Efficient estimation of conditional risk measures in a semiparametric GARCH model

Yang Yan
Dajing Shang
Oliver Linton

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP25/12

Efficient Estimation of Conditional Risk Measures in a Semiparametric GARCH Model*

Yang Yan † Dajing Shang‡
The London School of Economics

Oliver Linton§
University of Cambridge

May 10, 2012

Abstract

This paper proposes efficient estimators of risk measures in a semiparametric GARCH model defined through moment constraints. Moment constraints are often used to identify and estimate the mean and variance parameters and are however discarded when estimating error quantiles. In order to prevent this efficiency loss in quantile estimation, we propose a quantile estimator based on inverting an empirical likelihood weighted distribution estimator. It is found that the new quantile estimator is uniformly more efficient than the simple empirical quantile and a quantile estimator based on normalized residuals. At the same time, the efficiency gain in error quantile estimation hinges on the efficiency of estimators of the variance parameters. We show that the same conclusion applies to the estimation of conditional Expected Shortfall. Our comparison also leads to interesting implications of residual bootstrap for dynamic models. We find that these proposed estimators for conditional Value-at-Risk and expected shortfall are asymptotically mixed normal. This asymptotic theory can be used to construct

*Thanks to Xiaohong Chen for interesting discussions.

†Department of Statistics, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom.Email Address: y.yan3@lse.ac.uk

‡Financial Markets Group, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom. E-mail address: d.shang@lse.ac.uk

§Department of Economics, University of Cambridge, Austin Robinson Building, Sidgwick Avenue, Cambridge CB3 9DD, United Kingdom, e-mail: obl20@cam.ac.uk. Web Page: https://sites.google.com/site/oliverlinton/oliver-linton. This paper was partly written while I was at Universidad Carlos III de Madrid-Banco Santander Chair of Excellence, and I thank them for financial support.

confidence bands for these estimators by taking account of parameter uncertainty. Simulation evidence as well as empirical results are provided.

KEYWORDS: Empirical Likelihood; Empirical process; GARCH; Quantile; Value-at-Risk; Expected Shortfall.

JEL-CLASSIFICATION: C14, C22, G22.

1 Introduction

Many popular time series models specify some parametric or nonparametric structure for the conditional mean and variance. Often, these models are completed by a sequence of i.i.d errors ε_t.[1] For example, many models can be written in the form of $\mathcal{T}(y_t, y_{t-1}, \ldots; \theta) = \varepsilon_t$, where the parametric model $\mathcal{T}(\cdot; \theta)$ is used to remove the temporal dependence structure in y_t so that the error ε_t is i.i.d with certain distribution $F(\cdot)$. Parameters θ and $F(\cdot)$ together define the model. Often one assumes moment conditions on ε_t such as it being mean zero and variance one. These moment constraints are often used to identify and estimate the mean and variance parameters θ but are however often discarded when estimating the error distribution or quantile. This paper considers how best to utilize this information to estimate the distribution $F(\cdot)$, and further the quantiles of ε_t, so that one can construct an efficient estimator for the conditional distribution and hence quantiles of y_{t+1} given $\mathcal{F}_t = \{y_t, y_{t-1}, \ldots, y_0\}$.

Knowledge of the conditional distribution is very important in finance since all financial instruments are more or less pricing or hedging certain sections of the distribution of underlying assets. For example, mean-variance trade-off in portfolio management is concerned with the first and second moments; exotic derivatives are traded for transferring downside risks, which are lower portions of the asset's distribution. Other practical usage of conditional distribution estimation includes the risk-neutral density estimation and Value-at-Risk (VaR) estimation. VaR is defined as the maximum potential loss in value of a portfolio of financial instruments with a given confidence ratio $100\alpha\%$ (typically 1% or 5%) over a certain horizon. From a statistical point of view, VaR is the $100\alpha\%$ quantile of the conditional distribution of portfolio returns. VaR has becoming the industrial standard for risk measures due to the mandates of the Basel I capital accords. It is used to quantify market risks and set trading limits. An intuitive interpretation of VaR is uninsured loss tolerance, or willingness to pay for risks. Direct estimation of conditional VaR has been considered by Engle and Manganelli (2004), and pursued further in Koenker and Xiao (2009) from the point of view of quantile autoregression. Recently, it has been argued that Value at Risk is not a coherent measure of risk, specifically it can violate the subadditivity axiom of Artzner et al. (1999). Instead the expected shortfall (ES) is an alternative risk measure that does satisfy all of their axioms. ES is defined as the expected return on the portfolio in the worst $100\alpha\%$ of the cases. ES incorporates more information than VaR because ES gives the average loss in the tail below $100\alpha\%$. The estimation of unconditional ES has been considered in Scaillet (2004) and Chen (2008). The recent Basel Committee on Banking Supervision round III has suggested using expected shortfall in place of value at risk, so this measure is likely to gain in prominence in

[1] There are some notable exceptions to this including Engle and Manganelli (2004).

the future.

We consider the following popular AR(p)-GARCH(1,1) model

$$\begin{aligned} y_t &= \sum_{j=1}^{p} \rho_j y_{t-j} + h_t^{1/2} \varepsilon_t \\ h_t &= \omega + \beta h_{t-1} + \gamma u_{t-1}^2, \end{aligned} \tag{1}$$

where $u_t = h_t^{1/2} \varepsilon_t$, and $\{\varepsilon_t\}$ is an i.i.d sequence of innovations with mean zero and variance one and p is a finite and known integer. We suppose that ε_t has a density function $f(\cdot)$, which is unknown apart from the two moment conditions:

$$\int x f(x) dx = 0; \int x^2 f(x) dx = 1. \tag{2}$$

These moment conditions are standard in parametric settings and identify h_t as the conditional variance of y_t given $\mathcal{F}_{t-1}$. Furthermore, the error density and all the parameters are jointly identified in the semiparametric model. In this case, the conditional Value-at-Risk of y_t given $\mathcal{F}_{t-1}$ and the conditional expected shortfall of y_t given $\mathcal{F}_{t-1}$ are respectively,

$$\xi_t(\alpha) = \sum_{j=1}^{p} \rho_j y_{t-j} + h_t^{1/2} q_\alpha$$

$$\chi_t(\alpha) = E[y_t | 1(y_t \leq \xi_t(\alpha)), \mathcal{F}_{t-1}] = \sum_{j=1}^{p} \rho_j y_{t-j} + h_t^{1/2} E[\varepsilon_t | 1(\varepsilon_t \leq q_\alpha)] = \sum_{j=1}^{p} \rho_j y_{t-j} + h_t^{1/2} ES_\alpha,$$

where q_α is the α-quantile of ε_t, while $ES_\alpha = E[\varepsilon_t | 1(\varepsilon_t \leq q_\alpha)]$ is the α-expected shortfall of ε_t. In the sequel we assume that $p = 0$ for simplicity of notation. This is quite a common simplification in the literature; the main thrust of our results carry over to the more general p case.

Let $\theta = (\omega, \beta, \gamma)$. The goal of this paper is to estimate the parameters $(\theta, q_\alpha, ES_\alpha)$ efficiently and plug in these efficient estimators to obtain the conditional quantile $\widehat{\xi}_{n,t} = h_t^{1/2}(\widehat{\theta})\widehat{q}_\alpha$ and the conditional expected shortfall $\widehat{\chi}_t(\alpha) = h_t^{1/2}(\widehat{\theta})\widehat{ES}_\alpha$.

Since this model involves both finite dimensional parameters θ and infinite dimensional parameter $f(\cdot)$, we call it a semiparametric model. This paper constructs an efficient estimator for both θ and the α'th quantile of $f(\cdot)$, q_α, for model (1) under moment constraints (2). Consequently, the conditional quantile estimator and conditional expected shortfall estimator are efficient.

This dynamic model (1) is widely used for financial time series due to its empirical success in modelling financial time series. Financial time series exhibit variances that changes over time. This time-varying volatility is best described in GARCH models. Kuester, Mittnik and Paolella (2006) argued that quantile regression based on this model is the best compared to other parametric and nonparametric models. This quantile estimation method is often referred as historical simulation method (Kuester, Mittnik and Paolella (2006)). This method employs volatility estimator as a filter to transform the conditional correlated returns into iid errors, for which vast quantile estimators such as empirical quantile or extreme value theory based quantile can be readily applied. For example, Riskmetrics (1996) employs a GARCH model with normal errors; McNeil and Frey (2000) propose new VaR forecast methods by combining GARCH models with Extreme Value Theory (EVT); Engle (2001) illustrates VaR forecasts in GARCH models with empirical quantiles; Nyström and Skoglund (2004) use GMM-type volatility estimators for the GARCH based VaR forecasts. See Duffie and Pan (1997), Engle and Manganelli (2004) and Gourieroux and Jasiak (2002) for more detailed surveys for VaR forecasts.

Estimation of GARCH parameters has a long history. Consistency and asymptotic normality have been established under various conditions, see Weiss (1986), Lee and Hansen (1994), Hall and Yao (2003), and Jensen and Rahbek (2006). In contrast, there are only limited papers discussing the efficiency issues involved in estimating semiparametric GARCH models. The first attempt is due to Engel and Gonzalez-Rivera (1991), who showed partial success in achieving efficiency via Monte Carlo simulations. In their theoretical work, Linton (1993) and Drost and Klaassen (1997) explained that full adaptive estimation of θ is not possible and showed their efficient estimators for β via a reparamerization. Ling and McAleer (2003) further considers adaptive estimation in nonstationary ARMA-GARCH models.

We complement previous work on GARCH models by providing an efficient estimator for $F(\cdot)$ and thus the quantile of ε_t. It is well known that, in the absence of any auxilliary information about $F(\cdot)$, the empirical distribution function $\widehat{F}(x) = n^{-1}\sum_{t=1}^{n} 1(\varepsilon_t \le x)$ is semiparametrically efficient. However, $\widehat{F}(x)$ is no longer efficient when moment constraints (2) are available, see Bickel et al. (1993). The empirical likelihood (EL) weighted empirical distribution estimator is efficient with the existence of auxiliary information in the form of moments restrictions (2). The EL method was initiated by Owen (1990) and extended by Kitamura (1997) to time series. In i.i.d settings, Chen (1996) discovered second order improvement by empirical likelihood weighted kernel density estimation under moment restrictions. Zhao (1996) showed that there are variance gains by empirical likelihood weighted M-estimation when moment restrictions are available. Schick and Wefelmeyer (2002) provide an efficient estimator for the error distribution in nonlinear autoregressive models. However, the proposed estimator has the shortcoming that it is not a distribution itself. Müller et al. (2005) showed that the EL-weighted empirical distribution estimator is efficient in an autoregressive model. In this paper, we use EL weighted distribution estimator to construct estimates of VaR and ES in GARCH models. We show that, the resulting quantile and ES estimators for ε are efficient. Furthermore, the conditional VaR $\xi_t(\alpha)$ and ES estimators $\chi_t(\alpha)$ are asymptotically mixed-normal.

Various quantile estimators have been proposed recently, see Koenker, and Xiao (2009) and Chen, Koenker, and Xiao (2009). For fully nonparametric estimators, see Chen and Tang (2005) and Cai and Wang (2008). However, nonparametric estimators are subject to the curse of dimensionality and thus not widely applicable in practice. Furthermore, these nonparametric quantile estimators are too flexible to capture the stylized fact that financial returns are conditionally heteroskedastic. Given that this time-varying volatility is the key feature of financial time series, historical simulation method would be more advantageous than nonparametric methods in VaR forecasting. In our semiparametric model, the quantile estimator preserves the property of time-varying volatility and allows other aspect of conditional distribution unspecified. Model information is fully explored in the estimation so we gain by providing an efficient solution to conditional quantile estimation. Furthermore, the parametric filter (the GARCH model for volatility) bundle the conditioning set into a one-dimensional volatility so that there is no curse of dimensionality.

To the best of our knowledge, the only paper to address efficient conditional quantile estimation is Komunjer and Vuong (2010). However, their model is different from ours: they consider efficient conditional quantile estimation without moment constraints (2). Ai and Chen (2003) provide a very general framework for estimation and efficiency in semiparametric time series models defined through moment restrictions. No doubt some of our results can be replicated by their methodology using the sieve method.

We will discuss efficient estimation of θ in section 2 and efficient estimation of q_α in section 3. Once we collect efficient estimators for these parameters, we can construct the conditional quantile estimator $\xi_t(\alpha)$ and ES estimator $\chi_t(\alpha)$ and discuss their asymptotic distribution in section 4. We present our simulation results and empirical applications in section 5. Section 6 concludes with further extensions.

2 Efficient estimation of θ

Efficient estimation for semiparametric GARCH models was initially addressed by Engel and Gonzalez-Rivera (1991). Their Monte Carlo evidence showed that their estimation of GARCH parameters cannot fully capture the potential efficiency gain. Linton (1993) considered the ARCH(p) special case of (1) with no mean effect and assumed only that the errors were distributed symmetrically about zero. In that case, the error density is not jointly identified along with all the parameters, although the identified subvector is adaptively estimable. Drost and Klaassen (1997) consider a general case that allowed for different identification conditions. They showed that a subvector of the parameters can be adaptively estimated while a remaining parameter cannot be.

We rewrite the volatility model to reflect this. Specifically, now let $h_t = c^2 + ac^2y_{t-1}^2 + bh_{t-1}$. The finite dimensional parameter in this model $\theta = (c, a, b)^\top \in \Theta \subset \mathbb{R}^3$ is to be partitioned into two parts: $(c, \beta^\top)$ where $\beta = (a, b)^\top \in B$ for the reason that only β is adaptively estimable, see Linton (1993) and Drost and Klaassen (1997). As a result, we can rewrite the volatility as $h_t(\theta) = c^2 g_t(a, b)$, where $g_t(\beta) = 1 + au_{t-1}^2 + bg_{t-1}(\beta)$.

In the sequel we will use the following notations frequently: moment conditions $R_1(\varepsilon) = 1(\varepsilon \leq q_\alpha) - \alpha$, $R_2(\varepsilon) = (\varepsilon, \varepsilon^2 - 1)^\top$; the Fisher scale score $R_3(\varepsilon) = 1 + \varepsilon\frac{f'(\varepsilon)}{f(\varepsilon)}$ of the error density f; derivatives $G_t(\beta) = \partial \log g_t(\beta)/\partial\beta$, $G(\beta) = E[G_t(\beta)]$, $H_t(\theta) = \partial \log h_t(\theta)/\partial\theta$, $H(\theta) = E[H_t(\theta)]$,

$$G_2(\beta) = E\left[\frac{\partial \log g_t(\beta)}{\partial\beta}\frac{\partial \log g_t(\beta)}{\partial\beta^\top}\right] \quad ; \quad H_2(\theta) = E\left[\frac{\partial \log h_t(\theta)}{\partial\theta}\frac{\partial \log h_t(\theta)}{\partial\theta^\top}\right].$$

When the argument is evaluated at the true value, we use abbreviation: for example, $G = G(\beta_0)$ and $H_t = H_t(\theta_0)$.

The log-likelihood of observations $\{y_1, \ldots, y_n\}$ (given h_0) assuming that f is known is

$$L(\theta) = \sum_{t=1}^{n} \log f(c^{-1}g_t^{-1/2}(\beta)y_t) + \log c^{-1}g_t^{-1/2}(\beta).$$

Then the score function in the parametric model at time t as

$$l_t(\theta) = -\frac{1}{2}\left(1 + \varepsilon_t(\theta)\frac{f'(\varepsilon_t(\theta))}{f(\varepsilon_t(\theta))}\right)\frac{\partial \log h_t(\theta)}{\partial\theta}.$$

We now consider the semiparametric model where f is unknown. To see why the parameter θ is not adaptively estimable, we consider the density function $f(x; \eta)$ with a shape parameter $\eta, \eta \in \Upsilon$. It is clear from $E[\partial l_t(\theta; \eta)/\partial\eta] \neq 0$ that the estimation of η affects the efficiency of the estimates of θ. If we knew the density function $f(\cdot)$ and are interested in estimating β in presence of the nuisance parameter c, the efficient score function for β is the vector

$$l_{1t}^*(\beta) = -\frac{1}{2}\{G_t(\beta) - G(\beta)\}R_3(\varepsilon_t), \tag{3}$$

according to the Convolution Theorem 2.2 in Drost and Klaassen (1997). The density function $f(\cdot)$ is unknown. Drost and Klaassen (1997) showed that introduction of unknown $f(\cdot)$ in presence of unknown c does not change the efficient influence function for β.

We make the following assumptions:

Assumptions A

A1. $c > 0, a \geq 0$ and $b \geq 0$. $E[\ln\{b + ac^2\varepsilon_t^2\}] < 0$.

A2. The density function f satisfies the moment restrictions: $\int xf(x)dx = 0$ and $\int x^2 f(x)dx = 1$; it has finite fourth moment $\int x^4 f(x)dx < \infty$, and $E\varepsilon^4 - 1 - (E\varepsilon^3)^2 \neq 0$.

A3. The density function f is positive and f' is absolutely continuous with

$$||f||_\infty = \sup_{x\in\mathbb{R}} f(x) < \infty, \sup_{x\in\mathbb{R}} |x|f(x) < \infty, \int |x|f(x)dx < \infty.$$

A4. The density function f has positive and finite Fisher information for scale

$$0 < \int (1 + xf'(x)/f(x))^2 f(x)dx < \infty.$$

A5. The density function f for the initial value $h_1(\theta)$ satisfies that, the likelihood ratio for $h_1(\theta)$,

$$l_n(h_1) = \log\{f_{\widetilde{\theta}_n}/f_{\theta_n}(h_1(\theta)) \xrightarrow{P} 0, \quad as\ n \to \infty$$

where $\widetilde{\theta}_n$ and θ_n is defined similarly as in Drost and Klassen (1997)

Remark. Assumption A.1 ensures the positivity of h_t and the strict stationarity of y_t. Since $E[\ln\{b+ac^2\varepsilon_t^2\}] \leq b+ac^2-1$, a sufficient condition for strict stationarity is $b+ac^2 < 1$, see Nelson (1990). A.2 is introduced to make sure that the variance matrix $E[R_2(\varepsilon)R_2(\varepsilon)^\top]$ is invertible A.3 is made because we will need some boundedness of f to make a uniform expansion for the empirical distributions, see section 3. A.4 is typically assumed for efficiency discussion, see for example, Linton (1993) and Drost and Klaassen (1997). A.5 is assumed to obtain the uniform LAN theorem and the Convolution Theorem, as in Drost and Klaassen (1997).

We will suppose that there exists an initial $\sqrt{T}$-consistent estimator of all the parameters, for example the QMLE. The large sample property of GARCH parameters has been studied in different context. For example, Lee and Hansen (1994) and Berkes et. al. (2003) for detailed consistency discussion of Gaussian QMLE, and Weiss (1986) for OLS. Jensen and Rahbek (2004) considered the asymptotic theory of QMLE for nonstationary GARCH models. We have the following result which extends Drost and Klaasen (1997) and Drost, Klaassen, and Werker (1997).

Theorem 1. *Suppose that assumptions A hold. Then there exists an efficient estimator*

$\widehat{\theta}$ *that has the following expansion*

$$\sqrt{n}(\widehat{\theta}-\theta_0)=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\psi_t(\theta_0)+o_p(1), \tag{4}$$

$$\psi_t(\theta_0)=\begin{pmatrix} -\frac{1}{2}E[l_{1t}^{*}l_{1t}^{*\top}]^{-1}\{G_t-G\} & 0 \\ \frac{c_0}{4}G^{\top}E[l_{1t}^{*}l_{1t}^{*\top}]^{-1}\{G_t-G\} & \frac{c_0}{2}(-E\varepsilon^3,1) \end{pmatrix}\begin{pmatrix} R_3(\varepsilon_t) \\ R_2(\varepsilon_t)\end{pmatrix}.$$

Consequently,

$$\sqrt{n}(\widehat{\theta}-\theta_0)\Longrightarrow N(0,\Omega_\theta),$$

$$\Omega_\theta=\begin{pmatrix} E[l_{1t}^{*}l_{1t}^{*\top}]^{-1} & -\frac{c_0}{2}E[l_{1t}^{*}l_{1t}^{*\top}]^{-1}G \\ -\frac{c_0}{2}G^{\top}E[l_{1t}^{*}l_{1t}^{*\top}]^{-1} & \frac{c_0^2}{4}\{E\varepsilon^4-1-(E\varepsilon^3)^2+G^{\top}E[l_{1t}^{*}l_{1t}^{*\top}]^{-1}G\}\end{pmatrix}.$$

For technical reasons, the estimator employed in the theorem makes use of sample splitting, discretization, and trimming in order to facilitate the proof. In practice, none of these devices may be desirable. We have found the following estimator scheme works well in practice. Suppose that $k(\cdot)$ is a symmetric, second-order kernel function with $\int k(x)dx=1$ and $\int xk(x)dx=0$, and let h and b be bandwidths that satisfy $h\to 0, nh^4\to\infty, b\to 0, nb^4\to\infty$. We construct the efficient estimator for θ in three steps:

1. Let $\widehat{\theta}_1=(\widehat{\beta}_1^{\top},\widehat{c}_1)^{\top}$ be an initial $\sqrt{T}$-consistent estimator, for example the QMLE, and compute the residuals $\widehat{\varepsilon}_{1t}=y_t/h_t^{1/2}(\widehat{\theta}_1)$.

2. Update the estimator of β by using the Newton–Raphson method:

$$\widehat{\beta}=\widehat{\beta}_1+\left[\frac{1}{n}\sum\nolimits_{t=1}^{n}\widehat{l}_{1t}^{*}(\widehat{\beta}_1)\widehat{l}_{1t}^{*}(\widehat{\beta}_1)^{\top}\right]^{-1}\frac{1}{n}\sum\nolimits_{t=1}^{n}\widehat{l}_{1t}^{*}(\widehat{\beta}_1)$$

$$\widehat{l}_{1t}^{*}(\widehat{\beta}_1)=-\frac{1}{2}\left[G_t(\widehat{\beta}_1)-\frac{1}{n}\sum\nolimits_{s=1}^{n}G_s(\widehat{\beta}_1)\right]\widehat{R}_3(\widehat{\varepsilon}_{1t})$$

   with $\widehat{R}_3(x)=1+x\widehat{f}'(x)/\widehat{f}(x)$, $\widehat{f}(x)=n^{-1}h^{-1}\sum_{t=1}^{n}k(\frac{\widehat{\varepsilon}_{1t}-x}{h})$ and $\widehat{f}'(x)=-n^{-1}b^{-2}\sum_{t=1}^{n}k'(\frac{\widehat{\varepsilon}_{1t}-x}{b})$.

3. Denote $\widehat{e}_t=y_tg_t^{-1/2}(\widehat{\beta})$ and the efficient estimator for c is

$$\widehat{c}=\sqrt{\frac{1}{n}\sum\nolimits_{t=1}^{n}\widehat{e}_t^2-\frac{1}{n}\frac{\sum_{t=1}^{n}\widehat{e}_t^3}{\sum_{t=1}^{n}\widehat{e}_t^2}\sum\nolimits_{t=1}^{n}\widehat{e}_t}.$$

This procedure can be repeated until some convergence criterion is met, although for our theoretical purposes, one iteration is sufficient.

REMARK. It can be shown that the simpler estimator $\widetilde{c}=\sqrt{\frac{1}{n}\sum_{t=1}^{n}\widehat{e}_t^2}$ has an asymptotic variance $c_0^2\{E\varepsilon^4-1+G^{\top}E[l_{1t}^{*}l_{1t}^{*\top}]^{-1}G\}/4$, which is strictly larger than our efficient estimator

$\widehat{c}$ unless the error distribution is symmetric, i.e. $E\varepsilon^3 = 0$.

3 Efficient estimation of q_α and ES_α

We now turn to the estimation of the quantities of interest. To motivate our theory, we first discuss the estimation of q_α with the availability of true errors, and then discuss what to do in the case of estimation errors.

3.1 Quantile estimation with true errors available

In this subsection we estimate the quantile by inverting various distribution estimators. Because the unknown error distribution satisfies condition (2), it is desirable to construct distribution estimators that have this property.

The empirical distribution function $\widehat{F}(x) = n^{-1}\sum_{t=1}^n 1(\varepsilon_t \leq x)$ is commonly used but it does not impose these moment constraints. In practice, a common approach is to re-center the errors. Therefore, we also consider a modified empirical distribution, $\widehat{F}_N(x) = n^{-1}\sum_{t=1}^n 1((\varepsilon_t - \widehat{\mu}_\varepsilon)/\widehat{\sigma}_\varepsilon \leq x)$, where $\widehat{\mu}_\varepsilon = n^{-1}\sum_{t=1}^n \varepsilon_t$ and $\widehat{\sigma}_\varepsilon^2 = n^{-1}\sum_{t=1}^n \varepsilon_t^2 - (n^{-1}\sum_{t=1}^n \varepsilon_t)^2$. By construction, this distribution estimator satisfies the moment constraints (2). It is easy to see that the relationship between $\widehat{F}(x)$ and $\widehat{F}_N(x)$ is $\widehat{F}_N(x) = \widehat{F}(\widehat{\mu}_\varepsilon + x\widehat{\sigma}_\varepsilon)$.

In this paper, we consider a new weighted empirical distribution estimator $\widehat{F}_w(x) = \sum_{t=1}^n \widehat{w}_t 1(\varepsilon_t \leq x)$, where the empirical likelihood weights $\{\widehat{w}_t\}$ come from the following:

$$\max_{\{w_t\}} \Pi_{t=1}^n w_t$$
$$\text{s.t.} \sum\nolimits_{t=1}^n w_t = 1; \sum\nolimits_{t=1}^n w_t\varepsilon_t = 0; \sum\nolimits_{t=1}^n w_t(\varepsilon_t^2 - 1) = 0.$$

By construction, $\widehat{F}_w$ satisfies the moment restrictions.

In the absence of the moment constraints, it is easy to see that $\arg\max_{\{w_t\}}\{\Pi_{t=1}^n w_t + \lambda(1 - \sum_{t=1}^n w_t)\} = 1/n$. In this case our weighted empirical distribution estimator is the same as $\widehat{F}(x)$. Since the unknown distribution is in the family $\mathcal{P} = \{f(x) : \int xf(x)dx = 0, \int(x^2-1)f(x)dx = 0\}$, we expect $\widehat{F}_w(x)$ to be more efficient by incorporating these moment constraints, Bickel, Klaassen, Ritov, and Wellner (1993). Lemma 1 (appendix) which shows the uniform expansion for the distribution estimators $\widehat{F}(x)$, $\widehat{F}_N(x)$ and $\widehat{F}_w$ confirms our conjecture.

Remark. It is well-known that $\sqrt{n}(\widehat{F}(x) - F(x)) \Longrightarrow N(0, F(x)(1 - F(x)))$. The empirical distribution is the most efficient estimator without any auxiliary information about $F(\cdot)$. This is consistent with our result because $w_t = 1/n$ is the solution to the problem of $\max_{\{w_t\}}\{\Pi_{t=1}^n w_t + \lambda(1 - \sum_{t=1}^n w_t)\}$.

Remark. It can be seen from Lemma 1 that

$$\begin{aligned}\sqrt{n}(\widehat{F}_N(x) - F(x)) &\implies N(0, F(x)(1-F(x)) + C_x) \\ \sqrt{n}(\widehat{F}_w(x) - F(x)) &\implies N(0, F(x)(1-F(x)) - A_x^\top B^{-1} A_x).\end{aligned}$$

We can see that normalization introduced estimation error; see Durbin (1973). This estimation error has been cumulated and is reflected by the additional term C_x in the asymptotic variance. The sign of C_x function is indeterminate, see the Figure 1 in the appendix. It depends on the density $f(x)$ and point to be evaluated. For standard normal distribution and student distributions, $C_x \leq 0$, which means, for these two distributions, $\widehat{F}_N(x)$ is more efficient than $\widehat{F}(x)$. In contrast, for mixed normal distribution and Chi-squared distributions, the efficiency ranking depends on the point to be evaluated. On the other hand, weighting the empirical distribution takes into account the information in (2), which is reflected in the term $-A_x^\top B^{-1} A_x$. This term can be explained as the projection of $1(\varepsilon \leq x) - F(x)$ onto $R_2(\varepsilon)$. The covariance A_x measures the relevance of moment constraints (2) in estimating distribution function. The information content that helps in estimating unknown $F(x)$ is weak when A_x is small. In case of $A_x = 0$, the moment constraints (2) do not have any explanation power at all since $1(\varepsilon \leq x) - F(x)$ and $R_2(\varepsilon)$ is orthogonal. In the appendix we give conditions under which $\widehat{F}_N(x)$ and $\widehat{F}(x)$ can be as efficient as $\widehat{F}_w(x)$.

We now define our quantile and expected shortfall estimators. For an estimated c.d.f., $\widetilde{F}$, let

$$\widetilde{q}_\alpha = \sup\{t : \widetilde{F}(t) \leq \alpha\} = \widetilde{F}^{-1}(\alpha) \quad ; \quad \widetilde{ES}_\alpha = \frac{1}{\alpha}\int_{-\infty}^{\widetilde{q}_\alpha} x d\widetilde{F}(x). \tag{5}$$

See Scaillet (2004) and Chen (2008) for estimators of expected shortfall in a different setting. Let $\widehat{q}_\alpha, \widehat{q}_{N\alpha}, \widehat{q}_{w\alpha}, \widehat{ES}_\alpha, \widehat{ES}_{N\alpha}$, and $\widehat{ES}_{w\alpha}$ be defined from (5) using the $\widehat{F}(x)$, $\widehat{F}_N(x)$, and $\widehat{F}_w(x)$ as required. The next theorem presents the asymptotic distribution of these quantile estimators. Define:

$$\begin{aligned}V_1 &= \frac{\alpha(1-\alpha)}{f(q_\alpha)^2} \quad ; \quad V_2 = \frac{\alpha(1-\alpha)}{f(q_\alpha)^2} + \frac{C_{q_\alpha}}{f(q_\alpha)^2} \quad ; \quad V_3 = \frac{\alpha(1-\alpha)}{f(q_\alpha)^2} - \frac{A_{q_\alpha}^\top B^{-1} A_{q_\alpha}}{f(q_\alpha)^2} \\ V_4 &= \alpha^{-2}\text{var}((\varepsilon - q_\alpha)1(\varepsilon \leq q_\alpha)) \quad ; \quad V_5 = \alpha^{-2}\text{var}((\varepsilon - q_\alpha)1(\varepsilon \leq q_\alpha) - \alpha\varepsilon - \frac{\varepsilon^2}{2}\int_{-\infty}^{q_\alpha} x f(x) dx)) \\ V_6 &= \alpha^{-2}\text{var}((\varepsilon - q_\alpha)1(\varepsilon \leq q_\alpha) + R_2^\top(\varepsilon) B^{-1}\int_{-\infty}^{q_\alpha} A_x dx).\end{aligned}$$

Theorem 2. *Suppose that assumptions A.1-A.5 hold. The quantile and expected short-*

fall estimators are asymptotically normal:

$$\begin{aligned}
\sqrt{n}(\widehat{q}_\alpha - q_\alpha) &\Longrightarrow N(0, V_1) \quad ; \quad \sqrt{n}(\widehat{ES}_\alpha - ES_\alpha) \Longrightarrow N(0, V_4) \\
\sqrt{n}(\widehat{q}_{N\alpha} - q_\alpha) &\Longrightarrow N(0, V_2) \quad ; \quad \sqrt{n}(\widehat{ES}_{N\alpha} - ES_\alpha) \Longrightarrow N(0, V_5) \\
\sqrt{n}(\widehat{q}_{w\alpha} - q_\alpha) &\Longrightarrow N(0, V_3) \quad ; \quad \sqrt{n}(\widehat{ES}_{w\alpha} - ES_\alpha) \Longrightarrow N(0, V_6) \,.
\end{aligned}$$

Remark. It is clear from the comparison of asymptotic variances that $\widehat{q}_{w\alpha}$, which is based on inverting empirically weighted distribution estimators, is the most efficient one. The same conclusion holds for ES since ES is the aggregation of lower quantiles: $\widetilde{ES}_\alpha = \frac{1}{\alpha}\int_{-\infty}^{\widetilde{q}_\alpha} x d\widetilde{F}(x) = \frac{1}{\alpha}\int_0^\alpha \widetilde{q}_\alpha d\alpha$.

Remark. For improvement in mean squared efficiency, one could consider inverting the smoothed weighted empirical distribution $\widehat{F}_{sw}(x) = \sum_{t=1}^n \widehat{w}_t K(\frac{x-\varepsilon_t}{h})$ with $\widehat{F}_s(x) = n^{-1}\sum_{t=1}^n K(\frac{x-\varepsilon_t}{h})$ being a special case. However, the large sample property will be the same as the unsmoothed one here. The unsmoothed distribution estimators considered in this paper are free from the complication of bandwidth choice.

3.2 Quantile estimation with estimated parameters

We now assume that we don't know the true parameters θ, we don't observe ε_t. Instead we observe the polluted error, $\varepsilon_t(\theta_n) = y_t/h_t^{1/2}(\theta_n)$, where θ_n is an estimator sequence satisfying $\theta_n - \theta_0 = O_p(n^{-1/2})$. Now we construct an efficient estimator for residual distribution $F(x)$ and then invert to get back the quantile estimator $q_{n\alpha} = F_n^{-1}(\alpha)$. We treat a general class of estimators θ_n for completeness.

Motivated by the efficiency gain shown in Lemma 1, we estimate the quantile by inverting the following distribution function estimator:

$$\widehat{\widehat{F}}_w(x) = \sum_{t=1}^n \widehat{\widehat{w}}_t 1(\varepsilon_t(\theta_n) \leq x), \tag{6}$$

where $\{\widehat{\widehat{w}}_t\}$ are defined by the solution of the following optimization problem

$$\begin{aligned}
&\max_{\{w_t\}} \Pi_{t=1}^n w_t \\
\text{s.t. } \sum\nolimits_{t=1}^n w_t &= 1; \sum\nolimits_{t=1}^n w_t \varepsilon_t(\theta_n) = 0; \sum\nolimits_{t=1}^n w_t(\varepsilon_t^2(\theta_n) - 1) = 0.
\end{aligned}$$

For comparison purposes, we also consider the residual empirical distribution estimator $\widehat{\widehat{F}}(x) = n^{-1}\sum_{t=1}^n 1(\varepsilon_t(\theta_n) \leq x)$ and the standardized empirical distribution $\widehat{\widehat{F}}_N(x) = n^{-1}\sum_{t=1}^n 1((\varepsilon_t(\theta_n) - \widehat{\widehat{\mu}}_\varepsilon)/\widehat{\widehat{\sigma}}_\varepsilon \leq x)$, where $\widehat{\widehat{\mu}}_\varepsilon = n^{-1}\sum_{t=1}^n \varepsilon_t(\theta_n)$ and $\widehat{\widehat{\sigma}}_\varepsilon^2 = n^{-1}\sum_{t=1}^n \varepsilon_t^2(\theta_n) - (n^{-1}\sum_{t=1}^n \varepsilon_t(\theta_n))^2$.

Suppose that there is an estimator $\widetilde{\theta}$ that has influence function $\chi_t(\theta_0)$, i.e.

$$\sqrt{n}(\widetilde{\theta}-\theta_0)=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\chi_t(\theta_0)+o_p(1). \tag{7}$$

Then Lemma 2 (appendix) shows the uniform expansion of the distribution estimators $\widehat{\widehat{F}}(x), \widehat{\widehat{F}}_N(x)$ and $\widehat{\widehat{F}}_w(x)$ based on $\widetilde{\theta}$ and Corollary 3 are the covariance functions of the process.

We next explore the magnitude of the correction terms in some examples. Suppose that

$$\chi_t(\theta_0)=J_t(\theta_0)(\varepsilon_t^2-1), \tag{8}$$

where $J_t(\theta_0)\in\mathcal{F}_{t-1}$, so that $\chi_t(\theta_0)$ is a martingale difference sequence. Denote $J(\theta_0)=E[J_t(\theta_0)]$. Then the asymptotic variances of the three distribution estimators are

$$\begin{aligned}
\Omega_{1,J}(x) &= F(x)(1-F(x))+\frac{[E[\varepsilon^4]-1]x^2f(x)^2}{4}\{H(\theta_0)^\top J(\theta_0)\}^2+xf(x)H(\theta_0)^\top J(\theta_0)a_2(x)\\
\Omega_{2,J}(x) &= F(x)(1-F(x))+\frac{[E[\varepsilon^4]-1]x^2f(x)^2}{4}\{H(\theta_0)^\top J(\theta_0)\}^2+xf(x)H(\theta_0)^\top J(\theta_0)a_2(x)\\
&\quad +f(x)^2+\frac{x^2f(x)^2[E[\varepsilon^4]-1]}{4}+xf(x)^2E[\varepsilon^3]+xf(x)a_2(x)+2f(x)a_1(x)\\
&\quad +xf(x)^2E[\varepsilon^3]H(\theta_0)^\top J(\theta_0)+\frac{x^2f(x)^2[E[\varepsilon^4]-1]}{2}H(\theta_0)^\top J(\theta_0)\\
\Omega_{3,J}(x) &= F(x)(1-F(x))-A_x^\top B^{-1}A_x+\{E[\varepsilon^4]-1\}\left\{\frac{xf(x)}{2}+\begin{pmatrix}0 & 1\end{pmatrix}B^{-1}A_x\right\}^2\{H(\theta_0)^\top J(\theta_0)\}^2.
\end{aligned}$$

In the special case of the least squares estimator,

$$\chi_t(\theta_0)=H_1(\theta_0)^{-1}h_t(\theta_0)\frac{\partial h_t(\theta_0)}{\partial\theta}(\varepsilon_t^2-1),$$

where $H_1(\theta_0)=E[\frac{\partial h_t(\theta_0)}{\partial\theta}\frac{\partial h_t(\theta_0)}{\partial\theta^\top}]$. Denote $H_3(\theta_0)=E[h_t(\theta_0)\frac{\partial h_t(\theta_0)}{\partial\theta}]$, then $J_t(\theta_0)=H_1(\theta_0)^{-1}H_{3t}(\theta_0)$. In the special case of the Gaussian QMLE,

$$\chi_t(\theta_0)=\{H_2(\theta_0)\}^{-1}\frac{1}{h_t(\theta_0)}\frac{\partial h_t(\theta_0)}{\partial\theta}(\varepsilon_t^2-1),$$

then $J_t(\theta_0)=H_2(\theta_0)^{-1}H_t(\theta_0)$. In both cases the asymptotic variance is increased relative to Lemma 1. Since the QMLE residuals $\varepsilon_t(\widetilde{\theta})$ are obtained under the moment condition $n^{-1}\sum_{t=1}^{n}[H_t(\widetilde{\theta})(\varepsilon_t^2(\widetilde{\theta})-1)]=0$ with probability one, the first moment of $\widehat{\widehat{F}}(x)$ is $\int xd\widehat{\widehat{F}}(x)=n^{-1}\sum_{t=1}^{n}(\varepsilon_t^2(\widetilde{\theta})-1)$, which may not be zero with probability one.

Based on the asymptotic expansion of distribution estimators in Lemma 2, we construct quantile estimators by inverting these distribution estimators. We next give the main result of the paper. Let $\widehat{\widehat{q}}_\alpha, \widehat{\widehat{q}}_{N\alpha}, \widehat{\widehat{q}}_{w\alpha}, \widehat{\widehat{ES}}_\alpha, \widehat{\widehat{ES}}_{N\alpha}$, and $\widehat{\widehat{ES}}_{w\alpha}$ be defined from (5) using the estimated c.d.f.s $\widehat{\widehat{F}}(x)$, $\widehat{\widehat{F}}_N(x)$, and $\widehat{\widehat{F}}_w(x)$ as required. Define the asymptotic covariance matrices:

$$
\begin{aligned}
\Omega_\alpha &= \frac{\alpha(1-\alpha)}{f(q_\alpha)^2} + \frac{q_\alpha^2}{4}[E\varepsilon^4 - 1 - (E\varepsilon^3)^2] + \frac{q_\alpha(a_{2q_\alpha} - a_{1q_\alpha}E\varepsilon^3)}{f(q_\alpha)} \\
\Omega_{N\alpha} &= \frac{\alpha(1-\alpha)}{f(q_\alpha)^2} + \frac{C_{q_\alpha}}{f(q_\alpha)^2} + \frac{3q_\alpha^2[E\varepsilon^4 - 1 - (E\varepsilon^3)^2]}{4} + \frac{q_\alpha(a_{2q_\alpha} - a_{1q_\alpha}E\varepsilon^3)}{f(q_\alpha)} \\
\Omega_{w\alpha} &= \frac{\alpha(1-\alpha)}{f(q_\alpha)^2} - \frac{A_{q_\alpha}^\top B^{-1}A_{q_\alpha}}{f(q_\alpha)^2} + [\frac{q_\alpha}{2} + \frac{\begin{pmatrix} 0 & 1 \end{pmatrix} B^{-1}A_{q_\alpha}}{f(q_\alpha)}]^2[E\varepsilon^4 - 1 - (E\varepsilon^3)^2] \\
\Omega_{ES} &= \alpha^{-2}\mathrm{var}\left((\varepsilon - q_\alpha)1(\varepsilon \le q_\alpha) - \frac{\varepsilon^2 - \varepsilon E\varepsilon^3}{2}\int_{-\infty}^{q_\alpha} xf(x)dx\right) \\
\Omega_{ESN} &= \alpha^{-2}\mathrm{var}\left((\varepsilon - q_\alpha)1(\varepsilon \le q_\alpha) - \varepsilon\int_{-\infty}^{q_\alpha}[f(x) - \frac{xf(x)}{2}E\varepsilon^3]dx - \varepsilon^2\int_{-\infty}^{q_\alpha} xf(x)dx\right) \\
\Omega_{ESW} &= \alpha^{-2}\mathrm{var}\left((\varepsilon - q_\alpha)1(\varepsilon \le q_\alpha) - (\varepsilon^2 - \varepsilon E\varepsilon^3)\int_{-\infty}^{q_\alpha}[\frac{xf(x)}{2} + \begin{pmatrix} 0 & 1 \end{pmatrix} B^{-1}A_x]dx\right. \\
&\quad \left. + R_2^\top(\varepsilon)B^{-1}\int_{-\infty}^{q_\alpha} A_x dx\right).
\end{aligned}
$$

THEOREM 3. *Suppose assumptions A.1-A.5 hold. The quantile and expected shortfall estimators are asymptotically normal:*

$$
\begin{aligned}
\sqrt{n}(\widehat{\widehat{q}}_\alpha - q_\alpha) &\Longrightarrow N(0,\Omega_\alpha) \quad ; \quad \sqrt{n}(\widehat{\widehat{ES}}_\alpha - ES_\alpha) \Longrightarrow N(0,\Omega_{ES}) \\
\sqrt{n}(\widehat{\widehat{q}}_{N\alpha} - q_\alpha) &\Longrightarrow N(0,\Omega_{N\alpha}) \quad ; \quad \sqrt{n}(\widehat{\widehat{ES}}_{N\alpha} - ES_\alpha) \Longrightarrow N(0,\Omega_{ESN}) \\
\sqrt{n}(\widehat{\widehat{q}}_{w\alpha} - q_\alpha) &\Longrightarrow N(0,\Omega_{w\alpha}) \quad ; \quad \sqrt{n}(\widehat{\widehat{ES}}_{w\alpha} - ES_\alpha) \Longrightarrow N(0,\Omega_{ESW}).
\end{aligned}
$$

REMARK. For the same reason as above, we can see that $\widehat{\widehat{q}}_{w\alpha}$ is more efficient than $\widehat{\widehat{q}}_\alpha$. The same conclusion holds for ES.

REMARK. Notice that the asymptotic variances of VaRs and ESs do not contain any functional form of the heteroskedasticity. This is due to the orthogonality in information between estimators for the distribution $F(x)$ and variance estimator for β.

REMARK. With this large sample property available, we can compute standard errors by the obvious plug-in method.

4 Efficient estimation of conditional VaR and conditional expected shortfall

We have discussed the asymptotic property of efficient estimators $\widehat{\theta}$ and $\widehat{\widehat{q}}_{w\alpha}$. They are shown to be the best among competitors in terms of smallest asymptotic variances. Both are important ingredients to the conditional quantile estimator $\widehat{\xi}_{n,t}$ as $\widehat{\xi}_{n,t} = h_t^{1/2}(\widehat{\theta})\widehat{\widehat{q}}_{w\alpha}$ and the conditional expected shortfall $\widehat{\chi}_{n,t} = h_t^{1/2}(\widehat{\theta})\widehat{\widehat{ES}}_{w\alpha}$. In this section, we will show that these two quantities are asymptotically mixed normal. Define:

$$\omega_{\xi t} = h_t(\theta_0)\left\{\frac{q_\alpha^2}{4}(G_t^\top - G)E[l_{1t}^* l_{1t}^{*\top}]^{-1}(G_t - G) + \Omega_{w\alpha}\right\},$$

$$\omega_{\chi t} = h_t(\theta_0)\left\{\frac{ES_\alpha^2}{4}H_t^\top \Omega_\theta H_t + ES_\alpha \frac{(-E\varepsilon^3, 1)E\left[\left\{(\varepsilon_t - q_\alpha)1(\varepsilon_t \le q_\alpha) + R_2^\top(\varepsilon_t)C\right\}R_2(\varepsilon_t)\right]}{\alpha} + \Omega_{ESW}\right\}$$

$$C = \int_{-\infty}^{q_\alpha}[\frac{xf(x)}{2} + \begin{pmatrix} 0 & 1 \end{pmatrix} B^{-1}A_x]dx(E\varepsilon^3, -1) + \int_{-\infty}^{q_\alpha} A_x^\top dx B^{-1}.$$

Theorem 4. *Suppose assumptions A.1-A.5 hold. The conditional quantile estimator* $\widehat{\xi}_{n,t}$ and *conditional quantile estimator* $\widehat{\chi}_{n,t}$ *are asymptotically mixed normal*

$$\begin{aligned}\sqrt{n}(\widehat{\xi}_{n,t} - \xi_t) &\Longrightarrow MN(0, \omega_{\xi t}) \\ \sqrt{n}(\widehat{\chi}_{n,t} - \chi_t) &\Longrightarrow MN(0, \omega_{\chi t}),\end{aligned}$$

where the random positive scalars $\omega_{\xi t}$ *and* $\omega_{\chi t}$ *are independent of the underlying normals.*

Remark. From the influence functions of $(\widehat{a}, \widehat{b})$ and $\widehat{\widehat{q}}_{w\alpha}$, we can see that they are asymptotically orthogonal: this is anticipated as the parameter (a, b) is adaptively estimated with respect to the error distribution.

Remark. This mixed normal distribution asymptotics is parallel to results obtained in Barndorff-Nielsen and Shephard (2002) for estimation of the quadratic variation of Brownian semimartingales, see also Hall and Heyde (1980). It follows that $\sqrt{n}(\widehat{\xi}_{n,t} - \xi_t)/\omega_{\xi t}^{1/2} \Longrightarrow N(0,1)$ and $\sqrt{n}(\widehat{\xi}_{n,t} - \xi_t)/\widehat{\omega}_{\xi t}^{1/2} \Longrightarrow N(0,1)$, where $\widehat{\omega}_{\xi t}$ is a consistent estimator of $\omega_{\xi t}$. Therefore, one can conduct inference about $\xi_{n,t}$ with the usual confidence intervals.

5 Numerical Work

In this section we present some numerical evidence. The first part is Monte-Carlo simulation and the second is an empirical application.

5.1 Simulations

We follow Drost and Klaassen (1997) to simulate several GARCH (1,1) series from the model (1) with the following parameterizations:

1. $(c, a, b) \in \{(1, 0.3, 0.6), (1, 0.1, 0.8), (1, 0.05, 0.9)\}$;
2. $f(x) \in \{N(0,1), MN(2,-2), L, t(5), t(7), t(9), \chi^2_6, \chi^2_{12}\}$, which are, respectively, referred to the densities of standardized (mean 0 and variance 1) distributions from Normal, Mixed Normal with means $(2, -2)$, Laplace, student distributions with degree of freedom 5, 7 and 9 and chi-squared distribution with 6 and 12 degrees of freedom.

Sample size is set to $n = 500, 1000$. Simulations are carried out 2500 times to take averages. We consider the performance of the three distribution estimators and their associated quantile and ES estimators with α being 5% and 1%. We also have the simulation results for small samples $n = 25, 50, 100$, and for IGARCH models with $a + b = 1$. These results are similar to those in this paper and are available upon request.

The criterion for distribution estimator $\widehat{F}(x)$ is the integrated mean squared error (IMSE)

$$IMSE = \int E[\widehat{F}(x) - F(x)]^2 dx$$

and that for quantile and ES estimators ($\widehat{q}_\alpha$ and $\widehat{ES}_\alpha$) is the mean squared error

$$MSE = E[(\widehat{q}_\alpha - q_\alpha)^2]; MSE = E[(\widehat{ES}_\alpha - ES_\alpha)^2].$$

First, we consider the case where the true errors are available. The IMSEs of three distribution function estimators are summarized in Table 1. It is clear form this table that the weighted empirical distribution estimator $\widehat{F}_w(x)$ performs the best in all cases. The relative efficiency of $\widehat{F}_w(x)$ to the unweighted empirical distribution $\widehat{F}(x)$ is very large: it ranged from 50% in case of errors being Laplacian to 72% in the case of Mixed-normal. Figure 2 visualizes this gain by plotting the overlays of simulated distribution estimators with 100 replications. The colored region represents the possible paths of function estimators and it is clear that the magenta area (realizations of $\widehat{F}_w(x)$) is has the smaller width than blue area (realizations of $\widehat{F}(x)$). In order to compare the quantile estimators based on inverting these distribution estimators, we compute their average biases and mean squared errors under different distributional assumptions and in 500 and 1000 sample sizes. The average is taken over 2500 simulations. It is found that $\widehat{q}_{w\alpha}$ performs much better than $\widehat{q}_\alpha$ in all cases. This improvement is clearer in the case of $\alpha = 0.05$ than the case of $\alpha = 0.01$. This is because the further to the tail (when α is smaller), the smaller the covariance between $R_2(\varepsilon)$ and $1(\varepsilon \leq x) - \alpha$.

Next, we compare the distribution estimators when the errors are not observable and we use estimated errors from QMLE. Since QMLE is consistent in all above error distribution assumptions, we expect the QMLE residuals will behave close to the true errors, although with some estimation noises. Table 4-6 list the IMSE for distribution estimators under three different parameterizations. We find that, there are efficiency gains by weighting the empirical distribution estimator with empirical likelihoods. Figure 3 visualizes these gains, which vary across the assumptions of true error distributions. Table 7-12 compare the performance of residual quantile estimators. The conclusion is the same: empirical likelihood weightings reduces the variation of quantile estimators. However, these reductions are not of the same magnitude as in i.i.d case. The reason is because we use estimated errors in stead of true errors and the added estimation noise affect the performance of residual based estimators.

Thirdly, we compare different estimators for expected shortfall in the case of iid errors and GARCH residuals. As seen from table 15-18, the same conclusion holds for ES. For sample size n=500 and 1000, the proposed estimator does not do very well in the case of $\alpha = 0.01$, see table 18. This is expected because our efficient estimator (EL-weighted) involves an additional layer of numerical optimization, and for such low quantile/ES, the effective sample size is $n/100$. Therefore we tabulate the results for large sample $n = 10000$,which is the table 19(c). It's clear from table that our proposed VaR and ES estimators outperform other estimators in terms of smaller MSE.(The comparison of the estimators for $\widehat{q}_{0.01}$ and $\widehat{ES}_{0.01}$,when the true errors are available and $\widehat{\widehat{q}}_{0.01}$ and $\widehat{\widehat{ES}}_{0.01}$, when the polluted errors are calculated are provided in table 19(a) and 19(b)).

Finally, we consider the case of distribution and quantile estimation based on efficient residuals: the estimated errors are residuals from efficient estimation of parameter θ_0. As we notice that the performance of these estimators does not change much under different parameterization of θ_0, we only report the results in the case of $c = 1, a = 0.05, b = 0.9$. Table 13 summarizes the performance of quantile estimators for $q_{0.01}$ and $q_{0.05}$, while table 14 reports the true VaR and ES for distribution estimators and Figure 4 visualize the efficiency gains.

5.2 Empirical Work

The data used in our empirical study is the total market value of S&P 500 index, from 3rd Jan, 2001 to 31th Dec, 2010, which is available in the CRSP database. First of all, we examine some properties of the S&P 500 financial returns (Figure 4). And the summary of the statistics of the daily returns are provided in Table 20.

Figure 5(a) shows the ACF of the data with 95% confidence interval, most of the autocorrelations are within the range. While Figure 5(b) is the autocorrelation plot of squared returns, which has a strong evidence of the predictability of volatility. The kurtosis of the data is 10.35, which has a pretty strong evidence against the normality of the tail distribution. The formal procedure to test the normaility assumption of the tail has been done by Jarque-Bera test and Kolmogorov-Smirnov test and a graphic method is the QQ plot (Figure 6).

Table 21 shows the different threshold values of the conditional VaR and conditional ES of the three different models in our paper (GARCH-EL, GARCH-EL(recenter) and GARCH-ELW).

The main purpose of the empirical study is to see how our model performs in forecasting risk. This can be done by backtesting various VaR models. Backtesting evaluates VaR forecasts by checking how a VaR forecast model perform over a certain period of time. The number of the observations that are used to forecast the risk is called the estimation window, W_E and the data sample over which risk is forecast is called testing window, W_T. In our empirical study, we choose 1000 observations as our estimation window. (Figure 9,10 and 11). We later use a technique called violation ration (VR) to judge the quality of the VaR forecasts. If the actual return on a particular day exceeds the VaR forecast, we said the VaR limit is being violated. The VR is defined by the observed number of violation over the expected number of violation. If the VaR forecast of our model is accurate, the violation ration is expected to be equal to 1. A useful rule of thumb is that if the VR is between 0.8 and 1.2, the model is considered to be a good forecast. However, if the VR<0.5 or VR>1.5, the model is imprecise.

In the empirical analysis, we implement seven models: moving average (MA), historical simulation (HS), exponential moving average (EWMA), GARCH(1,1), GARCH-empirical distribution (GARCH-EL), GARCH-recenter empirical distribution (GARCH-ELR) and GARCH-weighted empirical distribution (GARCH-ELW). We first analyze the different VaR forecasting techniques by graphical methods.

The comparison results of the five VaR forecast models (MA, HS, EWMA,GARCH(1,1) and GARCH-ELW) is showed in Figure 9(a) while the violation ratios and conditional volatilities of these models are provided in Table 22. From Figure 9(a), we can see that HS and MS, the two methods that apply equal weight to historical data performs quite bad, while the other three conditional methods (EWMA, GARCH(1,1) and GARCH-ELW)are better and our method (GARCH-ELW) is the best. This result can also be asserted from Table 22 that our model (GARCH-ELW) performs much better that the others, providing that the violation ratio is 1.189 in our model (EWMA (2.2084) and GARCH(1,1)(2.3216)). However, there are not much difference between the three models mentioned in our paper (GARCH-EL, GARCH-ELR and GARCH-ELW), see Table 24.

From Figure 9(a), we can see that the volatility is relatively stable before 2007, and then the volatility steadily increase until it hits the biggest fluctuations starting from the end of 2008. After that, it drops sharply, but still much higher than the previous tranquil period. As we known, the largest volatility clustering happens during the period of the bankruptcy of Lehman Brothers and failure of AIG. To explore this in more detail, Figure 9(b) focus on such short period only. After the main crisis, the large volatility clusters again at the end of 2009 and also in the middle of 2010, which coincide with Greek government debt crisis and the announcement of the first bailout plan.

We graphically analyze the ES forecasting by using the same method. Figure 11(a) is the comparison of the five ES models (MA, HS, EWMA,GARCH(1,1) and GARCH-ELW) and Figure 11(b) only focus on the period of the global financial crisis too. As above, the HS and MA methods perform abysmally, and the other three conditional methods are much better. Our model (GARCH-ELW) is the best among them, which is similar as that of VaR models.

In addition, some formal statistical tests are helped to identify if a VR that is different from 1 is statistically significant. One of these is the Bernoulli coverage test, which ensures the theoretical confidence level p matches the empirical probability of violation. We denote a violation occurred on day t by δ_t, which takes value 1 or 0. 1 indicates a violation and 0 indicates no violation. The null hypothesis for VaR violations is

$$H_0 : \delta_t \ \sim \ Bernoulli(p).$$

Another test is called Independence of violation test. We are also interested in testing whether violations cluster. From the theoretical point of view, the violations, which indicates a sequence of losses, should spread out over time. The null hypothesis for this test is that there is no clustering. This test is important given the stylized evidence of volatility clustering.

The result for the statistical test for the significant of backtests is reported in Table 23. The violation ratio of 1 is strongly rejected for all the models except ours. And the independence test is also rejected for MA at 5%. The test result indicates that the VaR violation in our model is distributed by Bernoulli distribution and there is no violation clustering, given the p-value of the two tests are high separately, 0.4380 and 0.4770. (Comparison of the Bernoulli Coverage test & Independence test of model GARCH-EL, GARCH-ELR and GARCH-ELW are also provided in Table 25).

The significance of the backtest ES can be verified by a methodology called average normalized shortfall, the test procedure is sketched below. For days when VaR is violated, normalized shortfall (NS) is calculated as

$$NS_t = \frac{y_t}{ES_t},$$

where ES_t is the observed ES on day t. From the definition of ES, the expected y_t— given VaR is violated — is:

$$\frac{E(y_t \mid y_t < -VaR_t)}{ES_t} = 1.$$

The Null hypothesis: average NS ($\overline{NS}$)= 1. The test result is reported in Table 27. The average NS in our model is 0.9895, which is the nearest to 1.

6 Conclusion and Extension

This paper proposes and investigates new efficient conditional VaR and ES estimators in a semiparametric GARCH model. These proposed estimators for risk measures fully exploit the moment information which has been previously ignored in constructing innovation distribution estimators. We show they can achieve large efficiency improvement and quantify this magnitude in Monte Carlo simulations. At the same time, we present the asymptotic theory for one period ahead VaR and ES forecasts. The theory can be used as guidance as to constructing confidence intervals for point risk measure forecasts.

Even though we consider a simple GARCH(1,1) model in this paper, the efficient estimation method for both variance parameters and error quantile can be used for more complicated parametric volatility models. For example, one could consider GARCH with leverage effects or GARCH in mean models. Although the efficiency gain hinges on the efficiency of volatility estimators in theory, our MonteCarlo experiments show that this impact on efficiency improvement is quantitatively small.

Sometimes unconditional Value-at-Risk is also of interest to risk managers. Then the question in the current GARCH(1,1) context is whether we have efficiency gains from integrating the conditional VaR versus unconditional. This question is to be addressed in a separate paper.

References

[1] Ai, C. and X. Chen, (2003). "Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions," Econometrica, Econometric Society, vol. 71(6), pages 1795-1843

[2] Artzner, P.F., Delbaenm J, M. Eber, and D. Heath (1999). Coherent measures of risk. Mathematical Finance 9, 203-228.

[3] Barndorff-Nielsen, O. E. and N. Shephard (2002). Econometric analysis of realised volatility and its use in estimating stochastic volatility models. Journal of the Royal Statistical Society, Series B 64, 253-280.

[4] Basel Committee on Banking Supervision (2012). http://www.bis.org/publ/bcbs219.pdf

[5] Berkes, I., Horvăth, L. and Kokoszka, P. (2003). GARCH processes: structure and estimation. Bernoulli. 9(2), 201-227.

[6] Bickel, P. J., Klaassen, C. A. J., Ritov, Y., and Wellner, J. A. (1993). Efficient and Adaptive Estimation for Semiparametric Models. Johns Hopkins University Press, Baltimore. Reprinted (1998), Springer, New York.

[7] Cai and Wang (2008). Nonparametric methods for estimating conditional value-at-risk and expected shortfall. Journal of Econometrics, Volume 147, Issue 1, 120-130

[8] Chan, Deng, Peng and Xia (2007), Interval estimation of value-at-risk based on GARCH models with heavy-tailed innovations. Journal of Econometrics, vol. 137, issue 2, pages 556-576

[9] Chen, S.X. (1996) Empirical likelihood confidence intervals for nonparametric density estimation. Biometrika 83 , 329-341.

[10] Chen, S.X. (2008). Nonparametric estimation of expected shortfall. Journal of Financial Econometrics 1, 87-107.

[11] Chen, S.X. and C. Y. Tang (2005) Nonparametric Inference of Value at Risk for dependent Financial Returns, Journal of Financial Econometircs, 3, 227-255.

[12] Chen, X and Y. Fan (2006): Estimation and Model Selection of Semiparametric Copula-based Multivariate Dynamic Models under Copula Misspecification, Journal of Econometrics, 2006, Vol. 135, 125-154

[13] Chen, X., Z. Xiao, and R. Koenker (2009): Copula-Based Nonlinear Quantile Autoregression. The Econometrics Journal 12, S50-S67

[14] Drost, F.C. and C.A.J. Klaassen (1997). Efficient estimation in semiparametric GARCH models. Journal of Econometrics, 193–221.

[15] Drost, F.C.A. J. Klaassen and Bas Werker (1997) Aadaptive estimaiton in time series models. The Annals of Statistics Vol. 25, No. 2, 786-817

[16] Duffie, D. and J. Pan (1997), An Overview of Value at Risk, Journal of Derivatives, 4,7-49.

[17] Durbin, J. (1973). Weak convergence of the sample distribution when parameters are estimated. The Annals of Statistics 1, 279-290.

[18] Engle, R. F.(2001) Financial Econometrics - A New Discipline With New Methods Journal of Econometrics. V100 pp53-56

[19] Engel, R. F. and Gonzalez-Rivera (1991) Semiparametric ARCH Models, Journal of Business and Economic Statistics, 9:345-360.

[20] Engle, R. F. and Manganelli (2004) CAViaR: Conditional Value at Risk by Quantile Regression. *Journal of Business and Economic Statistics* 22, 367-81.

[21] Gill, R. D. (1989). Non- and semi-parametric maximum likelihood estimators and the von-Mises method-I. Scand. J. Statist. 16, 97-128.

[22] Hall, P and C. Heyde (1980). Martingale Limit Theory and Its Application. New York: Academic Press.

[23] Hall, P. and Yao, Q. (2003). Inference in ARCH and GARCH models with heavy-tailed errors. Econometrica. 71, 285-317.

[24] Jensen. S.T. and Rahbek, A 2004. Asymptotic Normality of the QMLE Estimator of ARCH in the Nonstationary Case, Econometrica, 72(2), pages 641-646.

[25] Kitamura, Y. (1997). Empirical likelihood methods for weakly dependent processes. Ann. Statist. 25, 2084-2102.

[26] Koenker and Xiao (2009). "Conditional Quantile Estimation for GARCH Models," Boston College Working Papers in Economics 725, Boston College Department of Economics.

[27] Komunjer, Ivana, 2005. "Quasi-maximum likelihood estimation for conditional quantiles," Journal of Econometrics, Elsevier, vol. 128(1), pages 137-164,

[28] Komunjer, I., Vuong, Q., Semiparametric efficiency bound in time-series models for conditional quantiles. Econometric Theory (2010) (forthcoming).

[29] Koul, H. and Ling, S. (2006). Fitting an error distribution in some heteroscedastic time series models. Annals of Statistics 34, 994-1012.

[30] Koul, H. and Ossiander, M. (1994). Weak convergence of randomly weighted dependent residual empiricals with applications to autoregression. Ann. Statist. 22 540–562.

[31] Kuester, K., Mittnik, S., Paolella, M. S. 2006, "Value-at-Risk prediction: A comparison of alternative strategies". Journal of. Financial Econometrics 4(1), p. 53-89.

[32] Lee, S.-W. and Hansen, B.E. (1994). Asymptotic theory for the GARCH (1,1) quasi-maximum likelihood estimator. Econometric Theory, 10, 29-52.

[33] Linton (1993). Adaptive estimation in ARCH models Econometric Theory (1993) 9, 539-569

[34] Linton, Pan and Wang (2009) Estimation for a non-stationary semi-strong GARCH(1,1) model with heavy-tailed errors Forthcoming in Econometric Theory

[35] Linton and Shang (2010) Semiparametric Value-at-Risk Forecasts for ARCH(∞) Models. Working paper, FMG, LSE

[36] Ling, S. and McAleer, M. (2003) Adaptive estimation in nonstationry ARMA models with GARCH noises. Annals of Statistics, 31, 642-674.

[37] McNiel, Alexander and Rudiger Frey(2000) Estimation of tail-related risk measures for heteroscedastic .nancial time series: an extreme value approach. Journal of Empirical Finance, 271-300

[38] U.U. Müller, A. Schick and W. Wefelmeyer (2005). Weighted residual-based density estimators for nonlinear autoregressive models. Statist. Sinica, 15, 177-195.

[39] Schick, A. and Wefelmeyer, W. (2002). Estimating the innovation distribution in nonlinear autoregressive models. Ann. Inst. Statist. Math. 54 245–260.

[40] Nelson, D.B. (1990), Stationarity and persistence in the GARCH(1,1) model, Econometric Theory, 6, 318–334.

[41] Nystrom and Jimmy Skoglund (2004/05) Efficient Filtering of Financial Time Series and Extreme Value Theory Journal of Risk, Vol. 7, No. 2, pp. 63-84,

[42] Owen, A. (1990). Empirical Likelihood Ratio Confidence Regions. Ann. Stat., 18 (1),. 90–120.

[43] Owen (2001). Empirical Likelihood. Chapman and Hall.

[44] Scaillet, O. (2004). Nonparametric estimation and sensitivity analysis of expected shortfall. Mathematical Finance 14, 115-129.

[45] Tong, H. (1990). Non-linear time series: A dynamical system approach. Oxford Univ. Press, Oxford.

[46] Zhang B (1995) M-estimation and quantile estimation in the presence of auxiliary information. Journal of Statistical Planning and Inference 44:77 94

[47] Weiss, A. (1986). Asymptotic theory for ARCH models: estimation and testing. Econometric Theory, 2, 107-131.

[48] Wiener, N. (1958). Nonlinear problems in random theory. MIT Press, Cambridge.

A Appendix

Proof of Theorem 1. Given θ and the observations $\{h_0, y_1, \ldots, y_n\}$, then the log likelihood is

$$L(\theta) = \sum_{t=1}^{n} [\log f(c^{-1} g_t^{-1/2}(\beta) y_t) + \log c^{-1} g_t^{-1/2}(\beta)].$$

Now we can write the conditional score at time t as

$$l_t(\theta) = -(1 + \varepsilon_t(\theta) \frac{f'(\varepsilon_t(\theta))}{f(\varepsilon_t(\theta))}) \begin{pmatrix} \frac{1}{2g_t(\beta)} \frac{\partial g_t(\beta)}{\partial \beta} \\ \frac{1}{c} \end{pmatrix}.$$

Then, according to Drost and Klaassen (1997), the efficient score and information matrix for β are

$$\begin{aligned} l_{1t}^*(\beta_0) &= -\frac{1}{2}\{G_t(\beta_0) - G(\beta_0)\}(1 + \varepsilon_t \frac{f'(\varepsilon_t)}{f(\varepsilon_t)}) \\ E[l_{1t}^*(\beta_0) l_{1t}^*(\beta_0)^\top] &= \frac{E[R_3(\varepsilon)^2]}{4}\{G_2(\beta_0) - G(\beta_0)G(\beta_0)^\top\}, \end{aligned}$$

and

$$\sqrt{n}(\widehat{\beta} - \beta_0) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} E[l_{1t}^*(\beta_0) l_{1t}^*(\beta_0)^\top]^{-1} l_{1t}^*(\beta_0) + o_p(1).$$

Next, as the efficient estimator for c is $\widehat{c} = \sqrt{\frac{1}{n}\sum_{t=1}^{n} \widehat{e}_t^2 - \frac{1}{n}\frac{\sum_{t=1}^{n}\widehat{e}_t^3}{\sum_{t=1}^{n}\widehat{e}_t^2}\sum_{t=1}^{n}\widehat{e}_t}$. Using delta method, we can see

$$\begin{aligned} \widehat{e}_t - e_t &= y_t[g_t^{-1/2}(\widehat{\beta}) - g_t^{-1/2}(\beta)] = -\frac{1}{2}\frac{e_t}{g_t(\beta_0)}\frac{\partial g_t(\beta_0)}{\partial \beta^\top}(\widehat{\beta} - \beta_0) + o_p(\frac{1}{n}) \\ \widehat{e}_t^2 - e_t^2 &= y_t^2[g_t^{-1}(\widehat{\beta}) - g_t^{-1}(\beta)] = -\frac{e_t^2}{g_t(\beta_0)}\frac{\partial g_t(\beta_0)}{\partial \beta^\top}(\widehat{\beta} - \beta_0) + o_p(\frac{1}{n}), \end{aligned}$$

consequently,

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n}\widehat{e}_t-\frac{1}{n}\sum_{t=1}^{n}e_t &= -\frac{1}{2}E[\frac{e_t}{g_t(\beta_0)}\frac{\partial g_t(\beta_0)}{\partial\beta^\top}](\widehat{\beta}-\beta_0)+o_p(n^{-1/2})\\ \frac{1}{n}\sum_{t=1}^{n}\widehat{e}_t^2-\frac{1}{n}\sum_{t=1}^{n}e_t^2 &= -E[\frac{e_t^2}{g_t(\beta_0)}\frac{\partial g_t(\beta_0)}{\partial\beta^\top}](\widehat{\beta}-\beta_0)+o_p(n^{-1/2}),\end{aligned}$$

as a result, by LLN and Ergodic Theorem,

$$\begin{aligned}&\sqrt{n}(\widehat{c}-c_0)\\ =\ &\sqrt{n}(\sqrt{\frac{1}{n}\sum\nolimits_{t=1}^{n}\widehat{e}_t^2-\frac{1}{n}\frac{\sum_{t=1}^{n}\widehat{e}_t^3}{\sum_{t=1}^{n}\widehat{e}_t^2}\sum\nolimits_{t=1}^{n}\widehat{e}_t}-\sqrt{\frac{1}{n}\sum\nolimits_{t=1}^{n}c_0^2\varepsilon_t^2}+\sqrt{\frac{1}{n}\sum\nolimits_{t=1}^{n}c_0^2\varepsilon_t^2}-c_0)\\ =\ &\frac{1}{2c_0}\{-c_0^2G^\top\sqrt{n}(\widehat{\beta}-\beta_0)-c_0^2\frac{E\varepsilon_t^3}{E\varepsilon_t^2}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\varepsilon_t+c_0^2\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\varepsilon_t^2-1)\}+o_p(1)\\ =\ &\frac{c_0}{2}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\{(\varepsilon_t^2-1)-\varepsilon_tE\varepsilon^3-G^\top E[l_{1t}^*l_{1t}^{*\top}]^{-1}l_{1t}^*\}+o_p(1).\end{aligned}$$

Since $E[(\varepsilon_t^2-1)l_{1t}^*]=0$ and $E[\varepsilon_t l_{1t}^*]=0$, we have

$$\Omega_c=\frac{c_0^2}{4}\{E\varepsilon^4-1-(E\varepsilon^3)^2+G^\top E[l_{1t}^*l_{1t}^{*}{}^\top]^{-1}G\}.$$

We can thus conclude that

$$\begin{aligned}&\sqrt{n}\begin{pmatrix}\widehat{\beta}-\beta_0\\ \widehat{c}-c_0\end{pmatrix}\\ =\ &\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\begin{pmatrix}-\frac{1}{2}E[l_{1t}^*l_{1t}^{*\top}]^{-1}\{G_t-G\} & 0\\ \frac{c_0}{4}G^\top E[l_{1t}^*l_{1t}^{*\top}]^{-1}\{G_t-G\} & \frac{c_0}{2}(-E\varepsilon^3,1)\end{pmatrix}\begin{pmatrix}R_3(\varepsilon_t)\\ R_2(\varepsilon_t)\end{pmatrix}+o_p(1)\end{aligned}$$

and

$$\Omega_\theta=\begin{pmatrix}E[l_{1t}^*l_{1t}^{*\top}]^{-1} & -\frac{c_0}{2}E[l_{1t}^*l_{1t}^{*\top}]^{-1}G\\ -\frac{c_0}{2}G^\top E[l_{1t}^*l_{1t}^{*\top}]^{-1} & \frac{c_0^2}{4}\{E\varepsilon^4-1-(E\varepsilon^3)^2+G^\top E[l_{1t}^*l_{1t}^{*\top}]^{-1}G\}\end{pmatrix}.$$

■

LEMMA 1. *Suppose that assumptions A.2-A.4 hold. Then* $\widehat{F}_N(x)$ *and* $\widehat{F}_w(x)$ *have the* *following expansion:*

$$\begin{aligned}\sup_{x\in\mathbb{R}}\left|\widehat{F}_N(x)-F(x)-\frac{1}{n}\sum_{t=1}^{n}\{1(\varepsilon_t\le x)-F(x)\}-f(x)\frac{1}{n}\sum_{t=1}^{n}\varepsilon_t-\frac{xf(x)}{2}\frac{1}{n}\sum_{t=1}^{n}(\varepsilon_t^2-1)\right| &= o_p(n^{-1/2}) \\ \sup_{x\in\mathbb{R}}\left|\widehat{F}_w(x)-F(x)-\frac{1}{n}\sum_{t=1}^{n}\{1(\varepsilon_t\le x)-F(x)\}+\frac{1}{n}\sum_{t=1}^{n}A_x^\top B^{-1}R_2(\varepsilon_t)\right| &= o_p(n^{-1/2}).\end{aligned}$$

Consequently, the process $\sqrt{n}(\widehat{F}_N-F)$ converges weakly to a zero-mean Gaussian process $\mathcal{Z}_N$ with covariance function Ω_N and the process $\sqrt{n}(\widehat{F}_w-F)$ converges weakly to a zero-mean Gaussian process $\mathcal{Z}_w$ with covariance function Ω_w, where:

$$\begin{aligned}\Omega_N(x,x') &= \mathrm{cov}\left(\mathcal{Z}_N(x),\mathcal{Z}_N(x')\right) \\ &= E\left[[1(\varepsilon\le x)-F(x)+f(x)\varepsilon+\frac{xf(x)}{2}(\varepsilon^2-1)]\right. \\ &\qquad \left.\times\,[1(\varepsilon\le x')-F(x')+f(x')\varepsilon+\frac{x'f(x')}{2}(\varepsilon^2-1)]\right] \\ \Omega_w(x,x') &= \mathrm{cov}\left(\mathcal{Z}_w(x),\mathcal{Z}_w(x')\right) \\ &= E\left[[1(\varepsilon\le x)-F(x)-A_x^\top B^{-1}R_2(\varepsilon)][1(\varepsilon\le x')-F(x')-A_{x'}^\top B^{-1}R_2(\varepsilon)]\right].\end{aligned}$$

Where we define the following quantities:

$$A_x=E[R_2(\varepsilon)1(\varepsilon\le x)];B=E[R_2(\varepsilon)R_2(\varepsilon)^\top];$$

$$C_x=f(x)^2\left\{\frac{E[\varepsilon^4]-1}{4}x^2+xE[\varepsilon^3]+1\right\}+f(x)\left\{2E[\varepsilon 1(\varepsilon\le x)]+xE[(\varepsilon^2-1)1(\varepsilon\le x)]\right\},$$

Proof of Lemma 1. We follow the proof of Theorem 4.1 in Koul and Ling (2006) closely. Define the empirical process

$$\nu_n(x,z_1,z_2)=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\{1(\varepsilon_t\le z_1+xz_2)-E[1(\varepsilon_t\le z_1+xz_2)]\}.$$

For any $z=(z_1,z_2)\in\mathbb{R}^2$, let $|z|=|z_1|\vee|z_2|$. In $\mathbb{R}^2$, we define a pseudo-metric

$$d_c(x,y)=\sup_{|z|\le c}|F(x(1+z_1)+z_2)-F(y(1+z_1)+z_2)|^{1/2},(x,y)\in\mathbb{R}^2,c>0.$$

Let $\mathcal{N}(\delta,c)$ be the cardinality of the minimal δ-net of and let

$$\mathcal{I}(c)=\int_0^1\{\ln\mathcal{N}(u,c)\}^{1/2}du$$

According to Theorem 4.1 in Koul and Ling (2006), assumptions imply that $\mathcal{I}(c) < \infty$ for any $c \in [0, 1)$. This combines with Koul and Ossiander (1994) show that the following stochastic equicontinuity condition holds:

$$\sup_{x\in\mathbb{R},|z_1|\le Cn^{-1/2},|z_2-1|\le Cn^{-1/2}} |\nu_n(x, z_1, z_2) - \nu_n(x, 0, 1)| = o_p(1).$$

As a result,

$$\nu_n(x, z_1, z_2) = \nu_n(x, 0, 1) + \nu_n(x, z_1, z_2) - \nu_n(x, 0, 1) = \nu_n(x, 0, 1) + o_p(1).$$

By LLN, we know that

$$\frac{1}{n}\sum_{t=1}^{n}\varepsilon_t = o_p(1), \sqrt{\frac{1}{n}\sum_{t=1}^{n}\varepsilon_t^2 - (\frac{1}{n}\sum\nolimits_{t=1}^{n}\varepsilon_t)^2} = o_p(1).$$

Therefore, $\widehat{F}_N(x)$ can be expanded as, uniformly in $x \in \mathbb{R}$,

$$\begin{aligned}
&\sqrt{n}(\widehat{F}_N(x) - F(x))\\
= &\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left\{1\left(\frac{\varepsilon_t - \widehat{\mu}_\varepsilon}{\widehat{\sigma}_\varepsilon} \le x\right) - F(x)\right\}\\
= &\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\{1\left(\varepsilon_t \le \widehat{\mu}_\varepsilon + x\widehat{\sigma}_\varepsilon\right) - F\left(\widehat{\mu}_\varepsilon + x\widehat{\sigma}_\varepsilon\right)\} + \sqrt{n}\{F(\widehat{\mu}_\varepsilon + x\widehat{\sigma}_\varepsilon) - F(x)\}\\
= &\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\{1(\varepsilon_t \le x) - F(x)\} + o_p(1) + \sqrt{n}\{F(\widehat{\mu}_\varepsilon + x\widehat{\sigma}_\varepsilon) - F(x)\}\\
= &\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\{1(\varepsilon_t \le x) - F(x)\} + f(x)\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\varepsilon_t + \sqrt{n}xf(x)(\sqrt{\frac{1}{n}\sum_{t=1}^{n}\varepsilon_t^2 - (\widehat{\mu}_\varepsilon)^2} - 1) + o_p(1)\\
= &\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\{1(\varepsilon_t \le x) - F(x)\} + f(x)\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\varepsilon_t + \sqrt{n}\frac{xf(x)}{2}[\frac{1}{n}\sum_{t=1}^{n}(\varepsilon_t^2 - 1) - (\widehat{\mu}_\varepsilon)^2] + o_p(1)\\
= &\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\{1(\varepsilon_t \le x) - F(x)\} + f(x)\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\varepsilon_t + \frac{xf(x)}{2}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\varepsilon_t^2 - 1) + o_p(1).
\end{aligned}$$

We know from Owen (2001) that

$$\widehat{w}_t = \frac{1}{n}\frac{1}{1 + \lambda_n' R_2(\varepsilon_t)}; \lambda_n = B^{-1}(\frac{1}{n}\sum_{t=1}^{n}R_2(\varepsilon_t)) + o_p(n^{-1/2}).$$

Consequently, uniformly in $x \in \mathbb{R}$,

$$
\begin{aligned}
& \sqrt{n}(\widehat{F}_w(x) - F(x)) \\
= \; & \sqrt{n}(\sum_{t=1}^{n} \widehat{w}_t 1(\varepsilon_t \leq x) - \frac{1}{n}\sum_{t=1}^{n} 1(\varepsilon_t \leq x) + \frac{1}{n}\sum_{t=1}^{n} 1(\varepsilon_t \leq x) - F(x)) \\
= \; & \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\{[n\widehat{w}_t - 1]1(\varepsilon_t \leq x)\} + \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\{1(\varepsilon_t \leq x) - F(x)\} \\
= \; & -\lambda_n^\top \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\{R_2(\varepsilon_t)1(\varepsilon_t \leq x)\} + \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\{1(\varepsilon_t \leq x) - F(x)\} + o_p(n^{-1/2}) \\
= \; & -\frac{1}{\sqrt{n}}\sum_{t=1}^{n} R_2(\varepsilon_t)^\top B^{-1}\frac{1}{n}\sum_{t=1}^{n}\{R_2(\varepsilon_t)1(\varepsilon_t \leq x)\} + \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\{1(\varepsilon_t \leq x) - F(x)\} + o_p(n^{-1/2}) \\
= \; & \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\{1(\varepsilon_t \leq x) - F(x)\} - \frac{1}{\sqrt{n}}\sum_{t=1}^{n} A_x B^{-1} R_2(\varepsilon_t) + o_p(n^{-1/2}),
\end{aligned}
$$

where the last equality holds because of ergodic theorem: $n^{-1}\sum_{t=1}^{n}\{R_2(\varepsilon_t)1(\varepsilon_t \leq x)\} = A_x + o_p(1)$.

■

Corollary 1. *Denote $E[\varepsilon 1(\varepsilon \leq x)] = a_1(x)$ and $E[(\varepsilon^2 - 1)1(\varepsilon \leq x)] = a_2(x)$. $\widehat{F}(x)$ is asymptotically less efficient than $\widehat{F}_w(x)$, and $\widehat{F}(x)$ achieves the efficiency bound iff*

$$a_1(x) = a_2(x) = 0.$$

$\widehat{F}_N(x)$ is asymptotically less efficient than $\widehat{F}_w(x)$ as $C_x \geq -A_x^\top B^{-1} A_x$. $\widehat{F}_N(x)$ achieves the efficiency bound iff

$$x = \frac{2(a_2(x) - a_1(x)E[\varepsilon^3])}{a_1(x)(E[\varepsilon^4] - 1) - a_2(x)E[\varepsilon^3]} \quad ; \quad f(x) = -\frac{2a_1(x) + xa_2(x)}{\frac{E[\varepsilon^4]-1}{2}x^2 + 2xE[\varepsilon^3] + 2}.$$

Proof of Corollary 1. Notice that

$$A_x^\top B^{-1} A_x = \frac{\{E[\varepsilon^4] - 1\}\{a_1(x) - a_2(x)\frac{E[\varepsilon^3]}{E[\varepsilon^4]-1}\}^2}{E[\varepsilon^4] - 1 - E[\varepsilon^3]^2} + \frac{a_2^2(x)}{E[\varepsilon^4] - 1},$$

and under the moment condition (2), $E[\varepsilon^4]-1=\mathrm{Var}(\varepsilon^2)\geq 0$ and

$$\begin{aligned} E[\varepsilon^4]-1-E[\varepsilon^3]^2 &= \{E[\varepsilon^4]-1\}\{1-\frac{E[\varepsilon^3]^2}{E[\varepsilon^4]-1}\} \\ &= \{E[\varepsilon^4]-1\}\{1-\mathrm{corr}(\varepsilon,\varepsilon^2)^2\} \\ &\geq 0 \end{aligned}$$

so $A_x^\top B^{-1}A_x\geq 0$ and $A_x^\top B^{-1}A_x=0\Leftrightarrow a_1(x)=a_2(x)=0$. As for the asymptotical efficiency comparison between $\widehat{F}_N(x)$ and $\widehat{F}_w(x)$, we have

$$\begin{aligned} & C_x+A_x^\top B^{-1}A_x \\ = & f(x)^2\{\frac{E[\varepsilon^4]-1}{4}x^2+xE[\varepsilon^3]+1\}+f(x)\{2a_1(x)+xa_2(x)\} \\ & +\frac{\{E[\varepsilon^4]-1\}\{a_1(x)-a_2(x)\frac{E[\varepsilon^3]}{E[\varepsilon^4]-1}\}^2}{E[\varepsilon^4]-1-E[\varepsilon^3]^2}+\frac{a_2^2(x)}{E[\varepsilon^4]-1} \\ = & f(x)^2\{\frac{E[\varepsilon^4]-1}{4}x^2+xE[\varepsilon^3]+1\}+f(x)\{2a_1(x)+xa_2(x)\} \\ & +\frac{\{E[\varepsilon^4]-1\}a_1(x)-2a_1(x)a_2(x)E[\varepsilon^3]+a_2^2(x)}{E[\varepsilon^4]-1-E[\varepsilon^3]^2} \\ = & \{\frac{E[\varepsilon^4]-1}{4}x^2+xE[\varepsilon^3]+1\}\{f(x)+\frac{2a_1(x)+xa_2(x)}{\frac{E[\varepsilon^4]-1}{2}x^2+2xE[\varepsilon^3]+2}\}^2 \\ & +\frac{\{x[a_1(x)(E[\varepsilon^4]-1)-a_2(x)E[\varepsilon^3]]+2(a_1(x)E[\varepsilon^3]-a_2(x))\}^2}{4\{E[\varepsilon^4]-1-E[\varepsilon^3]^2\}\{\frac{E[\varepsilon^4]-1}{4}x^2+xE[\varepsilon^3]+1\}}, \end{aligned}$$

additionally

$$\frac{E[\varepsilon^4]-1}{4}x^2+xE[\varepsilon^3]+1=\frac{E[\varepsilon^4]-1}{4}\{x+\frac{2E[\varepsilon^3]}{E[\varepsilon^4]-1}\}^2+\frac{E[\varepsilon^4]-1-E[\varepsilon^3]^2}{E[\varepsilon^4]-1}\geq 0,$$

so we can conclude $C_x\geq -A_x^\top B^{-1}A_x$, and $C_x=-A_x^\top B^{-1}A_x$ if and only if

$$x=\frac{2(a_2(x)-a_1(x)E[\varepsilon^3])}{a_1(x)(E[\varepsilon^4]-1)-a_2(x)E[\varepsilon^3]};f(x)=-\frac{2a_1(x)+xa_2(x)}{\frac{E[\varepsilon^4]-1}{2}x^2+2xE[\varepsilon^3]+2}.$$

■

LEMMA 2. *Suppose assumptions A.1-A.4 hold and there is an estimator $\widetilde{\theta}$ that has influence function $\chi_t(\theta_0)$,then the following expansion for distribution estimators based on $\widetilde{\theta}$ is*

$$\sup_{x\in\mathbb{R}}\left|\widehat{\widehat{F}}(x)-F(x)-\frac{1}{n}\sum_{t=1}^n\{1(\varepsilon_t\leq x)-F(x)\}-\frac{xf(x)}{2}H(\theta_0)^\top\frac{1}{n}\sum_{t=1}^n\chi_t(\theta_0)\right|=o_p(n^{-1/2})$$

$$\sup_{x\in\mathbb{R}}\left|\widehat{\widehat{F}}_N(x)-F(x)-\frac{1}{n}\sum_{t=1}^n\{1(\varepsilon_t\le x)-F(x)\}-\frac{xf(x)}{2}H(\theta_0)^\top\frac{1}{n}\sum_{t=1}^n\chi_t(\theta_0)-f(x)\frac{1}{n}\sum_{t=1}^n\varepsilon_t\right.$$
$$\left.-\frac{xf(x)}{2}\frac{1}{n}\sum_{t=1}^n\{\varepsilon_t^2-1\}\right|=o_p(n^{-1/2})$$
$$\sup_{x\in\mathbb{R}}\left|\widehat{\widehat{F}}_w(x)-F(x)-\frac{1}{n}\sum_{t=1}^n\{1(\varepsilon_t\le x)-F(x)\}-\left\{\frac{xf(x)}{2}+\begin{pmatrix}0&1\end{pmatrix}B^{-1}A_x\right\}H(\theta_0)^\top\frac{1}{n}\sum_{t=1}^n\chi_t(\theta_0)\right.$$
$$\left.+\frac{1}{n}\sum_{t=1}^n A_x^\top B^{-1}R_2(\varepsilon_t)\right|=o_p(n^{-1/2}).$$

Proof of Lemma 2. By Taylor expansion,

$$\begin{aligned}\sqrt{\frac{h_t(\widetilde{\theta})}{h_t(\theta_0)}}-1 &= \frac{1}{2}\frac{\partial\log h_t(\theta)}{\partial\theta^\top}(\widetilde{\theta}-\theta_0)\\ &+\frac{1}{4}(\widetilde{\theta}-\theta_0)^\top[\frac{1}{h_t(\theta)}\frac{\partial^2 h_t(\theta_1)}{\partial\theta\partial\theta^\top}-\frac{1}{2}\frac{\partial\log h_t(\theta_1)}{\partial\theta}\frac{\partial\log h_t(\theta_1)}{\partial\theta^\top}](\widetilde{\theta}-\theta_0),\end{aligned}$$

where θ_1 lies in between θ_0 and $\widetilde{\theta}$. Since $\widetilde{\theta}-\theta_0=\frac{1}{n}\sum_{t=1}^n\chi_t(\theta_0)+o_p(\frac{1}{\sqrt{n}})$, and $E_{\theta_0}\sup_{\theta\in U_{\theta_0}}||\frac{\partial\log h_t(\theta)}{\partial\theta}||^2<\infty$,which is due to Example 3.1 in Koul and Ling (2006), we have $\sum_{t=1}^n\left(\sqrt{\frac{h_t(\widetilde{\theta})}{h_t(\theta_0)}}-1\right)^2=o_p(1)$. This implies

$$\sup_{1\le t\le n}|\sqrt{\frac{h_t(\widetilde{\theta})}{h_t(\theta_0)}}-1|=o_p(1).$$

Using the same empirical process argument as in lemma , Lemma 4.1 in Koul and Ling

(2006), and the fact that it is clear that, uniformly in $x \in \mathbb{R}$,

$$\begin{aligned}
& \sqrt{n}(\widehat{\widehat{F}}(x) - F(x)) \\
= & \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \{1(\varepsilon_t(\widetilde{\theta}) \le x) - F(x)\} \\
= & \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \{1(\varepsilon_t \le \sqrt{\frac{h_t(\widetilde{\theta})}{h_t(\theta_0)}} x) - F(\sqrt{\frac{h_t(\widetilde{\theta})}{h_t(\theta_0)}} x) + F(\sqrt{\frac{h_t(\widetilde{\theta})}{h_t(\theta_0)}} x) - F(x)\} \\
= & \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \{1(\varepsilon_t \le \sqrt{\frac{h_t(\widetilde{\theta})}{h_t(\theta_0)}} x) - F(\sqrt{\frac{h_t(\widetilde{\theta})}{h_t(\theta_0)}} x)\} + \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \{F(\sqrt{\frac{h_t(\widetilde{\theta})}{h_t(\theta_0)}} x) - F(x)\} \\
= & \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \{1(\varepsilon_t \le x) - F(x)\} + \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \{F(\sqrt{\frac{h_t(\widetilde{\theta})}{h_t(\theta_0)}} x) - F(x)\} + o_p(1) \\
= & \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \{1(\varepsilon_t \le x) - F(x)\} + \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{x f(x)}{2h_t(\theta_0)} \frac{\partial h_t(\theta_0)}{\partial \theta^{\top}} (\widetilde{\theta} - \theta_0) + o_p(1) \\
= & \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \{1(\varepsilon_t \le x) - F(x)\} + \frac{1}{n} \sum_{t=1}^{n} \frac{x f(x)}{2h_t(\theta_0)} \frac{\partial h_t(\theta_0)}{\partial \theta^{\top}} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \chi_t(\theta_0) + o_p(1) \\
= & \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \{1(\varepsilon_t \le x) - F(x)\} + \frac{x f(x)}{2} H(\theta_0)^{\top} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \chi_t(\theta_0) + o_p(1),
\end{aligned}$$

where the last equation holds because of the ergodicity theorem $\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \frac{\partial \log h_t(\theta_0)}{\partial \theta} = E[\frac{\partial \log h_t(\theta_0)}{\partial \theta}]$.

The next is to show the asymptotic expansion for $\widehat{\widehat{F}}_N(x)$. Since $\varepsilon_t(\widetilde{\theta}) = \sqrt{\frac{h_t(\theta_0)}{h_t(\widetilde{\theta})}} \varepsilon_t$, the renormalized empirical distribution estimator can be shown, uniformly in $x \in \mathbb{R}$:

$$
\begin{aligned}
& \sqrt{n}(\widehat{\widehat{F}}_N(x) - F(x)) \\
= \ & \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \{ 1 \left(\frac{\varepsilon_t(\widetilde{\theta}) - \frac{1}{n}\sum\limits_{t=1}^{n} \varepsilon_t(\widetilde{\theta})}{\sqrt{\frac{1}{n}\sum\limits_{t=1}^{n} \varepsilon_t^2(\widetilde{\theta}) - (\frac{1}{n}\sum\limits_{t=1}^{n} \varepsilon_t(\widetilde{\theta}))^2}} \leq x \right) - F(x) \} \\
= \ & \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \{ 1(\varepsilon_t \leq \sqrt{\frac{h_t(\widetilde{\theta})}{h_t(\theta_0)}} \frac{1}{n} \sum_{t=1}^{n} \sqrt{\frac{h_t(\theta_0)}{h_t(\widetilde{\theta})}} \varepsilon_t + \sqrt{\frac{h_t(\widetilde{\theta})}{h_t(\theta_0)}} \sqrt{\frac{1}{n} \sum_{t=1}^{n} \frac{h_t(\theta_0)}{h_t(\widetilde{\theta})} \varepsilon_t^2 - (\frac{1}{n} \sum_{t=1}^{n} \varepsilon_t(\widetilde{\theta}))^2} x) \\
& - F(\sqrt{\frac{h_t(\widetilde{\theta})}{h_t(\theta_0)}} \frac{1}{n} \sum_{t=1}^{n} \sqrt{\frac{h_t(\theta_0)}{h_t(\widetilde{\theta})}} \varepsilon_t + \sqrt{\frac{h_t(\widetilde{\theta})}{h_t(\theta_0)}} \sqrt{\frac{1}{n} \sum_{t=1}^{n} \frac{h_t(\theta_0)}{h_t(\widetilde{\theta})} \varepsilon_t^2 - (\frac{1}{n} \sum_{t=1}^{n} \varepsilon_t(\widetilde{\theta}))^2} x) \\
& + F(\sqrt{\frac{h_t(\widetilde{\theta})}{h_t(\theta_0)}} \frac{1}{n} \sum_{t=1}^{n} \sqrt{\frac{h_t(\theta_0)}{h_t(\widetilde{\theta})}} \varepsilon_t + \sqrt{\frac{h_t(\widetilde{\theta})}{h_t(\theta_0)}} \sqrt{\frac{1}{n} \sum_{t=1}^{n} \frac{h_t(\theta_0)}{h_t(\widetilde{\theta})} \varepsilon_t^2 - (\frac{1}{n} \sum_{t=1}^{n} \varepsilon_t(\widetilde{\theta}))^2} x) - F(x) \} \\
= \ & \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \{ 1(\varepsilon_t \leq x) - F(x) \} + o_p(1) \\
& + \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \{ F(\sqrt{\frac{h_t(\widetilde{\theta})}{h_t(\theta_0)}} \frac{1}{n} \sum_{t=1}^{n} \sqrt{\frac{h_t(\theta_0)}{h_t(\widetilde{\theta})}} \varepsilon_t + \sqrt{\frac{h_t(\widetilde{\theta})}{h_t(\theta_0)}} \sqrt{\frac{1}{n} \sum_{t=1}^{n} \varepsilon_t^2(\widetilde{\theta}) - (\frac{1}{n} \sum_{t=1}^{n} \varepsilon_t(\widetilde{\theta}))^2} x) - F(x) \}
\end{aligned}
$$

where the last equation used empirical process approximation and

Now given that $\sqrt{n}(\widetilde{\theta} - \theta_0) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \chi_t(\theta_0) + o_p(1)$, we know that $\sqrt{\frac{1}{n}\sum\limits_{t=1}^{n} \varepsilon_t^2(\widetilde{\theta}) - (\frac{1}{n}\sum\limits_{t=1}^{n} \varepsilon_t(\widetilde{\theta}))^2}$ is of the same order as $\sqrt{\frac{1}{n}\sum\limits_{t=1}^{n} \varepsilon_t^2(\widetilde{\theta})}$, which is due to the fact that $(\frac{1}{n}\sum\limits_{t=1}^{n} \varepsilon_t(\widetilde{\theta}))^2$ is of higher order than $\frac{1}{n}\sum_{t=1}^{n}\varepsilon_t^2(\widetilde{\theta})$. As a result,

$$
\begin{aligned}
& F(\sqrt{\frac{h_t(\widetilde{\theta})}{h_t(\theta_0)}}\frac{1}{n}\sum_{t=1}^{n}\sqrt{\frac{h_t(\theta_0)}{h_t(\widetilde{\theta})}}\varepsilon_t+\sqrt{\frac{h_t(\widetilde{\theta})}{h_t(\theta_0)}}\sqrt{\frac{1}{n}\sum_{t=1}^{n}\frac{h_t(\theta_0)}{h_t(\widetilde{\theta})}\varepsilon_t^2}x)-F(x) \\
= \; & f(x)\{\sqrt{\frac{h_t(\widetilde{\theta})}{h_t(\theta_0)}}-1\}\frac{1}{n}\sum_{t=1}^{n}\sqrt{\frac{h_t(\theta_0)}{h_t(\widetilde{\theta})}}\varepsilon_t \\
& +f(x)\frac{1}{n}\sum_{t=1}^{n}\sqrt{\frac{h_t(\theta_0)}{h_t(\widetilde{\theta})}}\varepsilon_t \\
& +xf(x)\{\sqrt{\frac{h_t(\widetilde{\theta})}{h_t(\theta_0)}}-1\}\sqrt{\frac{1}{n}\sum_{t=1}^{n}\frac{h_t(\theta_0)}{h_t(\widetilde{\theta})}\varepsilon_t^2} \\
& +xf(x)\{\sqrt{\frac{1}{n}\sum_{t=1}^{n}\frac{h_t(\theta_0)}{h_t(\widetilde{\theta})}\varepsilon_t^2}-1\}+O_p(\frac{1}{n}) \\
= \; & I_{1t}+I_{2t}+I_{3t}+I_{4t},
\end{aligned}
$$

where

$$
\begin{aligned}
I_{1t} &= f(x)\{\frac{1}{2h_t(\theta_0)}\frac{\partial h_t(\theta_0)}{\partial\theta^{\top}}(\widetilde{\theta}-\theta_0)\}\frac{1}{n}\sum_{t=1}^{n}[1-\frac{1}{\sqrt{h_t(\theta_0)}}\frac{1}{2h_t(\theta_0)}\frac{\partial h_t(\theta_0)}{\partial\theta^{\top}}(\widetilde{\theta}-\theta_0)]\varepsilon_t \\
I_{2t} &= f(x)\frac{1}{n}\sum_{t=1}^{n}[1-\frac{1}{\sqrt{h_t(\theta_0)}}\frac{1}{2h_t(\theta_0)}\frac{\partial h_t(\theta_0)}{\partial\theta^{\top}}(\widetilde{\theta}-\theta_0)]\varepsilon_t \\
I_{3t} &= xf(x)\{\frac{1}{2h_t(\theta_0)}\frac{\partial h_t(\theta_0)}{\partial\theta^{\top}}(\widetilde{\theta}-\theta_0)\}\sqrt{\frac{1}{n}\sum_{t=1}^{n}\frac{h_t(\theta_0)}{h_t(\widetilde{\theta})}\varepsilon_t^2} \\
I_{4t} &= xf(x)\{\sqrt{\frac{1}{n}\sum_{t=1}^{n}\frac{h_t(\theta_0)}{h_t(\widetilde{\theta})}\varepsilon_t^2}-1\}.
\end{aligned}
$$

It is easy to see that

$$
\begin{aligned}
\sqrt{\frac{h_t(\widetilde{\theta})}{h_t(\theta_0)}}-1 &= \frac{1}{2h_t(\theta_0)}\frac{\partial h_t(\theta_0)}{\partial\theta^{\top}}(\widetilde{\theta}-\theta_0)+O_p(\frac{1}{n}) \\
\sqrt{\frac{h_t(\theta_0)}{h_t(\widetilde{\theta})}}-1 &= -\frac{1}{\sqrt{h_t(\theta_0)}}\frac{1}{2h_t(\theta_0)}\frac{\partial h_t(\theta_0)}{\partial\theta^{\top}}(\widetilde{\theta}-\theta_0)+O_p(\frac{1}{n}) \\
\frac{h_t(\theta_0)}{h_t(\widetilde{\theta})}-1 &= -\frac{1}{h_t(\theta_0)}\frac{\partial h_t(\theta_0)}{\partial\theta^{\top}}(\widetilde{\theta}-\theta_0)+O_p(\frac{1}{n})
\end{aligned}
$$

so now the four components can be rewritten as:

$$
\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{t=1}^{n} I_{1t} \\
= \;& \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\{f(x)\{\frac{1}{2h_t(\theta_0)}\frac{\partial h_t(\theta_0)}{\partial\theta^{\top}}(\widetilde{\theta}-\theta_0)\}\frac{1}{n}\sum_{t=1}^{n}[1-\frac{1}{\sqrt{h_t(\theta_0)}}\frac{1}{2h_t(\theta_0)}\frac{\partial h_t(\theta_0)}{\partial\theta^{\top}}(\widetilde{\theta}-\theta_0)]\varepsilon_t\} \\
= \;& \frac{f(x)}{2}\{\frac{1}{n}\sum_{t=1}^{n}\varepsilon_t-\frac{1}{n}\sum_{t=1}^{n}\frac{1}{2h_t^{3/2}(\theta_0)}\frac{\partial h_t(\theta_0)}{\partial\theta^{\top}}(\widetilde{\theta}-\theta_0)\varepsilon_t\}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{1}{h_t(\theta_0)}\frac{\partial h_t(\theta_0)}{\partial\theta^{\top}}(\widetilde{\theta}-\theta_0)
\end{aligned}
$$

$$
\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{t=1}^{n} I_{2t} &= f(x)\frac{1}{\sqrt{n}}\sum_{t=1}^{n}[1-\frac{1}{\sqrt{h_t(\theta_0)}}\frac{1}{2h_t(\theta_0)}\frac{\partial h_t(\theta_0)}{\partial\theta^{\top}}(\widetilde{\theta}-\theta_0)]\varepsilon_t \\
\frac{1}{\sqrt{n}}\sum_{t=1}^{n} I_{3t} &= xf(x)\sqrt{\frac{1}{n}\sum_{t=1}^{n}\frac{h_t(\theta_0)}{h_t(\widetilde{\theta})}\varepsilon_t^2}\,\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\{\frac{1}{2h_t(\theta_0)}\frac{\partial h_t(\theta_0)}{\partial\theta^{\top}}(\widetilde{\theta}-\theta_0)\} \\
\frac{1}{\sqrt{n}}\sum_{t=1}^{n} I_{4t} &= xf(x)\sqrt{n}\{\sqrt{\frac{1}{n}\sum_{t=1}^{n}\frac{h_t(\theta_0)}{h_t(\widetilde{\theta})}\varepsilon_t^2}-1\}.
\end{aligned}
$$

Consequently,

$$
\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\{F(\sqrt{\frac{h_t(\widetilde{\theta})}{h_t(\theta_0)}}\frac{1}{n}\sum_{t=1}^{n}\sqrt{\frac{h_t(\theta_0)}{h_t(\widetilde{\theta})}}\varepsilon_t+\sqrt{\frac{h_t(\widetilde{\theta})}{h_t(\theta_0)}}\sqrt{\frac{1}{n}\sum_{t=1}^{n}\frac{h_t(\theta_0)}{h_t(\widetilde{\theta})}\varepsilon_t^2}x)-F(x)\} \\
= \;& f(x)\frac{1}{\sqrt{n}}\sum_{t=1}^{n}[1-\frac{1}{2h_t^{3/2}(\theta_0)}\frac{\partial h_t(\theta_0)}{\partial\theta^{\top}}(\widetilde{\theta}-\theta_0)]\varepsilon_t+xf(x)\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{1}{2h_t(\theta_0)}\frac{\partial h_t(\theta_0)}{\partial\theta^{\top}}(\widetilde{\theta}-\theta_0) \\
&+\frac{xf(x)}{2}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\{\varepsilon_t^2-1\}+o_p(1) \\
= \;& f(x)\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\{[1-\frac{1}{2h_t^{3/2}(\theta_0)}\frac{\partial h_t(\theta_0)}{\partial\theta^{\top}}(\widetilde{\theta}-\theta_0)]\varepsilon_t+x\frac{1}{2h_t(\theta_0)}\frac{\partial h_t(\theta_0)}{\partial\theta^{\top}}(\widetilde{\theta}-\theta_0)+\frac{x}{2}[\varepsilon_t^2-1]\}+o_p(1),
\end{aligned}
$$

and by CLT and LLN,

$$
\begin{aligned}
\sqrt{n}(\widetilde{\theta}-\theta_0) &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\chi_t(\theta_0)+o_p(1) \\
\frac{1}{n}\sum_{t=1}^{n}\frac{1}{2h_t^{3/2}(\theta_0)}\frac{\partial h_t(\theta_0)}{\partial\theta}\varepsilon_t &= o_p(1)
\end{aligned}
$$

we have

$$\frac{1}{\sqrt{n}}f(x)\sum_{t=1}^{n}\{\frac{1}{\sqrt{h_t(\theta_0)}}\frac{1}{2h_t(\theta_0)}\frac{\partial h_t(\theta_0)}{\partial\theta^{\top}}(\widetilde{\theta}-\theta_0)\varepsilon_t\}=o_p(1).$$

Therefore, uniformly in $x\in\mathbb{R}$,

$$\begin{aligned}
&\sqrt{n}(\widehat{\widehat{F}}_N(x)-F(x))\\
=\ &n^{-1/2}\sum_{t=1}^{n}\{1(\varepsilon_t\le x)-F(x)\}\\
&+f(x)\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\{\varepsilon_t+x\frac{1}{2h_t(\theta_0)}\frac{\partial h_t(\theta_0)}{\partial\theta^{\top}}(\widetilde{\theta}-\theta_0)+\frac{x}{2}[\varepsilon_t^2-1]\}+o_p(1).
\end{aligned}$$

Since we know that

$$\widehat{\widehat{w}}_t=\frac{1}{n}\frac{1}{1+\widehat{\lambda}_n'R_2(\varepsilon_t(\widetilde{\theta}))};\widehat{\lambda}_n=B_n^{-1}(\frac{1}{n}\sum_{t=1}^{n}R_2(\varepsilon_t(\widetilde{\theta})))+o_p(n^{-1/2})$$

where $B_n=\frac{1}{n}\sum_{t=1}^{n}R_2(\varepsilon_t(\widetilde{\theta}))R_2(\varepsilon_t(\widetilde{\theta}))^{\top}$. Therefore,

$$\begin{aligned}
&\sqrt{n}(\widehat{\widehat{F}}_w(x)-F(x))\\
=\ &\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\{n\widehat{\widehat{w}}_t1(\varepsilon_t(\widetilde{\theta})\le x)-F(x)\}\\
=\ &\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\{n\widehat{\widehat{w}}_t1(\varepsilon_t(\widetilde{\theta})\le x)-1(\varepsilon_t(\widetilde{\theta})\le x)+1(\varepsilon_t(\widetilde{\theta})\le x)-F(x)\}\\
=\ &\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\{[n\widehat{\widehat{w}}_t-1]1(\varepsilon_t(\widetilde{\theta})\le x)\}+\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\{1(\varepsilon_t(\widetilde{\theta})\le x)-F(x)\}\\
=\ &I_5+\sqrt{n}(\widehat{\widehat{F}}(x)-F(x)).
\end{aligned}$$

Define $\varepsilon_t^*=\varepsilon_t+\frac{\varepsilon_t}{2}\frac{\partial\log h_t(\theta)}{\partial\theta^{\top}}(\widetilde{\theta}-\theta)$. From the $\sqrt{n}$-consistency of $\widetilde{\theta}$ and $E[\frac{\varepsilon_t}{2}\frac{\partial\log h_t(\theta)}{\partial\theta^{\top}}]=0$, we can see that

$$\sum_{t=1}^{n}\left(\varepsilon_t(\widetilde{\theta})-\varepsilon_t\right)^2=\sum_{t=1}^{n}\left(\frac{\varepsilon_t}{2}\frac{\partial\log h_t(\theta)}{\partial\theta^{\top}}(\widetilde{\theta}-\theta)+O_p((\widetilde{\theta}-\theta)^2)\right)^2=o_p(1),$$

which implies that $\max_{1\le t\le n}|\varepsilon_t(\widetilde{\theta})-\varepsilon_t|=o_p(1)$.

This means residuals $\varepsilon_t(\widetilde{\theta})=\sqrt{\frac{h_t(\theta_0)}{h_t(\widetilde{\theta})}}\varepsilon_t$ are uniformly close to ε_t. Therefore for the weights

$\widehat{\widehat{w}}_t = \frac{1}{n}\frac{1}{1+\widehat{\lambda}_n' R_2(\varepsilon_t(\widetilde{\theta}))}$, define $B_n = \frac{1}{n}\sum_{t=1}^{n} R_2(\varepsilon_t(\widetilde{\theta}))R_2(\varepsilon_t(\widetilde{\theta}))^\top$, and we can see

$$\begin{aligned}
\widehat{\lambda}_n &= B_n^{-1}(\frac{1}{n}\sum_{t=1}^{n} R_2(\varepsilon_t(\widetilde{\theta}))) + o_p(n^{-1/2}) \\
&= B^{-1}\frac{1}{n}\sum_{t=1}^{n}[\begin{pmatrix}\varepsilon_t \\ \varepsilon_t^2 - 1\end{pmatrix} - \frac{1}{2}\begin{pmatrix}\varepsilon_t \\ 2\varepsilon_t^2\end{pmatrix}\frac{1}{h_t(\theta_0)}\frac{\partial h_t(\theta_0)}{\partial \theta^\top}(\widetilde{\theta} - \theta_0)] + o_p(n^{-1/2})
\end{aligned}$$

so

$$\begin{aligned}
\sqrt{n}\widehat{\lambda}_n &= B^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}[\begin{pmatrix}\varepsilon_t \\ \varepsilon_t^2 - 1\end{pmatrix} - \frac{1}{2}\begin{pmatrix}\varepsilon_t \\ 2\varepsilon_t^2\end{pmatrix}\frac{1}{h_t(\theta_0)}\frac{\partial h_t(\theta_0)}{\partial \theta^\top}(\widetilde{\theta} - \theta_0)] + o_p(1) \\
&= B^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n} R_2(\varepsilon_t) - \frac{1}{2}B^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\begin{pmatrix}\varepsilon_t \\ 2\varepsilon_t^2\end{pmatrix}\frac{1}{h_t(\theta_0)}\frac{\partial h_t(\theta_0)}{\partial \theta^\top}(\widetilde{\theta} - \theta_0) + o_p(1).
\end{aligned}$$

Hence,

$$\begin{aligned}
I_5 &= -\sqrt{n}\widehat{\lambda}_n^\top \frac{1}{n}\sum_{t=1}^{n}\{R_2(\varepsilon_t(\widetilde{\theta}))1(\varepsilon_t(\widetilde{\theta}) \le x)\} \\
&= -\{\frac{1}{\sqrt{n}}\sum_{t=1}^{n} R_2(\varepsilon_t)^\top B^{-1}A_x - \frac{1}{2}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\begin{pmatrix}\varepsilon_t & 2\varepsilon_t^2\end{pmatrix}\frac{1}{h_t(\theta_0)}\frac{\partial h_t(\theta_0)}{\partial \theta^\top}(\widetilde{\theta} - \theta_0)B^{-1}A_x\} + o_p(1),
\end{aligned}$$

where the last equality holds because of ergodic theorem: $n^{-1}\sum_{t=1}^{n}\{R_2(\varepsilon_t(\widetilde{\theta}))1(\varepsilon_t(\widetilde{\theta}) \le x)\} = A_x + o_p(1)$. So, uniformly in $x \in \mathbb{R}$,

$$\begin{aligned}
&\sqrt{n}(\widehat{\widehat{F}}_w(x) - F(x)) \\
= &\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\{[n\widehat{\widehat{w}}_t - 1]1(\varepsilon_t(\widetilde{\theta}) \le x)\} + \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\{1(\varepsilon_t(\widetilde{\theta}) \le x) - F(x)\} \\
= &-\frac{1}{\sqrt{n}}\sum_{t=1}^{n} A_x^\top B^{-1}R_2(\varepsilon_t) + \frac{1}{2}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\begin{pmatrix}\varepsilon_t & 2\varepsilon_t^2\end{pmatrix}\frac{1}{h_t(\theta_0)}\frac{\partial h_t(\theta_0)}{\partial \theta^\top}(\widetilde{\theta} - \theta_0)B^{-1}A_x + \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\{1(\varepsilon_t \le x) - \\
&+\frac{xf(x)}{2}H(\theta_0)^\top\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\chi_t(\theta_0) + o_p(1) \\
= &\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\{1(\varepsilon_t \le x) - F(x)\} + \{\frac{xf(x)}{2} + \begin{pmatrix}0 & 1\end{pmatrix}B^{-1}A_x\}H(\theta_0)^\top\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\chi_t(\theta_0) \\
&-\frac{1}{\sqrt{n}}\sum_{t=1}^{n} A_x^\top B^{-1}R_2(\varepsilon_t) + o_p(1),
\end{aligned}$$

because

$$
\begin{aligned}
& \frac{1}{2}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\begin{pmatrix} \varepsilon_t & 2\varepsilon_t^2 \end{pmatrix}\frac{1}{h_t(\theta_0)}\frac{\partial h_t(\theta_0)}{\partial \theta^\top}(\widetilde{\theta}-\theta_0)B^{-1}A_x \\
= & \frac{1}{2}\frac{1}{n}\sum_{t=1}^{n}\begin{pmatrix} \varepsilon_t & 2\varepsilon_t^2 \end{pmatrix}\frac{1}{h_t(\theta_0)}\frac{\partial h_t(\theta_0)}{\partial \theta^\top}\sqrt{n}(\widetilde{\theta}-\theta_0)B^{-1}A_x \\
= & \begin{pmatrix} 0 & 1 \end{pmatrix}B^{-1}A_x H(\theta_0)^\top\sqrt{n}(\widetilde{\theta}-\theta_0)+o_p(1) \\
= & \frac{a_2(x)-a_1(x)E[\varepsilon^3]}{E[\varepsilon^4]-1-E[\varepsilon^3]^2}H(\theta_0)^\top\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\chi_t(\theta_0)+o_p(1).
\end{aligned}
$$

■

Proof of Theorem 2. Lemma 1 and the Proposition 1 of Gill (1989) imply the results regarding VaR. Notice that, for any consistent distribution function estimator $\widetilde{F}(x)$ with associated quantile estimator $\widetilde{q}_\alpha=\widetilde{F}^{-1}(\alpha)$, the expected shortfall can be expressed as

$$
\alpha\widetilde{ES}_\alpha=\int_{-\infty}^{\widetilde{q}_\alpha}xd\widetilde{F}(x)=\widetilde{q}_\alpha\widetilde{F}(\widetilde{q}_\alpha)-\int_{-\infty}^{\widetilde{q}_\alpha}\widetilde{F}(x)dx=\alpha\widetilde{q}_\alpha-\int_{-\infty}^{\widetilde{q}_\alpha}\widetilde{F}(x)dx
$$

we can see that

$$
\begin{aligned}
& \alpha(\widetilde{ES}_\alpha-ES_\alpha) \\
= & \int_{-\infty}^{\widetilde{q}_\alpha}xd\widetilde{F}(x)-\int_{-\infty}^{q_\alpha}xdF(x) \\
= & \alpha\widetilde{q}_\alpha-\int_{-\infty}^{\widetilde{q}_\alpha}\widetilde{F}(x)dx-\alpha q_\alpha+\int_{-\infty}^{q_\alpha}F(x)dx \\
= & \int_{-\infty}^{q_\alpha}(F(x)-\widetilde{F}(x))dx+\alpha(\widetilde{q}_\alpha-q_\alpha)-\int_{q_\alpha}^{\widetilde{q}_\alpha}\widetilde{F}(x)dx \\
= & \int_{-\infty}^{q_\alpha}(F(x)-\widetilde{F}(x))dx+(\widetilde{q}_\alpha-q_\alpha)(\alpha-\widetilde{F}(\bar{\widetilde{q}}_\alpha)) \\
= & \int_{-\infty}^{q_\alpha}(F(x)-\widetilde{F}(x))dx+o_p(n^{-1/2}).
\end{aligned}
$$

As a result:

$$
\begin{aligned}
& \alpha(\widehat{ES}_\alpha-ES_\alpha) \\
= & \int_{-\infty}^{q_\alpha}(F(x)-\widehat{F}(x))dx+o_p(n^{-1/2}) \\
= & \int_{-\infty}^{q_\alpha}F(x)dx-\frac{1}{n}\sum_{t=1}^{n}(q_\alpha-\varepsilon_t)1(\varepsilon_t\le q_\alpha)+o_p(n^{-1/2})
\end{aligned}
$$

$$
\begin{aligned}
&\alpha(\widehat{ES}_{N\alpha} - ES_\alpha) \\
= \;& \int_{-\infty}^{q_\alpha} (F(x) - \widehat{F}_N(x))dx + o_p(n^{-1/2}) \\
= \;& -\frac{1}{n}\sum_{t=1}^{n} \int_{-\infty}^{q_\alpha} \left\{ 1(\varepsilon_t \le x) - F(x) + f(x)\varepsilon_t + \frac{xf(x)}{2}(\varepsilon_t^2 - 1) \right\} dx + o_p(n^{-1/2}) \\
= \;& -\frac{1}{n}\sum_{t=1}^{n} \left\{ \int_{-\infty}^{q_\alpha} 1(\varepsilon_t \le x)dx - \int_{-\infty}^{q_\alpha} F(x)dx + \varepsilon_t \int_{-\infty}^{q_\alpha} f(x)dx + (\varepsilon_t^2 - 1)\int_{-\infty}^{q_\alpha} \frac{xf(x)}{2}dx \right\} + o_p(n^{-1/2}) \\
= \;& -\frac{1}{n}\sum_{t=1}^{n} \left\{ (q_\alpha - \varepsilon_t)1(\varepsilon_t \le q_\alpha) - \int_{-\infty}^{q_\alpha} F(x)dx + \alpha\varepsilon_t + \frac{\varepsilon_t^2 - 1}{2}\int_{-\infty}^{q_\alpha} xf(x)dx \right\} + o_p(n^{-1/2})
\end{aligned}
$$

$$
\begin{aligned}
&\alpha(\widehat{ES}_{w\alpha} - ES_\alpha) \\
= \;& \int_{-\infty}^{q_\alpha} (F(x) - \widehat{F}_w(x))dx + o_p(n^{-1/2}) \\
= \;& -\frac{1}{n}\sum_{t=1}^{n} \int_{-\infty}^{q_\alpha} \left\{ 1(\varepsilon_t \le x) - F(x) - A_x^\top B^{-1} R_2(\varepsilon_t) \right\} dx + o_p(n^{-1/2}) \\
= \;& -\frac{1}{n}\sum_{t=1}^{n} \left\{ \int_{-\infty}^{q_\alpha} 1(\varepsilon_t \le x)dx - \int_{-\infty}^{q_\alpha} F(x)dx - \int_{-\infty}^{q_\alpha} A_x^\top B^{-1} R_2(\varepsilon_t)dx \right\} + o_p(n^{-1/2}) \\
= \;& -\frac{1}{n}\sum_{t=1}^{n} \left\{ (q_\alpha - \varepsilon_t)1(\varepsilon_t \le q_\alpha) - \int_{-\infty}^{q_\alpha} F(x)dx - R_2^\top(\varepsilon_t)B^{-1}\int_{-\infty}^{q_\alpha} A_x dx \right\} + o_p(n^{-1/2}).
\end{aligned}
$$

■

COROLLARY 3. *Suppose that the semiparametric efficient estimator that has influence function $\chi_t(\theta_0) = \psi_t(\theta_0)$ is used. Then, the process $\sqrt{n}(\widehat{\widehat{F}} - F)$ converges weakly to a zero-mean Gaussian process $\mathcal{Z}$ with covariance function Ω, the process $\sqrt{n}(\widehat{\widehat{F}}_N - F)$ converges weakly to a zero-mean Gaussian process $\mathcal{Z}_{\widehat{N}}$ with covariance function $\Omega_{\widehat{N}}$, and the process $\sqrt{n}(\widehat{\widehat{F}}_w - F)$ converges weakly to a zero-mean Gaussian process $\mathcal{Z}_{\widehat{w}}$ with covariance function*

$\Omega_{\widehat{w}}$, *where:*

$$
\begin{aligned}
\Omega(x,x') &= \operatorname{cov}(\mathcal{Z}(x),\mathcal{Z}(x')) \\
&= E\Big[[1(\varepsilon\le x)-F(x)+\frac{xf(x)}{2}(\varepsilon^2-1-\varepsilon E\varepsilon^3)] \\
&\quad\times[1(\varepsilon\le x')-F(x')+\frac{x'f(x')}{2}(\varepsilon^2-1-\varepsilon E\varepsilon^3)]\Big] \\
\Omega_{\widehat{N}}(x,x') &= \operatorname{cov}\left(\mathcal{Z}_{\widehat{N}}(x),\mathcal{Z}_{\widehat{N}}(x')\right) \\
&= E\Big[[1(\varepsilon\le x)-F(x)+[f(x)-\frac{xf(x)}{2}E\varepsilon^3]\varepsilon+xf(x)(\varepsilon^2-1)] \\
&\quad\cdot[1(\varepsilon\le x')-F(x')+[f(x')-\frac{x'f(x')}{2}E\varepsilon^3]\varepsilon+x'f(x')(\varepsilon^2-1)]\Big] \\
\Omega_{\widehat{w}}(x,x') &= \operatorname{cov}(\mathcal{Z}_{\widehat{w}}(x),\mathcal{Z}_{\widehat{w}}(x')) \\
&= E\Big[[1(\varepsilon\le x)-F(x)+\{\frac{xf(x)}{2}+\begin{pmatrix}0 & 1\end{pmatrix}B^{-1}A_x\}(\varepsilon^2-1-\varepsilon E\varepsilon^3)-A_x^\top B^{-1}R_2(\varepsilon)] \\
&\quad\cdot[1(\varepsilon\le x')-F(x')+\{\frac{x'f(x')}{2}+\begin{pmatrix}0 & 1\end{pmatrix}B^{-1}A_{x'}\}(\varepsilon^2-1-\varepsilon E\varepsilon^3)-A_{x'}^\top B^{-1}R_2(\varepsilon)]\Big].
\end{aligned}
$$

Denote $E[\varepsilon 1(\varepsilon\le x)]=a_1(x)$ and $E[(\varepsilon^2-1)1(\varepsilon\le x)]=a_2(x)$. The pointwise asymptotic variances are $\Omega_j(x)$, where:

$$
\begin{aligned}
\Omega_1(x) &= F(x)(1-F(x))+\frac{x^2f(x)^2[E\varepsilon^4-1-(E\varepsilon^3)^2]}{4}+xf(x)(a_2(x)-a_1(x)E\varepsilon^3) \\
\Omega_2(x) &= F(x)(1-F(x))+C_x+\frac{3x^2f(x)^2[E\varepsilon^4-1-(E\varepsilon^3)^2]}{4}+xf(x)(a_2(x)-a_1(x)E\varepsilon^3) \\
\Omega_3(x) &= F(x)(1-F(x))-A_x^\top B^{-1}A_x+\left\{\frac{xf(x)}{2}+\begin{pmatrix}0 & 1\end{pmatrix}B^{-1}A_x\right\}^2[E\varepsilon^4-1-(E\varepsilon^3)^2].
\end{aligned}
$$

It can be shown that $\Omega_1(x)-\Omega_3(x)=a_1^2(x)\ge 0$. As a result, $\widehat{\widehat{F}}_w(x)$ is uniformly more efficient than $\widehat{\widehat{F}}(x)$ and they are equally efficient at x where $E[\varepsilon 1(\varepsilon\le x)]=0$. It can also be shown that $\Omega_2(x)\ge\Omega_3(x)$.

Proof of Corollary 3. Since $H_t(\theta_0)=\binom{G_t(\theta_0)}{2/c}$, we know that

$$
H(\theta_0)^\top\begin{pmatrix} -\frac{1}{2}E[l_{1t}^*l_{1t}^{*\top}]^{-1}\{G_t-G\} & 0 \\ \frac{c_0}{4}G^\top E[l_{1t}^*l_{1t}^{*\top}]^{-1}\{G_t-G\} & \frac{c_0}{2}(-E\varepsilon^3,1)\end{pmatrix}\begin{pmatrix}R_3(\varepsilon_t)\\ R_2(\varepsilon_t)\end{pmatrix}=\varepsilon_t^2-1-\varepsilon_tE\varepsilon^3.
$$

Plug

$$
\chi_t(\theta_0)=\begin{pmatrix} -\frac{1}{2}E[l_{1t}^*l_{1t}^{*\top}]^{-1}\{G_t-G\} & 0 \\ \frac{c_0}{4}G^\top E[l_{1t}^*l_{1t}^{*\top}]^{-1}\{G_t-G\} & \frac{c_0}{2}(-E\varepsilon^3,1)\end{pmatrix}\begin{pmatrix}R_3(\varepsilon_t)\\ R_2(\varepsilon_t)\end{pmatrix}
$$

into the expressions in lemma 2 and get, uniformly in $x \in \mathbb{R}$,

$$\begin{aligned}
\widehat{\widehat{F}}(x) - F(x) &= \frac{1}{n}\sum_{t=1}^{n}\{1(\varepsilon_t \le x) - F(x) + \frac{xf(x)}{2}[\varepsilon_t^2 - 1 - \varepsilon_t E\varepsilon^3]\} + o_p(n^{-1/2}) \\
\widehat{\widehat{F}}_N(x) - F(x) &= \frac{1}{n}\sum_{t=1}^{n}\{1(\varepsilon_t \le x) - F(x) + [f(x) - \frac{xf(x)}{2}E\varepsilon^3]\varepsilon_t + xf(x)[\varepsilon_t^2 - 1]\} + o_p(n^{-1/2}) \\
\widehat{\widehat{F}}_w(x) - F(x) &= \frac{1}{n}\sum_{t=1}^{n}\{1(\varepsilon_t \le x) - F(x)\} + \{\frac{xf(x)}{2} + \begin{pmatrix} 0 & 1 \end{pmatrix} B^{-1}A_x\}[\varepsilon_t^2 - 1 - \varepsilon_t E\varepsilon^3] \\
&\quad -\frac{1}{n}\sum_{t=1}^{n} A_x^\top B^{-1} R_2(\varepsilon_t) + o_p(n^{-1/2}).
\end{aligned}$$

Due to moment constraints (2), the following holds:

$$\begin{aligned}
E[(1+\varepsilon\frac{f'(\varepsilon)}{f(\varepsilon)})1(\varepsilon &\le x)] = F(x) + \int_{-\infty}^{x} \varepsilon df(\varepsilon) = xf(x) \\
E[\varepsilon\frac{f'(\varepsilon)}{f(\varepsilon)}] &= \int \varepsilon df(\varepsilon) = -1 \\
E[\varepsilon^2\frac{f'(\varepsilon)}{f(\varepsilon)}] &= \int \varepsilon^2 df(\varepsilon) = 0 \\
E[\varepsilon^3\frac{f'(\varepsilon)}{f(\varepsilon)}] &= \int \varepsilon^3 df(\varepsilon) = -3.
\end{aligned}$$

These equations and CLT show that they have asymptotic variance as follows:

$$\begin{aligned}
\Omega_1 &= F(x)(1-F(x)) + \frac{x^2 f(x)^2}{4}[E\varepsilon^4 - 1 - (E\varepsilon^3)^2] + xf(x)(a_2(x) - a_1(x)E\varepsilon^3) \\
\Omega_2 &= F(x)(1-F(x)) + C_x + \frac{3x^2 f(x)^2[E\varepsilon^4 - 1 - (E\varepsilon^3)^2]}{4} + xf(x)(a_2(x) - a_1(x)E\varepsilon^3) \\
\Omega_3 &= F(x)(1-F(x)) - A_x^\top B^{-1}A_x + \{\frac{xf(x)}{2} + \begin{pmatrix} 0 & 1 \end{pmatrix} B^{-1}A_x\}^2[E\varepsilon^4 - 1 - (E\varepsilon^3)^2].
\end{aligned}$$

■

Proof of Theorem 3. Lemma 2 and the Proposition 1 of Gill (1989) imply above results for VaR. Similar to the proof of corollary 2, we have

$$
\begin{aligned}
& \alpha(\widehat{\widehat{ES}}_\alpha - ES_\alpha) \\
= & \int_{-\infty}^{\widehat{\widehat{q}}_\alpha} x d\widehat{\widehat{F}}(x) - \int_{-\infty}^{q_\alpha} x dF(x) \\
= & \int_{-\infty}^{q_\alpha} (F(x) - \widehat{\widehat{F}}(x))dx + \alpha(\widehat{\widehat{q}}_\alpha - q_\alpha) - \int_{q_\alpha}^{\widehat{\widehat{q}}_\alpha} \widehat{\widehat{F}}(x)dx \\
= & \int_{-\infty}^{q_\alpha} (F(x) - \widehat{\widehat{F}}(x))dx + \alpha(\widehat{\widehat{q}}_\alpha - q_\alpha) - (\widehat{\widehat{q}}_\alpha - q_\alpha)\widehat{\widehat{F}}(\widehat{\widehat{q}}_\alpha) \\
= & \int_{-\infty}^{q_\alpha} (F(x) - \widehat{\widehat{F}}(x))dx + o_p(n^{-1/2}).
\end{aligned}
$$

Then the theorem holds because of the following:

$$
\begin{aligned}
& \int_{-\infty}^{q_\alpha} (F(x) - \widehat{\widehat{F}}(x))dx \\
= & -\frac{1}{n}\sum_{t=1}^{n} \int_{-\infty}^{q_\alpha} (1(\varepsilon_t \le x) - F(x) + \frac{xf(x)}{2}(\varepsilon_t^2 - 1 - \varepsilon_t E\varepsilon^3))dx \\
= & -\frac{1}{n}\sum_{t=1}^{n} \left\{ (q_\alpha - \varepsilon_t)1(\varepsilon_t \le q_\alpha) - \int_{-\infty}^{q_\alpha} F(x)dx + (\varepsilon_t^2 - 1 - \varepsilon_t E\varepsilon^3)\int_{-\infty}^{q_\alpha} \frac{xf(x)}{2}dx \right\},
\end{aligned}
$$

$$
\begin{aligned}
& \int_{-\infty}^{q_\alpha} (F(x) - \widehat{\widehat{F}}_N(x))dx \\
= & -\frac{1}{n}\sum_{t=1}^{n} \int_{-\infty}^{q_\alpha} \left\{ 1(\varepsilon_t \le x) - F(x) + [f(x) - \frac{xf(x)}{2}E\varepsilon^3]\varepsilon_t + xf(x)(\varepsilon_t^2 - 1) \right\} dx \\
= & -\frac{1}{n}\sum_{t=1}^{n} \left\{ (q_\alpha - \varepsilon_t)1(\varepsilon_t \le q_\alpha) - \int_{-\infty}^{q_\alpha} F(x)dx + \varepsilon_t \int_{-\infty}^{q_\alpha} [f(x) - \frac{xf(x)}{2}E\varepsilon^3]dx + (\varepsilon_t^2 - 1)\int_{-\infty}^{q_\alpha} xf(x)dx \right\}.
\end{aligned}
$$

$$
\begin{aligned}
& \int_{-\infty}^{q_\alpha} (F(x) - \widehat{\widehat{F}}_w(x))dx \\
= & -\frac{1}{n}\sum_{t=1}^{n} \int_{-\infty}^{q_\alpha} \left\{ 1(\varepsilon_t \le x) - F(x) + [\frac{xf(x)}{2} + \begin{pmatrix} 0 & 1 \end{pmatrix} B^{-1}A_x](\varepsilon_t^2 - 1 - \varepsilon_t E\varepsilon^3) - A_x^\top B^{-1} R_2(\varepsilon_t) \right\} dx \\
= & -\frac{1}{n}\sum_{t=1}^{n} \left\{ (q_\alpha - \varepsilon_t)1(\varepsilon_t \le q_\alpha) - \int_{-\infty}^{q_\alpha} F(x)dx + (\varepsilon_t^2 - 1 - \varepsilon_t E\varepsilon^3)\int_{-\infty}^{q_\alpha} [\frac{xf(x)}{2} + \begin{pmatrix} 0 & 1 \end{pmatrix} B^{-1}A_x]dx \right. \\
& \left. - R_2^\top(\varepsilon_t) B^{-1} \int_{-\infty}^{q_\alpha} A_x dx \right\}.
\end{aligned}
$$

$\blacksquare$

Proof of Theorem 4. Since $R_1(\varepsilon) = 1(\varepsilon \le q_\alpha) - \alpha$, $R_2(\varepsilon) = (\varepsilon, \varepsilon^2 - 1)^\top$, $R_3(\varepsilon) = 1 + \varepsilon\frac{f'(\varepsilon)}{f(\varepsilon)}$, and $R(\varepsilon) = (R_1(\varepsilon), R_2(\varepsilon)^\top, R_3(\varepsilon))^\top$. It is seen that

$$E[R_1(\varepsilon_t)|\mathcal{F}_{t-1}] = E[R_2(\varepsilon_t)|\mathcal{F}_{t-1}] = E[R_3(\varepsilon_t)|\mathcal{F}_{t-1}] = 0,$$

which implies that $\{Z_s\}$ is Martingale Difference Series. From Theorem 2, we have

$$\sqrt{n}\begin{pmatrix} \widehat{\theta} - \theta_0 \\ \widehat{\widehat{q}}_{w\alpha} - q_\alpha \end{pmatrix} = \frac{1}{\sqrt{n}}\sum_{s=1}^{n} Z_s + o_p(1),$$

$$\begin{aligned}
Z_s &= \Psi_s R(\varepsilon_s) \\
\Psi_t &= \begin{pmatrix} 0 & 0 & -\frac{1}{2}E[l_{1t}^* l_{1t}^{*\top}]^{-1}\{G_t - G\} \\ 0 & \frac{c_0}{2}(-E\varepsilon^3, 1) & \frac{c_0}{4}G^\top E[l_{1t}^* l_{1t}^{*\top}]^{-1}\{G_t - G\} \\ \frac{-1}{f(q_\alpha)} & \frac{A_{q_\alpha}^\top B^{-1}}{f(q_\alpha)} - [\frac{q_\alpha}{2} + \frac{\begin{pmatrix} 0 & 1 \end{pmatrix} B^{-1} A_{q_\alpha}}{f(q_\alpha)}](-E\varepsilon^3, 1) & 0 \end{pmatrix}.
\end{aligned}$$

Since: $E[R_1(\varepsilon)R_2(\varepsilon)] = A_{q_\alpha}$, $E[R_1(\varepsilon)R_3(\varepsilon)] = q_\alpha f(q_\alpha)$, and $E[R_2(\varepsilon)R_3(\varepsilon)] = \begin{pmatrix} 0 & -2 \end{pmatrix}^\top$, we have

$$\begin{aligned}
& \Omega_Z \\
= \; & E[\Psi_s R(\varepsilon_s) R(\varepsilon_s)^\top \Psi_s^\top] \\
= \; & E[\Psi_s \begin{pmatrix} \alpha(1-\alpha) & A_{q_\alpha}^\top & q_\alpha f(q_\alpha) \\ A_{q_\alpha} & B & \begin{pmatrix} 0 & -2 \end{pmatrix}^\top \\ q_\alpha f(q_\alpha) & \begin{pmatrix} 0 & -2 \end{pmatrix} & E[(1+\varepsilon f'/f)^2] \end{pmatrix} \Psi_s^\top] \\
= \; & \begin{pmatrix} E[l_{1t}^* l_{1t}^{*\top}]^{-1} & -\frac{c_0}{2}E[l_{1t}^* l_{1t}^{*\top}]^{-1}G & 0 \\ -\frac{c_0}{2}G^\top E[l_{1t}^* l_{1t}^{*\top}]^{-1} & \frac{c_0^2[E\varepsilon^4 - 1 - (E\varepsilon^3)^2] + G^\top E[l_{1t}^* l_{1t}^{*\top}]^{-1}G}{4} & \frac{c_0[E\varepsilon^4 - 1 - (E\varepsilon^3)^2][\frac{q_\alpha}{2} + \frac{\begin{pmatrix} 0 & 1 \end{pmatrix} B^{-1}A_{q_\alpha}}{f(q_\alpha)}]}{2} \\ 0 & \frac{c_0[E\varepsilon^4 - 1 - (E\varepsilon^3)^2][\frac{q_\alpha}{2} + \frac{\begin{pmatrix} 0 & 1 \end{pmatrix} B^{-1}A_{q_\alpha}}{f(q_\alpha)}]}{2} & \Omega_{\widehat{\widehat{q}}_{w\alpha}} \end{pmatrix}.
\end{aligned}$$

Due to Taylor expansion,

$$
\begin{aligned}
& \sqrt{n}(\widehat{\xi}_{n,t} - \xi_t) \\
= \; & \sqrt{n}(h_t^{1/2}(\widehat{\theta})\widehat{\widehat{q}}_{w\alpha} - h_t^{1/2}(\theta_0)\widehat{\widehat{q}}_{w\alpha} + h_t^{1/2}(\theta_0)\widehat{\widehat{q}}_{w\alpha} - h_t^{1/2}(\theta_0)q_\alpha) \\
= \; & \sqrt{n}\frac{q_\alpha}{2h_t^{1/2}(\theta_0)}\frac{\partial h_t^{1/2}(\theta_0)}{\partial \theta^\top}(\widehat{\theta} - \theta_0) + h_t^{1/2}(\theta_0)\sqrt{n}(\widehat{\widehat{q}}_{w\alpha} - q_\alpha) + o_p(1) \\
= \; & \sqrt{n}\begin{pmatrix} \frac{q_\alpha}{2h_t^{1/2}(\theta_0)}\frac{\partial h_t(\theta_0)}{\partial \theta^\top} & h_t^{1/2}(\theta_0) \end{pmatrix}\begin{pmatrix} \widehat{\theta} - \theta_0 \\ \widehat{\widehat{q}}_{w\alpha} - q_\alpha \end{pmatrix} + o_p(1) \\
= \; & \frac{1}{\sqrt{n}}\sum_{s=1}^{n} W_t^\top Z_s + o_p(1),
\end{aligned}
$$

where $W_t = \begin{pmatrix} \frac{q_\alpha}{2h_t^{1/2}(\theta_0)}\frac{\partial h_t(\theta_0)}{\partial \theta^\top} & h_t^{1/2}(\theta_0) \end{pmatrix}^\top$. Denote $X_{ns} = n^{-1/2}W_t^\top Z_s$, it follows that

$$
\sum_{s=1}^{n} X_{ns}^2 = W_t^\top \frac{1}{n}\sum_{s=1}^{n} Z_s Z_s^\top W_t \to^p W_t^\top \Omega_Z W_t.
$$

From Martingale Central Limit theorem, we can see that

$$
\frac{\sum_{s=1}^{n} X_{ns}}{\sqrt{\sum_{s=1}^{n} X_{ns}^2}} \Longrightarrow N(0,1)
$$

and $\sqrt{n}(\widehat{\xi}_{n,t} - \xi_t) \Longrightarrow N(0, \omega_{\xi t})$, where

$$
\begin{aligned}
\omega_{\xi t} &= W_t^\top \Omega_Z W_t \\
&= h_t(\theta_0)\{\frac{q_\alpha^2}{4}(G_t^\top - G)E[l_{1t}^* l_{1t}^{*\top}]^{-1}(G_t - G) \\
&\quad + \frac{\alpha(1-\alpha) - A_{q_\alpha}^\top B^{-1} A_{q_\alpha}}{f(q_\alpha)^2} + [q_\alpha + \frac{\begin{pmatrix} 0 & 1 \end{pmatrix} B^{-1} A_{q_\alpha}}{f(q_\alpha)}]^2[E\varepsilon^4 - 1 - (E\varepsilon^3)^2]\}.
\end{aligned}
$$

Denote a truncated version of h_{n+1}as

$$
h_{n+1}^* = \frac{c^2}{1-b} + ac^2\sum_{j=1}^{m} b^{j-1} y_{n+1-j}^2
$$

where the truncation order is $m = \log n$. As a result, the approximation error is of order $o_p(1)$:

$$
h_{n+1} - h_{n+1}^* = ac^2 \sum_{j=m+1}^{\infty} b^{j-1} y_{n+1-j}^2 = O_p(b^m)
$$

Similarly, we can show that $\frac{\partial h_{n+1}}{\partial\beta} - \frac{\partial h^*_{n+1}}{\partial\beta} = O_p(b^m)$. Consequently, $W_{n+1} - W^*_{n+1} = O_p(b^m)$.

At the same time, we have the following truncation approximation

$$\begin{aligned} & \frac{1}{n}\sum_{s=1}^{n} Z_s = \frac{1}{\sqrt{n}}\sum_{t=1}^{n-m}\Psi_s R(\varepsilon_s) + \sqrt{\frac{m-1}{n}}\frac{1}{\sqrt{m-1}}\sum_{t=n-m+1}^{n}\Psi_s R(\varepsilon_s) \\ = \quad & \frac{1}{\sqrt{n}}\sum_{t=1}^{n-m}\Psi_s R(\varepsilon_s) + o_p(1) \end{aligned}$$

As $\{\varepsilon_s\}$ is iid sequence, we can draw the conclusion that $W_{n+1} \rightarrow^p W^*_{n+1} \perp \frac{1}{\sqrt{n}}\sum\limits_{t=1}^{n-m}\Psi_s R(\varepsilon_s)$.

The above argument applies to $\widehat{\chi}_{n,t} = h_t^{1/2}(\widehat{\theta})\widehat{\widehat{ES}}_{w\alpha}$ as:

$$\begin{aligned} & \sqrt{n}(\widehat{\chi}_{n,t} - \chi_t) \\ = \quad & \sqrt{n}(h_t^{1/2}(\widehat{\theta})\widehat{\widehat{ES}}_{w\alpha} - h_t^{1/2}(\theta_0)\widehat{\widehat{ES}}_{w\alpha} + h_t^{1/2}(\theta_0)\widehat{\widehat{ES}}_{w\alpha} - h_t^{1/2}(\theta_0)ES_\alpha) \end{aligned}$$

$$\widehat{\widehat{ES}}_{w\alpha} - ES_\alpha = \frac{1}{n}\sum_{t=1}^{n}\left\{\frac{1}{\alpha}(\varepsilon_t - q_\alpha)1(\varepsilon_t \le q_\alpha) + \frac{1}{\alpha}\int_{-\infty}^{q_\alpha} F(x)dx + \frac{1}{\alpha}C^\top R_2(\varepsilon_t)\right\} + o_p(n^{-1/2}),$$

where $C = \int_{-\infty}^{q_\alpha}[\frac{xf(x)}{2} + \begin{pmatrix} 0 & 1 \end{pmatrix} B^{-1}A_x]dx(E\varepsilon^3, -1) + \int_{-\infty}^{q_\alpha} A_x^\top dx B^{-1}$. Notice that,

$$\operatorname{cov}(\widehat{\widehat{ES}}_{w\alpha} - ES_\alpha, \widehat{\theta} - \theta_0) = \begin{pmatrix} 0 \\ \frac{c_0(-E\varepsilon^3,1)}{2\alpha} E\left[\left\{(\varepsilon_t - q_\alpha)1(\varepsilon_t \le q_\alpha) + C^\top R_2(\varepsilon_t)\right\} R_2(\varepsilon_t)\right] \end{pmatrix},$$

so the conclusion regarding $\widehat{\chi}_{n,t}$ holds.

■

B Tables and Figures

Table 1. Integrated Mean Squared Error ($\times 10^{-3}$) of Distribution Function Estimators

	$n = 500$			$n = 1000$		
	$\widehat{F}(x)$	$\widehat{F}_N(x)$	$\widehat{F}_w(x)$	$\widehat{F}(x)$	$\widehat{F}_N(x)$	$\widehat{F}_w(x)$
N	0.3365	0.1199	0.1212	0.1616	0.0580	0.0583
MN	0.3286	0.1412	0.0916	0.1622	0.0687	0.0462
L	0.3313	0.2188	0.1603	0.1692	0.1092	0.0810
$t(5)$	0.3419	0.2157	0.1635	0.1657	0.1055	0.0797
$t(7)$	0.3255	0.1594	0.1458	0.1695	0.0791	0.0708
$t(9)$	0.3336	0.1439	0.1361	0.1664	0.0730	0.0687
χ^2_6	0.3308	0.1479	0.1217	0.1692	0.0721	0.0605
χ^2_{12}	0.3297	0.1335	0.1213	0.1692	0.0654	0.0595

Table 2. Comparison of quantile estimators for $q_{0.01}$ (true errors are available)

	$n = 500$						$n = 1000$					
	$Bias(\times 10^{-3})$			$MSE(\times 10^{-3})$			$Bias(\times 10^{-3})$			$MSE(\times 10^{-3})$		
	$\widehat{q}_\alpha$	$\widehat{q}_{N\alpha}$	$\widehat{q}_{w\alpha}$	$\widehat{q}_\alpha$	$\widehat{q}_{N\alpha}$	$\widehat{q}_{w\alpha}$	$\widehat{q}_\alpha$	$\widehat{q}_{N\alpha}$	$\widehat{q}_{w\alpha}$	$\widehat{q}_\alpha$	$\widehat{q}_{N\alpha}$	$\widehat{q}_{w\alpha}$
N	-2.1	-0.8	3.7	27.1	19.4	22.2	-0.7	-1.1	0.6	14.1	9.9	11.1
MN	-1.8	-2.5	-0.5	6.3	5.8	5.6	-1.3	-0.9	-0.4	3.3	3.1	2.9
L	-3.7	3.3	17.9	94.3	60.9	68.6	11.9	9.4	11.8	47.8	31.6	31.9
$t(5)$	25.1	25.3	40.7	102.5	67.5	72.8	8.4	11.4	22.2	50.4	34.1	35.1
$t(7)$	8.4	10.5	21.1	65.4	45.3	50.9	-1.7	0.1	4.9	34.9	23.1	23.6
$t(9)$	10.5	9.3	15.4	56.1	36.7	41.8	11.4	8.1	8.9	30.9	20.0	21.3
χ^2_6	-3.4	0.4	-1.7	1.7	3.7	1.7	-1.0	0.5	-0.4	0.9	1.8	0.8
χ^2_{12}	-4.6	-3.5	-3.1	4.9	6.0	4.8	-2.8	-0.9	-1.7	2.5	3.1	2.3

Table 3. Comparison of quantile estimators for $q_{0.05}$ (true errors are available)

	$n=500$						$n=1000$					
	$Bias(\times 10^{-3})$			$MSE(\times 10^{-3})$			$Bias(\times 10^{-3})$			$MSE(\times 10^{-3})$		
	$\widehat{q}_\alpha$	$\widehat{q}_{N\alpha}$	$\widehat{q}_{w\alpha}$	$\widehat{q}_\alpha$	$\widehat{q}_{N\alpha}$	$\widehat{q}_{w\alpha}$	$\widehat{q}_\alpha$	$\widehat{q}_{N\alpha}$	$\widehat{q}_{w\alpha}$	$\widehat{q}_\alpha$	$\widehat{q}_{N\alpha}$	$\widehat{q}_{w\alpha}$
N	-3.6	-1.6	-1.0	8.9	4.2	4.4	-1.2	-0.6	0.1	4.5	2.0	2.0
MN	-1.4	0.0	-0.1	2.4	2.0	1.4	-0.8	-1.1	0.0	1.3	1.0	0.7
L	0.8	4.6	7.9	18.2	8.8	8.8	2.6	2.7	4.6	9.4	4.9	4.6
$t(5)$	0.7	5.0	12.1	13.6	9.0	8.1	-1.4	2.7	4.9	6.9	4.8	3.9
$t(7)$	-2.2	1.3	5.3	12.2	6.3	6.7	0.6	2.9	3.5	6.3	3.3	3.2
$t(9)$	-2.6	0.2	1.9	11.7	6.0	6.0	3.5	2.1	3.1	5.8	2.9	2.8
χ^2_6	-1.4	1.3	0.4	1.5	2.1	1.1	-2.3	0.2	-1.0	0.7	1.1	0.6
χ^2_{12}	-1.7	-0.7	1.1	2.8	2.4	1.9	-2.0	-1.2	-0.8	1.4	1.3	1.0

Table 4. Integrated Mean Squared Error ($\times 10^{-3}$), $c = 1, a = 0.3, b = 0.6$.

	$n = 500$			$n = 1000$		
	$\widehat{\widehat{F}}(x)$	$\widehat{\widehat{F}}_N(x)$	$\widehat{\widehat{F}}_w(x)$	$\widehat{\widehat{F}}(x)$	$\widehat{\widehat{F}}_N(x)$	$\widehat{\widehat{F}}_w(x)$
N	0.3018	0.1196	0.1193	0.1498	0.0595	0.0595
MN	0.2954	0.1445	0.0940	0.1508	0.0726	0.0468
L	0.3249	0.2188	0.1699	0.1653	0.1089	0.0848
$t(5)$	0.3551	0.2069	0.1859	0.1751	0.1031	0.0951
$t(7)$	0.3211	0.1550	0.1490	0.1631	0.0794	0.0762
$t(9)$	0.3176	0.1428	0.1389	0.1599	0.0717	0.0703
χ^2_6	0.3999	0.1434	0.1548	0.2080	0.0739	0.0808
χ^2_{12}	0.3415	0.1302	0.1333	0.1745	0.0664	0.0678

Table 5. Integrated Mean Squared Error ($\times 10^{-3}$), $c = 1, a = 0.1, b = 0.8$.

	$n = 500$			$n = 1000$		
	$\widehat{\widehat{F}}(x)$	$\widehat{\widehat{F}}_N(x)$	$\widehat{\widehat{F}}_w(x)$	$\widehat{\widehat{F}}(x)$	$\widehat{\widehat{F}}_N(x)$	$\widehat{\widehat{F}}_w(x)$
N	0.3023	0.1195	0.1193	0.1499	0.0594	0.0595
MN	0.2960	0.1442	0.0937	0.1511	0.0726	0.0468
L	0.3250	0.2185	0.1700	0.1655	0.1088	0.0848
$t(5)$	0.3591	0.2067	0.1891	0.1763	0.1029	0.0962
$t(7)$	0.3222	0.1550	0.1495	0.1635	0.0792	0.0764
$t(9)$	0.3187	0.1428	0.1396	0.1600	0.0715	0.0702
χ^2_6	0.4014	0.1430	0.1547	0.2076	0.0738	0.0807
χ^2_{12}	0.3422	0.1301	0.1335	0.1740	0.0663	0.0672

Table 6. Integrated Mean Squared Error ($\times 10^{-3}$), $c = 1, a = 0.05, b = 0.9$.

	$n = 500$			$n = 1000$		
	$\widehat{\widehat{F}}(x)$	$\widehat{\widehat{F}}_N(x)$	$\widehat{\widehat{F}}_w(x)$	$\widehat{\widehat{F}}(x)$	$\widehat{\widehat{F}}_N(x)$	$\widehat{\widehat{F}}_w(x)$
N	0.3026	0.1193	0.1191	0.1500	0.0594	0.0595
MN	0.2964	0.1440	0.0937	0.1511	0.0727	0.0468
L	0.3255	0.2182	0.1700	0.1656	0.1088	0.0848
$t(5)$	0.3607	0.2069	0.1902	0.1769	0.1034	0.0968
$t(7)$	0.3232	0.1557	0.1502	0.1636	0.0791	0.0763
$t(9)$	0.3187	0.1428	0.1393	0.1602	0.0715	0.0702
χ^2_6	0.4021	0.1425	0.1548	0.2079	0.0737	0.0808
χ^2_{12}	0.3432	0.1299	0.1336	0.1741	0.0663	0.0673

Table 7. Comparison of quantile estimators for $q_{0.01}$, with $c = 1, a = 0.05, b = 0.9$

	$n = 500$						$n = 1000$					
	$Bias(\times 10^{-3})$			$MSE(\times 10^{-3})$			$Bias(\times 10^{-3})$			$MSE(\times 10^{-3})$		
	$\widehat{\widehat{q}}_{\alpha}$	$\widehat{\widehat{q}}_{N\alpha}$	$\widehat{\widehat{q}}_{w\alpha}$	$\widehat{\widehat{q}}_{\alpha}$	$\widehat{\widehat{q}}_{N\alpha}$	$\widehat{\widehat{q}}_{w\alpha}$	$\widehat{\widehat{q}}_{\alpha}$	$\widehat{\widehat{q}}_{N\alpha}$	$\widehat{\widehat{q}}_{w\alpha}$	$\widehat{\widehat{q}}_{\alpha}$	$\widehat{\widehat{q}}_{N\alpha}$	$\widehat{\widehat{q}}_{w\alpha}$
N	-2.6	-5.1	-3.5	21.2	18.6	20.6	-2.5	-4.9	-4.3	10.8	9.6	10.0
MN	-2.3	-3.9	-3.8	5.9	6.1	5.8	-0.8	-2.9	-1.8	3.0	3.2	3.0
L	0.8	-3.7	-1.8	61.2	54.0	58.1	3.3	1.2	1.1	35.1	31.0	29.8
$t(5)$	28.4	23.5	22.5	73.8	65.5	66.8	9.2	9.0	12.1	36.9	33.1	32.5
$t(7)$	5.5	1.3	1.2	48.9	43.1	45.7	1.4	0.3	-0.3	25.7	22.6	22.2
$t(9)$	1.0	-0.4	1.8	39.8	34.8	38.9	1.2	-1.2	-1.2	20.2	18.0	18.9
χ^2_6	27.3	21.4	25.8	7.0	4.5	5.9	13.5	10.5	13.3	3.4	2.1	2.8
χ^2_{12}	13.9	10.0	13.4	8.7	6.5	7.4	8.5	5.5	7.8	4.1	3.1	3.4

Table 8. Comparison of quantile estimators for $q_{0.01}$, with $c = 1, a = 0.3, b = 0.6$

	$n = 500$						$n = 1000$					
	$Bias(\times 10^{-3})$			$MSE(\times 10^{-3})$			$Bias(\times 10^{-3})$			$MSE(\times 10^{-3})$		
	$\widehat{\widehat{q}}_{\alpha}$	$\widehat{\widehat{q}}_{N\alpha}$	$\widehat{\widehat{q}}_{w\alpha}$	$\widehat{\widehat{q}}_{\alpha}$	$\widehat{\widehat{q}}_{N\alpha}$	$\widehat{\widehat{q}}_{w\alpha}$	$\widehat{\widehat{q}}_{\alpha}$	$\widehat{\widehat{q}}_{N\alpha}$	$\widehat{\widehat{q}}_{w\alpha}$	$\widehat{\widehat{q}}_{\alpha}$	$\widehat{\widehat{q}}_{N\alpha}$	$\widehat{\widehat{q}}_{w\alpha}$
N	-2.3	-4.3	-2.4	21.5	18.9	20.7	-1.6	-4.0	-3.8	11.0	9.9	10.3
MN	-2.4	-3.6	-3.7	5.8	6.1	5.7	-0.2	-2.2	-1.4	3.0	3.3	2.9
L	1.1	-3.1	-0.9	60.7	53.5	57.3	1.7	-0.2	0.4	34.8	30.5	29.7
$t(5)$	33.7	28.6	28.3	70.5	62.8	67.4	9.0	8.2	11.9	36.3	32.7	32.1
$t(7)$	4.5	0.8	-0.1	48.2	42.4	44.4	1.8	1.0	1.5	26.0	23.0	23.2
$t(9)$	2.1	1.5	3.2	40.0	35.1	38.8	0.6	-1.3	-1.9	20.4	18.2	18.8
χ^2_6	31.6	25.7	29.8	7.6	5.0	6.3	14.6	11.9	14.1	3.4	2.2	2.8
χ^2_{12}	16.2	12.6	15.8	8.7	6.6	7.5	9.8	7.2	9.1	4.2	3.2	3.5

Table 9. Comparison of quantile estimators for $q_{0.01}$, with $c = 1, a = 0.1, b = 0.8$

	$n = 500$						$n = 1000$					
	$Bias(\times 10^{-3})$			$MSE(\times 10^{-3})$			$Bias(\times 10^{-3})$			$MSE(\times 10^{-3})$		
	$\widehat{\widehat{q}}_{\alpha}$	$\widehat{\widehat{q}}_{N\alpha}$	$\widehat{\widehat{q}}_{w\alpha}$	$\widehat{\widehat{q}}_{\alpha}$	$\widehat{\widehat{q}}_{N\alpha}$	$\widehat{\widehat{q}}_{w\alpha}$	$\widehat{\widehat{q}}_{\alpha}$	$\widehat{\widehat{q}}_{N\alpha}$	$\widehat{\widehat{q}}_{w\alpha}$	$\widehat{\widehat{q}}_{\alpha}$	$\widehat{\widehat{q}}_{N\alpha}$	$\widehat{\widehat{q}}_{w\alpha}$
N	-3.1	-5.5	-3.6	21.4	18.9	20.5	-3.1	-4.0	-2.2	11.2	10.1	10.5
MN	-2.5	-4.0	-3.9	5.8	6.0	5.7	-1.5	-3.2	-2.1	2.9	3.2	2.8
L	-0.0	-3.8	-0.8	60.8	53.6	58.0	-0.9	-1.2	2.3	37.0	33.1	33.0
$t(5)$	29.7	25.7	22.1	71.7	64.0	67.4	9.2	9.1	14.0	37.0	33.3	34.4
$t(7)$	4.3	0.7	-0.1	48.4	42.7	45.4	4.7	3.6	4.9	26.4	23.5	23.3
$t(9)$	0.7	-0.1	2.3	39.9	35.0	38.4	3.3	1.1	1.2	20.3	18.2	18.7
χ^2_6	29.7	23.7	28.5	7.3	4.7	6.1	16.6	13.0	15.5	3.3	2.1	2.7
χ^2_{12}	15.3	11.6	14.8	8.7	6.6	7.6	8.7	5.8	8.0	4.4	3.2	3.6

Table 10. Comparison of quantile estimators for $q_{0.05}$, with $c = 1, a = 0.3, b = 0.6$

	$n = 500$						$n = 1000$					
	$Bias(\times 10^{-3})$			$MSE(\times 10^{-3})$			$Bias(\times 10^{-3})$			$MSE(\times 10^{-3})$		
	$\widehat{\widehat{q}}_{\alpha}$	$\widehat{\widehat{q}}_{N\alpha}$	$\widehat{\widehat{q}}_{w\alpha}$	$\widehat{\widehat{q}}_{\alpha}$	$\widehat{\widehat{q}}_{N\alpha}$	$\widehat{\widehat{q}}_{w\alpha}$	$\widehat{\widehat{q}}_{\alpha}$	$\widehat{\widehat{q}}_{N\alpha}$	$\widehat{\widehat{q}}_{w\alpha}$	$\widehat{\widehat{q}}_{\alpha}$	$\widehat{\widehat{q}}_{N\alpha}$	$\widehat{\widehat{q}}_{w\alpha}$
N	0.3	0.9	2.2	6.2	4.2	4.3	-0.7	-1.6	-0.6	3.0	2.1	2.1
MN	0.6	-1.6	0.1	1.8	2.1	1.5	-0.0	-1.0	-0.2	0.8	1.0	0.7
L	9.3	9.8	13.0	14.1	9.7	9.3	4.0	3.0	4.2	6.8	4.4	4.1
$t(5)$	19.5	15.9	18.4	12.3	8.6	8.4	8.2	7.1	9.0	6.1	4.5	4.4
$t(7)$	4.9	5.0	8.2	9.5	6.3	6.3	6.2	5.0	5.7	5.0	3.4	3.4
$t(9)$	5.5	4.6	6.6	8.4	5.7	5.8	4.5	3.1	4.1	4.6	3.1	3.0
χ^2_6	11.7	8.4	10.8	4.4	2.1	2.7	6.7	3.9	5.7	2.2	1.1	1.4
χ^2_{12}	7.0	2.4	5.5	4.7	2.4	2.7	4.1	2.2	3.6	2.3	1.3	1.4

Table 11. Comparison of quantile estimators for $q_{0.05}$, with $c = 1, a = 0.1, b = 0.8$

	$n = 500$						$n = 1000$					
	$Bias(\times 10^{-3})$			$MSE(\times 10^{-3})$			$Bias(\times 10^{-3})$			$MSE(\times 10^{-3})$		
	$\widehat{\widehat{q}}_{\alpha}$	$\widehat{\widehat{q}}_{N\alpha}$	$\widehat{\widehat{q}}_{w\alpha}$	$\widehat{\widehat{q}}_{\alpha}$	$\widehat{\widehat{q}}_{N\alpha}$	$\widehat{\widehat{q}}_{w\alpha}$	$\widehat{\widehat{q}}_{\alpha}$	$\widehat{\widehat{q}}_{N\alpha}$	$\widehat{\widehat{q}}_{w\alpha}$	$\widehat{\widehat{q}}_{\alpha}$	$\widehat{\widehat{q}}_{N\alpha}$	$\widehat{\widehat{q}}_{w\alpha}$
N	3.8	2.0	3.6	6.1	4.1	4.2	1.2	0.6	1.4	3.1	2.1	2.1
MN	-0.0	-1.2	-0.7	1.8	2.1	1.5	-0.4	-1.9	-0.9	0.9	1.1	0.7
L	14.4	11.7	13.9	13.7	9.0	8.7	6.4	6.6	8.2	7.3	4.9	4.6
$t(5)$	18.9	16.3	18.5	12.1	8.7	8.6	8.8	9.1	10.5	6.6	4.9	4.8
$t(7)$	6.0	4.9	8.0	9.2	6.3	6.4	4.8	4.2	5.3	4.9	3.3	3.2
$t(9)$	5.1	4.0	5.7	8.6	5.7	5.8	6.6	4.9	5.4	4.2	2.9	3.0
χ^2_6	11.9	7.5	10.8	4.5	2.3	2.8	7.9	4.4	6.6	2.2	1.0	1.3
χ^2_{12}	8.3	5.8	7.6	4.7	2.5	2.9	4.9	2.2	3.9	2.3	1.3	1.4

Table 12. Comparison of quantile estimators for $q_{0.05}$, with $c = 1, a = 0.05, b = 0.9$

	$n = 500$						$n = 1000$					
	$Bias(\times 10^{-3})$			$MSE(\times 10^{-3})$			$Bias(\times 10^{-3})$			$MSE(\times 10^{-3})$		
	$\widehat{\widehat{q}}_{\alpha}$	$\widehat{\widehat{q}}_{N\alpha}$	$\widehat{\widehat{q}}_{w\alpha}$	$\widehat{\widehat{q}}_{\alpha}$	$\widehat{\widehat{q}}_{N\alpha}$	$\widehat{\widehat{q}}_{w\alpha}$	$\widehat{\widehat{q}}_{\alpha}$	$\widehat{\widehat{q}}_{N\alpha}$	$\widehat{\widehat{q}}_{w\alpha}$	$\widehat{\widehat{q}}_{\alpha}$	$\widehat{\widehat{q}}_{N\alpha}$	$\widehat{\widehat{q}}_{w\alpha}$
N	1.2	1.3	3.3	6.1	4.2	4.3	-0.3	-1.4	-0.3	3.0	2.1	2.1
MN	1.0	-1.6	-0.1	1.8	2.0	1.5	0.1	-1.0	-0.5	0.8	1.0	0.7
L	8.9	9.0	12.1	14.1	9.7	9.2	4.1	2.9	4.1	6.9	4.4	4.2
$t(5)$	16.1	13.2	15.4	12.8	8.8	8.9	8.0	7.0	8.8	6.3	4.6	4.5
$t(7)$	4.2	3.8	7.1	9.4	6.2	6.2	5.7	4.2	5.5	5.1	3.4	3.4
$t(9)$	4.5	3.2	5.5	8.5	5.6	5.7	5.1	3.5	4.7	4.6	3.1	3.1
χ^2_6	11.2	7.8	10.1	4.5	2.1	2.7	6.0	3.2	5.1	2.2	1.1	1.4
χ^2_{12}	6.7	1.8	4.9	4.8	2.4	2.8	3.7	1.7	3.2	2.4	1.3	1.4

Table 13. Comparison of quantile estimators, with $c = 1, a = 0.05, b = 0.9, n = 1000, s = 500$

	$q_{0.01}$						$q_{0.05}$					
	$Bias(\times 10^{-3})$			$MSE(\times 10^{-3})$			$Bias(\times 10^{-3})$			$MSE(\times 10^{-3})$		
	$\widehat{\widehat{q}}_{\alpha}$	$\widehat{\widehat{q}}_{N\alpha}$	$\widehat{\widehat{q}}_{w\alpha}$	$\widehat{\widehat{q}}_{\alpha}$	$\widehat{\widehat{q}}_{N\alpha}$	$\widehat{\widehat{q}}_{w\alpha}$	$\widehat{\widehat{q}}_{\alpha}$	$\widehat{\widehat{q}}_{N\alpha}$	$\widehat{\widehat{q}}_{w\alpha}$	$\widehat{\widehat{q}}_{\alpha}$	$\widehat{\widehat{q}}_{N\alpha}$	$\widehat{\widehat{q}}_{w\alpha}$
N	-2.5	-4.9	-4.3	10.8	9.6	10.0	-0.3	-1.4	-0.3	3.0	2.1	2.1
MN	-0.8	-2.9	-1.8	3.0	3.2	3.0	0.1	-1.0	-0.5	0.8	1.0	0.7
L	3.3	1.2	1.1	35.1	31.0	29.8	4.1	2.9	4.1	6.9	4.4	4.2
$t(5)$	9.2	9.0	12.1	36.9	33.1	32.5	8.0	7.0	8.8	6.3	4.6	4.5
$t(7)$	1.4	0.3	-0.3	25.7	22.6	22.2	5.7	4.2	5.5	5.1	3.4	3.4
$t(9)$	1.2	-1.2	-1.2	20.2	18.0	18.9	5.1	3.5	4.7	4.6	3.1	3.1
χ^2_6	13.5	10.5	13.3	3.4	2.1	2.8	6.0	3.2	5.1	2.2	1.1	1.4
χ^2_{12}	8.5	5.5	7.8	4.1	3.1	3.4	3.7	1.7	3.2	2.4	1.3	1.4

Table 14. True VaRs and Expected Shortfalls for standardized distributions

	$q_{0.01}$	$q_{0.05}$	$ES_{0.01}$	$ES_{0.05}$
N	-2.3263	-1.6449	-2.6655	-2.0626
MN	-1.8129	-1.4676	-1.977	-1.679
L	-2.7662	-1.6282	-3.4734	-2.3352
$t(5)$	-2.6065	-1.5608	-3.4487	-2.2388
$t(7)$	-2.5337	-1.6012	-3.1863	-2.193
$t(9)$	-2.4883	-1.6167	-3.0524	-2.1643
χ^2_6	-1.4803	-1.2600	-1.5475	-1.3932
χ^2_{12}	-1.7207	-1.3827	-1.8472	-1.5880

Table 15. Comparison of estimators for $ES_{0.05}$ (true errors are available)

	$n = 500$						$n = 1000$					
	$Bias(\times 10^{-3})$			$MSE(\times 10^{-3})$			$Bias(\times 10^{-3})$			$MSE(\times 10^{-3})$		
	$\widehat{ES}$	$\widehat{ES}_N$	$\widehat{ES}_w$	$\widehat{ES}$	$\widehat{ES}_N$	$\widehat{ES}_w$	$\widehat{ES}$	$\widehat{ES}_N$	$\widehat{ES}_w$	$\widehat{ES}$	$\widehat{ES}_N$	$\widehat{ES}_w$
N	9.7	8.4	4.2	12.3	5.9	6.5	4.4	3.5	1.6	5.9	2.8	3.0
MN	5.1	3.6	2.3	3.0	2.8	2.4	1.7	2.3	1.2	1.5	1.4	1.1
L	17.6	13.9	8.4	38.3	14.3	12.7	6.3	5.6	3.1	20.1	7.8	6.5
$t(5)$	12.7	8.2	-0.0	45.7	17.3	15.6	3.3	3.1	-3.9	21.8	9.3	7.6
$t(7)$	9.0	8.9	3.8	29.5	12.5	12.3	5.5	5.0	1.9	14.4	6.0	5.8
$t(9)$	9.7	5.9	1.4	23.3	10.2	10.4	5.9	5.4	3.4	11.9	5.2	4.9
χ^2_6	3.6	2.3	2.7	1.1	2.6	1.1	1.2	-0.3	0.6	0.5	1.3	0.5
χ^2_{12}	5.3	3.7	3.4	2.8	3.0	2.3	2.3	1.6	1.3	1.3	1.5	1.1

Table 16. Comparison of estimators for $ES_{0.01}$ (true errors are available)

	$n=500$						$n=1000$					
	$Bias(\times 10^{-3})$			$MSE(\times 10^{-3})$			$Bias(\times 10^{-3})$			$MSE(\times 10^{-3})$		
	$\widehat{ES}$	$\widehat{ES}_N$	$\widehat{ES}_w$	$\widehat{ES}$	$\widehat{ES}_N$	$\widehat{ES}_w$	$\widehat{ES}$	$\widehat{ES}_N$	$\widehat{ES}_w$	$\widehat{ES}$	$\widehat{ES}_N$	$\widehat{ES}_w$
N	29.4	29.4	22.4	40.7	32.0	55.7	19.7	19.0	18.7	19.9	15.8	21.4
MN	13.6	14.8	10.3	9.7	9.0	20.2	9.0	9.2	8.8	4.8	4.8	7.2
L	65.3	67.7	64.0	195.3	129.8	166.4	27.2	25.4	18.6	97.8	64.6	70.4
$t(5)$	69.6	69.1	70.2	331.8	213.1	209.7	20.8	23.6	22.7	180.3	114.6	102.6
$t(7)$	59.4	59.3	55.2	179.4	123.0	147.3	18.6	22.3	25.4	94.8	64.6	65.5
$t(9)$	43.0	45.4	44.5	132.8	94.8	119.3	29.5	30.3	27.2	62.6	45.5	51.6
χ^2_6	8.9	4.0	7.5	1.4	4.0	9.5	4.6	2.6	5.1	0.7	2.0	2.5
χ^2_{12}	14.6	13.6	15.1	5.6	7.1	15.5	7.1	5.4	6.4	2.7	3.5	4.9

Table 17. Comparison of estimators for $ES_{0.05}$, with $c=1, a=0.05, b=0.9$

	$n=500$						$n=1000$					
	$Bias(\times 10^{-3})$			$MSE(\times 10^{-3})$			$Bias(\times 10^{-3})$			$MSE(\times 10^{-3})$		
	$\widehat{\widehat{ES}}$	$\widehat{\widehat{ES}}_N$	$\widehat{\widehat{ES}}_w$	$\widehat{\widehat{ES}}$	$\widehat{\widehat{ES}}_N$	$\widehat{\widehat{ES}}_w$	$\widehat{\widehat{ES}}$	$\widehat{\widehat{ES}}_N$	$\widehat{\widehat{ES}}_w$	$\widehat{\widehat{ES}}$	$\widehat{\widehat{ES}}_N$	$\widehat{\widehat{ES}}_w$
N	10.1	13.6	11.8	8.3	6.3	6.5	3.3	4.4	3.5	3.9	2.9	3.0
MN	4.0	5.2	5.2	2.3	2.7	2.3	1.4	2.7	1.9	1.1	1.4	1.1
L	8.3	11.6	8.3	20.8	14.5	12.6	5.6	6.5	3.8	10.9	7.7	6.5
$t(5)$	8.2	9.4	5.6	23.6	16.7	14.5	1.9	3.9	3.3	11.7	8.4	7.2
$t(7)$	11.2	12.8	8.9	16.8	12.0	11.5	5.7	8.0	7.5	8.2	5.9	5.5
$t(9)$	9.3	12.3	10.3	14.6	10.3	10.2	3.6	4.2	3.1	7.1	5.1	4.9
χ^2_6	-18.9	-14.7	-18.0	5.0	2.7	3.9	-11.9	-9.7	-11.6	2.6	1.4	1.9
χ^2_{12}	-12.2	-6.7	-10.3	5.7	3.3	4.1	-4.1	-2.4	-3.8	2.6	1.6	1.8

Table 18. Comparison of estimators for $ES_{0.01}$, with $c=1, a=0.05, b=0.9$

	$n=500$						$n=1000$					
	$Bias(\times 10^{-3})$			$MSE(\times 10^{-3})$			$Bias(\times 10^{-3})$			$MSE(\times 10^{-3})$		
	$\widehat{ES}$	$\widehat{ES}_N$	$\widehat{ES}_w$	$\widehat{ES}$	$\widehat{ES}_N$	$\widehat{ES}_w$	$\widehat{ES}$	$\widehat{ES}_N$	$\widehat{ES}_w$	$\widehat{ES}$	$\widehat{ES}_N$	$\widehat{ES}_w$
N	38.8	41.6	40.9	37.4	34.7	55.4	22.8	25.4	20.4	17.5	16.3	20.5
MN	19.0	20.7	16.8	9.3	9.5	19.7	9.8	11.9	11.7	4.7	4.9	7.2
L	102.6	107.8	101.2	134.9	127.5	144.2	48.7	51.3	49.5	69.0	64.0	67.9
$t(5)$	91.7	98.4	100.9	219.9	204.3	214.4	51.9	52.6	51.1	115.1	109.2	109.6
$t(7)$	82.6	87.5	86.4	130.3	122.5	141.9	39.1	40.7	42.5	67.8	63.0	64.5
$t(9)$	69.5	71.4	65.7	99.7	93.6	115.8	39.4	42.1	43.8	51.3	48.8	53.0
χ^2_6	-39.6	-33.5	-38.9	9.3	6.5	15.6	-22.2	-19.1	-20.7	4.2	2.8	5.7
χ^2_{12}	-12.4	-8.4	-13.6	10.2	8.0	19.3	-9.4	-6.3	-7.5	5.0	3.9	6.6

Table 19. Comparison of estimators for $q_{0.01}$ and $ES_{0.01}$. (a) when the true errors are available, $n=1000$

	$n=1000$						$n=1000$					
	$Bias(\times 10^{-3})$			$MSE(\times 10^{-3})$			$Bias(\times 10^{-3})$			$MSE(\times 10^{-3})$		
	$\widehat{q}_\alpha$	$\widehat{q}_{N\alpha}$	$\widehat{q}_{w\alpha}$	$\widehat{q}_\alpha$	$\widehat{q}_{N\alpha}$	$\widehat{q}_{w\alpha}$	$\widehat{ES}$	$\widehat{ES}_N$	$\widehat{ES}_w$	$\widehat{ES}$	$\widehat{ES}_N$	$\widehat{ES}_w$
N	-0.7	-1.1	0.6	14.1	9.9	11.1	19.7	19.0	18.7	19.9	15.8	21.4
MN	-1.3	-0.9	-0.4	3.3	3.1	2.9	9.0	9.2	8.8	4.8	4.8	7.2
L	11.9	9.4	11.8	47.8	31.6	31.9	27.2	25.4	18.6	97.8	64.6	70.4
$t(5)$	8.4	11.4	22.2	50.4	34.1	35.1	20.8	23.6	22.7	180.3	114.6	102.6
$t(7)$	-1.7	0.1	4.9	34.9	23.1	23.6	18.6	22.3	25.4	94.8	64.6	65.5
$t(9)$	11.4	8.1	8.9	30.9	20.0	21.3	29.5	30.3	27.2	62.6	45.5	51.6
χ^2_6	-1.0	0.5	-0.4	0.9	1.8	0.8	4.6	2.6	5.1	0.7	2.0	2.5
χ^2_{12}	-2.8	-0.9	-1.7	2.5	3.1	2.3	7.1	5.4	6.4	2.7	3.5	4.9

(b) with $c=1, a=0.05, b=0.9, n=1000$

	$n=1000$						$n=1000$					
	$Bias(\times 10^{-3})$			$MSE(\times 10^{-3})$			$Bias(\times 10^{-3})$			$MSE(\times 10^{-3})$		
	$\widehat{\widehat{q}}_\alpha$	$\widehat{\widehat{q}}_{N\alpha}$	$\widehat{\widehat{q}}_{w\alpha}$	$\widehat{\widehat{q}}_\alpha$	$\widehat{\widehat{q}}_{N\alpha}$	$\widehat{\widehat{q}}_{w\alpha}$	$\widehat{\widehat{ES}}$	$\widehat{\widehat{ES}}_N$	$\widehat{\widehat{ES}}_w$	$\widehat{\widehat{ES}}$	$\widehat{\widehat{ES}}_N$	$\widehat{\widehat{ES}}_w$
N	-2.5	-4.9	-4.3	10.8	9.6	10.0	22.8	25.4	20.4	17.5	16.3	20.5
MN	-0.8	-2.9	-1.8	3.0	3.2	3.0	9.8	11.9	11.7	4.7	4.9	7.2
L	3.3	1.2	1.1	35.1	31.0	29.8	48.7	51.3	49.5	69.0	64.0	67.9
$t(5)$	9.2	9.0	12.1	36.9	33.1	32.5	51.9	52.6	51.1	115.1	109.2	109.6
$t(7)$	1.4	0.3	-0.3	25.7	22.6	22.2	39.1	40.7	42.5	67.8	63.0	64.5
$t(9)$	1.2	-1.2	-1.2	20.2	18.0	18.9	39.4	42.1	43.8	51.3	48.8	53.0
χ^2_6	13.5	10.5	13.3	3.4	2.1	2.8	-22.2	-19.1	-20.7	4.2	2.8	5.7
χ^2_{12}	8.5	5.5	7.8	4.1	3.1	3.4	-9.4	-6.3	-7.5	5.0	3.9	6.6

(c) with $c=1, a=0.05, b=0.9, n=10000$

	$n=10000$						$n=10000$					
	$Bias(\times 10^{-3})$			$MSE(\times 10^{-3})$			$Bias(\times 10^{-3})$			$MSE(\times 10^{-3})$		
	$\widehat{\widehat{q}}_\alpha$	$\widehat{\widehat{q}}_{N\alpha}$	$\widehat{\widehat{q}}_{w\alpha}$	$\widehat{\widehat{q}}_\alpha$	$\widehat{\widehat{q}}_{N\alpha}$	$\widehat{\widehat{q}}_{w\alpha}$	$\widehat{\widehat{ES}}$	$\widehat{\widehat{ES}}_N$	$\widehat{\widehat{ES}}_w$	$\widehat{\widehat{ES}}$	$\widehat{\widehat{ES}}_N$	$\widehat{\widehat{ES}}_w$
N	0.9	0.6	0.8	1.1	1.0	1.0	1.6	1.7	1.4	1.8	1.7	1.7
MN	0.3	0.1	0.2	0.3	0.3	0.3	0.5	0.6	0.4	0.4	0.5	0.4
L	0.0	0.2	0.8	3.8	3.4	3.2	3.6	4.2	4.6	7.2	6.8	6.6
$t(5)$	1.0	1.3	2.2	4.0	3.5	3.3	4.8	5.0	5.3	12.3	11.7	10.9
$t(7)$	-0.1	-0.2	0.5	2.7	2.5	2.4	4.3	4.4	3.8	7.1	6.7	6.6
$t(9)$	-0.6	-0.4	-0.0	2.2	2.0	2.0	1.6	1.9	1.8	5.1	4.8	4.7
χ^2_6	1.4	1.4	1.4	0.3	0.2	0.2	-2.6	-2.3	-2.6	0.3	0.2	0.3
χ^2_{12}	0.5	0.1	0.3	0.4	0.3	0.3	-1.4	-1.2	-1.4	0.5	0.4	0.4

Table 20. S&P 500 daily return summary statistics, $2001-2010$.

Mean	-1.0800e-005
Standard deviation	0.0138
Min	-0.0944
Max	0.1090
Skewness	-0.1015
Kurtosis	10.3500
Autocorrelation(one lag) of returns	-0.0845
Autocorrelation(one lag) of squared returns	0.1918
Ljung-Box of returns 20 lags(p-value)	0.0000
Ljung-Box of squared returns 20 lags(p-value)	0.0000
Kolmogorov-Smirnov test	1.0000

Table 21. Conditional VaR and Conditional ES

Model	$\alpha = 0.05$		$\alpha = 0.01$	
	VaR	ES	VaR	ES
GARCH-EL	-1.6737	-2.2447	-2.4791	-3.1284
GARCH-ELR	-1.6799	-2.2501	-2.4842	-3.1327
GARCH-ELW	-1.6857	-2.2500	-2.4791	-3.0782

Table 22. Backtest VaR Models comparison 1. EWMA, MA, HS, GARCH(1,1) and GARCH-ELW

Model	Violation Ratio	Volatility
EWMA	2.2084	0.0185
MA	3.4541	0.0080
HS	2.2650	0.0130
GARCH(1,1)	2.3216	0.0174
GARCH-ELW(our model)	1.1891	0.0185

Table 23. Comparison of the Bernoulli Coverage test & Independence test.

Model	Bernoulli Coverage Test		Independence Test	
	Test statistics	p-value	Test statistics	p-value
EWMA	19.3778	0.0000	1.7626	0.1843
MA	65.6311	0.0000	5.3027	0.0213
HS	21.0129	0.0000	3.2245	0.0725
GARCH(1,1)	22.6991	0.0000	1.9503	0.1626
GARCH-ELW(our model)	0.6016	0.4380	0.5057	0.4770

Table 24. Backtest VaR Model comparison 2. GARCH-EL, GARCH-ELR, GARCH-ELW

Model	Violation Ratio	Volatility
GARCH-EL	1.1891	0.0185
GARCH-ELR	1.1891	0.0185
GARCH-ELW	1.1891	0.0185

Table 25. Comparison of the Bernoulli Coverage test & Independence test

Model	Bernoulli Coverage Test		Independence Test	
	Test statistics	p-value	Test statistics	p-value
GARCH-EL	0.6016	0.4380	0.5057	0.4770
GARCH-ELR	0.6016	0.4380	0.5057	0.4770
GARCH-ELW	0.6016	0.4380	0.5057	0.4770

Table 26. Backtest ES Models comparison. EWMA, MA, HS, GARCH(1,1) and GARCH-ELW

model	$\overline{NS}$
EWMA	1.0908
MA	1.4064
HS	1.1481
GARCH(1,1)	1.0796
GARCH-ELS(our model)	0.9895

Figure 1: Efficiency comparison, Empirical CDF v.s. Normalized CDF

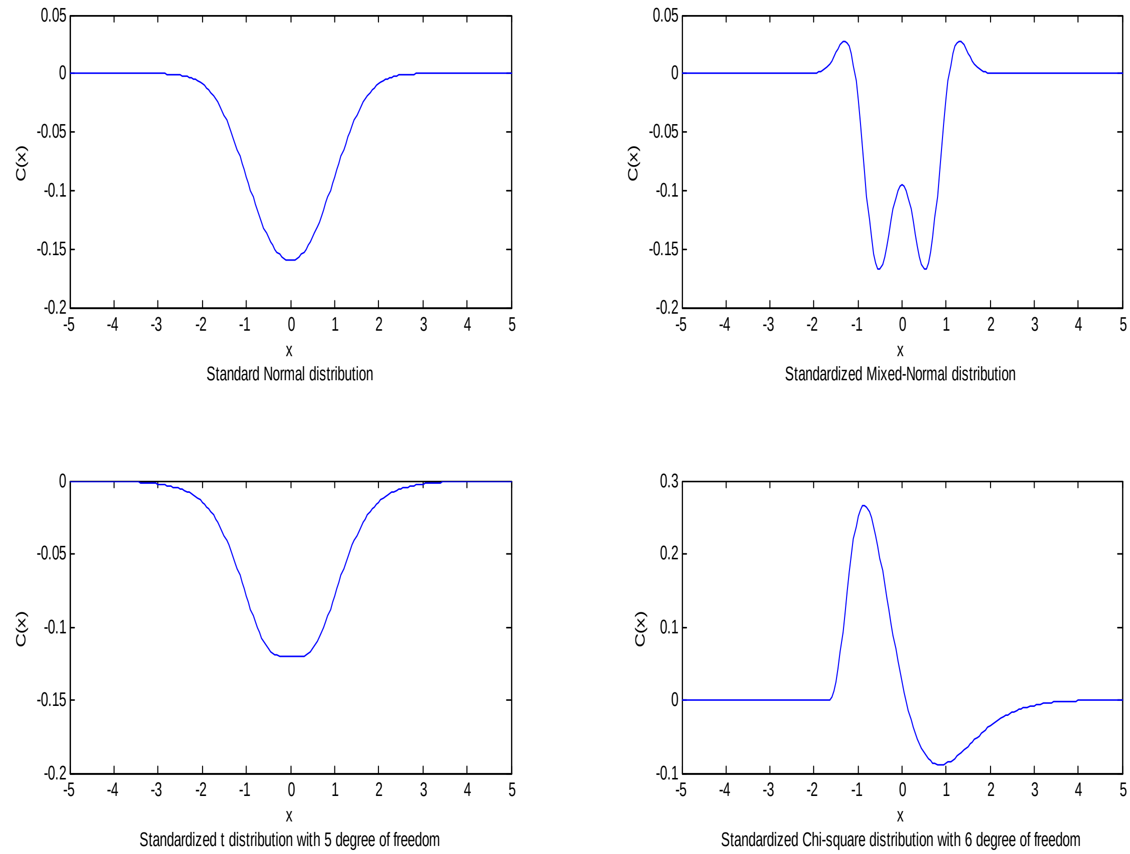
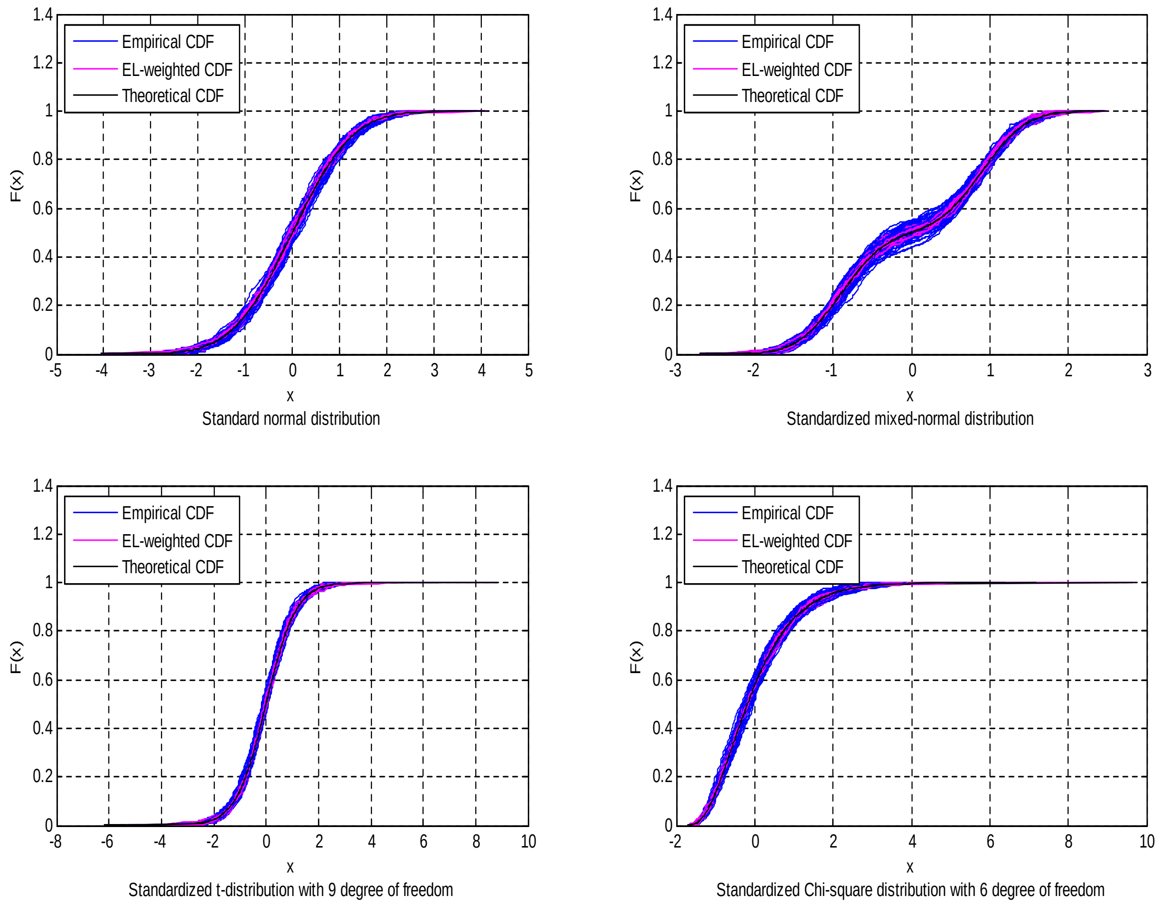


Figure 2: Overlay of two estimators using iid errors: EmpiricalCDF v.s. EL-weighted CDF. N=500.



.

Figure 3: Overlay of two estimators using GARCH errors: Empirical CDF v.s. EL-weighted CDF.N=500

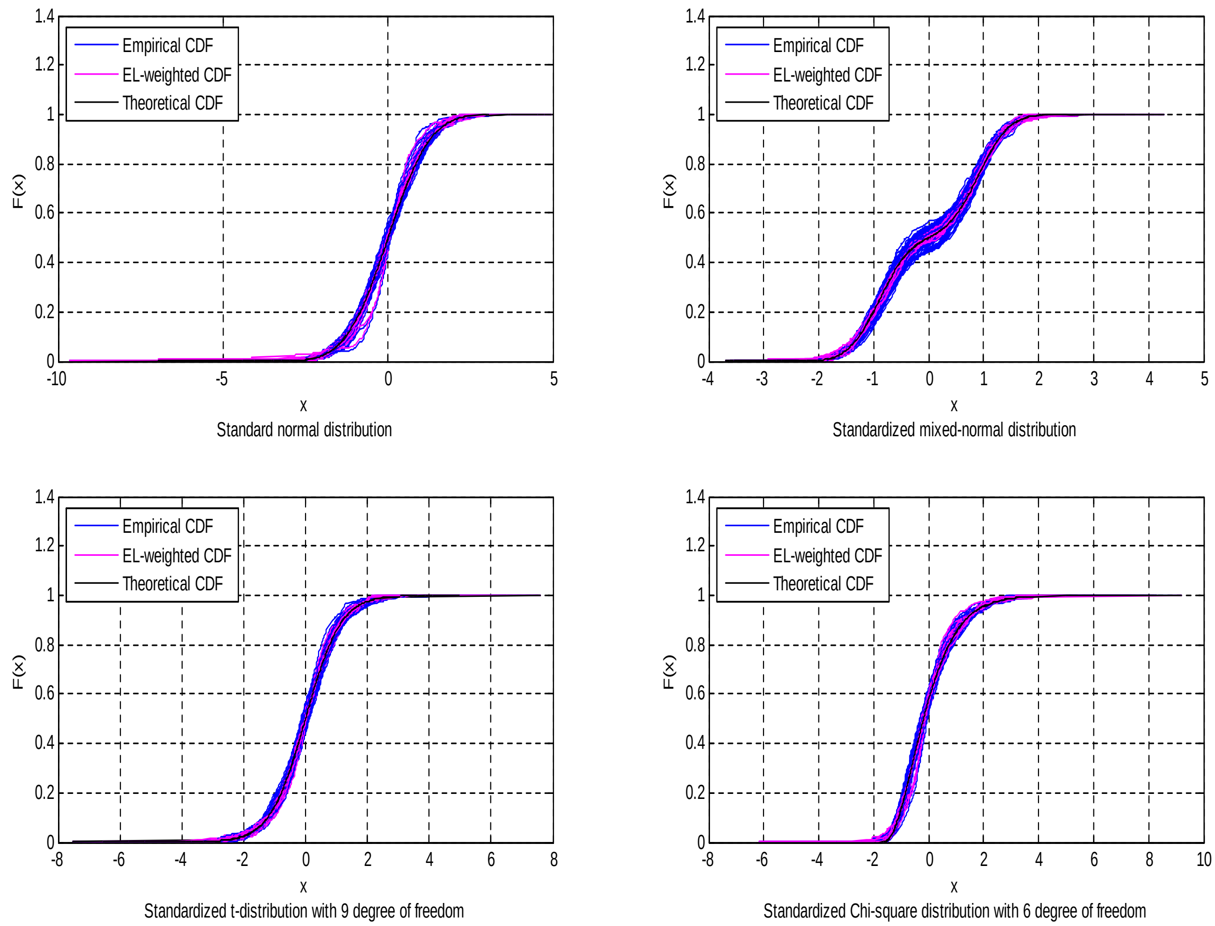


Figure 4. Price and Return Series (2000-2010)

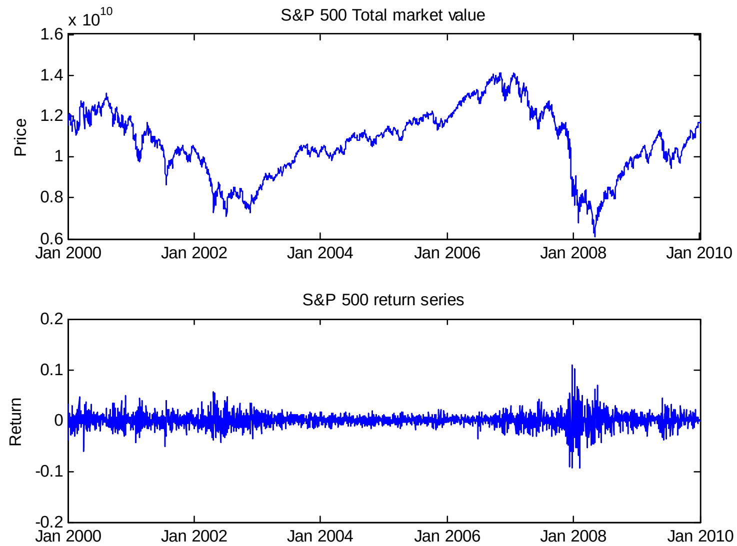


Figure 5. ACF of return and squared return (2000-2010)

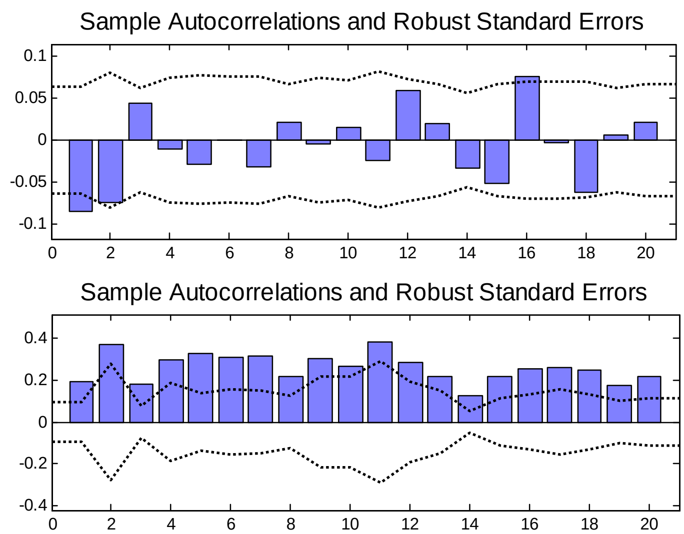


Figure 6. QQ plot for returns series and standardized residuals (2001-2010)

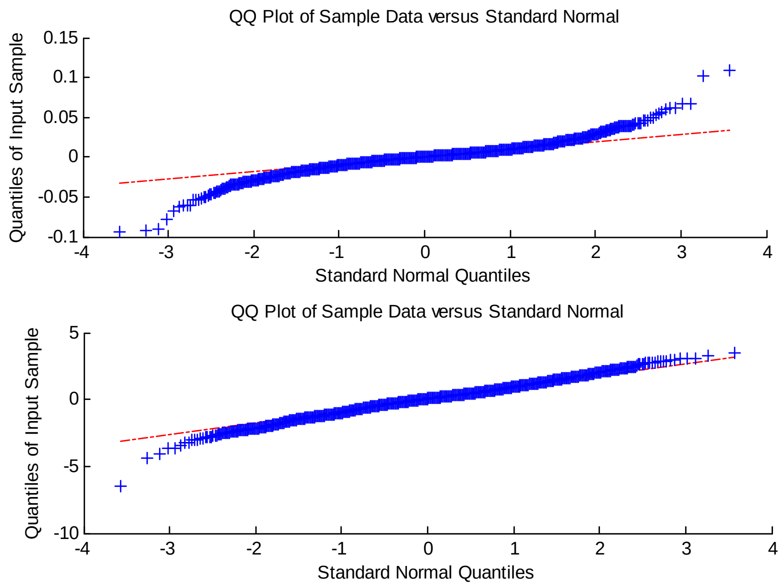


Figure 7. Conditional VaR and ES of model GARCH-ELW

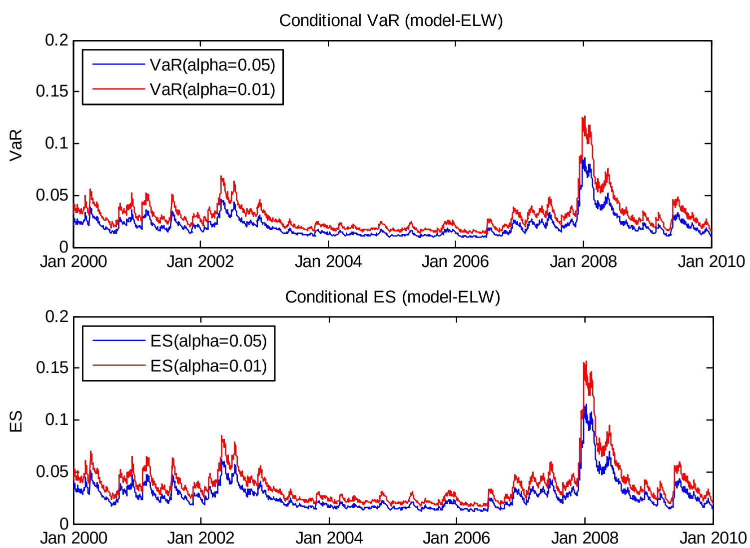


Figure 8. Returns, Conditional Var and Conditional ES(2000-2010)

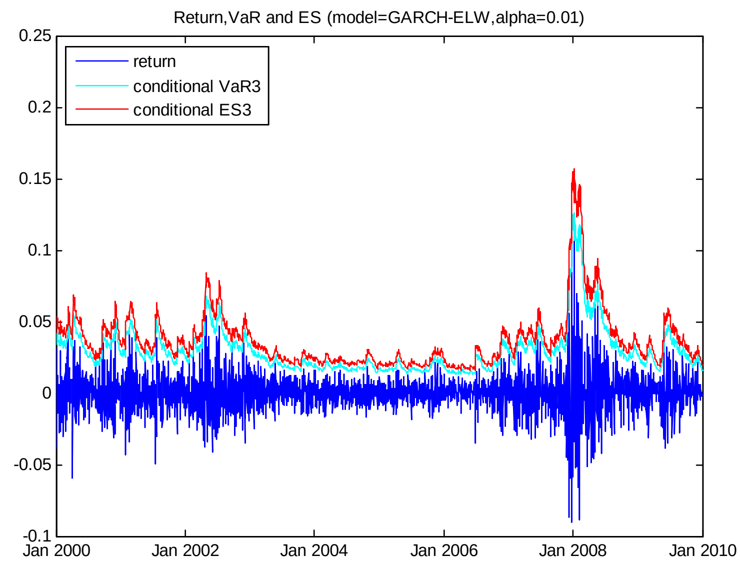
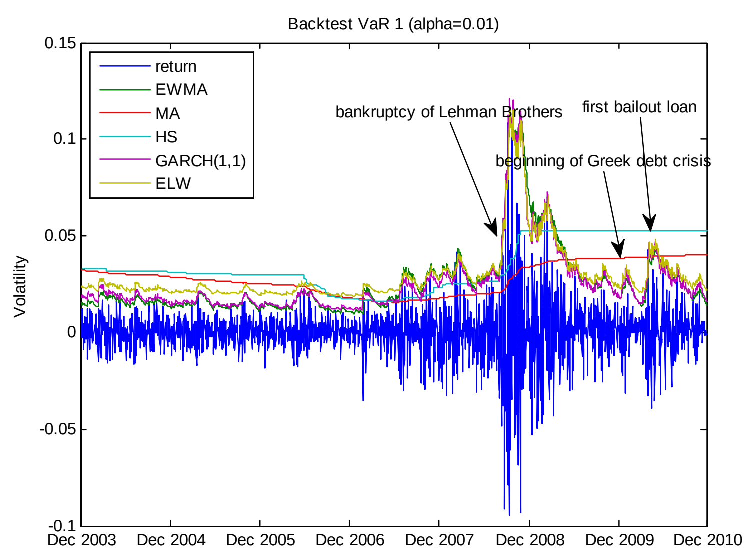


Figure 9. Backtest VaR model comparison 1. EWMA, MA, HS, GARCH(1,1) and GARCH-ELW (2003-2010). (a) Returns and VaR for the whole sample



(b) Focus on Sep 2008 to Dec 2008 — bankruptcy of Lehman Brothers

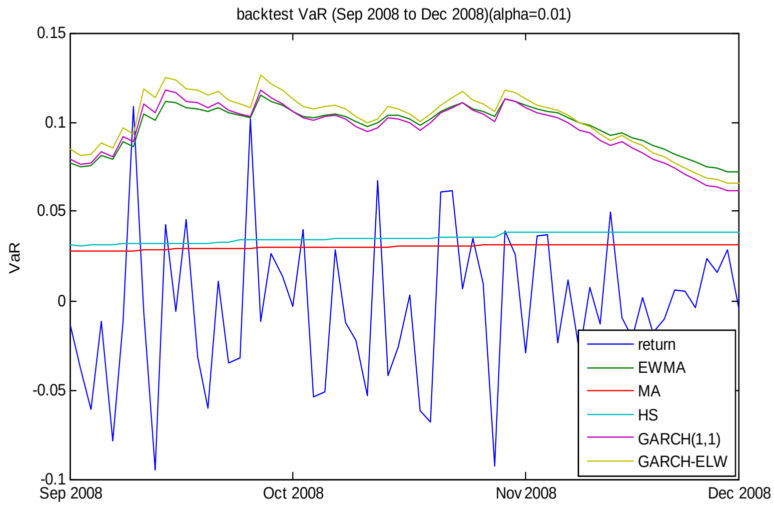


Figure 10. Backtest VaR model comparison 2. GARCH-EL, GARCH-ELR,GARCH-ELW(2003-2010)

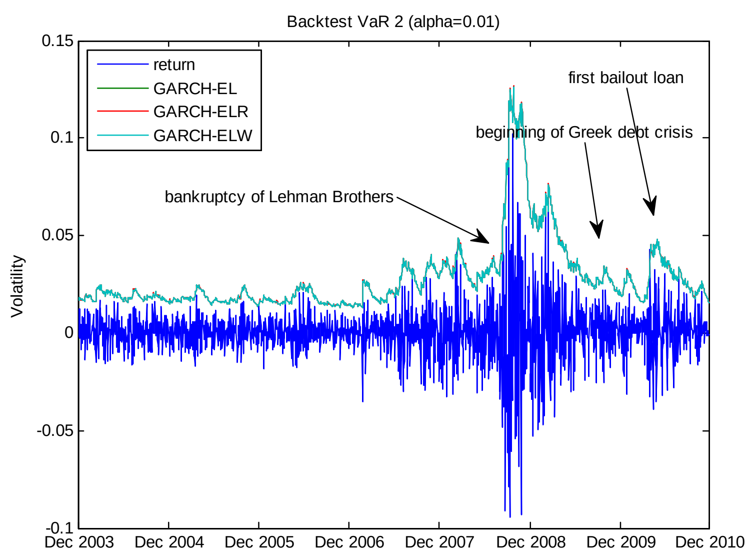
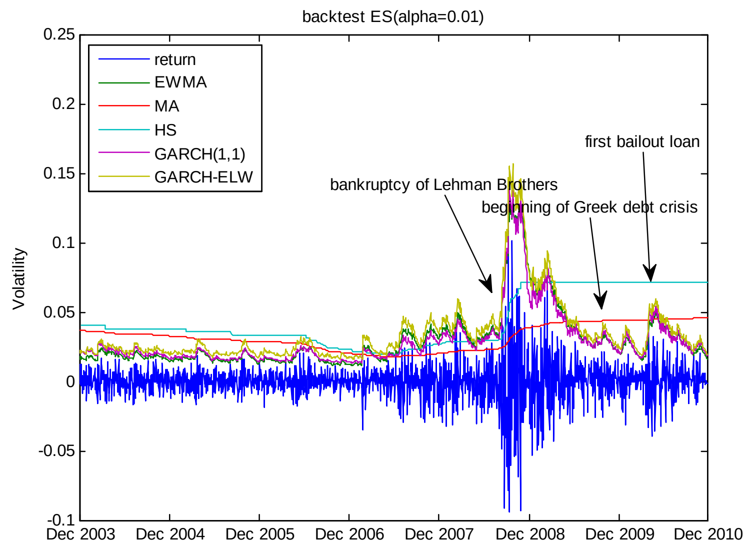


Figure 11. Backtest ES model comparison. EWMA, MA, HS, GARCH(1,1) and GARCH-ELW(2003-2010). (a) Returns and ES for the whole sample



(b) Focus on Sep 2008 to Dec 2008 — bankcruptcy of Lehman Brothers