Assessing Value at Risk with CARE, the Conditional AutoRegressive Expectile Models

Chung-Ming Kuan

Institute of Economics

Academia Sinica

Jin-Huei Yeh

Department of Finance

National Central University

Yu-Chin Hsu

Department of Economics

University of Texas, Austin

Author for correspondence: Chung-Ming Kuan, Institute of Economics, Academia Sinica, Taipei 115, Taiwan; ckuan@econ.sinica.edu.tw

† This research is part of the project "Advancement of Research on Econometric Theory and Applications" (AREMA). The first version of this paper was presented in the First Symposium on Econometric Theory and Applications (SETA) on May 19, 2005 in Academia Sinica, Taiwan. The authors would like to thank two referees and SETA participants for their constructive comments and suggestions.

Abstract

In this paper we propose a downside risk measure, the expectile-based Value at Risk (EVaR), which is more sensitive to the magnitude of extreme losses than the conventional quantile-based VaR (QVaR). The index θ of an EVaR is the relative cost of the expected margin shortfall and hence reflects the level of prudentiality. It is also shown that a given expectile corresponds to the quantiles with distinct tail probabilities under different distributions. Thus, an EVaR may be interpreted as a flexible QVaR, in the sense that its tail probability is determined by the underlying distribution. We further consider conditional EVaR and propose various Conditional AutoRegressive Expectile models that can accommodate some stylized facts in financial time series. For model estimation, we employ the method of asymmetric least squares proposed by Newey and Powell (1987, *Econometrica*) and extend their asymptotic results to allow for stationary and weakly dependent data. We also derive an encompassing test for non-nested expectile models. As an illustration, we apply the proposed modeling approach to evaluate the EVaR of stock market indices.

JEL classification: C22, C51

Keywords: asymmetric least squares, CARE model, expectile, quantile, prudentiality, Value at Risk

1 Introduction

Finding a proper risk measure is crucial in financial risk management. Distinct risk measures have different impacts on asset pricing, portfolio hedging, capital allocation, and investment performance evaluation. When downside risk is of primary concern, the upside and downside movements of returns may be treated differently; see, e.g., Markowitz (1952), Fishburn (1977) and Kahneman and Tversky (1979). A leading downside risk measure is Value at Risk (VaR). A VaR with the confidence level $(1-\alpha)$, $\alpha \in (0,1)$, is defined as the possible maximum loss for a given holding period with probability $(1-\alpha)$; see, e.g., Jorion (2000). Clearly, VaR is the negative of the α-th quantile of the underlying return distribution, and it can be obtained by minimizing asymmetrically weighted mean absolute deviations, with the weights α and $(1-\alpha)$ assigned to positive and negative deviations, respectively. Bassett, Koenker and Kordas (2004) show that such asymmetric weighting scheme is in line with certain distorted probability assessment employed in Choquet expected theory and capable of describing pessimism.

An undesirable property of the existing VaR measure is that it is insensitive to the magnitude of extreme losses. This is so because a VaR, as the quantile with a given tail probability, depends only on the probability (relative frequency) of more extreme realizations but not on their values. It is therefore easy to construct two return distributions that have very different tail behaviors and the same VaR. When the magnitude of loss matters, a quantile-based VaR (henceforth QVaR) may be considered too liberal or too conservative, depending on the tail shape of the underlying distribution. This suggests that QVaR with a given tail probability may not always be an appropriate downside risk measure. Indeed, practitioners and regulators are usually more concerned with the risk exposure in terms of the size of potential losses, for a catastrophic event may completely wipe out an investment.

To avoid the aforementioned problem with QVaR, we propose a downside risk measure that is more tail sensitive. This measure is defined on the *expectile* introduced in Newey and Powell (1987) and will be referred to as expectile-based VaR (henceforth EVaR).[1] The θ-th expectile is the solution to the minimization of asymmetrically weighted mean

[1]Our EVaR is different from the E-VaR of Aït-Sahalia and Lo (2000) which is based on economic valuation of VaR. Note that Taylor (2008) proposes estimating QVaR from expectiles but does not define EVaR; Taylor's work was brought to our attention at the final stage of this paper. Efron (1991) and Sin and Granger (1999) also consider estimating quantiles from expectiles.

squared errors, with the weights θ and $(1-\theta)$ assigned to positive and negative deviations, respectively. Owing to the quadratic loss function, expectiles, and hence EVaR's, are sensitive to extreme values of the distribution.

Taking EVaR as a margin requirement, it will be shown that θ is the relative cost of the expected margin shortfall. A larger (smaller) EVaR is a more (less) prudential margin and results in a smaller (larger) expected margin shortfall. As such, an EVaR is a risk measure under a given level of prudentiality. Moreover, it can be seen that the EVaR with a given θ corresponds to the QVaR's with distinct tail probabilities α under different distributions. Thus, EVaR may be interpreted as a flexible QVaR, in the sense that its confidence level (or tail probability) is not specified *a priori* but is determined by the underlying return distribution. This is in contrast with the conventional QVaR with a given α.

In this paper, we extend EVaR to conditional EVaR and propose various Conditional AutoRegressive Expectile (CARE) models that are capable of accommodating some stylized facts in financial time series. These CARE models are similar but not the same as the CAViaR models proposed by Engle and Manganelli (2004). While CAViaR models rely on the quantile regression method of Koenker and Bassett (1978), the CARE models can be estimated using the method of asymmetric least squares (ALS) proposed by Newey and Powell (1987). To make the ALS method applicable in the dynamic context, we extend the asymptotic results of Newey and Powell (1987) to allow for stationary and weakly dependent data. We also derive an encompassing test for non-nested CARE model specifications, which is analogous to the conditional mean encompassing test of Wooldridge (1990).[2] As an illustration, we apply the proposed CARE modeling approach to assess the EVaR of various stock indices.

This paper is organized as follows. We discuss the properties of expectiles and introduce the EVaR measure in Section 2. We present CARE model specifications, establish asymptotic properties of the ALS estimator, and derive an encompassing test in Section 3. The empirical results are reported in Section 4. Section 5 concludes the paper. All technical proofs are deferred to Appendix.

[2]Taylor (2008) also proposes CARE models for expectiles. As Taylor (2008) is concerned with the estimation of QVaR, his CARE models are the same as the CAViaR models. On the other hand, Taylor (2008) does not discuss the asymptotic properties of the ALS estimator and model specification test.

2 Expectile-Based VaR

Let Y denote an asset return with the distribution function F_Y. Given an $\alpha \in (0,1)$, the QVaR of Y with the confidence level $1-\alpha$ (or the tail probability α) is the negative of the α-th quantile of F_Y: $\mathrm{QVaR}(\alpha) = -q(\alpha)$. It is well known that the α-th quantile can be obtained by minimizing asymmetrically weighted mean absolute deviations:

$$\mathbb{E}\big[|\alpha - \mathbf{1}_{\{Y \leq q\}}| \cdot |Y - q|\big], \tag{1}$$

where $\mathbf{1}_A$ is the indicator of the event A. Thus, a QVaR is a natural product of an optimization problem with an asymmetric linear loss function. The first order condition of minimizing (1) is $\alpha \int_q^\infty \,\mathrm{d}F_Y(y) + (\alpha - 1)\int_{-\infty}^q \,\mathrm{d}F_Y(y) = 0$, which implies

$$\frac{\int_{-\infty}^q \,\mathrm{d}F(y)}{\int_{-\infty}^q \,\mathrm{d}F(y) + \int_q^\infty \,\mathrm{d}F(y)} = \int_{-\infty}^q \,\mathrm{d}F(y) = \alpha. \tag{2}$$

This shows that $q(\alpha)$ depends only on the probability of extreme losses but not their magnitude.

That QVaR is insensitive to the magnitude of extreme losses is a serious drawback in assessing tail risk. To be sure, consider two returns Y_A and Y_B with the following probability functions:

$$f_{Y_A}(y) = \begin{cases} 0.45, & y \in [0,2), \\ 0.05, & y \in [-2,0), \\ 0, & \text{otherwise}; \end{cases} \qquad f_{Y_B}(y) = \begin{cases} 0.45, & y \in [0,2), \\ 0.05, & y \in [-1,0), \\ 0.025, & y \in [-3,-1), \\ 0, & \text{otherwise}. \end{cases}$$

Despite that Y_B may have a larger loss than Y_A, it is easily seen that $\mathrm{QVaR}_{Y_A}(0.1) = \mathrm{QVaR}_{Y_B}(0.1) = 0$ and $\mathrm{QVaR}_{Y_A}(0.05) = \mathrm{QVaR}_{Y_B}(0.05) = 1$. In fact, for any $c > 1$, the return Y_C with

$$f_{Y_C}(y) = \begin{cases} 0.45, & y \in [0,2), \\ 0.05, & y \in [-1,0), \\ 0.05/(c-1), & y \in [-c,-1), \\ 0, & \text{otherwise}, \end{cases}$$

also yields the same QVaRs with the tail probabilities 10% and 5%, even though it may have much larger losses with a positive probability.

2.1 Expectile vs. Quantile

Newey and Powell (1987) considered a quadratic loss function with a weighting scheme similar to that in (1):

$$\mathbb{E}\big[\rho_\theta(Y-\nu)\big] := \mathbb{E}\big[|\theta - \mathbf{1}_{\{Y\le\nu\}}|\cdot|Y-\nu|^2\big], \tag{3}$$

where $\theta \in [0,1]$. The minimizer of (3), $\nu(\theta)$, is known as the θ-th expectile of Y. Clearly, (3) reduces to the standard least-squares objective function when $\theta = 0.5$, and $\nu(0.5)$ is just the expectation of Y. An expectile is also a quantile. Similar to $q(\alpha)$, Newey and Powell (1987) show that $\nu(\theta)$ is monotonically increasing in θ and is location and scale equivariant, in the sense that for $\tilde{Y} = aY + b$ and $a > 0$, $\nu_{\tilde{Y}}(\theta) = a\,\nu_Y(\theta) + b$.

The first order condition of minimizing (3) is

$$\theta\int_\nu^\infty |y-\nu|\,\mathrm{d}F_Y(y) + (\theta-1)\int_{-\infty}^{\nu}|y-\nu|\,\mathrm{d}F_Y(y) = 0.$$

Straightforward calculation shows that the expectile $\varepsilon(\theta)$ satisfies

$$\frac{\int_{-\infty}^{\nu}|y-\nu|\,\mathrm{d}F(y)}{\int_{-\infty}^{\nu}|y-\nu|\,\mathrm{d}F(y) + \int_{\nu}^{\infty}|y-\nu|\,\mathrm{d}F(y)} = \frac{\int_{-\infty}^{\nu}|y-\nu|\,\mathrm{d}F(y)}{\int_{-\infty}^{\infty}|y-\nu|\,\mathrm{d}F(y)} = \theta, \tag{4}$$

which is the ratio of the deviations of Y below ν to the overall deviations of Y from ν, both weighted by the distribution function. Hence, $\nu(\theta)$ depends on both the tail realizations of Y and their probability, whereas $q(\alpha)$ is determined solely by the tail probability.

From (4), it can also be verified that

$$\nu(\theta) = \gamma\,\mathbb{E}\big[Y|\nu(\theta) < Y\big] + (1-\gamma)\,\mathbb{E}\big[Y|Y\le\nu(\theta)\big], \tag{5}$$

where $\gamma = \theta[1-F_Y(\nu(\theta))]/\{\theta[1-F_Y(\nu(\theta))] + (1-\theta)F_Y(\nu(\theta))\}$. This shows that $\nu(\theta)$ is a balance between $\mathbb{E}(Y|\nu(\theta) < Y)$ (*conditional upside mean*) and $\mathbb{E}(Y|Y \le \nu(\theta))$ (*conditional downside mean*). As such, an expectile is less extreme than the expected shortfall because the latter is determined only by the conditional downside mean.

For any $\alpha \in (0,1)$, let $\theta(\alpha)$ be such that $\nu_Y(\theta(\alpha)) = q_Y(\alpha)$. Yao and Tong (1996) showed that $\theta(\alpha)$ is related to $q(\alpha)$ via:

$$\theta(\alpha) = \frac{\alpha\cdot q(\alpha) - \int_{-\infty}^{q(\alpha)} y\,dF(y)}{\mathbb{E}[Y] - 2\int_{-\infty}^{q(\alpha)} y\,dF(y) - (1-2\alpha)q(\alpha)}.$$

For example, when Y has a uniform distribution on $[-a,a]$, $q(\alpha) = 2\alpha a - a$ and $\theta(\alpha) = \alpha^2/(2\alpha^2 - 2\alpha + 1)$. Thus, for $\alpha = 1\%, 5\%, 10\%, 25\%, 50\%$, the corresponding $q(\alpha)$ are $\nu(\theta)$

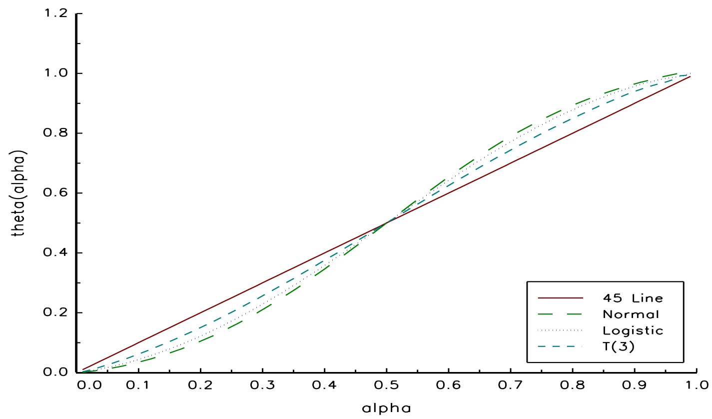

Figure 1: The correspondence between α and θ: $\theta(\alpha)$ function.

with $\theta = 0.01\%, 0.27\%, 1.2\%, 10\%, 50\%$, respectively. For other distributions, we examine the correspondence between α and $\theta(\alpha)$ via Monte Carlo simulations. We plot $\theta(\alpha)$ for the standard normal, logistic and $t(3)$ distributions in Figure 1, with α on the horizontal axis and $\theta(\alpha)$ on the vertical axis.

We can see that for $\alpha < (>)\ 0.5$, the $\theta(\alpha)$ curves all lie below (above) the $45°$ line where $\alpha = \theta(\alpha)$.[3] For a given $\alpha < 0.5$, $\theta(\alpha)$ is larger for the distribution with thicker tails. For the example discussed in the beginning of this section, $\theta(0.05)$ is approximately 0.011 for Y_A and 0.027 for Y_B. That is, although $q(0.05)$ is the same for Y_A and Y_B, it is an expectile corresponding to different θ for Y_A and Y_B, and hence different risk exposures in terms of weighted magnitude of extreme losses. Similarly, for a given $\theta < 0.5$, the corresponding α would be smaller if the distribution has thicker tails. Thus, an expectile with a given θ corresponds to quantiles with different α under distinct distributions, and hence represent different risks exposures in terms of the probability (frequency) of tail losses. Table 1 summarizes the α values implied by a given θ under various distributions.

To illustrate the sensitiveness of different risk measures to tail events, we compare the relative performance of quantile, expectile, and conditional downside (tail) mean in

[3]The $45°$ line represents the distribution whose expectiles agree with quantiles when $\theta = \alpha$. Koenker (1992) showed that its distribution function is

$$F(y) = \begin{cases} \frac{1}{2}\left(1 + \sqrt{1 - \frac{4}{4+y^2}}\right), & y \geq 0, \\ \frac{1}{2}\left(1 - \sqrt{1 - \frac{4}{4+y^2}}\right), & y < 0, \end{cases}$$

which has finite mean, infinite variance, and algebraic tails.

Table 1: Implied α values under different distributions.

θ	$U(-a,a)$	$\mathcal{N}(0,1)$	$t(30)$	$t(10)$	$t(5)$	$t(3)$
1%	9.2%	4.3%	4.0%	3.5%	3.0%	2.4%
3%	15.0%	9.1%	8.8%	8.0%	6.8%	5.6%
5%	18.6%	12.6%	12.3%	11.5%	10.0%	8.5%
10%	25.0%	19.5%	19.0%	18.3%	16.6%	14.5%
25%	36.6%	33.2%	32.8%	32.2%	31.9%	29.4%

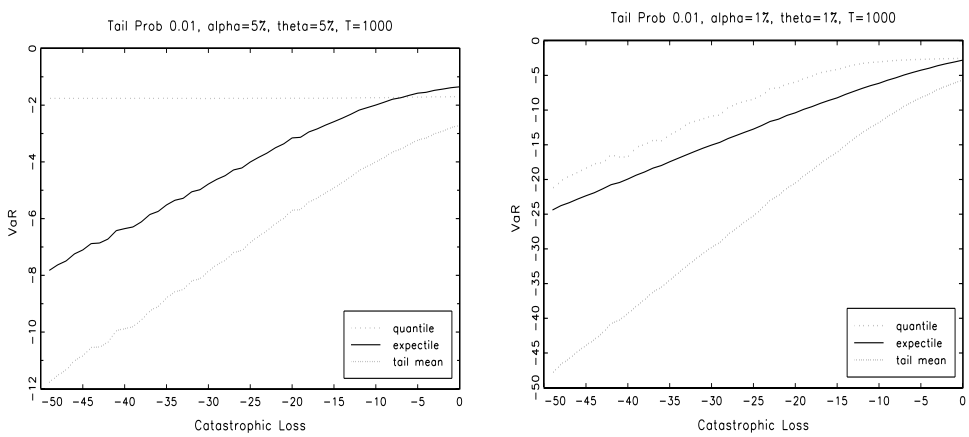

Figure 2: The catastrophic loss sensitivity of quantile, expectile and conditional tail mean

the presence of catastrophic loss, using Monte Carlo experiments. Similar to Duffie and Pan (1997), the data are independently drawn from $\mathcal{N}(0,\ 1/\sqrt{1-P})$ with probability $1-P$ or from $\mathcal{N}(c,\ 1/\sqrt{P})$ with probability P, cf. Gourieroux and Jasiak (2002). By setting P to a value close to 0, the observations are often drawn from $\mathcal{N}(0,\ 1/\sqrt{1-P})$, and there may be infrequent catastrophic losses taken from the more disperse distribution $\mathcal{N}(c,\ 1/\sqrt{P})$. In our simulations, $c \in [-1, -50]$, the sample size is 1000, and the number of replications is 1000. In Figure 2, we plot the quantiles with $\alpha = 0.01$ and 0.05, the expectiles with $\theta = 0.01$ and 0.05, and the conditional downside means based on $q(0.01)$ and $q(0.05)$. The left panel is the case that $P = \alpha = \theta = 0.01$, and the right panel is $P = 0.01$ with $\alpha = \theta = 0.05$.

From Figure 2 it is clear that the expectile and conditional downside mean vary with c, but the corresponding quantile may not. When $P < \alpha$, the quantile is not affected by the extreme values from $\mathcal{N}(c,\ 1/\sqrt{P})$ and hence remains constant across c. A qunatile would change with c when the chosen α level happens to be the same as (or smaller than) the probability of the tail distribution, P, yet its magnitude is smaller than that of the expectile for all c. These results show that the danger of basing a risk measure on the quantile with a given α level, as it may not respond properly to catastrophic losses. It is also clear that the conditional downside mean depends only on the tail event and hence is much larger (more conservative) than corresponding expectile and quantile.

2.2 Expectile-Based VaR

The properties discussed above suggest that an expectile, which takes into account the magnitude of loss, may serve as a better measure for tail risk. We thus define EVaR, expectile-based VaR, with the index θ as $\mathrm{EVaR}(\theta) = -\nu(\theta)$.

We now give an intuitive interpretation for θ. Taking $-\nu(\theta)$ as a margin (capital requirement), $\int_{-\infty}^{\nu(\theta)} |y - \nu(\theta)|\, \mathrm{d}F(y)$ is the expected margin shortfall and a potential cost for more extreme losses, and $\int_{\nu(\theta)}^{\infty} |y - \nu(\theta)|\, \mathrm{d}F(y)$ is an opportunity cost due to the expected margin overcharge. The sum of these two costs, $\int_{-\infty}^{\infty} |y - \nu(\theta)|\, \mathrm{d}F(y)$, is thus the expected total cost of holding the capital requirement $-\nu(\theta)$. In view of (4), θ can be understood as the relative cost of the expected margin shortfall. A larger $|\nu(\theta)|$ is a more prudential margin requirement and results in smaller expected margin shortfall and hence θ; see also Lam, Sin, and Leung (2004) for related discussion. As such, θ will be referred to as an *index of prudentiality*.

To summarize, an EVaR is a risk measure that balances between the cost of margin shortfall and the opportunity cost due to margin overcharge under a given level of prudentiality. This is in line with a major task of the clearinghouse of futures market; see, e.g., Baer, France and Moser (1994) and Booth, Broussard, Martikainen and Puttonen (1997). It is worth mentioning that, by (4), the index of prudentiality θ for an EVaR takes into account both the probability and the magnitude of return, whereas α of a QVaR is just its tail probability.

We emphasize that EVaR can be interpreted as a flexible QVaR for the underlying return distribution. Ideally, one would choose a smaller (larger) α for QVaR if the left tail of the return distribution were known to be thicker (thinner). Yet, the shape of a return distribution is rarely known in practice, and α is typically set by regulators and/or the management level. For example, J. P. Morgan reveals its daily QVaR at the tail level of 5%; the Bank of International Settlements sets QVaR for evaluating the adequacy of bank capital at 1% level. These choices of α are rigid and may not be able to reveal the potential risk when the return distribution exhibits different shapes over time. By contrast, the expectile with a given θ corresponds to the quantiles with distinct α values under different distributions. Thus, instead of finding the QVaR with a pre-determined α, we may identify the EVaR with a given θ and allow the data to reveal their risk in terms of the tail probability α, as shown in Figure 1.

3 CARE Model Specification and Estimation

The concept of expectiles is readily extended to conditional expectiles. In this section we first introduce conditional expectile models for EVaR, which are similar to but different from those of Engle and Manganelli (2004) and Taylor (2008). We shall also establish the asymptotic properties of the ALS estimator under more general conditions and derive an encompassing test for non-nested models.

3.1 Model Specifications

Given a collection of k variables, X, in the information set $\mathcal{F}$, let $\mu_\theta(X)$ denote the θ-th expectile of Y conditional on $\mathcal{F}$. We shall consider the linear specification $X'\boldsymbol{\beta}(\theta)$, with $\boldsymbol{\beta}(\theta)$ a $k\times 1$ parameter vector. When the data $(y_t, \boldsymbol{x}_t')'$ are available, the linear specification can be expressed as:

$$y_t = \boldsymbol{x}_t'\boldsymbol{\beta}(\theta) + e_t(\theta), \quad t = 1, \dots, T, \tag{6}$$

where $e_t(\theta)$ denotes the error term. We say $X'\boldsymbol{\beta}(\theta)$ is a correct specification of $\mu_\theta(X)$ if there exists $\boldsymbol{\beta}_o(\theta)$ such that $X'\boldsymbol{\beta}_o(\theta) = \mu_\theta(X)$ with probability one. Under correct specification, we have $y_t = \boldsymbol{x}_t'\boldsymbol{\beta}_o(\theta) + \varepsilon_t(\theta)$.

In the dynamic context, to model the conditional expectile of y_t, we consider the information set up to time $t-1$: $\mathcal{F}^{t-1}$. It is natural to include lagged returns in $\boldsymbol{x}_t$, so as to accommodate potential return correlation (dependence) over time. By the definition of expectile, it is also reasonable to expect that past positive return ($y_{t-1}^+ = \max(y_{t-1}, 0)$) and negative return ($y_{t-1}^- = \max(-y_{t-1}, 0)$) exert different effects on conditional expectiles, especially for tail expectiles. As such, we shall allow for asymmetric effects of return magnitude on tail expectiles by including the magnitude (square or absolute value) of positive and negative lagged returns in the model. Such asymmetry is in line with Black (1976) and Christie (1982); Nelson (1991), Glosten, et al. (1993), and Engle and Ng (1993) also allow for such effects in modeling conditional variance.

It is well known that $y_{t-1} = y_{t-1}^{+} - y_{t-1}^{-}$, $|y_{t-1}| = y_{t-1}^{+} + y_{t-1}^{-}$, and $y_{t-1}^2 = (y_{t-1}^{+})^2 + (y_{t-1}^{-})^2$. In the first CARE model specification, $\boldsymbol{x}_t = (1, y_{t-1}, (y_{t-1}^{+})^2, (y_{t-1}^{-})^2)'$, so that (6) reads:

$$\begin{aligned} y_t &= a_0(\theta) + a_1(\theta) y_{t-1} + b_1(\theta) y_{t-1}^2 + c_1(\theta)(y_{t-1}^{-})^2 + e_t(\theta) \\ &= a_0(\theta) + a_1(\theta) y_{t-1} + b_1(\theta)(y_{t-1}^{+})^2 + \gamma_1(\theta)(y_{t-1}^{-})^2 + e_t(\theta), \end{aligned} \tag{7}$$

where $\gamma_1(\theta) = b_1(\theta) + c_1(\theta)$. The positive and negative parts of y_{t-1} would exert the same magnitude effect on the θ-th conditional expectile when $b_1(\theta) = \gamma_1(\theta)$ (or $c_1(\theta) = 0$). The resulting conditional expectiles, however, may not be as smooth as the conditional quantiles modeled using a CAViaR model, because the former are more sensitive to the magnitude of past observations.[4]

Alternatively, we may use $|y_{t-1}|$ to represent the magnitude of y_{t-1}. This leads to the CARE specification with $\boldsymbol{x}_t = (1, y_{t-1}^{+}, y_{t-1}^{-})'$, so that (6) is

$$\begin{aligned} y_t &= a_0(\theta) + a_1(\theta) y_{t-1} + d_1(\theta)|y_{t-1}| + e_t(\theta) \\ &= a_0(\theta) + \delta_1(\theta) y_{t-1}^{+} + \lambda_1(\theta) y_{t-1}^{-} + e_t(\theta), \end{aligned} \tag{8}$$

with $\delta_1(\theta) = d_1(\theta) + a_1(\theta)$ and $\lambda_1(\theta) = d_1(\theta) - a_1(\theta)$. Clearly, y_{t-1}^{+} and y_{t-1}^{-} would not have the same effect on the θ-th conditional expectile unless $\delta_1(\theta) = \lambda_1(\theta)$ (or $a_1(\theta) = 0$). The right-hand side of (8) looks similar to the "asymmetric slope" specification of the CAViaR model, yet it does not involve a lagged conditional expectile.

A natural extension of (7) is the following CARE model:

$$\begin{aligned} y_t = a_0(\theta) + a_1(\theta) y_{t-1} + \cdots + a_q(\theta) y_{t-q} + b_1(\theta)(y_{t-1}^{+})^2 + \gamma_1(\theta)(y_{t-1}^{-})^2 + \cdots \\ + b_q(\theta)(y_{t-q}^{+})^2 + \gamma_q(\theta)(y_{t-q}^{-})^2 + e_t(\theta). \end{aligned} \tag{9}$$

[4]From (7) we can see that $\boldsymbol{x}_t'\boldsymbol{\beta}(\theta)$ has an AR structure:

$$\boldsymbol{x}_t'\boldsymbol{\beta}(\theta) = a_0(\theta) + a_1(\theta)\big(\boldsymbol{x}_{t-1}'\boldsymbol{\beta}(\theta)\big) + b_1(y_{t-1}^{+})^2 + \gamma_1(y_{t-1}^{-})^2 + a_1 e_{t-1}(\theta),$$

which is similar to a CAViaR specification with possibly asymmetric magnitude effects. Yet, the magnitude of lagged return and error also affect the behavior of conditional expectiles in our model.

The positive and negative lagged returns would have the same magnitude effect if $b_i(\theta) = \gamma_i(\theta)$, $i = 1, \ldots, q$. An extension of (8) is the CARE model:

$$y_t = a_0(\theta) + \delta_1(\theta)y_{t-1}^{+} + \lambda_1(\theta)y_{t-1}^{-} + \cdots + \delta_q(\theta)y_{t-q}^{+} + \lambda_q(\theta)y_{t-q}^{-} + e_t(\theta), \tag{10}$$

for which the positive and negative lagged returns would have the same magnitude effect if $\delta_i(\theta) = \lambda_i(\theta)$, $i = 1, \ldots, q$.

3.2 Model Estimation

The specification (6) can be estimated by the ALS method proposed by Newey and Powell (1987). Let $\boldsymbol{\beta}^*(\theta)$ be the minimizer of the loss function: $\mathbb{E}\left[\rho_\theta(Y - X'\boldsymbol{\beta}(\theta))\right]$, so that $y_t = \boldsymbol{x}_t'\boldsymbol{\beta}^*(\theta) + e_t^*(\theta)$. The ALS estimator for $\boldsymbol{\beta}^*(\theta)$, denoted as $\hat{\boldsymbol{\beta}}_T(\theta)$, can then be obtained by minimizing the sample counterpart: $T^{-1}\sum_{t=1}^{T}\rho_\theta(y_t - \boldsymbol{x}_t'\boldsymbol{\beta}(\theta))$.

The first order condition of the ALS minimization problem is

$$\frac{1}{T}\sum_{t=1}^{T}|\theta - \mathbf{1}_{\{y_t - \boldsymbol{x}_t'\boldsymbol{\beta}(\theta)\le 0\}}|\boldsymbol{x}_t(y_t - \boldsymbol{x}_t'\boldsymbol{\beta}(\theta)) =: \frac{1}{T}\sum_{t=1}^{T} w(e_t(\theta);\theta)\boldsymbol{x}_t e_t(\theta) = \mathbf{0},$$

where $w(e_t(\theta);\theta) = |\theta - \mathbf{1}_{\{e_t(\theta)\le 0\}}|$. The ALS estimator $\hat{\boldsymbol{\beta}}_T(\theta)$ thus satisfies:

$$\hat{\boldsymbol{\beta}}_T(\theta) = \left(\sum_{t=1}^{T} w(\hat{e}_t(\theta);\theta)\boldsymbol{x}_t\boldsymbol{x}_t'\right)^{-1}\left(\sum_{t=1}^{T} w(\hat{e}_t(\theta);\theta)\boldsymbol{x}_t y_t\right), \tag{11}$$

where $\hat{e}_t(\theta) = y_t - \boldsymbol{x}_t'\hat{\boldsymbol{\beta}}_T(\theta)$. Although (11) is not a closed form solution, it can be computed as an iterated weighted least squares estimator. For notation simplicity, we shall write $w_t^*(\theta) = w(e_t^*(\theta);\theta)$ and $\hat{w}_t(\theta) = w(\hat{e}_t(\theta);\theta)$.

Newey and Powell (1987) established consistency and asymptotic normality of the ALS estimator (11) under the condition that the data are i.i.d. Their results are readily extended to allow for stationary and weakly dependent data under suitable regularity conditions. These conditions are similar to those in Newey and Powell (1987) and are deferred to Appendix to reduce technicality. In what follows, we shall write $\xrightarrow{\mathbb{P}}$ and $\xrightarrow{D}$ for convergence in probability and convergence in distribution, respectively. The consistency result follows easily from Theorem 4.3 of Wooldridge (1994).

Theorem 3.1 *Given [A1]–[A3] in Appendix,* $\hat{\boldsymbol{\beta}}_T(\theta) \xrightarrow{\mathbb{P}} \boldsymbol{\beta}^*(\theta)$ *as* $T \to \infty$.

The proof of the asymptotic normality of normalized $\hat{\boldsymbol{\beta}}_T(\theta)$ is similar to that of Theorem 3 of Newey and Powell (1987), *mutatis mutandis*.

Theorem 3.2 *Given [A1]–[A3] in Appendix,*

$$\sqrt{T}\big(\hat{\boldsymbol{\beta}}_T(\theta) - \boldsymbol{\beta}^*(\theta)\big) \xrightarrow{D} \mathcal{N}\big(\mathbf{0},\ \boldsymbol{\Sigma}(\theta)\big),$$

as $T \to \infty$, *where* $\boldsymbol{\Sigma}(\theta) = \boldsymbol{\Xi}(\theta)^{-1}\boldsymbol{V}(\theta)\boldsymbol{\Xi}(\theta)^{-1}$ *with* $\boldsymbol{\Xi}(\theta) = \mathbb{E}[w_t^*(\theta)\boldsymbol{x}_t\boldsymbol{x}_t']$,

$$\boldsymbol{V}(\theta) = \lim_{T\to\infty} \boldsymbol{V}_T(\theta) := \lim_{T\to\infty} \operatorname{var}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T} w_t^*(\theta)\boldsymbol{x}_t e_t^*(\theta)\right),$$

and $e_t^*(\theta) = y_t - \boldsymbol{x}_t'\boldsymbol{\beta}^*(\theta)$.

When (6) is correctly specified for the θ-th conditional expectile, we have $\boldsymbol{\beta}^*(\theta) = \boldsymbol{\beta}_o(\theta)$, which also minimizes $\mathbb{E}[\rho_\theta(y_t - \boldsymbol{x}_t'\boldsymbol{\beta}(\theta)) \mid \mathcal{F}^{t-1}]$ (Newey and Powell, 1987, p. 824). Thus, $\boldsymbol{\beta}_o(\theta)$ satisfies the first order condition:

$$\mathbb{E}\big[w_t^o(\theta)\boldsymbol{x}_t\varepsilon_t(\theta) \mid \mathcal{F}^{t-1})\big] = \boldsymbol{x}_t\,\mathbb{E}\big[w_t^o(\theta)\varepsilon_t(\theta) \mid \mathcal{F}^{t-1})\big] = \mathbf{0};$$

where $\varepsilon_t(\theta) = y_t - \boldsymbol{x}_t'\boldsymbol{\beta}_o(\theta)$ and $w_t^o(\theta) = w\big(\varepsilon_t(\theta);\theta\big)$. Without loss of generality, $\boldsymbol{x}_t$ contains the constant one, so that the weighted errors, $w_t^o(\theta)\varepsilon_t(\theta)$, have the martingale difference property:

$$\mathbb{E}\big[w_t^o(\theta)\varepsilon_t(\theta) \mid \mathcal{F}^{t-1})\big] = \mathbf{0}. \tag{12}$$

Clearly, (12) reduces to the conventional martingale difference condition for least-squares errors when $w_t^o(\theta) = 1/2$ for all t. It follows that Theorem 3.2 holds as:

$$\sqrt{T}\big(\hat{\boldsymbol{\beta}}_T(\theta) - \boldsymbol{\beta}_o(\theta)\big) \xrightarrow{D} \mathcal{N}\big(\mathbf{0},\ \boldsymbol{\Sigma}(\theta)\big),$$

where $\boldsymbol{\Sigma}(\theta) = \boldsymbol{\Xi}(\theta)^{-1}\boldsymbol{V}(\theta)\boldsymbol{\Xi}(\theta)^{-1}$ with $\boldsymbol{V}(\theta) = \operatorname{var}\big(w_t^o(\theta)\boldsymbol{x}_t\varepsilon_t(\theta)\big)$, by the martingale difference property (12).

As in Newey and Powell (1987), the asymptotic covariance matrix $\boldsymbol{\Sigma}(\theta)$ can be consistently estimated by $\widehat{\boldsymbol{\Sigma}}_T(\theta) = \widehat{\boldsymbol{\Xi}}_T(\theta)^{-1}\widehat{\boldsymbol{V}}_T(\theta)\widehat{\boldsymbol{\Xi}}_T(\theta)^{-1}$, where

$$\widehat{\boldsymbol{\Xi}}_T(\theta) = \frac{1}{T}\sum_{t=1}^{T}\hat{w}_t(\theta)\boldsymbol{x}_t\boldsymbol{x}_t' \xrightarrow{\mathbb{P}} \boldsymbol{\Xi}(\theta)$$

$$\widehat{\boldsymbol{V}}_T(\theta) = \frac{1}{T}\sum_{t=1}^{T}\hat{w}_t^2(\theta)\hat{e}_t^2(\theta)\boldsymbol{x}_t\boldsymbol{x}_t' \xrightarrow{\mathbb{P}} \boldsymbol{V}(\theta) = \operatorname{var}\big(w_t^o(\theta)\boldsymbol{x}_t\varepsilon_t(\theta)\big).$$

It can be shown that the proof in Newey and Powell (1987) in fact carries over under stationarity and the martingale difference property (12); we omit the details.

3.3 Model Specification Test

In section 3.1, there are two CARE specifications, (9) and (10), for tail conditional expectiles. To determine an appropriate model, we construct an encompassing test of the following null model:

$$H_0 : \boldsymbol{x}_t'\boldsymbol{\beta}_o(\theta) = \mu_\theta(\boldsymbol{x}_t), \quad \text{with probability one,}$$

against the alternative:

$$H_1 : \boldsymbol{\zeta}_t'\boldsymbol{\gamma}_o(\theta) = \mu_\theta(\boldsymbol{x}_t), \quad \text{with probability one,}$$

where $\boldsymbol{x}_t$ $(k \times 1)$ and $\boldsymbol{\zeta}_t$ $(m \times 1)$ are in $\mathcal{F}^{t-1}$, and $\boldsymbol{x}_t$ and $\boldsymbol{\zeta}_t$ contain different elements. For example, $\boldsymbol{x}_t$ includes the constant one, y_{t-i}, $(y_{t-i}^+)^2$, and $(y_{t-i}^-)^2$, $i = 1, \ldots, q$, when (9) is the null model, whereas $\boldsymbol{\zeta}_t$ includes the constant one, y_{t-i}^+, and y_{t-i}^-, $i = 1, \ldots, q$, when (10) is the alternative model.

In view of (12), we may test the null hypothesis by checking if the weighted errors of the null model are uncorrelated with the variables in the alternative model:

$$\mathbb{E}\big[\boldsymbol{\zeta}_t w_t^o(\theta)\varepsilon_t(\theta)\big] = \mathbf{0}. \tag{13}$$

We can then base a test of (13) on:

$$\begin{aligned}
&\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{\zeta}_t\hat{w}_t(\theta)\hat{e}_t(\theta) \\
&\qquad = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{\zeta}_t\hat{w}_t(\theta)\varepsilon_t(\theta) - \frac{1}{T}\sum_{t=1}^{T}\hat{w}_t(\theta)\boldsymbol{\zeta}_t\boldsymbol{x}_t'\sqrt{T}\big(\hat{\boldsymbol{\beta}}_T(\theta) - \boldsymbol{\beta}_o(\theta)\big) \\
&\qquad = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\boldsymbol{\zeta}_t\hat{w}_t(\theta)\varepsilon_t(\theta) \\
&\qquad\quad - \left(\frac{1}{T}\sum_{t=1}^{T}\hat{w}_t(\theta)\boldsymbol{\zeta}_t\boldsymbol{x}_t'\right)\left(\frac{1}{T}\sum_{t=1}^{T}\hat{w}_t(\theta)\boldsymbol{x}_t\boldsymbol{x}_t'\right)^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\hat{w}_t(\theta)\boldsymbol{x}_t\varepsilon_t(\theta).
\end{aligned}$$

By (A.24) of Newey and Powell (1987),

$$\left|\frac{1}{T}\sum_{t=1}^{T}\hat{w}_t(\theta)\boldsymbol{x}_t\boldsymbol{x}_t' - \frac{1}{T}\sum_{t=1}^{T}w_t^o(\theta)\boldsymbol{x}_t\boldsymbol{x}_t'\right| \xrightarrow{\mathbb{P}} 0,$$

where $|A|$ denotes the maximum norm of the matrix A. Similarly,

$$\left|\frac{1}{T}\sum_{t=1}^{T}\hat{w}_t(\theta)\boldsymbol{\zeta}_t\boldsymbol{x}_t' - \frac{1}{T}\sum_{t=1}^{T}w_t^o(\theta)\boldsymbol{\zeta}_t\boldsymbol{x}_t'\right| \xrightarrow{\mathbb{P}} 0,$$

A suitable law of large numbers ensure that $T^{-1}\sum_{t=1}^{T} w_t^o(\theta)\boldsymbol{x}_t\boldsymbol{x}_t' \xrightarrow{\mathbb{P}} \boldsymbol{\Xi}(\theta)$ and

$$\frac{1}{T}\sum_{t=1}^{T} w_t^o(\theta)\boldsymbol{\zeta}_t\boldsymbol{x}_t' \xrightarrow{\mathbb{P}} \mathbb{E}[w_t^o(\theta)\boldsymbol{\zeta}_t\boldsymbol{x}_t'] =: \boldsymbol{\Gamma}(\theta).$$

It follows that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} \hat{w}_t(\theta)\boldsymbol{\zeta}_t\hat{e}_t(\theta) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\boldsymbol{\zeta}_t - \boldsymbol{\Gamma}(\theta)\boldsymbol{\Xi}(\theta)^{-1}\boldsymbol{x}_t\right)\hat{w}_t(\theta)\varepsilon_t(\theta) + o_{\mathbb{P}}(1). \tag{14}$$

This is the basis of the proposed non-nested test.

Recall that

$$\sqrt{T}\big(\hat{\boldsymbol{\beta}}_T(\theta) - \boldsymbol{\beta}^*(\theta)\big) = -\boldsymbol{\Xi}(\theta)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\hat{w}_t(\theta)\boldsymbol{x}_t e_t^*(\theta)\right) + o_p(1).$$

In view of the proof of Theorem 3.2, we conclude that $T^{-1/2}\sum_{t=1}^{T}\hat{w}_t(\theta)\boldsymbol{x}_t e_t^*(\theta)$ is asymptotically equivalent to $T^{-1/2}\sum_{t=1}^{T} w_t^*(\theta)\boldsymbol{x}_t e_t^*(\theta)$ which is asymptotically normally distributed. A similar conclusion also holds for $T^{-1/2}\sum_{t=1}^{T}\hat{w}_t(\theta)\boldsymbol{\zeta}_t e_t^*(\theta)$. Under the null hypothesis, $e_t^*(\theta) = \varepsilon_t(\theta)$, and (14) is such that

$$\begin{aligned}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\hat{w}_t(\theta)\boldsymbol{\zeta}_t\hat{e}_t(\theta) &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\big(\boldsymbol{\zeta}_t - \boldsymbol{\Gamma}(\theta)\boldsymbol{\Xi}(\theta)^{-1}\boldsymbol{x}_t\big)w_t^o(\theta)\varepsilon_t(\theta) + o_{\mathbb{P}}(1)\\ &\xrightarrow{D} \mathcal{N}(\mathbf{0},\ \boldsymbol{\Omega}(\theta)),\end{aligned} \tag{15}$$

where $\boldsymbol{\Omega}(\theta) = \mathbb{E}\big[w_t^o(\theta)^2\varepsilon_t^2(\theta)\big(\boldsymbol{\zeta}_t - \boldsymbol{\Gamma}(\theta)\boldsymbol{\Xi}(\theta)^{-1}\boldsymbol{x}_t\big)\big(\boldsymbol{\zeta}_t - \boldsymbol{\Gamma}(\theta)\boldsymbol{\Xi}(\theta)^{-1}\boldsymbol{x}_t\big)'\big]$ by the martingale difference property (12). Note that $\boldsymbol{\Omega}$ has rank $q \le m$, where m is the dimension of $\boldsymbol{\zeta}_t$. For example, q may be the number of elements in $\boldsymbol{\zeta}_t$ that are not included in $\boldsymbol{x}_t$.

It follows from (15) that the proposed test statistic is:

$$\frac{1}{T}\left(\sum_{t=1}^{T}\hat{w}_t(\theta)\boldsymbol{\zeta}_t\hat{e}_t(\theta)\right)\left(\widehat{\boldsymbol{\Omega}}(\theta)^-\right)\left(\sum_{t=1}^{T}\hat{w}_t(\theta)\boldsymbol{\zeta}_t\hat{e}_t(\theta)\right)' \xrightarrow{D} \chi^2(q), \tag{16}$$

where $\widehat{\boldsymbol{\Omega}}(\theta)^-$ is the generalized inverse of the consistent estimator, $\widehat{\boldsymbol{\Omega}}(\theta)$, for $\boldsymbol{\Omega}(\theta)$. This is a conditional expectile encompassing test, analogous to the conditional mean encompassing test of Wooldrdige (1990). Note that a consistent estimator of $\boldsymbol{\Omega}(\theta)$ is

$$\begin{aligned}\widehat{\boldsymbol{\Omega}}_T(\theta) = \frac{1}{T}\sum_{t=1}^{T}\Bigg\{\hat{w}_t^2(\theta)\hat{e}_t^2&\left[\boldsymbol{\zeta}_t - \left(\frac{1}{T}\sum_{t=1}^{T}\hat{w}_t(\theta)\boldsymbol{\zeta}_t\boldsymbol{x}_t'\right)\left(\frac{1}{T}\sum_{t=1}^{T}\hat{w}_t(\theta)\boldsymbol{x}_t\boldsymbol{x}_t'\right)^{-1}\boldsymbol{x}_t\right]\\ &\left[\boldsymbol{\zeta}_t - \left(\frac{1}{T}\sum_{t=1}^{T}\hat{w}_t(\theta)\boldsymbol{\zeta}_t\boldsymbol{x}_t'\right)\left(\frac{1}{T}\sum_{t=1}^{T}\hat{w}_t(\theta)\boldsymbol{x}_t\boldsymbol{x}_t'\right)^{-1}\boldsymbol{x}_t\right]'\Bigg\}.\end{aligned}$$

Table 2: Summary statistics of stock market indices.

Index	Mean	Median	Max	Min	S. Dev.	Skew.	Kurt.
S&P500	0.0162	0.0262	2.420	−3.089	0.469	−0.112	6.560
NASDAQ	0.0167	0.0577	5.756	−4.416	0.740	0.012	7.539

$\boldsymbol{\Omega}(\theta)$ may also be estimated using a suitable bootstrap method.

Remark: Let $\tilde{\boldsymbol{\zeta}}_t$ denote the sub-vector of $\boldsymbol{\zeta}_t$ that is not in the linear space spanned by the variables in $\boldsymbol{x}_t$. Then, the encompassing test (16) may be computed as

$$\frac{1}{T}\left(\sum_{t=1}^{T}\hat{w}_t(\theta)\tilde{\boldsymbol{\zeta}}_t\hat{e}_t(\theta)\right)\left(\widetilde{\boldsymbol{\Omega}}_T(\theta)^{-1}\right)\left(\sum_{t=1}^{T}\hat{w}_t(\theta)\tilde{\boldsymbol{\zeta}}_t\hat{e}_t(\theta)\right)' \xrightarrow{D} \chi^2(q),$$

where $\widetilde{\boldsymbol{\Omega}}_T(\theta)$ has rank q and is computed as $\widehat{\boldsymbol{\Omega}}(\theta)$, with $\boldsymbol{\zeta}_t$ replaced by $\tilde{\boldsymbol{\zeta}}_t$.

4 Empirical Study

To illustrate the proposed CARE model, we conduct a simple empirical study to assess the value at risk of some stock indices. For each index, we shall select an appropriate CARE model specification and then evaluate both in-sample and out-of-sample performance of the selected model.

4.1 Data and Computation

Our study focuses on two indices: S&P500 and NASDAQ. The daily data of these indices are taken from Datastream; the sample period is from Jan 03, 1995 to Dec. 29, 2006 with 3023 observations. The daily return of an index is computed as 100 times the first difference of the log transformation of the index. Table 2 collects the summary statistics of these daily returns. We find that both returns have mean close to zero and standard deviations less than one. Also, they are slightly skewed and have excess kurtosis. In particular, NASDAQ has a wider range (longer tails) and also a larger standard deviation and kurtosis coefficient. This can also be seen from its histogram and estimated density in Figure 3, where the densities are obtained from STATA based on the Epanechnikov kernel. The return series plotted in Figure 4 also reveal that large values of NASDAQ index return mainly occur during 1999–2001, the period of dot-com bubble.

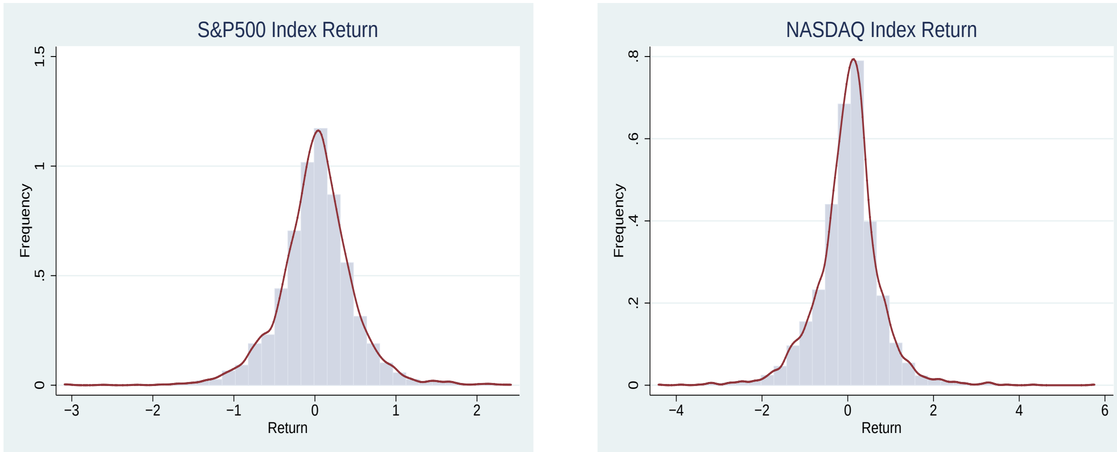

Figure 3: Kernel densities of stock index returns

In our empirical analysis, the first 2522 observations from 1995 to 2004 are used for model estimation and the remaining 500 observations are reserved for evaluating the out-of-sample performance of the selected model. As far as model estimation is concerned, we follow Newey and Powell (1987) and adopt the iterated weighted least squares (IWLS) algorithm. For each model, we use the OLS estimates as the initial values for the IWLS estimates and iterate till the estimates converge (the convergence criterion is set at 10^{-12}). The estimation program is coded in GAUSS.

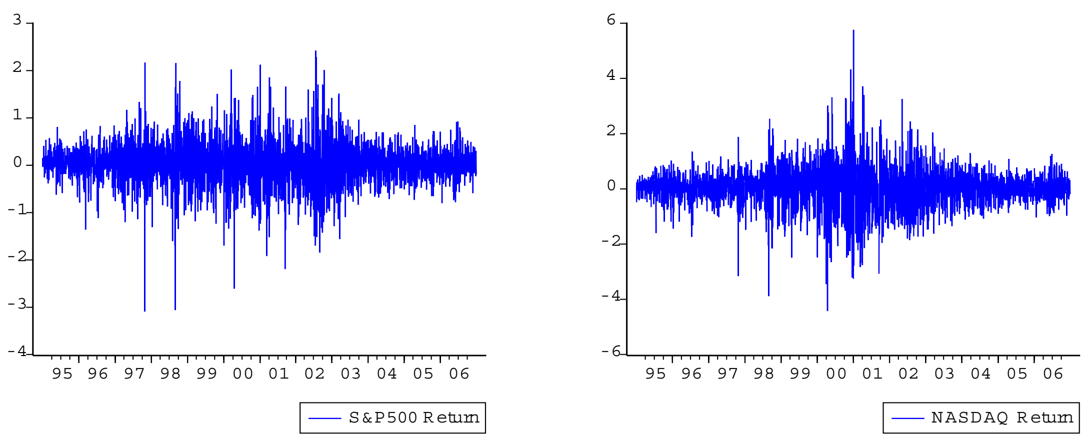

Figure 4: Stock return series: 1995–2006

4.2 Empirical Results

For the empirical study, we consider two class of CARE models discussed in Section 3.1. The first class is a simpler form of model (9):

$$y_t = a_0(\theta) + a_1(\theta)y_{t-1} + b_1(\theta)(y_{t-1}^+)^2 + \gamma_1(\theta)(y_{t-1}^-)^2 + \cdots$$
$$+ b_q(\theta)(y_{t-q}^+)^2 + \gamma_q(\theta)(y_{t-q}^-)^2 + e_t(\theta),$$

where y_{t-1} is admitted; higher order lags enter the model only in terms of squares. This will be referred to as an SQ(q) model. We do not include other y_{t-i}, $i \geq 2$, in SQ models because they are typically insignificant and also because their presence may affect the significance of other included variables. The second class is model (10):

$$y_t = a_0(\theta) + \delta_1(\theta)y_{t-1}^+ + \lambda_1(\theta)y_{t-1}^- + \cdots + \delta_q(\theta)y_{t-q}^+ + \lambda_q(\theta)y_{t-q}^- + e_t(\theta),$$

which will be referred to as an ABS(q) model.

We first determine the number of lags in each class of models. To this end, we estimate each model with $q = 4$ and test the significance of parameter estimates. When the estimates of b_4 and γ_4 in the SQ(4) model (or δ_4 and λ_4 in the ABS(4) model) are *both* insignificant, we drop the lag-4 variables and re-estimate the SQ(3) (or ABS(3)) model. Otherwise, we keep the SQ(4) (or ABS(4)) model. Note that, for a given lag, the positive and negative parts of the lagged variable are both kept in the model as long as at least one of their parameter estimates is significant. This allows us to examine the asymmetry effects on conditional expectiles. We can proceed to check if we want to stay with the SQ(3) (or ABS(3)) model, and so on. When appropriate SQ and ABS models are chosen, we apply the encompassing test introduced in Section 3.3 to determine the final model.

For $\theta = 0.05$, our estimation and significance test (at 5% level) results lead to SQ(2) and ABS(3) models for S&P500 and SQ(3) and ABS(4) models for NASDAQ. For S&P500, the encompassing test of SQ(2) against ABS(3) yields a statistic of 9.09 with p-value 16.9%, and the test statistic of ABS(3) against SQ(2) is 8.34 with p-value 8.0%. Hence, the encompassing test rejects ABS(3) model at 10% level but can not reject SQ(2) model at the same level. For NASDAQ, the encompassing test statistic of SQ(3) against ABS(4) is 37.275 with p-value less than 0.1%, and the statistic of ABS(4) against SQ(3) is 9.651 with p-value 14.01%. This shows that SQ(3) model is rejected at a very small level and that ABS(4) model can not be rejected at 10% level. Thus, we set the final models for

Table 3: The parameter estimates of the CARE models: $\theta = 0.05$.

S&P500: SQ(2)			NASDAQ: ABS(4)		
Variable	Estimate	(s.e.)	Variable	Estimate	(s.e.)
cons.	-1.113	(.049)***	cons.	-0.565	(.060)***
y_{t-1}	0.360	(.085)***	y_{t-1}^{+}	-0.022	(.060)
$(y_{t-1}^{+})^2$	-0.121	(.028)***	y_{t-1}^{-}	-0.269	(.088)***
$(y_{t-1}^{-})^2$	0.085	(.039)**	y_{t-2}^{+}	-0.182	(.055)***
$(y_{t-2}^{+})^2$	-0.018	(.014)	y_{t-2}^{-}	-0.389	(.078)***
$(y_{t-2}^{-})^2$	-0.126	(.054)**	y_{t-3}^{+}	-0.107	(.061)*
			y_{t-3}^{-}	-0.168	(.069)**
			y_{t-4}^{+}	-0.220	(.069)***
			y_{t-4}^{-}	-0.218	(.074)***
in-sample tail prob.:		10.2%	in-sample tail prob.:		11.6%
out-of-sample tail prob.:		3.2%	out-of-sample tail prob.:		4.2%
out-of-sample θ:		0.7%	out-of-sample θ:		1.1%

Note: $*$, $**$ and $***$ label significance at 10%, 5% and 1% levels, respectively.

S&P500 and NASDAQ as SQ(2) and ABS(4), respectively; their parameter estimates are summarized in Table 3.

For S&P500, it can be seen that the effects of $(y_{t-1}^{+})^2$ and $(y_{t-1}^{-})^2$ in the SQ(2) model have opposite signs, but the effects of $(y_{t-2}^{+})^2$ and $(y_{t-2}^{-})^2$ have the same negative sign. We also find that, apart from the sign, the effects of $(y_{t-i}^{+})^2$ and $(y_{t-i}^{-})^2$ are not significantly different at 5% level (but the effects of $(y_{t-2}^{+})^2$ and $(y_{t-2}^{-})^2$ are significantly different at 10% level). For NASDAQ, most coefficients in the ABS(4) model are significantly negative. The test statistics indicate that only the effects of y_{t-1}^{+} and y_{t-1}^{-} are significantly different at 5% level; other positive and negative parts virtually have the same effect (the effects of y_{t-2}^{+} and y_{t-2}^{-} are significantly different at 10% level).

We also calculate the in-sample tail probability for the estimated expectile, i.e., the percentage that y_t falls below the estimated expectiles. For S&P500, the tail probability is 10.2%; for NASDAQ, it is about 11.6%. Hence, when the index of prudentiality $\theta = 5\%$ is our concern, QVaR at 5% level would be too conservative. In the light of Table 1, we may also infer that the tail of the conditional distribution for S&P500 is close to that of $t(5)$ and the tail for NASDAQ is close to that of $t(10)$. We also find that the out-of-sample tail probabilities for both indices are much smaller: 3.2% for S&P500 and 4.2% for NASDAQ. The out-of-sample θ for these indices are about 1% and also smaller than the 5% level in the in-sample period. From Figure 4 we can see that both indices are more volatile and have quite extreme observations during the in-sample period. But these indices only fluctuate slightly in the out-of-sample period. This may explain why the out-of-sample tail probabilities and θ are all much smaller than their in-sample counterparts.

To be sure, we re-estimate CARE models and evaluate their performance based on a more volatile sub-sample. The in-sample period is from Jan. 2, 1996 to Dec. 31, 2002; the data in 2003 are reserved for out-of-sample evaluation. Following the procedure above, we come up with SQ(3) model for S&P500 and ABS(5) model for NASDAQ. To save space, we do not report the detailed estimation and test results (these results are available upon request). For S&P500 and NASDAQ, the in-sample tail probabilities are, respectively, 10.7% and 12.1%, which are similar to those obtained from the complete sample. The out-of-sample tail probabilities for both indices are about 7.1%, which are closer to their in-sample counterparts. The out-of-sample θ's (3.0% for S&P500 and 2.4% for NASDAQ) are also closer to the pre-set 5% level. These results confirm that the performance of CARE models hinges on the sample used for model estimation.

5 Concluding Remarks

In this paper we propose a new downside risk measure, EVaR. Compared with the conventional QVaR measure, EVaR has two advantages. First, EVaR is more sensitive to the magnitude of extreme losses. Second, EVaR with a given index of prudentiality may be viewed as a flexible QVaR, in the sense that its tail probability is not set a priori but is determined by the underlying distribution. To implement this measure, we propose various CARE models for EVaR and discuss model estimation and specification test. These constitute an alternative to existing methods for assessing downside risk, such as the CAViaR model for conditional QVaR. In practice, our approach may be further improved by finding other CARE model specifications that can better characterize the dynamic behavior of conditional expectiles. Moreover, it is also important to establish an intuitive criterion to determine a proper index of prudentiality for EVaR. These topics are not fully addressed in this paper and are left to future research.

Appendix

Regularity Conditions:

[A1] $\boldsymbol{z}_t = (y_t, \boldsymbol{x}_t')'$ is strictly stationary and ergodic and has the probability density function $f(\boldsymbol{z}_t) = g(y_t|\boldsymbol{x}_t)h(\boldsymbol{x}_t)$ with respect to the measure $\nu_z = \eta \times \nu_x$, where $f(\boldsymbol{z}_t)$ is continuous in y_t for almost all $\boldsymbol{x}_t$, and η denotes the Lebesque measure on the real line. Also, $\mathbb{E}(\boldsymbol{x}_t\boldsymbol{x}_t')$ is of full rank k.

[A2] There is $\delta > 0$ such that $\int |\boldsymbol{z}|^{4+\delta} f(y|\boldsymbol{x})h(\boldsymbol{x})d\nu_z < \infty$.

[A3] $\boldsymbol{\beta}(\theta) \in \mathcal{B} \subseteq \mathcal{R}^k$, where $\mathcal{B}$ is compact.

[A4] There is a positive K such that $\boldsymbol{V}_T = \operatorname{var}\left(T^{-1/2}\sum_{t=1}^T w_t^*(\theta)\boldsymbol{x}_t e_t^*(\theta)\right) \leq K$, where $e_t^*(\theta) = y_t - \boldsymbol{x}_t'\boldsymbol{\beta}^*(\theta)$.

Proof of Theorem 3.1: We verify the conditions M.1–M.3 imposed in Theorem 4.3 of Wooldridge (1994) for $\rho_\theta(y_t - \boldsymbol{x}_t'\boldsymbol{\beta}(\theta))$. First, it is easy to see that M.1 holds under [A1] and [A3]. For M.2, we must show that $\rho_\theta(y_t - \boldsymbol{x}_t'\boldsymbol{\beta}(\theta))$ obeys a weak uniform law of large numbers. In the light of Theorem 4.1 of Wooldridge (1994), it remains to show that $\rho_\theta(y_t - \boldsymbol{x}_t'\boldsymbol{\beta}(\theta))$ is dominated by an integrable function for all $\boldsymbol{\beta}(\theta) \in \mathcal{B}$. To this end, note that there exist constants $d_1, d_2, M > 0$,

$$|\rho_\tau(y_t - \boldsymbol{x}_t'\boldsymbol{\beta}(\theta))| \leq |\boldsymbol{z}_t|^2(d_1 + d_2|\boldsymbol{\beta}(\theta)|^2) \leq |\boldsymbol{z}_t|^2 M,$$

where the last inequality follows because $\mathcal{B}$ is compact. The right-hand side is clearly integrable by [A2] and does not depend on $\boldsymbol{\beta}(\theta)$, so that M.2 holds. Note that by strict stationarity of $\boldsymbol{z}_t$ and an argument similar to that of Theorem 3 in Newey and Powell (1987), there exists a unique minimizer, $\boldsymbol{\beta}^*(\theta)$, of $\mathbb{E}[\rho_\theta(y_t - \boldsymbol{x}_t'\boldsymbol{\beta}(\theta))]$, as required by M.3. The assertion follows from Theorem 4.3 of Wooldridge (1994). $\square$

Proof of Theorem 3.2: The proof is similar to that for Theorem 3 of Newey and Powell (1987). When the order of expectation and differentiation can be exchanged, let

$$\boldsymbol{\lambda}_\theta(\boldsymbol{\beta}(\theta)) := \nabla_{\boldsymbol{\beta}}\mathbb{E}\big[\rho_\theta(y_t - \boldsymbol{x}_t'\boldsymbol{\beta}(\theta))\big]/2 = -\mathbb{E}\big[w(e_t(\theta);\theta)\boldsymbol{x}_t e_t(\theta)\big],$$
$$\nabla_{\boldsymbol{\beta}}\boldsymbol{\lambda}_\theta(\boldsymbol{\beta}(\theta)) = \mathbb{E}[w(e_t(\theta);\theta)\boldsymbol{x}_t\boldsymbol{x}_t'].$$

Clearly, $\boldsymbol{\lambda}_\theta(\boldsymbol{\beta}^*(\theta)) = \boldsymbol{0}$. It can be verified that conditions [A1]–[A3] are sufficient for (N-1)–(N-3) of Huber (1967) and hence Lemma 3 of Huber (1967). Note that Huber's proof requires only the first order Chebyshev's inequality and hence is not affected by weak dependence of the data imposed in [A1]. Lemma 3 of Huber (1965) and [A4] together imply:

$$\sqrt{T}\boldsymbol{\lambda}_\theta(\hat{\boldsymbol{\beta}}_T(\theta)) + \frac{1}{\sqrt{T}}\sum_{t=1}^{T} w_t^*(\theta)\boldsymbol{x}_t e_t^*(\theta) = o_{\mathbb{P}}(1).$$

The proof of this result requires the second order Chebyshev's inequality. Hence, the uniform boundedness of $\boldsymbol{V}_T(\theta)$ imposed in [A4] is needed; see also Theorem 3 of Huber (1967). By mean value expansion of $\boldsymbol{\lambda}_\theta(\hat{\boldsymbol{\beta}}_T(\theta))$ around $\boldsymbol{\beta}^*(\theta)$,

$$\begin{aligned}\sqrt{T}\boldsymbol{\lambda}_\theta(\hat{\boldsymbol{\beta}}_T(\theta)) &= -\frac{1}{\sqrt{T}}\sum_{t=1}^{T} w_t^*(\theta)\boldsymbol{x}_t e_t^*(\theta)\\ &= \nabla_{\boldsymbol{\beta}}\boldsymbol{\lambda}_\theta(\ddot{\boldsymbol{\beta}}_T(\theta))\sqrt{T}\big(\hat{\boldsymbol{\beta}}_T(\theta) - \boldsymbol{\beta}^*(\theta)\big) + o_p(1),\end{aligned}$$

where $\ddot{\boldsymbol{\beta}}_T(\theta)$ denotes the mean value. Hence,

$$\sqrt{T}\big(\hat{\boldsymbol{\beta}}_T(\theta) - \boldsymbol{\beta}^*(\theta)\big) = -\Big(\nabla_{\boldsymbol{\beta}}\boldsymbol{\lambda}_\theta(\ddot{\boldsymbol{\beta}}_T(\theta))\Big)^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T} w_t^*(\theta)\boldsymbol{x}_t e_t^*(\theta) + o_{\mathbb{P}}(1).$$

The consistency of $\hat{\boldsymbol{\beta}}_T(\theta)$ implies that $\ddot{\boldsymbol{\beta}}_T(\theta))$ also converges to $\boldsymbol{\beta}^*(\theta)$. By the continuity of $\nabla_{\boldsymbol{\beta}}\boldsymbol{\lambda}_\theta(\boldsymbol{\beta}(\theta))$, we have $\nabla_{\boldsymbol{\beta}}\boldsymbol{\lambda}_\theta(\ddot{\boldsymbol{\beta}}_T(\theta)) \xrightarrow{\mathbb{P}} \boldsymbol{\Xi}(\theta)$. It follows that

$$\sqrt{T}\big(\hat{\boldsymbol{\beta}}_T(\theta) - \boldsymbol{\beta}^*(\theta)\big) = -\boldsymbol{\Xi}(\theta)^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T} w_t^*(\theta)\boldsymbol{x}_t e_t^*(\theta) + o_{\mathbb{P}}(1).$$

By [A1] and [A2], a central limit theorem for stationary sequence yields:

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} w_t^*(\theta)\boldsymbol{x}_t e_t^*(\theta) \xrightarrow{D} \mathcal{N}(\mathbf{0},\ \boldsymbol{V}(\theta)).$$

These results together ensure the desired conclusion. □

References

Aït-Sahalia, Y. and A. Lo (2000). Nonparametric risk management and implied risk aversion, *Journal of Econometrics*, **94**, 9–51.

Baer, H. L., V. G. France, and J. T. Moser (1994). Opportunity cost and prudentiality: An analysis of futures clearinghouse behavior, Policy Research Working Paper 1340, The World Bank Policy Research Department, New York.

Bassett, G. W., R. Koenker, and G. Kordas (2004). Pessimistic portfolio allocation and Choquet expected utility, *Journal of Financial Econometrics*, **2**, 477–492.

Black, F. (1976). Studies of stock price volatility changes, in *Proceedings of the 1976 Meetings of the Business and Economic Statistics Section, American Statistical Association*, 177–181.

Booth, G. G., J. P. Broussard, T. Martikainen, and V. Puttonen (1997). Prudent margin levels in the Finnish stock index futures market, *Management Science*, **43**, 1177–1188.

Christie, A. (1982). The stochastic behavior of common stock variances: Value, leverage, and interest rate effects, *Journal of Financial Economics*, **10**, 407–432.

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

Efron, B. (1991). Regression percentiles using asymmetric squared error loss, *Statistica Sinica*, **1**, 93–125.

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

Engle R. F. and V. Ng (1993). Measuring and testing the impact of news in volatility, *Journal of Finance*, **48**, 1749–1778.

Fishburn, P. (1977). Mean-risk analysis with risk associated with below-target returns, *American Economic Review*, **67**, 116–125.

Glosten, L., R. Jagannathan, and D. Runkle (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks, *Journal of Finance*, **48**, 1779–1801.

Gourieroux C. and J. Jasiak (2002). Value at Risk, Working paper, CREST.

Huber, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions, *Proceedings of the 5th Berkeley Symposium*, **1**, 221–233.

Jorion, P. (2000). *Value at Risk: The New Benchmark for Managing Financial Risk*, Chicago: McGraw-Hill.

Kahneman, I. and A. Tversky (1979). Prospect theory: An analysis of decision under risk, *Econometrica*, **47**, 263–290.

Koenker, R. and G. Bassett (1978). Regression quantiles, *Econometrica*, **46**, 33–50.

Koenker, R. (1992). When are expectiles percentiles, *Econometric Theory*, **8**, Problem: 423–424; Solutions: 526–527.

Lam, K., C.-Y. Sin, and R. Leung (2004). A theoretical framework to evaluate different margin-setting methodologies, *Journal of Futures Markets*, **24**, 117–145.

Markowitz, H.˜(1952). Portfolio selection, *Journal of Finance*, **7**, 77–91.

Nelson, D. (1991). Conditional heteroskedasticity in asset returns: A new approach, *Econometrica*, **59**, 347–370.

Newey, W. K. and J. L. Powell (1987). Asymmetric least squares estimation and testing, *Econometrica*, **55**, 819–847.

Sin, C.-Y. and C. W. J. Granger (1999). Estimating and forecasting quantiles with asymmetric least squares, Working paper, University of California, San Diego.

Taylor, J. W. (2008). Estimating value at risk and expected shortfall using expectiles, *Journal of Financial Econometrics*, **6**, 231–252.

Woodridge, J. M. (1990). An encompassing approach to conditional mean tests with applications to testing non-nested hypotheses, *Journal of Econometrics*, **45**, 331–350.

Woodridge, J. M. (1994). Estimation and inferences for dependent processes, in *Handbook of Econometrics*, Volume 4, R. F. Engle and D. L. McFadden (eds.), Chap. 45, pp. 2638–2691, Elsevier.

Yao, Q. and H. Tong (1996). Asymmetric least squares regression and estimation: A nonparametric approach, *Nonparametric Statistics*, **6**, 273–292.