INTRODUCTION TO LONG MEMORY TIME SERIES

Although most economic time series are nonstationary and do require differencing of some kind, it is not necessarily true that taking first differences and then using an ARMA model will be the best remedy. In Box-Jenkins analysis, it is assumed that if the series is nonstationary, the first difference will be well behaved, as long as there are no seasonal components. In particular, it is hoped that $\Delta x_t$ will have rapidly decaying autocorrelations and be free of trend-like behavior, so that it can be well described by a stationary invertible ARMA model. This is not always the case.

Consider, for example, the interest rates on Treasury Bills (monthly) from Jan 1983 to July 1992. The raw series is clearly nonstationary: the interest rates tend to wander, and the sample autocorrelations decay very slowly. The periodogram is completely dominated by the low frequency components. The first differences are better behaved, in that the sample autocorrelations decay more rapidly than for the raw data, but there still seems to be some trend-like behavior. A look at the periodogram reveals that most of the power is still concentrated near zero frequency. This is consistent with the trend-like behavior of the time domain plot. Differencing the series a second time seems to go too far in removing the trend-like behavior. The periodogram of $\Delta^2 x_t$ shows almost no power near zero frequency, indicating that the low frequency components have not merely been attenuated; they have virtually been annihilated. This will make long-term forecasting extremely difficult. The low spectral power near zero frequency indicates that $\Delta^2 x_t$ may be non-invertible, a sign that the data may have been over-differenced. Somehow, it seems that the "right" number of differences ($d$) is somewhere between 1 and 2. The long memory models provide us with a way to define such a fractional difference, and would provide a useful alternative to using an ARMA model for the first difference of the data in this example.

We will say that a stationary time series $\{x_t\}$ has long memory if there is a nonzero $d \in (-.5,.5)$ such that the spectral density obeys a power law, $f(\lambda) \sim k \lambda^{-2d}$ as $\lambda \to 0^+$. Thus, as $\lambda \to 0$, $f(\lambda)$ tends either to $\infty$ (if $d > 0$) or to zero (if $d < 0$). If $d = 0$, we say that $\{x_t\}$ has short memory. In this case, $f(0)$ will be positive and finite. All stationary invertible ARMA processes are short memory.
Another phenomenon indicating the inadequacy of ARMA models for some stationary time series and showing the need for long memory models is the failure of the central limit theorem. If $x_0, \ldots, x_{n-1}$ are iid with finite variance, then the sample mean $\bar{x}$ is asymptotically normal, and has a variance proportional to $1/n$. Standard statistical inference (confidence intervals, hypothesis tests, etc.) for the mean $\mu$ is based on this result. The CLT will still hold even if the observations are not iid, as long as the autocovariances decay rapidly. Examples are provided by the stationary invertible ARMA processes, all of which have autocovariance functions which decay exponentially fast to zero. It has been found, however, that certain observed time series, although apparently stationary, seem to violate the central limit theorem in that the variance of $\bar{x}$ seems to go to zero more slowly than $1/n$. In fact, it is common for a graph of $\log(\text{var } \bar{x})$ versus $\log n$ to be reasonably linear, but with a slope between 0 and $-1$. A famous example is the yearly ranges in the level of the Nile River, for which the variance of $\bar{x}$ seems to decrease approximately as $n^{-2}$. Mandelbrot and Wallis (1969) found similar behavior in a variety of geophysical time series including rainfall, earthquake frequencies and sunspot numbers. If indeed the variance of $\bar{x}$ goes to zero more slowly than $1/n$, then inferences about the mean $\mu$ of the series would be wrong (anti-conservative) by orders of magnitude: we will thank that $\mu$ is known much more precisely than what can actually be supported by the data.

Further empirical study of this phenomenon revealed that, although none of the autocorrelations was particularly large when considered individually, there seemed to be a very large number of relatively small contributions, even at very long lags, which were working together to break down the assumptions needed for the central limit theorem to hold.

A theoretical analysis shows that if the autocovariances decay to zero slowly enough, then the central limit theorem can indeed fail. Suppose that $c_r \sim k r^{2d-1}$ for some nonzero $d \in (-.5,.5)$, so that there is non-negligible correlation even between distant past and distant future. It can be shown that this condition is equivalent to the frequency-domain definition of a long memory series given earlier. (The only exception is for $d = 0$, in which case the frequency domain definition is more reasonable.) Then, as we now show, $\text{var } \bar{x} \sim k_1 n^{2d-1}$, where $k_1$ is a constant. To prove this, we will use the fact that the periodogram $I(\omega)$ and the sample autocovariances $\hat{c}_r$ are related by
Thus, at zero frequency we have

\[ I(0) = \frac{1}{2\pi n} \sum_{r=0}^{n-1} x_r^2 = \frac{1}{2\pi} \sum_{|r| \leq n} \hat{c}_r. \]

Therefore, assuming for simplicity that \( E[x_r] = 0 \), we have

\[ \text{var } \bar{x} = E[\bar{x}^2] = \frac{1}{n} \sum_{|r| \leq n} E[\hat{c}_r] = \frac{1}{n} \sum_{|r| \leq n} (1 - |r|/n) c_r \]

\[ -\frac{2k}{n} \sum_{r=1}^{n-1} (1 - r/n) r^{2d-1} = -n^{2d-1} \frac{2k}{n} \sum_{r=1}^{n-1} (1 - r/n) (r/n)^{2d-1} \]

\[ \sim n^{2d-1} 2k \int_0^1 (1-x)x^{2d-1}dx. \]

Thus, \( \text{var } \bar{x} = k_1 n^{2d-1} \), where \( k_1 = 2k \int_0^1 (1-x)x^{2d-1}dx \) is a constant.

In the case \( d > 0 \), the autocovariances decay to zero so slowly that they are not summable, i.e. \( \sum |c_r| = \infty \), and \( \text{var } \bar{x} \) decays to zero more slowly than \( 1/n \). If \( d < 0 \), then the autocovariances are summable, \( \sum |c_r| < \infty \), but they still decay to zero more slowly than the exponential rate achieved by stationary invertible ARMA processes. In this case, \( \text{var } \bar{x} \) goes to zero more quickly than \( 1/n \). For this reason, some authors refer to the case \( d < 0 \) as intermediate memory, reserving the term long memory for the case \( d > 0 \). We will follow the terminology of Brockwell and Davis, however, and use "long memory" whenever \( d \neq 0 \).

One of the first proposers of long memory models for time series was D.R. Cox, who used the models to describe the variation in yarn diameters in his early consulting work for the textile industry. The models were originally invented by Kolmogorov, and widely popularized by Mandelbrot because of their connection with fractals. We will describe this in more detail later, but we note here that it has been observed that stock market fluctuations look more or less the same whether they are observed every week, every day, every hour or every minute. That is, the structure at every level of magnification
is similar. More recently, Hosking established the relationship between long memory and fractional differencing. He proposed a class of models called **fractional ARIMA**, in which the degree of differencing can be any real number. The simplest of these models is called **fractionally integrated noise**, also known as the fractional ARIMA (0,d,0) model, and is described in the next section.