TIME AGGREGATION OF NORMAL MIXTURE GARCH MODELS

Carol Alexander and Emese Lazar
ISMA Centre, School of Business, The University of Reading
Whiteknights Park, PO Box 242
United Kingdom
C.Alexander@ismacentre.rdg.ac.uk and E.Lazar@ismacentre.rdg.ac.uk

ABSTRACT
Normal mixture GARCH models capture the time-variation of variance, skewness and kurtosis that characterizes financial data. These models are more flexible and have been shown to offer a better fit than symmetric and asymmetric $t$-GARCH models. In this paper we give a weak definition for the normal mixture GARCH(1,1) model that is aggregating in time. This result paves the way for the analysis of the continuous time limits of such models.

KEY WORDS
GARCH process, normal mixture, time aggregation

1. Introduction
The empirical distributions of returns to financial assets are characterized by skewness and excess kurtosis ([1], [2], [3], [4] and [5]). One of the most flexible and tractable families of non-normal distributions that are able to capture the skewness and kurtosis commonly found in financial data is the normal mixture family ([6] and [7]).

The normal mixture density function is given by:

$$
\eta(x) = \sum_{i=1}^{K} p_i \phi_i(x) \quad \sum_{i=1}^{K} p_i = 1
$$

where $[p_1, p_2, ..., p_K]$ is the positive mixing law and $\phi_i(x) = \phi(x; \mu_i, \sigma_i^2)$ are normal density functions.

A random variable whose distribution is characterized by a density function of this form is denoted

$$
X \sim NM(p_1, ..., p_K; \mu_1, ..., \mu_K; \sigma_1^2, ..., \sigma_K^2)
$$

Normal mixtures have a wide range of interpretations ([6], [4], [8]). From a behavioral point of view two major directions can be differentiated: (i) the mixing law determines the relative frequencies of the different behavioral types in the population and (ii) it gives the probabilities that any element of the population behaves according to each of the behavioral types ([9]).

Another feature of financial data that is widely accepted is the volatility clustering of returns. To capture this, the broad class of normal GARCH models was introduced by Engle (1982) and Bollerslev (1986) ([10] and [11]). They assume:

$$
y_t = X_t \gamma + \epsilon_t, \quad \epsilon_t | I_{t-1} \sim N(0, \sigma_t^2)
$$

where $I_{t-1}$ represents the information set available at time $t-1$ and the conditional variance follows a deterministic process with autoregressive characteristics. The basic GARCH(1,1) conditional variance process for the errors is defined by:

$$
\sigma_t^2 = \omega + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2, \quad \omega > 0, \quad \alpha, \beta > 0, \quad \alpha + \beta < 1
$$

The normal GARCH(1,1) process can account for some, but not all of the unconditional kurtosis in the data. Although the unconditional kurtosis can be greater under the $t$-GARCH process, where the conditional density of the errors is $t$-distributed ([12]) this model cannot account for the time variation in skewness and in kurtosis that characterizes the conditional distributions of financial asset returns ([13], [14], [15]) except if these are explicitly modeled as in [14], [16] and [17].

Many authors have considered simple normal mixture GARCH models, where the error term follows a normal mixture distribution ([18], [19], [20], [21], [22] and [23]). However most of these papers impose severe constraints on the dynamics of conditional skewness and kurtosis. Consequently Haas, Mittnik and Paolella (2004) introduced the general normal mixture NM(K)-GARCH(p, q) process ([24]).

Alexander and Lazar (2005) provide an extensive analysis of the theoretical and empirical properties of the NM(K)-GARCH(1,1) process and show that the general NM(2)-GARCH(1,1) model offers a better fit than symmetric and asymmetric $t$-GARCH models ([25]). Moreover, time variation in conditional skewness and kurtosis is endogenous to the model.
The NM(K)-GARCH(1,1) process is defined in [25] by:

\[ Y_t = X_t' \gamma + \varepsilon_t \]  

(1)

\[ \sigma_t^2 = \sum_{i=1}^{K} p_i \sigma_{it-1}^2 + \sum_{i=1}^{K} p_i \mu_i^2, \]

(2)

\[ \sum_{i=1}^{K} p_i = 1 \quad \text{and} \quad \sum_{i=1}^{K} p_i \mu_i = 0 \]

(3)

\[ \sigma_{it}^2 = \omega_t + \alpha_t \varepsilon_{it-1}^2 + \beta_t \sigma_{it-1}^2, \quad i = 1, \ldots, K \]

(4)

and according to the strong definition, the error term follows, conditionally, a normal mixture distribution:

\[ \varepsilon_t \mid t-1 \sim \text{NM} \left( p_1, \ldots, p_K ; \mu_1, \ldots, \mu_K ; \sigma_{11}^2, \ldots, \sigma_{K1}^2 \right) \]

The application of this model should not be limited to econometric analysis. Since the commonly observed smile and/or skew in implied volatility surfaces results in excess kurtosis and/or skewness in the conditional densities of the underlying returns ([5]) neither the normal GARCH model, nor the t-GARCH model can be ‘smile consistent’. But since the NM(K)-GARCH(1,1) process (1) – (4) does capture time variation in the higher conditional moments, it is a more realistic candidate for GARCH option pricing.

To this end, the continuous-time limit of strong NM(K)-GARCH models is derived in [26]. However, when the continuous-time limit of the model is under discussion, time aggregation of a time variant model naturally comes into question: is the model keeping its properties when analyzed under a different frequency?

Obviously, the assumption (4) does not guarantee time aggregation: a simple counter example is that the sum of two NM(K) distributed variables will, in general, be NM(2K) distributed and not NM(K). So already shifting from a time-step of length \( m \) to a step of length \( 2m \) we loose time aggregation. Also, a special technique is needed to take care of the dependence of the conditional variance on all previous disturbances. Exactly the same problem arises even in the simple normal GARCH(1,1) process. This problem has been studied extensively by Drost and Nijman ([27]) who subsequently defined a ‘weak’ normal GARCH(1,1) process that is invariant under changing the length of the time interval of observations. Their result has no straightforward generalization to the normal mixture GARCH(1,1) process.

The purpose of this short paper is to derive a definition that is weaker than (4) under which the time aggregation property for normal mixture GARCH(1,1) processes does hold. This definition can later be used to study the continuous-time limit of the model.

2. Time Aggregation

Taking expectations in (2) shows that the total unconditional variance in the model can be expressed as a sum of two terms: the mean (under the discrete mixing law distribution) of the unconditional variances plus the variance (under the mixing law) of the conditional means of the components:

\[ E \left( \sigma_t^2 \right) = \sum_{i=1}^{K} p_i E \left( \sigma_{it}^2 \right) + \sum_{i=1}^{K} p_i \mu_i^2 \]

(5)

In this decomposition of the total variance we are interested in the relative contribution of each component variance, and to this end we define:

\[ q_i = \frac{E \left( \sigma_{it}^2 \right)}{E \left( \sigma_t^2 \right)} \]

(6)

\[ \eta_{it} = q_i \varepsilon_{it}^2 - \sigma_{it}^2 \]

(7)

It is easy to see that \( E(\eta_{it}) = 0 \). Re-writing (7) in the form

\[ \eta_{it} = q_i \left( \varepsilon_{it}^2 - \sigma_t^2 \right) + \left( q_i \sigma_t^2 - \sigma_{it}^2 \right) \]

we see that \( \eta_{it} \) captures two time-varying effects for each variance component: the deviation of the squared return from the overall conditional variance, and the difference between the conditional variance of the component and a proportion \( q_i \) of the overall conditional variance.

In the following definition \( f \) can be any single-valued function. We also adopt the behavioral interpretation (ii) above, so that \( p_i \) is the probability that the conditional variance process at time \( t \) is governed by the \( i \)-th component variance. To formalize this notation we introduce a variable \( s_t \in [1, 2, \ldots K] \) to denote the ‘state of nature’ at time \( t \).

\[ P(s_t = i) = p_i \]

(8)

\[ E(\varepsilon_t \mid s_t = i) = \mu_i \]

(9)

\[ E(\varepsilon_t f(\varepsilon_{t-1}, \ldots, \varepsilon_0)) = 0 \]

(10)

\[ \text{Corr} \left( \eta_{it}, \eta_{it-j} \right) = 0, \quad j > 0 \]

(11)

\[ \text{Corr} \left( \eta_{it}, \varepsilon_{u-j} \eta_{u-j} \right) = 0, \quad j > 0 \]

(12)

**Definition**

A random variable \( Y_t \) is said to follow a *weak NM(K)-GARCH(1,1) process* if, in addition to (1), (2), (3), (6) and (7) the following properties hold:

\[ P(s_t = i) = p_i \]

(8)

\[ E(\varepsilon_t \mid s_t = i) = \mu_i \]

(9)

\[ E(\varepsilon_t f(\varepsilon_{t-1}, \ldots, \varepsilon_0)) = 0 \]

(10)

\[ \text{Corr} \left( \eta_{it}, \eta_{it-j} \right) = 0, \quad j > 0 \]

(11)

\[ \text{Corr} \left( \eta_{it}, \varepsilon_{u-j} \varepsilon_{u-j} \right) = 0, \quad j > 0 \]

(12)
Note that the weak NM(K)-GARCH(1,1) process is not an exact generalization of the strong NM(K)-GARCH(1,1) process. It can be shown that under (4) some, but not all the correlations defined in (11) and (12) are exactly zero. However, those that are not zero are extremely small (less than 0.01 and 0.0005 respectively) and we believe they tend to zero as the sample size tends to infinity, although this has yet to be proved.

In the following theorem the variance process for the $i^{th}$ component, based on time-steps of length $m$ is defined by:

$$m \sigma_{i,m}^2 = m \omega_i + m \alpha_i \cdot m \sigma_{i-1,m}^2 + m \beta_i \cdot m \sigma_{i,(i-1)m}^2$$

for $i = 1, \ldots, K$.

**Theorem**

The class of weak NM(K)-GARCH(1,1) processes is closed under temporal aggregation. After annualization, the parameters of $m \sigma_{i,m}^2$ are given by:

$$m \omega_i = \omega_i \frac{1 - \left( \frac{\alpha_i}{q_i} + \beta_i \right)^m}{1 - \left( \frac{\alpha_i}{q_i} + \beta_i \right)}$$

$$m \alpha_i = q_i \left( \frac{\alpha_i}{q_i} + \beta_i \right)^m - m \beta_i$$

and $m \beta_i \in (0,1)$ is the solution to

$$\frac{m \beta_i}{1 + m \beta_i^2} = -\text{Corr}(v_{it}, v_{i,t-m})$$

where

$$v_{it} = \frac{1}{2} \sum_{p=0}^{2m-1} \min(p,m) \eta_{i,p} + \sum_{0 \leq s < t \leq m} c_k \eta_{i,t-p} + \sum_{0 \leq s < t \leq m} \left( \frac{\alpha_i}{q_i} + \beta_i \right)^m \sum_{0 \leq s < t \leq m} \sigma_{i,s} \sigma_{i,m-s}$$

and

$$c_k = \begin{cases} 1 & \text{if } k = 0, \\ \frac{\alpha_i}{q_i} \left( \frac{\alpha_i}{q_i} + \beta_i \right)^{k-1} & \text{if } 1 \leq k \leq m-1, \\ -\beta_i \left( \frac{\alpha_i}{q_i} + \beta_i \right)^{m-1} & \text{if } k = m. \end{cases}$$

**Example (m=2):**

$$2 \omega_i = \omega_i \left(1 + \frac{\alpha_i}{q_i} + \beta_i \right)$$

$$2 \alpha_i = q_i \left( \frac{\alpha_i}{q_i} + \beta_i \right)^2 - 2 \beta_i$$

and $2 \beta_i \in (0,1)$ is the solution to:

$$\frac{2 \beta_i}{1 + 2 \beta_i^2} = -\frac{\alpha_i}{q_i} + \beta_i$$

where:

$$c = \left( \frac{\alpha_i}{q_i} + \beta_i \right)^2$$

$$d = \frac{\alpha_i}{q_i} - \beta_i \left( \frac{\alpha_i}{q_i} + \beta_i \right) - \beta_i \left( \frac{\alpha_i}{q_i} + \beta_i \right)$$

$$e = \frac{4q_i^2 \text{E}(\epsilon_i^2 \epsilon_{i-1}^2)}{\text{E}(\epsilon_i^2)}$$

$$f = 1 + \frac{\alpha_i}{q_i} + \beta_i$$

**Proof of Theorem:**

From (3) we have:

$$\sigma_{i,t}^2 = \omega_i + \left( \frac{\alpha_i}{q_i} + \beta_i \right) q_i \epsilon_{i-1}^2 - \beta_i \eta_{i,t-1}$$

This can be rearranged as:

$$q_i \epsilon_{i,t}^2 = \omega_i \left(1 + \frac{\alpha_i}{q_i} + \beta_i \right) + \left( \frac{\alpha_i}{q_i} + \beta_i \right) \epsilon_{i,t-1}^2 + \eta_{i,t-1} \left( \frac{\alpha_i}{q_i} + \beta_i \right) \eta_{i,t-2}$$

Repeating this equation for $t-1$ and summing yields:

$$q_i \epsilon_{i,t}^2 = \omega_i \left(1 + \frac{\alpha_i}{q_i} + \beta_i \right) + \left( \frac{\alpha_i}{q_i} + \beta_i \right)^2 \epsilon_{i,t-2}^2 + \eta_{i,t-1} \left( \frac{\alpha_i}{q_i} + \beta_i \right) \eta_{i,t-2}$$

$$+ q_i \epsilon_{i,t} \epsilon_{i,t-1} - q_i \left( \frac{\alpha_i}{q_i} + \beta_i \right) \epsilon_{i,t-2} \epsilon_{i,t-3}$$

$$+ \frac{1}{2} \left( \eta_{i,t} \left( \frac{\alpha_i}{q_i} + \beta_i \right) \eta_{i,t-1} + \left( \frac{\alpha_i}{q_i} - \beta_i \left( \frac{\alpha_i}{q_i} + \beta_i \right) \right) \eta_{i,t-2} \right)$$

$$+ \frac{1}{2} \left( \beta_i \left( \frac{\alpha_i}{q_i} + \beta_i \right) \eta_{i,t-3} \right)$$

$$\epsilon_{i,t} = \epsilon_{i,t} - \alpha_i \epsilon_{i,t-1} - \beta_i \epsilon_{i,t-2}$$

$$= \epsilon_{i,t} - \alpha_i \epsilon_{i,t-1} - \beta_i \epsilon_{i,t-2} + \frac{1}{2} \left( \eta_{i,t} \left( \frac{\alpha_i}{q_i} + \beta_i \right) \eta_{i,t-1} + \left( \frac{\alpha_i}{q_i} - \beta_i \left( \frac{\alpha_i}{q_i} + \beta_i \right) \right) \eta_{i,t-2} \right)$$

$$+ \frac{1}{2} \left( \beta_i \left( \frac{\alpha_i}{q_i} + \beta_i \right) \eta_{i,t-3} \right)$$

$$+ \left( \frac{\alpha_i}{q_i} + \beta_i \right) \epsilon_{i,t-1}^2 + \left( \frac{\alpha_i}{q_i} + \beta_i \right) \epsilon_{i,t-2}^2 + \left( \frac{\alpha_i}{q_i} + \beta_i \right) \epsilon_{i,t-3}^2$$

$$+ \eta_{i,t-1} \left( \frac{\alpha_i}{q_i} + \beta_i \right) \eta_{i,t-2}$$

$$+ \eta_{i,t-1} \left( \frac{\alpha_i}{q_i} + \beta_i \right) \eta_{i,t-2}$$

$$+ \eta_{i,t-1} \left( \frac{\alpha_i}{q_i} + \beta_i \right) \eta_{i,t-2}$$
where we have used the following relationship for annualized returns:

\[
2 \varepsilon_i^2 = (\varepsilon_i + \varepsilon_{i-1})^2 / 2 = \varepsilon_i^2 / 2 + \varepsilon_{i-1}^2 / 2 + \varepsilon_i \varepsilon_{i-1}
\]

Letting

\[
v_{it} = q_i \varepsilon_i \varepsilon_{i-1} - q_i \left( \frac{\alpha_i}{q_i} + \beta_i \right) \varepsilon_{i-2} \varepsilon_{i-3} + \frac{1}{2} \left( \eta_{it} + \left( 1 + \frac{\alpha_i}{q_i} \right) \eta_{i-1} + \left( \frac{\alpha_i}{q_i} - \beta_i \left( \frac{\alpha_i}{q_i} + \beta_i \right) \right) \eta_{i-2} \right) - \beta_i \left( \frac{\alpha_i}{q_i} + \beta_i \right) \eta_{i-3}
\]

we can now write

\[
q_{i-2} \varepsilon_i^2 = \alpha_i \left( 1 + \frac{\alpha_i}{q_i} + \beta_i \right) + \left( \frac{\alpha_i}{q_i} + \beta_i \right)^2 q_{i-2} \varepsilon_{i-2}^2 + v_{it}
\]

and it is easy to see that:

\[
E(v_{it}) = 0
\]

\[
E(v_{it}v_{it-2k}) = 0 \quad \text{for } k > 1
\]

\[
\text{Corr}(v_{it}, v_{it-2}) = -\frac{ce + d}{e + ec^2 + f}
\]

Denoting the above correlation by \( a \), we have \( a \in (1/2, 0) \). \(^1\) Consider the function \( f(x) = x^2 + a^{-1}x + 1 \) for \( a \in (1/2, 0) \). Since \( f \) is continuous, \( f(0) = 1 \) and \( f(1) < 0 \), \( f \) must have a root \( x \) between 0 and 1. And if \( f(x) = 0 \),

\[
\frac{x}{1+x^2} = -a.
\]

Hence there exists \( 0 < \lambda_i < 1 \) such that:

\[
-Corr(v_{it}, v_{it-2}) = \frac{\lambda_i}{1 + \lambda_i^2}
\]

We now define \( w_{it} \) in the following way:

\[
w_{i0} = v_{i0} \quad \text{and} \quad w_{it} = v_{it} + \lambda_i w_{it-2}
\]

Obviously \( E(w_{it}) = 0 \). Also, it can be shown that: \(^2\)

\[
\text{Corr}(w_{it}, w_{it-2k}) = 0 \quad \text{for } k > 1.
\]

Rewriting (17) as

\[
v_{it} = w_{it} - \lambda_i w_{it-2}
\]

we have:

\[
q_{i-2} \varepsilon_i^2 = \alpha_i \left( 1 + \frac{\alpha_i}{q_i} + \beta_i \right) + \left( \frac{\alpha_i}{q_i} + \beta_i \right)^2 q_{i-2} \varepsilon_{i-2}^2 + w_{it} - \lambda_i w_{it-2}
\]

that is:

\[
q_{i-2} \varepsilon_i^2 - w_{it} = \alpha_i \left( 1 + \frac{\alpha_i}{q_i} + \beta_i \right) + \left( \frac{\alpha_i}{q_i} + \beta_i \right)^2 q_{i-2} \varepsilon_{i-2}^2 - \lambda_i w_{it-2} - q_{i-2} \varepsilon_{i-2}^2
\]

Now, similarly to (7), we define

\[
2 \sigma_{it}^2 = q_{i-2} \varepsilon_i^2 - w_{it}.
\]

Also, let:

\[
2 \sigma_{it}^2 = \sum_{i=1}^{K} p_i \varepsilon_i^2 \sigma_{it}^2 + \sum_{i=1}^{K} \mu_i^2
\]

Hence (19) becomes:

\[
2 \sigma_{it}^2 = \alpha_i \left( 1 + \frac{\alpha_i}{q_i} + \beta_i \right) + \left( \frac{\alpha_i}{q_i} + \beta_i \right)^2 q_{i-2} \varepsilon_{i-2}^2 + \lambda_i \sigma_{it-2}^2
\]

and this is the updating formula for the new two-period conditional variances of each component. Equations (10)

\(^2\) Let \( c_{2k} = \text{Corr}(w_{it}, w_{it-2k}) \). From (16) and (18) we have that \( c_4 = c_2 \left( 1 + \lambda_i^2 \right) \left( \lambda_i \left( 1 + \lambda_i^2 \right) \right) \). Assuming \( c_2 \geq 0 \) (the case for \( c_2 \leq 0 \) follows similarly) we have \( c_4 \geq c_2 \). Applying (15) similarly we obtain that \( c_{2(k+1)} \geq c_{2k} \), with equality only if \( c_2 = c_4 = \ldots = 0 \). Summing (15) for \( k = 2, 3, \ldots \) gives

\[
(1 - \lambda_i)^2 \sum_{k=2}^{\infty} c_{2k} + \lambda_i (c_4 - c_2) = 0
\]

which holds only if \( c_{2k} = 0 \) for \( k \geq 1 \).
and (11) for the new two-step NM(K)-GARCH(1,1) model (replacing \( \eta \) by \( w \) and using steps of length 2) are obviously satisfied. Equation (12) can also be proved easily using the results derived above. The proof for \( m \)-period time intervals follows by induction. QED

**Corollary:**
The transformation of the parameters within the time aggregation keeps the unconditional (individual and overall) annualized variances constant.

**Proof of Corollary:**
Taking expectations in (14) gives:

\[
(22) \quad E\left(\varepsilon_i^2\right) = E\left(\varepsilon_i^2\right)
\]

Equations (7) and (20) result in:

\[
(23) \quad E\left(\sigma_i^2\right) = q_iE\left(\varepsilon_i^2\right) \quad \text{and} \quad E\left(\sigma_i^2\right) = q_iE\left(\varepsilon_i^2\right)
\]

Combining (2), (20), (21) and (22) will lead to:

\[
E\left(\varepsilon_i^2\right) = E\left(\varepsilon_i^2\right)
\]

Hence the unconditional overall and individual variances remain unchanged as the time interval changes.

Note that the Theorem holds irrespective of the new \( m \)-step values for \( p \) and \( \mu \). In theory these could change in any way we like. For instance, we could introduce a simple parameterization for these parameters so that they change systematically over time. Naturally, we would require their definition to preserve the stylized properties of moments in accordance with the central limit theorem.

Suppose we define

\[
(23) \quad \mu_{\text{annualized}} = \mu_i \quad \text{and} \quad p_{\text{annualized}} = p_i
\]

for annualized returns. Then the unconditional skewness, which only depends on the unconditional variances and the parameters \( p \) and \( \mu \), will be unaffected by the time step change. However, we still have a term structure for kurtosis, even under the simple assumptions (23).

The analytical relationship between the kurtosis's computed using different time steps is fairly complex, but we have found that the kurtosis is a convex, generally decreasing function of the step length. To illustrate this, we estimated a NM(2)-GARCH(1,1) model for the Dow Jones index for the period of 2 Jan 1980 – 7 Sep 2004. Based on the estimated parameters and using the updating formulae given above we computed kurtosis values for different time steps (see Fig. 1). The excess kurtosis decreases when returns are measured at longer time steps, in accordance with the central limit theorem.

### 3. Conclusion

NM(K)-GARCH(1,1) models are gaining popularity due to their ability to capture time-varying higher moments. Still, as with simple normal GARCH(1,1) models, the usual ‘strong’ definition has the deficiency that when different time steps are considered, the number of normals in the normal mixture distribution changes.

This paper has introduced a weak definition for NM(K)-GARCH(1,1) models that is time aggregating, i.e. when discrete-time returns are measured at different time intervals, the model remains a valid description of the data using any step length. We obtained that the unconditional variance is not affected by the step length, but the kurtosis decreases as longer steps are considered.

Consequently, it is legitimate to study the continuous-time limit of this model using the weak definition. The advantage of using diffusion limits of discrete-time models is that they offer more advanced tools for option pricing and hedging than the discrete-time model on its own.

An empirical study testing the accuracy of the formulae derived is desirable. More than this, the study of the continuous-time limit of weak NM(K)-GARCH(1,1) models is of major interest (work in progress) and the main test will be to see how well these continuous models perform for option pricing and hedging.
References:


