Chapter 12: Time Series Models of Heteroscedasticity

- Our ARIMA models that we have studied have modeled the conditional mean of our time series: The mean of Y_t given the previous observations.
- Our ARIMA models have assumed that the conditional variance is *constant* and equal to the noise variance, σ^2.
- For example, our $AR(1)$ model assumes that:

$$E(Y_t|Y_{t-1}, Y_{t-2}, \ldots) = \phi Y_{t-1} \text{ and}$$
$$var(Y_t|Y_{t-1}, Y_{t-2}, \ldots) = var(e_t) = \sigma^2$$

- If our time series exhibits nonconstant variance, sometimes this can be remedied by transforming the data (for example, by working with $\log(Y_t)$).
- But in other data sets, especially financial time series, the conditional variance is nonconstant in some irregular, random pattern that cannot be remedied by a transformation.
- A data set with nonconstant variance is called *heteroscedastic*.

The Return Series $\{r_t\}$

- In financial data, the *return* is a kind of growth rate of the series.
- If Y_t is the value or price of an asset at time t, then the return (also called the relative gain) of the asset at time t is denoted r_t and is defined as

$$r_t = \frac{Y_t - Y_{t-1}}{Y_{t-1}}.$$

- Note that $r_t = Y_t/Y_{t-1} - 1$, and so $1 + r_t = Y_t/Y_{t-1}$. Thus

$$\begin{aligned} \log(1 + r_t) &= \log\left(\frac{Y_t}{Y_{t-1}}\right) \\ &= \log(Y_t) - \log(Y_{t-1}) = \nabla \log(Y_t) \end{aligned}$$

Approximating the Return Series

- And if r_t is a very small number (near 0), it turns out that $r_t \approx \log(1 + r_t) \approx \log(1) = 0$ so we can approximate r_t by $\nabla \log(Y_t)$.
- This approximation is often done, and sometimes $\nabla \log(Y_t)$ is itself called the return.
- In addition, the return is sometimes multiplied by 100 so that it can be interpreted as the percent change in price.

An Example of a Heteroscedastic Time Series

- See the time series plot of the daily CREF stock values from August 26, 2004 until August 15, 2006.
- The time series plot shows an increasing trend, and possibly nonconstant variance.
- A time series plot of the returns $100 \times \nabla \log(Y_t), t = 1, \ldots, n$ shows that at certain time periods, the stock price is more volatile (more variable) than in other time periods, and "quiet periods" tend to alternate with "volatile periods."
- This phenomenon is called *volatility clustering*.
- There is particular volatility near the end of the plot (July 12-August 14, 2006), when there was a war in Lebanon.

More Analysis of the CREF Time Series

- An initial investigation of the returns indicates they could be modeled as white noise.
- Neither the ACF nor PACF indicate significant autocorrelation, and the mean of the return process is not significantly different from zero.
- However, the volatility clustering indicates that the variance is not constant over time, so the returns may not be independent and identically distributed (iid).
- If the data are iid, then *nonlinear* transformations of the data should resemble white noise as well.
- We can take the squared returns or absolute returns, and if these have significant autocorrelation, this is evidence that the original data were not iid.

Detailed Analysis of the CREF Time Series

- A glance at the ACF and the PACF of the absolute returns of the CREF series shows there are significant autocorrelations at several lags (not a white noise-like pattern).
- A look at the ACF and the PACF of the squared returns of the CREF series tells a similar story.
- The McLeod-Li test is a version of the Ljung-Box test for autocorrelation based on the squared data.
- The alternative hypothesis is that the data have autoregressive conditional heteroscedasticity (ARCH).

Testing for ARCH in the CREF Time Series

- Under the null hypothesis of no ARCH, the test statistic has a chi-square distribution with degrees of freedom equal to the number of autocorrelations used in the test.
- The `McLeod.Li.test` function plots the p-values of this test for a variety of different lags.
- For the CREF data, the test is significant when there are more than 3 lags used, indicating that ARCH does exist for these data.

The $ARCH(1)$ Model

- The simplest model for ARCH data is the $ARCH(1)$ model.
- Let $\sigma^2_{t|t-1}$ denote the conditional variance (or conditional volatility) of r_t, given all returns though time $t-1$.
- Then the $ARCH(1)$ model for the return process $\{r_t\}$ is:

$$r_t = \sigma_{t|t-1}\epsilon_t$$
$$\sigma^2_{t|t-1} = \omega + \alpha r^2_{t-1}$$

where $\alpha \geq 0$ and $\omega \geq 0$ are unknown parameters, and the ϵ_i's are iid random variables with mean zero and variance 1.

More on the $ARCH(1)$ Model

- The conditional distribution of $r_t | r_{t-1}$ has mean zero and variance $\omega + \alpha r_{t-1}^2$.
- So we see the conditional variance of r_t is not constant, and it depends on the previous return.
- Otherwise, the returns follow a type of white noise process (that has nonconstant conditional variance).
- A more general $ARCH(q)$ model is possible:

$$r_t = \sigma_{t|t-1} \epsilon_t$$
$$\sigma_{t|t-1}^2 = \omega + \alpha_1 r_{t-1}^2 + \cdots + \alpha_q r_{t-q}^2,$$

but we will not cover this in detail.

Properties of the $ARCH(1)$ Model

- The $ARCH(1)$ model for r_t can be written as an $AR(1)$ model for the *squared* returns r_t^2, where the noise process is non-normal:

$$r_t^2 = \omega + \alpha r_{t-1}^2 + \sigma_{t|t-1}^2(\epsilon_t^2 - 1)$$

- Therefore, we can specify an $ARCH(1)$ model for r_t if our tools suggest an $AR(1)$ model for the squared returns r_t^2.
- Also, it can be shown that $E(r_t) = 0$, and $var(r_t) = \sigma^2 = \omega/(1 - \alpha)$.
- This implies that $0 \leq \alpha < 1$, if (and only if) the return series is weakly stationary.

Properties of the $ARCH(1)$ Model

- Oddly, the $ARCH(1)$ model is weakly stationary even though the conditional variance is not constant (but note the unconditional variance is constant).
- Also, for any $h > 0$, $cov(r_{t+h}, r_t) = 0$, so $\{r_t\}$ has zero autocorrelation at lags greater than 0.
- Finally, the kurtosis of r_t is always greater than that of a normal distribution, so the distribution of r_t has "fatter tails" than a normal.

Estimating the $ARCH(1)$ Model

- The conditional variance $\sigma^2_{t|t-1}$ is a parameter and is not observable, but note that r_t^2 is an unbiased estimator of $\sigma^2_{t|t-1}$.
- The parameters ω and α of the $ARCH(1)$ model can be estimated by conditional ML.
- The `garch` function in the `tseries` package can estimate the $ARCH(1)$ model on real data.
- One issue is that the ARCH likelihood tends to be fairly flat unless n is large, so it can be difficult for numerical methods to find the true maximum.

Example with a Simulated $ARCH(1)$ Time Series

- See the R example for the plot of the simulated $ARCH(1)$ series with $\omega = 0.01$, $\alpha = 0.9$.
- We can clearly see the volatility clustering and nonconstant variance.
- The `garch` function estimates ω and α to produce the fitted model for $\sigma^2_{t|t-1}$.
- The diagnostic tests show that the model residuals appear uncorrelated.
- Note that we may wish to forecast this conditional variance h time units into the future.
- It can be shown that $\sigma^2_{t+h|t} = \omega + \alpha\sigma^2_{t+h-1|t}$, where we let $\sigma^2_{t+h|t} = r^2_{t+h}$ for $h < 0$.
- Plugging in the estimates for ω and α, we then have a recursive formula for the forecasted conditional variance.

Modeling the Conditional Mean and the Conditional Variance

- In the previous example, we just assumed the conditional mean was zero.
- For some data, the mean process is not constant, and we can combine a regression model or ARMA model for the mean with an ARCH model for the errors.
- For example, an $AR(1)$ model (with an intercept) for the mean process, with $ARCH(1)$ errors, would be

$$Y_t = \theta_0 + \phi Y_{t-1} + Z_t,$$

  where Z_t follows an $ARCH(1)$ model.
- This model can be fit with the `garchFit` function in the `fGarch` package.
- See the R example on the U.S. GNP data.

The GARCH Model

- The ARCH model can be extended to the *generalized autoregressive conditional heteroscedasticity*, or GARCH, model.
- The GARCH model introduces one or more lags of the conditional variance into the model.
- The $GARCH(1,1)$ model can be expressed as:

$$
\begin{aligned}
r_t &= \sigma_{t|t-1}\epsilon_t \\
\sigma^2_{t|t-1} &= \omega + \beta\sigma^2_{t-1|t-2} + \alpha r^2_{t-1}
\end{aligned}
$$

The $GARCH(p, q)$ Model

- This model can be generalized to the $GARCH(p, q)$ model by adding more lags of the conditional variance and/or squared returns:

$$
\begin{aligned}
r_t &= \sigma_{t|t-1}\epsilon_t \\
\sigma^2_{t|t-1} &= \omega + \beta_1\sigma^2_{t-1|t-2} + \cdots + \beta_p\sigma^2_{t-p|t-p-1} + \\
&\quad \alpha_1 r^2_{t-1} + \cdots + \alpha_q r^2_{t-q}
\end{aligned}
$$

- In this notation and in the `garch` function in the `tseries` package, the first subscript is the number of GARCH terms (number of β's) and the second subscript is the number of ARCH terms (number of α's).
- In some books and software (like the `garchFit` function in the `fGarch` package), the ordering of the subscripts is reversed.

Specifying and Estimating a GARCH Model

- If the returns follow a $GARCH(1,1)$ model, then the squared returns r_t^2 behave like an $ARMA(1,1)$ process.
- So if our model specification tools indicate the squared returns are $ARMA(1,1)$, we can use a $GARCH(1,1)$ model for the return process $\{r_t\}$.
- In general, if $\{r_t\}$ is $GARCH(p,q)$, then $\{r_t^2\}$ follows an ARMA model with orders $\max(p,q)$ and p.
- Note that if $q < p$, then q is impossible to identify using this rule, so the typical approach in that case is to fit a $GARCH(p,p)$ model and remove the terms whose estimated coefficients are not significant.
- In GARCH models, the β's and α's are usually constrained to be nonnegative, which is a sufficient (though not necessary) condition for the conditional variances to be nonnegative.
- Parameter estimation is carried out through ML.

Example on a Simulated $GARCH(1,1)$ Time Series

- See the R example for a simulated $GARCH(1,1)$ series with $\omega = 0.02$, $\alpha = 0.05$, and $\beta = 0.9$.
- The ACF and PACF of this series show little autocorrelation (just a bit at lags 3 and 20).
- But the ACFs and PACFs of the absolute values and squared values show lots of autocorrelations (not resembling white noise).
- The sample EACF of the squared values is unclear, possibly suggesting $ARMA(2,2)$, which would imply a $GARCH(2,2)$ for the original series.
- Alternatively, we can examine the sample EACF of the absolute values, which suggests an $ARMA(1,1)$, which would imply a $GARCH(1,1)$ for the original series.
- Fitting both models, the $GARCH(1,1)$ looks better, based on the AICs and on the significance of the estimated coefficients.

Example: The CREF Time Series

- If we fit a GARCH model to the CREF return series, we must decide on the appropriate orders.
- The EACF of the series of squared returns indicates an $ARMA(1,1)$ model, which implies a $GARCH(1,1)$ model for the CREF returns.
- The EACF of the absolute returns is a bit unclear, but possibly suggests the same model.
- The GARCH(1,1) model fit yields the estimates $\hat{\omega} = 0.01633$, $\hat{\alpha} = 0.04414$, $\hat{\beta} = 0.91704$.
- Note that this model's estimate of the long-term variance is $\hat{\omega}/(1 - \hat{\alpha} - \hat{\beta}) = 0.01633/(1 - 0.04414 - 0.91704) = 0.4206$, which is very close to the sample variance of 0.4161.

Diagnostics: The CREF Time Series

- The usual diagnostics can be used to check the fit of the GARCH model.
- The Jarque-Bera test and Ljung-Box test on the residuals both indicate no problems.
- The Q-Q plot of the residuals shows approximate normality, and the residual time plot and ACF of the squared residuals show that the squared residuals resemble white noise, which indicates a good model.
- A general Box-type test of the autocorrelations of the squared residuals indicates no problems (large P-value).
- A plot of the ACF of the absolute residuals and the `gBox` test on the absolute residuals shows some possible lag-2 autocorrelation.
- But overfitting with a $GARCH(1, 2)$ model produces worse results, and the AIC of the $GARCH(1, 1)$ model is better, so we conclude the $GARCH(1, 1)$ model gives a good fit.

An AR(1)-GARCH(1,1) Model Example

- The Dow Jones Industrial Average from April 20, 2006 to April 20, 2016 exhibits some autocorrelation in the original series of returns, and lots of autocorrelation in the squared returns.
- We can fit an $AR(1)$ model for the mean process, and a $GARCH(1,1)$ model for the conditional variance, using the `garchFit` function in the `fGarch` package.
- The plot of the predicted conditional standard deviation over time shows lots of volatility during the time of the financial crisis of 2008.