Statistics and Its Interface Volume 8 (2015) 295–303

A regression analysis of expected shortfall*

Zongwu Cai†, Jia Su, and Sufianti

To estimate the expected shortfall, a coherent risk measure, this paper proposes an easily implemented regression technique based on a proportional mean residual life regression model with explanatory (lagged) variables. The parameters are estimated by using a quasi-likelihood method and the asymptotic normality of the proposed estimator is derived under an α-mixing process assumption. Based on a simulation study, the proposed estimator performs fairly well. In the empirical study, the backtesting procedure is conducted based on the daily and weekly return of S&P 500 Index using the 95% confidence level. The performance of the model is evaluated by its ability to accurately estimate ES compared with two more alternative models. The results generally favor the proposed model over the alternative models.

AMS 2000 subject classifications: Primary 91B30; secondary 37M10.
Keywords and phrases: Backtesting method, Coherent risk, Expected shortfall, Proportional mean residual life function, Quasi-likelihood estimation, Risk management, Semiparametric modeling.

1. INTRODUCTION

In today's financial world, risk management has taken an increasingly important role in sustaining an institution's financial stability and its ability to remain economically viable. Following several financial crises in global financial markets, accurate and reliable risk measures have become essential to cope with future adverse events. Value-at-risk (VaR) and expected shortfall (ES) are two of the more well-known quantitative risk measures for a portfolio of asset(s) adopted by financial institutions and regulators.

VaR, being conceptually simpler, is described as

$$\begin{aligned} \text{VaR} &= -\inf\{x \in \mathbb{R} : P(x_t > x) \leqslant 1-\alpha\} \\ &= -\inf\{x \in \mathbb{R} : F(x) > \alpha\} = -F^{-1}(\alpha) \end{aligned}$$

at a given $\alpha \in (0,1)$, where $F(\cdot)$ represents the distribution function of x_t (say, the log return, where $-x_t$ is the loss) with its inverse $F^{-1}(\cdot)$, and $(1-\alpha)\times 100\%$ is the confidence level. However, by simply being the threshold of possible losses, VaR does not give any information about the severity of losses by which it is exceeded. On the other hand, ES, formally defined as

$$\text{ES} = -E[x_t \mid -x_t > \text{VaR}] = -\frac{1}{\alpha}\int_{-\infty}^{-\text{VaR}} x f(x)dx \geq \text{VaR}$$

or the conditional expected value of losses given that the losses are larger than VaR, takes into consideration the magnitude of the losses beyond VaR. Furthermore, VaR violates the subadditivity principle, one of the four axioms of a coherent risk measure (Artzner, Delbaen, Eber and Heath, 1999), in the sense that it does not promote portfolio diversification, while ES is coherent; see Acerbi and Tasche (2002) for details. Thus, ES is preferred for practical applications.

Various quantitative models, from parametric to nonparametric, characterize ES. For a fully parametric estimation of ES, the unconditional return, or log-return series, is often assumed to follow a certain distribution, e.g., a normal or Student's t-distribution, or an other type of distribution. Alternatively, the conditional return and variance can be assumed to follow some time series model, such as the autoregressive moving average (ARMA) model (Box and Jenkins, 1970) for returns and generalized autoregressive conditional heteroskedasticity (GARCH) model (Bollerslev, 1986) for variance. Using parametric models is simple and requires little information, but misspecification of models and distributions will produce misleading results. For nonparametric ES estimation, the commonly used methods for nonparametric ES estimation include historical simulation (HS) and Monte Carlo simulation. Other models found in the literature include Scaillet (2004, 2005), Cai and Wang (2008), and Chen (2008), and the references therein. While no distributional assumption is needed, the biggest drawbacks of using pure nonparametric models are the complexity and high computational cost.

One might prefer using an easily implemented semiparametric model which incorporates parametric and nonparametric components with possible explanatory variables. By requiring no distributional assumption and working efficiently when only little information is available, a semiparametric model combines the merits of both paramet-

*We thank the participants of the 2012 FERM Conference in Changsha, China. Cai's research is partially supported by the National Natural Science Foundation of China grants #71131008, #70971113 and #71271179. Su's research was supported by the Scholarship Award for Excellent Doctoral Student granted by the Ministry of Education and the Excellent Doctoral Student Training Program granted by Xiamen University.
†Corresponding author.

ric and nonparametric models. Extreme value theory (Embrechts, Kluppeberg and Mikosch, 1999) and filtered historical simulation (Barone-Adesi, Bourgoin and Giannopoulos, 1998, 1999) are two semiparametric approaches that can be applied to estimate ES. A new semiparametric model is adopted from a popular concept in actuarial studies. The model, termed the proportional mean residual life (PMRL) model, is originally proposed by Oakes and Dasu (1990) to measure human life expectancy. To account for the presence of covariates in regression analysis, Maguluri and Zhang (1994) extended the PMRL model to operate in a more general framework.

However, the original PMRL regression model was developed for independent and identically distributed (i.i.d.) data. As seen by practitioners in the financial market, making an i.i.d. assumption for financial returns is completely unreasonable. A financial return series shows little or no serial correlation while its absolute and squared counterparts are often found to be highly serially correlated (with positive and significantly greater than zero autocorrelations), resulting in time-varying volatility and volatility clustering. Since small (large) returns are almost always followed by small (large) returns, returns cannot be assumed to be i.i.d.

We address this by proposing a modification to the asymptotic normality of the PMRL regression model to account for the peculiarities of financial data. The resulting asymptotic properties are derived under an α-mixing process assumption. Next, a simulation study is performed to evaluate the performance of the regression parameter estimates with a time series assumption. Finally, we perform an empirical study with real market outcomes to compare the accuracy and reliability of the original and the modified PMRL regression models in ES estimation, under both the i.i.d. and time series assumptions, along with other generally known models using the backtesting method.

The rest of the paper is structured as follows. Section 2 describes the model and derives the asymptotic properties of the estimator, and also presents a significance test to see whether the explanatory variables can be used to explain the risk in the model. Section 3 displays the simulation and real data application of the model. Section 4 outlines the main findings and conclusions of our study. All technical details are given in the Appendix.

2. ANALYSIS FRAMEWORK

2.1 The model and estimation procedure

The mean residual life function $e(x)$ of a non-negative random variable x_t with survival function $S(x)$ and finite mean μ is

$$e(x) = E(x_t - x | x_t > x) = S(x)^{-1} \int_x^\infty S(u)du.$$

Then, $S(x)$ can be derived from $e(x)$ by the inversion formula

$$S(x) = \frac{e(0)}{e(x)} \exp\left\{ - \int_0^x e(u)^{-1} du \right\}.$$

Two survival functions $S_0(x)$ and $S_1(x)$ are said to have proportional mean residual life if in an obvious notation

$$(1) \qquad e_1(x) = \theta e_0(x) \ \text{ for all } \ x \geqslant 0, \ \theta > 0.$$

Therefore,

$$S_1(x) = S_0(x) \left\{ \int_x^\infty S_0(u)du/\mu_0 \right\}^{1/\theta - 1}.$$

It is well known that a necessary and sufficient condition for $S_1(x)$ to be a survival function for all $\theta > 0$ is that $e_0(x)$ is nondecreasing; see Maguluri and Zhang (1994) for details. Oakes and Dasu (1990) proposed a semparametric PMRL model, whereas Maguluri and Zhang (1994) extended this model to a regression context. Model (1) can be extended to a more general setting with explanatory variable z,

$$(2) \qquad e(x|z) = \exp(-\beta^T z) e_0(x).$$

Clearly, $\beta = \ln[e(x|z = 0)/e(x|z = 1)]$ which measures the risk so that β is commonly called the log of risk. It follows from (2) that for some baseline survival function $S_0(\cdot)$,

$$(3) \qquad S(x|z) = S_0(x) \left\{ \int_x^\infty S_0(u)du/\mu_0 \right\}^{\exp(\beta^T z) - 1},$$

where the corresponding baseline mean is $\mu_0 = E[\exp(\beta^T z)x|z]$; see Maguluri and Zhang (1994) for details. In the PMRL model, Dasu and Oakes (2003) proposed a class of weighted ratio estimators. Suppose that we observe time series sample $\{(x_t, z_t)\}_{t=1}^T$ from a population satisfying (2). Our method of constructing an estimator $\hat{\beta}$ of β is based on the solution of

$$(4) \ U(\hat{\beta}) = \frac{\frac{1}{T}\sum_{t=1}^T x_t z_t \exp(\hat{\beta}^T z_t)}{\frac{1}{T}\sum_{t=1}^T x_t \exp(\hat{\beta} z_t)} - \frac{1}{T}\sum_{t=1}^T z_t = 0.$$

Clearly, when the baseline survival function is the exponential distribution $S_0(x) = \exp(x/\mu_0)$, then (4) is the true maximum likelihood equation of this exponential regression model. Therefore, the estimator given by (4) is called the quasi-likelihood estimate.

2.2 Asymptotic theories

In estimating VaR and ES, the financial data are commonly assumed to follow a certain time series model such as an ARMA or GARCH process. We consider a more general structure in the form of an α-mixing process. The asymptotic results of the paper are derived under the α-mixing assumption. In this section, to establish the large sample theory, we require the following assumptions.

Assumptions

A1. Parameter space $\Theta \subset \mathbb{R}^p$ is compact. Function $U(\beta)$ is a measurable function of observations for each $\beta \in \Theta$, and $U(\beta)$ is continuous in $\beta \in \Theta$. Objective function $\hat{U}(\beta)$ uniformly converges over Θ to $U_0(\beta) = E(U(\beta))$.
A2. Time series $\{(x_t, z_t^T)\}$ is stationary α-mixing. Further, assume that there exists some $\delta > 0$, such that $E|x_t \exp(\beta^T z_t)|^{2+\delta} < \infty$ and $E|x_t z_t \exp(\beta^T z_t)|^{2+\delta} < \infty$, and that there exists $r > 2$, such that the mixing coefficient $\alpha(i) = O(i^{-\tau})$, where $\tau = (\delta + r)/\delta$.
A3. There is a unique $\beta_0 \in \Theta$ such that $U(\beta_0) = 0$.
A4. The first order derivative $U'(\beta)$ is a negative semidefinite matrix that uniformly converges to the negative definite matrix $U_0' = E(U'(\beta))$.

Theorem 1 presents the consistency and asymptotic normality of the estimator. This result is new in the sense that it generalizes the result of Maguluri and Zhang (1994).

Theorem 1: Assume that Assumptions A1–A4 are satisfied. Then,

$$\sqrt{T}(\hat{\beta} - \beta_0) \rightarrow^d N\ (0, \Sigma_\beta),$$

where $\Sigma_\beta = A^{-1}V(A^{-1})^T$, $A = \mathrm{Var}(z_t)$ and $V = \mu_0^{-2}[\mathrm{Var}(\xi_t) + 2\sum_{j=1}^{\infty} \mathrm{Cov}(\xi_1, \xi_{j+1})]$ with $\xi_t = [x_t \exp(\beta_0^T z_t) - \mu_0](z_t - \mu_z)$.

The variance-covariance matrix Σ_β can be consistently estimated by using the heteroskedasticity and autocorrelation consistent (HAC) estimation method of Newey and West (1987), given by $\hat{\Sigma}_\beta = \hat{A}^{-1}\hat{V}(\hat{A}^{-1})^T$, where $\hat{A} = \frac{1}{T}\sum_{t=1}^{T}(z_t - \widehat{\mu}_z)(z_t - \widehat{\mu}_z)^T$ with $\widehat{\mu}_z = \frac{1}{T}\sum_{t=1}^{T} z_t$, and $\hat{V} = \widehat{\Gamma}_0 + \sum_{j=1}^{l} w_j\,[\widehat{\Gamma}_j + \widehat{\Gamma}_j^T]$ with $w_j = w(j)$, $\widehat{\Gamma}_j = \frac{1}{(T-j-p)\hat{\mu}_0^2}\sum_{t=1}^{T-j}\widehat{\xi}_t\widehat{\xi}_{t+j}^T$, l denoting the lag truncation parameter, and $w(\cdot)$ being the weighting function. In particular, when the data are independent, the variance-covariance matrix reduces to $\Sigma_\beta = A^{-1}\mathrm{Var}(\xi_t)(A^{-1})^T\mu_0^{-2}$, similar to that in Maguluri and Zhang (1994), and it further reduces to $\Sigma_\beta = A^{-1}$ when the baseline survival function is exponential. In our time series model, the variance-covariance matrix plays an important role in the inference. For the HAC covariance matrix, besides the weighting function $w(\cdot)$ such as the Barlett function defined by $w_j = w(j) = 1 - j/(l+1)$, other weighting functions can also be considered. Newey and West (1987) showed that if $l = l(T) \rightarrow \infty$ and $l^4/T \rightarrow 0$, $\hat{V}$ is a consistent estimator of V. Newey and West (1994) also pointed out that the choice of the kernel is not particularly important. The truncated parameter is a more important determinant of the finite sample property. In our paper, we propose using the Barlett function and choose the truncated parameter as in Schwert (1987), $l = \mathrm{int}\{4(T/100)^{1/4}\}$. To implement the HAC estimator, we can simply use the package *sandwich* in **R** with the command vcovHAC() or meatHAC().

2.3 Hypothesis testing

It is natural to investigate whether certain variables are statistically significant after fitting the model. This leads to the testing problem $H_0 : \beta = \beta_0$ versus $H_1 : \beta \neq \beta_0$. More generally, one may consider the pair of null and alternative hypotheses as $H_0 : R\beta = r$ versus $H_1 : R\beta \neq r$, where R is a $J \times p$ matrix with full rank and $J \leqslant p$. The testing problem can be validated by the Wald statistic as $W_T = T(R\hat{\beta} - r)^T(R\hat{A}^{-1}\hat{V}\hat{A}^{-1T}R^T)^{-1}(R\hat{\beta} - r)$. It is straightforward to show that when Assumptions A1–A4 are satisfied, under the null hypothesis, the distribution of the Wald statistic W_T can be approximated by a χ_J^2 distribution with J degrees of freedom. In particular, for testing $H_0 : \beta_j = \beta_{0j}$ versus $H_1 : \beta_j \neq \beta_{0j}$ for any $1 \leq j \leq p$, the test statistic is $t_j^2 = T(\hat{\beta}_j - \beta_{0j})^2/(\hat{A}^{-1}\hat{V}\hat{A}^{-1T})_{j,j}$, and $t_j^2 \rightarrow^d \chi_1^2$.

3. EMPIRICAL STUDIES

3.1 A simulated example

In this section, we evaluate the performance of estimator $\hat{\beta}$. A simulation study is carried out using sample sizes of $T = 100$, 200, and 500. We consider the case when $z_t = (z_{1t}, z_{2t})^T$ is generated by the AR(1) model as $z_{1t} = \alpha_1 z_{1(t-1)} + w_{1t}$ and $z_{2t} = \alpha_2 z_{2(t-1)} + w_{2t}$, where w_{1t} and w_{2t} are independently generated from $N(0, 1)$ and are mutually independent. Here α_1 and α_2 measure the degree of the dependency of $\{z_t\}$. To consider different dependencies, we consider cases where (α_1, α_2) take the values $(0, 0)$, $(0.5, 0.5)$ and $(0.9, 0.9)$, respectively. From (3), we have

$$S(x|z) = S_0(x)\left\{\int_x^\infty S_0(u)du/\mu_0\right\}^{\exp(\beta^T z)-1}.$$

Given covariate z_t and baseline survival function $S_0(x)$, x_t can be generated from this equation. Throughout the paper, we assume that the baseline distribution follows an exponential distribution, or $S_0(x) = \exp(-x)$. Therefore, (4) is the maximum likelihood equation of this exponential regression model and $\hat{\beta}$ is asymptotically efficient. The β values used in our simulation are taken to be $\beta_1 = 0.5$ and $\beta_2 = 1$.

Table 1 provides a summary of the simulation results when the baseline distribution is assumed to follow a unit exponential distribution. The mean and standard deviation of $\hat{\beta}$ for each case are computed based on 1,000 simulations. As we can see from Table 1, the second column gives the mean of $\hat{\beta}$, the third column comprises the standard deviation of $\hat{\beta}$, denoted by $\mathrm{SD}_{\hat{\beta}}$, and the fourth and fifth columns contain the mean of estimated standard error, denoted by SE (see Theorem 1 for the detailed formulation) for $C = 0$ and $C = 1$, respectively.

When estimating the standard error, we carefully choose the appropriate kernel and lag to ensure a consistent variance-covariance matrix, given the time series specification. We use the Bartlett kernel and lag $l =$

Table 1. Simulation Results for the Exponential Distribution

(α_1, α_2)	T	Mean of $\hat{\beta}$	$SD_{\hat{\beta}}$	SE $(C = 0)$	SE $(C = 1)$
(0, 0)	100	0.5018	0.1070	0.1028	0.1023
		0.9985	0.1119	0.1017	0.1006
	200	0.5021	0.0692	0.0707	0.0704
		1.0010	0.0719	0.0710	0.0709
	500	0.5006	0.0440	0.0447	0.0445
		0.9995	0.0454	0.0445	0.0445
(0.5, 0.5)	100	0.5049	0.0912	0.0897	0.0883
		1.0029	0.0943	0.0910	0.0890
	200	0.5010	0.0619	0.0609	0.0609
		0.9982	0.0630	0.0612	0.0610
	500	0.4997	0.0388	0.0385	0.0382
		1.0001	0.0391	0.0386	0.0384
(0.9, 0.9)	100	0.5025	0.0582	0.0593	0.0572
		0.9998	0.0594	0.0588	0.0573
	200	0.4993	0.0363	0.0356	0.0349
		1.0006	0.0368	0.0359	0.0352
	500	0.4994	0.0201	0.0204	0.0200
		0.9994	0.0209	0.0206	0.0200

$C\ \mathrm{int}\{4(T/100)^{1/4}\}$ as in Schwert (1987), where C is a constant, either $C = 0$ or 1. As presented in Table 1 for all settings, when the sample size is increased, $\hat{\beta}$ converges to the true value of β and both $SD_{\hat{\beta}}$ and SE decrease. This implies that the proposed estimator is consistent. When $C = 0$ (the i.i.d. case), the resulting estimated standard errors are surprisingly close to the standard deviation of the estimated coefficient in all cases for C, although the independent case ($\alpha_1 = \alpha_2 = 0$) has estimated standard errors slightly closer to $SD_{\hat{\beta}}$ than do the other two cases ($\alpha_1 = \alpha_2 = 0.5$) and ($\alpha_1 = \alpha_2 = 0.9$) where the data are dependent. Therefore, we can conclude that the proposed estimation procedure performs fairly well.

To demonstrate the power of the proposed misspecification test, we consider the null hypothesis $H_0 : \beta_j = \theta_j$, $j = 1, 2$ versus the alternative H_a: at least one $\beta_j \neq \theta_j$. The power function is evaluated under a family of alternative models indexed by γ, $H_a : \beta_j = \theta_j + \gamma\theta_j$, where $\gamma \geqslant 0$. For each sample $T = 100,\ 200$, and 500, the test described in Section 3 is implemented by running the simulation by performing 1,000 replications. The resulting simulated power functions against γ are presented in Table 2 for $\alpha_1 = \alpha_2 = 0$, Table 3 for $\alpha_1 = \alpha_2 = 0.5$, and Table 4 for $\alpha_1 = \alpha_2 = 0.9$, respectively. Note that when $\gamma = 0$, the specified alternative hypothesis collapses into the null hypothesis.

First, let us look at Table 2, where we can see that when $C = 1$, there is a minor size distortion when the sample size equals 100. This distortion may be due to the small sample size and using HAC with dependence. Overall, $C = 0$ has a better performance compared to the other case. Obviously, for $C = 0$ case, performance improves as the sample size increases.

Table 2. Simulated Power Function of the Test When $\alpha_1 = 0$ and $\alpha_2 = 0$

		$C = 0$			$C = 1$		
T	γ	1%	5%	10%	1%	5%	10%
100	0	0.017	0.043	0.127	0.010	0.046	0.094
	0.1	0.112	0.243	0.336	0.136	0.260	0.367
	0.2	0.338	0.515	0.666	0.395	0.580	0.685
	0.3	0.727	0.871	0.926	0.769	0.900	0.940
	0.4	0.946	0.980	0.991	0.946	0.982	0.991
200	0	0.011	0.054	0.106	0.004	0.041	0.088
	0.1	0.137	0.295	0.398	0.153	0.317	0.435
	0.2	0.619	0.783	0.866	0.637	0.802	0.870
	0.3	0.947	0.986	0.994	0.943	0.987	0.996
	0.4	0.999	1.000	1.000	1.000	1.000	1.000
500	0	0.010	0.051	0.103	0.006	0.044	0.104
	0.1	0.351	0.586	0.703	0.361	0.583	0.712
	0.2	0.975	0.994	0.999	0.971	0.995	0.999
	0.3	1.000	1.000	1.000	0.997	1.000	1.000
	0.4	1.000	1.000	1.000	1.000	1.000	1.000

Table 3. Simulated Power Function of the Test When $\alpha_1 = 0.5$ and $\alpha_2 = 0.5$

		$C = 0$			$C = 1$		
T	γ	1%	5%	10%	1%	5%	10%
100	0	0.019	0.071	0.125	0.007	0.051	0.105
	0.1	0.109	0.239	0.336	0.152	0.292	0.394
	0.2	0.411	0.616	0.728	0.474	0.664	0.763
	0.3	0.838	0.937	0.988	0.842	0.937	0.967
	0.4	0.984	0.997	0.998	0.988	0.999	0.999
200	0	0.015	0.068	0.124	0.008	0.051	0.109
	0.1	0.179	0.388	0.504	0.231	0.427	0.544
	0.2	0.778	0.912	0.961	0.808	0.942	0.968
	0.3	0.984	0.995	0.998	0.986	0.994	0.998
	0.4	1.000	1.000	1.000	1.000	1.000	1.000
500	0	0.009	0.063	0.115	0.008	0.058	0.106
	0.1	0.488	0.696	0.800	0.492	0.714	0.806
	0.2	0.996	0.999	1.000	0.995	1.000	1.000
	0.3	1.000	1.000	1.000	0.997	1.000	1.000
	0.4	1.000	1.000	1.000	1.000	1.000	1.000

From Table 3 when $\alpha_1 = \alpha_2 = 0.5$ for which the data are somewhat dependent, one can see that there is a severe positive size distortion at all three significance levels for the independent case $C = 0$. However, when $C = 1$, the empirical sizes generated from simulation are close to the significance levels, in particular, when the sample size equals 500. This demonstrates that the estimate of the null distribution is approximately correct. Also, the power functions for $C = 1$ are larger than those for $C = 0$. Furthermore, we can see that the power functions reach one quickly when γ departs slightly from zero. This shows that our test is indeed powerful.

To evaluate the power performance for strongly correlated data, Table 4 presents the simulated power function against γ when $\alpha_1 = \alpha_2 = 0.9$. The results in Table 4 for the

Table 4. Simulated Power Function of the Test When $\alpha_1 = 0.9$ and $\alpha_2 = 0.9$

		$C = 0$			$C = 1$		
T	γ	1%	5%	10%	1%	5%	10%
100	0	0.0269	0.085	0.121	0.011	0.057	0.107
	0.1	0.315	0.505	0.609	0.391	0.570	0.660
	0.2	0.816	0.905	0.943	0.848	0.926	0.953
	0.3	0.974	0.994	0.996	0.977	0.994	0.998
	0.4	1.000	1.000	1.000	1.000	1.000	1.000
200	0	0.013	0.062	0.130	0.009	0.042	0.108
	0.1	0.634	0.821	0.876	0.679	0.823	0.888
	0.2	0.992	1.000	1.000	0.992	0.997	1.000
	0.3	1.000	1.000	1.000	1.000	1.000	1.000
	0.4	1.000	1.000	1.000	1.000	1.000	1.000
500	0	0.014	0.051	0.108	0.006	0.051	0.096
	0.1	0.990	0.999	1.000	0.989	1.000	1.000
	0.2	1.000	1.000	1.000	1.000	1.000	1.000
	0.3	1.000	1.000	1.000	1.000	1.000	1.000
	0.4	1.000	1.000	1.000	1.000	1.000	1.000

$C = 0$ case are similar to those when $\alpha_1 = \alpha_2 = 0.5$; it is obvious that a very serious positive size distortion for $C = 0$ exists. Particular for $\alpha_1 = \alpha_2 = 0.9$, $C = 1$ predominantly outperforms the other case.

3.2 A real example

For an empirical application using real market data, the performance of the PMRL regression model in ES estimation will be compared under both the i.i.d. ($C = 0$) and time series case ($C = 1$) as well as with two alternative models, namely Historical Simulation and GARCH$(1, 1)$ with constant mean and Gaussian error. For the PMRL regression model, to obtain the ES estimates, VaR is calculated using the Conditional Autoregressive Value-at-Risk (CAViaR) model with asymmetric slope specification of Engle and Manganelli (2004),

$$f_t(\gamma) = \gamma_1 + \gamma_2 f_{t-1}(\gamma) + \gamma_3 (y_{t-1})^+ + \gamma_4 (y_{t-1})^-,$$

where $\text{VaR}_t = -f_t(\gamma)$, $(y)^+ = \max(y, 0)$, and $(y)^- = -\min(y, 0)$.

We evaluate performance using the backtesting method applied to a portfolio consisting of a single stock index. The historical data used in this paper are daily and weekly log returns of S&P 500 Index (GSPC) from January 1990 to October 2011. We use a confidence level of 95% and testing window sizes of 500 and 250 observations as well as 15 and 12 lags for the daily and weekly data, respectively. The backtesting criterion is as follows.

The performance measure proposed by Kerkhof and Melenberg (2004) is implemented by backtesting ES estimates using their violation rate. This method relies on approximating the specific quantile level that the ES falls at, i.e., the nominal level δ_α. The nominal levels (using the 95% confidence level) are 1.96% for parametric models and 1.8% for semiparametric and nonparametric models. The violation rate is defined as the proportion of observations for which the actual loss exceeds the estimated ES, or

Table 5. Testing for Autocorrelations

Daily S&P 500			Weekly S&P 500		
Lags	p-value	χ^2	Lags	p-value	χ^2
5	$1.701e^{-07}$	39.7197	4	0.09477	7.9141
10	$4.496e^{-08}$	54.1777	8	0.05672	15.1273
15	$4.478e^{-11}$	81.0545	12	0.01918	24.1881

$$\text{ESRate} = \frac{\sum_{i=1}^{n} I_{\{L_i > \widehat{ES}_i^\alpha\}}}{n}.$$

The series of ES estimates should have an ESRate close to the nominal level, δ_α. Thus, ESRatio = ESRate/δ_α, is used to compare the models. Models with ESRatio ≈ 1 are most desirable. When the ratio is less than one, it means that the ES model is conservative (has overestimated the risk) while the risk is underestimated (lower than the actual) by the model when the ratio is greater than one. Formal testing with the 95% confidence level is achieved using

- Kupiec's proportion of failures (POF) test (Kupiec, 1995), to test whether the observed frequency of violations is consistent with the frequency of expected violations estimated by the ES model and chosen confidence level.
- Christoffersen's interval forecast test (Christoffersen, 1998), to test whether the violations estimated by the ES model between two successive days are independent.
- Mixed Kupiec test (Haas, 2001), to measure the time between violations and hence is useful to capture the more general form of dependence between violations.

Before estimating ES, we briefly perform a preliminary analysis on the correlations of the real market data. For both the daily and weekly return of S&P 500 Index, the Ljung Box Q-test is applied to examine the correlations in returns data. Using the 95% confidence level, Table 5 indicates that the daily return of S&P 500 Index exhibits serial correlations for the first 15 lags, while the weekly returns are serially uncorrelated for the first 10 lags and then subsequent lags are correlated with one another

After testing for correlations in the historical data, we carry out a coefficient estimation based on the PMRL regression model as the first step in forecasting ES. The coefficients for both the i.i.d and time series specifications of the PMRL regression model are estimated by using the numerical optimization method, the Newton-Raphson method with the following conditions: (1) Initial number of lags is 10; (2) Error bound of 10^{-8} to ensure accuracy; (3) Maximum number of iterations: 500; and (4) The initial value for all lags is zero.

Table 6. t-Statistic Values of All Lags Using Two-sided t-test

Daily S&P 500				Weekly S&P 500							
$\alpha = 95\%$				$\alpha = 95\%$				$\alpha = 90\%$			
$C = 0$		$C = 1$		$C = 0$		$C = 1$		$C = 0$		$C = 1$	
Lag	$\|t\|$	Lag	$\|t\|$	Lag	$\|t\|$	Lag	$\|t\|$	Lag	$\|t\|$	Lag	$\|t\|$
1	**2.393**	**1**	**3.136**	1	1.491	1	1.755	1	1.491	**1**	**1.755**
2	**1.989**	**2**	**2.337**	2	1.102	2	1.381	2	1.102	2	1.381
3	0.508	3	0.525	3	0.533	3	0.534	3	0.533	3	0.534
4	0.628	4	0.695	4	0.987	4	1.051	4	0.987	4	1.051
5	1.158	5	1.300	5	1.171	5	1.111	5	1.171	5	1.111
6	0.266	6	0.271	6	1.450	6	1.529	6	1.450	6	1.529
7	1.676	7	1.914	7	1.112	7	1.038	7	1.112	7	1.038
8	0.581	8	0.526	8	1.337	8	1.461	8	1.337	8	1.461
9	0.161	9	0.164	9	0.258	9	0.247	9	0.258	9	0.247
10	0.994	10	1.141	10	1.590	10	1.471	10	1.590	10	1.471
11	0.820	11	0.682	**11**	**2.200**	**11**	**2.151**	**11**	**2.200**	**11**	**2.151**
12	1.798	**12**	**1.983**	12	0.743	12	0.769	12	0.743	12	0.769
13	0.916	13	1.178								
14	0.457	14	0.502								
15	1.088	15	0.995								

Table 7. Final Estimation Results for Coefficients β's (two-sided test, critical value = 1.96)

Daily S&P 500				Weekly S&P 500			
$C = 0$		$C = 1$		$C = 0$		$C = 1$	
Lag	Coeff.	Lag	Coeff.	Lag	Coeff.	Lag	Coeff.
1	0.5551	1	0.5478	11	−1.0993	11	−1.0993
2	0.5629	2	0.5661				
		12	−0.4501				

The resulting t-statistics of all coefficient estimates for both the daily and weekly return of S&P 500 Index are given in Table 6, based on the testing procedure in Section 2.3.

For the daily return, using a time series specification results in more significant coefficients than using the i.i.d. specification. For the weekly return, both specifications produce the same number of significant coefficients at the 95% confidence level. However, Table 6 shows that the $C = 1$ specification results in larger t-statistic value for the first lag, which is significant if the 90% confidence level is used instead. Therefore, it is more appropriate to use the time series specification as it is generally known that a financial time series is serially correlated with its first lag. For backtesting purposes, we use the final significant coefficient estimate found in Table 7 (tested for significance using the 95% confidence level).

Aside from estimating the coefficients of the PMRL regression model, to get the ES estimates we have to conduct a parameter estimation for the CAViaR model to calculate VaR. The estimation uses the slightly modified code from Engle and Manganelli (2004). The resulting CAViaR equations for both return series are

$$\begin{aligned}\mathrm{VaR}_t(\gamma) &= 0.0003 + 0.9406\mathrm{VaR}_{t-1}(\gamma) \\ &\quad - 0.0008\max(y_{t-1}, 0) + 0.1878(-\min(y_{t-1}, 0))\end{aligned}$$

for the daily return and for the weekly return,

$$\begin{aligned}\mathrm{VaR}_t(\gamma) &= 0.0016 + 0.9047\mathrm{VaR}_{t-1}(\gamma) \\ &\quad - 0.0733\max(y_{t-1}, 0) + 0.2814(-\min(y_{t-1}, 0)).\end{aligned}$$

The backtesting results are given in Table 8. As stated in Table 8, for the daily return of S&P 500 Index, our PMRL regression model with a time series specification slightly outperforms its i.i.d. counterpart based on the ESRatio. Furthermore, both PMRL models typically produce an ESRatio closer to 1, compared to other alternative models, and simultaneously pass all three formal tests. Meanwhile, both the HS and GARCH(1,1) models are deemed less adequate models for ES estimation, given the data, with the GARCH(1,1) model being the most inadequate model as indicated by a very small ESRatio and being rejected by all three tests.

The results for daily return and the weekly return of S&P 500 Index are similar as the PMRL regression model with both i.i.d. and time series specifications ($C = 0$ and $C = 1$) produce ESRatios closest to 1 and pass all three formal tests. They outperform the two alternative models. The i.i.d. specification produces an identical ESRatio to the time series case due to using the same significant coefficient. The HS and GARCH(1,1) models are again seen as inadequate models, however, in this case the HS model is deemed as the worst model with an ESRatio farthest from 1 and also being rejected by all tests.

For the empirical example using the daily and weekly return of S&P 500 Index, backtesting results have shown that both the i.i.d. and time series specifications of the PMRL

Table 8. Backtesting Results

	Daily S&P 500				Weekly S&P 500			
	$C = 0$	$C = 1$	HS	GARCH(1,1)	$C = 0$	$C = 1$	HS	GARCH(1,1)
ESRatio	1.2222	**1.1111**	1.4444	0.1020	**1.5556**	**1.5556**	2.8889	0.4082
Kupiec POF	**Accept**	**Accept**	**Accept**	Reject	**Accept**	**Accept**	Reject	**Accept**
Christoffersen	**Accept**	**Accept**	**Accept**	Reject	**Accept**	**Accept**	Reject	Reject
Mixed Kupiec	**Accept**	**Accept**	Reject	Reject	**Accept**	**Accept**	Reject	Reject

regression model generate ES estimates which are closer to the actual losses within the testing window compared to the alternative models (the HS and GARCH(1,1) with constant mean and normal error distribution), while the time series case of the model performs slightly better than the i.i.d case. The superiority of time series specification over the i.i.d one is more prominent for daily return series. As generally known, daily return series is more serially correlated compared to weekly return, and hence for data with higher serial correlations, the proposed estimator performs better with respect to the applied data and backtesting criterion.

4. CONCLUSION

We introduce the proportional mean residual life (PMRL) regression model, commonly used in reliability and survival analysis, as a new semiparametric model in estimating ES. To account for the dependence characteristic of financial data, the asymptotic normality of the proposed estimator, based on the PMRL regression model, is derived under an α-mixing assumption. To evaluate the performance of our new model, we conduct a simulation study and an empirical application using real market data. A simulation study is conducted under several cases to evaluate the regression parameter estimates under a non-i.i.d. assumption. The result shows that when the sample size is small our estimation and testing perform reasonably well for both empirical size and power.

For the empirical application, we perform a backtesting procedure using the daily and weekly log-return series of S&P 500 Index and compare the performance of our model using a time series specification to its i.i.d. counterpart and to two other models based on the criterion proposed by Kerkhof and Melenberg (2004). The backtesting results for both series generally indicate that the PMRL model with time series specification performs better than the other models. The superiority of using the non-i.i.d. assumption over the i.i.d. case is more apparent for data with higher serial correlations, shown by the daily return series results as the daily return exhibits stronger serial correlation than does the weekly return series.

APPENDIX

Lemma 1: If Assumptions A1 and A3 are satisfied, then $\hat{\beta}$ is a consistent estimator of β_0.

Proof: It follows by the extreme value lemma of White (1994). This proof is complete. ■

Lemma 2: If Assumptions A1–A4 are satisfied, and $E[x_t^2(1 + z_t z_t^T)\exp(2\beta^T z_t)] < \infty$, then, $U^{'}(\beta_0) = \mathrm{Var}(z_t) + o_p(1)$.

Proof: Re-express $U^{'}(\beta_0)$ as $U^{'}(\beta_0) = [D_{zz}D_x - D_{xz}D_{xz}^T]/D_x^2$, where $D_x = \frac{1}{T}\sum_{t=1}^T x_t \exp(\beta_0^T z_t)$, $D_{zz} = \frac{1}{T}\sum_{t=1}^T x_t \exp(\beta_0^T z_t) z_t z_t^T$ and $D_{xz} = \frac{1}{T}\sum_{t=1}^T x_t z_t \exp(\beta_0^T z_t)$. It suffices to show that $D_x = \mu_0 + o_p(1)$, $D_{xz} = \mu_0\mu_z + o_p(1)$ and $D_{zz} = \mu_0 E[z_t z_t^T] + o_p(1)$. Then, the lemma is proved. When $\{(x_t, z_t)\}$ is stationary α-mixing, so are $\{x_t \exp(\beta_0^T z_t)\}$, $\{x_t z_t \exp(\beta_0^T z_t)\}$, $\{x_t z_t z_t^T \exp(\beta_0^T z_t)\}$ with the same mixing coefficient. Clearly, $E[D_x] = \mu_0$ and

$$
\begin{aligned}
&T\,\mathrm{Var}(D_x) = \mathrm{Var}(x_t \exp(\beta_0^T z_t))\\
&+ \frac{1}{T}\sum_{t\neq s} \mathrm{Cov}(x_t \exp(\beta_0^T z_t), x_s \exp(\beta_0^T z_s))\\
&= \mathrm{Var}(x_t \exp(\beta_0^T z_t))\\
&+ \frac{2}{T}\sum_{l\geqslant 1}^{T-1}(T-l)\mathrm{Cov}(x_1 \exp(\beta_0^T z_1), x_{1+l}\exp(\beta_0^T z_{1+l})))\\
&\leqslant C_1 + C_2 \sum_{l\geqslant 1}^{T-1}(1-l/T)\alpha(l)^{\delta/(2+\delta)}\\
&\leq C_1 + C_2 \sum_{l\geqslant 1}^{T-1}(1-l/T)l^{-(\delta+r)/(\delta+2)} = O(1)
\end{aligned}
$$

by using Davydov's inequality (Hall and Heyde 1980, Corrolly A.2), and Assumption A2. Therefore, $\mathrm{Var}(D_x) \to 0$, which implies that $D_x = \mu_0 + o_p(1)$ holds true. By the same token, for D_{xz}, $E(x_t z_t \exp(\beta_0^T z_t)) = E\{E[x_t \exp(\beta_0^T z_t)|z_t] z_t)\} = \mu_0 E(z_t) = \mu_0\mu_z$ and $\mathrm{Var}(D_{xz}) \to 0$. Similarly, $E(x_t \exp(\beta_0^T z_t) z_t z_t^T) = E\{E[\exp(\beta_0^T z_t) x_t|z_t] z_t z_t^T\} = \mu_0 E(z_t z_t^T)$ and $\mathrm{Var}(D_{zz}) \to 0$. Therefore, $U^{'}(\beta_0) = \mathrm{Var}(z_t) + o_p(1)$. This proof is complete. ■

Lemma 3: If Assumptions A1–A4 are satisfied, and $E[x_t^2(1 + z_t z_t^T)\exp(2\beta^T z_t)] < \infty$, we have

$$\sqrt{T}U(\beta_0) = \sqrt{T}\bar{\xi}/\mu_0 + o_p(1) \longrightarrow^d N(0, V).$$

Proof: It is clear that

$$D_x\, U(\beta_0) = \frac{1}{T}\sum_{t=1}^{T} x_t \exp(\beta_0^T z_t)(z_t - \bar{z}) = \bar{\xi} - [D_x - \mu_0](\bar{z} - \mu_z),$$

which implies that $\sqrt{T}U(\beta_0) = \sqrt{T}\bar{\xi}/D_x - [D_x - \mu_0]\sqrt{T}(z_t - \bar{z})/D_x = \sqrt{T}\bar{\xi}/\mu_0 + o_p(1) \longrightarrow^d N(0, V)$ by the law of large numbers and the central limit theorem for the α-mixing process in Ibragimov and Linnik (1971). This proof is complete. ∎

Proof of Theorem 1: Using the Taylor mean value theorem and letting $\beta^* = \lambda\beta_0 + (1-\lambda)\hat{\beta}$, where $0 \leqslant \lambda \leqslant 1$, we can obtain that $U(\hat{\beta}) = U(\beta_0) + U^{'}(\beta^*)(\hat{\beta} - \beta_0)$. To get the asymptotical properties, it suffices to show that consistency $\hat{\beta} \longrightarrow^p \beta_0$ and asymptotic normality $\sqrt{T}U(\beta_0) \longrightarrow^d N(0, V)$, where $V = \lim_{T\to\infty} TE(U(\beta_0)U(\beta_0)^T)/\mu_0^2$. By Lemma 1, $\hat{\beta}$ is a consistent estimator of β. By $\beta^* = \lambda\beta_0 + (1-\lambda)\hat{\beta} = \beta_0 + (1-\lambda)(\hat{\beta} - \beta_0) \to^p \beta_0$, and the continuity of $U_0^{'}(\beta_0)$, $U_0^{'}(\beta^*) \to^p U_0^{'}(\beta_0)$ holds true. Also, by the triangle inequality and the uniform weak law of large numbers for $U^{'}(\beta_0)$,

$$\begin{aligned} ||U^{'}(\beta^*) - U_0^{'}(\beta_0)|| &\leqslant \sup_{\beta^*\in\mathcal{R}} ||U^{'}(\beta^*) - U_0^{'}(\beta^*)|| \\ &\quad + ||U_0^{'}(\beta^*) - U_0^{'}(\beta_0)|| \;\to^p\; 0. \end{aligned}$$

Because $U_0^{'}(\beta_0)$ is nonsingular, so is $U^{'}(\beta^*)$ for T sufficiently large. Therefore, we obtain

$$\sqrt{T}(\hat{\beta} - \beta_0) \approx -U^{'}(\beta^*)^{-1}\sqrt{T}U(\beta_0).$$

A combination of these results with the Slutsky theorem and Lemmas 2 and 3 gives

$$\sqrt{T}(\hat{\beta} - \beta_0) \approx -U^{'}(\beta^*)^{-1}\sqrt{T}U(\beta_0) \to_d N(0, A^{-1}V(A^{-1})^T).$$

Thus, the proof is complete. ∎

Received 24 January 2014

REFERENCES

[1] Acerbi, C. and Tasche, D. (2002). On the coherence of expected shortfall. *Journal of Banking and Finance*, **26**, 1487–1503.

[2] Artzner, P., Delbaen, F., Eber, J. M., and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, **9**, 203–228. MR1850791

[3] Barone-Adesi, G., Bourgoin, F., and Giannopoulos, K. (1998). Do not look back. *Risk*, **11**, 100–104.

[4] Barone-Adesi, G., Giannopoulos, K., and Vosper, L. (1999). VaR without correlations for nonlinear portfolios. *Journal of Futures Markets*, **19**, 583–602.

[5] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, **31**, 307–327. MR0853051

[6] Box, G. E. P. and Jenkins, G. M. (1970). *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco. MR0272138

[7] Cai, Z. and Wang, X. (2008). Nonparametric estimation of conditional VaR and expected shortfall. *Journal of Econometrics*, **147**, 120–130. MR2472986

[8] Chen, S. X. (2008). Nonparametric estimation of expected shortfall. *Journal of Financial Econometrics*, **6**, 87–107.

[9] Christoffersen, P. (1998). Evaluating interval forecasts. *International Economic Review*, **39**, 841–862. MR1661906

[10] Dasu, T. and Oakes, D. (2003). Inference for the proportional mean residual life model. *Lecture Notes-Monograph Series*, **43**, 105–116. MR2125050

[11] Embrechts, P., Kluppelberg, C., and Mikosch, T. (1999). *Modeling Extremal Events for Finance and Insurance*, 2nd edition. Springer-Verlag, New York. MR1458613

[12] Engle, R. F. and Manganelli, S. (2004). CAViaR: Conditional autoregressive value at risk by regression quantile. *Journal of Business and Economics Statistics*, **22**, 367–381. MR2091566

[13] Haas, M. (2001). New methods in backtesting. *Working Paper*, Financial Engineering Research Center Caesar, Friedensplatz, Bonn, Germany.

[14] Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and Its Applications*. Academic Press, New York. MR0624435

[15] Ibragimov, I. A. and Linnik, Yu. V. (1971). *Independent and Stationary Sequences of Random Variables*. Walters-Noordhoff, Groningen, the Netherlands. MR0322926

[16] Kerkhof, F. L. J. and Melenberg, B. (2004). Backtesting for risk-based regulatory capital. *Journal of Banking and Finance*, **28**, 1845–1865.

[17] Kupiec, P. H. (1995). Techniques for verifying the accuracy of risk measurement models. *The Journal of Derivatives*, **3**, 73–84.

[18] Maguluri, G. and Zhang, C. H. (1994). Estimation in the mean residual life regression model. *Journal of the Royal Statistical Society, Series B*, **56**, 477–489. MR1278221

[19] Newey, W. K. and West, K. D. (1987). A simple, positive-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, **55**, 703–708. MR0890864

[20] Newey, W. K. and West, K. D. (1994). Automatic lag selection in covariance matrix estimation. *Review of Economic Studies*, **61**, 631–653. MR1299308

[21] Oakes, D. and Dasu, T. (1990). A note on residual life. *Biometrika*, **77**, 409–410. MR1064816

[22] Scaillet, O. (2004). Nonparametric estimation and sensitivity analysis of expected shortfall. *Mathematical Finance*, **14**, 115–129. MR2030838

[23] Scaillet, O. (2005). Nonparametric estimation of conditional expected shortfall. *Revue Assurances et Gestion des Risques/Insurance and Risk Management Journal*, **74**, 639–660.

[24] Schwert, G. W. (1987). Effects of model specification on tests for unit roots in macroeconomic data. *Journal of Monetary Economics*, **20**, 73–103.

[25] White, H. (1994). *Estimation, Inference, and Specification Analysis*. Cambridge University Press, New York. MR1292251

Zongwu Cai
Department of Economics
University of Kansas
Lawrence, KS 66045
USA

Wang Yanan Institute for Studies in Economics
MOE Key Laboratory of Econometrics
and Fujian Key Laboratory of Statistical Sciences
Xiamen University
Xiamen, Fujian 361005
China
E-mail address: caiz@ku.edu

Jia Su
Institute for Quantitative Economics
Huaqiao University
Xiamen, Fujian 361021
China
E-mail address: jia43zhu@163.com

Sufianti
Wang Yanan Institute for Studies in Economics
MOE Key Laboratory of Econometrics
and Fujian Key Laboratory of Statistical Sciences
Xiamen University
Xiamen, Fujian 361005
China
E-mail address: sufianti88@gmail.com