Volatility Prediction Using Kernel Regression: Supplementary Material

Jussi Klemelä

August 23, 2017

Abstract

We give supplementary material for the article "Volatility prediction using kernel regression".

1 The Sum of Squared Prediction Errors

We assume to have returns $R_1, \ldots, R_T$. The sequential (or recursive) out-of-sample sum of squares of prediction errors for the whole sample is defined as

$$\mathrm{SSPE}(\hat{f}) = \sum_{t=t_0}^{T-\eta} \left(R_{t+\eta}^2 - \hat{f}(t) \right)^2, \tag{1}$$

where $\eta < t_0 \le T - \eta$, $\eta \ge 1$ is the prediction horizon, and $\hat{f}(t)$ is estimated using the data $R_1, \ldots, R_t$.

A computationally less expensive sum of squared prediction errors can be defined by dividing the sample into an estimation set and into a test set. The predictor is constructed using the estimation set and the sum of squared prediction errors is computed using the test set:

$$\mathrm{SSPE}_{test}(\hat{f}) = \sum_{t=t_0}^{T-\eta} \left(R_{t+\eta}^2 - \hat{f}(t_0) \right)^2,$$

where $\hat{f}(t_0)$ is computed using the estimation data R_t, $t = 1, \ldots, t_0$. The test data is $R_{t+\eta}$, $t = t_0, \ldots, T - \eta$.

A third version of the out-of-sample sum of squared prediction errors is obtained when the predictor is updated at every time point, but the predictor uses always the same number of past observations. The predictor uses windows of observations which are rolled over the available data. Let predictor $\hat{f}(s,t)$ be constructed using the data R_i, $i = s, \ldots, t$. Define

$$\mathrm{SSPE}_{roll}(\hat{f}) = \sum_{t=t_0}^{T-\eta} \left(R_{t+\eta}^2 - \hat{f}(t - t_0 + 1, t) \right)^2 .$$

Now the sum of squared prediction errors is computed for the estimator which is constructed using exactly t_0 observations at every time point.

2 Leverage Effect and GARCH Models

Andersen et al. (2006) identify the three most common GARCH formulations for describing the leverage effect being (1) asymmetric GARCH models, (2) threshold GARCH models, and (3) exponential GARCH models.

Heston and Nandi (2000) define model

$$\begin{aligned} \sigma_t^2 &= \alpha_0 + \alpha_1 (\epsilon_{t-1} - \lambda \sigma_{t-1})^2 + \beta \sigma_{t-1}^2 \qquad (2) \\ &= \alpha_0 + \alpha_1 \frac{(R_{t-1} - \lambda \sigma_{t-1}^2)^2}{\sigma_{t-1}^2} + \beta \sigma_{t-1}^2, \end{aligned}$$

where $\lambda \in \mathbf{R}$ is the skewness parameter. The model is an example of an asymmetric GARCH model. Engle and Ng (1993) define the nonlinear asymmetric GARCH model

$$\sigma_t^2 = \alpha_0 + \alpha_1 \sigma_{t-1}^2 (\epsilon_{t-1} - \lambda)^2 + \beta \sigma_{t-1}^2, \qquad (3)$$

which is for $\lambda = 0$ equal to the GARCH$(1,1)$ model. Engle and Ng (1993) have defined the VGARCH model

$$\sigma_t^2 = \alpha_0 + \alpha_1 (\epsilon_{t-1} - \lambda)^2 + \beta \sigma_{t-1}^2. \qquad (4)$$

A threshold GARCH model was defined in Glosten et al. (1993) and Zakoïan (1994). In this model

$$\sigma_t^2 = \alpha_0 + \alpha_1 Y_{t-1}^2 + \lambda Y_{t-1}^2 I_{(Y_{t-1}<0)} + \beta \sigma_{t-1}^2.$$

The exponential GARCH model was defined in Nelson (1991). In this model

$$\log \sigma_t^2 = \alpha + \alpha_1 (|\epsilon_{t-1}| - E(|\epsilon_{t-1}|)) + \lambda \epsilon_{t-1} + \beta \log \sigma_{t-1}^2.$$

Menn and Rachev (2009) propose the GARMAX model which also can cope with the leverage effect. Chorro et al. (2012) account for the leverage effect by considering GARCH models where the innovations follow a generalized hyperbolic distribution, instead of the standard normal distribution.

3 Testing Statistical Significance

Giacomini and White (2006) proposes a generalization of the test proposed by Diebold and Mariano (1995) and West (1996). Giacomini and White (2006) test the hypothesis

$$H_0 : E_i(d_{i+\eta}) = 0, \qquad H_1 : E_i(d_{i+\eta}) \neq 0.$$

where E_i means the conditional expectation with respect to the information available at time i. Now $E_i(d_{i+\eta}) = 0$ is equivalent to the fact that $E(hd_{i+\eta}) = 0$ for all h which are measurable with respect to $\mathcal{F}_i$, where $\mathcal{F}_i$ is the information set at time i. Let h_i be a $\mathcal{F}_i$-measurable random vector. Let

$$Z_i = h_{i-\eta} d_i.$$

Now (Z_t) is a vector time series, with $Z_t \in \mathbf{R}^d$. If the time series $(a'Z_t)_{t\in\mathbf{Z}}$ satisfies the conditions for the univariate central limit theorem for all $a \in \mathbf{R}^d$, then the multivariate central limit theorem holds. A central limit theorem states that

$$S_t = t^{-1/2} \sum_{i=1}^{t} (Z_i - EZ_i) \xrightarrow{d} N(0, \Sigma), \tag{5}$$

where

$$\Sigma = \sum_{j=-\infty}^{\infty} \Gamma(j) = \Gamma(0) + \sum_{j=1}^{\infty} (\Gamma(j) + \Gamma(j)'),$$

and the autocovariance matrix $\Gamma(j)$ is defined as $\Gamma(j) = \mathrm{Cov}(Z_i, Z_{i+j})$. Note that we used the property $\Gamma(j) = \Gamma(-j)'$. To estimate Σ in (5) we use

$$\hat{\Sigma} = \hat{\Gamma}(0) + \sum_{j=1}^{t-1} w(j) \left(\hat{\Gamma}(j) + \hat{\Gamma}(j)' \right),$$

where $\hat{\Gamma}(j) = \frac{1}{t} \sum_{i=1}^{t-j} \left(Z_i - \bar{Z} \right) \left(Z_{i+j} - \bar{Z} \right)'$, for $j = 0, \ldots, t-1$. Chen and Ghysels (2012) apply the test in volatility prediction by taking

$$h_i = \left(1, (R_i^2 - \hat{f}^b(i-\eta))^2 \right)',$$

where $\hat{f}^b$ is the predictor used as a benchmark. In our case $\hat{f}^b = \hat{f}^{garch}$. The test statistics is

$$\mathrm{TS} = S_t' \hat{\Sigma}^{-1} S_t.$$

Under H_0, TS $\xrightarrow{d} \chi^2(1)$, as $t \to \infty$. Large values of the test statistics lead to the rejection of the null hypothesis. Let the observed value of TS be equal to y. The p-value is equal to $P(\mathrm{TS} > y) \approx 1 - F(y)$, where F is the distribution function of the $\chi^2(1)$ distribution.

4 Performance Measurement in Quantile Estimation

The number of exceedances can be used to measure the performance of a quantile estimator. Let

$$\hat{p} = \frac{1}{T - t_0} \sum_{t=t_0}^{T-1} I_{(R_{t+1} \le \hat{q}_{t+1})},$$

where $1 < t_0 \le T - 1$. Now $\hat{p}$ should be close to p. We can measure the quality of the quantile estimator with the difference $p - \hat{p}$, or we can express the quality of the quantile estimator by using p-values of a test statistics for

$$H_0 : E(\hat{p}) = p, \qquad H_1 : E(\hat{p}) \ne p.$$

Assuming that random variable $N = \sum_{t=t_0}^{T-1} I_{(R_{t+1} \le \hat{q}_{t+1})}$ has a binomial distribution under H_0 with parameters $n = T - t_0$ and p, we can use the likelihood ratio test statistics

$$\mathrm{LR} = -2 \log[(1-p)^{n-N} p^N] + 2 \log[(1 - N/n)^{n-N} (N/n)^N],$$

which is asymptotically $\chi^2(1)$ distributed; see Jorion (2006). The use of the number of exceedances is not as easy as the use of the loss function when we want to measure the performance simultaneously over all time intervals.

References

Andersen, T. G., Bollerslev, T., Christoffersen, P. F. and Diebold, F. X. (2006), Volatility and correlation forecasting, *in* G. Elliott, C. W. J. Granger and A. Timmerman, eds, 'Handbook of Economic Forecasting', North-Holland, Amsterdam, pp. 777–878.

Chen, X. and Ghysels, E. (2012), 'News – good or bad – and its impact on volatility predictions over multiple horizons', *Rev. Financial Studies* **24**(1), 46–81.

Chorro, C., Guegan, D. and Ielpo, F. (2012), 'Option pricing for GARCH-type models with generalized hyperbolic innovations', *Quantitative Finance* **12**(7), 1079–1094.

Diebold, F. X. and Mariano, R. S. (1995), 'Comparing predictive accuracy', *J. Bus. Econ. Statist.* **13**, 225–263.

Engle, R. F. and Ng, V. (1993), 'Measuring and testing the impact of news on volatility', *J. Finance* **43**, 1749–1778.

Giacomini, R. and White, H. (2006), 'Tests of conditional predictive ability', *Econometrica* **74**, 1545–1578.

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

Heston, S. L. and Nandi, S. (2000), 'A closed form GARCH option valuation model', *Rev. Financial Studies* **13**, 585–625.

Jorion, P. (2006), *Value at Risk*, McGraw-Hill, New York.

Menn, C. and Rachev, S. T. (2009), 'Smoothly truncated stable distributions, GARCH-models, and option pricing', *Math. Methods Operations Research* **69**(3), 411–438.

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

West, K. D. (1996), 'Asymptotic inference about predictive ability', *Econometrica* **64**, 1067–1084.

Zakoïan, J.-M. (1994), 'Threshold heteroskedastic models', *J. Economic Dynamics and Control* **18**, 931–955.