Forecasting for Financial Stock Returns Using a Quantile Function Model

Yuzhi Cai

Abstract—In this talk, we introduce a newly developed quantile function model that can be used for estimating conditional distributions of financial returns and for obtaining multi-step ahead out-of-sample predictive distributions of financial returns. Since we forecast the whole conditional distributions, any predictive quantity of interest about the future financial returns can be obtained simply as a by-product of the method. We also show an application of the model to the daily closing prices of Dow Jones Industrial Average (DJIA) series over the period from 2 January 2004 - 8 October 2010. We obtained the predictive distributions up to 15 days ahead for the DJIA returns, which were further compared with the actually observed returns and those predicted from an AR-GARCH model. The results show that the new model can capture the main features of financial returns and provide a better fitted model together with improved mean forecasts compared with conventional methods. We hope this talk will help audience to see that this new model has the potential to be very useful in practice.

Keywords—DJIA, Financial returns, predictive distribution, quantile function model.

I. Introduction

QUANTILE regression method has been used widely in many areas [1]. This approach estimates a sequence of quantiles of a response variable, leading to a discrete version of the distribution of the response variable. Another quantile approach to statistical modelling is to estimate the whole conditional quantile function of a response variable, see for example, [2] and [3]-[7]. In this talk we will explain how a quantile function model [8] could be used to make predictions for financial returns.

A general quantile function model may be defined by

$$Q_Y(\tau \mid \xi, \mathbf{x}) = h_1(\eta_1, x_1, \ldots, x_p)$$
$$+h_2(\eta_2, x_1, \ldots, x_p)Q(\tau, \gamma),$$

where $\xi = (\eta_1, \eta_2, \gamma)$ is the model parameter vector, h_i $(i = 1, 2)$ are known functions of $\mathbf{x}$ and η_i, $h_2(\eta_2, x_1, \ldots, x_p) > 0$, $Q(\tau, \gamma)$ is the quantile function of the error term with explicit mathematical expression, and $\tau \in (0, 1)$ is the probability that Y takes values that are less than $Q_Y(\tau \mid \xi, \mathbf{x})$.

The specific model we will use is given by

$$Q_{y_t}(\tau \mid \beta, \mathbf{y}_{t-1}) = a_0 + a_1 y_{t-1} + \cdots + a_{k_1} y_{t-k_1}$$
$$+\sqrt{b_0 + b_1 y_{t-1}^2 + \cdots + b_{k_2} y_{t-k_2}^2}\, Q(\tau, \gamma),$$

where

$$Q(\tau, \gamma) = \frac{\tau^{\gamma_1} - 1}{\gamma_1} - \frac{(1-\tau)^{\gamma_2} - 1}{\gamma_2}.$$

Y. Cai is with the School of Management, Swansea University, United Kingdom (e-mail: y.cai@swansea.ac.uk).

So, this model says that not only the location of the distribution of y_t depends on the past values of the series but also the scale of the distribution of y_t also depends on the past values of the series.

It is noticed that once the model has been estimated we can use the fitted model for forecasting. So before applying the model to the Dow Jones Industrial Average (DJIA) series, we first briefly describe the forecasting method. For more details please see [8].

II. The Forecasting Method

Suppose we have estimated the model by using the MCMC method developed in [8]. The posterior samples of the model parameters collected from the MCMC method are denoted by $\beta^{(u)}$ for $u = 1, \ldots, U$. Suppose the length of the observed series is n, we want to forecast the distribution of y_{n+m}, where $m = 1, 2, \ldots$. The forecasting method is a simulation based method and makes a full use of the posterior samples.

Specifically, for 1-step ahead forecasting, i.e. $m = 1$, we have

$$f(y_{n+1} \mid \mathbf{y}_n)$$
$$= \int_\beta f(y_{n+1} \mid \beta, \mathbf{y}_n)\pi(\beta \mid \mathbf{y}_n)d\beta$$
$$\approx \tfrac{1}{U}\textstyle\sum_{u=1}^U f(y_{n+1} \mid \beta^{(u)}, \mathbf{y}_n).$$

This defines a density function of $f(y_{n+1} \mid \mathbf{y}_n)$. So we can obtain a random sample of size I from $f(y_{n+1} \mid \beta^{(u)}, \mathbf{y}_n)$, denoted by $y_{n+1}^{(u,i_1)}$, $i_1 = 1, \ldots, I$, which will be used for the 2-step ahead predictive density function. The sample mean or median may be used as a point forecast.

For 2-step ahead forecasting, we have

$$f(y_{n+2} \mid \mathbf{y}_n)$$
$$= \int_{y_{n+1}} \int_\beta f(y_{n+2} \mid \beta, y_{n+1}, \mathbf{y}_n)$$
$$\times f(y_{n+1} \mid \beta, \mathbf{y}_n)\pi(\beta \mid \mathbf{y}_n)d\beta dy_{n+1}$$
$$\approx \tfrac{1}{U}\textstyle\sum_{u=1}^U \tfrac{1}{I}\sum_{i_1=1}^I f(y_{n+2} \mid \beta^{(u)}, y_{n+1}^{(u,i_1)}, \mathbf{y}_n).$$

So we can also obtain a random sample of size I from $f(y_{n+2} \mid \beta^{(u)}, y_{n+1}^{(u,i_1)}, \mathbf{y}_n)$, which will be used for the 3-step ahead predictive density function. The sample mean or sample median may be used as a 2-step ahead point forecast.

By repeating this procedure we have a random sample from each step ahead distribution, these samples allow us to estimate any predictive quantity of interest.

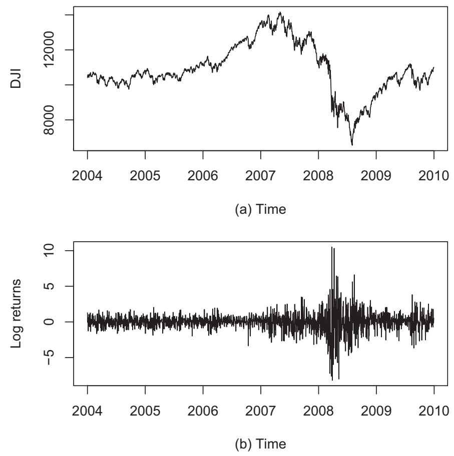

Fig. 1 (a) Time series plot of the DJIA between 2/1/2004-8/10/2010. (b) Time series plot of the DJIA returns

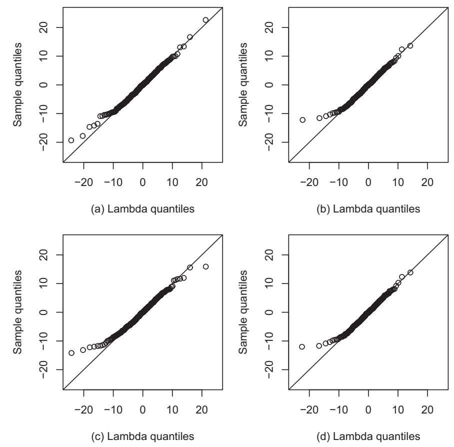

Fig. 2 (a) Time series plot of the DJIA between 2/1/2004-8/10/2010. (b) Time series plot of the DJIA returns

III. Application

We consider Daily Dow Jones Industrial Average (DJIA) in the period 2 January 2004 - 8 October 2010 and we consider its log returns. Figs. 1 (a) and (b) show the time series plots of the DJIA and its log-returns respectively. It is clear that the figure shows some common features of a financial time series.

We fitted four quantile function models with $k_1 = 0, 1$ and $k_2 = 0, 1$ to the DJIA returns and we found that the best fitted model has $k_1 = k_2 = 1$.

The fitted model is

$$Q_{x_t}(\tau \mid \hat{\beta}, \mathbf{x}_{t-1}) = 0.0623 - 0.077x_{t-1}$$

$$+\sqrt{0.113 + 0.042x_{t-1}^2}\left(\frac{\tau^{-0.301}-1}{-0.301} - \frac{(1-\tau)^{-0.209}-1}{-0.209}\right).$$

so the standardized residuals is given by

$$r_t = \frac{x_t - (0.0623 - 0.077x_{t-1})}{\sqrt{0.113 + 0.042x_{t-1}^2}}.$$

A good fitted model is suggested if the distribution of r_t approximately follows the distribution defined by

$$\hat{Q}(\tau, \hat{\gamma}) = \frac{\tau^{-0.301} - 1}{-0.301} - \frac{(1-\tau)^{-0.209} - 1}{-0.209},$$

which may be check by using a QQ-plot between r_t and $\hat{Q}(\tau, \hat{\gamma})$. Fig. 2 shows the QQ-plots of the four model with different orders. It confirms that the model with $k_1 = 1$ and $k_2 = 1$ is the best.

For comparison purpose, we also fitted a sequence of ARMA-GARCH models to the same data. Figure 3 shows the QQ-plots of all the fitted models with different orders. It seems that they all behave very similarly, but the AIC values suggest that the estimated AR(1)-GARCH(1,1) with t-innovations is a

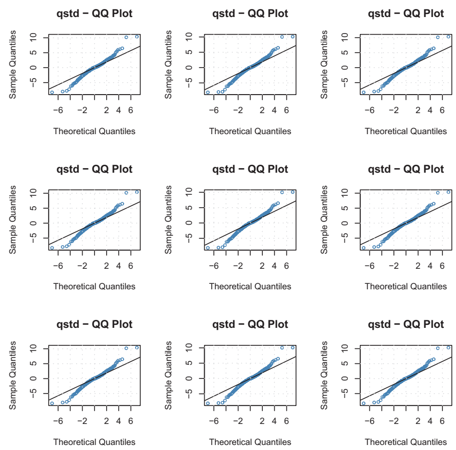

Fig. 3 (a) Time series plot of the DJIA between 2/1/2004-8/10/2010. (b) Time series plot of the DJIA returns

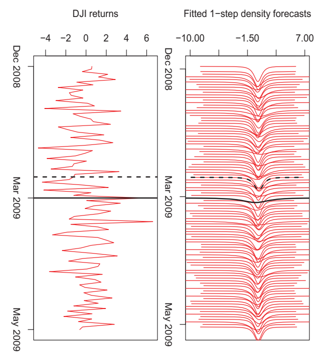

Fig. 4 (a) Time series plot of the DJIA between 2/1/2004-8/10/2010. (b) Time series plot of the DJIA returns

better one. So we consider the estimated AR-GARCH model

TABLE I
OUT-OF-SAMPLE POINT FORECASTS FOR THE LOG RETURNS OF THE DJIA

Steps	Observed	Predicted (Q-AR)	2.5% quantile	97.5% quantile	Predicted (AR-GARCH)	Lower CI	Upper CI
1	0.035	-0.034	-0.465	0.366	0.005	-1.646	1.656
2	0.091	-0.008	-0.514	0.451	0.034	-1.622	1.691
3	0.684	-0.004	-0.466	0.443	0.033	-1.630	1.695
4	0.008	-0.007	-0.543	0.457	0.033	-1.635	1.701
5	-0.308	-0.013	-0.562	0.470	0.033	-1.641	1.707
6	0.729	0.007	-0.516	0.520	0.033	-1.647	1.712
7	-1.492	-0.014	-0.507	0.483	0.033	-1.653	1.718
8	1.171	-0.002	-0.451	0.481	0.033	-1.658	1.724
9	0.347	0.004	-0.517	0.504	0.033	-1.664	1.729
10	-0.126	-0.018	-0.510	0.452	0.033	-1.669	1.735
11	0.282	0.023	-0.463	0.511	0.033	-1.675	1.740
12	0.048	-0.010	-0.494	0.440	0.033	-1.680	1.746
13	-0.387	-0.001	-0.514	0.473	0.033	-1.686	1.751
14	-0.111	0.004	-0.501	0.478	0.033	-1.691	1.756
15	0.040	-0.002	-0.4898	0.419	0.033	-1.696	1.762
MSE		0.3346521			0.3357334		

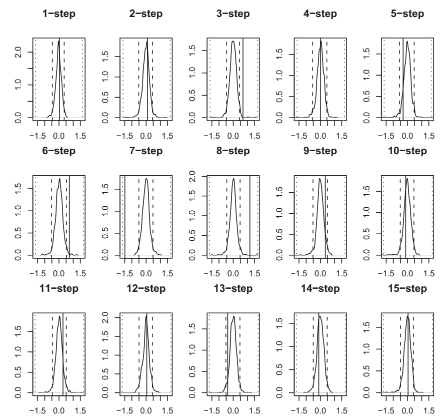

Fig. 5 (a) Time series plot of the DJIA between 2/1/2004-8/10/2010. (b) Time series plot of the DJIA returns

with t-innovations given below.

$$y_t = 0.0327 - 0.0557 y_{t-1} + \sqrt{h_t}\varepsilon_t\,, \qquad (1)$$
$$\quad (0.0184) \quad (0.0242)$$

where

$$h_t = 0.0085 + 0.0808 v_{t-1}^2 + 0.9143 h_{t-1},$$
$$(0.0036) \quad (0.0133) \quad (0.0133)$$

where

$$v_t = y_t - 0.0327 + 0.0557 y_{t-1}$$

and ε_t follows the t-distribution with 7.2150(1.3335) degrees of freedom, and the numbers in brackets are the standard errors of the estimated parameter values.

Fig. 4 shows the one-step ahead predictive density functions during the period from 23 December 2008 to 19 May 2009.

The differences between the predictive density functions indicate the effects of the differences in information sets. Larger absolute returns implies a higher level of uncertainty leading to a very flat predictive distribution of the returns on the next day.

Fig. 5 shows out-of-sample predictive probability distributions up to 15 steps ahead. where continuous vertical lines represent the actually observed returns on these days, the dashed vertical lines give a 95% probability interval of the estimated distributions, and the dotted vertical lines give a 95% confidence interval of the distribution obtained from the estimated AR-GARCH model. It is seen that the 95% confidence intervals of the distribution obtained from the estimated AR-GARCH model are much wider than those obtained from the quantile function model, which may suggest some uncertainties that are involved in the estimation of the AR-GARCH model. It is also seen that these predictive distributions also enable us to study any multi-step ahead predictive quantity about the DJIA returns.

For example, Table I shows the out-of-sample point forecasts for the log returns of the DJIA. The MSE values show that our model has a slightly better performance than the other model.

IV. CONCLUSIONS

We showed how to use a quantile function model to analyze financial returns. We found that this model can provide an improved fit, which suggests that this new model can capture the main features of most financial return series including extreme returns, skewness and volatility clustering.

Our results show that the predictive distributions of the DJIA returns depend on the past information, they are skewed and they have thicker tails compared with those obtained from other models.

Although the observed returns covers the 2008 economic crisis period, our model dealt with this situation well. All these show that the quantile function model has the potential to be very useful for financial time series in practice.

We have not compared the quantile function model with other models including the important CAViaR model [9]. We will carry out such comparisons in the future.

REFERENCES

[1] Koenker, R. (2005). *Quantiles Regression*. Cambridge University Press.
[2] Gilchrist, W.G. (2000). *Statistical Modelling with Quantile Functions*. Chapman & Hall/CRC.
[3] Cai, Y. (2015). A general quantile function model for economic and financial time series. *Econometric Reviews*. Accepted.
[4] Cai, Y. (2013). Quantile function models for survival data analysis. *Australian and New Zealand Journal of Statistics* 55, 155-172.
[5] Cai, Y. (2010a). Multivariate quantile function models. *Statistica Sinica* 20, 481-496.
[6] Cai, Y. (2010b). Polynomial power-Pareto quantile function models. *Extremes* 13, 291-314.
[7] Cai, Y. (2009). Autoregression with non-Gaussian Innovations. *Journal of Time Series Econometrics*, Vol.1, Iss.2, Article 2. DOI: 10.2202/1941-1928.1016.
[8] Cai, Y, Montes-Rojas, G. and Olmo, J. (2013). Quantile double AR time series models for financial returns. *Journal of Forecasting* 32, 551-560.
[9] Engle, R.F. and Manganelli, S. (2004). CAViaR: Conditional autoregressive value at risk by regression quantiles. *Journal of Business and Economic Statistics* 22, 367-381.