2005 Proceedings of the Autumn Conference, Korea Statistical Society

Nonparametric estimation of conditional Value-at-Risk

Seok-Oh Jeong[1] and Kee-Hoon Kang [2]

1 The risk model

Consider a univariate asset price process $(S_t)_{t=0,1,\dots}$, which is typically a stock price or a foreign exchange rate. Define the returns, $(r_t)_{t=1,2,\dots}$, of (S_t) by its log-returns:

$$r_t = \log S_t - \log S_{t-1}.$$

Assume that this return process is of the form

$$r_t = \sigma_t \varepsilon_t \tag{1}$$

where σ_t is the predictable volatility process, and ε_t are independently and identically distributed (IID) random variables having zero mean zero and unit variance. The volatility coefficient σ_t is often considered as the square root of the conditional variance of the returns:

$$\sigma_t^2 = \mathrm{Var}(r_t|\mathcal{F}_{t-1}). \tag{2}$$

Under the conditional heteroscedasticity model in (1), (conditional) Value-at-Risk at level α is defined by the value $\mathrm{VaR}_{\alpha,t}$ which satisfies the following relation:

$$P(r_t < \mathrm{VaR}_{\alpha,t} \,|\, \mathcal{F}_{t-1}) = \alpha. \tag{3}$$

Since σ_t is assumed predictable, we may put

$$\mathrm{VaR}_{\alpha,t} = \sigma_t q_\alpha$$

where q_α is the α-th quantile of the distribution of ε_t. Hence, as well as the accurate volatility estimation, it is essential for obtaining the value of VaR to specify the distribution of the innovations ε_t correctly.

When we think of σ_t as in (2), ε_t's have unit variance. However, we don't have to stick to this normalization. In order to obtain properly standardized versions of σ_t and q_α, one may simply multiply σ_t and divide q_α by $\sqrt{\mathrm{Var}(\varepsilon_t)}$. Therefore there is nothing to be changed in the value of VaR. Now we discuss the estimation of σ_t and q_α.

[1] Department of Statistics, Hankuk University of Foreign Studies, South Korea.
E-mail: seokoh@stat.ucl.ac.be
[2] Department of Statistics, Hankuk University of Foreign Studies, South Korea.
E-mail: khkang@hufs.ac.kr

2 Local homogeneity model for volatility

Fix a time point τ. The local homogeneity of volatility process (σ_t) on a time interval $I = [\tau - m, \tau)$ means that σ_t varies little in t on the interval I. Once the local homogeneity on I is proven, σ_τ^2 can be simply estimated by the average of the squared returns in the time interval I:

$$\widehat{\sigma}_\tau^2 = \sum_{t \in I} r_t^2, \tag{4}$$

This is basically a moving average with varying window size.

How can we determine whether an interval is the homogeneity interval or not? Mercurio and Spokoiny (2004) provided a martingale deviation probability bound to test the homogeneity, but it has a drawback that it is not tractable to control the level of test. We are to present another approach which enables us to control the level of the homogeneity test but does not require such a high-math.

Consider a power transform on r_t to get a regression like model as in Mercurio and Spokoiny (2004): for some $\gamma > 0$,

$$|r_t|^\gamma = \sigma_t^\gamma |\varepsilon_t|^\gamma = \theta_t + \theta_t s_\gamma \zeta_t$$

where $\theta_t = c_\gamma \sigma_t^\gamma$, $\zeta_t = d_\gamma^{-1}(|\varepsilon_t|^\gamma - c_\gamma)$, $c_\gamma = \mathrm{E}|\varepsilon_t|^\gamma$, $d_\gamma = \sqrt{\mathrm{Var}|\varepsilon_t|^\gamma}$ and $s_\gamma = d_\gamma / c_\gamma$. Under the homogeneity of σ_t on an interval I, there is a constant trend θ_I such that $\theta_t \equiv \theta_I$ holds for all $t \in I$, and we naturally estimate this constant trend by the average of $|r_t|^\gamma$'s in the interval I:

$$\widetilde{\theta}_I = \frac{1}{N_I} \sum_{t \in I} |r_t|^\gamma$$

where N_I is the number of time points in the interval I. Then, under the homogeneity hypothesis on the interval I, one may show that, for any subinterval $J \subset I$

$$T_{I,J} = \frac{\widetilde{\theta}_J - \widetilde{\theta}_{I \setminus J}}{S_{I,J}\sqrt{N_J^{-1} + N_{I \setminus J}^{-1}}}$$

is approximately $t(N_I - 2)$-distributed, where $t(n)$ denotes the t-distribution with degrees of freedom n, and

$$S_{I,J}^2 = \frac{1}{N_I - 2} \left\{ \sum_{t \in J} \left(|r_t|^\gamma - \widetilde{\theta}_J \right)^2 + \sum_{t \in I \setminus J} \left(|r_t|^\gamma - \widetilde{\theta}_{I \setminus J} \right)^2 \right\}.$$

Note that, by construction, we have

$$\widetilde{\theta}_I = \theta_I \left(1 + s_\gamma \bar{\zeta}_I \right)$$

where $\bar{\zeta}_I = N_I^{-1} \sum_{t \in I} \zeta_t$. It follows that

$$\sum_{t \in I} \left(|r_t|^\gamma - \widetilde{\theta}_I \right)^2 = \theta_I^2 s_\gamma^2 \sum_{t \in I} (\zeta_t - \bar{\zeta}_I)^2$$

Thus, under the homogeneity on the interval I, the distribution of test statistic $T_{I,J}$ does not depend on θ:

$$T_{I,J} = \frac{\bar{\zeta}_J - \bar{\zeta}_{I \setminus J}}{S_{\zeta,I,J}\sqrt{N_J^{-1} + N_{I \setminus J}^{-1}}}.$$

where

$$S^2_{\zeta,I,J} = \frac{1}{N_I - 2}\left\{\sum_{t\in J}(\zeta_t - \bar{\zeta}_J)^2 + \sum_{t\in I\setminus J}(\zeta_t - \bar{\zeta}_{I\setminus J})^2\right\}.$$

Since ζ_t's are iid random variables having zero mean and unit variance, one may execute t-test to test the homogeneity using the statistic $T_{I,J}$.

Algorithm for the local homogeneity interval Fix $m_0 > 1$ and $\alpha > 0$.

[Initialization] Set $k = 1$.

[Update] Set $I = [\tau - (k+1)m_0, \tau)$.

[Test] Compute $T_{I,J}$ for all $J \in \{[\tau - \ell m_0, \tau) \,|\, \ell = 1, 2, ..., k\}$. If there is a subinterval J such that $|T_{I,J}| > t_{\alpha/(2k)}(N_I - 2)$, then set $\hat{m} = km_0$ and $\hat{I} = [\tau - \hat{m}, \tau)$ and stop.

[Loop] Set $k = k + 1$ and goto [Update].

Note that $t_{\alpha/(2k)}(N_I - 2)$ is used for the critical value instead of $t_{\alpha/2}(N_I - 2)$, which is the Bonferroni correction for multiple tests in [Test] step.

3 Quantile estimation

Now we discuss the nonparametric estimation of quantiles. Consider a random sample $X_1, \cdots, X_n$. Define the quantile function as, for $\alpha \in (0,1)$

$$Q(\alpha) = \inf\{x \,|\, F(x) \geq \alpha\},$$

where F is the (common) distribution function of X_i's. Let us assume that the quantile function $Q(\alpha)$ is smooth in α. The empirical quantile estimator is

$$\begin{aligned} Q_n(\alpha) &= X_{(s)}, \text{ if } (s-1)/n < \alpha \leq s/n,\ s = 1, \ldots, n \\ &= X_{(1)}, \text{ if } \alpha = 0, \end{aligned}$$

where $X_{(1)} \leq \cdots \leq X_{(n)}$ is the order statistics of $X_1, \cdots, , X_n$. The kernel quantile estimator, proposed by Parzen (1979), is given by

$$\hat{Q}_1(\alpha) = \int_0^1 \frac{1}{h} K\left(\frac{s-\alpha}{h}\right) Q_n(s)\, ds,$$

where K is a kernel function, and $h > 0$ is a bandwidth. This estimator has been known to suffer from a severe boundary problem, i.e. to be problematic when estimating extreme quantiles for α less than h. Cheng and Peng (2002) tackled this problem by introducing the local quadratic regression idea to the quantile estimation. They considered the minimizing problem with respect to a, b and c of the weighted integral of squared error of a quadratic approximation to the empirical quantile estimator Q_n

$$\int_0^1 \left\{Q_n(s) - a - b(\alpha - s) - c(\alpha - s)^2\right\}^2 K\left(\frac{\alpha - s}{h}\right).$$

The local quadratic estimator is then given by the value of a in the above solution, which has the precise form

$$\hat{Q}_2(\alpha) = \frac{A_0(\alpha)(a_2a_4 - a_3^2) - A_1(\alpha)(a_1a_4 - a_2a_3) + A_2(\alpha)(a_1a_3 - a_2^2)}{a_0(a_2a_4 - a_3^2) - a_1(a_1a_4 - a_2a_3) + a_2(a_1a_3 - a_2^2)}, \tag{5}$$

where $a_i = \int_0^1 (\alpha - s)^i K((\alpha - s)/h)ds$, and $A_i(\alpha) = \int_0^1 (\alpha - s)^i K((\alpha - s)/h)Q_n(s)ds$. The local quandratic estimator $\hat{Q}_2$ is free of the boundary problem, and it gives a robust estimate over the choice of the bandwidth $h > 0$.

4 Estimation of value-at-risk

Based on two pillars of the adaptive volatility estimation and the local quadratic quantile estimation, now we can build the nonparametric estimator of VaR as follows:

$$\widehat{\mathrm{VaR}}_{\alpha,t} = \hat{\sigma}_t \hat{q}_\alpha. \tag{6}$$

where $\hat{\sigma}_t$ is the volatility estimate in (4) and $\hat{q}_\alpha$ is the local quadratic quantile estimate in (5) from the devolatilized returns $\hat{\varepsilon}_t = r_t/\hat{\sigma}_t$'s.

5 Empirical study: KOSPI

5.1 Data

Figure 1 depicts the series of daily KOSPI from January 03, 1990 to September 30, 2002 containing the economic crisis of South Korea in 1997. Its daily returns R_t are given below the series, which clearly shows the impact of the crisis in 1997. In Figure 2 we note the first two significant autoregressive coefficients of KOSPI return process. Due to this market microstructure, we assume AR(2)-model for R_t as follows:

$$R_t = \phi_1 R_{t-1} + \phi_2 R_{t-2} + r_t,$$
$$r_t = \sigma_t \varepsilon_t.$$

The estimates of AR-coefficients are ϕ_1 and ϕ_2 are given in Table 1.

$\hat{\phi}_1$	$\hat{\phi}_2$
0.1031	-0.0549

Table 1: Autoregressive coefficients of KOSPI returns R_t.

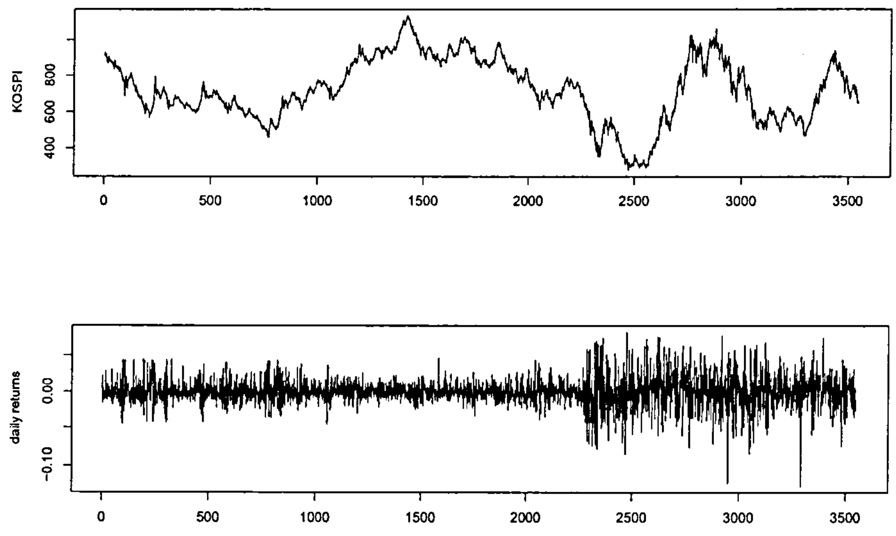

Figure 1: The time series plots for KOSPI and its returns. Jan-03-1990 ~ Sep-30-2002.

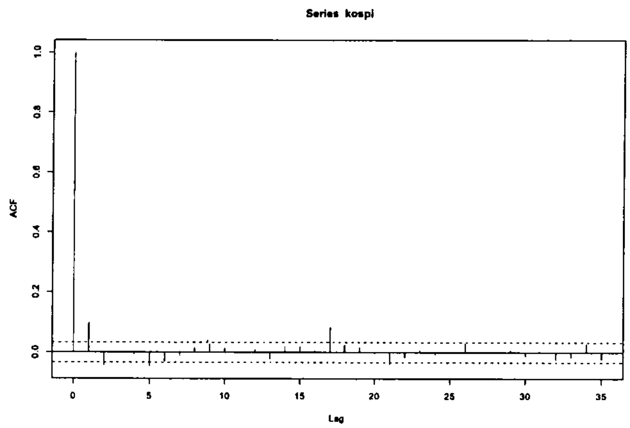

Figure 2: ACF plot for KOSPI returns R_t.

5.2 Volatility estimation

In Figure 3 we show the time series plot of the residuals r_t from the fitted AR(2)-model for R_t, together with the estimated volatility coefficients $\hat{\sigma}_t$ of r_t which were obtained according to the proposed procedure. Hereafter we call r_t the returns again. As expected, the proposed procedure captured the abrupt changes in the volatility, i.e. the magnitude of the returns r_t very well.

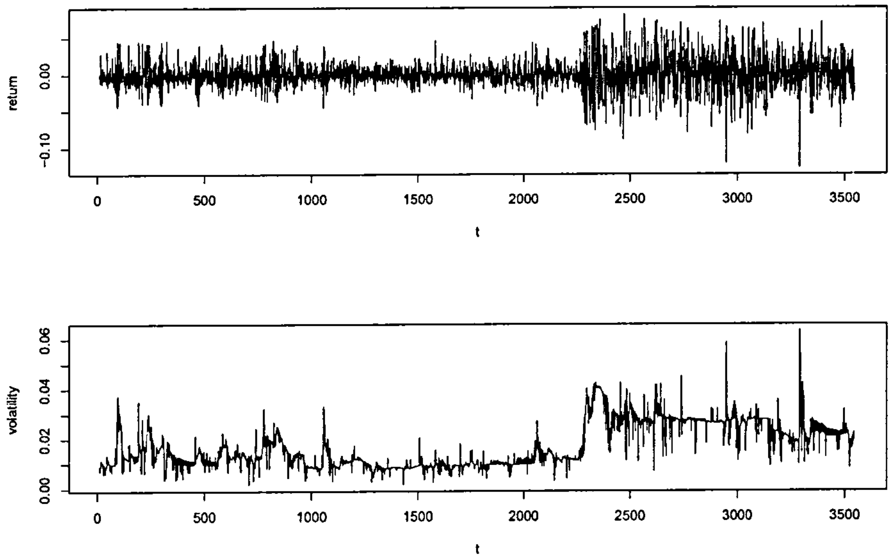

Figure 3: KOSPI returns r_t and the estimated volatility.

5.3 Devolatilized returns

Besides this feature on the volatility estimates, the devolatilized returns $\hat{\varepsilon}_t$ computed by $r_t/\hat{\sigma}_t$ should be iid if the proposed approach is reliable. In Figure 4 we plot the autoregressive coefficients for the returns r_t and their squared value as well as for the devolatilized returns and the squared devolatilized returns. There we see that the dependency structure in the squared returns are successfully filtered by the proposed procedure. Furthermore, Figure 5 shows that the inhomogeneity in the return process r_t has been removed successfully.

5.4 Value-at-Risk

Now the Value-at-Risk at time t of KOSPI data is estimated as follows:

$$\widehat{\mathrm{VaR}}_{\alpha,t} = \hat{\phi}_1 R_{t-1} + \hat{\phi}_2 R_{t-2} + \hat{\sigma}_t \hat{q}_\alpha$$

where $\hat{q}_\alpha$ is the estimated α-th quantile of the distribution of ε_t.

5.5 Backtesting VaR

By definition, a good VaR model should ensure that the proportion of the VaR violation is close to its VaR level α. Let T be the total number of time points where VaR has been computed, and let N be the number of VaR violations. Since N is a binomial random variable under

$$H_0 : E[N] = T\alpha,$$

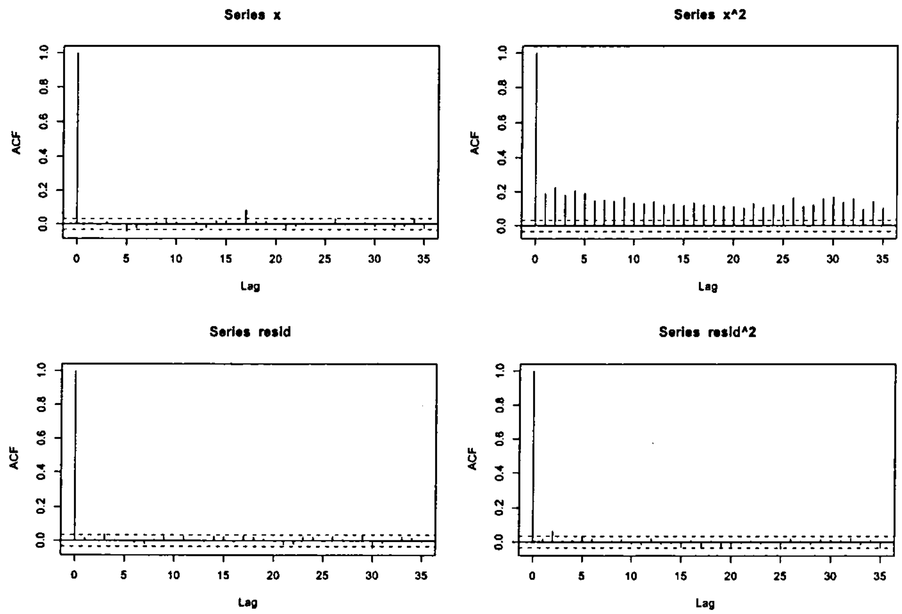

Figure 4: ACF plots for KOSPI returns r_t and the squared returns r_t^2 (upper two panels) and the devolatilized returns $\hat{\varepsilon}_t = r_t/\hat{\sigma}_t$ and the squared devolatilized returns $\hat{\varepsilon}_t^2$ (lower panels).

one may consider the likelihood ratio test statistic given by

$$LR = -2\log\left\{(1-\alpha)^{T-N}\alpha^N\right\} + 2\log\left\{(1-N/T)^{T-N}(N/T)^N\right\}$$

which is asymptotically $\chi^2(1)$-distributed under H_0, see Christoffersen (1998).

α	N	T	N/T	LR	p-value
0.05	152	3039	0.05002	1.73170e-05	0.9967
0.01	29	3039	0.00954	0.06521	0.7984

Table 2: Backtesting results for KOSPI data.

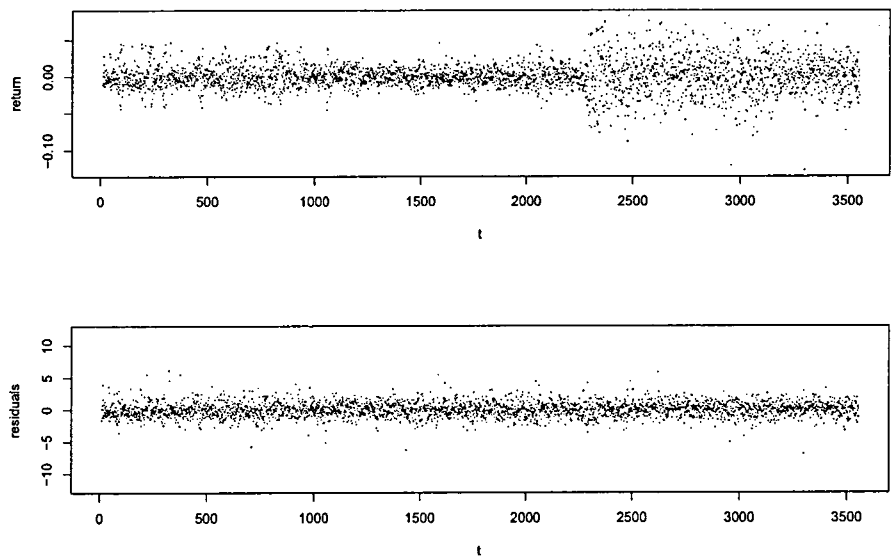

Figure 5: KOSPI returns r_t and the devolatilized returns $r_t/\hat{\sigma}_t$.

References

Basel Committee on Banking Supervision (1995). *An internal model-based approach to market risk capital requirements*, Basel: Bank for International Settlements.

Cheng, M.-Y. and Peng, L. (2002). Regression modeling for nonparametric estimation of distribution and quantile functions, *Statistica Sinica* **12**, 1043–1060.

Christoffersen, P. F. (1998). Evaluating Interval Forecast, *International Economic Review* **39**, 841–862.

Eberlein, E., Kallsen, J. and Kristen, J. (2003). Risk Management Based on Stochastic Volatility, *Journal of Risk* **5**, 19–44.

Chen, Y., Härdle, W. and Jeong, S.-O. (2005). Nonparametric risk management with generalized hyperbolic distribution, Sonderforschungsbereich 649, DP2005-001.

McNeil, A. J. and Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series : an extreme value approach, *Journal of Empirical Finance* **7**, 271–300.

Mercurio, D. and Spokoiny, V. (2004). Statistical Inference for Time Inhomogeneous Volatility Models, *The Annals of Statistics* **32**, 577–602.