HÖGSKOLAN
Dalarna

Conditional Skewness and Kurtosis in GARCH Model

Author:
HENG WANG

Supervisor:
CHANGLI HE

June 21, 2011

Abstract

This paper proposes the GARCH with skewness (GARCHS) process for estimating time-varying conditional variance and skewness. A skew normal distribution is assumed in GARCHS model. I also extend the GARCHS process to a GARCH with skewness and kurtosis (GARCHSK) process for conditional kurtosis. This model is based on the assumption that error terms follow a distribution of Gram-Charlier series expansion of the normality, in which conditional skewness and kurtosis are parameters can be used directly. I apply the two methods to financial daily returns of Nordic exchange market. Parameters are estimated through MLE by BHHH algorithm. The results indicate the performance of conditional skewness and kurtosis and show a difference between GARCHS and GARCHSK model.

Contents

1 Introduction

The second moment of financial returns, variance, is an important measurement of the market fluctuation. Engel (1988) introduced an autoregressive conditional heteroskedasticity (ARCH) process to model variance and for a long time, and the ARCH model had been widely used. As more research were focused on the relation between variance and expected return, the self-related property of variance was mostly believed. Bollerslev (1986) developed the ARCH model to a more general one, the GARCH model, with the effect of self regressive considered.

Skewness, measured by the third moment, is found in many economic data such as stock or exchange index returns. The existence of skewness made an asymmetric conditional distribution of returns. The negative skewness indicates a higher probability of negative returns and positive skewness the opposite. Due to the important impact of skewness a lot of asymmetric distribution classed were introduced to the GARCH model. Similar to skewness, what the fourth moment of returns concern, kurtosis, has also drawn attention. A large kurtosis implies a sharp peak of the distribution, i.e. the probability of limited around the mean is extremely large as well.

However, in most empirical work, the conditional skewness and kurtosis are observed not consistent within a long time period. The original GARCH model is not able to capture the dynamics of time-varying skewness and kurtosis. Harvey (1999) presented a new methodology, GARCH with skewness (GARCHS) model, to estimate dynamic conditional skewness by introducing the third moment equation. And Brooks (2005) set a GARCH with kurtosis (GARCHK) model to capture dynamic kurtosis via constructing a fourth moment equation. Both of them adapted a noncentral t distribution to the error terms which brought also a large amount of calculation. Instead, Leon (2004) used a Gram-Charlier series expansion of normality distribution (GCD) to model the conditional skewness and kurtosis simultaneously. The new method, GARCH with skewness and kurtosis (GARCHSK) model, made estimation more easily and a more comprehensive description of financial returns.

This paper studies the conditional skewness of financial returns through a GARCHS model under the assumption of skew normal distribution and a GARCHSK model introduced by Leon (2004). Specifically, I present a maximum likelihood frame work for estimating time-varying variance, skewness and kurtosis under the assumption of the two models. I use the models to model daily returns on Nordic 40 OMX Index. Both results show a strong evidence that conditional skewness and kurtosis exist and perform an important role as time varies. And the comparison of the results indicate either model has some better properties than the other.

In the diagnostic section, I carry out some illustration to examine the model fitness by comparing the differences between the short-period sample variance, skewness, kurtosis and these estimated by models. The figures are clearly to point out how good either model fits, as well as the specifically conditional moment test.

2 Method

2.1 AR-GARCH process

Before Engle (1982) introduced the ARCH (Autoregressive Conditional Heteroskedastic) process, the classical assumptions of time series and econometric models usually regarded the variance as constant terms. The ARCH model allows the conditional variance to change over time as a function of past errors while unconditional variance remain constant. Bollerslev (1986) developed the model to a generalized ARCH model by adding the lagged conditional variance to the equation, which is the widely used GARCH model.

Let ϵ_t denote a discrete-time stochastic process, and I_t the information set through time t. The GARCH(p,q) process can be given as:

$$\epsilon_t | I_{t-1} \sim N(0, h_t),$$

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i h_{t-i}$$

where

$$p \geq 0, q > 0$$

$$\alpha_0 > 0, \alpha_i \geq 0, i = 1, \ldots, q,$$

$$\beta_i \geq 0, i = 1, \ldots, p.$$

The equation with only first order of ϵ_t is mean equation and other variance equation. In the above equation, conditional distribution of ϵ_t is considered to be a normal distribution with zero mean and variance of h_t. When $p = 0$, the process is just ARCH(q) process without its own lag effect; when $p = q = 0$, ϵ_t is white noise with variance of α_0. Bollerslev gave the stationary condition of GARCH process with the parameter constrains,

$$\sum_{i=1}^{q} \alpha_i + \sum_{i=1}^{p} \beta_i < 1$$

We usually generate the residual from mean equation. The GARCH(p,q) regression model is obtained by letting the ϵ_t be innovations in a linear regression,

$$\epsilon_t = y_t - x_t' b,$$

where y_t and x_t are observed variable and b a vector of unknown parameters. In the case of financial returns, we only have one variable of r_t, so an autoregressive model is a appropriate way.Thus the mean equation is expressed as an AR(s) process

$$\epsilon_t = r_t - \sum_{i=1}^{s} b_i r_{t-i},$$

where $b = (b_1, \ldots, b_s)$ is autoregressive coefficients to be estimated.

In the after years, when GARCH was widely used, more and more time series data were found asymmetric against the normality assumption, so a new adjustment was needed. Researchers have tried kinds of skewed distribution such as skewed t distribution, Standardized Skewed-Generalized Error Distribution (GED), etc and results indicate a more fitted model. In Bollerslev's paper, he estimated the model by BHHH algorithm to maximize the log-likelihood function. Similar to the normal distribution, the estimation can be worked out by the same method under the assumption of more complicated skewed distribution.

2.2 Time varying skewness and kurtosis

Engle and Bollerslev constructed conditional kurtosis via the conditional variance under the assumption of Gaussian density. When modeling the term structure of interest rates, Hansen (1994) extended the GARCH model to allow for time-varying skewness and kurtosis by an alternative parameterization of non-central t distribution. Thereafter, Harvey and Siddique (1999) proposed a methodology for estimating time-varying conditional skewness. The new model shows that the conditional variance and skewness are both autoregressive independently, which means not tied together. In Harvey's model, if we use a skew normal distribution (see Appendix 1) instead of non-central t distribution, then we get a new GARCHS model. In this model, given a series of stock prices $\{P_0, P_1, \ldots, P_T\}$, we define continuously financial returns at time t as

$$r_t = lnP_t - lnP_{t-1}, t = 1, 2, \ldots, T.$$

Specifically, we present a financial return model through GARCH(1,1) structure for conditional variance and skewness. For the mean equation, we adapt a AR(1) process. Thus the new model is expressed as following:

$$\begin{aligned} r_t &= \alpha r_{t-1} + \epsilon_t \\ h_t &= \beta_0 + \beta_1 \epsilon_{t-1}^2 + \beta_2 h_{t-1} \\ s_t &= \gamma_0 + \gamma_1 \epsilon_{t-1}^3 + \gamma_2 s_{t-1} \end{aligned}$$

where h_t is the conditional variance of r_t,
s_t is the conditional skewness of r_t,
$\epsilon_t | I_{t-1} \sim SN(0, w_t, p_t)$.

Brooks et al (2005) use the Student's t distribution to model conditional kurtosis separated from conditional variance. Leon at el (2005) developed a GARCH-type model assuming a Gram-Charlier series expansion (see Appendix 2) of the normal density function for the error term, which is easier to estimate than the model by Harvey & Siddique and Brooks. Following Leon (2005), the GARCHSK model is expressed as following:

$$\begin{aligned} r_t &= \alpha r_{t-1} + \epsilon_t \\ h_t &= \beta_0 + \beta_1 \epsilon_{t-1}^2 + \beta_2 h_{t-1} \\ s_t &= \gamma_0 + \gamma_1 \eta_{t-1}^3 + \gamma_2 s_{t-1} \\ k_t &= \delta_0 + \delta_1 \eta_{t-1}^4 + \delta_2 k_{t-1} \end{aligned}$$

where h_t is the conditional variance of r_t,

s_t is the conditional skewness of η_t,

k_t is the conditional kurtosis of η_t,

$\eta_t = h_t^{-\frac{1}{2}} \epsilon_t$.

Suppose η_t follows a conditional distribution of Gram-Charlier series expansion of normal density function. Therefore the conditional distribution of η_t can be expressed as

$$f(\eta_t | I_{t-1}) = \phi(\eta_t)\psi(\eta_t)^2/\Gamma_t,$$

where

$$\psi(\eta_t) = 1 + \frac{s_t}{6}(\eta_t^3 - 3\eta_t) + \frac{k_t - 3}{24}(\eta_t^4 - 6\eta_t^2 + 3),$$

$$\Gamma_t = 1 + \frac{s_t^2}{6} + \frac{(k_t - 3)^2}{24}.$$

We call this specification of variance, skewness and kurtosis the ARGARCHSK(1,1,1,1) model. The parameters need to be constrained to ensure that conditional variance and kurtosis are positive and the three properties stationary. Harvey and Siddique (1999) imposed the constraints of variance and skewness equation that $\beta_0 > 0$, $0 < \beta_1 < 1$, $0 < \beta_2 < 1$, $-1 < \gamma_1 < 1$, $-1 < \gamma_2 < 1$, and $\beta_1 + \beta_2 < 1$ and $-1 < \gamma_1 + \gamma_2 < 1$. Similar to variance equation, we have the constraints of kurtosis equation that $0 < \delta_1 < 1$, $0 < \delta_2 < 1$ and $\delta_1 + \delta_2 < 1$.

3 Estimation

In this section we adapt maximum likelihood estimation of the ARGARCHSK regression model. The method is extended from that of GARCH regression model so the process will be very similar (see Appendix 3).

Denote parameters to be estimated as θ, then the first order derivative of log-likelihood function is $\frac{\partial l_t}{\partial \theta} = (\frac{\partial l_t}{\partial \alpha}, \frac{\partial l_t}{\partial \beta}, \frac{\partial l_t}{\partial \gamma}, \frac{\partial l_t}{\partial \delta})$. To calculate maximum likelihood estimates, we need an iterative procedure. In this case, the second order derivative is more complicate to work out compared to the GARCH process. The Berndt, Hall, Hall and Hausman (BHHH, 1974) algorithm works a more efficient way using the summary of products of first order derivative instead of Hessian matrix. Denote $\theta^{(i)}$ as the parameter estimates after the ith iteration. $\theta^{(i+1)}$ is calculated from the recursive from

$$\theta^{(i+1)} = \theta^{(i)} + \lambda_i \left(\sum_{t=1}^{T} \frac{\partial l_t}{\partial \theta}|_{\theta=\theta^{(i)}} \frac{\partial l_t}{\partial \theta'}|_{\theta=\theta^{(i)}} \right) \sum_{t=1}^{T} \frac{\partial l_t}{\partial \theta}|_{\theta=\theta^{(i)}}$$

where λ_i is a variable step length freely chosen to maximize the likelihood function, accelerating or slowing down the recursion.

According to the asymptotic theory of maximum likelihood estimation, when sample size T is sufficiently large, the estimates can be well approximated by the following distribution:

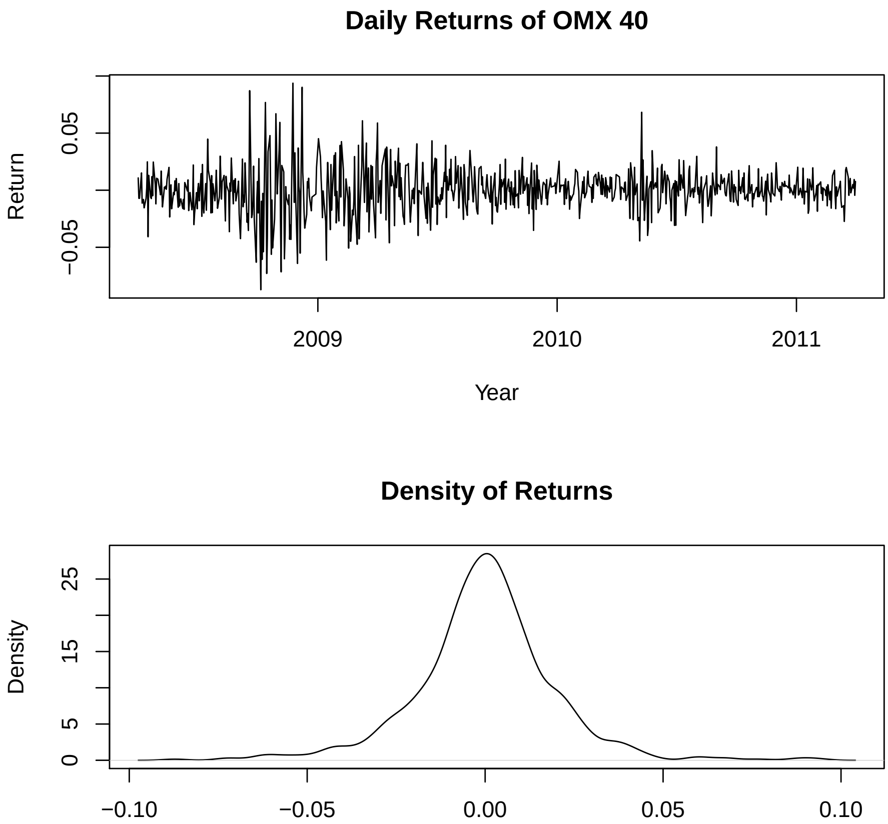
$$\hat{\theta} \sim N(\theta_0, T^{-1}I^{-1})$$

I is know as the information matrix with one outerproduct estimate as

$$I_{OP} = T^{-1} \sum_{t=1}^{T} \frac{\partial l}{\partial \theta} \frac{\partial l}{\partial \theta'}$$

We usually get I_{OP} from the last BHHH iteration for the final estimates.

Figure 1



4 Empirical Results

4.1 Data

Our data is daily OMX Nordic 40 index for pan-regional Nordic Stock Exchange from May.1, 2008 to April.30, 2011. It is a market value-weighted index that consists of the 40 most-traded stock classes of shares from the four stock markets in the Nordic countries - Denmark, Finland, Iceland and Sweden. We calculate the daily returns from the index data. The first part of Figure 1 illustrates the variation of daily returns. It is obviously variation varies with the return, when the absolute value of return is quite large or small, so is the variance. This indicates a time-varying conditional variance model. From the second part of Figure 1, an unconditional distribution similar to normal distribution is suggested, however with high peak and a little asymmetry. Supplementary Table shows monthly return, variance, skewness and kurtosis of continuously compounded daily returns. These are unconditional moments directly from the sample within a period of one month. Clearly, the sample summary shows significantly a unconditional skewness of 0.08985 and a unconditional kurtosis of 6.18477 as an evidence of existence of skewness and kurtosis. And from the long period monthly summary, skewness and kurtosis vary frequently through time. Skewness is sometimes negative sometimes positive implying an asymmetry of returns. However kurtosis doesn't change a lot due to the standardized moment.

Table 1
Model Estimates for ARGARCHSK(1,1,1,1)

Parameter	GARCHS		GARCHSK	
α_0	0.046629	***	-0.00327	***
	(2.40649×10^{-9})		(2.21279×10^{-13})	
β_0	0.049181	***	0.00000	***
	(9.02159×10^{-9})		(2.00997×10^{-15})	
β_1	0.02891	***	0.03525	***
	(3.69838×10^{-10})		(2.01441×10^{-07})	
β_2	0.93223	***	0.94352	***
	(7.16530×10^{-10})		(4.10786×10^{-07})	
γ_0	0.00467	***	0.00581	*
	(1.03314×10^{-10})		(3.92010×10^{-05})	
γ_1	0.00693	***	0.00410	**
	(1.48426×10^{-12})		(1.32925×10^{-05})	
γ_2	0.94584	***	0.07108	
	(2.82919×10^{-9})		(7.61839×10^{-01})	
δ_0			0.10496	***
			(4.08610×10^{-05})	
δ_1			0.00273	***
			(1.28198×10^{-08})	
δ_2			0.95955	***
			(5.45480×10^{-06})	

The t-Statistics are reported with * denoting significance at 40%,
** denoting significance at 30%, and *** denoting significance at 5%.

4.2 Results

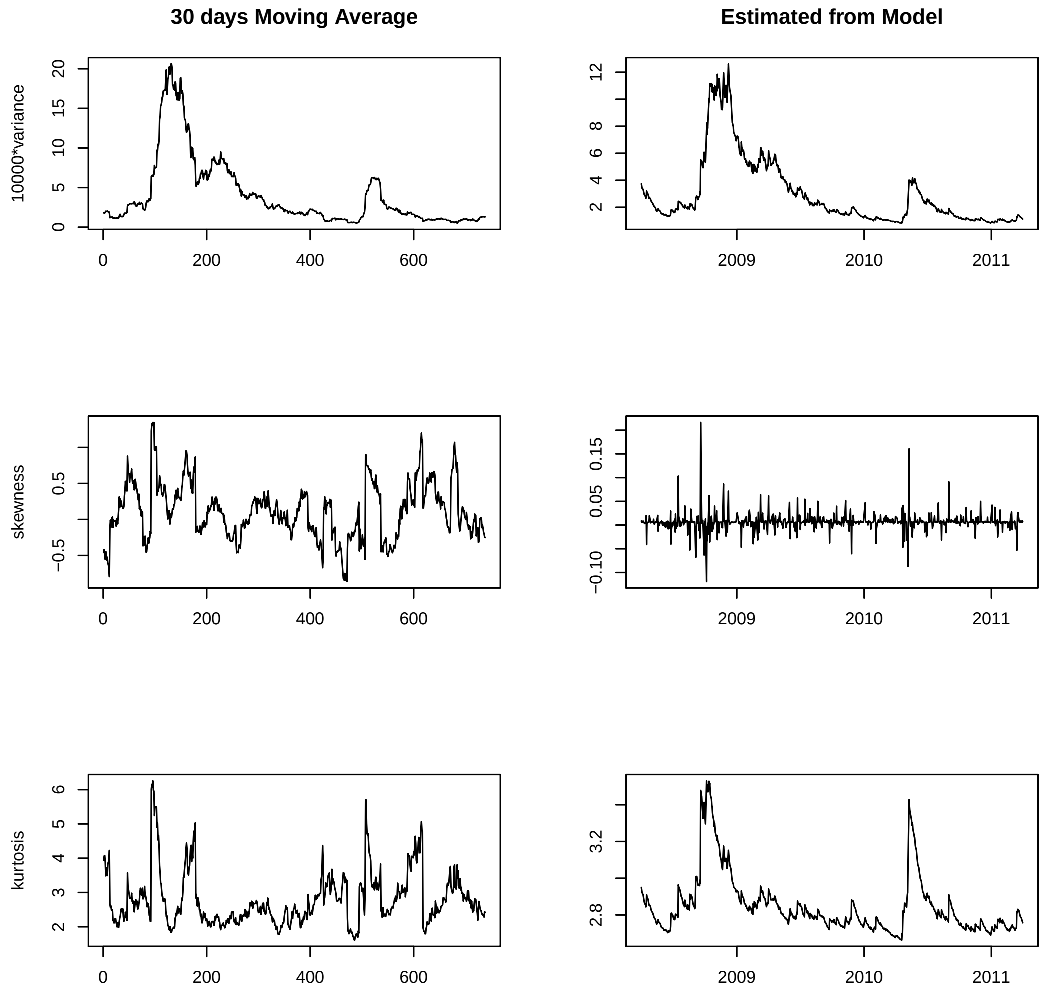
Table 1 gives results for estimation of ARGARCHS(1,1,1) and ARGARCHSK(1,1,1,1) model. $|\hat{\alpha}| < 1$, $\hat{\beta}_1 + \hat{\beta}_2 < 1$, $|\hat{\gamma}_1 + \hat{\gamma}_2| < 1$, $\hat{\delta}_1 + \hat{\delta}_2 < 1$, all estimates satisfy the constraints of basic assumption and stationary condition. Based on the asymptotic theory, all parameters of GARCHS are significant at confidence level of 5% through a simple t test. Compared to GARCHS, 3 parameters of 4 equations in GARCHSK are insignificant at 5% and still 1 parameters are insignificant at 40%. Notice that in GARCHSK, $\hat{\gamma}_2$ is much smaller than other self-regressive coefficients, $\hat{\beta}_2$ and $\hat{\delta}_2$, so the self-regressive impact of conditional skewness in this model may not be obvious. From the results, both $|\hat{\alpha}|$ is quite small, indicating a little effect of autoregressive on returns itself. In this situation, the error terms nearly equal to the returns. Recall the distribution of returns, it can be approximately regarded as the conditional distribution of error terms. More details about it will be discussed in next sector. Large $\hat{\beta}_2$, $\hat{\gamma}_2$ and $\hat{\delta}_2$ approach to 1 explain the highly self-correlation of conditional variance, skewness and kurtosis. When the self-regressive coefficient is small, the change between conditional moment is more or less effected by ϵ_{t-1}.

4.3 Diagnostic Tests

For diagnostics, first we check the properties of the residuals. For daily returns on OMX 40 Nordic Index, the residuals from the model have a skewness of 0.08799 and a kurtosis of 6.17594 approximately to those of returns. And the standardized residuals have a skewness of 0.00277 and a kurtosis of 2.96401.

Then we focus on the fitted value of the conditional variance, skewness and kurtosis. We graph Figure 2 consisting of six figures, in which three figures in first column illustrate the 30 days moving average variance, skewness and kurtosis and the related other three figures in second column illustrate the estimated conditional variance, skewness and kurtosis through the ARGARCHSK(1,1,1,1) model. The comparison in each row demonstrates that the model fitted the trend of conditional moments exactly and capture most of the peaks. In Figure 3 without conditional kurtosis we graph the illustration of ARGARCH(1,1,1) model.

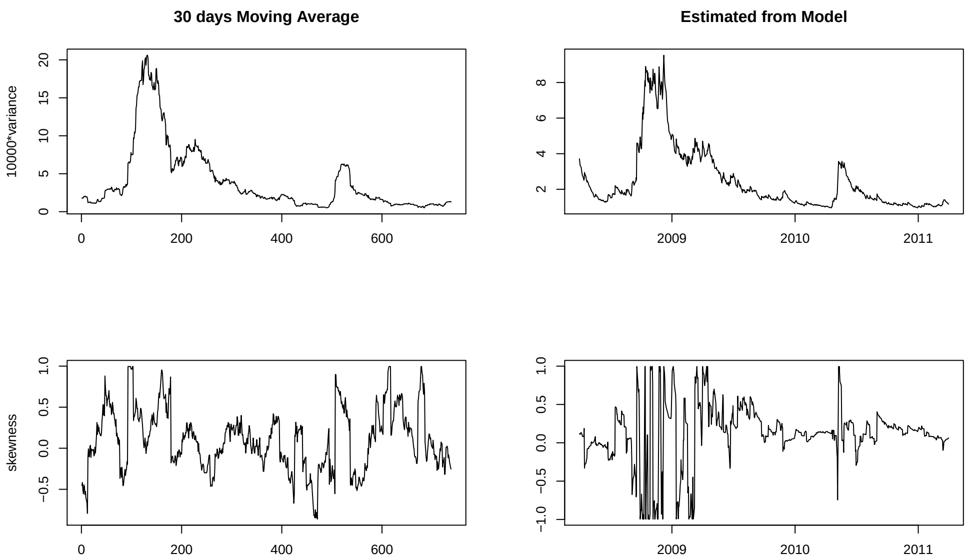
Figure 2



The estimated trend of conditional variance is similar to that of ARGARCHSK, showing a good fitting for Moving Average. However the estimated skewness misses a lot of peaks within the whole period.

To test if the time series model with higher conditional moments is correctly specified, Neywey (1985) and Nelson (1991) introduced a conditional moments test through the standardized residuals $\hat{\eta}_t$ from the estimated model. We construct a Wald test based on a proper set of orthogonality conditions. The Wald statistic follows a χ^2 distribution with the degrees of freedom which equals to the

Figure 3



number of orthogonality conditions. The 13 conditions for the GARCHS need to be examined in this case are as follows:

$$
\begin{aligned}
E[\hat{\eta}_t] &= 0, \\
E[\hat{\eta}_t \hat{\eta}_{t-j}] &= 0 \quad for \quad j = 1, 2, 3, 4, \\
E[(\hat{\eta}_t^2 - 1)(\hat{\eta}_{t-j}^2 - 1)] &= 0 \quad for \quad j = 1, 2, 3, 4, \\
E[(\hat{\eta}_t^3 - s_t)(\hat{\eta}_{t-j}^3 - s_{t-j})] &= 0 \quad for \quad j = 1, 2, 3, 4,
\end{aligned}
$$

The first five conditions consider the specification of conditional mean and the next four examine the conditional variance. $10th$ to $13rd$ conditions are related to the conditional skewness. If the test statistics is significantly different from zero, it shows great evidence that the model capture all of the dynamic features of the conditional moments. After calculation, we get a $\chi^2_{(13)}$ statistic (see Appendix 4) of 14.94. The probability to accept null hypothesis is 0.31 which can not reject the specification of the model at the 5% confidence level. For GARCHSK model the higher moment related to the conditional kurtosis need to be examined as well. Therefore 4 additional conditions are supplemented to the test:

$$E[(\hat{\eta}_t^4 - k_t)(\hat{\eta}_{t-j}^4 - k_{t-j})] \quad = 0 \quad for \quad j = 1, 2, 3, 4.$$

We get a $\chi^2_{(17)}$ statistic of 19.53 of this test. And the probability to accept null hypothesis is 0.30 which can not reject the specification of the model at the 5% confidence level either.

5 Conclusions

This article is based on the approach of GARCH theory and has applied a GARCHS model to allow conditional skewness to financial returns of the stock market. The primary GARCH model and one additional equation for skewness is nested to the new model. For conditional kurtosis, a more general ARGARCHSK model with another additional equation for kurtosis works in this case. Both models are estimated in a maximum likelihood framework, assuming a conditional distribution of skew normal distribution and Gram-Charlier expansion density. We fit the models to daily returns on the OMX Nordic 40 exchange index and the results prove a strong evidence of the existence of conditional skewness and kurtosis. In diagnostic, though GARCHS model gets larger proportion of significant coefficients than those of GARCHSK model, the GARCHSK model fits data better than GARCHS. The estimated conditional variance, skewness and kurtosis of GARCHSK show a similar trend to the those of the sample which shows a good property of the model to capture dynamic changes on returns.

References

[1] Campbell R. Harvey and Akhtar Siddique. Autoregressive Conditional Skewness. *Journal of Financial and Quantitative Analysis*, Vol. 34, No. 4, Dec 1999.

[2] Chris Brooks, Simon P. Burke, Saeed Heram, and Gita Persand. Autoregressive Conditional Kurtosis. *Journal of Financial Econometrics*, Vol. 3, No.3, 2005.

[3] Jonathan Dark. Time varying skewness and kurtosis and a new model. Department of Econometrics and Business Statistics, Monash University, Australia.

[4] Angel Leon, Gonzalo Rubio, and Gregorio Serna. Autoregressive Conditional Volatility, Skewness And Kurtosis. *Working Papers. Serie AD*, 2004-13.

[5] Eric Jondeau and Michael Rockinger. Gram-Charlier densities. *Journal of Economic Dynamics & Control*, Vol. 25, 2001.

[6] A. Ronald Gallant and George Tauchen. Seminonparametric Estimation of Conditionally Constrained Heterogeneous Processes: Asset Pricing Applications. *Econometrica*, Vol. 57, No. 5, Sep 1989.

[7] Whitney K. Newey. Maximum Likelihood Specification Testing and Conditional Moment Tests. *Econometrica*, Vol. 53, No. 5, Sep 1985.

[8] Peter A. Abken, Dilip B. Madan, and Sailesh Ramamurtie. Estimation of Risk-Neutral and Statistical Densities by Hermite Polynomial Approximation: With An Application to Eurodollar Futures Options. *Federal Reserve Bank of Atlanta Working Paper*, 96-5, Jun 1996.

[9] Daniel B. Nelson. Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica*, Vol. 59, No. 2, Mar 1991.

[10] David M. Durkker. Bootstrapping A Conditional Moments Test for Normality after Tobit Estimation. *The Stata Journal*, Vol. 2, No. 2, 2002.

[11] James D. Hamilton. *Time Series Analysis*. Princeton University Press, 1st edition, Jan 1994.

[12] Tim Bollerslev. Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, Vol. 31, 1986.

[13] Bruce E. Hansen. Autoregressive Conditional Density Estimation. *International Economic Review*, Vol. 35, No. 3, Aug 1994.

Appendix 1

Denote ξ as location parameter, w as scale parameter and α as skew parameter, a skew normal distribution, $SN(\xi, w, \alpha)$, is expressed as

$$f(x) = \frac{2}{w}\phi\left(\frac{x-\xi}{w}\right)\Phi\left(\alpha\left(\frac{x-\xi}{w}\right)\right)$$

where ϕ and Φ are the probability density function and cumulative density function of standard normal.

Variance and skewness of the distribution are

$$h = w^2\left(1 - \frac{2\delta^2}{\Pi}\right)$$

$$s = \frac{4-\Pi}{2}\frac{(\delta\sqrt{2/\Pi})^3}{(1-2\delta^2/\Pi)^{3/2}}.$$

where

$$\delta = \frac{\alpha}{\sqrt{1+\alpha^2}}.$$

The skew normal distribution is right skewed if $\alpha > 0$ and left skewed if $\alpha < 0$. Notice that when $\alpha = 1$, skewness reaches to its maximum value of approximately 0.9952717. Therefore we usually adjust the skewness in the estimation as

$$s_{adjust} = min(0.995, s).$$

Appendix 2

According to the method of Eric and Micheal (1999), if the distribution of a random variable η is believed similar to a normal one but probability distribution function (pdf) unknown, a proper way to approximate it is to express the pdf as the form

$$g(\eta) = p_n(\eta)\phi(\eta),$$

where $\phi(\eta)$ is the standard zero mean and unit variance normal density and $p_n(\eta)$ is chosen to guarantee the same first moments as the distribution of η. Abken and Madan (1996) described $\phi(\eta)$ as the reference measure density and $p_n(\eta)$ as the change of measure density. Given the assumption and that of Gaussian random process generating uncertainty, the change of measure could be constructed by Hermite polynomials:

$$p_n(\eta) = \sum_{i=0}^{n} c_i He_i(\eta),$$

where $He_i(\eta)$ are the Hermite polynomials. These polynomials are defined in terms of the normal density as

$$He_i(\eta) = (-1)^i \frac{\partial^i \phi(\eta)}{\partial \eta^i} \phi(\eta)^{-1}.$$

The first several Hermite polynomials are computed as the following expression:

$$\begin{aligned}
He_0(\eta) &= 1 \\
He_1(\eta) &= \eta \\
He_2(\eta) &= \eta^2 - 1 \\
He_3(\eta) &= \eta^3 - 3\eta \\
He_4(\eta) &= \eta^4 - 6\eta^2 + 3 \\
\ldots &
\end{aligned}$$

When η is standardized with zero mean and unit variance, there is a generally used representation:

$$p_4(\eta) = 1 + \frac{C_1}{6} He_3(\eta) + \frac{C_2}{24} He_4(\eta)$$

This is the Gram-Charlier type-A expansion. Eric and Machael (1999) proved that $C_1$and $C_2$are correspond, respectively, to the skewness and kurtosis of $g(\eta)$. $C_1 = \int_{-\infty}^{+\infty} \eta^3 g(\eta) d\eta = s$ and $C_2 = \int_{-\infty}^{+\infty} \eta^4 g(\eta) d\eta - 3 = k - 3$. Thus pdf of η can be written as

$$g(\eta) = p_4(\eta)\phi(\eta) = [1 + \frac{s}{6}(\eta^3 - 3\eta) + \frac{k-3}{24}(\eta^4 - 6\eta^2 + 3)]\phi(\eta)$$

However in some case this pdf is not a real pdf because for some parameters the value of $g(\eta)$ might be negative without constrain of $p_4(\eta)$. To insure positivity, Gallant and Tauchen (1989) square the polynomial part. And to insure the density integrates to one they devided by the integral over $\mathbb{R}$. The improved pdf is

$$f(\eta) = p_4^2(\eta)\phi(\eta) / \int_{\mathbb{R}} p_4^2(z)\phi(z) dz$$

Based on the properties of Hermite polynomials, it can be proved that $\int_{\mathbb{R}} p_4^2(z)\phi(z) dz = 1 + \frac{s^2}{6} + \frac{(k-3)^2}{24}$. Therefore we get the common expression of $f(\eta)$

$$f(\eta) = \phi(\eta)\psi(\eta)^2 / \Gamma$$

where

$$\psi(\eta) = 1 + \frac{s}{6}(\eta^3 - 3\eta) + \frac{k-3}{24}(\eta^4 - 6\eta^2 + 3),$$

$$\Gamma = 1 + \frac{s^2}{6} + \frac{(k-3)^2}{24}.$$

Appendix 3

Since GARCHSK process is more complicate than GARCHS, only estimation of GARCHSK model is showed here as an demonstration: Let $z'_{1t} = (1, \epsilon^2_{t-1}, h_{t-1})$, $z'_{2t} = (1, \eta^3_{t-1}, s_{t-1})$, $z'_{3t} = (1, \eta^4_{t-1}, k_{t-1})$, $\beta' = (\beta_0, \beta_1, \beta_2)$, $\gamma' = (\gamma_0, \gamma_1, \gamma_2)$, $\delta' = (\delta_0, \delta_1, \delta_2)$ and $\theta \in \Theta$, where $\theta = (\alpha', \beta', \gamma', \delta')$ and Θ is a compact Euclidean subspace to garantee finite second, third and fourth moments of ϵ_t and η_t. Denote true parameters by θ_0, where $\theta_0 \in \Theta$.

Rewrite the ARGARCHSK(1,1,1,1) as

$$\begin{aligned} \epsilon_t &= r_t - \alpha r_{t-1} \\ h_t &= z'_{1t}\beta \\ s_t &= z'_{2t}\gamma \\ k_t &= z'_{3t}\delta \end{aligned}$$

Since $\eta_t = h_t^{-\frac{1}{2}}\epsilon_t$, then the pdf of ϵ_t is $f(\epsilon_t|I_{t-1}) = h_t^{\frac{1}{2}} f(\eta_t|I_{t-1})$. Therefore, the log likelihood function for a sample of T observations is, apart from some constant,

$$L_T(\theta) = T^{-1}\sum_{t=1}^{T} l_t(\theta),$$

$$l_t(\theta) = -\frac{1}{2}logh_t - \frac{1}{2}\eta_t^2 + log(\psi^2(\eta_t)) - ln\Gamma_t$$

Differentiating with respect to the mean parameters yields

$$
\begin{aligned}
\frac{\partial \epsilon_t}{\partial \alpha} &= -r_{t-1} \\
\frac{\partial h_t}{\partial \alpha} &= 2\beta_1 \epsilon_{t-1} \frac{\partial \epsilon_{t-1}}{\partial \alpha} + \beta_2 \frac{\partial h_{t-1}}{\partial \alpha} \\
\frac{\partial \eta_t}{\partial \alpha} &= \frac{\partial (\epsilon_t h_t^{-\frac{1}{2}})}{\partial \alpha} = h_t^{-\frac{1}{2}} \frac{\partial \epsilon_t}{\partial \alpha} - \frac{1}{2} \epsilon_t h_t^{-\frac{3}{2}} \frac{\partial h_t}{\partial \alpha} \\
\frac{\partial s_t}{\partial \alpha} &= 3\gamma_1 \eta_{t-1}^2 \frac{\partial \eta_{t-1}}{\partial \alpha} + \gamma_2 \frac{\partial s_{t-1}}{\partial \alpha} \\
\frac{\partial k_t}{\partial \alpha} &= 4\delta_1 \eta_{t-1}^3 \frac{\partial \eta_{t-1}}{\partial \alpha} + \delta_2 \frac{\partial k_{t-1}}{\partial \alpha} \\
\frac{\partial \psi_t}{\partial \alpha} &= \frac{\eta_t^3 - 3\eta_t}{6} \frac{\partial s_t}{\partial \alpha} + s_t \frac{3\eta_t^2 - 3}{6} \frac{\partial \eta_t}{\partial \alpha} + \frac{\eta_t^4 - 6\eta_t^2 + 3}{24} \frac{\partial k_t}{\partial \alpha} + (k_t - 3) \frac{\eta_t^3 - 3\eta_t}{6} \frac{\partial \eta_t}{\partial \alpha} \\
\frac{\partial \Gamma_t}{\partial \alpha} &= \frac{s_t}{3} \frac{\partial s_t}{\partial \alpha} + \frac{k_t - 3}{12} \frac{\partial k_t}{\partial \alpha}
\end{aligned}
$$

Thus

$$
\frac{\partial l_t}{\partial \alpha} = -\frac{h_t^{-1}}{2} \frac{\partial h_t}{\partial \alpha} - \eta_t \frac{\partial \eta_t}{\partial \alpha} + 2\psi_t^{-1} \frac{\partial \psi_t}{\partial \alpha} - \Gamma_t^{-1} \frac{\partial \Gamma_t}{\partial \alpha}
$$

Then differentiating with respect to the variance parameters yields

$$
\begin{aligned}
\frac{\partial h_t}{\partial \beta} &= z_{1t} + \beta_2 \frac{\partial h_{t-1}}{\partial \beta} \\
\frac{\partial \eta_t}{\partial \beta} &= \frac{\partial (\epsilon_t h_t^{-\frac{1}{2}})}{\partial \beta} = -\frac{1}{2} \epsilon_t h_t^{-\frac{3}{2}} \frac{\partial h_t}{\partial \beta} \\
\frac{\partial s_t}{\partial \beta} &= 3\gamma_1 \eta_{t-1}^2 \frac{\partial \eta_{t-1}}{\partial \beta} + \gamma_2 \frac{\partial s_{t-1}}{\partial \beta} \\
\frac{\partial k_t}{\partial \beta} &= 4\delta_1 \eta_{t-1}^3 \frac{\partial \eta_{t-1}}{\partial \beta} + \delta_2 \frac{\partial k_{t-1}}{\partial \beta} \\
\frac{\partial \psi_t}{\partial \beta} &= s_t(\eta_t^2 - 1) \frac{\partial \eta_t}{\partial \beta} + \frac{\eta_t^3 - 3\eta_t}{6} \frac{\partial s_t}{\partial \beta} + (k_t - 3) \frac{\eta_t^3 - 3\eta_t}{6} \frac{\partial \eta_t}{\partial \beta} + \frac{\eta_t^4 - 6\eta_t^2 + 3}{24} \frac{\partial k_t}{\partial \beta} \\
\frac{\partial \Gamma_t}{\partial \beta} &= \frac{s_t}{3} \frac{\partial s_t}{\partial \beta} + \frac{k_t - 3}{12} \frac{\partial k_t}{\partial \beta}
\end{aligned}
$$

Thus we get the log likelihood differential with β

$$\frac{\partial l_t}{\partial \beta} = -\frac{h_t^{-1}}{2}\frac{\partial h_t}{\partial \beta} - \eta_t \frac{\partial \eta_t}{\partial \beta} + 2\psi_t^{-1}\frac{\partial \psi_t}{\partial \beta} - \Gamma_t^{-1}\frac{\partial \Gamma_t}{\partial \beta}$$

Differentiating with respect to the skewness parameters yields

$$\begin{aligned}
\frac{\partial s_t}{\partial \gamma} &= z_{2t} + \gamma_2 \frac{\partial s_{t-1}}{\partial \gamma} \\
\frac{\partial \psi_t}{\partial \gamma} &= \frac{\eta_t^3 - 3\eta_t}{6}\frac{\partial s_t}{\partial \gamma} \\
\frac{\partial \Gamma_t}{\partial \gamma} &= \frac{s_t}{3}\frac{\partial s_t}{\partial \gamma}
\end{aligned}$$

Thus

$$\frac{\partial l_t}{\partial \gamma} = 2\psi_t^{-1}\frac{\partial \psi_t}{\partial \gamma} - \Gamma_t^{-1}\frac{\partial \Gamma_t}{\partial \gamma}$$

Differentiating with respect to the kurtosis parameters yields

$$\begin{aligned}
\frac{\partial k_t}{\partial \delta} &= z_{3t} + \delta_2 \frac{\partial k_{t-1}}{\partial \delta} \\
\frac{\partial \psi_t}{\partial \delta} &= \frac{\eta_t^4 - 6\eta_t^2 + 3}{24}\frac{\partial k_t}{\partial \delta} \\
\frac{\partial \Gamma_t}{\partial \delta} &= \frac{k_t - 3}{12}\frac{\partial k_t}{\partial \delta}
\end{aligned}$$

Thus

$$\frac{\partial l_t}{\partial \delta} = 2\psi_t^{-1}\frac{\partial \psi_t}{\partial \delta} - \Gamma_t^{-1}\frac{\partial \Gamma_t}{\partial \delta}$$

Then the first order derivative of log-likelihood function is

$$\frac{\partial l_t}{\partial \theta} = (\frac{\partial l_t}{\partial \alpha}, \frac{\partial l_t}{\partial \beta}, \frac{\partial l_t}{\partial \gamma}, \frac{\partial l_t}{\partial \delta})$$

Then take BHHH recursion

$$\theta^{(i+1)} = \theta^{(i)} + \lambda_i \left(\sum_{t=1}^{T} \frac{\partial l_t}{\partial \theta}|_{\theta=\theta^{(i)}} \frac{\partial l_t}{\partial \theta'}|_{\theta=\theta^{(i)}} \right) \sum_{t=1}^{T} \frac{\partial l_t}{\partial \theta}|_{\theta=\theta^{(i)}}$$

Appendix 4

Neway (1985) and Tauchen (1985) found such a matrix, denoted as $\hat{Q}^{-1}$, so that the Wald statistic

$$W = L'\widehat{M}\widehat{Q}^{-1}\widehat{M}'L \xrightarrow{d} \chi^2(r)$$

where L is an $(N \times 1)$ vector of ones, $\widehat{M}$ is the $(N \times r)$ matrix of sample realization of the r moment restrictions, and $\widehat{Q}^{-1}$ is a weighting matrix that scales the inner product of sample averages $L'\widehat{M}$.

In this paper, if the ith moment restriction under hypothesis of model parameters of θ is $E[g_i(\hat{\eta}_t)] = 0$, where $g_i(\hat{\eta}_t)$ is a function of any order moment of $\hat{\eta}_t$, then the jth sample realization on this moment is

$$m_{i,j}(\theta) = g_i(\hat{\eta}_j).$$

Therefore we take the realization matrix, $\widehat{M}$, with dimension of $(T \times r)$.

Let $f(r_t;\theta)$ be the contribution of observation t to the log likelihood, $S(\theta)$ be the score matrix and $H(\theta)$ the average information matrix:

$$S(\theta) = \sum_{t=1}^{T} \frac{\partial f(r_t;\theta)}{\partial \theta}$$

$$H(\theta) = \frac{1}{T}\sum_{t=1}^{T} E\left[\frac{\partial^2 f(r_t;\theta)}{\partial \theta \partial \theta'}\right].$$

Specifically, choose the consistent estimators

$$\widehat{W} = T^{-1}\widehat{S}'\widehat{M}$$

and

$$\widehat{H} = T^{-1}\widehat{S}'\widehat{S},$$

then $\widehat{Q}$ is expressed as

$$\widehat{Q} = \left(\widehat{M} - \widehat{S}\widehat{H}^{-1}\widehat{W}\right)'\left(\widehat{M} - \widehat{S}\widehat{H}^{-1}\widehat{W}\right).$$

Supplementary Table

Summary Statistics for Monthly returns

Month	Mean$\times 10^4$	Var$\times 10^4$	Skewness	Kurtosis
Apr08	-9.51003	2.30518	-0.33667	3.27093
May08	7.38010	1.10199	-0.39267	2.81211
Jun08	-72.26921	1.48543	0.47184	2.74732
Jul08	7.36835	2.90394	0.48782	2.74732
Aug08	2.72098	2.54393	-0.26318	2.87204
Sep08	-86.92392	9.35501	1.17889	5.16340
Oct08	-71.73474	20.01259	0.06616	1.99330
Nov08	-31.34573	16.65568	0.43137	2.67429
Dec08	-19.92064	10.96391	0.61411	4.33128
Jan09	-19.88948	7.09344	-0.14195	2.28927
Feb09	-48.77096	6.56000	0.18771	2.18517
Mar09	22.41504	9.01366	-0.00924	1.92635
Apr09	95.56487	7.16806	-0.29331	2.24715
May09	28.76205	4.12669	0.09218	2.00069
Jun09	-5.81535	4.40558	0.03819	2.39886
Jul09	33.52271	2.55191	0.21006	2.82044
Aug09	5.37884	2.88003	0.27038	1.81868
Sep09	-1.0941	1.74647	-0.36491	2.33056
Oct09	2.42404	1.63701	0.31614	2.14546
Nov09	9.15637	2.75781	-0.21746	2.17132
Dec09	20.56920	0.89460	0.34509	2.72869
Jan10	10.79023	0.77489	-0.04531	2.40953
Feb10	10.32572	1.09183	-0.70700	2.90462
Mar10	37.41278	0.56086	-0.11913	1.64519
Apr10	6.75699	1.69626	-0.22654	2.60682
May10	-35.01033	7.99995	0.60776	2.85811
Jun10	-12.06817	2.70315	-0.35023	1.91272
Jul10	48.29743	1.69496	0.30410	2.40957
Aug10	-8.74808	2.00498	0.49102	3.71225
Sep10	16.45150	0.65799	0.22357	1.90994
Oct10	4.91681	1.01595	0.53207	2.18763
Nov10	16.91914	1.18375	0.13806	2.45243
Dec10	31.26393	0.48113	0.62450	3.11096
Jan11	-2.91451	1.01273	0.10636	2.75883
Feb11	-10.20461	0.82864	-0.08878	2.43587
Mar11	-0.79266	1.34252	-0.39424	2.49379
Total	-0.39022	3.89235	0.08984	6.18477