International Journal of the Physical Sciences Vol. 6(30), pp. 6867 - 6871, 23 November, 2011
Available online at http://www.academicjournals.org/IJPS
DOI: 10.5897/IJPS11.1502

Full Length Research Paper

Wind speed forecasting based on autoregressive moving average- exponential generalized autoregressive conditional heteroscedasticity-generalized error distribution (ARMA-EGARCH-GED) model

Hongkui Li [1], Ranran Li[1]* and Yanlei Zhao[2]

[1]School of Science, Shandong University of Technology, Shandong, 255049, People's Republic of China.
[2]School of Electrical and Electronic Engineering, Shandong University of Technology, Shandong, 255049, People's Republic of China.

Accepted 19 October, 2011

With the increase of wind power as a renewable energy source in many countries, wind speed forecasting has become more and more important to the planning of wind speed plants, the scheduling of dispatchable generation and tariffs in the day-ahead electricity market, and the operation of power systems. However, the uncertainty of wind speed makes troubles in them. For this reason, a wind speed forecasting method based on time-series is proposed in this paper. We adopt exponential GARCH (EGARCH) models as asymmetric specifications and GARCH-GED for distribution assumptions. The wind speed series are forecasted by using the autoregressive moving average (ARMA)-GARCH model, ARMA-GARCH-M model and ARMA-GARCH-GED model, respectively, after which the forecasting precision of ARMA-GARCH, ARMA-GARCH-M and ARMA-EGARCH-GED models are compared. The results show that ARMA-EGARCH-GED model possesses higher accuracy than ARMA-GARCH-M model (Lalarukh and Yasmin, 1997), and is of certain practical value. However, this study confirms that the conditional generalized error distribution (GED) can better describe the possibility of fat-tailed, non-normal conditional distribution of returns.

Key words: Wind speed, forecasting, ARMA, GARCH, GARCH-GED.

INTRODUCTION

Electricity is essential in our daily life and more scholars dedicate their study to the power research (Hadi and Azah, 2011; Al-Ammar and El-Kady, 2011). Wind power is widely seen as a renewable energy source with the best chance of competing with fossil-fuel power stations in the near term. In the year 2009, by the World Wind Energy Report (World Wind Energy Association.World Wind Energy Report, 2009), altogether 82 countries used wind energy on a commercial basis, out of which 49 countries increased their installed capacity, the capacity is more than the installed capacity, doubling every third year. The growth rate went up steadily in the market for new wind turbines since the year 2004, reaching 29.0% in 2008, after 26.6% in 2007, 25.6% in the year 2006 and 23.8% in 2005 (World Wind Energy Association. World Wind Energy

*Corresponding author. E-mail: fordarkblue@yahoo.com.cn.

Abbreviation: ARMA, Autoregressive moving average; **GED,** generalized error distribution; **GARCH-M,** GARCH-in-mean.

Report, 2008). It reached 38'312 MW in 2009, which increased to 42.1%. The size of 2009 is 10 times of the size of ten years ago. The growth rate of wind power is 31,7% in 2009, the highest rate since 2001 (World Wind Energy Association.World Wind Energy Report, 2009).

However, wind is an intermittent energy source, which means that there exists large variability in the production of energy due to various factors, and is regarded as a non-dispatchable energy source. It makes troubles in ensuring the reliability of wind power generation since the wind speed is uncertain. So, power generation from wind has some drawbacks. For these reason, the wind speed forecasting method is attracting more attention. In the past decades, pepole have developed different wind power forecasting models by using physical or statistical approaches. By using the physical approach, they estimate local wind speed using the physical laws governing atmospheric behaviour. By using the statistical approach, they determine the relationship between a set of explanatory variables and the power generated at the wind. Lalarukh and Yasmin (1997) presented the stochastic simulation and forecast models of hourly average wind speeds. Several basic features of wind speed data, such as autocorrelation, non-Gaussian distribution, diurnal nonstationarit and the positive correlation between consecutive wind speed observations are taken into account by fitting an ARMA (p,q) process to wind speed data transformed to make their distribution approximately Gaussian and standardized to remove scattering of transformed data.

Since accurately modeling the mean and volatility of wind speed can be beneficial to effective wind energy utilization, Liu et al. (2011) evaluates the effectiveness of autoregressive moving average generalized autoregressive conditional heteroscedasticity (ARMA-GARCH) approaches for modeling the mean and volatility of wind speed. It includes five different GARCH approaches, and each approach consists of an original form and a modified form, GARCH-in-mean (GARCH-M). They proved that the ARMA-GARCH(-M) approaches can effectively catch the trend change of the mean and volatility of wind speed, and the ARMA-GARCH-M structures can also consistently improve the modeling sufficiency of mean wind speed.

Based on time-series, a wind speed forecasting method is proposed in this paper. We adopt EGARCH models as asymmetric specifications and GARCH-GED for distribution assumptions. The wind speed series are forecasted by using ARMA-GARCH model, ARMA-GARCH-M model and ARMA-GARCH-GED model, respectively. Forecasting precision of ARMA-GARCH model, ARMA-GARCH-M model and ARMA-EGARCH-GED model are compared. The results show that ARMA-EGARCH-GED model possesses higher accuracy than ARMA-GARCH-M model in Lalarukh and Yasmin (1997), and is of certain practical value. We confirms that the conditional generalized error distribution (GED) can better account for the possibility of fat-tailed, non-normal conditional distribution of returns, and performs exceedingly well at all common risk levels (ranging from 0.25 to 10%).

BASIC PRINCIPLE OF ARMA-EGARCH-GED MODEL

Here, the basic concept of ARMA model, ARCH model, GARCH model and EGARCH model are introduced.

ARMA model

A typical linear ARMA model for conditional mean which is denoted by $ARMA(r,m)$ is expressed as:

$$y_t = c + \sum_{i=1}^{r} \phi_i y_{t-i} + \sum_{j=1}^{m} \theta_j \varepsilon_{t-j} + \varepsilon_t \quad (1)$$

where y_t is the time series needed to be modeled, c is a constant term of the ARMA model, r is the number of autoregressive orders, m is the number of moving average orders, ϕ_i is i th autoregressive coefficients, θ_j is j th moving average coefficients and ε_t is the error term at time period t.

Stationarity test

Sometimes, time series has to be stationary since not every measurement is suited for a dimension estimate. The stationarity test decides whether the data requires differencing. The data must be transformed in ARMA modeling to stationary form prior to analysis.

But different stationary models of the same time series may suggest very different predictions. Therefore, it is tremendously important for applied forecasters to decide on which model to use. Since one of the early motivations for unit root tests was precisely to help determine whether to use forecasting models in differences or levels in a particular applications (Dickey et al., 1986), unit root can test often be used as a diagnostic tool to guide the decision.

The most frequently utilized test for unit roots is the augmented Dickey-Fuller regression (ADF) (Dickey and Fuller, 1981) which is a generalized auto-regression model formulated in the following regression equation:

$$\nabla y_t = \gamma y_{t-1} + \xi_1 \nabla y_{t-1} + \xi_2 \nabla y_{t-2} + \cdots + \xi_{p-2} \nabla y_{t-p} + \xi_{p-1} \nabla y_{t-p+1} + \varepsilon_t \quad (2)$$

where p is determined by minimizing the Akaike information criterion.

Akaike's information criterion (AIC)

Akaike's information criterion, proposed by Akaike (1973), is a way of selecting the best fitted model from a set of models by using a function based on likelihood. The selected model is the one that minimizes the Kullback-Leibler distance between the model and the truth, and has a good fit to the truth but few parameters. It is defined as:

$$AIC = -2\ln(likehood) + 2K$$

where $\ln$ is the natural logarithm, $likehood$ is the natural logarithm, K is the number of parameters in the model.

The ARCH model is extremely useful for modelling economic data possessing high volatility. It provides a systematic framework for volatility, modeling statistical methods for time series models with ARCH errors been rapidly developed since the ARCH model was introduced by Engle in 1982.

The general $GARCH(p,q)$ model for the conditional variance of innovations ε_t is:

$$\sigma_t^2 = k + \sum_{i=1}^{p}\alpha_i\sigma_{t-i}^2 + \sum_{j=1}^{q}\beta_j\varepsilon_{t-j}^2 \quad (3)$$

With constraints:

$$\sum_{i=1}^{p}\alpha_i + \sum_{j=1}^{q}\beta_j < 1, \qquad k > 0, \quad \alpha_i \geq 0(i = 1,2,\cdots,p) \quad ,$$
$$\beta_j \geq 0(j = 1,2,\cdots,q) \quad (4)$$

By Equation 4, we know that the GARCH process is weakly stationary since the mean, variance and auto covariance are finite and constant over time.

The GARCH model can capture the stylized facts of financial time series, such as time-varying volatility, persistence and volatility clustering. However, it fails to capture asymmetric volatility. For overcoming this limitation, Nelson (1991) presented the exponential GARCH (EGARCH) model. Nelson and Cao (1992) found that the nonnegativity constraints in the linear GARCH model are too restrictive. It imposes the nonnegative constraints on the parameters, α_i and β_i. But there are no restrictions on these parameters in the EGARCH model. In the EGARCH model:

$$\ln(\sigma_t^2) = \omega + \sum_{i=1}^{p}\alpha_i\left[\left|\varepsilon_{t-i}\right|/\sigma_{t-i} - \gamma_i\varepsilon_{t-i}/\sigma_{t-i} - \sqrt{\frac{2}{\pi}}\right] + \sum_{j=1}^{p}\beta_j\ln(\sigma_{t-j}^2)$$

The EGARCH model has the following properties:

1. The function $\left|\varepsilon_{t-i}\right|/\sigma_{t-i} - \gamma_i\varepsilon_{t-i}/\sigma_{t-i} - \sqrt{\frac{2}{\pi}}$ is linear in ε_t with slope coefficient $\theta+1$ if ε_t is positive, while $\left|\varepsilon_{t-i}\right|/\sigma_{t-i} - \gamma_i\varepsilon_{t-i}/\sigma_{t-i} - \sqrt{\frac{2}{\pi}}$ is linear in ε_t with slope coefficient $\theta-1$ if ε_t is negative;
2. When $\theta = 0$, then the large innovations increase the conditional variance if $\left|\varepsilon_t\right| - E\left|\varepsilon_t\right| > 0$, and decrease the conditional variance if $\left|\varepsilon_t\right| - E\left|\varepsilon_t\right| < 0$;
3. When $\theta < 1$ and the innovations z_t are less than $\sqrt{2/\pi}\Big/\theta-1$, then the innovation in variance, $\left|\varepsilon_{t-i}\right|/\sigma_{t-i} - \gamma_i\varepsilon_{t-i}/\sigma_{t-i} - \sqrt{\frac{2}{\pi}}$, is positive. Thus, the negative innovations in returns, ε_t, cause the innovation to the conditional variance to be positive if ϑ is much less than 1.

WIND SPEED FORECASTING BASED ON ARMA-EGARCH-GED MODEL

For the wind speed time series, when the sampling time is 1 min, we first construct the ARMA model, then we use Lagrange multiplier method to test their residuals, but we find that there is obvious heteroscedasticity. After verifying, heteroscedasticity of residuals can be fit by EGARCH-M-GED model. Then, the ARMA-EGARCH-M-GED model is consrtructed.

We use ARMA-ARCH model and ARMA-EGARCH-M-GED model to do the sample forecasting, respectively, and compare forecasts with the measured data of October 14, then we calculated the forecast absolute average error. The compared predictive power is as shown in Table 1, and the predictive is as shown in Figure 1.

By the comparison of Table 1 and Figure 1, we can see that the ARMA (3,1) -EGARCH(1,0)-M-GED model has the following properties:

Table 1. The comparison of predictive power.

Model	The absolute average error (%)	Skewness	Variance
ARMA-ARCH	31.11	0.029677	0.167012
ARMA (3,1)-EGARCH (1,0)-M-GED	30.09	0.009122	0.134312

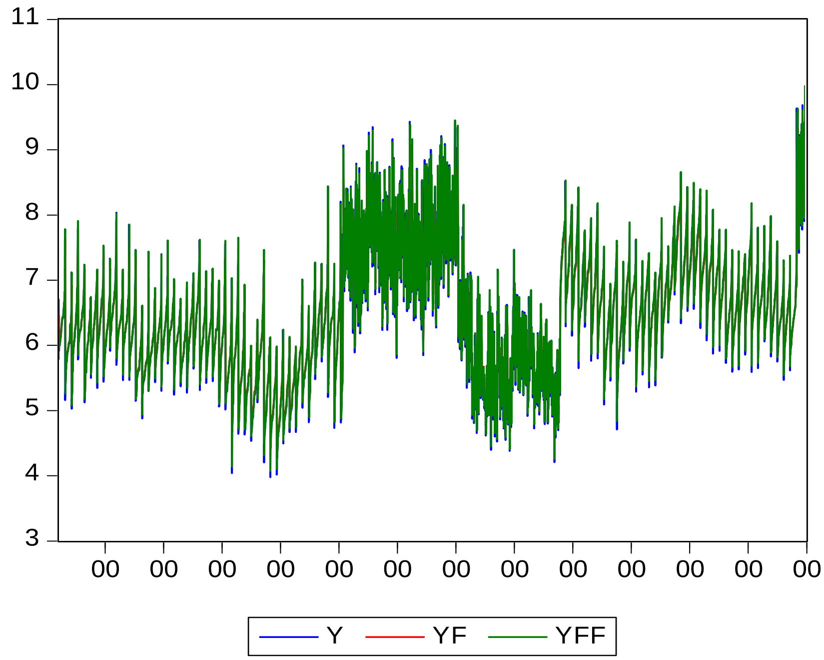

Figure 1. Predictions: Y, Original data; YF, ARMA-ARCH predictions; YFF, ARMA(3,1)-EGARCH(1,0)-M-GED predictions.

1. Wind speed time series tend to have volatility clustering phenomenon (that is, large fluctuations is immediately followed by large fluctuations and small fluctuation is immediately followed by smaller fluctuations), and its variance changes over time, which does not match the ARMA model requests with the same variance. EGARCH-M model can well describe the characteristics of data's heteroscedasticity, it can not only overcome the traditional GARCH model parameters' non-negative constraints, but also reflects the asymmetric effect caused by positive and negative shocks ($\gamma > 0$, that is, the impact of positive shocks is greater than negative shocks; $\gamma < 0$, the impact of negative shocks is greater than positive shocks), and it has great flexibility.
2. The data kurtosis of general normal distribution is 3 (the data are calculated by dividing the fourth-order moments by the square of the second-order moment), when the data kurtosis is large than the normal one, the data are found with a thick tail. Wind speed time series is usually with a certain thick tail. The data sequence with a thick tail, the probability of extreme values of which is larger than the probability of the normal data. Generalized error distribution (GED) is the most typical one of several thick-tailed distributions, and it can better describe the wind speed time series with a thick tail.
3. ARMA-EGARCH-M-GED model is the promotion and correction of classical ARMA model, and for wind speed time series prediction, the prediction has higher accuracy, which has some practical value.

ACKNOWLEDGEMENTS

This work was supported financially by the National Natural Science Foundation of China Grants (50807034) Shandong Province Natural Science Foundation (ZR2010EQ014).

REFERENCES

Akaike H (1973). Information theory and extension of the maximum likelihood principle. In: Petrov BN, Cszaki F, editors. Second international symposium on information theory. Akademiai Kiado, Budapest, pp. 267-81.

Al-Ammar EA, El-Kady MA (2011). Efficient calculation of operating security regions in power systems. Int. J. Phys. Sci., 6(14): 3475–3486.

Hadi Z, Azah M (2011). Hussain Shareef and Marjan Mohammadjafari Impact of distributed generations on power system protection performance. Int. J. Phys. Sci., 6(16): 3999-4007.

Lalarukh K, Yasmin ZJ (1997). Time series models to simulate and forecast hourly averaged wind speed in Quetta, Pakistan, Solar Energy,61: 23-32.

Liu HP, Ergin E, Shi J (2011). Comprehensive evaluation of ARMA-GARCH(-M) approaches for modeling the mean and volatility of wind speed, Appl. Energy, 88: 724-732.

World Wind Energy Association.World Wind Energy Report (2009). <http://www.wwindea.org/home/images/stories/worldwindenergyreport_s.pdf>

World Wind Energy Association.World Wind Energy Report (2008). <http://www.wwindea.org/home/images/stories/worldwindenergyreports.pdf>