH(1,1) errors. For replication purpose, we use the time

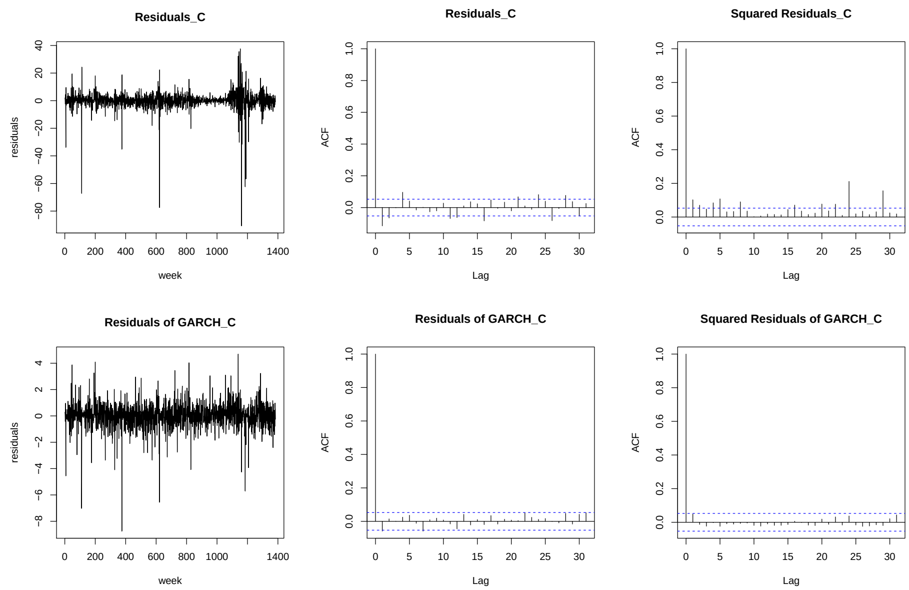
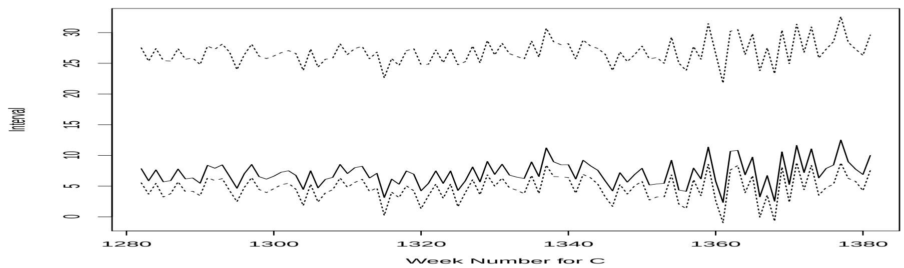


Fig 3: Time series and autocorrelation plots for Citigroup. The top plots are for the least-squares residuals $\{\epsilon_t^{(j)}\}$ from the model (6.3), and bottom plots are for the corresponding GARCH residual $\{\eta_t^{(j)}\}$ from the model (6.2). The middle plots are for their nominal levels, and right plots are for their squares.

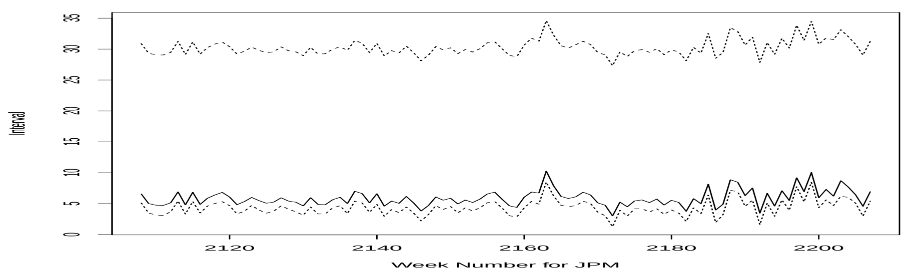
index as the week number in the dataset.

To calculate the empirical-likelihood-based interval for the conditional Value-at-Risk $\mathrm{VaR}_{\boldsymbol{x}}(0.95)$ based on independent errors, first, we compute the weighted least squares estimate of $\boldsymbol{\beta}^{(j)}$. Second, we estimate the conditional Value-at-Risk by combining the predicted mean and the empirical quantile of residuals. Third, we use these estimates as an initial value to solve $\sum_{t=1}^{n} \boldsymbol{W}_t(\boldsymbol{\beta}, \theta) = 0$ in Theorem 3, which gives $\widehat{\mathrm{VaR}}_{\boldsymbol{x}}(0.95)$. Finally, the interval is obtained by increasing and decreasing θ from $\widehat{\mathrm{VaR}}_{\boldsymbol{x}}(0.95)$ with step 0.01 until $-2\log L^P(\theta) > \chi^2_{1,0.95}$. The predicted values and the lower and upper endpoints of the empirical likelihood intervals with level 95% are plotted in Figure 4. We use the bandwidth $h = n^{-1/3}$ throughout the analysis.

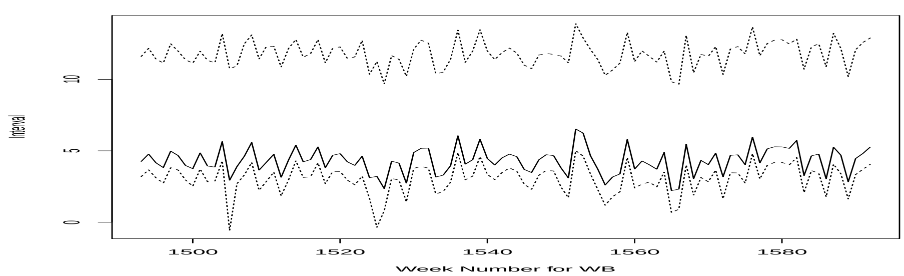
Similarly, to calculate the empirical-likelihood-based interval for the conditional Value-at-Risk $\mathrm{VaR}_{\boldsymbol{x},\sigma}(0.95)$ based on GARCH errors, first, we obtain weighted least squares estimate of $\boldsymbol{\beta}^{(j)}$ and residuals in (6.3). Second, we fit the GARCH model (6.2) to these residuals using the fGarch package in R.



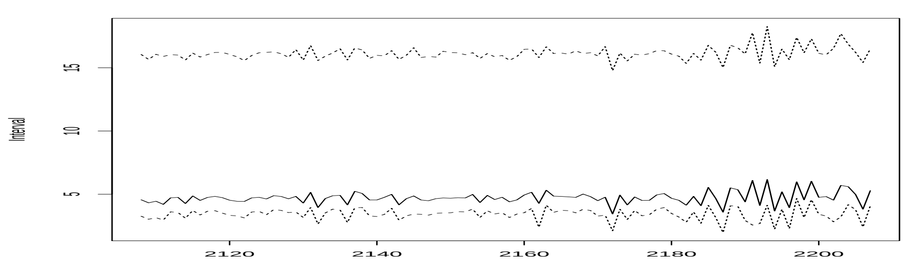
(a) Citigroup



(b) JPMorgan Chase

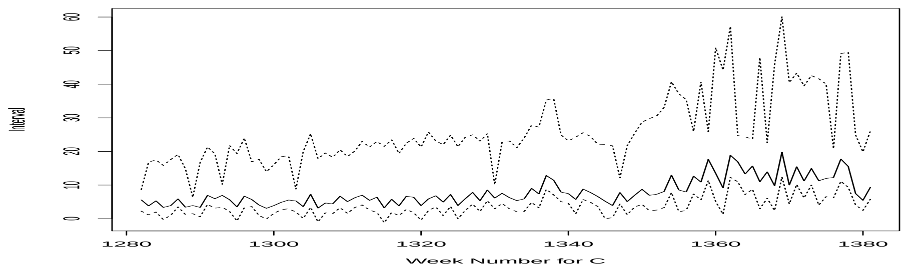


(c) Wachovia

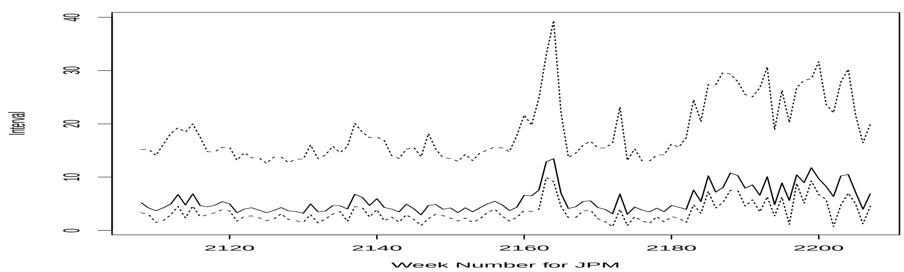


(d) Wells Fargo

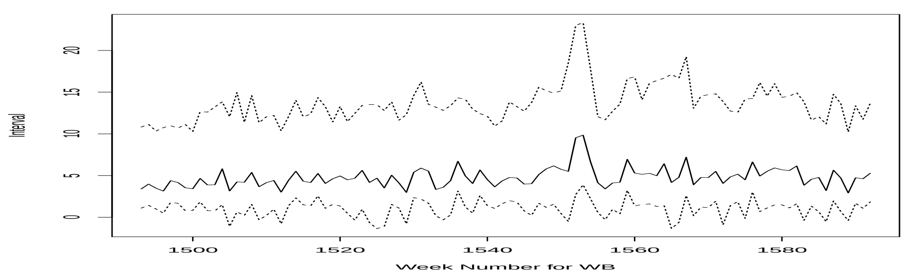
Fig 4: Empirical-likelihood-based 95% confidence intervals (dashed) and predicted values (solid) for the conditional VaR(0.95) of weekly bank loss (in percentage) based on regression Model (6.3) with independent errors. The time index is the week number in our dataset.



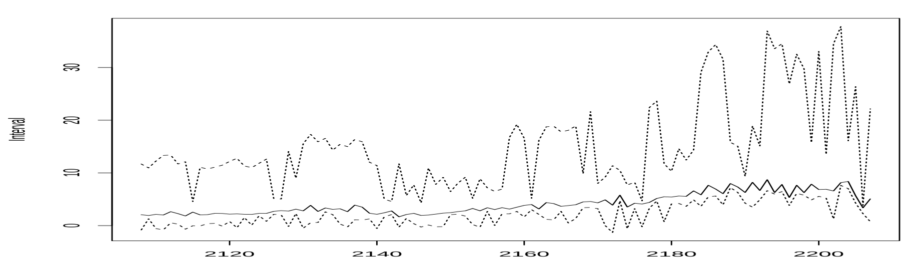
(a) Citigroup



(b) JPMorgan Chase



(c) Wachovia



(d) Wells Fargo

Fig 5: Empirical-likelihood-based 95% confidence intervals (dashed) and predicted values (solid) for the conditional VaR(0.95) of weekly bank loss (in percentage) based on the regression model (6.3) with GARCH errors (6.2). The time index is the week number in our dataset.

Third, we obtain an estimate of $\mathrm{VaR}_{\boldsymbol{x},\sigma}(0.95)$ by using the estimated parameters, and the empirical quantile of the GARCH residuals. Fourth, we use this Value-at-Risk estimate as an initial value to solve $\sum_{t=1}^{n} \bar{\boldsymbol{W}}_t(\boldsymbol{\zeta}, \boldsymbol{\beta}, \theta) = 0$ in Theorem 4, which gives $\widehat{\mathrm{VaR}}_{\boldsymbol{x},\sigma}(0.95)$. Finally, the interval is obtained by increasing and decreasing θ in Theorem 4 from $\widehat{\mathrm{VaR}}_{\boldsymbol{x},\sigma}(0.95)$ with step 0.01 until $-2\log \bar{L}^P(\theta) > \chi^2_{1,0.95}$. The point estimates and the lower and upper endpoints of the empirical likelihood intervals with level 95% are plotted in Figure 5.

One can observe that the empirical likelihood method produces asymmetric confidence intervals in both independent-errors and GARCH models, and the upper parts are typically much wider than the lower part. This may indicate the different severity of the estimation risk compared to the point estimates. The confidence interval by normal approximation can hardly capture such asymmetric property driven from the data. Also the independent-errors model produces relatively flatter point estimates and confidence intervals than the GARCH model across time. This indicates that the GARCH model is preferred when the data express the feature of heteroskedastic volatility. For example, the peak values in the plots of Figure 5 are not detected by the independent-errors model in the plots of Figure 4. Finally, Citigroup and JPMorgan Chase have wider intervals than Wachovia and Wells Fargo in general, and the intervals for Wachovia seems most robust against the model specifications. The upper endpoints of Well Fargo are bumpiest in the GARCH model.

7. Conclusions. When an asset or a financial variable is predicted by some known market variables or risk factors, a conditional VaR becomes more meaningful for risk management. This paper starts from revisiting the distribution-free estimation procedure proposed by Chun et al. (2012) and extends their asymptotic results for the conditional VaR estimator under more general model assumptions. We also correct their miscalculated asymptotic variance. We show that the asymptotic variance depends on whether there are predictors with infinite variance. Since the derived asymptotic variance is very complicated, this paper further proposes an empirical likelihood method to effectively construct a confidence interval for the conditional VaR. Our approach does not need to explicitly estimate the asymptotic variance nor know whether there exist infinite variance predictors. Furthermore, we apply the same idea to quantile regression method, which allows infinite variance predictors and extends the study in Whang (2006) to conditional VaR in the supplementary material. Numerical studies demonstrate that the proposed intervals are accurate and useful in monitoring/managing the risk of the underlying asset or financial variable.

Acknowledgments. We thank the editor Prof. Ming Yuan, an associate editor and two reviewers for their helpful comments.

SUPPLEMENTARY MATERIAL

Supplement to "Inference for Conditional Value-at-Risk of a Predictive Regression"
(doi: 10.1214/00-AOSXXXXSUPP; .pdf). In this supplement, we provide more simulation and empirical analysis results, and we prove all the theorems stated in this paper.

References.

Adrian, T. and Brunnermeier, M. (2016), CoVaR, *American Economic Review* **106** 1705–1741.

Adrian, T., Boyarchenko, N. and Giannone, D. (2019), Vulnerable growth, *American Economic Review*, **109**, 1263–1289.

Bartolucci, F. (2007), A penalized version of the empirical likelihood ratio for the population mean. *Statistics & probability letters*, **77**, 104–110.

Baysal, R. E. and Staum, J. (2008), Empirical likelihood for value-at-risk and expected shortfall, *The Journal of Risk* **11** 3–32.

Bollerslev, T. (1986), Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics* **31** 307–327.

Chernozhukov, V. (2005), Extremal quantile regression, *The Annals of Statistics* **33** 806–839.

Chen, S. X. and Hall, P. (1993), Smoothed empirical likelihood confidence intervals for quantiles, *The Annals of Statistics* **21** 1166–1181.

Chen, J., Variyath, A.M. and Abraham, B. (2008), Adjusted empirical likelihood and its properties. *Journal of Computational and Graphical Statistics*, **17**, 426–443.

Cochrane, J.H. (2009). *Asset pricing: Revised edition*, Princeton University Press.

Christoffersen, P. and Gonçalves, S. (2005), Estimation risk in financial risk management, *The Journal of Risk* **7** 1–28.

Chun, S. Y., Shapiro, A. and Uryasev, S. (2012), Conditional value-at-risk and average value-at-risk: estimation and asymptotics, *Operations Research* **60** 739–756.

Davis, R. A. and Hsing, T. (1995), Point process and partial sum convergence fo weakly dependent random variables with infinite variance, *The Annals of Probability* **23** 879–917.

Davis, R. A. and Wu, W. (1997), M-estimation for linear regression with infinite variance, *Probability and Mathematical Statistics* **17** 1–20.

Duffie, D. and Pan, J. (1997), An overview of value at risk, *The Journal of derivatives* **4** 7–49.

Duffie, D. and Pan, J. (2001), Analytical value-at-risk with jumps and credit risk, *Finance and Stochastics* **5** 155–80.

Emerson, S. C., and Owen, A. B. (2009), Calibration of the empirical likelihood method for a vector mean. *Electronic Journal of Statistics*, **3**, 1161–1192.

Engle, R. (2001), GARCH 101: The use of ARCH/GARCH models in applied econometrics. *Journal of Economic Perspectives*, **15** 157–168.

Engle, R. (2004), Risk and volatility: Econometric models and financial practice. *American Economic Review*, **94**, 405–420.

Engle, R. F. and Manganelli, S. (2004), CAViaR: conditional autoregressive value at risk by regression quantiles, *Journal of Business and Economic Statistics* **22** 367–381.

Fama, E. F. and Roll, R. (1971), Parameter estimates for symmetric stable distributions, *Journal of the American Statistical Association* **66** 331–338.

Fan, Y., Härdle, W.K., Wang, W. and Zhu, L. (2018), Single index based CoVaR with very high dimensional covariates, *Journal of Business and Economic Statistics*, **36**, 212–226.

Feller, W. (1971), *An Introduction to Probability Theory and Its Applications*, Vol. 2, 2nd Edition, Wiley.

Gao, F. and Song, F. (2008), Estimation risk in GARCH VaR and ES estimates, *Econometric Theory* **24** 1404–1424.

Granger, C. W. J. and Orr, D. (1972), Infinite variance and research strategy in time series analysis, *Journal of the American Statistical Association* **67** 275–285.

Hall, P. and Heyde, C. C. (1980), *Martingale Limit Theory and Its Applications*. Academic Press.

Hall, P., and Yao, Q. (2003), Inference in ARCH and GARCH models with heavy–tailed errors, *Econometrica* **71** 285–317.

He, Y., Hou, Y., Peng, L., and Shen, H. (2019), Supplement to “Inference for Conditional Value-at-Risk of a Predictive Regression”.

Hill, B.M. (1975), A simple general approach to inference about the tail of a distribution,” *The Annals of Statistics* **3** 1163–1174.

Horowitz, J. (1998), Bootstrap Methods for Median Regression Models, *Econometrica* **66** 1327-1351.

Jorion, P. (1996), Risk^2: Measuring the risk in value at risk, *Financial Analysts Journal* **52** 47–56.

Jorion, P. (2007), *Value at risk: The New Benchmark for Managing Financial Risk,* 3rd ed, McGraw-Hill, Boston.

Kitamura, Y. (2007), Empirical Likelihood Methods in Econometrics: Theory and Practice. In *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress*. Cambridge University Press.

Koenker, R. and Bassett, G. (1978), Regression Quantiles, *Econometrica: Journal of the Econometric Society* **46** 33–50.

Koenker, R. (2005), *Quantile Regression*. Cambridge University Press.

Kou, S. and Peng, X. (2016), On the measurement of economic tail risk, *Operations Research* **64** 1056–1072.

Kou, S., Peng, X. and Heyde, C. C. (2013), External risk measures and Basel accords, *Mathematics of Operations Research* **38** 393–417.

Koul, H.L. (2002), *Weighted Empirical Processes in Dynamic Nonlinear Models.* Lecture Notes in Statistics, Vol. **166**. Springer, New York.

Lahiri, S. N., and Mukhopadhyay, S. (2012), A penalized empirical likelihood method in high dimensions. *The Annals of Statistics*, **40** 2511–2540.

Lai, Tze Leung, and Xing, H. (2013), Stochastic change-point ARX-GARCH models and their applications to econometric time series, *Statistica Sinica* **23** 1573–1594.

Lee, S. and Noh, J. (2013), Quantile Regression Estimator for GARCH Models, *Scandinavian Journal of Statistics* **40** 2–20.

Ling, S. (2007), Self-weighted and local quasi-maximum likelihood estimators for ARMA-GARCH/IGARCH models, *Journal of Econometrics* **140** 849–873.

Mammen, E. (1996). Empirical processes of residuals for high-dimensional linear models, *The Annals of Statistics* **24** 307–335.

Mandelbrot, B. (1963), The Variation of Certain Speculative Prices, *The Journal of Business* **36** 394-419.

McNeil, A. J., Frey, R. and Embrechts, P. (2015), *Quantitative risk management: Concepts, techniques and tools.* Princeton University Press.

Müller, U.U., Schick, A. and Wefelmeyer, W. (2012). Estimating the error distribution function in semiparametric additive regression models. *Journal of Statistical Planning and Inference* **142** 552–566.

Owen, A. (1990), Empirical Likelihood Ratio Confidence Regions, *The Annals of Statistics* **18** 90–120.

Owen, A. (2001), *Empirical Likelihood*. Chapman & Hall/CRC, New York.

Qin, J. and Lawless, J. (1994), Empirical likelihood and general estimating equations, *The Annals of Statistics* **22** 300–325.

Rachev, S. T. and Mittnik, S. (2000), *Stable Paretian Models in Finance*, Wiley, New York.

Rachev, S. T., Menn, C. and Fabozzi, F. J. (2005), *Fat-tailed and skewed asset return distributions: implications for risk management, portfolio selection, and option pricing*, Wiley, Hoboken, New Jersey.

Resnick, S. (2007), *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*, Springer, New York.

Sherman, R. P. (1993), The limiting distribution of the maximum rank correlation estimator, *Econometrica* **61** 123–137.

Galen R. Shorack and Jon A. Wellner (1986), *Empirical Processes with Applications to Statistics*, SIAM.

Spokoiny, Vladimir.(2009), Multiscale local change point detection with applications to Value-at-Rrisk. *The Annals of Statistics* **37** 1405–1436.

Tsao, M. and Wu, F.(2013). Empirical likelihood on the full parameter space. *The Annals of Statistics* **41**, 2176–2196.

van der Vaart, A. W. (2000), *Asymptotic Statistics*, Cambridge University Press.

Whang, Y. J. (2006), Smoothed empirical likelihood methods for quantile regression models, *Econometric Theory* **22** 173-205.

Xu, K. (2016), Model-free inference for tail risk measures, *Econometric Theory* **32** 122–153.

Amsterdam School of Economics
University of Amsterdam
Amsterdam 1018 WB, The Netherlands
E-mail: y.he2@uva.nl

School of Data Science
Fudan University
Shanghai 200433, China
E-mail: oishihou@yahoo.com

Department of Risk Management and Insurance
Georgia State University
Atlanta, GA 30303, USA
E-mail: lpeng@gsu.edu

Faculty of Business and Economics
The University of Hong Kong
Pok Fu Lam Road, Hong Kong
E-mail: haipeng@hku.hk