Technische Universiteit Eindhoven University of Technology

Conditional Correlation Modeling using Machine Learning and Multivariate GARCH

Paul Melkert

Operations, Planning, Accounting and Control Group,
Department of Industrial Engineering,
Eindhoven University of Technology,
The Netherlands

A new semiparametric multivariate volatility model is proposed, which integrates parametric univariate GARCH specifications for the conditional volatilities with nonparametric machine learning estimators for the conditional correlations. Consider a k-dimensional financial log return series $r_t = (r_{1,t}, \ldots, r_{k,t})'$:

$$r_t = H_t^{1/2} z_t,$$

where conditional variance-covariances of log returns r_t is denoted by a $k \times k$ positive definite matrix H_t, $H_t^{1/2}$ its Cholesky decomposition, and z_t denotes a $k \times 1$ vector of i.i.d. standard normal random variables. We consider a reparameterization of H_t, i.e. $H_t = D_t R_t D_t$, where D_t is a diagonal matrix with conditional volatility of each marginal log return series on its diagonal and matrix R_t includes all pairwise correlations $\rho_{i,j,t}$:

$$\begin{pmatrix} h_{1,1,t}^{1/2} & & 0 \\ & \ddots & \\ 0 & & h_{k,k,t}^{1/2} \end{pmatrix} \begin{pmatrix} 1 & \rho_{1,2,t} & \cdots & \rho_{1,k,t} \\ \rho_{2,1,t} & 1 & & \rho_{2,k,t} \\ \vdots & & \ddots & \vdots \\ \rho_{k,1,t} & \rho_{k,2,t} & \cdots & 1 \end{pmatrix} \begin{pmatrix} h_{1,1,t}^{1/2} & & 0 \\ & \ddots & \\ 0 & & h_{k,k,t}^{1/2} \end{pmatrix},$$

Diagonal elements of matrix D_t are modeled parametrically using GARCH(1,1) and GJR-GARCH(1,1) specifications and conditional correlation matrix R_t is estimated in a nonparametric manner using machine learning algorithms k-nearest neighbor (KNN) and random forest (RF) of Breiman [1]. A parsimonious learner model specification, with the covariate space constructed using Pearson and Kendall sample correlations from moving windows, ensures that the model remains flexible and easy to estimate in high dimensional systems. Proposed semiparametric multivariate volatility model is applied to the 30 constituents included in the Dow Jones Industrial Average index and model adequacy is verified in terms of a common downside market risk measure Value-at-Risk (VaR):

$$VaR_\alpha = inf\{l \in \mathbb{R} : P(L > l) \leq 1 - \alpha\},$$

where $\alpha \in (0, 1)$ and $F_L(l) = P(L \leq l)$, denotes the distribution function of corresponding loss distribution. A failure test of unconditional coverage (Kupiec) and an independence test (Christoffersen) are considered to statistically evaluate proposed model against a parametric conditionally heteroscedastic class of models, the dynamic conditional correlation model of Engke [2], and to detect any misspecification of the risk model. It is demonstrated that proposed model is able to capture interesting conditional properties in correlations and that it shows competitive performance with the DCC model, both in tranquil market conditions (mid-1990's) and volatile market conditions (2000-2001 Dot-com bubble); the likelihood of a VaR exceedance does not depend greatly on whether a VaR exceedance occurred on the previous day. Moreover, the null hypothesis that the total number of VaR exceedances equals the expected number of VaR exceedances, given independence, is satisfied for a variety of model specifications. This can also be seen in Figure 1, which shows the backtest results for unconditional coverage (Kupiec test).

Model	VaR Exceedances 1994-1995			VaR Exceedances 2000-2001		
	I(0.95)	I(0.975)	I(0.99)	I(0.95)	I(0.975)	I(0.99)
DCC-Garch	**11**	7	2	33	16	5
DCC-GJR	**11**	7	2	31	14	5
KNN(5)-Pearson-Garch	17	11	7	**39**	**22**	**12**
KNN(5)-Pearson-GJR	19	13	7	**38**	**21**	9
KNN(5)-Kendall-Garch	25	15	10	**58**	**32**	**17**
KNN(5)-Kendall-GJR	27	17	**11**	**55**	**32**	**17**
KNN(100)-Pearson-Garch	**13**	**6**	**1**	29	18	6
KNN(100)-Pearson-GJR	**13**	7	**1**	29	17	5
KNN(100)-Kendall-Garch	17	12	7	**49**	**25**	**13**
KNN(100)-Kendall-GJR	18	12	7	**46**	**24**	**12**
KNN(idw)-Pearson-Garch	**11**	**6**	2	30	15	7
KNN(idw)-Pearson-GJR	**11**	7	2	28	15	6
KNN(idw)-Kendall-Garch	**12**	10	6	**43**	**21**	**10**
KNN(idw)-Kendall-GJR	**13**	11	6	**39**	18	9
RF(10)-Pearson-Garch	17	10	3	**39**	18	9
RF(10)-Pearson-GJR	17	11	4	**38**	18	8
RF(10)-Kendall-Garch	18	12	8	**54**	**28**	**15**
RF(10)-Kendall-GJR	21	12	10	**53**	**29**	**16**
RF(100)-Pearson-Garch	**11**	9	3	**37**	**20**	6
RF(100)-Pearson-GJR	**13**	10	4	**36**	**20**	6
RF(100)-Kendall-Garch	17	12	8	**55**	**30**	**15**
RF(100)-Kendall-GJR	20	12	10	**55**	**29**	**15**

Figure 1: Test results of unconditional coverage (Kupiec test) at higher quantiles of the loss distribution in tranquil and volatile market conditions. Bold face indicates rejection.

References

[1] Leo Breiman. Random forests. Machine Learning, 45(1):5-32, 2001.

[2] Robert Engle. Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. Journal of Business & Economic Statistics, 20 (3):339-350, 2002.