Ryan Johnston

1 Introduction

Although my project proposal stated that I would be working on aggregated time series models, there is a considerable amount of information and background needed before getting there. Thus, this final report consists mostly of "setting the table" or giving the proper background information to understand the models I will be getting into. This part of my paper consists of showing recursive substitution, working with the backshift operator, defining what a difference equation is, and what an autoregressive process is. I am also aware that there may be some unclear statements since I continue to work with my advisor on this matter.

2 Autoregressive Process

Time series that are encountered in practice are often modeled and well approximated by the autoregressive time series. The *pth* order autoregressive time series (AR(p)) is given by the following equation

$$\sum_{i=0}^{p} \rho_i X_{t-i} = \varepsilon_t, \qquad t = 0, \pm 1, \pm 2, \ldots$$

where $\rho_0 \neq 0, \rho_p \neq 0$ and the ε_t are typically assumed to be uncorrelated $(0, \sigma^2)$ random variables.

In this project we will consider and explore the first order autoregressive time series (AR(1)) process most commonly represented in the form,

$$X_t = \rho X_{t-1} + \varepsilon_t \qquad t = 0, \pm 1, \pm 2, \ldots$$

Before we begin exploring the AR(1) process however, we will first want to establish some building blocks to better understand what the AR(1) process is.

3 Difference Equations

A *difference equation* is an expression relating a variable X_t to its previous values. For example, say we want to know the value that X takes on at time t denoted by X_t. A difference equation gives us an equation directly relating the value X at t to the value of X at a previous period, say $t-1$, plus another variable ε_t:

$$X_t = \rho X_{t-1} + \varepsilon_t.$$

The ε_t are often referred to as the error terms and generally make up the variability that is generated in the random variables from one time period to the next. We will also further explore the constant ρ shortly. We can easily see from the formula above that the AR(1) process is simply a first order difference equation, and upon further inspection we observe that we can solve the equation by recursive substitution

$$\begin{aligned}
X_t =& \rho X_{t-1} + \varepsilon_t \\
=& \rho(\rho X_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\
=& \rho^2 X_{t-2} + \rho\varepsilon_{t-1} + \varepsilon_t \\
=& \rho^3 X_{t-3} + \rho^2\varepsilon_{t-2} + \rho\varepsilon_{t-1} + \varepsilon_t \\
&\vdots \\
=& \rho^{t+1} X_{-1} + \rho^t\varepsilon_0 + \rho^{t-1}\varepsilon_1 + \cdots + \rho\varepsilon_{t-1} + \varepsilon_t,
\end{aligned}$$

where X_0 is the initial value. This of course can be generalized to the case where the initial value is X_{t-N}. For example, say we want to know the value of X at time t and we know the value of X N periods prior (X_{t-N}). Recursive substitution would give

$$\begin{aligned}
X_t =& \rho X_{t-1} + \varepsilon_t \\
=& \rho(\rho X_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\
=& \rho^2 X_{t-2} + \rho\varepsilon_{t-1} + \varepsilon_t \\
=& \rho^3 X_{t-3} + \rho^2\varepsilon_{t-2} + \rho\varepsilon_{t-1} + \varepsilon_t \\
&\vdots \\
=& \rho^N X_{t-N} + \rho^{N-1}\varepsilon_{t-(N-1)} + \cdots + \rho\varepsilon_{t-1} + \varepsilon_t.
\end{aligned}$$

Now we come back to the constant ρ. We observe that the effect of $\varepsilon_{t-(N-1)}$ on X_t is ρ^{N-1}. Thus for $|\rho| < 1$, ρ^{N-1} geometrically goes to zero. Whereas when $|\rho| > 1$, ρ^{N-1} grows exponentially over time. Thus, if $|\rho| < 1$ the system is considered stable, in other words the further back in time that a given change occurs the less it will effect the present. The given change will eventually die out over time. For $|\rho| > 1$ the system blows up. A given change from the past increasingly effects the future as time goes on. It is probably intiuitively obvious that we want a system that is effected less the further we go in the past. Thus,throughout this paper we will use the assumption that $|\rho| < 1$.

4 Backshift operator

A useful operator often used in time series is known as the backshift operator denoted by the symbol L. Let $\{a_n\}_{n=-\infty}^{\infty}$ be a sequence. By applying the backshift operator we obtain a sequence $\{a_{n-1}\}_{n=-\infty}^{\infty}$:

$$La_n = a_{n-1}.$$

noindent Applying the backshift operator twice would give:

$$L(La_n) = La_{n-1} = a_{n-2}.$$

Applying the backshift operator twice is denoted by "L^2", therefore, we could rewrite the above equation as

$$L^2 a_n = a_{n-2}.$$

In general shifting the time series back k steps is represented by

$$L^k a_n = a_{n-k}.$$

We now return to our difference equation:

$$X_t = \rho X_{t-1} + \varepsilon_t$$

and observe this can be rewritten using the backshift operator:

$$X_t = \rho L X_t + \varepsilon_t$$

using a little algebra we obtain the equations:

$$\begin{aligned} X_t - \rho L X_t &= \varepsilon_t \\ \text{or } (1-\rho L)X_t &= \varepsilon_t. \\ \Rightarrow X_t &= \frac{1}{1-\rho L}\varepsilon_t. \end{aligned}$$

Now lets apply the operation $(1 + \rho L + \rho^2 L^2 + \cdots + \rho^{N-1}L^{N-1})$ to both sides of the equation to get

$$(1 + \rho L + \rho^2 L^2 + \cdots + \rho^{N-1}L^{N-1})(1-\rho L) = (1 + \rho L + \rho^2 L^2 + \cdots + \rho^{N-1}L^{N-1})\varepsilon_t.$$

We now expand the left side of the equation

$$\begin{aligned} (1+&\rho L + \rho^2 L^2 + \cdots + \rho^{N-1}L^{N-1})(1-\rho L) \\ =&(1 + \rho L + \rho^2 L^2 + \cdots + \rho^{N-1}L^{N-1}) \\ &- (\rho L + \rho^2 L^2 + \cdots + \rho^N L^N) \\ =&(1 - \rho^N L^N) \end{aligned}$$

and substitute it back in to get

$$\begin{aligned} (1 - \rho^N L^N)X_t =&(1 + \rho L + \rho^2 L^2 + \cdots + \rho^{N-1}L^{N-1})\varepsilon_t \\ \Rightarrow X_t =&\rho^N X_{t-N} + \rho^{N-1}\varepsilon_{t-(N-1)} + \cdots + \rho\varepsilon_{t-1} + \varepsilon_t. \end{aligned}$$

Note that this equation is exactly the same equation obtained using recursive substitution.

5 AR(1) Process

As noted previously the AR(1) process satisfies the difference equation

$$X_t = \rho X_{t-1} + \varepsilon_t \qquad\qquad t = 0, \pm 1, \pm 2, \ldots$$

In order to better work with this equation we would like to find a unique solution for X_t that is both stationary and a function of the error terms. We'll first need to start with some definitions.

Let (Ω, F, P) be a probability space and let T be an index set. A time series can be considered a collection of random variables $\{X_t : t \in T\}$. The joint distribution function of a finite set of random variables $(X_{t_1}, X_{t_2}, \ldots, X_{t_k})$ from the collection $\{X_t : t \in T\}$ is defined by

$$F_{X_{t_1}, X_{t_2}, \ldots, X_{t_k}}(x_{t_1}, x_{t_2}, \ldots, x_{t_k}) = P\{\omega : X_{t_1}(\omega) \leq x_{t_1}, \ldots, X_{t_k}(\omega) \leq x_{t_k}\}.$$

Definition 1 (Stict Stationarity): *The time series is said to be strict stationary if the joint distribution of* $(X_{t_1}, X_{t_2}, \ldots, X_{t_k})$ *is the same as that of* $(X_{t_{1+h}}, X_{t_{2+h}}, \ldots, X_{t_{k+h}})$*, that is if*

$$F_{X_{t_1},X_{t_2},\ldots,X_{t_k}}(x_{t_1}, x_{t_2}, \ldots, x_{t_k}) = F_{X_{t_1},X_{t_2},\ldots,X_{t_k}}(x_{t_1}, x_{t_2}, \ldots, x_{t_k})$$

In other words, the joint distribution only depends on the difference h and not on the time $(t_1, \ldots, t_k)$.

Definition 2 (Weak Stationarity): *The time series* $\{X_t : t \in T\}$ *is said to be weakly stationary if*

$$\begin{aligned} &(i)\ \ E[X_t] = C \ \text{ with some constant } \ C \\ &(ii)\ \ \mathrm{Cov}(X_t, X_{t+h}) \ \text{ does not depend on } \ t. \end{aligned}$$

Definition 3 (Almost Sure Convergence): *Let* X_n *be a sequence of random variables and let* X *be a random variable. Let* $(\Omega,\ F,\ P)$ *be a probability space where* X_n *and* X *are defined. We say that the sequence* X_n *converges almost surely or with probability 1 if*

$$P(\lim_{n\to\infty} X_n = X) = 1 \ \text{ or equivalently } \quad P(\{\omega \in \Omega \mid \lim_{n\to\infty} X_n(\omega) = X(\omega)\}) = 1.$$

Essentially this means that we are almost guaranteed that the values of X_n approach (almost surely) the value of X in the sense that the events for which X_n does not converge to X have probability zero. An example to illustrate this would be to suppose we are throwing a dart at a unit square. We could divide this sqaure into two subregions by drawing a vertical line down the middle. The probability of hitting either subregion is just equal to the area of that subregion. The area of the line that divides the regions is zero, so the probability that the dart lands on the line is zero. The possibility, however, that the dart lands on the line does exist. In essence, as the number of throws of the dart tends to infinity, the ratio of times that the line is hit tends to zero so the probability of hitting one of the two regions is one.

Lemma 1: *If* $|\rho| > 1$ *and* $E(|\varepsilon_i|) \le C$ *then*

$$\sum_{i=0}^{\infty} |\rho|^i |\varepsilon_{t-i}|$$

converges almost surely.

Proof: Since $\sum_{i=1}^{N} |\rho|^i |\varepsilon_{t-i}|$ is monotone, then

$$\lim_{N\to\infty} \sum_{i=1}^{N} |\rho|^i |\varepsilon_{t-i}| \to \xi \text{ almost surely}$$

where ξ can either be finite or ∞.

Next we show that

$$\begin{aligned}
E(\sum_{i=1}^{N}|\rho|^i|\varepsilon_{t-i}|) &= \sum_{i=1}^{N}|\rho|^i E(|\varepsilon_{t-i}|) \\
&\leq C\sum_{i=1}^{N}|\rho|^i \\
&\leq C\sum_{i=1}^{\infty}|\rho|^i \\
&= \frac{C}{1-|\rho|}
\end{aligned}$$

and by Fatou's Lemma

$$\begin{aligned}
E(\liminf_{N\to\infty}\sum_{i=1}^{N}|\rho|^i|\varepsilon_{t-i}|) &\leq \liminf_{N\to\infty} E(\sum_{i=1}^{N}|\rho|^i|\varepsilon_{t-i}|) \\
&\leq \frac{C}{1-|\rho|}
\end{aligned}$$

We are able to finish the proof.

Lemma 2: *if $\{X_t\}$ is strictly stationary and $E|X_0|^\nu < \infty$ for some $\nu > 0$ then*

$$\rho^N X_{t-N} \to 0 \text{ almost surely.}$$

To show the proof of this we'll need the to use the following lemma and inequality.

Borel-Cantelli Lemma: *if for all $\epsilon > 0$*

$$\sum_n P(|X_n - X| > \epsilon) < \infty \Rightarrow X_n \to X \text{ almost surely}$$

Markov's Inequality: *Given $Y \geq 0$*

$$P(Y \geq x) \leq \frac{E(Y)}{x}$$

Proof of lemma 2*:*

$$\begin{aligned}
P(|\rho^N X_{t-N}| > \epsilon) &= P(|X_{t-N}| > \epsilon |rho|^{-N}) \\
&= P(|X_{t-N}|^\nu > \epsilon^\nu |\rho^{-N\nu} \\
&= P(|X_0|^\nu > \epsilon^\nu |\rho|^{-N\nu}) \textit{ by strict stationarity} \\
&\leq \frac{E(|X_0|^\nu)}{\epsilon^\nu} |\rho|^{N\nu} \textit{ by Markov's Inequality} \\
&\Rightarrow \sum_{N=1}^{\infty} P(|\rho^N X_{t-N}| > \epsilon) < \infty \\
&\Rightarrow |\rho^N X_{t-N}| \to 0 \textit{ almost surely}
\end{aligned}$$