THREE IMPORTANT CONCEPTS IN TIME SERIES ANALYSIS: STATIONARITY, CROSSING RATES, AND THE WOLD REPRESENTATION THEOREM

Prof. Thomas B. Fomby
Department of Economics
Southern Methodist University
June 2008

I. Definition of Covariance Stationarity

A very important concept in time series analysis is the concept of the stationarity of a time series. Consider the time series process $y_t, t = 1,2,\cdots,T$. This process is called **covariance stationary** (or simply stationary) if the following conditions hold:

(i) $E(y_t) = \mu$ for all t

(ii) $E(y_t - \mu)^2 = Var(y_t) = \sigma_y^2$ for all t (1)

(iii) $E(y_t - \mu)(y_{t-j} - \mu) = Cov(y_t, y_{t-j}) = \gamma_j$ for all t and a given value of $j \in \{1,2,\cdots\}$

Condition (i) states that the time series y_t has a constant mean irrespective of the time period at which it is observed. Condition (ii) states that y_t has a constant variance irrespective of the time period at which it is observed. Condition (iii) states that the covariance between observations y_t and y_{t-j} is only a function of j and not t. The covariance between two observations that are j-periods apart is constant no matter when the j-period apart observations are observed.

For example consider the so-called AR(1) model

$$y_t = \phi_0 + \phi_1 y_{t-1} + a_t \qquad (2)$$

where $|\phi_1| < 1$ is assumed and the error term a_t has the so-called "**white noise**" properties

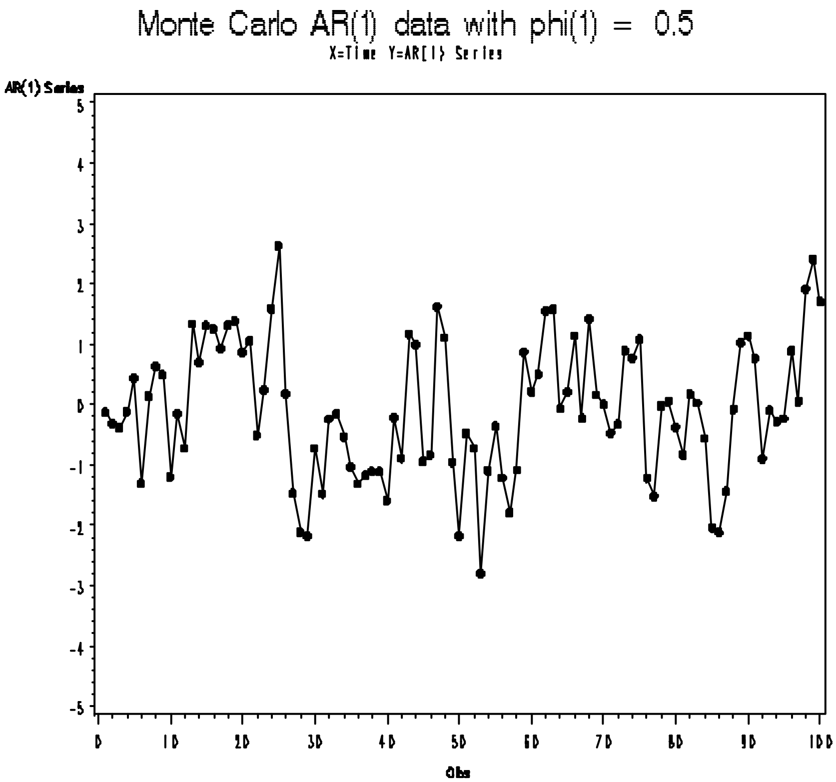
(i) $E(a_t) = 0$ for all t

(ii) $E(a_t^2) = \sigma_a^2$ for all t (3)

(iii) $E(a_s a_t) = 0$ for all $s \neq t$.

Thus the error term is white noise because it has a constant zero mean, a constant variance, σ_a^2, and the errors are uncorrelated with each other as long as their time subscripts do not match. A **Monte Carlo** realization of this AR(1) process for $\phi_0 = 0.0$, $\phi_1 = 0.5$, $\sigma_a^2 = 1.0$, and T = 100 is reproduced below in Figure 1.

Figure 1

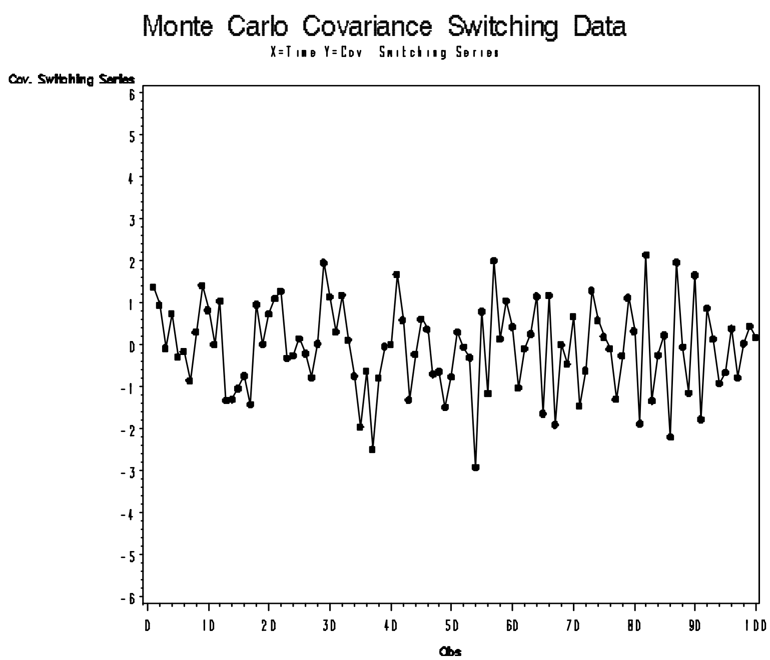


For the AR(1) process of (2) we show in the Appendix that $E(y_t) = \dfrac{\phi_0}{1-\phi_1}$,

$Var(y_t) = \dfrac{\sigma_a^2}{1-\phi_1^2}$, and $Cov(y_t, y_{t-j}) = \dfrac{\phi_1^j \sigma_a^2}{1-\phi_a^2}$. For the above Monte Carlo data this implies $E(y_t) = 0$, $Var(y_t) = \dfrac{1}{1-0.5^2} = 1.33$, and $Cov(y_t, y_{t-j}) = 0.5^j (1.33)$. Visually, one can see the constancy of the mean and possibly the constancy of the variance but it is difficult to visually gauge the constancy of the covariance as prescribed by model (2). Nevertheless, the data does seem to have a gentle rolling nature that seems to repeat itself as would be implied by the positive covariances of the series.

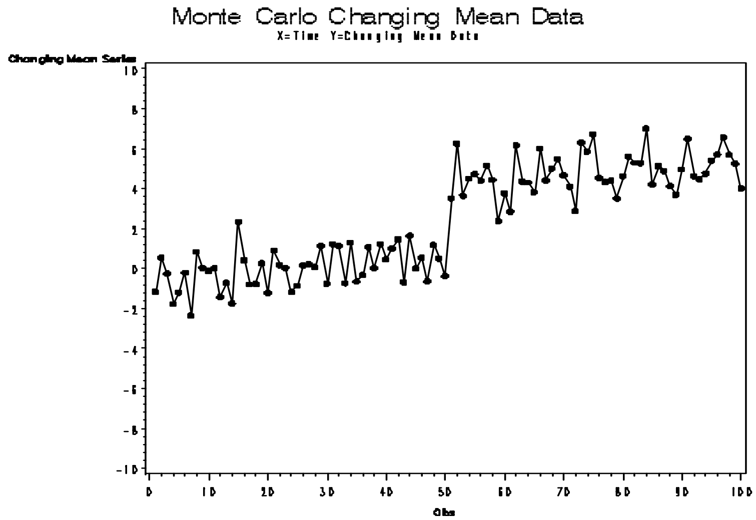
The above time series is to be compared to a graph where for t = 1 to 50 the model is the AR(1) model $y_t = 0.8y_{t-1} + a_t$ whereas for t = 51 to 100 the AR(1) model is $y_t = -0.8y_{t-1} + a_t$. See Figure 2 below. The means of both "half" processes are equal to zero and their variances are both 1/0.36 = 2.778 while their covariances are different in sign but equal in magnitude for odd values of j but equal to each other for even values of j. For example, for the first part of the data $\text{cov}(y_t, y_{t-1}) = 0.8(2.778)$ while for the second half of the data, $\text{cov}(y_t, y_{t-1}) = -0.8(2.778)$. That is, for the first half of the data, the covariance of the observations one period apart is positive. When last period's observation is **above the mean**, this period's observation tends also to be **above the mean** and when last period's observation is **below the mean**, this period's observation tends to be **below the mean**. In contrast, for the second half of the data, the covariance of the observations one period apart is negative. When last period's observation is **above the mean**, this period's observation tends also to be **below the mean** and when last period's observation is **below the mean**, this period's observation tends to be **above the mean**. This covariance-schizophrenic behavior is reflected in the slow-rolling nature of the first half of the data and the "saw-tooth" pattern of the data in the second half. Obviously this data is not covariance stationary although the data has a constant mean and a constant variance. It just doesn't have a constant covariance function.

Figure 2



Another case that turns up sometimes is the "Changing Mean" case as presented in Figure 3 below.

Figure 3



In this case the data must also be split into two groups with the groups being modeled separately.

One of the substantial benefits of having a stationary time series is that **with one realization of the time series** (which is usually all we ever have!), one can **consistently** estimate the population mean, variance, and covariances of the time series process using the following sample statistics:

$$\bar{y} = \frac{\sum_{t=1}^{T} y_t}{T} \quad \textbf{(sample mean)} \tag{4}$$

$$s_y^2 = \frac{\sum_{t=1}^{T} (y_t - \bar{y})^2}{T} \quad \textbf{(sample variance)} \tag{5}$$

$$c_j = \frac{\sum_{t=j+1}^{T} (y_t - \bar{y})(y_{t-j} - \bar{y})}{T} \quad \textbf{(sample covariance)} \tag{6}$$

$$r_j = \frac{c_j}{s_y^2} = \frac{\sum_{t=j+1}^{T} (y_t - \bar{y})(y_{t-j} - \bar{y})}{\sum_{t=1}^{T} (y_t - \bar{y})^2}, \; j = 1, 2, \cdots. \tag{7}$$

(sample autocorrelation) .

The last sample statistic, r_j, is the sample statistic for estimating the population autocorrelation at lag j, namely, $\rho_j = Cov(y_t, y_{t-j})/Var(y_t)$. The display of all of the autocorrelations ρ_j on the y-axis with the lag j on the x-axis of a graph constitutes what is called the **population autocorrelation function**. The display of all of the sample autocorrelations r_j on the y-axis with the lag j on the x-axis of a graph constitutes what is called the **sample autocorrelation function**.

By the term "consistency" we mean that, as the number of observations goes to infinity, the following probability statements hold:

$$\Pr ob(|\bar{y} - \mu|) < \varepsilon) \to 1 \text{ as } T \to \infty \quad (8)$$

$$\Pr ob(|s_y^2 - \sigma_y^2|) < \delta) \to 1 \text{ as } T \to \infty \quad (9)$$

$$\Pr ob(|c_j - \gamma_j|) < \tau) \to 1 \text{ as } T \to \infty \quad (10)$$

$$\Pr ob(|r_j - \rho_j|) < \kappa) \to 1 \text{ as } T \to \infty \quad (11)$$

for arbitrarily small values of $\varepsilon > 0, \delta > 0, \tau > 0$ and $\kappa > 0$. Thus, in reality the conditions (8) – (11) mean that as the sample size on a stationary time series goes to infinity, the exact value of the population parameters, $\mu, \sigma_y^2, \gamma_j$ and ρ_j will become known with probability one. But for sample sizes that are "large" the probability will be high that the corresponding sample statistics will be in a small neighborhood of the actual population values. Determining these population parameters with consistency is important because they play a crucial part in providing forecasting formulas that yield accurate forecasts.

II. Transforming Non-Stationary Time Series to Stationary Form

Time series that have non-constant means, non-constant variances, and non-constant covariances are exceeding difficult to forecast. However, as we will see, a non-stationary time series can often be transformed to a stationary time series, say by taking the natural logarithm of the data and/or differencing the data. In the above non-constant covariance case represented in Figure 2 above, all we have to do is split the data into two parts, part I being observations 1 – 50 and part II being observations 51 – 100 and then treating them separately as two distinct stationary time series. The first part of course would be of interest to the economic or business historian while the second part would be more relevant for the forecaster. The following three subsections will discuss three **stationarity transformations** that often used to transform non-stationary time series into stationary ones. The first two transformations are often recommended in **the Box-Jenkins approach** to building time series forecasting models.

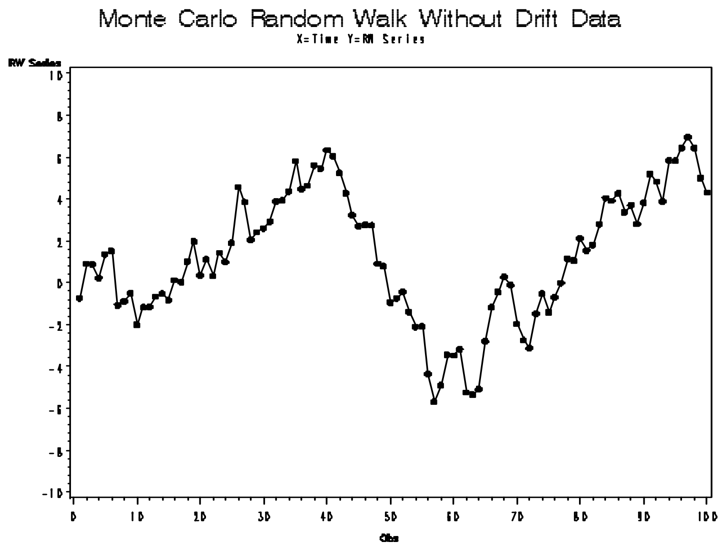
Differencing as a Solution

A very interesting special case of the first order autoregressive (AR(1)) model of equation (2) is when the first-order autoregressive coefficient is equal one $(\phi_1 = 1)$. Consider the Random Walk without drift model

$$y_t = y_{t-1} + a_t \tag{12}$$

where we let $y_0 = 0$. A Monte Carlo realization of this process is plotted below in Figure 4.

Figure 4



By repeated substitution it is easy to show that (12) yields

$$y_t = a_t + a_{t-1} + \cdots + a_1 + y_0 \tag{13}$$

where y_0 is fixed at zero. It follows that

$$E(y_t) = E(a_t + a_{t-1} + \cdots + a_1)$$

$$= E(a_t) + E(a_{t-1}) + \cdots + E(a_1) = 0 \ , \tag{14}$$

and

$$Var(y_t) = Var(a_t + a_{t-1} + \cdots + a_1)$$

$$= Var(a_t) + Var(a_{t-1}) + \cdots + Var(a_1) = t\sigma_a^2 , \tag{15}$$

the next-to-last equality following from the independence of the a_t's. Also

$$Cov(y_t, y_{t-j}) = E(a_t + \cdots + a_1)(a_{t-j} + \cdots + a_1)$$

$$= E(a_{t-j}^2) + E(a_{t-j-1}^2) + \cdots E(a_1^2) = (t-j)\sigma_a^2 \quad (16)$$

the next-to-last equality following from the independence of the $a_t's$. Finally it follows that the autocorrelation at lag j is given by

$$\rho_j = Corr(y_t, y_{t-j}) = Cov(y_t, y_{t-j}) / Var(y_t)$$

$$= (t-j)\sigma_a^2 / t\sigma_a^2 = \frac{t-j}{t} \ . \quad (17)$$

Obviously, the Random Walk without drift process (12) is non-stationary. Although the process has a constant mean of zero, its variances is ever-changing from one time period to the next and, in fact, it approaches infinity as the sample size goes to infinity. (See (15).) Moreover, the covariance at lag j is a function of time as well. (See (16).) Also for a fixed lag j, the correlation between observations that are j periods apart approaches one as the sample size goes to infinity. This is rather peculiar behavior but that is what is implied by this random walk (unit root) case. For a **fixed sample size** it should be noted that the autocorrelation function will be very slowly damping starting at 1 for j = 0 and only slowly approaching zero at the rate of (1/t) as j goes to infinity. In fact, one might expect the sample autocorrelation function to behave similarly when the autoregressive process (2) has a unit root $(\phi_1 = 1)$. It will tend to be very slowly damping as well.

Fortunately, the problem of non-stationarity can be remedied simply by differencing the data and instead focusing on the time series $y_t^* = y_t - y_{t-1} \equiv \Delta y_t$. That is because, from (12), we have

$$\Delta y_t = a_t \quad (18)$$

and, therefore,

$$E(\Delta y_t) = (E(a_t) = 0 \quad (19)$$

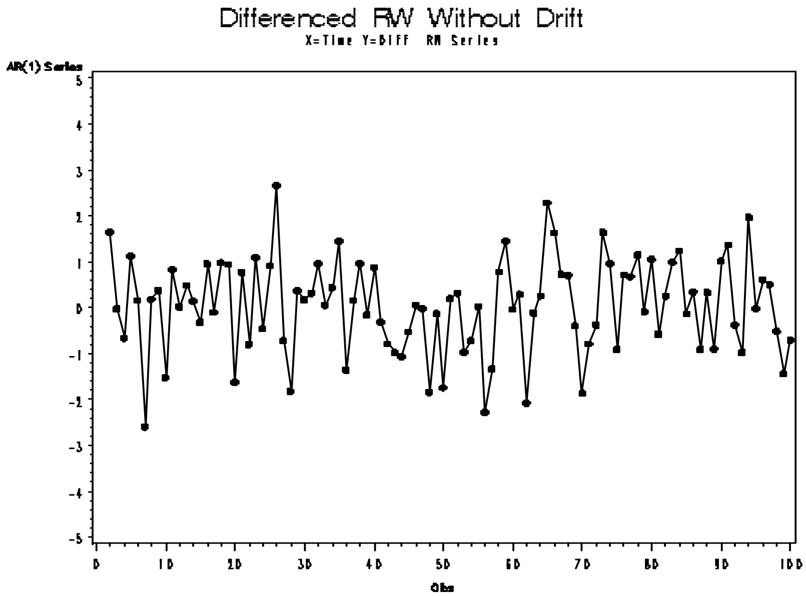
$$Var(\Delta y_t) = Var(a_t) = \sigma_a^2 \quad (20)$$

and

$$Cov(y_t, y_{t-j}) = E(a_t a_{t-j}) = 0 \text{ for all j and t.} \quad (21)$$

We can see that the transformed (differenced) series, Δy_t, is stationary because it has a constant mean, constant variance, and the covariance function is not a function of t. The differenced Random Walk of the previous plot is presented below in Figure 5. Notice its roughly contant mean, constant variance, and constant covariance.

Figure 5



Transforming Exponential Growth Data to Percentage Changes

Another frequently occurring non-stationary time series in business and economic data is depicted below. See Figure 5 below. This is called **exponential growth data** such that taking the natural logarithm of the data and then differencing the logarithmically transformed data produces a stationary time series. See the below two graphs, where the original data (y_t) is plotted in Figure 6 and in Figure 7 we have plotted the transformed series $\Delta\log(y_t)$. This latter transformation can be thought of a percentage change transformation since, for small percentage changes in y_t,

$$\Delta\log(y_t) \approx \frac{y_t - y_{t-1}}{y_{t-1}} \quad . \tag{22}$$

Figure 6

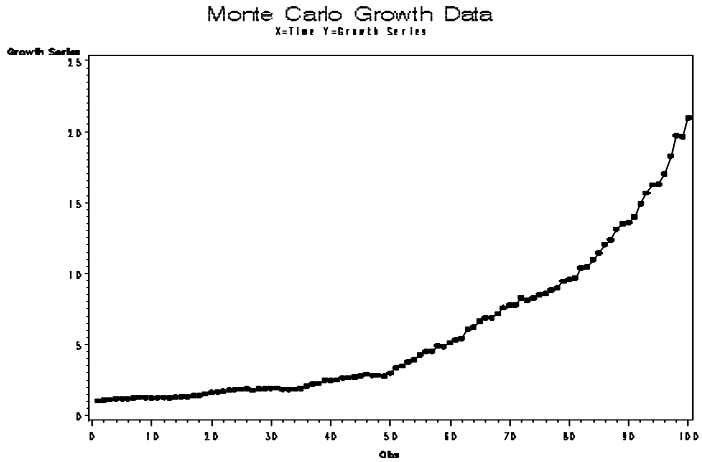
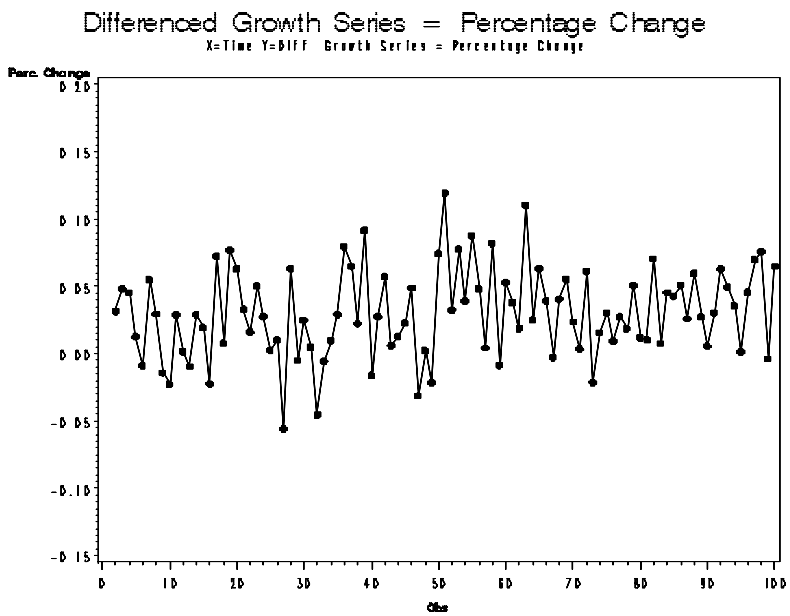
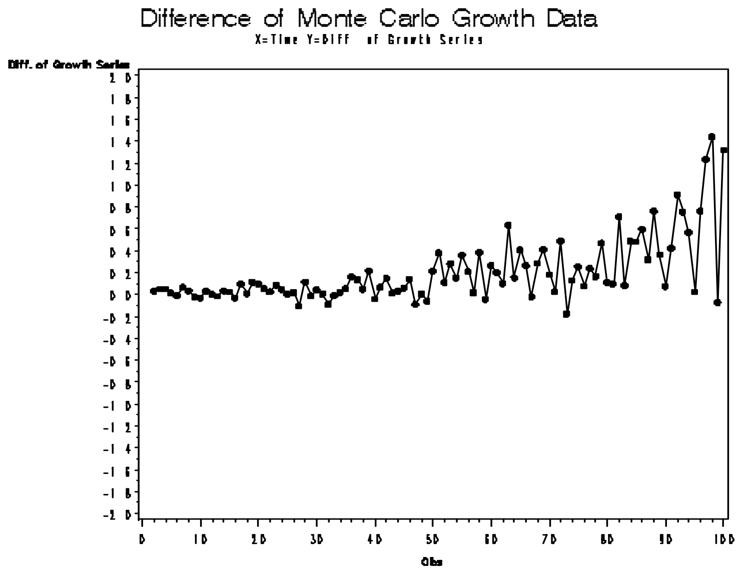


Figure 7



Later on we will learn how to use the SAS Macro %logtest to distinguish between the cases where we can simply difference the data versus situations where we need to first take the logarithms of the data and then difference the logarithms of the data thereafter. In fact you can see that if one is careful in the inspection of the proposed transformed series one can detect whether an appropriate transformation to stationarity has been applied or not. See Figure 8 below where the simple differences (Δy_t) of the exponential growth data have been plotted.

Figure 8



Obviously, in this case the transformed data are not stationary as the data trends to drift up over time and a constant mean and constant variance are not maintained in the data.

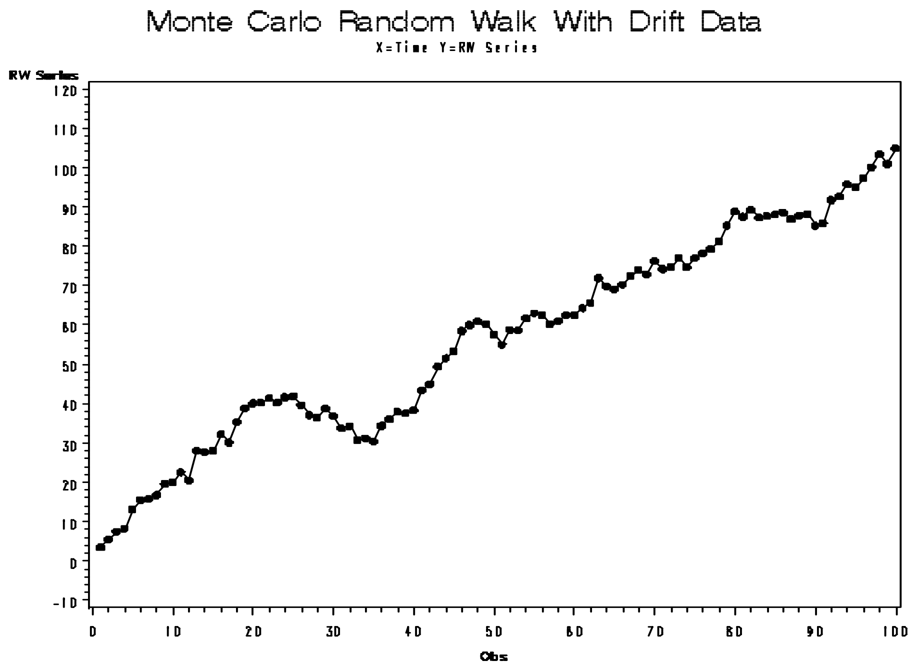
Deterministic Trend versus Stochastic Trend

Random Walk data can also have a drift in it. Consider the data plotted in Figure 9. It has been generated by the process

$$y_t = \phi_0 + y_{t-1} + a_t \tag{23}$$

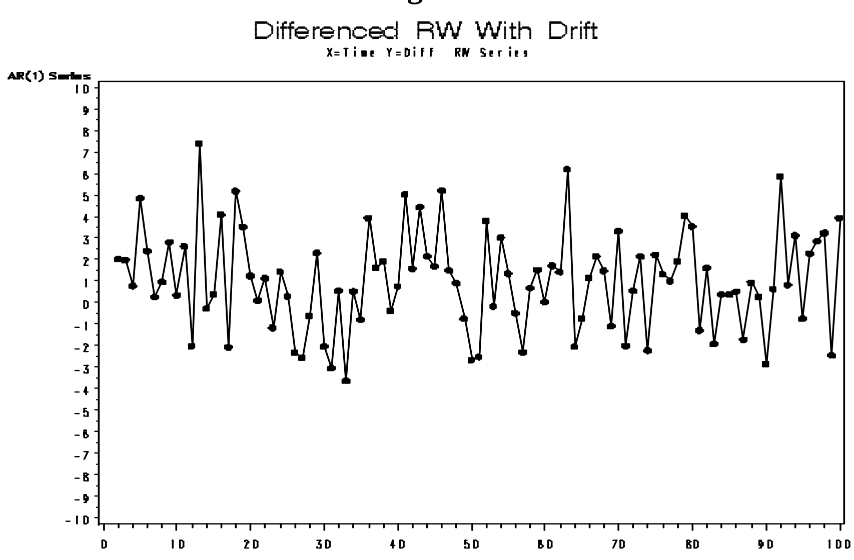
where $\phi_0 = 1$ and $\sigma_a^2 = 4$. If one were to draw a straight line through this trending data one would notice that the crossing rate of that line would be low hence hinting of the unit root $(\phi_0 = 1)$ in the data. Again, the stationarity transformation of this data is the first differencing operation Δy_t since $\Delta y_t = \phi_0 + a_t$ represents a constant mean, constant variance, and constant covariance process.

Figure 9



See Figure 10 below for the result of applying the difference operation to the data in figure 9. The data has been rendered stationary by taking first differences.

Figure 10

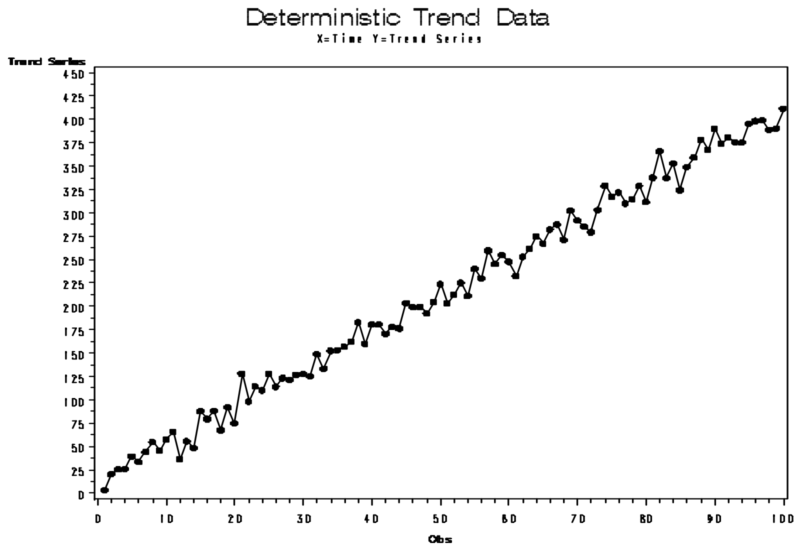


However, there exist time series in business and economics that might be better characterized as following what is called a **deterministic time trend** as in

$$y_t = \beta_0 + \beta_1 t + a_t \tag{24}$$

where t is the time index $t = 1,2,\cdots,T$. Such a **trend stationary** (TS) time series is plotted below in Figure 11 below.

Figure 11



Here you would see many crossing of the trend should you put a trend line through the above data. In later study we will see that **Crossing-rate** and **Dickey-Fuller unit root tests** can be used to distinguish between these cases. An important point to note here is that in the stationary AR(1) of Figure 1, the data reverts to the mean quite often whereas the Random Walk data of Figure 3 rarely does. Likewise the Random Walk with drift data of Figure 9 rarely reverts to a trend line one would place through the data whereas the deterministic trend data reverts often to the trend line.

To the extent that time series data reverts frequently to a mean or trend we call such data **mean-reverting** and having a **deterministic trend (mean)**. In contrast, to the extent that time series data reverts infrequently to a trend we call such data **non-mean reverting** and having a **stochastic trend**. To **motivate** the term stochastic trend, consider the following random-walk-like model called the **Random Walk structural model** by Harvey (1989).

$$\begin{aligned} y_t &= \mu_t + a_t \\ \mu_t &= \mu_{t-1} + \varepsilon_t \end{aligned} \tag{25}$$

We will study more about such models when we study **Unobserved Components models** later. Here we specify that the mean of the series μ_t is **stochastic** and, in fact, follows a random walk without drift. We can rewrite this model as

$$y_t = y_{t-1} + \nu_t \tag{26}$$

that is almost like the Random Walk without drift model (12) except now

$$\nu_t = a_t - a_{t-1} + \varepsilon_t \tag{27}$$

is a so-called MA(1) error term instead of being white noise. However, the point is made. Random Walk data behave much like data that has a stochastic random walk mean in it, hence the term **stochastic mean** as compared to a **deterministic mean**.

The question might be, "Why make such a big deal over whether a time series has a **stochastic trend** in it which can be treated by differencing or a **deterministic trend** that can be modeled as a constant mean or deterministic time trend?" The short answer is that when forecasting with models that have stochastic trends in them the prediction confidence intervals become ever wider the further out into the future one predicts. In contrast, in the deterministic mean or trend models the forecast confidence intervals approach a limit equal to the variance of the data around the mean or trend. Thus **the conception of trend** that one chooses to adopt is an important decision to make and can have a substantial effect on how one conveys the uncertainty of predicting a time series in the future. We will see this point more clearly later when we separately examine the confidence limits of the deterministic trend model and the Box-Jenkins model.

III. The Wold Decomposition (Representation) Theorem

In 1954, Herman Wold proved a very important theorem concerning stationary time series data. Here is a formal statement of Wold's theorem without proof:

Wold Decomposition (Representation) Theorem: Any stationary process y_t can be uniquely represented as the sum of two mutually uncorrelated processes D_t and Z_t where D_t is linearly deterministic and Z_t is a $MA(\infty)$ process. That is,

$$y_t = D_t + Z_t \tag{28}$$

where

$$Z_t = \psi_0 a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \cdots \tag{29}$$

with $\psi_0 = 1$ and the "psi" weights $\psi_1, \psi_2, \cdots$ being absolutely summable in that $\sum_{j=0}^{\infty} |\psi_j| < \infty$ and the a_t's are uncorrelated errors that have zero mean and constant variance. *******

As we shall see, D_t in the Box-Jenkins models is equal to the mean μ of the stationary form of the time series. However, in the Wold Theorem D_t can be anything ranging from a **deterministic time trend** $(\beta_0 + \beta_1 t)$, to a **deterministic cycle** like

$$D_t = a\cos(\omega t + \theta) \qquad (30)$$

where a is the amplitude of the cycle, θ is the phase, and the period of the cycle (p) such that p = $2\pi / \omega$ and a **deterministic seasonal** D_t as in

$$D_t = \delta_1 S_{t1} + \delta_2 S_{t2} + \cdots + \delta_s S_{ts} \qquad (31)$$

where S_{tk} is a season dummy variable taking the value of 1 during season k and zero otherwise with s number of seasons during the year.

The very useful result that the Wold Representation provides us is that all stationary Box-Jenkins models of stationary time series y_t can be written as a $MA(\infty)$ process as in

$$y = \mu + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \cdots \quad . \qquad (32)$$

This form will be very helpful to use, especially as it relates to deriving the properties of the models and their forecast functions and confidence intervals.

APPENDIX

Consider the AR(1) model

$$y_t = \phi_0 + \phi_1 y_{t-1} + a_t \quad (2)$$

where $|\phi_1| < 1$ is assumed and the error term a_t has the so-called "**white noise**" properties

(i) $E(a_t) = 0$ for all t

(ii) $E(a_t^2) = \sigma_a^2$ for all t (3)

(iii) $E(a_s a_t) = 0$ for all $s \neq t$.

We are to show that

(A) $E(y_t) = \dfrac{\phi_0}{1-\phi_1}$

(B) $Var(y_t) = \dfrac{\sigma_a^2}{1-\phi_1^2}$

(C) $Cov(y_t, y_{t-j}) = \dfrac{\phi_1^j \sigma_a^2}{1-\phi_a^2}$

(D) $Corr(y_t, y_{t-j}) = \phi_1^j$

Result (A):

By backward substitution we can confirm the Wold Representation theorem result that

$$y_t = \phi_0 + \phi_0\phi_1 + \phi_0\phi_1^2 + \cdots + a_t + \phi_1 a_{t-1} + \phi_1^2 a_{t-2} + \cdots$$

$$y_t = \frac{\phi_0}{1-\phi_1} + a_t + \phi_1 a_{t-1} + \phi_1^2 a_{t-2} + \cdots \quad . \quad \text{(A.1)}$$

In getting (A.1) we used the result that if $|\phi_1| < 1$, the geometric series $1 + \phi_1 + \phi_1^2 + \cdots$

converges to $1/(1-\phi_1)$. It follows from (A.1) that

$$E(y_t) = E(\frac{\phi_0}{1-\phi_1} + a_t + \phi_1 a_{t-1} + \phi_1^2 a_{t-2} + \cdots)$$

$$= E(\frac{\phi_0}{1-\phi_1}) + E(a_t) + \phi_1 E(a_{t-1}) + \phi_1^2 E(a_{t-2}) + \cdots$$

$$= \frac{\phi_0}{1-\phi_1}$$

as desired.

Results (B):

From (A.1) and the independence of the a_t's it follows that

$$Var(y_t) = Var(\frac{\phi_0}{1-\phi_1} + a_t + \phi_1 a_{t-1} + \phi_1^2 a_{t-2} + \cdots)$$

$$= Var(\frac{\phi_0}{1-\phi_1}) + Var(a_t) + \phi_1^2 Var(a_{t-1}) + \phi_1^4 Var(a_{t-2}) + \cdots$$

$$= \sigma_a^2 + \phi_1^2 \sigma_a^2 + \phi_1^4 \sigma_a^2 + \cdots$$

$$= \sigma_a^2 (\frac{1}{1-\phi_1^2}) = \frac{\sigma_a^2}{1-\phi_1^2}$$

as desired.

Result (C):

Using (A.1) we have

$$Cov(y_t, y_{t-j}) = E(y_t - \mu)(y_{t-j} - \mu)$$

$$= E(a_t + \phi_1 a_{t-1} + \phi_1^2 a_{t-2} \cdots)(a_{t-j} + \phi_1 a_{t-j-1} + \phi_1^2 a_{t-j-2} + \cdots)$$

$$= E(\phi_1^j a_{t-j}^2) + E(\phi_1^{j+2} a_{t-j-1}^2) + E(\phi_1^{j+4} a_{t-j-2}^2) + \cdots$$

$$= \phi_1^j (\sigma_a^2 + \phi_1^2 \sigma_a^2 + \phi_1^4 \sigma_a^2 + \cdots) = \phi_1^j \sigma_a^2 (1 + \phi_1^2 + \phi_1^4 + \cdots)$$

$$= \frac{\phi_1^j \sigma_a^2}{(1-\phi_1^2)}$$

as desired.

Result (D):

$$Corr(y_t, y_{t-j}) = \frac{Cov(y_t, y_{t-j})}{Var(y_t)} = \frac{\dfrac{\phi_1^j \sigma_a^2}{(1-\phi_1^2)}}{\dfrac{\sigma_a^2}{(1-\phi_1^2)}} = \phi_1^j$$

as desired.

End of Appendix.