Non-Stationarity

March 24, 2020

The material in this set of notes is based on S&S Chapter 3, Sections 3.7-3.9, and S&S Chapter 5, with the exception of the material on non-stationarity tests based on **ARIMA**(p, d, q) models. This material is based on Section 2.7.5 of Tsay (2010), as well as several published journal articles introducing and reviewing these tests:

- Dickey and Fuller (1981), "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root";
- Said and Dickey (1984), "Testing for unit roots in autoregressive-moving average models of unknown order";
- Said and Dickey (1985), "Hypothesis Testing in **ARIMA**$(p, 1, q)$ Models";
- Phillips and Perron (1988), "Testing for a unit root in time series regression";
- Schwert (1989), "Tests for Unit Roots: A Monte Carlo Investigation".

I don't expect you to all to read them, but I thought they actually offered clearer explanations of what these tests are doing than the textbooks I have looked at so I wanted to share them as useful resources for anyone who is curious.

Review

Let's think back to when we first introduced the concept of stationarity. We defined a **stationary** time series y_t as having finite second moments, i.e. $\mathbb{E}[y_t^2] < \infty$ for all t, a constant mean function, $\mu_{y,t} = \mu_y$, and an autocovariance function $\kappa_y(s,t)$ that depends on s and t only through their absolute difference $h = |s - t|$. So far, we've just been using stationary models and crossing our fingers and hoping that our data is stationary. Now we'll start thinking about:

- How to assess whether or not a specific observed time series $\boldsymbol{y}$ is stationary;
- What kind of models to use if we conclude that an observed time series $\boldsymbol{y}$ is not stationary.

We'll see that these ideas are related.

A basic first step to assessing stationarity is always to just plot the time series and examine it carefully, assessing whether or not it looks like the mean and variance are constant over the entire time interval. Some further exploratory analysis can be performed by binning the data and examining how the bin means and variances change - if the time series corresponds to a stationary process and each bin contains enough observations, the bin means and variances should all be very similar and should not display any systematic trends.

If we want to take a more sophisticated approach to assessing stationarity, we might want to consider a parametric model for the data, where the parameter values determine whether or not the model is stationary. This will allow us to develop a hypothesis test for non-stationarity. When we do this, we'll tend to consider two different types of non-stationarity (1) non-stationarity of the mean and (2) non-stationarity of the variance.

Mean Non-Stationarity

The ARIMA(p, d, q) Model

The workhorse (most commonly used) model for data that displays non-stationarity of the mean is the **ARIMA**(p, d, q) model, which generalizes the **ARMA**(p, q) model that we are already familiar with. Let's introduce some new notation first, because we're going use the idea of differencing a time series to define the **ARIMA**(p, d, q) model. We introduce a differencing operator ∇^d defined according to:

$$\begin{aligned}
\nabla y_t &= y_t - y_{t-1} \\
\nabla^2 y_t &= \nabla y_t - \nabla y_{t-1} = y_t - 2y_{t-1} + y_{t-2} \\
\nabla^3 y_t &= \nabla^2 y_t - \nabla^2 y_{t-1} = y_t - 3y_{t-1} + 3y_{t-2} - y_{t-3} \\
&\vdots \\
\nabla^k y_t &= \nabla^{k-1} y_t - \nabla^{k-1} y_{t-1}.
\end{aligned}$$

Differencing is a very useful concept because it can be used to address certain kinds of mean non-stationarity. We can show this via a few simple examples. Suppose that we observe a nonstationary time series with a linear trend in time

$$y_t = a + bt + w_t,$$

where $w_t \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_w^2)$. What happens if we difference the observed time series? We obtain

$$\begin{aligned}
\nabla y_t =& a + bt + w_t - (a + b(t-1) + w_{t-1}) \\
=& b + w_t - w_{t-1}.
\end{aligned}$$

This is a **MA**(1) process with mean b, so we get a stationary process!

What if we'd had an even more complicated trend over time, e.g. a quadratic trend,

$$y_t = a + bt + ct^2 + w_t.$$

What happens if we difference the observed time series in this case? We obtain

$$\begin{aligned}\nabla y_t =& a + bt + ct^2 + w_t - \left(a + b\left(t-1\right) + c\left(t-1\right)^2 + w_{t-1}\right) \\ =& b + 2ct + c + w_t - w_{t-1}.\end{aligned}$$

This isn't stationary yet - we still have a linear trend in time. What if we difference again?

$$\begin{aligned}\nabla^2 y_t =& \nabla y_t - \nabla y_{-1} \\ =& \left(b + 2ct + c + w_t - w_{t-1}\right) - \left(b + 2c\left(t-1\right) + c + w_{t-1} - w_{t-2}\right) \\ =& 2c + w_t - 2w_{t-1} + w_{t-2}.\end{aligned}$$

This is a $\mathbf{MA}\left(2\right)$ process with mean $2c$, so again we get a stationary process!

In general, $\nabla^k y_t$ will be stationary if

$$y_t = a_0 + a_1 t + a_2 t^2 + \cdots + a_k t^k + w_t,$$

where $a_0, a_1, \ldots, a_k$ are the coefficients of a degree-k polynomial time trend and w_t is stationary!

This leads us to the definition of the $\mathbf{ARIMA}(p,d,q)$ model. A process y_t is said to be $\mathbf{ARIMA}(p,d,q)$ if

$$\phi\left(B\right)\left(\nabla^d y_t - \mu\right) = \theta\left(B\right) w_t, \tag{1}$$

where $w_t \overset{i.i.d.}{\sim} \mathcal{N}\left(0, \sigma_w^2\right)$ and $\mu = \mathbb{E}\left[\nabla^d y_t\right]$. This means that the differenced time series $\nabla^d y_t$ is an $\mathbf{ARMA}(p,q)$ process. The $\mathbf{ARIMA}(p,d,q)$ model generalizes the $\mathbf{ARMA}(p,q)$ model, insofar as setting $d = 0$ yields an $\mathbf{ARMA}(p,q)$ model. The $\mathbf{ARIMA}(p,d,q)$ model gives us a parametric framework for assessing stationarity - if $d = 0$, y_t is a stationary process, whereas if $d > 0$ y_t is non-stationary. This leads us to the three tests of non-stationarity.

- The **Dickey-Fuller** test tests the null hypothesis H that y_t is an $\mathbf{ARIMA}\left(0,1,0\right)$ process against the alternative hypothesis that y_t is a stationary $\mathbf{ARMA}\left(1,0,0\right)$ process.

It assumes that

$$\nabla y_t = a + \kappa y_{t-1} + w_t, \tag{2}$$

where $w_t \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_w^2)$. The null and alternative hypotheses can be equivalently expressed as $\kappa = 0$ and $\kappa \neq 0$, respectively. The test statistic is the t-statistic $\hat{\kappa}/\text{se}(\hat{\kappa})$ for the the least-squares estimate $\hat{\kappa}$ based on (2), and the approximate distribution of the test statistic under the null as $n \to \infty$ is the Dickey-Fuller distribution. This is a non-standard distribution, but the relevant quantiles have been derived. Although useful, this test is of limited utility because the alternative hypothesis is very restrictive - what if y_t is a stationary $\mathbf{ARMA}(2, 0, 0)$ or $\mathbf{ARMA}(1, 0, 1)$ process?

- The **Augmented Dickey-Fuller** test addresses this limitation by testing the null hypothesis that y_t is an $\mathbf{ARIMA}(p, 1, 0)$ process against the alternative hypothesis that y_t is an $\mathbf{ARMA}(p + 1, 0, 0)$ process. It assumes that

$$\nabla y_t = a + \kappa y_{t-1} + \phi_1 \nabla y_{t-1} + \cdots + \phi_p \nabla y_{t-p} + w_t, \tag{3}$$

where $w_t \sim \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_w^2)$. Clearly, if $\kappa = 0$, then y_t is an $\mathbf{ARIMA}(p, 1, 0)$. If $\kappa \neq 0$, we can rewrite the equation as a stationary $\mathbf{ARIMA}(p + 1, 0, 0)$ process,

$$y_t = a + (1 + \kappa + \phi_1) y_{t-1} + (\phi_2 - \phi_1) y_{t-2} + \cdots + (\phi_p - \phi_{p-1}) y_{t-p} - \phi_p y_{t-p-1} + w_t.$$

The order p is usually chosen by using AIC, AICc, or SIC/BIC to choose select the "best" $\mathbf{ARIMA}(p, 1, 0)$ for y_t. Once the order p has been chosen, the test statistic is the t-statistic $\hat{\kappa}/\text{se}(\hat{\kappa})$ for the the least-squares estimate $\hat{\kappa}$ based on (3), and the approximate distribution of $\hat{\kappa}$ under the null as $n \to \infty$ is still the Dickey-Fuller distribution. This is a much more useful test than the original Dickey-Fuller test, because the alternative hypothesis is the larger class of $\mathbf{ARIMA}(p + 1, 0, 0)$ models with $p > 0$. Furthermore, because many $\mathbf{ARIMA}(p + 1, 0, q)$ processes can be well approximated using $\mathbf{ARIMA}(k, 0, 0)$ processes for some value of k, we can think of

the alternative hypothesis also approximately including **ARIMA** $(p+1,0,q)$ models! This means that the augmented Dickey-Fuller test basically approximately works as a test of the null hypothesis that y_t is an **ARIMA** $(p,1,q)$ process against the alternative hypothesis that y_t is an **ARMA** $(p+1,0,q)$ process.

- The **Phillips-Perron** test is motivated by the concern that although many **ARIMA** $(p+1,0,q)$ processes can be well approximated using **ARIMA** $(k,0,0)$ processes for some value of k, k may need to be too large for this to be practically useful. It tests an even more general null hypothesis against a more general alternative hypothesis, specifically it tests the null hypothesis that ∇y_t is a stationary process (and y_t is not) against the alternative hypothesis that y_t is a stationary process. We can think of this assuming that

$$y_t - y_{t-1} = a + \kappa y_{t-1} + w_t, \tag{4}$$

  where w_t is a stationary process, with a finite number L of nonzero autocorrelations and slightly nonconstant variance. The number L is often chosen in practice as a deterministic function of the length of the time series, n. The test statistic is computed from the residuals from the linear regression based on (4), and the approximate distribution of the test statistic under the null as $n \to \infty$ is still the Dickey-Fuller distribution.

  The greater generality offered by the Phillips-Perron test does not come for free! Making fewer assumptions about how y_t behaves under the alternative can mean that we may need more data for the test statistic to be approximately Dickey-Fuller distributed under the null, i.e. for the test to actually perform the way we want it to.

These tests are often used to choose the differencing parameter d of **ARIMA**(p,d,q) given an observed time series $\boldsymbol{x}$. Note that it isn't appropriate to choose d using AIC, AICc, SIC/BIC, because differencing reduces the number of available observations, so **ARIMA** models with different values of d are fit using different numbers of observations. We can use

one of these non-stationarity tests to select d by specifying a desired level α, e.g. $\alpha = 0.05$, and then proceeding as follows:

(i) Conduct a level-α test of non-stationarity of y_t.

- Fail to reject non-stationarity $\implies$ Set $k = 1$, proceed to (ii).
- Reject non-stationarity $\implies$ STOP, set $d = 0$.

(ii) Conduct a level-α test of non-stationarity of $\nabla^k y_t$.

- Fail to reject non-stationarity $\implies$ Set $k = k + 1$, proceed to (ii).
- Reject non-stationarity $\implies$ STOP, set $d = k$.

Once an order d has been selected, the same approaches that we used to select p and q for $\mathbf{ARMA}(p, q)$ models, e.g. minimizing AIC, AICc, SIC/BIC, can be used because we can assume that $\nabla^d y_t$ can be treated as an r $\mathbf{ARMA}(p, q)$ process.

Forecasting based on an $\mathbf{ARIMA}(p, d, q)$ model is not as straightforward. Recall that we previously derived forecasts by minimizing expected forecast error:

$$\mathbb{E}\left[\left(y_{n+1} - \left(\sum_{j=1}^{n} c_{nj} y_{n+1-j}\right)\right)^2\right] = \mathbb{E}\left[x_{n+1}^2\right] + \boldsymbol{c}_n' \boldsymbol{A}_n \boldsymbol{c}_n - 2\boldsymbol{c}_n' \boldsymbol{b}_n,$$

where $a_{n,ij} = \mathbb{E}\left[y_{n+1-i} y_{n+1-j}\right]$ and $b_{n,i} = \mathbb{E}\left[y_{n+1} y_{n+1-i}\right]$. When we did this previously, we assumed that $\boldsymbol{x}$ was stationary which made $\boldsymbol{A}_n$ and $\boldsymbol{b}_n$ easy to compute because both have nice, simple forms that only depend on how far apart any two values y_t and y_s are in time: $\mathbb{E}\left[y_{n+1-i} y_{n+1-j}\right] = \gamma_y(i - j)$ and $\mathbb{E}\left[y_{n+1} y_{n+1-i}\right] = \gamma_y(i)$. When we allow $\boldsymbol{x}$ to be non-stationary and assume an $\mathbf{ARIMA}(p, d, q)$ model for $\boldsymbol{x}$ with $d > 0$, then the entries of $\boldsymbol{A}_n$ and $\boldsymbol{b}_n$ may not have the same nice form because the autocovariances may depend on the actual time indices of the values. Fortunately, we can rewrite each value of the time series y_t as:

$$y_t = y_0 + \sum_{i=1}^{t} \nabla^d y_i.$$

Then the forecasting equation is still quadratic in $\boldsymbol{c}_n$ and can be expressed as

$$\mathbb{E}\left[\left(y_0 + \sum_{i=1}^{n+1} \nabla^d y_i - \left(\sum_{j=1}^{n} c_{nj}\left(y_0 + \sum_{i=1}^{n+1-j} \nabla^d y_i\right)\right)\right)^2\right],$$

which will depend on the covariances $\mathbb{E}\left[y_0 \nabla^d y_i\right]$ for $i > 0$ and $\mathbb{E}\left[\nabla^d y_i \nabla^d y_j\right]$. If we assume that $\mathbb{E}\left[y_0 \nabla x_i^d\right] = 0$, then it will just depend on the autocovariances $\mathbb{E}\left[\nabla^d y_i \nabla^d y_j\right]$, which we can obtain easily because we have assumed that $\nabla^d y_t$ is stationary. We won't do this by hand in class because expanding all the terms of the forecasting equation for an $\mathbf{ARIMA}(p, d, q)$ model is very tedious even when $\mathbb{E}\left[y_0 \nabla x_i^d\right] = 0$, instead we'll rely on statistical software to obtain forecasts and forecast errors for us.

Using ARIMA to Address Seasonality with SARIMA$(p, d, q) \times (p_l, d_l, q_l)$

In practice, we often encounter data that has seasonal trends in time that induce mean non-stationarity. Letting $\boldsymbol{s}$ be an $n \times 1$ vector with entries s_t that take on l possible consecutive values $c_1, \ldots, c_l$ and satisfy $s_{t-l} = s_t$ for all $l > (t-1)$ and which indicate the season during which observation y_t was observed.

$$y_t = \sum_{j=1}^{l} b_j \mathbb{1}_{\{s_t = c_j\}} + w_t,$$

where w_t is stationary. Then the seasonally differenced process

$$\nabla^l y_t = \sum_{j=1}^{l} b_j \mathbb{1}_{\{s_t = c_j\}} + w_t - \sum_{j=1}^{l} b_j \mathbb{1}_{\{s_{t-l} = c_j\}} - w_{t-l} = w_t - w_{t-l},$$

is stationary! We introduce some new notation for this - if l is the number of season values $\boldsymbol{s}$ can take on, then

$$\begin{aligned}
\nabla_l y_t &= y_t - y_{t-l} \\
\nabla_l^2 y_t &= \nabla_s y_t - \nabla_s y_{t-l}, \\
&\dots \\
\nabla_l^{d+1} y_t &= \nabla_s^d y_t - \nabla_s^d y_{t-l},
\end{aligned}$$

and so on.

One way to address this to extend our $\mathbf{ARIMA}(p,d,q)$ models to include seasonal differencing. We call this a $\mathbf{SARIMA}(p,d,q)\times(p_l,d_l,q_l)$ model, and it is given by

$$\tilde{\phi}\left(B^l\right)\phi\left(B\right)\left(\nabla_l^{d_l}\nabla^d y_t - \mu\right) = \tilde{\theta}\left(B^l\right)\theta\left(B\right)w_t, \tag{5}$$

where $w_t \overset{i.i.d.}{\sim} \mathcal{N}\left(0,\sigma_w^2\right)$. This becomes a bit complicated, as there are now the set of parameters is much larger, specifically μ, σ_w^2, $\theta_1,\dots,\theta_q$, $\phi_1,\dots,\phi_p$, $\tilde{\theta}_1,\dots,\tilde{\theta}_{q_l}$, and $\tilde{\phi}_1,\dots,\tilde{\phi}_{pl}$.

Let's work through an example for a time series y_t that is observed monthly and displays a linear trend in time as well as seasonality, insofar as measurements from the same month across different years y_t and y_{t-12}, are similar. We might want to consider an $\mathbf{SARIMA}(1,1,1)\times(1,1,1)$ model with $\mu=0$, which assumes

$$\left(1-\tilde{\phi}_1 B^{12}\right)\left(1-\phi_1 B\right)\nabla_{12}\nabla y_t = \left(1+\tilde{\theta}_1 B^{12}\right)\left(1+\theta B\right)w_t, \tag{6}$$

where $w_t \overset{i.i.d.}{\sim} \mathcal{N}\left(0,\sigma_w^2\right)$. We can simplify the terms one at a time.The term $\nabla_{12}\nabla y_t$ first computes the first difference $\nabla y_t = y_t - y_{t-1}$ and then computes the annual differences of the first differences,

$$\nabla_{12}\nabla y_t = \nabla_{12}\left(y_t - y_{t-1}\right) = y_t - y_{t-12} - \left(y_{t-1} - y_{t-13}\right).$$

This means that the $\mathbf{SARIMA}(1,1,1)\times(1,1,1)$ model assumes that the annual differences of the first differences are stationary. The polynomials can be expanded $\left(1-\tilde{\phi}_1 B^{12}\right)\left(1-\phi_1 B\right)$

and $\left(1+\tilde{\theta}_1 B^{12}\right)$ can simply be expanded out using the fact that $B^i B^j = B^{i+j}$,

$$\left(1-\tilde{\phi}_1 B^{12}\right)(1-\phi_1 B) = 1-\tilde{\phi}_1 B^{12}-\phi_1 B+\phi_1\tilde{\phi}_1 B^{13}$$
$$\left(1+\tilde{\theta}_1 B^{12}\right)(1-\theta_1 B) = 1+\tilde{\theta}_1 B^{12}+\theta_1 B+\theta_1\tilde{\theta}_1 B^{13}.$$

Putting it altogether yields one big complicated model! We can expand the terms in (6) to

$$\left(1-\tilde{\phi}_1 B^{12}-\phi_1 B+\phi_1\tilde{\phi}_1 B^{13}\right)(y_t - y_{t-12} - (y_{t-1} - y_{t-13})) = \left(1+\tilde{\theta}_1 B^{12}+\theta_1 B+\theta_1\tilde{\theta}_1 B^{13}\right) w_t.$$

We could expand this out further, but this is a much more interpretable form that looks more like the models we have seen before and further simplification would make things much messier.

Adding a Time Trend

Another way of addressing nonstationarity of the mean is to explicitly model a trend over time, and assume that deviations from that trend are stationary. We alluded to this when we introduced the idea of differencing. We saw that if a time series has a linear trend in time,

$$y_t = a + bt + w_t, \tag{7}$$

and errors $w_t \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_w^2)$ that are stationary about the mean, then the first differences $\nabla y_t = y_t - y_{t-1}$ will be stationary. Alternatively, we could continue to analyze the data on the original scale but explicitly model and then extract a linear trend in time.

However, we need to be careful when deciding whether or not to include a linear trend, because a time series that is stationary about a linear trend can look very similar to a non-stationary **ARMA** (p, q) time series. Fortunately, all of the tests we discussed previously have been extended to allow us to test the null hypothesis that y_t is non-stationary about

a linear trend against the alternative hypothesis that y_t is stationary about a linear trend! We won't go into the details, but most statistical software that implements the Augmented Dickey-Fuller or Phillips-Perron test the has an option to include or exclude a time trend.

Variance Non-Stationarity

Variance-Stablizing Transformations

The simplest thing we can do to address non-stationarity of the variances over the time series is to perform a variance stabilizing transformation on the raw data y_t before continuing any further and applying any time series models. This is appropriate when the variances are monotonically increasing or decreasing over time. The Box-Cox power family of transformations

$$y_t = \begin{cases} \left(y_t^\lambda - 1\right)/\lambda & \lambda \neq 0 \\ \log\left(y_t\right) & \lambda = 0 \end{cases}$$

are often used, where the parameter λ is chosen to make the variance of y_t as nearly constant as possible over time. In practice, $\lambda = 0$ or $\lambda = 1/2$ are often chosen. We can then apply any of our time series models to the transformed time series y_t. Forecasting can be performed by computing forecasts $\hat{y}_{n+k}$ for $k > 0$ and then using the inverse of whatever transformation function was applied to obtain $\hat{x}_{n+k}$. Variance-stabilizing transformations can be very useful, but be warned that applying them *can* lead to biased forecasts on the original scale, $\hat{x}_{n+k}$. See Guerrero (1993), "Time-series analysis supported by power transformations" for a discussion of one popular approach to choosing a transformation and debiasing the corresponding forecasts.

GARCH and ARCH Models

When we observe non-stationary variances that they are *not* monotonically increasing or decreasing in time but rather alternating between low- and high-variance periods we may want to explicitly model the variances. A common model for this kind of non-stationarity is the **generalized ARCH (GARCH)** model, which includes the simpler **ARCH** model as a special case. A $\mathbf{GARCH}(m, r)$ model for a time series y_t is given by:

$$\begin{aligned} y_t =& \sigma_t e_t \\ \sigma_t^2 =& \alpha_0 + \alpha_1 x_{t-1}^2 + \cdots + \alpha_m x_{t-m}^2 + \beta_1 \sigma_{t-1}^2 + \cdots + \beta_r \sigma_{t-r}^2, \end{aligned} \tag{8}$$

where $e_t \overset{i.i.d.}{\sim} \mathcal{N}(0, 1)$. Note that without imposing some conditions on $\alpha_0, \alpha_1, \ldots, \alpha_m$, we cannot be sure that the $\mathbf{GARCH}(m, r)$ model is well defined. By well defined, we mean that $\mathbf{GARCH}(m, r)$ model only produces nonnegative conditional variances σ_t^2 and corresponds to a causal model for the time series y_t.

The $\mathbf{ARCH}(m)$ model is obtained by setting $r = 0$. If we want to ensure that the $\mathbf{ARCH}(m)$ model is well defined, we need to impose conditions on the coefficients $\alpha_0, \alpha_1, \ldots, \alpha_m$. One set of such conditions is:

($\star$) $\alpha_0, \alpha_1, \ldots, \alpha_m \geq 0$;

(†) $\sum_{i=1}^{m} \alpha_i < 1$.

The condition ($\star$) ensures that the variances σ_t^2 are nonnegative, whereas the condition (†) ensures that the model for y_t is causal. The latter condition (†) makes sense intuitively when we realize that we can think of the $\mathbf{ARCH}(m)$ as an $\mathbf{AR}(m)$ model for the variances σ_t^2.

Let's work through a little example of an $\mathbf{ARCH}(1)$ model first to get some intuition.

- The $\mathbf{ARCH}(1)$ process is mean-zero,

$$\mathbb{E}[y_t] = \mathbb{E}[\sigma_t e_t] = \mathbb{E}[\sigma_t]\,\mathbb{E}[e_t] = 0.$$

- $\mathbb{V}[y_t | \sigma_t^2] = \mathbb{E}[\sigma_t^2 e_t^2 | \sigma_t^2] = \sigma_t^2$;

- The variance process over time is stationary with unconditional variance

$$\begin{aligned}\mathbb{E}\left[\sigma_t^2\right] &= \alpha_0 + \alpha_1 \mathbb{E}\left[\mathbb{E}\left[x_{t-1}^2 | \sigma_t^2\right]\right] \\ &= \alpha_0 + \alpha_1 \mathbb{E}\left[\sigma_{t-1}^2\right] \\ \mathbb{E}\left[\sigma_t^2\right] &= \frac{\alpha_0}{1-\alpha_1} \\ \gamma_y(0) &= \frac{\alpha_0}{1-\alpha_1}.\end{aligned}$$

- Consecutive values of y_t are *uncorrelated*,

$$\gamma_y(h) = \mathbb{E}\left[y_t y_{t-h}\right] = \mathbb{E}\left[\sigma_t \sigma_{t-h} e_t e_{t-h}\right] = \mathbb{E}\left[\sigma_t \sigma_{t-h}\right] \mathbb{E}\left[e_t\right] \mathbb{E}\left[e_{t-h}\right] = 0.$$

- Consecutive values of y_t^2 are *correlated*,

$$\gamma_{y^2}(h) = \alpha_1^h.$$

For $\mathbf{GARCH}(m, r)$ models with orders m and $r > 0$, we are not able to easily characterize the conditions that must be imposed for the $\mathbf{GARCH}(m, r)$ model to be well defined. We can think of a $\mathbf{GARCH}(m, r)$ model for y_t as being closey related to an $\mathbf{ARMA}(\max(m, r), r)$ process for x_t^2, where the noise w_t corresponds to $v_t = x_t^2 - \sigma_t^2 = \sigma_t^2(e_t^2 - 1)$. Showing this takes some algebra:

$$\begin{aligned}x_t^2 =& \sigma_t^2 + v_t \\ x_t^2 =& \alpha_0 + \alpha_1 x_{t-1}^2 + \cdots + \alpha_m x_{t-m}^2 + \beta_1 \sigma_{t-1}^2 + \cdots + \beta_r \sigma_{t-r}^2 + v_t \\ x_t^2 =& \alpha_0 + \alpha_1 x_{t-1}^2 + \cdots + \alpha_m x_{t-m}^2 + \beta_1\left(x_{t-1}^2 - v_{t-1}\right) + \cdots + \beta_r\left(x_{t-r}^2 - v_{t-r}\right) + v_t \\ x_t^2 =& \begin{cases} \alpha_0 + \left(\sum_{i=1}^r \beta_i v_{t-i} + (\alpha_i + \beta_i) x_{t-i}^2\right) + \left(\sum_{i=r+1}^m \alpha_i x_i^2\right) + v_t & m > r \\ \alpha_0 + \left(\sum_{i=1}^r \beta_i v_{t-i} + (\alpha_i + \beta_i) x_{t-i}^2\right) + v_t & m = r \\ \alpha_0 + \left(\sum_{i=1}^m \beta_i v_{t-i} + (\alpha_i + \beta_i) x_{t-i}^2\right) + \left(\sum_{i=m+1}^r \beta_i v_{t-i} + \beta_i x_{t-i}^2\right) + v_t & m < r. \end{cases}\end{aligned}$$

Note that this is not quite an $\mathbf{ARMA}(\max(m, r), r)$ process because the noise is not

$\sigma_t^2 \left(e_t^2 - 1\right)$ is not normal.

Maximum Likelihood Estimation

Let $k = \max(m, r)$. Because each y_t is conditionally normally distributed given $y_1, \ldots, y_{t-k}$ for $t > k$ with mean zero and variance σ_t^2, we can estimate the unknown parameters by maximum likelihood, where the likelihood is given by:

$$p\left(y_{k+1}, \ldots, y_n\right) = \prod_{t=k+1}^{m} \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left\{-\frac{1}{2}\left(\frac{x_t^2}{\sigma_t^2}\right)\right\}.$$

This can be difficult to do computationally in practice, because the parameters enter into the likelihood nonlinearly. Alternatively, we can just use our **ARMA** tools to get an approximate estimate of the $\mathbf{GARCH}(m, r)$ parameters, because a $\mathbf{GARCH}(m, r)$ process for y_t is closely related to an $\mathbf{ARMA}(\max(m, r), r)$ process for x_t^2 with parameters that are simple linear functions of the $\mathbf{GARCH}(m, r)$ parameters.

Combining ARCH and GARCH with ARMA

We can add additional time dependence into a $\mathbf{GARCH}(m, r)$ model by adding an $\mathbf{ARMA}(p, q)$ model as well:

$$\begin{aligned} \phi(B)\left(y_t - \mu_y\right) =& \theta(B)\left(w_t\right) + \sigma_t e_t \\ \sigma_t^2 =& \alpha_0 + \alpha_1 e_{t-1}^2 + \cdots + \alpha_m e_{t-m}^2 + \beta_1 \sigma_{t-1}^2 + \cdots + \beta_r \sigma_{t-r}^2, \end{aligned} \tag{9}$$

where $e_t \overset{i.i.d.}{\sim} \mathcal{N}(0, 1)$ and $w_t \overset{i.i.d.}{\sim} \mathcal{N}(0, \sigma_w^2)$.

This looks very complicated, but we can actually estimate the parameters of this model reasonably well in an ad-hoc but effective way just using the `arima` and `garch` functions from the `stats` and `tseries` packages for `R` by:

– Fitting an $\mathbf{ARMA}(p, q)$ model to the data y_t and computing residuals r_t;

– Fitting a **GARCH**(m, r) model to the residuals r_t.

More sophisticated simultaneous estimation is more computationally challenging, but can be performed using the `rugarch` package for `R`.