Kurtosis of GARCH and Stochastic Volatility Models with Non-normal Innovations

Xuezheng Bai  Jeffrey R. Russell  George C. Tiao
Graduate School of Business
University of Chicago
July 27, 2001

Abstract

Both volatility clustering and conditional non-normality can induce the leptokurtosis typically observed in financial data. In this paper, the exact representation of kurtosis is derived for both GARCH and stochastic volatility models when innovations may be conditionally non-normal. We find that, for both models, the volatility clustering and non-normality contribute interactively and symmetrically to the overall kurtosis of the series.
1 Introduction

Many financial series, such as returns on stocks and foreign exchange rates, exhibit leptokurtosis and time-varying volatility. These two features have been the subject of extensive studies ever since Mandelbrot (1963) and Fama (1965) first reported them.

The Autoregressive Conditional Heteroscedastic (ARCH) Model by Engle (1982) and its generalization, the GARCH model by Bollerslev (1986), provide a convenient framework to study time-varying volatility in financial markets. In practice, a common assumption in applying GARCH models to financial data is that the return series is conditionally normally distributed. We shall refer to this as the normal GARCH model.

It is well known that the normal GARCH model is consistent with volatility clustering pattern. However, the kurtosis implied by the normal GARCH model tends to be far less than the the sample kurtosis observed for most financial return series. For example, Bollerslev (1987) finds evidence of conditional leptokurtosis in monthly S&P 500 Composite Index returns and advocates use of the t-distribution, Hong (1988) rejects the normality assumption for daily NYSE stock returns. Thus, the normal GARCH model is inconsistent with the large leptokurtosis typically observed in asset returns. The stochastic volatility model can also be used to model volatility clustering, see Polson et al. (1994). Like the normal GARCH model, the conditional normal stochastic volatility model also generates leptokurtosis but often not sufficiently large to explain the sample kurtosis.

This paper derives expressions for the kurtosis of GARCH and stochastic volatility models in the presence of conditionally non-normal leptokurtic innovations. Previously, He and Terasvirta (1999b) examined the forth moment structure of the GARCH(1, 1) model with conditionally non-normal innovations and He and Terasvirta (1999a) extend their results to the GARCH(p, q) model. In both of these papers, the kurtosis is expressed as a function of the underlying model parameters. We take a somewhat different approach in this paper. By working with the well known ARMA representations of the squares of the error term we are able to extend their results to a broader class of models that include stochastic volatility models. Second, working with the ARMA representation provides an interesting perspective on the interaction between the volatility clustering and the non-normality in the innovation. In particular, we decompose the overall kurtosis into the kurtosis induced by volatility clustering and the kurtosis induced by conditionally non-normal innovations. We find the interesting result that both of these kurtosis enter symmetrically and interactively to
determine the overall kurtosis of the process. Additionally, working with the ARMA representations greatly simplifies the derivations.

This paper is organized as follows. We present the results for GARCH model in Sections 2 and 3 with illustrative examples. In Section 4, we study the kurtosis of linear stochastic volatility models. Section 5 summarizes our findings.

2 The kurtosis of GARCH models with non-normal innovations

Consider a GARCH($p,q$) model for the time series $\epsilon_t$,

\[
\begin{align*}
\epsilon_t &= \sqrt{h_t}z_t \\
h_t &= \omega + \sum_{i=1}^p \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^q \beta_j h_{t-j}
\end{align*}
\]  

where $h_t$ is the conditional variance of $\epsilon_t$ given $V_t = \epsilon_{t-1}, \epsilon_{t-2}, \ldots$, and $z_t$ are iid with mean 0 and variance 1. The parameters $(\omega, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)$ are restricted such that $h_t > 0$ for all $t$. We shall assume that the fourth moment of $z_t$ exists. This section derives the kurtosis of $\epsilon_t$ when it exists.

In the following, we call the kurtosis (or the excess kurtosis in the econometric literature) of $z_t$ the iid kurtosis and denote it by $\gamma_2(z)$, and call the kurtosis of $\epsilon_t$ the overall kurtosis and denote it by $\gamma_2(\epsilon)$ if it exists.

Rewriting the equation for $h_t$ in (1) with $u_t = \epsilon_t^2 - h_t$, we have the following well-known ARMA($r,q$) representation for $\epsilon_t^2$:

\[
\phi(B)\epsilon_t^2 = \omega + \beta(B)u_t
\]  

where $B$ is the backward shift operator, $\phi(B) = 1 - \sum_{i=1}^r \phi_i B^i, \phi_i = (\alpha_i + \beta_i)$, $r = \text{max}(p,q)$, and $\beta(B) = 1 - \sum_{i=1}^q \beta_i B^i$. It is understood that $\alpha_i = 0$ for $i > p$ and $\beta_i = 0$ for $i > q$. We shall make the following two further assumptions:

A.1. all the zeroes of the polynomial $\phi(B)$ are lying outside of the unit circle

A.2. $0 < (\gamma_2(z) + 2)a_1 < 1$, with $a_1 = \sum_{i=1}^\infty \psi_i^2$

where the $\psi_i$’s are obtained from the relation $\psi(B) \phi(B) = \beta(B)$ with $\psi(B) = 1 + \sum_{i=1}^\infty \psi_i B^i$. These two assumptions ensure that the $u_t^i$’s are uncorrelated with zero mean and finite variance and that
the $\epsilon_t^2$ process is weakly stationary. In this case, the autocorrelation function of $\epsilon_t^2$ will be exactly the same as that for a stationary ARMA($p$, $q$) model.

If $z_t$ follows a normal distribution, $\gamma_2^{(z)} = 0$ and the process defined by Equation (1) is the normal GARCH($p$, $q$) process. The kurtosis of this normal GARCH process is called the GARCH kurtosis, and is denoted by $\gamma_2^{(g)}$ when it exists.

To calculate the GARCH kurtosis $\gamma_2^{(g)}$ and to understand the relation of the overall kurtosis $\gamma_2^{(e)}$ with volatility clustering and conditional non-normality, we have the following theorem:

**Theorem 2.1** If $\epsilon_t$ follows the GARCH($p$, $q$) process specified by (1) and satisfies Assumptions A.1 and A.2, then we have

(a) $\gamma_2^{(g)} = \frac{6a_1}{1 - 2a_1}$

(b) $\gamma_2^{(e)} = \frac{\gamma_2^{(g)} + \gamma_2^{(z)} + \gamma_2^{(g)} \gamma_2^{(z)}}{1 - \frac{1}{6} \gamma_2^{(g)} \gamma_2^{(z)}}$.

**Proof:**

Assuming $E(h_t)$ and $E(h_t^2)$ exist, the GARCH model in Equation (1) implies,

$E(\epsilon_t^2) = E(h_t)$

$E(\epsilon_t^4) = E(h_t^2 z_t^4) = (\gamma_2^{(z)} + 3)E(h_t^2)$

$var(\epsilon_t^2) = E(\epsilon_t^4) - [E(\epsilon_t^2)]^2 = (\gamma_2^{(z)} + 3)E(h_t^2) - [E(h_t)]^2$

and from the definition of $u_t$,

$E(u_t) = 0$

$E(\epsilon_t^4) = E(h_t^2 z_t^4) = (\gamma_2^{(z)} + 3)E(h_t^2) - [E(h_t)]^2$

$var(u_t) = var(\epsilon_t^2 - h_t) = E(\epsilon_t^4 - h_t^2) = (\gamma_2^{(z)} + 2)E(h_t^2)$

From the ARMA representation (2), we have, under assumption A.1,

$E(\epsilon_t^2) = E(h_t) = \omega/(1 - \phi_1 - \ldots - \phi_p)$

and, again assuming that $E(h_t^2)$ exists,

$var(\epsilon_t^2) = var(u_t)(1 + a_1) = (\gamma_2^{(z)} + 2)E(h_t^2)(1 + a_1)$
Equating the extreme right hand side of (5) to that of (10) for the two expressions of \( \text{var}(\epsilon_t^2) \), and making use of the expression for \( E(h_t) \) in (9), we obtain, under assumption A.2, the following result for \( E(h_t^2) \)

\[
E(h_t^2) = \frac{\omega / (1 - \phi_1 - \cdots - \phi_p)^2}{1 - (\gamma_2^{(z)} + 2)a_1}
\]

Now the kurtosis of \( \epsilon_t \) is,

\[
\gamma_2^{(c)} = \frac{E(\epsilon_t^4)}{[E(\epsilon_t^2)]^2} - 3 = \frac{(\gamma_2^{(z)} + 3)E(h_t^2)}{[E(h_t)]^2} - 3
\]

\[
= \frac{6a_1 + \gamma_2^{(z)}(1 + 3a_1)}{1 - (\gamma_2^{(z)} + 2)a_1}
\]

(11)

Part (a) of the theorem follows from Equation (11) by letting \( \gamma_2^{(z)} = 0 \). Now from part (a) we can express \( a_1 \) in terms of \( \gamma_2^{(g)} \) as

\[
a_1 = \frac{\gamma_2^{(g)}}{2(3 + \gamma_2^{(g)})}
\]

(12)

Substituting the right hand side of (12) for \( a_1 \) in (11), we obtain part (b) of the theorem.

### 3 Discussion and Examples

We now discuss some interesting issues stemmed from Theorem 2.1. In the Theorem, part (a) relates the normal GARCH kurtosis \( \gamma_2^{(g)} \) to the GARCH parameters \( \alpha 's \) and \( \beta 's \) in (1) characterizing the volatility clustering. Specifically, it confirms the well-known fact that volatility clustering introduces leptokurtosis. For the normal GARCH(1,1) model, the expression reduces to the result in Bollerslev (1986), i.e.,

\[
\gamma_2^{(g)} = \frac{6\alpha^2}{1 - \phi^2 - 2\alpha^2}
\]

(13)

where \( \phi^2 + 2\alpha^2 < 1 \), and \( \phi = \alpha + \beta \). On the other hand from (1) and (2), and under conditions A.1 and A.2, the autocorrelation function of \( \epsilon_t^2 \) for a GARCH(1, 1) model is

\[
corr(\epsilon_t^2, \epsilon_{t-\ell}^2) = \rho_1 \cdot \phi^{\ell-1}
\]

(14)

where \( \rho_1 \) is the lag-one autocorrelation (\( \ell = 1 \)) and is related to \( \phi \) and \( \alpha \) as

\[
\rho_1 = \alpha \cdot \frac{1 - \phi^2 + \phi \cdot \alpha}{1 - \phi^2 + \alpha^2} = \alpha (1 + \alpha \cdot \frac{\phi - \alpha}{1 - \phi^2 + \alpha^2})
\]

(15)
Thus, $\rho_1$ determines the lag-one value and $\phi$ the rate of decay of the autocorrelation function. For financial returns data, the sample autocorrelations are generally positive and persistent, and $\rho_1$ is far below 1 with values near .2 common. This phenomenon corresponds to the case $\rho_1$ is about .2 and $\phi$ is large, often close to but less than 1. Using Equation (15) we can express $\gamma_2^{(g)}$ in Equation (13) in terms of $\rho_1$ and $\phi$. Figure 1 shows the contours of $\gamma_2^{(g)}$ in the region $0.00 < \rho_1 < .50$ and $.70 < \phi < 1.00$. Consider the region of $0.10 < \rho_1 < .25$ and $0.85 < \phi < 0.99$, in which most financial returns locate. In this region, the implied kurtosis $\gamma_2^{(g)}$ values between 0.15 and 2.85, which are substantially below the sample kurtosis commonly found in the returns data. See the empirical examples shown below. It is well known that under fairly general conditions the parameters of the volatility dynamics can be consistently estimated even when the wrong distribution has been assumed for $z_t$, see Newey and Steigerwald (1997). Thus, MLE estimates of the normal GARCH(1, 1) model frequently imply autocorrelations of $\epsilon_t^2$ that match closely the sample autocorrelation function, but is not able to match the large leptokurtosis typically found in the data. This led Bollerslev (1987) to suggest use t-distribution to match the excessive sample kurtosis and Engle and Gonzalez-Rivera (1991) proposing a semi-parametric model with non-parametric density for iid innovations $z_t$. It is simply not possible for the normal GARCH(1, 1) model to match both volatility dynamics and kurtosis. A similar plot and conclusion are given in Terasvirta (1996).

Now, part (b) of Theorem 2.1 shows that the overall kurtosis $\gamma_2^{(e)}$ can be expressed as a function of the kurtosis $\gamma_2^{(g)}$ induced by time varying volatility, and the kurtosis $\gamma_2^{(z)}$ of $z_t$. Specifically, part (b) suggests that $\gamma_2^{(g)}$ and $\gamma_2^{(z)}$ contribute interactively and symmetrically to the increase of the overall kurtosis. To further understand the nonlinear effect, let’s consider the partial derivative of $\gamma_2^{(e)}$ with respect to $\gamma_2^{(g)}$:

$$
\frac{\partial \gamma_2^{(e)}}{\partial \gamma_2^{(g)}} = \frac{1 + \frac{5}{6} \gamma_2^{(z)} + \frac{1}{6} [\gamma_2(z)]^2}{(1 - \frac{1}{6} \gamma_2^{(g)} \gamma_2^{(z)})^2}
$$

We first notice that the overall kurtosis $\gamma_2^{(e)}$ is increasing with $\gamma_2^{(g)}$ from the positivity of the partial derivative $\frac{\partial \gamma_2^{(e)}}{\partial \gamma_2^{(g)}}$. Furthermore, we see that the partial derivative $\frac{\partial \gamma_2^{(e)}}{\partial \gamma_2^{(g)}}$ increases with $\gamma_2^{(g)}$ through the denominator but is greatly exacerbated by $\gamma_2^{(z)}$. Similar conclusions can be drawn by considering the partial derivative of $\gamma_2^{(e)}$ with respect to $\gamma_2^{(z)}$.

Figure (2) shows the contours of $\gamma_2^{(e)}$ as a function of $\gamma_2^{(g)}$ and $\gamma_2^{(z)}$. 

5
Figure 1: Contour Plot of Normal GARCH Kurtosis $\gamma_2^{(g)}$
Figure 2: Contour Plot of the overall Kurtosis $\gamma_2^{(e)}$
From the above plot, we see that for each fixed value of $\gamma_{2}^{(g)}$, leptokurtic departure from normality $\gamma_{2}^{(z)}$ contributes greatly to the non-linear increase in the overall kurtosis. Typical parameter estimates will produce kurtosis in the upper left corner of the plot. This will be the case for the data we examine below.

To conclude this section, we consider two examples that illustrate the effects of conditional leptokurtosis and volatility clustering on the overall kurtosis. The first series we consider is the IBM daily returns ($IBM_{t}$) from January 2, 1990 to December 29, 2000.\footnote{The IBM daily returns are obtained from CRSP. For estimation convenience, the series are standardized to have unit sample variance.} This series shows little dynamic dependence in the level but strong volatility clustering as shown by the sample autocorrelation function of $IBM_{t}^{2}$ in Figure 3, and its sample kurtosis is 6.5306. The second series consists of half-hour foreign exchange rate returns of the Dollar/Deutsch Mark (DD) for the one year period January to December, 1996.\footnote{This data set is obtained from Olsen and Associates. This series shows the usual characteristics of high frequency data, i.e., deterministic intra-day volatility pattern and an MA(1) structure in the conditional mean. For a detailed description of the data and consequent analysis, see Bai et al. (2001).} For simplicity, we concentrate on the volatility clustering and leptokurtosis of DD. Here we use the adjusted series $\tilde{DD}$ after accounting for both the deterministic intra-day volatility pattern and an MA(1) structure as in Bai et al. (2001). The sample kurtosis for the adjusted return series is 8.1551.

We model these two series using a normal GARCH(1, 1) model

\[
\begin{align*}
    y_{t} &= \sqrt{h_{t}}z_{t}, \quad z_{t} \sim \text{iid } N(0, 1) \\
    h_{t} &= \omega + \alpha \epsilon_{t-1}^{2} + \beta h_{t-1}
\end{align*}
\]

where $y_{t}$ can be taken as either the IBM daily returns $IBM_{t}$ or the adjusted $\tilde{DD}$ series. The maximum likelihood estimation results are given in Table 1:

The values in the parentheses show the estimated standard errors. Diagnostic tests (not provided here for brevity) show no sign of correlation in either the fitted innovation $\hat{z}_{t}$ and its squares $\hat{z}_{t}^{2}$. That is, the volatility clustering for both return series can be captured through the normal GARCH(1, 1) model. However, applying part (a) of Theorem 2.1 to the estimates given in Table 1, we obtain the following implied normal GARCH kurtosis:

\[
\gamma_{2}^{(g)}(IBM) = 0.6076, \quad \gamma_{2}^{(g)}(\tilde{DD}) = 0.4966
\]
Figure 3: Plot of ACF Function for Series $IBM_t^2$
which are much smaller than the sample kurtosis of 6.5306 and 8.1551, respectively. It is clear that the normal GARCH(1, 1) model is inconsistent with the leptokurtosis.

To better explain the leptokurtosis observed in the series, we allow for conditional non-normality. Instead of using normal innovations in (17), we assume that $z_t$ are iid and follow a mixture of two normal distributions as in Bai et al. (2001):

$$z_t \sim \text{mixture normal } (\lambda, \eta) = \begin{cases} 
N(0, \sigma^2) & \text{with probability } (1 - \eta) \\
N(0, \lambda \sigma^2) & \text{with probability } \eta 
\end{cases} \quad (19)$$

where $\sigma^2 = (1 - \eta + \lambda \eta)^{-1}$ so that $\text{var}(z_t) = 1$. For the mixture normal $(\lambda, \eta)$, the kurtosis is,

$$\gamma_2(z) = \frac{3\eta(1-\eta)(\lambda - 1)^2}{(1-\eta + \lambda \eta)^2} \quad (20)$$

The resulting GARCH(1, 1) model fits the returns data very well in both cases. Table 2 gives the estimation results.

Table 2: MLE Estimates for IBM Daily Return with Mixture Normal Innovations

<table>
<thead>
<tr>
<th>Series</th>
<th>$\omega$</th>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>$\eta$</th>
<th>$\lambda$</th>
</tr>
</thead>
<tbody>
<tr>
<td>IBM</td>
<td>0.0068</td>
<td>0.0267</td>
<td>0.9656</td>
<td>0.0777</td>
<td>7.2077</td>
</tr>
<tr>
<td></td>
<td>(0.0023)</td>
<td>(0.0047)</td>
<td>(0.0060)</td>
<td>(0.0159)</td>
<td>(0.7621)</td>
</tr>
<tr>
<td>$\tilde{DD}$</td>
<td>0.1328</td>
<td>0.1296</td>
<td>0.7372</td>
<td>0.1163</td>
<td>6.9950</td>
</tr>
<tr>
<td></td>
<td>(0.0157)</td>
<td>(0.0125)</td>
<td>(0.0230)</td>
<td>(0.0103)</td>
<td>(0.3159)</td>
</tr>
</tbody>
</table>
Applying Equation (20) and part (b) of Theorem 2.1 to the above estimates, we obtain the following estimates of the components and overall kurtosis for two return series:

<table>
<thead>
<tr>
<th>Series</th>
<th>volatility clustering $\gamma_2^{(g)}$</th>
<th>non-normality $\gamma_2^{(z)}$</th>
<th>overall $\gamma_2^{(c)}$</th>
<th>sample kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>IBM</td>
<td>0.3074</td>
<td>3.7703</td>
<td>6.2510</td>
<td>6.53</td>
</tr>
<tr>
<td>$\tilde{D}\tilde{D}$</td>
<td>0.4686</td>
<td>3.8469</td>
<td>8.3163</td>
<td>8.16</td>
</tr>
</tbody>
</table>

For both series, we notice that the non-normality kurtosis is much larger than the kurtosis induced by volatility clustering and that the overall kurtosis is about twice the sum of the component kurtosis. Thus, once we allow for conditional non-normality, the implied overall kurtosis for IBM daily returns and Dollar/Deutsch Mark exchange rate returns are close to the sample kurtosis, indicating that a mixture GARCH(1, 1) model can explain the overall leptokurtosis in addition to volatility clustering.

4 The Kurtosis of Linear Stochastic Volatility Models

In this section, we extend the results in Section 2 on the GARCH($p$, $q$) model to the linear stochastic volatility LSV($p$, $q$) model (see Polson et al. (1994)) for the observable time series $\epsilon_t$,

\[
\begin{align*}
\epsilon_t &= \sqrt{h_t}z_t \\
h_t &= \omega + \sum_{i=1}^{p} \alpha_i h_{t-i} + u_t - \sum_{j=1}^{q} \beta_j u_{t-j}
\end{align*}
\]

where $(z_t, u_t)'$ is a bivariate unobservable time series, $z'_t$ are iid with mean 0 and unit variance, and $u'_t$ are iid with zero mean, variance $\sigma^2_u$ and independent of the $z'_t$'s. Similar to the GARCH process, $h_t$ can be regarded as the variance of $\epsilon_t$ conditional on the information set $\{u_t, u_{t-1}, u_{t-2}, \cdots\}$. We shall assume that the fourth moment of $z_t$ exists. In this section we derive the kurtosis of $\epsilon_t$ when it exists. Following the convention in Section 2, $\gamma_2^{(z)}$ denotes the kurtosis of $z_t$, and $\gamma_2^{(c)}$ the overall kurtosis of $\epsilon_t$. Writing $\alpha(B) = 1 - \sum_{i=1}^{p} \alpha_i B^i$, and $\beta(B) = 1 - \sum_{i=1}^{q} \beta_i B^i$ we now make the following assumption,

A.3 all the zeroes of the polynomial $\alpha(B)$ are lying outside of the unit circle

Under this assumption, the $h_t$ is weakly stationary with

\[
\begin{align*}
E(h_t) &= \frac{\omega}{(1 - \sum_{i=1}^{p} \alpha_i)} \\
E(h_t^2) &= \sigma^2_u (1 + a_1^*) + [E(h_t)]^2, a_1^* = \sum_{i=1}^{\infty} \pi_i^2 < \infty
\end{align*}
\]
where the $\pi'_i$'s are obtained from the relation $\pi(B)\alpha(B) = \beta(B)$ with $\pi(B) = 1 + \sum_{i=1}^{\infty} \pi_i B^i$.

If $z_t$ follows a normal distribution, the process defined by Equation (21) is called a conditional normal LSV($p,q$) process, or simply a normal LSV($p,q$) process. In the following, we call the kurtosis for the normal LSV process the SV kurtosis and denote it by $\gamma_2^{(sv)}$ when it exists.

The following theorem relates the LSV kurtosis with the volatility clustering parameters of the LSV model, and it also relates the overall kurtosis $\gamma_2^{(\epsilon)}$ of the process to iid kurtosis $\gamma_2^{(z)}$ and LSV kurtosis $\gamma_2^{(sv)}$.

**Theorem 4.1** If $\epsilon_t$ follows the process defined by equation (21), then under assumption A.3 we have

(a). $\gamma_2^{(sv)} = 3a_1^{*}\sigma_u^2 [E(h_t)]^{-2}$

(b). $\gamma_2^{(\epsilon)} = \gamma_2^{(z)} + \gamma_2^{(sv)} + \frac{1}{3}\gamma_2^{(z)} \gamma_2^{(sv)}$

where $E(h_t)$ is given in (22).

**Proof:** The LSV model in equation (21) implies that $E(\epsilon_t) = 0$, and under assumption A.3

\[ E(\epsilon_t^2) = E(h_t) \quad (24) \]
\[ E(\epsilon_t^4) = (\gamma_2^{(z)} + 3)E(h_t^2) \quad (25) \]

Thus,

\[ \gamma_2^{(\epsilon)} = \frac{3E(\epsilon_t^4)}{[E(\epsilon_t^2)]^2} - 3 = (\gamma_2^{(z)} + 3)\frac{E(h_t^2)}{[E(h_t)]^2} - 3 \]

\[ = (\gamma_2^{(z)} + 3)\{1 + a_1^{*}\sigma_u^2 [E(h_t)]^{-2}\} - 3 \quad (26) \]

Part (a) of the theorem follows from Equation (27) by letting $\gamma_2^{(z)} = 0$. Substituting $a_1^{*}\sigma_u^2 [E(h_t)]^{-2}$ by $\frac{1}{3}\gamma_2^{(sv)}$ in Equation (27), we obtain part (b) of the theorem.

As Theorem 2.1 for GARCH process, part (b) of Theorem 4.1 relates the overall kurtosis $\gamma_2^{(\epsilon)}$ of a LSV process to the volatility clustering property through the normal LSV kurtosis $\gamma_2^{(sv)}$, and to the conditional non-normality property through the iid kurtosis $\gamma_2^{(z)}$. Interestingly, although LSV and GARCH processes have different data generating mechanism, their overall kurtosis are determined symmetrically and interactively by two component kurtosis one of which represents the conditional non-normality property and the other the volatility clustering property. It is also
interesting to note that, for both models, the overall kurtosis is determined solely by the two component kurtosis.

Furthermore, part (b) of Theorem (4.1) suggests that $\gamma_2^{(sv)}$ and $\gamma_2^{(z)}$ have a non-linear combined effect on the overall kurtosis $\gamma_2^{(e)}$. When one of the $\gamma_2^{(sv)}$ and $\gamma_2^{(z)}$ is fixed, the overall kurtosis $\gamma_2^{(e)}$ is a linear function of the other.

5 Conclusion

This paper examines the kurtosis of a process which exhibits both volatility clustering and conditional non-normality. For the GARCH model and the linear stochastic volatility model, we find that the overall kurtosis of the process is a non-linear symmetric function of the kurtosis due to the volatility clustering and that due to conditional non-normality, respectively. In other words, volatility clustering and conditional non-normality contribute symmetrically and non-linearly to the overall kurtosis.

Applying the result to the classic normal GARCH model, we find that although the typical parameters can fit the volatility clustering very well, the implied kurtosis is far less than the sample kurtosis observed in the financial data. By allowing for the non-normality in the innovations, the non-normal GARCH model not only fit the dynamic relation well but also gives kurtosis estimate much closer to the sample kurtosis. Through the two examples in Section 3, we have demonstrated that a conditional non-normal GARCH model can capture both the volatility clustering and the exceeding large kurtosis typically observed in the financial series.

References


