Ch. 19 Models of Nonstationary Time Series

In time series analysis we do not confine ourselves to the analysis of stationary time series. In fact, most of the time series we encounter are $non-stationary$. How to deal with the nonstationary data and use what we have learned from stationary model are the main subjects of this chapter.

1 Integrated Process

Consider the following two process

$$\begin{aligned} X_t &= \phi X_{t-1} + u_t, \quad |\phi| < 1; \\ Y_t &= Y_{t-1} + v_t, \end{aligned}$$

where u_t and v_t are mutually uncorrelated white noise process with variance σ_u^2 and σ_v^2, respectively. Both X_t and Y_t are $AR(1)$ process. The difference between two models is that Y_t is a special case of a X_t process when $\phi = 1$ and is called a *random walk* process. It is also refereed to as a $AR(1)$ model with a *unit root* since the root of the $AR(1)$ process is 1. When we consider the statistical behavior of the two processes by investigating the mean (the first moment), and the variance and autocovariance (the second moment), they are completely different. Although the two process belong to the same $AR(1)$ class, X_t is a stationary process, while Y_t is a *nonstationary* process.

The process Y_t explain the use of the term "unit root process." Another expression that is sometimes used is that the process Y_t is **integrated** of order 1. This is indicated as $Y_t \sim I(1)$. The term "integrated" comes from calculus; if $dY(t)/dt = v(t)$, then $Y(t)$ is the integral of $v(t)$.[1] In discrete time series, if $\triangle Y_t = v_t$, then Y might also be viewed as the integral, or sum over t, of v.

[1]To see this, recall from the basic property of Riemann integral that suppose that v is bounded on $I : \{a \leq s \leq b\}$ and let $Y(t)$ is defined on I by $Y(t) \equiv \int_a^t v(s)ds$, then $Y'(t_0) = \lim_{h\to 0} \frac{Y(t_0+h)-Y(t_0)}{h} = v(t_0)$, $\forall t_o \in (a, b]$. In discrete time, setting $h = 1$ we obtain the unit root process.

1.1 First Two Moments of Random Walk (without drift) Process

Assume that $t \in \mathcal{T}^*, \mathcal{T}^* = \{0, 1, 2, ...\}$,[2] the first stochastic processes can be expressed ad

$$X_t = \phi^t X_0 + \sum_{i=0}^{t-1} \phi^i u_{t-i}.$$

Similarly, in the unit root case

$$Y_t = Y_0 + \sum_{i=0}^{t-1} v_{t-i}.$$

Suppose that the initial observation is zero, $X_0 = 0$ and $Y_0 = 0$. The means of the two process are

$$E(X_t) = 0 \quad and \quad E(Y_t) = 0,$$

and variance are

$$Var(X_t) = \sum_{i=0}^{t-1} \phi^{2i} Var(u_{t-i}) \longrightarrow \frac{1}{1-\phi^2}\sigma_u^2$$

and

$$Var(Y_t) = \sum_{i=0}^{t-1} Var(v_{t-i}) = t \cdot \sigma_v^2 \longrightarrow \infty.$$

The autocovariance of the two series are

$$\begin{aligned}
\gamma_\tau^X &= E(X_t X_{t-\tau}) = E\left[\left(\sum_{i=0}^{t-1} \phi^i u_{t-i}\right)\left(\sum_{i=0}^{t-\tau-1} \phi^i u_{t-\tau-i}\right)\right] \\
&= E[(u_t + \phi^1 u_{t-1} + ... + \phi^\tau u_{t-\tau} + ... + \phi^{t-1} u_1)(u_{t-\tau} + \phi^1 u_{t-\tau-1} + ... + \phi^{t-\tau-1} u_1) \\
&= \sum_{i=0}^{t-\tau-1} \phi^i \phi^{\tau+i} \sigma_u^2 \\
&= \sigma_u^2 \phi^\tau (\sum_{i=0}^{t-\tau-1} \phi^{2i}) \\
&\longrightarrow \frac{\phi^\tau}{1-\phi^2}\sigma_u^2 \\
&= \phi^\tau \gamma_0^X.
\end{aligned}$$

[2] This assumption that the starting data being 0 is required to derive the convergence of integrated process to standard Brownian Motion. A standard Brown Motion is defined on $t \in [0, 1]$.

and

$$\begin{aligned}
\gamma_\tau^Y &= E(Y_t Y_{t-\tau}) = E\left[\left(\sum_{i=0}^{t-1} v_{t-i}\right)\left(\sum_{i=0}^{t-\tau-1} v_{t-\tau-i}\right)\right] \\
&= E[(v_t + v_{t-1} + ... + v_{t-\tau} + v_{t-\tau-1} + ... + v_1)(v_{t-\tau} + v_{t-\tau-1} + ... + v_1)] \\
&= (t-\tau)\sigma_v^2.
\end{aligned}$$

We may expect that the autocorrelation functions are

$$r_\tau^X = \frac{\gamma_\tau^X}{\gamma_0^X} = \phi^\tau \longrightarrow 0$$

and

$$r_\tau^Y = \frac{\gamma_\tau^Y}{\gamma_0^Y} = \frac{(t-\tau)}{t} \longrightarrow 1, \ \forall\, \tau.$$

The mean of X_t and Y_t are the same, but the variance (including autocovariance) are different. The important thing to note is that the variance and autocovariance of Y_t **are function of** t, while those of X_t converge to a constant asymptotically. Thus as t increase the variance of Y_t increase, while the variance of X_t converges to a constant.

1.2 First Two Moments of Random Walk With Drift Process

If we add a constant to the $AR(1)$ process, then the means of two processes also behave differently. Consider the $AR(1)$ process with a constant (or drift) as follows

$$X_t = \alpha + \phi X_{t-1} + u_t, \quad |\phi| < 1$$

and

$$Y_t = \alpha + Y_{t-1} + v_t.$$

The successive substitution yields

$$X_t = \phi^t X_0 + \alpha \sum_{i=0}^{t-1} \phi^i + \sum_{i=0}^{t-1} \phi^i u_{t-i}$$

and

$$Y_t = Y_0 + \alpha t + \sum_{i=0}^{t-1} v_{t-i}. \tag{1}$$

Note that Y_t contains a (deterministic) trend t. If the initial observations are zero, $X_0 = 0$ and $Y_0 = 0$, then the means of two process are

$$\begin{aligned} E(X_t) &\longrightarrow \frac{\alpha}{1-\phi} \quad \textit{and} \\ E(Y_t) &= \alpha t \end{aligned}$$

but the variance and the autocovariance are the same as those derived from $AR(1)$ model without the constant. By adding a constant to the $AR(1)$ processes, the means of two processes as well the variance are different. Both mean and variance of Y_t are time varying, while those of X_t are constant.

Since the variance (the second moment) and even mean (the first moment) of the nonstationary series is not constant over time, the conventional asymptotic theory cannot be applied for these series (Recall the moment condition in CLT on p.22 of Ch. 4).

1.3 Random Walk with Reflecting Barrier

While the unit root hypothesis is evident in time series analysis to describe random wandering behavior of economic and financial variables, the data containing a unit root are in fact possibly been censored before they are observed. For example, in the financial literatures, price subject to price limits imposed in stock markets, commodity future exchanges, the positive nominal interest rate, and foreign exchange futures markets have been treated as censored variables. Censored data are also common in commodity markets where the government has historically intervened to support prices or to impose quotas (see de Jong and Herrera, 2004).

Lee et. al. (2006) study how to test the unit root hypothesis when the time series data is censored. The underlying latent equation that contained a unit root is a general $I(1)$ process:

$$Z_t = \rho Z_{t-1} + v_t, \quad t = 1, 2, ..., T,$$

where $\rho = 1$, v_t is zero-mean stationary and invertible process to be specified below. However, Z_t is censored from the left and is observed only through the censored sample Y_t, i.e.

$$W_t = \max(c, Z_t),$$

where c is a known constant. Let $\mathbf{C}$ be the set of censored observations. For those $Z_t, t \in \mathbf{C}$, we call them latent variables. Our problem is to inference ρ on the basis of T observations on W_t.

Equations on Z_t and W_t constitute a censored unit root process[3] which is one example of the dynamic censored regression model except for the absence of all exogenous explanatory variables in the right-hand side of Z_t. The static censored regression model are standard tools in econometrics. Statistical theory of this model in cross-section situations has been long been understood; see for example the treatment in Maddala (1983). Dynamic censored regression model have also become popular for time series when lags of the dependent variable have been included among the regressors. Lee (1999) and de Jong and Herrera (2004) propose maximum likelihood method for estimation of such model. However, the assumption is made in their studies that the lag polynomial $1 - \rho b$ in Z_t has its root outside the unit circle, which completely excludes the possibility of a unit root in the latent equation.

[3]When the special case that v_t is a white noise process, then Z_t is called a random walk with a reflecting barriers at c. The content of a dam is one example of this process.

2 Deterministic Trend and Stochastic Trend

Many economic and financial times series do trended upward over time (such as GNP, M2, Stock Index etc.). See the plots of Hamilton, p.436. For a long time each trending (nonstationary) economic time series has been decomposed into a deterministic trend and a stationary process. In recent years the idea of stochastic trend has emerged, and enriched the framework of analysis to investigate economic time series.

2.1 Detrending Methods

2.1.1 Differencing-Stationary

One of the easiest ways to analyze those nonstationary-trending series is to make those series stationary by differencing. In our example, the random walk series with drift Y_t can be transformed to a stationary series by differencing once

$$\triangle Y_t = Y_t - Y_{t-1} = (1-L)Y_t = \alpha + v_t.$$

Since v_t is assumed to be a white noise process, the first difference of Y_t is stationary.[4] The variance of $\triangle Y_t$ is constant over the sample period. In the $I(1)$ process,

$$Y_t = Y_0 + \alpha t + \sum_{i=0}^{t-1} v_{t-i}, \tag{2}$$

αt is a **deterministic trend** while $\sum_{i=0}^{t-1} v_{t-i}$ is a **stochastic trend**.

When the nonstationary series can be transformed to the stationary series by differencing once, the series is said to be *integrated of order* 1 and is denoted by $I(1)$, or in common, a unit root process. If the series needs to be differenced d times to be stationery, then the series is said to be $I(d)$. The $I(d)$ series ($d \neq 0$) is also called a $differencing - stationary\ process$ (DSP). When $(1-L)^d Y_t$ is a stationary and invertible series that can be represented by an $ARMA(p,q)$ model, i.e.

$$(1-\phi_1 L - \phi_2 L^2 - ... - \phi_p L^p)(1-L)^d Y_t = \alpha^* + (1+\theta_1 L + \theta_2 L^2 + ... + \theta_q L^q)\varepsilon_t \tag{3}$$

[4]It is simple to see this result: $E(\triangle Y_t) = \alpha$, $Var(\triangle Y_t) = \sigma_v^2$ and $E[(\triangle Y_t - \alpha)(\triangle Y_s - \alpha)] = v_t v_s = 0$ for $t \neq s$.

or

$$\phi(L)\triangle^d Y_t = \alpha^* + \theta(L)\varepsilon_t,$$

where all the roots of $\phi(L) = 0$ and $\theta(L) = 0$ lie outside the unit circle, we say that Y_t is an autoregressive integrated moving-average $ARIMA(p, d, q)$ process. In particular an unit root process, $d = 1$ or an $ARIMA(p, 1, q)$ process is therefore

$$\phi(L)\triangle Y_t = \alpha^* + \theta(L)\varepsilon_t$$

or

$$(1 - L)Y_t = \alpha + \psi(L)\varepsilon_t, \tag{4}$$

where $\alpha = \phi^{-1}(1)\alpha^*$, $\psi(L) = \phi^{-1}(L)\theta(L)$ and is absolutely summable.

Successive substitution of (4) yields a generalization of (2):

$$Y_t = Y_0 + \alpha t + \psi(L)\sum_{i=0}^{t-1}\varepsilon_{t-i}. \tag{5}$$

2.1.2 Trend-Stationary

Another important class that accommodate the trend in the model is the *trend−stationary process* (TSP). Consider the series

$$X_t = \mu + \alpha t + \psi(L)\varepsilon_t, \tag{6}$$

where the coefficients of $\psi(L)$ is absolute summable.

The mean of X_t is $E(X_t) = \mu + \alpha t$ and is not constant over time, while the variance of X_t is $Var(X_t) = (1_+\psi_1^2+\psi_2^2+...)\sigma^2$ and constant. Although the mean of X_t is not constant over the period, it can be forecasted perfectly whenever we know the value of t and the parameters μ and α. In the sense it is stationary around the deterministic trend t and X_t can be transformed to stationarity by regressing it on time. Note that both DSP model equation (5) and the TSP model equation (6) exhibit a linear trend, but the appropriated method of eliminating the trend differs. (It can be seen that the DSP is only *trend − nonstationary* from the definition of TSP.)

Most econometric analysis is based the variance and covariance among the variables. For example, the OLS estimator from the regression Y_t on X_t is the ratio of the covariance between Y_t and X_t to variance of X_t. Thus if the variance of the variables behave differently, the conventional asymptotic theory cannot be applicable. When the order of integration is different, the variance of each process behave differently. For example, if Y_t is an $I(0)$ variable and X_t is $I(1)$, the OLS estimator from the regression Y_t on X_t converges to zero asymptotically, since the denominator of the OLS estimator, the variance of X_t, increase as t increase, and thus it dominates the numerator, the covariance between X_t and Y_t. That is, the OLS estimator does not have an asymptotic distribution. (It is degenerated with the conventional normalization of $\sqrt{T}$. See Ch. 21 for details)

2.2 Comparison of Trend-stationary and Differencing -Stationary Process

The best way to under the meaning of stochastic and deterministic trend is to compare their time series properties. This section compares a trend-stationary process (6) with a unit root process (4) in terms of forecasts of the series, variance of the forecast error, dynamic multiplier, and transformations needs to achieve stationarity.

2.2.1 Returning to a Central Line ?

The TSP model (6) has a central line $\mu + \alpha t$, around which, X_t oscillates. Even if shock let X_t deviate temporarily from the line there takes place a force to bring it back to the line. On the other hand, the unit root process (5) has no such a central line. One might wonder about a deterministic trend combined with a random walk. The discrepancy between Y_t and the line $Y_0 + \alpha t$, became unbounded as $t \to \infty$.

2.2.2 Forecast Error

The TSP and unit root specifications are also very different in their implications for the variance of the forecast error. For the trend-stationary process (6), the s

ahead forecast is

$$\hat{X}_{t+s|t} = \mu + \alpha(t+s) + \psi_s\varepsilon_t + \psi_{s+1}\varepsilon_{t-1} + \psi_{s+2}\varepsilon_{t-2} + ...$$

which are associated with forecast error

$$\begin{aligned} X_{t+s} - \hat{X}_{t+s|t} &= \{\mu + \alpha(t+s) + \varepsilon_{t+s} + \psi_1\varepsilon_{t+s-1} + \psi_2\varepsilon_{t+s-2} + ... \\ &\quad +\psi_{s-1}\varepsilon_{t+1} + \psi_s\varepsilon_t + \psi_{s+1}\varepsilon_{t-1} + ...\} \\ &\quad -\{\mu + \alpha(t+s) + \psi_s\varepsilon_t + \psi_{s+1}\varepsilon_{t-1} + ...\} \\ &= \varepsilon_{t+s} + \psi_1\varepsilon_{t+s-1} + \psi_2\varepsilon_{t+s-2} + ... + \psi_{s-1}\varepsilon_{t+1}. \end{aligned}$$

The MSE of this forecast is

$$E[X_{t+s} - \hat{X}_{t+s|t}]^2 = \{1 + \psi_1^2 + \psi_2^2 + ... + \psi_{s-1}^2\}\sigma^2.$$

The MSE increases with the forecasting horizon s, though as s becomes large, the added uncertainty from forecasting farther into the future becomes negligible:[5]

$$\lim_{s\to\infty} E[X_{t+s} - \hat{X}_{t+s|t}]^2 = \{1 + \psi_1^2 + \psi_2^2 + ...\}\sigma^2 < \infty.$$

Note that the limiting MSE is just the unconditional variance of the stationary component $\psi(L)\varepsilon_t$.

To forecast the unit root process (4), recall that the change $\triangle Y_t$ is a stationary process that can be forecast using the standard formula:

$$\begin{aligned} \triangle\hat{Y}_{t+s|t} &= \hat{E}[(Y_{t+s} - Y_{t+s-1})|Y_t, Y_{t-1}, ...] \\ &= \alpha + \psi_s\varepsilon_t + \psi_{s+1}\varepsilon_{t-1} + \psi_{s+2}\varepsilon_{t-2} + ... \end{aligned}$$

The level of the variable at date $t+s$ is simply the sum of the change between t and $t+s$:

$$\begin{aligned} Y_{t+s} &= (Y_{t+s} - Y_{t+s-1}) + (Y_{t+s-1} - Y_{t+s-2}) + ... + (Y_{t+1} - Y_t) + Y_t \qquad (7) \\ &= \triangle Y_{t+s} + \triangle Y_{t+s-1} + ... + \triangle Y_{t+1} + Y_t. \qquad (8) \end{aligned}$$

Therefore the s period ahead forecast error for the unit root process is

$$\begin{aligned} Y_{t+s} - \hat{Y}_{t+s|t} &= \{\triangle Y_{t+s} + \triangle Y_{t+s-1} + ... + \triangle Y_{t+1} + Y_t\} \\ &\quad -\{\triangle\hat{Y}_{t+s|t} + \triangle\hat{Y}_{t+s-1|t} + ... + \triangle\hat{Y}_{t+1|t} + Y_t\} \\ &= \{\varepsilon_{t+s} + \psi_1\varepsilon_{t+s-1} + ... + \psi_{s-1}\varepsilon_{t+1}\} \\ &\quad +\{\varepsilon_{t+s-1} + \psi_1\varepsilon_{t+s-2} + ... + \psi_{s-2}\varepsilon_{t+1}\} + ... + \{\varepsilon_{t+1}\} \\ &= \varepsilon_{t+s} + [1 + \psi_1]\varepsilon_{t+s-1} + [1 + \psi_1 + \psi_2]\varepsilon_{t+s-2} + ... \\ &\quad +[1 + \psi_1 + \psi_2 + ... + \psi_{s-1}]\varepsilon_{t+1}, \end{aligned}$$

[5]Since ψ_j is absolutely summable.

with MSE

$$E[Y_{t+s} - \hat{Y}_{t+s|t}]^2 = \{1 + [1 + \psi_1]^2 + [1 + \psi_1 + \psi_2]^2 + ... + [1 + \psi_1 + \psi_2 + ... + \psi_{s-1}]^2\}\sigma^2.$$

The MSE again increase with the length of the forecasting horizon s, though in contrast to the trend-stationary case. The MSE does not converge to any fixed value as s goes to infinity. See Figures 15.2 on p. 441 of Hamilton.

The model of a TSP and the model of DSP have totally different views about how the world evolves in future. In the former the forecast error is bounded even in the infinite horizon, but in the latter the error become unbounded as the horizon extends.

One result is very important to understanding the asymptotic statistical properties to be presented in the subsequent chapter. The (deterministic) trend introduced by a nonzero drift α, (αt is $O(T)$) asymptotically dominates the increasing variability arising over time due to the unit root component. ($\sum_{i=0}^{t-1} \varepsilon_{t-i}$ is $O(T^{1/2})$.) This means that data from a unit root with positive drift are certain to exhibit an upward trend if observed for a sufficiently long period.[6]

2.2.3 Impulse Response

Another difference between TSP and unit root process is the persistence of innovations. Consider the consequences for X_{t+s} if ε_t were to increase by one unit with ε's for all other dates unaffected. For the TSP process (4), this impulse response is given by

$$\frac{\partial X_{t+s}}{\partial \varepsilon_t} = \psi_s.$$

For a trend-stationary process, then, the effect of any stochastic disturbance eventually wears off:

$$\lim_{s\to\infty} \frac{\partial X_{t+s}}{\partial \varepsilon_t} = 0.$$

[6] Hamilton p. 442: " Figure 15.3 plots realization of a Gaussian random walk without drift and with drift. The random walk without drift, shown in panel (a), shows no tendency to return to its starting value or any unconditional mean. The random walk with drift, shown in panel (b), shows no tendency to return to a fixed deterministic trend line, though the series is asymptotically dominated by the positive drift term." That is, the random walk is trending up, but it is not causing by the purpose of returning to the trend line.

By contrast, for a unit root process, the effect of ε_t on Y_{t+s} is seen from (8) and (4) to be

$$\begin{aligned}\frac{\partial Y_{t+s}}{\partial \varepsilon_t} &= \frac{\partial \triangle Y_{t+s}}{\partial \varepsilon_t} + \frac{\partial \triangle Y_{t+s-1}}{\partial \varepsilon_t} + ... + \frac{\partial \triangle Y_{t+1}}{\partial \varepsilon_t} + \frac{\partial Y_t}{\partial \varepsilon_t} \\ &= \psi_s + \psi_{s-1} + ... + \psi_1 + 1 \quad (since \frac{\partial \triangle Y_{t+s}}{\partial \varepsilon_t} = \psi_s \; from \; (4))\end{aligned}$$

An innovation ε_t has a permanent effect on the level of Y that is captured by

$$\lim_{s\to\infty} \frac{\partial Y_{t+s}}{\partial \varepsilon_t} = 1 + \psi_1 + \psi_2 + ... = \psi(1).$$

Example:
The following $ARIMA(4,1,0)$ model was estimated for Y_t:

$$\triangle Y_t = 0.555 + 0.312\triangle Y_{t-1} + 0.122\triangle Y_{t-2} - 0.116\triangle Y_{t-3} - 0.081\triangle y_{t-4} + \hat{\varepsilon}_t.$$

For this specification, the permanent effect of a one-unit change in ε_t on the level of Y_t is estimated to be

$$\psi(1) = \frac{1}{\phi(1)} = \frac{1}{(1 - 0.312 - 0.122 + 0.116 + 0.081)} = 1.31.$$

2.2.4 Transformations to Achieve Stationarity

A final difference between trend-stationary and unit root process that deserves comment is the transformation of the data needed to generate a stationary time series. If the process is really trend stationary as in (6), the appropriate treatment is to subtract αt from X_t to produce a stationary representation. By contrast, if the data were really generated by the unit root process (5), subtracting αt from Y_t, would succeed in removing the time-dependence of the mean but not the variance as seen in (5).

There have been several papers that have studied the consequence of $overdiffer - encing$ and $underdifferencing$:
(a). If the process is really TSP as in (6), difference it would be

$$\triangle X_t = \mu + \alpha t - \mu - \alpha(t-1) + \psi(L)(1-L)\varepsilon_t = \alpha + \psi^*(L)\varepsilon_t. \tag{9}$$

In this representation, this look like a DSP however, a unit root has been introduced into the moving average representation, $\psi^*(L)$ which violates the definition of $I(d)$ process as in (4). This is the case of over-differencing.

(b). If the process is really DSP as in (6), and we treat it as TSP, we have a case of under-differencing.

3 Other Approaches to Trended Time Series

3.1 Fractional Integration

3.1.1 Fractional White Noise

We formally defined an $ARFIMA(0, d, 0)$, or a **fractional white noise** process to be a discrete-time stochastic process Y_t which may be represented as

$$(1 - L)^d Y_t = \varepsilon_t, \tag{10}$$

where ε_t is a mean-zero white noise and d is possibly non-integer. The following theorem give some of the basic properties of the process, assuming for convenience that $\sigma_\varepsilon^2 = 1$.

Theorem 1:
Let Y_t be an $ARFIMA(0, d, 0)$ process.
(a) When $d < \frac{1}{2}$, Y_t is a stationary process and has the infinite moving average representation

$$Y_t = \varphi(L)\varepsilon_t = \sum_{k=0}^{\infty} \varphi_k \varepsilon_{t-k}, \tag{11}$$

where

$$\varphi_k = \frac{d(1+d)...(k-1+d)}{k!} = \frac{(k+d-1)!}{k!(d-1)!} = \frac{\Gamma(k+d)}{\Gamma(k+1)\Gamma(d)}.$$

Here, $\Gamma(\cdot)$ is a *Gamma* function. As $k \to \infty$, $\varphi_k \sim k^{d-1}/(d-1)! \equiv \frac{1}{\Gamma(d)} \cdot k^{d-1}$.

(b) When $d > -\frac{1}{2}$, Y_t is invertible and has the infinite autoregressive representation

$$\phi(L)Y_t = \sum_{k=0}^{\infty} \phi_k Y_{t-k} = \varepsilon_t, \tag{12}$$

where

$$\phi_k = \frac{-d(1-d)...(k-1-d)}{k!} = \frac{(k-d-1)!}{k!(-d-1)!} = \frac{\Gamma(k-d)}{\Gamma(k+1)\Gamma(-d)}.$$

As $k \to \infty$, $\varphi_k \sim k^{-d-1}/(-d-1)! \equiv \frac{1}{\Gamma(-d)} \cdot k^{-d-1}$.

(c) When $-\frac{1}{2} < d < \frac{1}{2}$, the autocovariance of Y_t ($\sigma_\varepsilon^2 = 1$) is

$$\gamma_k = E(Y_t Y_{t-k}) = \frac{\Gamma(k+d)\Gamma(1-2d)}{\Gamma(k+1-d)\Gamma(1-d)\Gamma(d)} \tag{13}$$

and the autocorrelations functions is

$$r_k = \frac{\gamma_k}{\gamma_0} = \frac{\Gamma(k+d)\Gamma(1-d)}{\Gamma(k-d+1)\Gamma(d)}. \tag{14}$$

As $k \to \infty$, $r_k \sim \frac{\Gamma(1-d)}{\Gamma(d)} \cdot k^{2d-1}$.

Proof:
For part (a).
Using the standard binomial expansion

$$(1-z)^{-d} = \sum_{k=0}^{\infty} \frac{\Gamma(k+d)z^k}{\Gamma(d)\Gamma(k+1)}, \quad (how?) \tag{15}$$

it follows that

$$\varphi_k = \frac{\Gamma(k+d)}{\Gamma(k+1)\Gamma(d)}, \quad k \geq 1.$$

Using the standard approximation derived from Sheppard's formula, that for large k, $\Gamma(k+a)/\Gamma(k+b)$ is well approximated by j^{a-b}, it follows that

$$\varphi_k \sim k^{d-1}/(d-1)! \simeq Ak^{d-1} \tag{16}$$

for k large and an appropriate constant A.

Consider now an $MA(\infty)$ model given exactly by (11), i.e.,

$$Y_t = A \sum_{k=1}^{\infty} k^{d-1} \varepsilon_{t-k} + \varepsilon_t$$

so that $\varphi_0 = 1$. This series has variance

$$Var(Y_t) = A^2 \sigma_\varepsilon^2 \left(1 + \sum_{k=1}^{\infty} k^{2(d-1)}\right).$$

From the theory of infinity series, it is known that

$$\sum_{k=1}^{\infty} k^{-s} \quad converges\ for\ s > 1 \tag{17}$$

but otherwise diverges. It follows that the variance of Y_t is finite provided $d < \frac{1}{2}$, but is infinite if $d \geq \frac{1}{2}$. Also, since $\sum_{k=0}^{\infty} \varphi_k^2 < \infty$, the fractional white noise process is mean square summable and stationary for $d < \frac{1}{2}$.[7]

The proofs of part (b) is analogous to part (a) and is omitted.

For part (c), See Hosking (1981) and Granger and Joyeux (1980) for the proof of γ_k and r_k. It is note that

$$r_k = \frac{\gamma_k}{\gamma_0} = \frac{\frac{\Gamma(k+d)\Gamma(1-2d)}{\Gamma(k+1-d)\Gamma(1-d)\Gamma(d)}}{\frac{\Gamma(d)\Gamma(1-2d)}{\Gamma(1-d)\Gamma(1-d)\Gamma(d)}} = \frac{\Gamma(k+d)\Gamma(1-d)}{\Gamma(k-d+1)\Gamma(d)} \simeq \frac{\Gamma(1-d)}{\Gamma(d)} k^{2d-1}. \tag{18}$$

3.1.2 Relations to the Definitions of Long-Memory Process

For $0 < d < \frac{1}{2}$, the fractionally integrated process, $I(d)$, Y_t is long memory in the sense of the condition (1), its autocorrelations are all positive ($\frac{\Gamma(1-d)}{\Gamma(d)} k^{2d-1}$) such that condition (1) is violated [8] and decay at a hyperbolic rate.

For $-\frac{1}{2} < d < 0$, the sum of absolute values of the processes autocorrelations tend to a constant, so that it has **short memory** according to definition (1).[9] In this situation, the $ARFIMA(0, d, 0)$ process is said to be 'antipersistent' or to have 'intermediate memory', and all its autocorrelations, excluding lag zero, are negative and decay hyperbolically to zero.

The relation of the second definition of long memory with $I(d)$ process can be illustrated with the behavior of the partial sum S_T in (2), when Y_t is a fractional white noise as in (5). Sowell (1990) shows that

$$\lim_{T\to\infty} Var(S_T)T^{-(1+2d)} = \lim_{T\to\infty} E(S_T^2)T^{-(1+2d)} = \sigma_\varepsilon^2 \frac{\Gamma(1-2d)}{(1+2d)\Gamma(1+d)\Gamma(1-d)}.$$

Hence,

$$Var(S_T) = O(T^{1+2d}),$$

[7] Brockwell and Davis (1987) show that Y_t is convergent in mean square through its spectral representation.

[8] Suppose that $\sum a_n$ converges. Then $\lim a_n = 0$. See Fulks (1978), p.465.

[9] From (12), $\sum_{k=0}^{\infty} \frac{\Gamma(1-d)}{\Gamma(d)} k^{2d-1}$ converges for $1 - 2d > 1$ or that $d < 0$.

which implies that the variance of the partial sum of an $I(d)$ process, with $d = 0$, grows linearly,i.e., at rate of $O(T^1)$. For a process with intermediate memory with $-\frac{1}{2} < d < 0$, the variance of the partial sum grows at a **slower rate** than the linear rate, while for a long memory process with $0 < d < \frac{1}{2}$, the rate of growth is faster than a linear rate.

The relation of the third definition of long memory with $I(d)$ process can be illustrated beginning with the definition of fractional Brownian motion.

Brownian motion is a continuous time stochastic process $B(t)$ with independent Gaussian increments. Its derivatives is the continuous-time white noise process.

Fractional Brownian motion $B_H(t)$ is a generalization of these process. The fractional Brownian motion with parameter H, usually $0 < H < 1$, is the $(\frac{1}{2} - H)th$ fractional derivatives of Brownian motion. The continuous-time fractional noise is then defined as $B'_H(t)$, the derivative of fractional Brownian motion; it may also be thought of as the $(\frac{1}{2} - H)th$ fractional derivative of the continuous time white noise, to which it reduces when $H = \frac{1}{2}$.

We seek a discrete time analogue of continuous time fractional white noise. One possibility is discrete time fractional Gaussian noise, which is defined to be a process whose correlation is the same as that of the process of unit increments $\triangle B_H(t) = B_H(t) - B_H(t-1)$ of fractional Brownian motion.

The discrete time analogue of Brownian motion is the random walk, X_t defined by

$$(1 - L)X_t = \varepsilon_t,$$

where ε_t is $i.i.d.$. The first difference of X_t is the discrete-time white noise process ε_t. By analogy with the above definition of continuous time white noise we defined **fractionally differenced white noise** with parameter H to be the $(\frac{1}{2} - H)th$ fractional difference of discrete time white noise. The fractional difference operator $(1 - L)^d$ is defined in the natural way, by a binomial series:

$$\begin{aligned}(1-L)^d &= \sum_{k=0}^{\infty} \binom{d}{k} (-L)^k = 1 - dL - \frac{1}{2}d(1-d)L^2 - \frac{1}{6}d(1-d)(2-d)L^3 - ... \\ &= \sum_{k=0}^{\infty} \frac{\Gamma(k-d)z^k}{\Gamma(-d)\Gamma(k+1)}. \end{aligned} \tag{19}$$

We write $d = H - \frac{1}{2}$, so that the continuous time fractional white noise with parameters H has as its discrete time analogue the process $X_t = (1 - L)^{-d}\varepsilon_t$, or

$(1-L)^d X_t = \varepsilon_t$, where ε_t is a white noise process.

With the results above, the fractional white $I(d)$ process is also a long memory process according to definition 3 by substitution $d = H - \frac{1}{2}$ into (9).

3.1.3 ARFIMA process

A natural extension of the fractional white noise model (5) is the **fractional** $ARMA$ model or the $ARFIMA(p,d,q)$ model

$$\phi(L)(1-L)^d Y_t = \theta(L)\varepsilon_t, \tag{20}$$

where d denotes the fractional differencing parameter, $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p$, $\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + ... + \theta_q L^q$ and ε_t is white noise. The properties of an $ARFIMA$ process is summarized in the following theorem.

Theorem 2:
Let Y_t be an $ARFIMA(p,d,q)$ process. Then
(a) Y_t is stationary if $d < \frac{1}{2}$ and all the roots of $\phi(L) = 0$ lie outside the unit circle.

(b) Y_t is invertible if $d > -\frac{1}{2}$ and all the roots of $\theta(L) = 0$ lie outside the unit circle.

c) If $-\frac{1}{2} < d < \frac{1}{2}$, the autocovariance of Y_t, $\gamma_k = E(Y_t Y_{t-k}) \sim B \cdot k^{2d-1}$, as $k \to \infty$, where B is a function of d.

Proof:
(a). Writing $Y_t = \varphi(L)\varepsilon_t$, we have $\varphi(z) = (1-z)^{-d}\theta(z)\phi(z)^{-1}$. Now the power series expansion of $(1-z)^{-d}$ converges for all $|z| \leq 1$ when $d < \frac{1}{2}$, that of $\theta(z)$ converges for all z and θ_i since $\theta(z)$ is polynomial, and that of $\phi(z)^{-1}$ converges for all $|z| \leq 1$ when all the roots of $\phi(z) = 0$ lie outside the unit circle. Thus when all these conditions are satisfied, the power series expansion of $\varphi(z)$ converges for all $|z| \leq 1$ and so Y_t is stationary.

(b). The proof is similar to (a) except that the conditions are required on the convergence of $\pi(z) = (1-z)^d \phi(z)\theta(z)^{-1}$.

(c). See Hosking (1981) p.171.

The reason for choosing this family of $ARFIMA(p,d,q)$ process for modeling purposes is therefore obvious from Theorem 2. The effect of the d parameter on distant observation decays hyperbolically as the lag increases, while the effects of the ϕ_i and θ_j parameters decay exponentially. Thus d may be chosen to describe the **high-lag correlation** structure of a time series while the ϕ_i and θ_j parameters are chosen to describe the **low-lag correlation** structure. Indeed the long-term behavior of an $ARFIMA(p,d,q)$ process may be expected to be similar to that of an $ARFIMA(0,d,0)$ process with the same value of d, since for very distant observations the effects of the ϕ_i and θ_j parameters will be negligible. Theorem 2 shows that this is indeed so.

Exercise:

Plot the autocorrelation function for lags 1 to 50 under the following process:

(a) $(1-0.8L)Y_t = \varepsilon_t$;

(b) $(1-0.8L)Y_t = (1-0.3L)(1-0.2L)(1-0.7L)\varepsilon_t$;

(c) $(1-L)^{0.25}Y_t = \varepsilon_t$;

(d) $(1-L)^{-0.25}Y_t = \varepsilon_t$.

3.2 Occasional Breaks in trend

According to the unit root specification (6), events are occurring all the time that permanently affect Y. Perron (1989) and Rappoport and Reichlin (1989) have argued that economic events that have large permanent effects are relatively rare. This idea can be illustrated with the following model, in which Y_t is a TSP but with a single break:

$$Y_t = \begin{cases} \mu_1 + \alpha t + \varepsilon_t & for\ t < T_0 \\ \mu_2 + \alpha t + \varepsilon_t & for\ t \geq T_0. \end{cases} \tag{21}$$

We first difference (10) to obtain

$$\triangle Y_t = \xi_t + \alpha + \varepsilon_t - \varepsilon_{t-1}, \tag{22}$$

where $\xi_t = (\mu_2 - \mu_1)$ when $t = T_0$ and is zero otherwise. Suppose that ξ_t is viewed as a random variable with Bernoulli distribution,

$$\xi_t = \begin{cases} \mu_2 - \mu_1 & with\ probability\ p \\ 0 & with\ probability\ 1 - p. \end{cases}$$

Then, ξ_t is a white noise with mean $E(\xi_t) = p(\mu_2 - \mu_1)$. (11) could be rewritten as

$$\triangle Y_t = \kappa + \eta_t, \tag{23}$$

where

$$\begin{aligned} \kappa &= p(\mu_2 - \mu_1) + \alpha \\ \eta_t &= \xi_t - p(\mu_2 - \mu_1) + \varepsilon_t - \varepsilon_{t-1}. \end{aligned}$$

But η_t is the sum of a zero mean white noise process $\xi_t^* = [\xi_t - p(\mu_2 - \mu_1)]$ and an independent $MA(1)$ process $[\varepsilon_t - \varepsilon_{t-1}]$. η_t has mean zero $E(\eta_t = 0)$ and autocovariance functions, $\gamma_\tau^\eta = 0$ for $\tau \geq 2$. Therefore an $MA(1)$ representation for η_t exists, say $\eta_t = \upsilon_t - \theta\upsilon_{t-1}$. From this perspective, (11) could be viewed as an $ARIMA(0,1,1)$process,

$$\triangle Y_t = \kappa + \upsilon_t - \theta\upsilon_{t-1},$$

with a non-Gaussian distribution for the innovation υ_t which is a sum of Gaussian and Bernoulli distribution. See the plot on p.451 of Hamilton.