TOPIC 6

EXTENSIONS OF GARCH PROCESSES

1. Integrated GARCH Process

Recall from Topic 4 (p. 8) that if ε_t follows a GARCH(1,1) process, then it can be shown that ε_t^2 has an ARMA(1,1) representation, namely,

$$\varepsilon_t^2 = \alpha_0 + (\alpha_1 + \beta_1)\varepsilon_{t-1}^2 - \beta_1 v_{t-1} + v_t \tag{1}$$

where $v_t = \varepsilon_t^2 - \sigma_t^2$ is the difference between the squared innovation and the conditional variance at time *t*. In many applications, we find that $\alpha_1 + \beta_1$ is approximately one. When $\alpha_1 + \beta_1 = 1$, equation (1) becomes

$$\varepsilon_t^2 = \alpha_0 + \varepsilon_{t-1}^2 - \beta_1 v_{t-1} + v_t \tag{2}$$

so that there is a unit root in the squared residuals ε_t^2. Equation (2) can be written as:

$$\Delta\varepsilon_t^2 = \alpha_0 - \beta_1 v_{t-1} + v_t \quad where \quad \Delta\varepsilon_t^2 = \varepsilon_t^2 - \varepsilon_{t-1}^2$$

Because the there is a unit root in the squared residuals (they are stationary in first differences), the model is called an Integrated GARCH(1,1), also known as the IGARCH(1,1) model.

Recall from Topic 4 notes (p. 8), that the *h*-step ahead forecast of the conditional variance from a GARCH(1,1) model is:

$$E(\sigma_{t+h}^2 \mid \Omega_t) = \alpha_0[1 + (\alpha_1 + \beta_1) + (\alpha_1 + \beta_1)^2 + \ldots + (\alpha_1 + \beta_1)^{h-2}] + (\alpha_1 + \beta_1)^{h-1}\sigma_{t+1}^2$$

and that

$$\lim_{h\to\infty} E(\sigma_{t+h}^2 \mid \Omega_t) = \frac{\alpha_0}{1-(\alpha_1 + \beta_1)} = \mathrm{var}(\varepsilon_t) \tag{3}$$

When $\alpha_1 + \beta_1 = 1$,

$$E(\sigma_{t+h}^2 \mid \Omega_t) = \alpha_0(h-1) + \sigma_{t+1}^2 \tag{4}$$

so that the forecast of the conditional variance becomes larger and larger as *h* increases. In the limit, as $h \to \infty$, the forecast of the conditional variance becomes infinitely large, meaning that the unconditional variance of the process is infinite (or undefined) as can be seen from equation (3) upon substituting $\alpha_1 + \beta_1 = 1$.

2. Asymmetric GARCH Models

In the GARCH (or ARCH) models that we have discussed so far, a positive or negative shock last period (that is, ε_{t-1}) will have the same impact on today's volatility because the squared of ε_{t-1} enters the model only. However, negative shocks appear to contribute more to stock market volatility than do positive shocks. This is called the leverage effect. A negative shock to aggregate stock prices reduces the aggregate market value of equity relative to the aggregate market value of corporate debt. Thus the likelihood of corporate bankruptcy increases as firms are more highly leveraged. This increases the risk of holding stocks.

The simplest GARCH model allowing for asymmetric response is the threshold GARCH or the TGARCH model. In this model the GARCH(1,1) conditional variance function is replaced with:

$$
\begin{aligned}
&\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \gamma \varepsilon_{t-1}^2 D_{t-1} + \beta_1 \sigma_{t-1}^2 \\
&\textit{where} \\
&D_t = \begin{cases} 1, & \textit{if } \varepsilon_t < 0 \\ 0, & \textit{otherwise} \end{cases} \\
&\textit{and} \\
&\alpha_0 > 0, \alpha_1 \geq 0, \alpha_1 + \gamma \geq 0, \beta_1 \geq 0
\end{aligned}
\tag{5}
$$

The dummy variable D_t keeps track of whether the lagged residual is positive or negative. When $\varepsilon_{t-1} \geq 0$, the effect of the lagged squared residual on the current conditional variance (σ_t^2) is simply α_1. In contrast, when $\varepsilon_{t-1} < 0$, $D = 1$ so that the effect of the lagged squared residual on the current conditional variance is $\alpha_1 + \gamma$. If $\gamma = 0$, the response is symmetric and we have the standard GARCH(1,1) model. If $\gamma \neq 0$, there is an asymmetric response of the conditional variance to "news", the lagged residual. If there are leverage effects, $\gamma > 0$ so that negative shocks have a bigger impact on the conditional variance than do positive shocks.

Asymmetric response may also be introduced by way of the exponential GARCH or EGARCH model:

$$
\ln(\sigma_t^2) = \alpha_0 + \alpha_1 \left| \frac{\varepsilon_{t-1}}{\sigma_{t-1}} \right| + \gamma \frac{\varepsilon_{t-1}}{\sigma_{t-1}} + \beta_1 \ln(\sigma_{t-1}^2) \tag{6}
$$

There are three important characteristics of the EGARCH model. First, the log of the conditional variance is being modeled not the conditional variance itself. Regardless of the magnitude of $\ln(\sigma_t^2)$, the implied value of σ_t^2 can never be negative. Thus, it is permissible for the coefficients (in equation (6)) to be negative. In other words, the log specification ensures that the conditional variance is always positive because σ_t^2 is obtained by exponentiating $\ln(\sigma_t^2)$. Second, instead of using the value of ε_{t-1}^2, the

EGARCH model uses the absolute value of the standardized value of ε_{t-1} (that is, ε_{t-1} divided by it standard error σ_{t-1}) as the measure of the size of a shock. Note that the standardized value of ε_{t-1} is a unit free measure. Third, the EGARCH model allows for asymmetric response of the log of the conditional variance to "news". The sign of the "news" is captured by the term $\varepsilon_{t-1}/\sigma_{t-1}$. If $\varepsilon_{t-1}/\sigma_{t-1}$ is positive, the effect of the standardized shock on the conditional variance is $\alpha_1 + \gamma$. If $\varepsilon_{t-1}/\sigma_{t-1}$ is negative, the effect of the standardized shock on the conditional variance is $\alpha_1 - \gamma$. If $\gamma < 0$, the effect of a negative standardized shock is larger than that of a positive shock so that there is evidence for a leverage effect.

3. Tests for Leverage Effects

First estimate the mean equation with, say, a GARCH(1,1) specification for the variance equation, by maximum likelihood methods and form the standardized residuals

$$s_t = \frac{\varepsilon_t}{\sigma_t}$$

To test for leverage effects, one could estimate a regression of the form

$$s_t^2 = a_0 + a_1 s_{t-1} + a_2 s_{t-2} + \cdots + a_n s_{t-k} + \eta_t \tag{7}$$

where η_t is the regression disturbance. If there are no leverage effects, the squared standardized residuals should be uncorrelated with the levels of the standardized residuals. If the regression slope coefficients were negative and statistically significant, that would indicate negative shocks are associated with large values of the conditional variance and, thus, there are leverage effects.

Engle and Ng (1993) developed a second way to determine whether positive and negative shocks have different effects on the conditional variance. Let

$$D_t = \begin{cases} 1 & if \;\; \varepsilon_t < 0 \\ 0 & if \;\; \varepsilon_t \geq 0 \end{cases}$$

The Sign Bias test uses the regression equation of the form

$$s_t^2 = a_0 + a_1 D_{t-1} + \eta_t \tag{8}$$

where η_t is the regression disturbance. If a *t*-test indicates that a_1 is statistically different from zero, the sign of the current period shock is helpful in predicting volatility. In particular, if a_1 is positive and statistically different from zero, negative shocks tend to increase the conditional variance. To generalize the test, one could estimate the regression:

$$s_t^2 = a_0 + a_1 D_{t-1} + a_2 D_{t-1} s_{t-1} + a_3 (1 - D_{t-1}) s_{t-1} + \eta_t$$

Note that $(1 - D_{t-1})$ assigns a value of one to positive or zero shocks. The presence of $D_{t-1}s_{t-1}$ and $(1 - D_{t-1})s_{t-1}$ is designed to determine whether the effects of positive and negative shocks on the conditional variance depend on their size. Statistical significance of a_2 and a_3 would suggest the presence of size bias, where not only the sign (indicated by the statistical significance of a_1) but also the magnitude or size of the shock is important for predicting the conditional variance.

4. Leverage Effects in the Composite NYSE Index

Recall from Topic 4 notes that we estimated an MA(1)-GARCH(1,1) model for the percentage daily logarithmic change in the NYSE index, denoted sr_t, over the period January 3, 1995 to August 30, 2002, a total of 1,931 observations. Having done this, we now save the standardized residuals from this model (denoted s_t) and estimate the regression given by equation (7) for three lags. The results are shown in Table 1.

Table 1: Estimation of Regression Equation for Leverage Effects

Dependent Variable: S2
Method: Least Squares
Sample (adjusted): 5 1931
Included observations: 1927 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.974715	0.046069	21.15789	0
S(-1)	-0.15996	0.045941	-3.48179	0.0005
S(-2)	-0.25772	0.045936	-5.6104	0
S(-3)	-0.0882	0.045937	-1.92	0.055

R-squared	0.024002	Mean dependent var	1.000512
Adjusted R-squared	0.022479	S.D. dependent var	2.037487
S.E. of regression	2.014457	Akaike info criterion	4.24065
Sum squared resid	7803.602	Schwarz criterion	4.252199
Log likelihood	-4081.87	F-statistic	15.76358
Durbin-Watson stat	2.075408	Prob(F-statistic)	0

The coefficients on s_{t-1}, s_{t-2} and s_{t-3} are negative and statistically significant. Thus, negative shocks are associated with large values of the conditional variance, suggesting the presence of leverage effects. Table 2 reports the results of the sign bias test given by equation (8).

Table 2: Results of the Sign Bias Test

Dependent Variable: S2
Method: Least Squares
Sample (adjusted): 3 1931
Included observations: 1929 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.637753	0.126028	5.060401	0
D(-1)	0.418212	0.135489	3.086678	0.0021

R-squared	0.00492	Mean dependent var	0.999597
Adjusted R-squared	0.004404	S.D. dependent var	2.036632
S.E. of regression	2.032143	Akaike info criterion	4.257095
Sum squared resid	7957.747	Schwarz criterion	4.262864
Log likelihood	-4103.97	F-statistic	9.527584
Durbin-Watson stat	1.967774	Prob(F-statistic)	0.002053

Since the coefficient on $D(-1)$ is positive and significant, we again conclude that negative shocks tend to increase the conditional variance of sr_t.

In view of these findings, we estimated the MA(1)-TGARCH(1,1) model. The results are reported in table 3. The coefficient on the asymmetric term is 0.1948. It is positive and statistically significant. Thus, there is evidence for leverage effects in the returns to the NYSE Composite Index.

It is interesting to compare the value of the likelihood function from the MA(1)-TGARCH model, which is -2475.02 with that from the MA(1)-GARCH(1,1) model, which is -2516.63. It is valid to make such a comparison since the MA(1)-TGARCH(1,1) model nests the MA(1)-GARCH(1,1). In other words, the MA(1)-GARCH(1,1) model can be viewed as a restricted model with respect to the MA(1)-TGARCH(1,1) model since it is obtained from the latter when the coefficient on the asymmetric term is restricted to be zero. Clearly the maximized value of the likelihood function from the MA(1)-TGARCH(1,1) model is larger than that from the MA(1)-GARCH(1,1) model. We would expect this since the coefficient on the asymmetric term in the TGARCH model is highly statistically significant. Nevertheless, we could perform a likelihood ratio test of the restriction that the coefficient on the asymmetric term is zero as follows:

$$LR = -2(LL_R - LL_U)$$
$$= -2(-2516.63 - (-2475.02))$$
$$= 83.2$$

The *LR* statistic is distributed as a $\chi^2(1)$ since there is only one restriction here. Since $83.2 > \chi^2_{0.05}(1) = 3.841$, we reject the null that there is no asymmetric response and conclude that the MA(1)-TGARCH(1,1) model is better than the MA(1)-GARCH(1,1) model.

Table 3: Results of Estimation of MA(1)-TGARCH(1,1) Model

Dependent Variable: SR
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 2 1931
Included observations: 1930 after adjustments
Convergence achieved after 19 iterations
MA backcast: 1, Variance backcast: ON
GARCH = C(3) + C(4)*RESID(-1)^2 + C(5)*RESID(-1)^2*(RESID(-1)<0)
+ C(6)*GARCH(-1)

	Coefficient	Std. Error	z-Statistic	Prob.
Mean Equation				
C	0.041635	0.020227	2.05835	0.0396
MA(1)	0.114205	0.023915	4.775395	0
Variance Equation				
C	0.018143	0.002799	6.482478	0
RESID(-1)^2	-0.00644	0.009726	-0.66206	0.5079
RESID(-1)^2*(RESID(-1)<0)	0.194754	0.015796	12.32918	0
GARCH(-1)	0.893656	0.009227	96.85522	0

R-squared	0.003704	Mean dependent var	0.0353
Adjusted R-squared	0.001115	S.D. dependent var	1.006207
S.E. of regression	1.005646	Akaike info criterion	2.571005
Sum squared resid	1945.788	Schwarz criterion	2.588306
Log likelihood	-2475.02	F-statistic	1.430784
Durbin-Watson stat	2.081274	Prob(F-statistic)	0.209991

The results of estimating the MA(1)-EGARCH(1,1) model are shown in table 4. The coefficient on the asymmetric term (shown in the table as C(5)) is -0.15524. Since this coefficient is negative and statistically significant, there is evidence for a leverage effect, that is negative shocks have a bigger impact on the log of the conditional variance than do positive shocks. We cannot compare the maximized log-likelihood value from the MA(1)-EGARCH(1,1) model with that from the MA(1)-GARCH(1,1) model since the models are not nested: the GARCH model cannot be viewed as a restricted EGARCH model since in the EGARCH the log of the conditional variance is being modeled whereas in the GARCH, the level of the conditional variance is being modeled.

Table 4: Results of Estimating MA(1)-EGARCH(1,1) Model

Dependent Variable: SR
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 2 1931
Included observations: 1930 after adjustments
Convergence achieved after 16 iterations
MA backcast: 1, Variance backcast: ON
LOG(GARCH) = C(3) + C(4)*ABS(RESID(-1)/@SQRT(GARCH(-1))) +
C(5)*RESID(-1)/@SQRT(GARCH(-1)) + C(6)*LOG(GARCH(-1))

	Coefficient	Std. Error	z-Statistic	Prob.
Mean Equation				
C	0.027938	0.019166	1.457706	0.1449
MA(1)	0.11858	0.023734	4.996317	0
Variance Equation				
C(3)	-0.10572	0.013741	-7.69377	0
C(4)	0.126307	0.016369	7.716313	0
C(5)	-0.15524	0.011328	-13.7042	0
C(6)	0.96458	0.004179	230.8339	0

R-squared	0.003355	Mean dependent var	0.03353
Adjusted R-squared	0.000765	S.D. dependent var	1.006207
S.E. of regression	1.005822	Akaike info criterion	2.555701
Sum squared resid	1946.47	Schwarz criterion	2.573003
Log likelihood	-2460.25	F-statistic	1.295357
Durbin-Watson stat	2.089712	Prob(F-statistic)	0.263051

5. Exogenous Variables in the GARCH Specification

Sometimes it is useful to include an exogenous variable in the variance equation. For example, financial market volume often helps to explain financial market volatility. In this case, the standard GARCH(1,1) model would be augmented in the following way

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 + \gamma x_t$$

where γ is a parameter and x_t is a positive exogenous variable, for example, the volume of trades on the NYSE today.

6. GARCH-in-Mean Models

The GARCH(1,1)-in-Mean model (which is written in abbreviated form as GARCH(1,1)-M) is:

$$y_t = a_0 + a_1\sigma_t + \varepsilon_t$$

$$\varepsilon_t \mid \Omega_{t-1} \sim N(0, \sigma_t^2)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \beta_1\sigma_{t-1}^2$$

Let y_t be the return on a financial asset or portfolio. Then $E(y_t \mid \Omega_{t-1}) = a_0 + a_1\sigma_t$. Thus, the conditional mean return depends on the conditional standard deviation. Since the conditional standard deviation can be viewed as a measure of the risk associated with the asset or portfolio, the specification for the mean equation captures the notion in finance of a trade-off between mean return and risk. The mean return is time-varying since σ_t is time-varying. Only in the case of where $a_1 = 0$ is the mean return constant, although there is time-varying volatility in the model given by the GARCH(1,1) specification. Note that in some empirical applications the conditional variance rather than the conditional standard deviation appears in the mean equation.

As a practical matter, if there appears to be a shift in the conditional mean of y_t in response to changing volatility, then that is indicative of a GARCH-M process.

Table 5 presents a GARCH-in-Mean model for the term premium between the three and six month U.S. zero coupon bonds.

Table 5: GARCH(1,1)-in-Mean Model for the Term Premium

Dependent Variable: TERM
Method: ML - ARCH (Marquardt) - Normal distribution
Sample (adjusted): 1947M01 1987M02
Included observations: 482 after adjustments
Convergence achieved after 32 iterations
Variance backcast: OFF
GARCH = C(4) + C(5)*RESID(-1)^2 + C(6)*GARCH(-1)

	Coefficient	Std. Error	z-Statistic	Prob.
	Mean Equation			
STDDEV	0.380607	0.076123	4.999884	0
C	0.003369	0.001901	1.772456	0.0763
TERM(-1)	0.712542	0.039053	18.24564	0
	Variance Equation			
C	2.75E-06	6.21E-06	0.443379	0.6575
RESID(-1)^2	0.385388	0.038328	10.05507	0
GARCH(-1)	0.756158	0.015696	48.17537	0

R-squared	0.484561	Mean dependent var	0.223071
Adjusted R-squared	0.479147	S.D. dependent var	0.220262
S.E. of regression	0.158963	Akaike info criterion	-1.72331
Sum squared resid	12.0282	Schwarz criterion	-1.6713
Log likelihood	421.3169	F-statistic	89.49705
Durbin-Watson stat	2.176771	Prob(F-statistic)	0

The term premium is defined as the yield to maturity on six month bills less the yield to maturity on three month bills. The data cover the period December 1946 to February 1987. The coefficient on the conditional standard deviation is positive and statistically significant as expected since the higher the risk, the higher the term premium required on the long bond relative to the short bond. Also included in the mean equation is the lagged term premium to account for serial correlation. It is apparent that the term premium is quite persistent. Finally, $\alpha_1 + \beta_1 = 1.15$, which is quite a bit larger than one, violating the sign restrictions on the model.

7. Maximum Likelihood Estimation of the ARMA-GARCH Models

Consider the ARMA(1,1)-GARCH(1,1) model:

$$
\begin{aligned}
y_t &= \gamma + \phi y_{t-1} + \theta\varepsilon_{t-1} + \varepsilon_t \\
\varepsilon_t \mid \Omega_{t-1} &\sim N(0, \sigma_t^2) \\
\sigma_t^2 &= c + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2
\end{aligned}
$$

The only observed series we have is $\{y_t\}$. Thus we will have to reconstruct $\{\varepsilon_t\}$ and $\{\sigma_t^2\}$ from observed $\{y_t\}$. We do so iteratively and need to assume values for *t*=0: ε_0 and σ_0^2. Usually, we set $\varepsilon_0 = 0$ and $\sigma_0^2 = \bar{\sigma}^2$, where $\bar{\sigma}^2$ is the (unconditional) sample variance. Then, for given values of parameters $\gamma, \phi, \theta, c, \alpha, \beta$ and given $\varepsilon_0 = 0$, $\sigma_0^2 = \bar{\sigma}^2$ and y_0 we can compute ε_1 and σ_1^2:

$$
\begin{aligned}
\varepsilon_1 &= y_1 - (\gamma + \phi y_0 + \theta\varepsilon_0) \\
\sigma_1^2 &= c + \alpha\varepsilon_0^2 + \beta\sigma_0^2
\end{aligned}
$$

All subsequent values of $\{\varepsilon_t\}$ and $\{\sigma_t^2\}$ are reconstructed in a similar way:

$$
\begin{aligned}
\varepsilon_t &= y_t - (\gamma + \phi y_{t-1} + \theta\varepsilon_{t-1}) \\
\sigma_t^2 &= c + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2
\end{aligned}
\qquad (9)
$$

Next step is to specify the likelihood, that is the joint probability to observe specific values of $\{y_t\}$ for given $\gamma, \phi, \varphi, c, \alpha, \beta$ and $\varepsilon_0 = 0$, $\sigma_0^2 = \bar{\sigma}^2$ and maximize it for given $\{y_t\}, \varepsilon_0 = 0$, $\sigma_0^2 = \bar{\sigma}^2$ with respect to the parameters $\gamma, \phi, \varphi, c, \alpha, \beta$.

In order to specify the likelihood we need to know the joint (unconditional) distribution of $\{y_t\}$. However what we are given, instead, is the conditional distributions of $\{y_t \mid \Omega_{t-1}\}$. Moreover $\{y_t\}$ are not independent.

There are two way around this problem both of which lead to the same solution.

One way it to consider maximizing the joint likelihood of the standardized innovations $\xi_t = \dfrac{y_t - (\gamma + \phi y_{t-1} + \theta\varepsilon_{t-1})}{\sigma_t^2}$. By assumption $\{\xi_t\}$ are iid standard normal random variables and their joint pdf is

$$f(\xi_1, \xi_2, \ldots, \xi_T \mid \gamma, \phi, \theta, c, \alpha, \beta, \varepsilon_0, \sigma_0^2) = \prod_{t=1}^{T} f(\xi_t \mid \varepsilon_t, \sigma_t^2) =$$
$$= \frac{1}{(\sqrt{2\pi})^T \prod_{t=1}^{T} \sigma_t} \exp\left\{-\frac{1}{2}\sum_{t=1}^{T}\frac{\varepsilon_t^2}{\sigma_t^2}\right\},$$

where ε_t and σ_t^2 are computed iteratively as in Eq. (9).

The other (I would say more proper) way is to use the following decomposition (applying the Bayes formula for conditional probability):

$$f(y_T, y_{T-1}, \ldots, y_1, y_0) = f(y_T \mid y_{T-1}, \ldots, y_1, y_0) f(y_{T-1}, \ldots, y_1, y_0) =$$
$$= f(y_T \mid y_{T-1}, \ldots, y_1, y_0) f(y_{T-1} \mid y_{T-2}, \ldots, y_1, y_0) f(y_{T-2}, \ldots, y_1, y_0) =$$
$$= f(y_T \mid \Omega_{T-1}) f(y_{T-1} \mid \Omega_{T-2}) \cdots f(y_1 \mid \Omega_0) f(\Omega_0)$$

and specify pdf conditional on parameters as

$$f(y_T, y_{T-1}, \ldots, y_1, y_0 \mid \gamma, \phi, \theta, c, \alpha, \beta, \varepsilon_0, \sigma_0^2) = \prod_{t=1}^{T} f(y_t \mid \Omega_{t-1}) f(\Omega_0) =$$
$$= \frac{1}{(\sqrt{2\pi})^T \prod_{t=1}^{T} \sigma_t} \exp\left\{-\frac{1}{2}\sum_{t=0}^{T}\frac{\varepsilon_t^2}{\sigma_t^2}\right\},$$

where ε_t and σ_t^2 are computed iteratively as in Eq. (9).

The LHS of this expression can be interpreted as the joint probability[1] to observe a given set of y_t conditional and $\gamma, \phi, \theta, c, \alpha, \beta, \varepsilon_0, \sigma_0^2$. Alternatively, it may be interpreted as a

---

[1] We talk about continuous random variables here and, therefore, technically the probability to observe any specific value is equal to zero. Instead, we need to talk about the probability of observing a range of values around Y_i:

$P\left(y_t - \dfrac{\delta}{2} \le Y \le y_t + \dfrac{\delta}{2}\right) \approx \delta f(y_t)$, where δ is a small constant

This point is often omitted since δ is a constant, which does not affect the optimization procedure.

function of parameters $\gamma, \phi, \theta, c, \alpha, \beta$, conditional on the sample outcomes y_t. In the latter interpretation it is referred to as a likelihood function and written

$$\text{Likelihood function } = L(\gamma, \phi, \theta, c, \alpha, \beta, \varepsilon_0, \sigma_0^2; y_t, t = 1, \ldots, T)$$

with the order of the symbols in brackets reflecting the emphasis on the parameters being conditioned on the observations. Maximizing the likelihood function with respect to the three parameters ($\gamma, \phi, \varphi, c, \alpha, \beta$) gives estimators of the parameters which maximize the probability of obtaining the sample values that have actually been observed.

In most applications, it is computationally easier to maximize the log of the likelihood function rather than maximizing the likelihood function itself. It makes no difference since the resulting expressions will yield estimators of the parameters that maximize both the likelihood and the log of the likelihood since the log is a monotonic transformation that does not alter the location of the maximum. The log-likelihood corresponding to equation is

$$l(\gamma, \phi, \theta, c, \alpha, \beta, \varepsilon_0, \sigma_0^2; y_t, t = 1, \ldots, T) = -\frac{T}{2}\ln(2\pi) - \sum_{t=1}^{T} \ln \sigma_t - \frac{1}{2}\sum_{t=1}^{T} \frac{\varepsilon_t^2}{\sigma_t^2}$$

with

$$\begin{aligned} \varepsilon_t &= y_t - (\gamma + \phi y_{t-1} + \theta \varepsilon_{t-1}) \\ \sigma_t^2 &= c + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2. \end{aligned}$$

The log-likelihood is computed using loop. The resulting expression is nonlinear and too complex to use analytic optimization (by setting derivative equal to zero). Instead, numerical optimization is used.

Since we use MLE, our estimators of the parameters are consistent and asymptotically normal and efficient.

Note: for multivariate models and Granger causality see lecture slides.