Chapter 4

Nonlinear Time Series Models

Prerequisites

- A basic understanding of expectations, conditional expectations and how one can use conditioning to obtain an expectation.

Objectives:

- Use relevant results to show that a model has a stationary, solution.
- Derive moments of these processes.
- Understand the differences between linear and nonlinear time series.

So far we have focused on linear time series, that is time series which have the representation

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j}, \tag{4.1}$$

where $\{\varepsilon_t\}$ are iid random variables. Such models are extremely useful, because they are designed to model the autocovariance structure and are straightforward to use for forecasting. These are some of the reasons that they are used widely in several applications. Note that all stationary Gaussian time series have a linear form (of the type given in (4.1)), where the innovations $\{\varepsilon_t\}$ are Gaussian.

A typical realisation from a linear time series, will be quite regular with no suddent bursts or jumps. This is due to the linearity of the system. However, if one looks at financial data, for example, there are sudden bursts in volatility (variation) and extreme values, which calm down after a while. It is not possible to model such behaviour well with a linear time series. In order to capture 'nonlinear behaviour several nonlinear models have been proposed. The models typically consists of products of random variables which make possible the sudden irratic bursts seen in the data. Over the past 30 years there has been a lot research into nonlinear time series models. Probably one of the first nonlinear models proposed for time series analysis is the bilinear model, this model is used extensively in signal processing and engineering. A popular model for modelling financial data are (G)ARCH-family of models. Other popular models are random autoregressive coefficient models and threshold models, to name but a few (see, for example, Subba Rao (1977), Granger and Andersen (1978), Nicholls and Quinn (1982), Engle (1982), Subba Rao and Gabr (1984), Bollerslev (1986), Terdik (1999), Fan and Yao (2003), Straumann (2005) and Douc et al. (2014)).

Once a model has been defined, the first difficult task is to show that it actually has a solution which is almost surely finite (recall these models have dynamics which start at the $-\infty$, so if they are not well defined they could be 'infinite'), with a stationary solution. Typically, in the nonlinear world, we look for causal solutions. I suspect this is because the mathematics behind existence of non-causal solution makes the problem even more complex.

We state a result that gives sufficient conditions for a stationary, causal solution of a certain class of models. These models include ARCH/GARCH and Bilinear models. We note that the theorem guarantees a solution, but does not give conditions for it's moments. The result is based on Brandt (1986), but under stronger conditions.

Theorem 4.0.1 (Brandt (1986)) *Let us suppose that $\{\boldsymbol{X}_t\}$ is a d-dimensional time series defined by the stochastic recurrence relation*

$$\boldsymbol{X}_t = A_t \boldsymbol{X}_{t-1} + \boldsymbol{B}_t, \tag{4.2}$$

where $\{A_t\}$ and $\{B_t\}$ are iid random matrices and vectors respectively. If $\mathrm{E}\log\|A_t\| < 0$ and $\mathrm{E}\log\|\boldsymbol{B}_t\| < \infty$ (where $\|\cdot\|$ denotes the spectral norm of a matrix), then

$$\boldsymbol{X}_t = \boldsymbol{B}_t + \sum_{k=1}^{\infty}\left(\prod_{i=0}^{k-1} A_{t-i}\right)\boldsymbol{B}_{t-k} \tag{4.3}$$

converges almost surely and is the unique strictly stationary causal solution.

Note: The conditions given above are very strong and Brandt (1986) states the result under *which weaker conditions, we outline the differences here. Firstly, the assumption* $\{A_t, B_t\}$ *are iid can be relaxed to their being Ergodic sequences. Secondly, the assumption* $\mathrm{E}\log\|A_t\| < 0$ *can be relaxed to* $\mathrm{E}\log\|A_t\| < \infty$ *and that* $\{A_t\}$ *has a negative Lyapunov exponent, where the Lyapunov exponent is defined as* $\lim_{n\to\infty}\frac{1}{n}\|\prod_{j=1}^{n} A_j\| = \gamma$*, with* $\gamma < 0$ *(see Brandt (1986)).*

The conditions given in the above theorem may appear a little cryptic. However, the condition $\mathrm{E}\log|A_t| < 0$ (in the unvariate case) becomes quite clear if you compare the SRE model with the AR(1) model $X_t = \rho X_{t-1} + \varepsilon_t$, where $|\rho| < 1$ (which is the special case of the SRE, where the coefficients is deterministic). We recall that the solution of the AR(1) is $X_t = \sum_{k=1}^{\infty}\rho^j\varepsilon_{t-j}$. The important part in this decomposition is that $|\rho^j|$ decays geometrically fast to zero. Now let us compare this to (4.3), we see that ρ^j plays a similar role to $\prod_{i=0}^{k-1} A_{t-i}$. Given that there are similarities between the AR(1) and SRE, it seems reasonable that for (4.3) to converge, $\prod_{i=0}^{k-1} A_{t-i}$ should converge geometrically too (at least almost surely). However analysis of a product is not straight forward, therefore we take logarithms to turn it into a sum

$$\frac{1}{k}\log\prod_{i=0}^{k-1} A_{t-i} = \frac{1}{k}\sum_{i=0}^{k-1}\log A_{t-i} \stackrel{\text{a.s.}}{\to} \mathrm{E}[\log A_t] := \gamma,$$

since it is the sum of iid random variables. Thus taking anti-logs

$$\prod_{i=0}^{k-1} A_{t-i} \approx \exp[k\gamma],$$

which only converges to zero if $\gamma < 0$, in other words $\mathrm{E}[\log A_t] < 0$. Thus we see that the condition $\mathrm{E}\log|A_t| < 0$ is quite a logical conditional afterall.

4.1 Data Motivation

4.1.1 Yahoo data from 1996-2014

We consider here the closing share price of the Yahoo daily data downloaded from `https://uk.finance.yahoo.com/q/hp?s=YHOO`. The data starts from from 10th April 1996 to 8th August 2014 (over 4000 observations). A plot is given in Figure 4.1. Typically the logarithm of such data taken, and in order to remove linear and/or stochastic trend the first difference of the logarithm is taken (ie. $X_t = \log S_t - \log S_{t-1}$). The hope is that after taking differences the data has been stationarized (see Example 2.3.2). However, the data set spans almost 20 years and this assumption is rather precarious and will be investigated later. A plot of the data after taking first differences together with the QQplot is given in Figure 4.2. From the QQplot in Figure 4.2, we observe that log

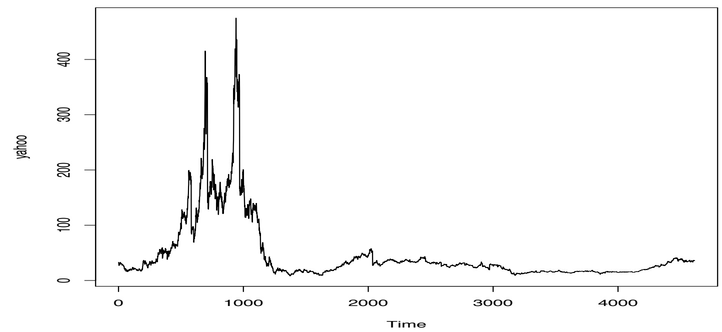


Figure 4.1: Plot of daily closing Yahoo share price 1996-2014

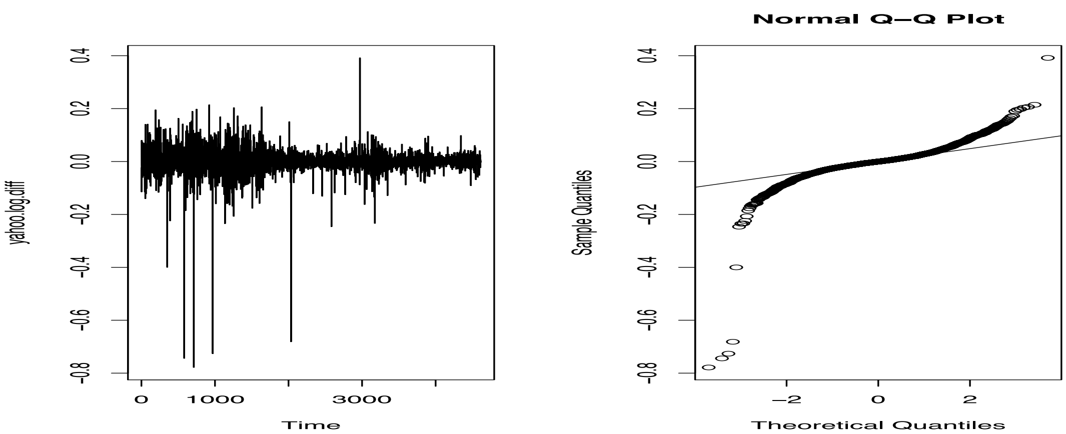


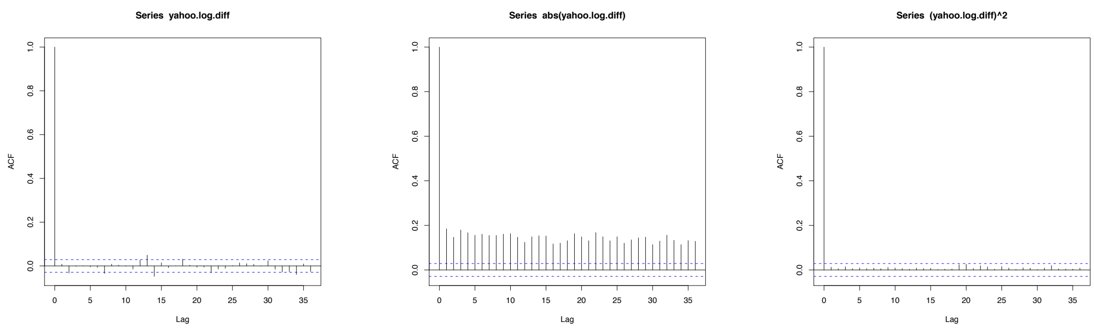
Figure 4.2: Plot of log differences of daily Yahoo share price 1996-2014 and the corresponding QQplot

differences $\{X_t\}$ appears to have very thick tails, which may mean that higher order moments of the log returns do not exist (not finite).

In Figure 4.3 we give the autocorrelation (ACF) plots of the log differences, absolute log differences and squares of the log differences. Note that the sample autocorrelation is defined as

$$\widehat{\rho}(k) = \frac{\widehat{c}(k)}{\widehat{c}(0)}, \quad \text{where} \quad \widehat{c}(k) = \frac{1}{T}\sum_{t=1}^{T-|k|}(X_t - \bar{X})(X_{t+k} - \bar{X}). \tag{4.4}$$

The dotted lines are the errors bars (the 95% confidence of the sample correlations constructed under the assumption the observations are independent, see Section 6.2.1). From Figure 4.3a we see that there appears to be no correlation in the data. More precisely, most of the sample correlations are within the errors bars, the few that are outside it could be by chance, as the error bars are constructed pointwise. However, Figure 4.3b the ACF plot of the absolutes gives significant large correlations. In contrast, in Figure 4.3c we give the ACF plot of the squares, where there does not appear to be any significant correlations.



(a) ACF plot of the log differences
(b) ACF plot of the absolute of the log differences
(c) ACF plot of the square of the log differences

Figure 4.3: ACF plots of the transformed Yahoo data

To summarise, $\{X_t\}$ appears to be uncorrelated (white noise). However, once absolutes have been taken there does appear to be dependence. This type of behaviour cannot be modelled with a linear model. What is quite interesting is that there does not appear to be any significant correlation in the squares. However, on explanation for this can be found in the QQplot. The data has extremely thick tails which suggest that the forth moment of the process may not exist (the empirical variance of X_t will be extremely large). Since correlation is defined as (4.4) involves division by $\widehat{c}(0)$, which could be extremely large, this would 'hide' the sample covariance.

`R` code for Yahoo data

Here we give the `R` code for making the plots above.

```
yahoo <- scan("~/yahoo304.96.8.14.txt")
yahoo <- yahoo[c(length(yahoo):1)] # switches the entries to ascending order 1996-2014
yahoo.log.diff <- log(yahoo[-1]) - log(yahoo[-length(yahoo)])
# Takelog differences
par(mfrow=c(1,1))
```

```
plot.ts(yahoo)
par(mfrow=c(1,2))
plot.ts(yahoo.log.diff)
qqnorm(yahoo.log.diff)
qqline(yahoo.log.diff)
par(mfrow=c(1,3))
acf(yahoo.log.diff) # ACF plot of log differences
acf(abs(yahoo.log.diff)) # ACF plot of absolute log differences
acf((yahoo.log.diff)**2) # ACF plot of square of log differences
```

4.1.2 FTSE 100 from January - August 2014

For completeness we discuss a much shorter data set, the daily closing price of the FTSE 100 from 20th January - 8th August, 2014 (141 observations). This data was downloaded from `http://markets.ft.com/research//Tearsheets/PriceHistoryPopup?symbol=FTSE:FSI`.

Exactly the same analysis that was applied to the Yahoo data is applied to the FTSE data and the plots are given in Figure 4.4-4.6.

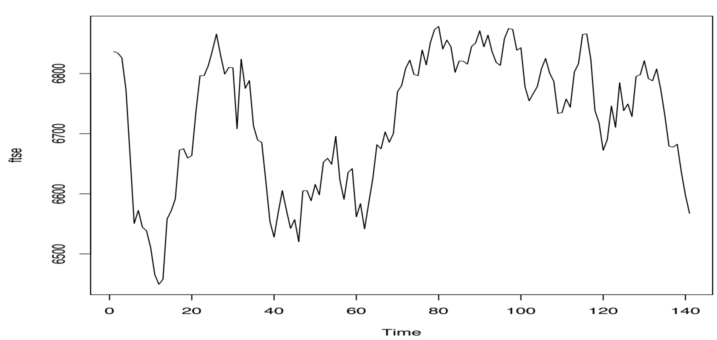


Figure 4.4: Plot of daily closing FTSE price Jan-August, 2014

We observe that for this (much shorter) data set, the marginal observations do not appear to deviate much from normality (note just because the marginal is Gaussian does not mean the entire time series is Gaussian). Furthermore, the ACF plot of the log differences, absolutes and squares do not suggest any evidence of correlation. Could it be, that after taking log differences, there is no dependence in the data (the data is a realisation from iid random variables). Or that there is dependence but it lies in a 'higher order structure' or over more sophisticated transformations.

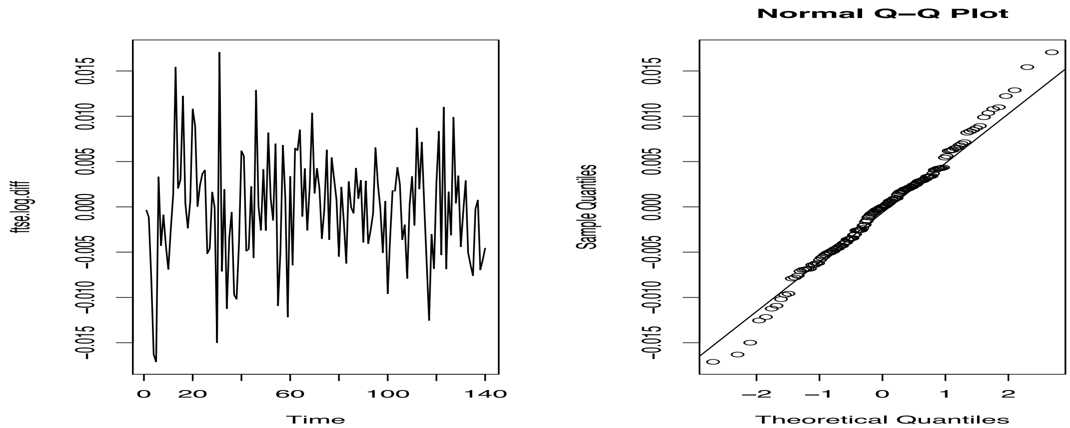
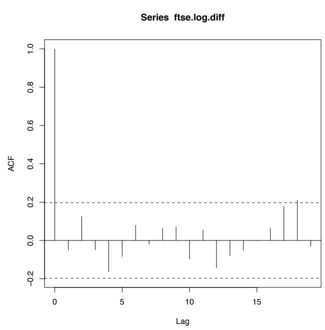
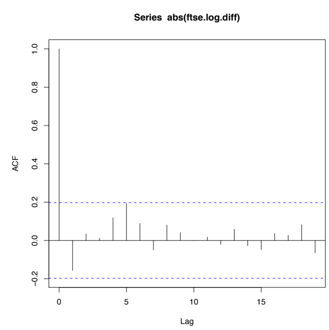


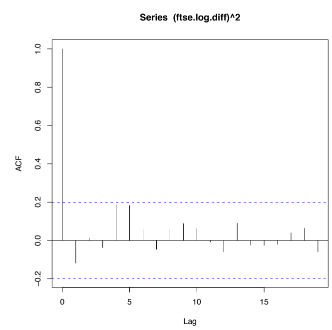
Figure 4.5: Plot of log differences of daily FTSE price Jan-August, 2014 and the corresponding QQplot



(a) ACF plot of the log differences



(b) ACF plot of the absolute of the log differences



(c) ACF plot of the square of the log differences

Figure 4.6: ACF plots of the transformed FTSE data

Comparing this to the Yahoo data, may be we 'see' dependence in the Yahoo data because it is actually nonstationary. The mystery continues (we look into this later). It would be worth while conducting a similar analysis on a similar portion of the Yahoo data.

4.2 The ARCH model

During the early 80s Econometricians were trying to find a suitable model for forecasting stock prices. They were faced with data similar to the log differences of the Yahoo data in Figure 4.2. As Figure 4.3a demonstrates, there does not appear to be any linear dependence in the data, which makes the best linear predictor quite useless for forecasting. Instead, they tried to predict the variance of future prices given the past, $\mathrm{var}[X_{t+1}|X_t, X_{t-1}, \ldots]$. This called for a model that has a zero autocorrelation function, but models the conditional variance.

To address this need, Engle (1982) proposed the autoregressive conditionally heteroskadastic (ARCH) model (note that Rob Engle, together with Clive Granger, in 2004, received the Noble prize for Economics for Cointegration). He proposed the ARCH(p) which satisfies the representation

$$X_t = \sigma_t Z_t \qquad \sigma_t^2 = a_0 + \sum_{j=1}^{p} a_j X_{t-j}^2,$$

where Z_t are iid random variables where $\mathrm{E}(Z_t) = 0$ and $\mathrm{var}(Z_t) = 1$, $a_0 > 0$ and for $1 \leq j \leq p$ $a_j \geq 0$.

Before, worrying about whether a solution of such a model exists, let us consider the reasons behind why this model was first proposed.

4.2.1 Features of an ARCH

Let us suppose that a causal, stationary solution of the ARCH model exists (X_t is a function of $Z_t, Z_{t-1}, Z_{t-1}, \ldots$) and all the necessary moments exist. Then we obtain the following.

(i) The first moment:

$$\begin{aligned} \mathrm{E}[X_t] &= \mathrm{E}[Z_t\sigma_t] = \mathrm{E}[\mathrm{E}(Z_t\sigma_t|X_{t-1}, X_{t-2}, \ldots)] = \underbrace{\mathrm{E}[\sigma_t \mathrm{E}(Z_t|X_{t-1}, X_{t-2}, \ldots)]}_{\sigma_t \text{ function of } X_{t-1},\ldots,X_{t-p}} \\ &= \mathrm{E}[\sigma_t \underbrace{\mathrm{E}(Z_t)}_{\text{by causality}}] = \mathrm{E}[0 \cdot \sigma_t] = 0. \end{aligned}$$

Thus the ARCH process has a zero mean.

(ii) The conditional variance:

$$\begin{aligned} \mathrm{var}(X_t|X_{t-1}, X_{t-2}, \ldots, X_{t-p}) &= \mathrm{E}(X_t^2|X_{t-1}, X_{t-2}, \ldots, X_{t-p}) \\ &= \mathrm{E}(Z_t^2\sigma_t^2|X_{t-1}, X_{t-2}, \ldots, X_{t-p}) = \sigma_t^2 \mathrm{E}[Z_t^2] = \sigma_t^2. \end{aligned}$$

Thus the conditional variance is $\sigma_t^2 = a_0 + \sum_{j=1}^{p} a_j X_{t-j}^2$ (a weighted sum of the squared past).

(iii) The autocovariance function:

Without loss of generality assume $k > 0$

$$\begin{aligned} \text{cov}[X_t, X_{t+k}] &= \text{E}[X_t X_{t+k}] = \text{E}[X_t \text{E}(X_{t+k}|X_{t+k-1}, \ldots, X_t)] \\ &= \text{E}[X_t \sigma_{t+k} \text{E}(Z_{t+k}|X_{t+k-1}, \ldots, X_t)] = \text{E}[X_t \sigma_{t+k} \text{E}(Z_{t+k})] = \text{E}[X_t \sigma_{t+k} \cdot 0] = 0. \end{aligned}$$

The autocorrelation function is zero (it is a white noise process).

(iv) We will show in Section 4.2.2 that $\text{E}[X^{2d}] < \infty$ iff $[\sum_{j=1}^{p} a_j]\text{E}[Z_t^{2d}]^{1/d} < 1$. It is well known that even for Gaussian innovations $\text{E}[Z_t^{2d}]^{1/d}$ grows with d, therefore if any of the a_j are non-zero (recall all need to be positive), there will exist a d_0 such that for all $d \geq d_0$ $\text{E}[X_t^d]$ will not be finite. Thus the we see that the ARCH process has thick tails.

Usually we measure the thickness of tails in data using the Kurtosis measure (see wiki).

Points (i-iv) demonstrate that the ARCH model is able to model many of the features seen in the stock price data.

In some sense the ARCH model can be considered as a generalisation of the AR model. That is the squares of ARCH model satisfy

$$X_t^2 = \sigma^2 Z_t^2 = a_0 + \sum_{j=1}^{p} a_j X_{t-j}^2 + (Z_t^2 - 1)\sigma_t^2, \tag{4.5}$$

with characteristic polynomial $\phi(z) = 1 - \sum_{j=1}^{p} a_j z^j$. It can be shown that if $\sum_{j=1}^{p} a_j < 1$, then the roots of the characteristic polynomial $\phi(z)$ lie outside the unit circle (see Exercise 2.1). Moreover, the 'innovations' $\epsilon_t = (Z_t^2 - 1)\sigma_t^2$ are *martingale differences* (see wiki). This can be shown by noting that

$$\text{E}[(Z_t^2 - 1)\sigma_t^2 | X_{t-1}, X_{t-2}, \ldots] = \sigma_t^2 \text{E}(Z_t^2 - 1 | X_{t-1}, X_{t-2}, \ldots) = \sigma_t^2 \underbrace{\text{E}(Z_t^2 - 1)}_{=0} = 0.$$

Thus $\text{cov}(\epsilon_t, \epsilon_s) = 0$ for $s \neq t$. Martingales are a useful asymptotic tool in time series, we demonstrate how they can be used in Chapter 10.

To summarise, in many respects the ARCH(p) model resembles the AR(p) except that the innovations $\{\epsilon_t\}$ are martingale differences and not iid random variables. This means that despite the resemblence, it is not a linear time series.

We show that a unique, stationary causal solution of the ARCH model exists and derive conditions under which the moments exist.

4.2.2 Existence of a strictly stationary solution and second order stationarity of the ARCH

To simplify notation we will consider the ARCH(1) model

$$X_t = \sigma_t Z_t \qquad \sigma_t^2 = a_0 + a_1 X_{t-1}^2. \tag{4.6}$$

It is difficult to directly obtain a solution of X_t, instead we obtain a solution for σ_t^2 (since X_t can immediately be obtained from this). Using that $X_{t-1}^2 = \sigma_{t-1}^2 Z_{t-1}^2$ and substituting this into (4.6) we obtain

$$\sigma_t^2 = a_0 + a_1 X_{t-1}^2 = (a_1 Z_{t-1}^2)\sigma_{t-1}^2 + a_0. \tag{4.7}$$

We observe that (4.7) can be written in the stochastic recurrence relation form given in (4.2) with $A_t = a_1 Z_{t-1}^2$ and $B_t = a_0$. Therefore, by using Theorem 4.0.1, if $\mathrm{E}[\log a_1 Z_{t-1}^2] = \log a_1 + \mathrm{E}[\log Z_{t-1}^2] < 0$, then σ_t^2 has the strictly stationary causal solution

$$\sigma_t^2 = a_0 + a_0 \sum_{k=1}^{\infty} a_1^k \prod_{j=1}^{k} Z_{t-j}^2.$$

The condition for *existence* using Theorem 4.0.1 and (4.7) is

$$\mathrm{E}[\log(a_1 Z_t^2)] = \log a_1 + \mathrm{E}[\log Z_t^2] < 0, \tag{4.8}$$

which is immediately implied if $a_1 < 1$ (since $\mathrm{E}[\log Z_t^2] \leq \log \mathrm{E}[Z_t^2] = 0$), but it is also satisfied under weaker conditions on a_1.

To obtain the moments of X_t^2 we use that it has the solution is

$$X_t^2 = Z_t^2 \left(a_0 + a_0 \sum_{k=1}^{\infty} a_1^k \prod_{j=1}^{k} Z_{t-j}^2 \right), \tag{4.9}$$

therefore taking expectations we have

$$\mathrm{E}[X_t^2] = \mathrm{E}[Z_t^2]\mathrm{E}\left(a_0 + a_0\sum_{k=1}^{\infty} a_1^k \prod_{j=1}^{k} Z_{t-j}^2\right) = a_0 + a_0\sum_{k=1}^{\infty} a_1^k.$$

Thus $\mathrm{E}[X_t^2] < \infty$ if and only if $a_1 < 1$ (heuristically we can see this from $\mathrm{E}[X_t^2] = \mathrm{E}[Z_2^2](a_0 + a_1\mathrm{E}[X_{t-1}^2])$).

By placing stricter conditions on a_1, namely $a_1\mathrm{E}(Z_t^{2d})^{1/d} < 1$, we can show that $\mathrm{E}[X_t^{2d}] < \infty$ (note that this is an iff condition). To see why consider the simple case d is an integer, then by using (4.9) we have

$$\begin{aligned} X_t^{2d} &\geq Z_t^{2d} a_0^d \sum_{k=1}^{\infty} a_1^{dk}\left(\prod_{j=1}^{k} Z_{t-j}^2\right)^{2d} \\ \Rightarrow \mathrm{E}[X_t^{2d}] &\geq \mathrm{E}[Z_t^{2d}] a_0^d \sum_{k=1}^{\infty} a_1^{dk} \prod_{j=1}^{k} \mathrm{E}[Z_{t-j}^{2d}] = \mathrm{E}[Z_t^{2d}] a_0^d \sum_{k=1}^{\infty} a_1^{dk} \mathrm{E}[Z_t^{2d}]^k \\ &= \mathrm{E}[Z_t^{2d}] a_0^d \sum_{k=1}^{\infty} \left(a_1^d \mathrm{E}[Z_t^{2d}]\right)^k. \end{aligned}$$

It is immediately clear the above is only finite if $a_1\mathrm{E}[Z_t^{2d}]^{1/d} < 1$.

The ARCH(p) model

We can generalize the above results to ARCH(p) processes (but to show existence of a solution we need to write the ARCH(p) process as a vector process similar to the Vector AR(1) representation of an AR(p) given in Section 2.4.1). It can be shown that under sufficient conditions on the coefficients $\{a_j\}$ that the stationary, causal solution of the ARCH(p) model is

$$\begin{aligned} X_t^2 &= a_0 Z_t^2 + \sum_{k\geq 1} m_t(k) \qquad (4.10) \\ \text{where } m_t(k) &= \sum_{j_1,\ldots,j_k\geq 1} a_0\Big(\prod_{r=1}^{k} a_{j_r}\Big)\prod_{r=0}^{k} Z^2_{t-\sum_{s=0}^{r} j_s} \quad (j_0 = 0). \end{aligned}$$

The above solution belongs to a general class of functions called a Volterra expansion. We note that $\mathrm{E}[X_t^2] < \infty$ iff $\sum_{j=1}^{p} a_j < 1$.

4.3 The GARCH model

A possible drawback of the ARCH(p) model is that the conditional variance only depends on finite number of the past squared observations/log returns (in finance, the share price is often called the return). However, when fitting the model to the data, analogous to order selection of an autoregressive model (using, say, the AIC), often a large order p is selected. This suggests that the conditional variance should involve a large (infinite?) number of past terms. This observation motivated the GARCH model (first proposed in Bollerslev (1986) and Taylor (1986)), which in many respects is analogous to the ARMA. The conditional variance of the GARCH model is a weighted average of the squared returns, the weights decline with the lag, but never go completely to zero. The GARCH class of models is a rather parsimonous class of models and is extremely popular in finance. The GARCH(p,q) model is defined as

$$X_t = \sigma_t Z_t \qquad \sigma_t^2 = a_0 + \sum_{j=1}^{p} a_j X_{t-j}^2 + \sum_{i=1}^{q} b_i \sigma_{t-i}^2 \tag{4.11}$$

where Z_t are iid random variables where $\mathrm{E}(Z_t) = 0$ and $\mathrm{var}(Z_t) = 1$, $a_0 > 0$ and for $1 \leq j \leq p$ $a_j \geq 0$ and $1 \leq i \leq q$ $b_i \geq 0$.

Under the assumption that a causal solution with sufficient moments exist, the same properties defined for the ARCH(p) in Section 4.2.1 also apply to the GARCH(p,q) model.

It can be shown that under suitable conditions on $\{b_j\}$ that X_t satisfies an ARCH(∞) representation. Formally, we can write the conditional variance σ_t^2 (assuming that a stationarity solution exists) as

$$(1 - \sum_{i=1}^{q} b_i B^i)\sigma_t^2 = (a_0 + \sum_{j=1}^{p} a_j X_{t-j}^2),$$

where B denotes the backshift notation defined in Chapter 2. Therefore if the roots of $b(z) = (1 - \sum_{i=1}^{q} b_i z^i)$ lie outside the unit circle (which is satisfied if $\sum_i b_i < 1$) then

$$\sigma_t^2 \;=\; \frac{1}{(1 - \sum_{j=1}^{q} b_j B^j)}(a_0 + \sum_{j=1}^{p} a_j X_{t-j}^2) = \alpha_0 + \sum_{j=1}^{\infty} \alpha_j X_{t-j}^2, \tag{4.12}$$

where a recursive equation for the derivation of α_j can be found in Berkes et al. (2003). In other words the GARCH(p,q) process can be written as a ARCH(∞) process. This is analogous to the invertibility representation given in Definition 2.2.2. This representation is useful when estimating the parameters of a GARCH process (see Berkes et al. (2003)) and also prediction. The expansion in (4.12) helps explain why the GARCH(p,q) process is so popular. As we stated at the start of this section, the conditional variance of the GARCH is a weighted average of the squared returns, the weights decline with the lag, but never go completely to zero, a property that is highly desirable.

Example 4.3.1 (Inverting the GARCH$(1,1)$) *If $b_1 < 1$, then we can write σ_t^2 as*

$$\sigma_t^2 = \left[\sum_{j=0}^{\infty} b^j B^j\right] \cdot \left[a_0 + a_1 X_{t-1}^2\right] = \frac{a_0}{1-b} + a_1 \sum_{j=0}^{\infty} b^j X_{t-1-j}^2.$$

This expansion offers us a clue as to why the GARCH$(1,1)$ is so popular in finance. In finance one important objective is to predict future volatility, this is the variance of say a stock tomorrow given past information. Using the GARCH model this is σ_t^2, which we see is

$$\sigma_t^2 = \frac{a_0}{1-b} + a_1 \sum_{j=0}^{\infty} b^j X_{t-1-j}^2.$$

This can be viewed as simply an exponentially weighted average of X_{t-j}^2. Some researchers argue that other models can lead to the same predictor of future volatility and there is nothing intrinsically specially about the GARCH process. We discuss this in more detail in Chapter 5.

In the following section we derive conditions for existence of the GARCH model and also it's moments.

4.3.1 Existence of a stationary solution of a GARCH$(1,1)$

We will focus on the GARCH$(1,1)$ model as this substantially simplifies the conditions. We recall the conditional variance of the GARCH$(1,1)$ can be written as

$$\sigma_t^2 = a_0 + a_1 X_{t-1}^2 + b_1 \sigma_{t-1}^2 = \left(a_1 Z_{t-1}^2 + b_1\right) \sigma_{t-1}^2 + a_0. \tag{4.13}$$

We observe that (4.13) can be written in the stochastic recurrence relation form given in (4.2) with $A_t = (a_1 Z_{t-1}^2 + b_1)$ and $B_t = a_0$. Therefore, by using Theorem 4.0.1, if $\mathrm{E}[\log(a_1 Z_{t-1}^2 + b_1)] < 0$, then σ_t^2 has the strictly stationary causal solution

$$\sigma_t^2 = a_0 + a_0 \sum_{k=1}^{\infty} \prod_{j=1}^{k} (a_1 Z_{t-j}^2 + b_1). \tag{4.14}$$

These conditions are relatively weak and depend on the distribution of Z_t. They are definitely satisfied if $a_1 + b_1 < 1$, since $\mathrm{E}[\log(a_1 Z_{t-1}^2 + b_1)] \leq \log \mathrm{E}[a_1 Z_{t-1}^2 + b_1] = \log(a_1 + b_1)$. However existence of a stationary solution does not require such a strong condition on the coefficients (and there can still exist a stationary solution if $a_1 + b_1 > 1$, so long as the distribution of Z_t^2 is such that $\mathrm{E}[\log(a_1 Z_t^2 + b_1)] < 0$).

By taking expectations of (4.14) we can see that

$$\mathrm{E}[X_t^2] = \mathrm{E}[\sigma_t^2] = a_0 + a_0 \sum_{k=1}^{\infty} \prod_{j=1}^{k} (a_1 + b_1) = a_0 + a_0 \sum_{k=1}^{\infty} (a_1 + b_1)^k.$$

Thus $\mathrm{E}[X_t^2] < \infty$ iff $a_1 + b_1 < 1$ (noting that a_1 and b_1 are both positive). Expanding on this argument, if $d > 1$ we can use Minkowski inequality to show

$$(\mathrm{E}[\sigma_t^{2d}])^{1/d} \leq a_0 + a_0 \sum_{k=1}^{\infty} (\mathrm{E}[\prod_{j=1}^{k} (a_1 Z_{t-j}^2 + b_1)]^d)^{1/d} \leq a_0 + a_0 \sum_{k=1}^{\infty} (\prod_{j=1}^{k} \mathrm{E}[(a_1 Z_{t-j}^2 + b_1)^d])^{1/d}.$$

Therefore, if $\mathrm{E}[(a_1 Z_{t-j}^2 + b_1)^d] < 1$, then $\mathrm{E}[X_t^{2d}] < \infty$. This is an iff condition, since from the definition in (4.13) we have

$$\mathrm{E}[\sigma_t^{2d}] = \mathrm{E}[\underbrace{a_0 + (a_1 Z_{t-1}^2 + b_1)\sigma_{t-1}^2}_{\text{every term is positive}}]^d \geq \mathrm{E}[(a_1 Z_{t-1}^2 + b_1)\sigma_{t-1}^2]^d = \mathrm{E}[(a_1 Z_{t-1}^2 + b_1)^d]\mathrm{E}[\sigma_{t-1}^{2d}],$$

since σ_{t-1}^2 has a causal solution, it is independent of Z_{t-1}^2. We observe that by stationarity and if $\mathrm{E}[\sigma_t^{2d}] < \infty$, then $\mathrm{E}[\sigma_t^{2d}] = \mathrm{E}[\sigma_{t-1}^{2d}]$. Thus the above inequality only holds if $\mathrm{E}[(a_1 Z_{t-1}^2 + b_1)^d] < 1$. Therefore, $\mathrm{E}[X_t^{2d}] < \infty$ iff $\mathrm{E}[(a_1 Z_{t-1}^2 + b_1)^d] < 1$.

Indeed in order for $\mathrm{E}[X_t^{2d}] < \infty$ a huge constraint needs to be placed on the parameter space of a_1 and b_1.

Exercise 4.1 *Suppose $\{Z_t\}$ are standard normal random variables. Find conditions on a_1 and b_1 such that $\mathrm{E}[X_t^4] < \infty$.*

The above results can be generalised to GARCH(p, q) model. Conditions for existence of a stationary solution hinge on the random matrix corresponding to the SRE representation of the GARCH model (see Bougerol and Picard (1992a) and Bougerol and Picard (1992b)), which are nearly impossible to verify. Sufficient and necessary conditions for both a stationary (causal) solution and second order stationarity ($\mathrm{E}[X_t^2] < \infty$) is $\sum_{j=1}^{p} a_j + \sum_{i=1}^{q} b_i < 1$. However, many econometricians believe this condition places an unreasonable constraint on the parameter space of $\{a_j\}$ and $\{b_j\}$. A large amount of research has been done on finding consistent parameter estimators under weaker conditions. Indeed, in the very interesting paper by Berkes et al. (2003) (see also Straumann (2005)) they derive consistent estimates of GARCH parameters on far milder set of conditions on $\{a_j\}$ and $\{b_i\}$ (which don't require $\mathrm{E}(X_t^2) < \infty$).

Definition 4.3.1 *The IGARCH model is a GARCH model where*

$$X_t = \sigma_t Z_t \qquad \sigma_t^2 = a_0 + \sum_{j=1}^{p} a_j X_{t-j}^2 + \sum_{i=1}^{q} b_i \sigma_{t-i}^2 \tag{4.15}$$

where the coefficients are such that $\sum_{j=1}^{p} a_j + \sum_{i=1}^{q} b_i = 1$. This is an example of a time series model which has a strictly stationary solution but it is not second order stationary.

Exercise 4.2 *Simulate realisations of ARCH(1) and GARCH(1,1) models. Simulate with both iid Gaussian and t-distribution errors ($\{Z_t\}$ where $\mathrm{E}[Z_t^2] = 1$). Remember to 'burn-in' each realisation.*

In all cases fix $a_0 > 0$. Then

(i) Simulate an ARCH(1) with $a_1 = 0.3$ and $a_1 = 0.9$.

(ii) Simulate a GARCH(1,1) with $a_1 = 0.1$ and $b_1 = 0.85$, and a GARCH(1,1) with $a_1 = 0.85$ and $b_1 = 0.1$. Compare the two behaviours.

4.3.2 Extensions of the GARCH model

One criticism of the GARCH model is that it is 'blind' to negative the sign of the return X_t. In other words, the conditional variance of X_t only takes into account the magnitude of X_t and does not depend on increases or a decreases in S_t (which corresponds to X_t being positive or negative). In contrast it is largely believed that the financial markets react differently to negative or positive X_t. The general view is that there is greater volatility/uncertainity/variation in the market when the price goes down.

This observation has motivated extensions to the GARCH, such as the EGARCH which take into account the sign of X_t. Deriving conditions for such a stationary solution to exist can be difficult task, and the reader is refered to Straumann (2005) and more the details.

Other extensions to the GARCH include an Autoregressive type model with GARCH innovations.

4.3.3 `R code`

`install.packages("tseries"), library("tseries")` recently there have been a new package developed `library("fGARCH")`.

4.4 Bilinear models

The Bilinear model was first proposed in Subba Rao (1977) and Granger and Andersen (1978) (see also Subba Rao (1981)). The general Bilinear (BL(p,q,r,s)) model is defined as

$$X_t - \sum_{j=1}^{p}\phi_j X_{t-j} = \varepsilon_t + \sum_{i=1}^{q}\theta_i\varepsilon_{t-i} + \sum_{k=1}^{r}\sum_{k'=1}^{s} b_{k,k'}X_{t-k}\varepsilon_{t-k'},$$

where $\{\varepsilon_t\}$ are iid random variables with mean zero and variance σ^2.

To motivate the Bilinear model let us consider the simplest version of the model BL$(1,0,1,1)$

$$X_t = \phi_1 X_{t-1} + b_{1,1}X_{t-1}\varepsilon_{t-1} + \varepsilon_t = [\phi_1 + b_{1,1}\varepsilon_{t-1}]X_{t-1} + \varepsilon_t. \quad (4.16)$$

Comparing (4.18) with the conditional variane of the GARCH$(1,1)$ in (4.13) we see that they are very similar, the main differences are that (a) the bilinear model does not constrain the coefficients to be positive (whereas the conditional variance requires the coefficients to be positive) (b) the ε_{t-1} depends on X_{t-1}, whereas in the GARCH$(1,1)$ Z_{t-1}^2 and σ_{t-1}^2 are independent coefficients and (c) the innovation in the GARCH$(1,1)$ model is deterministic, whereas in the innovation in the Bilinear model is random. (b) and (c) makes the analysis of the Bilinear more complicated than the GARCH model.

4.4.1 Features of the Bilinear model

In this section we assume a causal, stationary solution of the bilinear model exists, the appropriate number of moments and that it is invertible in the sense that there exists a function g such that $\varepsilon_t = g(X_{t-1}, X_{t-2}, \ldots)$.

Under the assumption that the Bilinear process is invertible we can show that

$$\begin{aligned} \mathrm{E}[X_t|X_{t-1}, X_{t-2}, \ldots] &= \mathrm{E}[(\phi_1 + b_{1,1}\varepsilon_{t-1})X_{t-1}|X_{t-1}, X_{t-2}, \ldots] + \mathrm{E}[\varepsilon_t|X_{t-1}, X_{t-2}, \ldots] \\ &= (\phi_1 + b_{1,1}\varepsilon_{t-1})X_{t-1}, \end{aligned} \quad (4.17)$$

thus unlike the autoregressive model the conditional expectation of the X_t given the past is a nonlinear function of the past. It is this nonlinearity that gives rise to the spontaneous peaks that we see a typical realisation.

To see how the bilinear model was motivated in Figure 4.7 we give a plot of

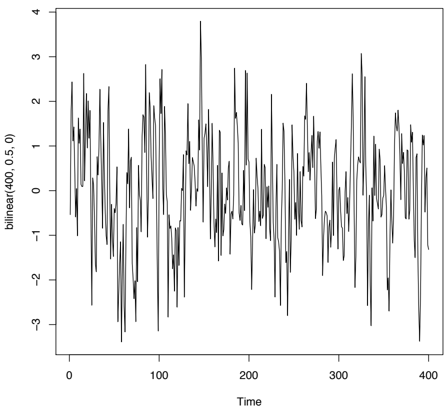
$$X_t = \phi_1 X_{t-1} + b_{1,1}X_{t-1}\varepsilon_{t-1} + \varepsilon_t, \quad (4.18)$$

where $\phi_1 = 0.5$ and $b_{1,1} = 0, 0.35, 0.65$ and -0.65. and $\{\varepsilon_t\}$ are iid standard normal random variables. We observe that Figure 4.7a is a realisation from an AR(1) process and the subsequent plots are for different values of $b_{1,1}$. Figure 4.7a is quite 'regular', whereas the sudden bursts in activity become more pronounced as $b_{1,1}$ grows (see Figures 4.7b and 4.7c). In Figure 4.7d we give a plot realisation from a model where $b_{1,1}$ is negative and we see that the fluctation has changed direction.

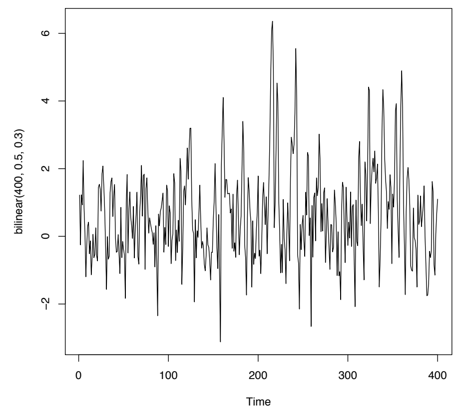
Remark 4.4.1 (Markov Bilinear model) *Some authors define the* $BL(1, 0, 1, 1)$ *as*

$$Y_t = \phi_1 Y_{t-1} + b_{1,1}Y_{t-1}\varepsilon_t + \varepsilon_t = [\phi_1 + b_{11}\varepsilon_t]Y_{t-1} + \varepsilon_t.$$

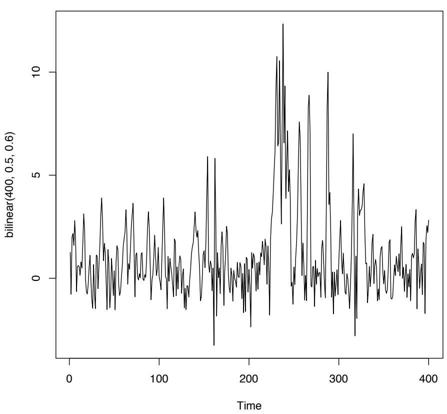
The fundamental difference between this model and (4.18) is that the multiplicative innovation (using ε_t *rather than* ε_{t-1}*) does not depend on* Y_{t-1}*. This means that* $\mathrm{E}[Y_t|Y_{t-1}, Y_{t-2}, \ldots] = \phi_1 Y_{t-1}$ *and the autocovariance function is the same as the autocovariance function of an* $AR(1)$ *model with the same AR parameter. Therefore, it is unclear the advantage of using this version of the model if the aim is to forecast, since the forecast of this model is the same as a forecast using the corresponding* $AR(1)$ *process* $X_t = \phi_1 X_{t-1} + \varepsilon_t$*. Forecasting with this model does not take into*



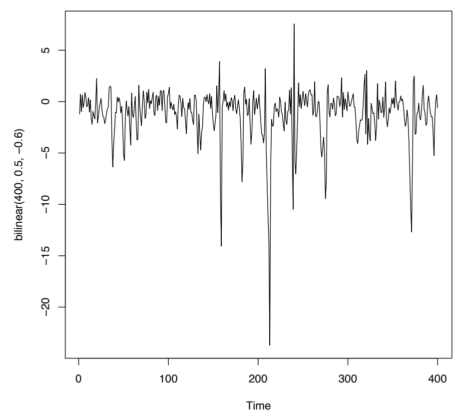
(a) $\phi_1 = 0.5$ and $b_{1,1} = 0$



(b) $\phi_1 = 0.5$ and $b_{1,1} = 0.35$



(c) $\phi_1 = 0.5$ and $b_{1,1} = 0.65$



(d) $\phi_1 = 0.5$ and $b_{1,1} = -0.65$

Figure 4.7: Realisations from different BL$(1, 0, 1, 1)$ models

account its nonlinear behaviour.

4.4.2 Solution of the Bilinear model

We observe that (4.18) can be written in the stochastic recurrence relation form given in (4.2) with $A_t = (\phi_1 + b_{11}\varepsilon_{t-1})$ and $B_t = a_0$. Therefore, by using Theorem 4.0.1, if $\mathrm{E}[\log(\phi_1 + b_{11}\varepsilon_{t-1})] < 0$ and $\mathrm{E}[\varepsilon_t] < \infty$, then X_t has the strictly stationary, causal solution

$$X_t = \sum_{k=1}^{\infty} \left[\prod_{j=1}^{k-1} (\phi_1 + b_{1,1}\varepsilon_{t-j}) \right] \cdot [(\phi_1 + b_{1,1}\varepsilon_{t-k})\varepsilon_{t-k}] + \varepsilon_t. \tag{4.19}$$

To show that it is second order stationary we require that $\mathrm{E}[X_t^2] < \infty$, which imposes additional conditions on the parameters. To derive conditions for $\mathrm{E}[X_t^2]$ we use (4.20) and the Minkowski inequality to give

$$
\begin{aligned}
(\mathrm{E}[X_t^2])^{1/2} &\leq \sum_{k=1}^{\infty} \mathrm{E}\left(\left[\prod_{j=1}^{k-1}(\phi_1 + b_{1,1}\varepsilon_{t-j})\right]^2\right)^{1/2} \cdot \left(\mathrm{E}\left[(\phi_1 + b_{11}\varepsilon_{t-k})\varepsilon_{t-k}\right]^2\right)^{1/2} \\
&= \sum_{k=1}^{\infty}\prod_{j=1}^{k-1} \mathrm{E}\left(\left[(\phi_1 + b_{1,1}\varepsilon_{t-j})\right]^2\right)^{1/2} \cdot \left(\mathrm{E}\left[(\phi_1 + b_{1,1}\varepsilon_{t-k})\varepsilon_{t-k}\right]^2\right)^{1/2}. \qquad (4.20)
\end{aligned}
$$

Therefore if $\mathrm{E}[\varepsilon_t^4] < \infty$ and

$$
\mathrm{E}\left[(\phi_1 + b_{1,1}\varepsilon_t)\right]^2 = \phi^2 + b_{11}^2 \mathrm{var}(\varepsilon_t) < 1,
$$

then $\mathrm{E}[X_t^2] < \infty$ (note that the above equality is due to $\mathrm{E}[\varepsilon_t] = 0$).

Remark 4.4.2 (Inverting the Bilinear model) *We note that*

$$
\varepsilon_t = -(bX_{t-1})\varepsilon_{t-1} + [X_t - \phi X_{t-1}],
$$

thus by iterating backwards with respect to ε_{t-j} *we have*

$$
\varepsilon_t = \sum_{j=0}^{\infty}\left((-b)^{j-1}\prod_{i=0}^{j} X_{t-1-j}\right)[X_{t-j} - \phi X_{t-j-1}].
$$

This invertible representation is useful both in forecasting and estimation (see Section 5.5.3).

Exercise 4.3 *Simulate the* $BL(2,0,1,1)$ *model (using the* $AR(2)$ *parameters* $\phi_1 = 1.5$ *and* $\phi_2 = -0.75$*). Experiment with different parameters to give different types of behaviour.*

Exercise 4.4 *The random coefficient AR model is a nonlinear time series proposed by Barry Quinn (see Nicholls and Quinn (1982) and Aue et al. (2006)). The random coefficient* $AR(1)$ *model is defined as*

$$
X_t = (\phi + \eta_t)X_{t-1} + \varepsilon_t
$$

where $\{\varepsilon_t\}$ *and* $\{\eta_t\}$ *are iid random variables which are independent of each other.*

(i) *State sufficient conditions which ensure that* $\{X_t\}$ *has a strictly stationary solution.*

(ii) State conditions which ensure that $\{X_t\}$ is second order stationary.

(iii) Simulate from this model for different ϕ and $\text{var}[\eta_t]$.

4.4.3 `R code`

Code to simulate a $\text{BL}(1, 0, 1, 1)$ model:

```
# Bilinear Simulation
# Bilinear(1,0,1,1) model, we use the first n0 observations are burn-in
# in order to get close to the stationary solution.
bilinear <- function(n,phi,b,n0=400) {
y <- rnorm(n+n0)
w <- rnorm(n + n0)
for (t in 2:(n+n0)) {
y[t] <- phi * y[t-1] + b * w[t-1] * y[t-1] + w[t]
}
return(y[(n0+1):(n0+n)])
}
```

4.5 Nonparametric time series models

Many researchers argue that fitting parametric models can lead to misspecification and argue that it may be more realistic to fit nonparametric or semi-parametric time series models instead. There exists several nonstationary and semi-parametric time series (see Fan and Yao (2003) and Douc et al. (2014) for a comprehensive summary), we give a few examples below. The most general nonparametric model is

$$X_t = m(X_{t-1}, \ldots, X_{t-p}, \varepsilon_t),$$

but this is so general it looses all meaning, especially if the need is to predict. A slight restriction is make the innovation term additive (see Jones (1978))

$$X_t = m(X_{t-1}, \ldots, X_{t-p}) + \varepsilon_t,$$