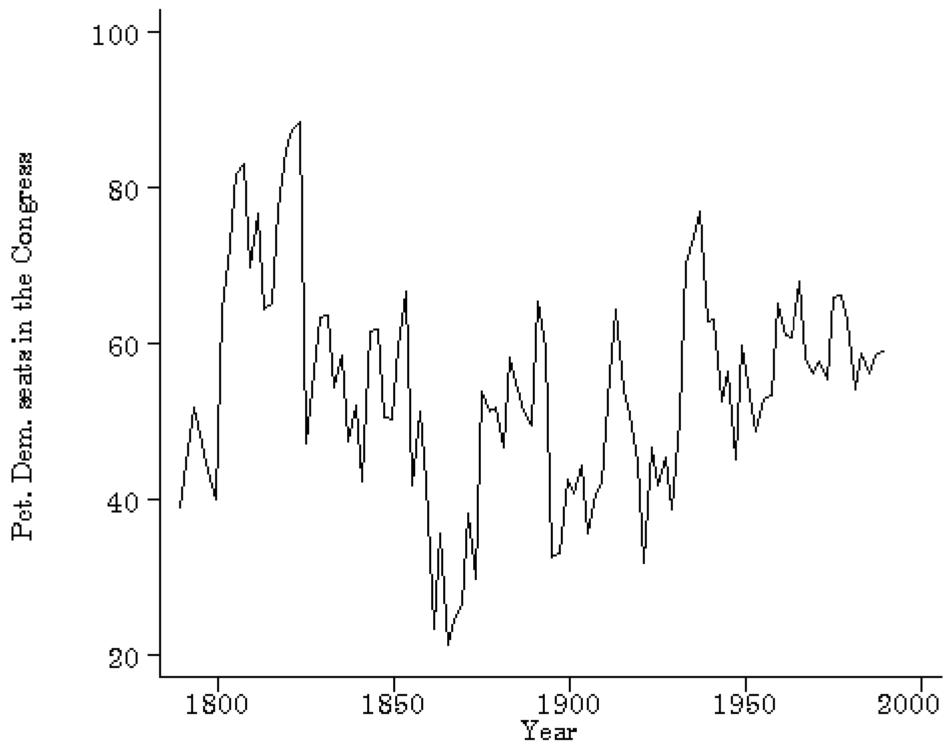
POLS571 - Longitudinal Data Analysis
September 6, 2001

FYI, we'll be using an example dataset today: The percentage of Democratically-held seats in the U.S. Congress, for the 1st - 101st Congresses (1789-1990). A plot of the series is in Figure 1.

Figure 1: Democratic Percentage of the U.S. Congress, 1789-1990



1 Properties of Time Series

1.1 Stationarity

We covered this in the last class...

1.2 Asymptotic Independence/Ergodicity

We might generally expect that the covariance between two observations Y_t and Y_{t-s} would decrease as the distance between them (the lag) increases (that is, as s increases). In the limit, we could say that two observations in a series are asymptotically uncorrelated if:

$$Cov(Y_t, Y_{t-s}) = \gamma_s \rightarrow 0 \;\; as \;\; s \rightarrow \infty \tag{1}$$

This condition is often referred to as *ergodicity*, and is a lot like stationarity.

2 Autocovariance, Autocorrelation and Partial Autocorrelation

If a series is stationary and ergoditic, then there are easy, consistent estimates for the mean, variance and autocovariance:

$$\hat{\mu} = \bar{Y} = T^{-1} \sum_{t=1}^{T} Y_t \tag{2}$$

$$\hat{\sigma^2} = T^{-1} \sum_{t=1}^{T} (Y_t - \bar{Y})^2 \tag{3}$$

$$\hat{\gamma_s} = T^{-1} \sum_{t=s+1}^{T} (Y_t - \bar{Y})(Y_{t-s} - \bar{Y}), \;\; s = 1, 2, 3, ... \tag{4}$$

One way to characterize a series as to the extent of dependence over time is by plotting its autocovariances against the number of lags s. In practice, however, we often want to "standardize" this measure, to get the *autocorrelation* function. We do so by dividing the autocovariance by the estimated variance:

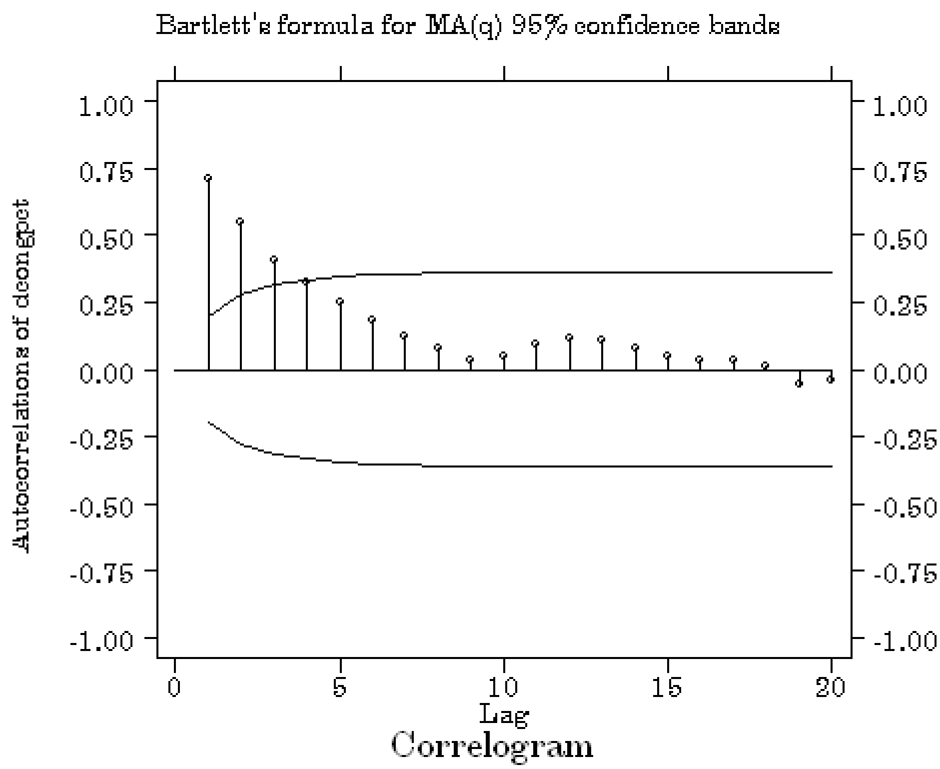
$$\hat{\rho_s} = \frac{\hat{\gamma_s}}{\hat{\sigma^2}}, \;\; s = 0, \pm 1, \pm 2, ... \tag{5}$$

Note that $\rho_0 = 1$, by definition. Plotting ρ_s against s gives the *autocorrelation function* (often abbreviated ACF). This plot is also occasionally referred to as a *correlogram.* The Stata command for this is `-ac-`.

Example...

```
.  ac dcongpct, lags(20) needle
```

Figure 2: ACF Plot, Democratic Percentage of the U.S. Congress, 1789-1990



The correlogram indicates that there's significant autocorrelation out to about four lags, and that this autocorrelation decays relatively slowly over time (which isn't all that surprising).

2.1 Partial Autocorrelation

The partial autocorrelation coefficient (typically "PAC" or "PACF") is the correlation between Y_t and Y_{t-s} after controlling for ("partialling out") the the common linear effects of the intermediate lags. The formulas for doing this are complex,[1] but the general result is that the PACFs allow us to distinguish between AR(1) and MA(q) processes (more on this later...). The relevant Stata command is `-pac-`.

Example: The PACF from the Democratic Congressional Percentage series.

```
.  pac dcongpct, lags(20) needle
```

The PACF plot suggests that the only significant partial autocorrelation occurs at one lag.

What does this mean? We'll get to that in a bit...

3 AR, MA and Integrated Processes

Time series are usually categorized according to the nature of the data-generating process which underlies them. There are three general characteristics: *autoregressive* series, *moving-average* series, and *integrated* series, plus combinations of the three (hence, ARIMA).

3.1 Integrated Processes and Random Walks

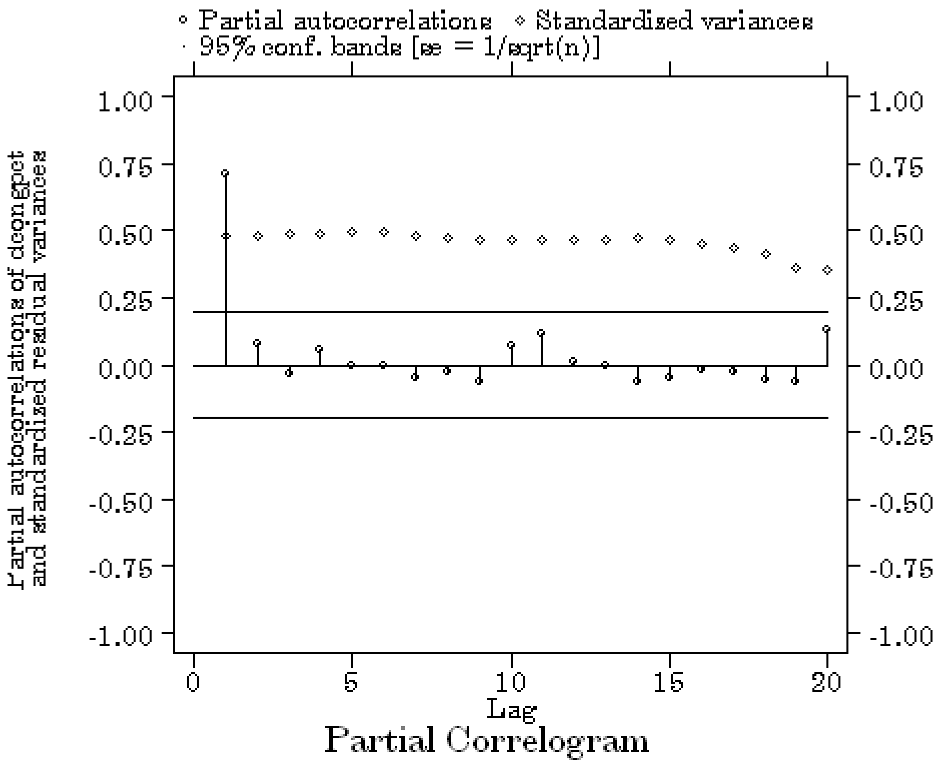
In many respects, integrated series are the simplest. Simply put, an integrated series is a series in which the value of Y_t is simply a sum of random "shocks".

3.1.1 I(1) Series

An example is the *integrated of order one* (I(1)) series :

[1] Specifically, the PACF is estimated from a solution to the Yule-Walker system of equations. A good mathematical treatment of this is in Box et. al. 1994, 64-69

Figure 3: PACF Plot, Democratic Percentage of the U.S. Congress, 1789-1990



$$Y_t = Y_{t-1} + u_t, \; u_t \sim i.i.d.(0, \sigma_u^2) \tag{6}$$

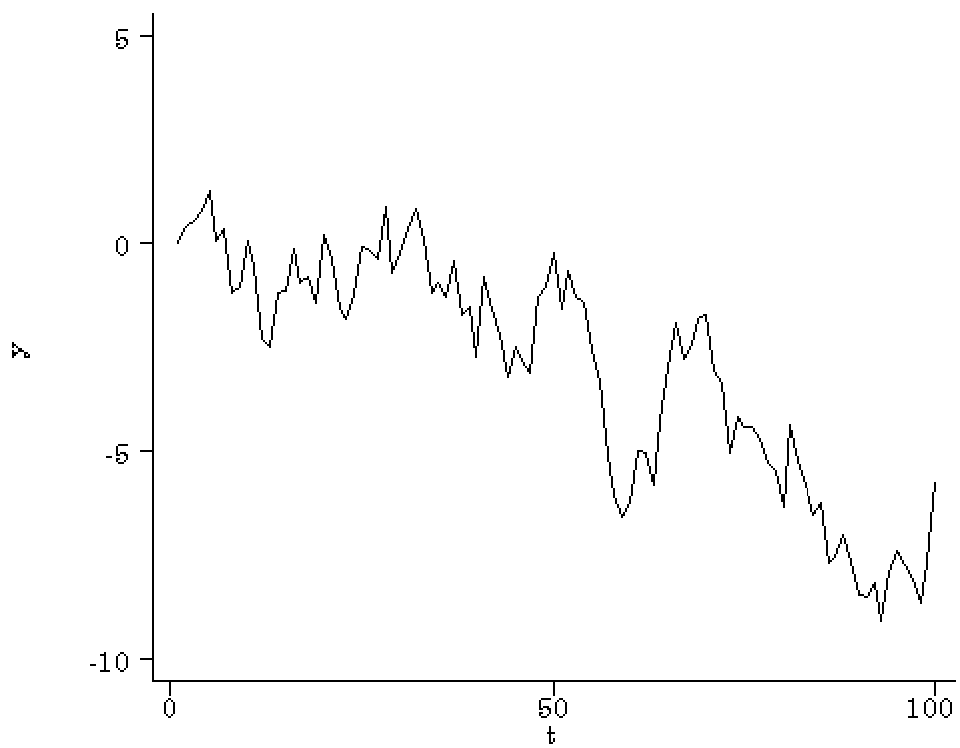
This series is also known as a *random walk*. By recursively substituting Y_{t-k-1} for Y_{t-k}, we can see that:

$$\begin{aligned} Y_t &= Y_{t-2} + u_{t-1} + u_t \\ &= Y_{t-3} + u_{t-2} + u_{t-1} + u_t \\ &= \sum_{t=0}^{T} u_t \end{aligned} \tag{7}$$

that is, that the process is simply a *sum of all past random shocks*. This means that the effect of any one shock *persists*; a large increase or decrease in Y at some t (due to a value of u_t that is especially large in magnitude) will cause the series to shift up or down until a countervailing shock comes along. As a result of this persistence, integrated series tend to drift; that is, they can take on consistently high or low values for long periods of time.

Example: Here's a random integrated series of the form $Y_t = Y_{t-1} + u_t$, where $u \sim N(0, 1)$, $Y_0 = 0$ and $T = 100$). Note that, while $E(Y_t) = 0$,[2] the value of Y_t drifts far away from zero for long periods of time.

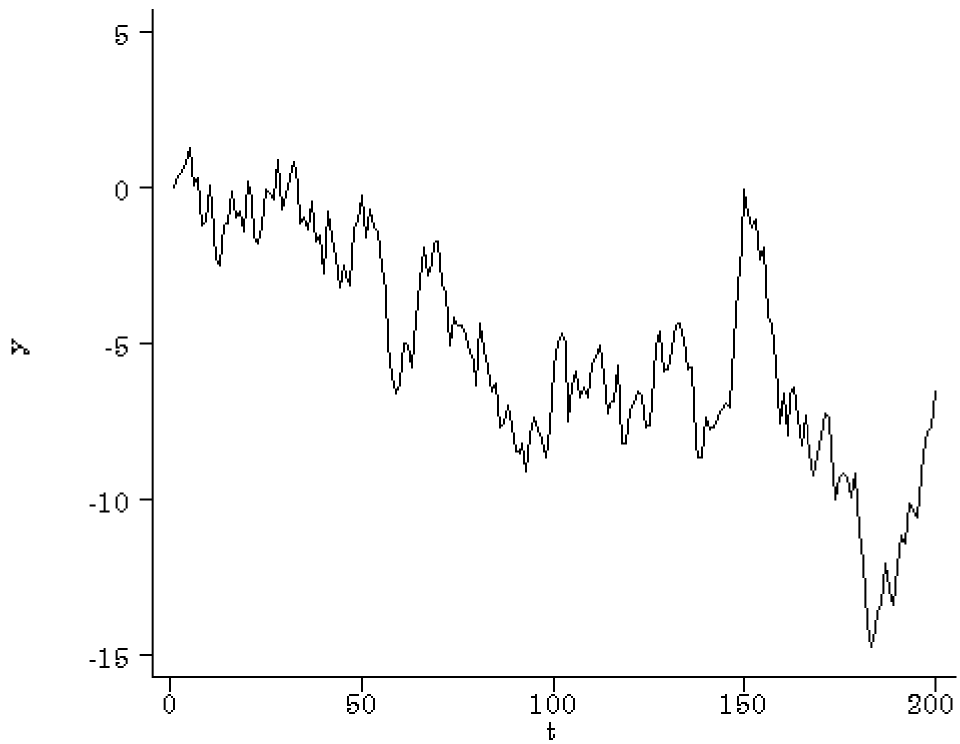
Figure 4: Random I(1) series with $u_t \sim N(0, 1)$, $T = 100$



If we extend the series to $T = 200$, we see a similar pattern.

[2] If $Y_0 \equiv u_0 \neq 0$, then the expected value of Y is Y_0.

Figure 5: Random I(1) series with $u_t \sim N(0,1)$, $T = 200$



In both of these series, we might be tempted to say that they are "trending" downwards. In fact, they aren't; the quality of persistence in an I(1) process only makes it look as if they are. In reality, they are *drifting*, and integrated series are capable of drifting away from their (theoretical) means for long periods of time. This tendency to drift means that the I(1) series is not stationary; in particular, the variance of an I(1) process is equal to:

$$Var(Y_t) \equiv E(Y_t)^2 = t\sigma^2 \tag{8}$$

and the autocovariance is equal to:

$$Cov(Y_t, Y_{t-s}) = |t-s|\sigma^2. \tag{9}$$

Both values depend on t, which means that the I(1) series is not stationary.

Because an I(1) series is simply a sum of all previous shocks, there's an easy way to model it: *differencing*. We can rearrange (6) to:

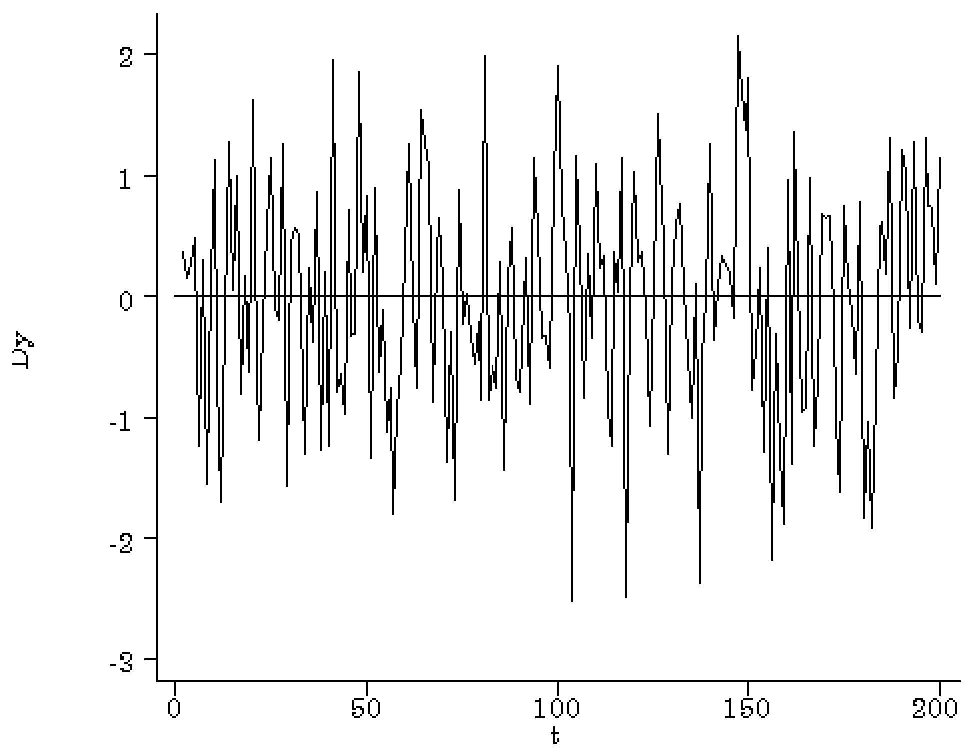
$$Y_t - Y_{t-1} = u_t \tag{10}$$

which we often write in terms of the difference operator Δ (or sometimes ∇):

$$\Delta Y_t = u_t$$

The differenced series is just the white-noise process u_t. For example, differencing the example series above gives the following graph:

Figure 6: Differenced I(1) series with $u_t \sim N(0, 1)$, $T = 200$



3.1.2 Higher-order Integrated Series

In general, the order of integration can be thought of as the number of differencings a series requires to be made stationary. A nonstationary I(1) series, differenced once, becomes a stationary one. Similarly, an I(d) series is one which, differenced d times, becomes a stationary series. One example is the I(2) process:

$$Y_t = u_t + 2u_{t-1} + 3u_{t-2} + ... \tag{12}$$

If we difference this equation, we get:

$$\begin{aligned} \Delta Y_t &= [u_t + 2u_{t-1} + 3u_{t-2} + ...] - [u_{t-1} + 2u_{t-2} + 3u_{t-3} + ...] \\ &= \sum_{j=0}^{T} u_{t-j} \end{aligned} \tag{13}$$

If we further difference this series, we get:

$$\begin{aligned} \Delta^2 Y_t \equiv (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) &= [u_t + u_{t-1} + ...] - [u_{t-1} + u_{t-2} + ...] \\ &= u_t \end{aligned} \tag{14}$$

which is a stationary series.

Another kind of series that leads to higher-order integrated processes is one with a polynomial trend. Consider the simple deterministic process:

$$Y_t = t^2 \tag{15}$$

Differencing this yields:

$$\begin{aligned} \Delta Y_t &= t^2 - (t-1)^2 \\ &= t^2 - t^2 + 2t - 1 \\ &= 2t - 1 \end{aligned} \tag{16}$$

and further differencing gives:

$$
\begin{aligned}
\Delta^2 Y_t &= (2t-1)-(2t-3) \\
&= 2 \qquad (17)
\end{aligned}
$$

that is, a series of constants. More generally, a kth order polynomial series can generally be made stationary by differencing k times. Generally, though, there aren't a lot of practical applications which involved orders of integration higher than I(1).

3.2 AR(p) Series

3.2.1 The AR(1) Model

An autoregressive series is what the name implies: a series in which past values of Y_t directly influence current values. Such models have a lot of intuitive appeal, in that they seem to reflect what we think of as direct temporal dependence. The simplest AR series is the AR(1):

$$Y_t = \phi Y_{t-1} + u_t, \;\; u_t \sim i.i.d.(0, \sigma_u^2) \qquad (18)$$

An alternative way to write (18) is simply:

$$Y_t - \phi Y_{t-1} = u_t \qquad (19)$$

Assume that the series started with a single constant value Y_0 sometime in the distant past, T periods ago. By repeatedly substituting lagged values of Y into the equation, we can rewrite this as:

$$
\begin{aligned}
Y_t &= \phi[\phi Y_{t-2} + u_{t-1}] + u_t \\
&= \phi^2 Y_{t-2} + \phi u_{t-1}] + u_t \\
&= \phi^2[\phi Y_{t-3} + u_{t-2}] + \phi u_{t-1} + u_t \\
&= \phi^3 Y_{t-3} + \phi^2 u_{t-2} + \phi u_{t-1} + u_t \\
&= ... \\
&= \sum_{j=0}^{T-1} \phi^j u_{t-j} + \phi^T Y_0 \qquad (20)
\end{aligned}
$$

The first part of this is a "moving average" of (exponentially ϕ-weighted) lagged values of the us. The second is a function of the starting value Y_0 (a.k.a. Y_{t-T}).

What is the expected value (mean) of this series?

$$\begin{aligned} E(Y_t) &= E(\sum_{j=0}^{T-1} \phi^j u_{t-j}) + E(\phi^T Y_{t-T}) \\ &= \phi^T Y_0 \end{aligned} \tag{21}$$

That is, the expected value depends on the starting value and (more importantly) ϕ.

- If $|\phi| > 1$, then the mean of the series depends on the starting value Y_0, and is increasing (for $\phi > 1$) or alternating between positive and negative with increasing amplitude (for $\phi < -1$) over time.

- If $|\phi| = 1$, then the mean of the series is exactly equal to the starting value Y_0.

- If $|\phi| < 1$, then the importance of Y_0 decreases over time; asymptotically (as $T \to \infty$), the influence of the starting value disappears, and we can write (20) as:

$$Y_t = \sum_{j=0}^{\infty} \phi^j u_{t-j} \tag{22}$$

This suggests one interpretation for an AR(1) series: As the sum of an exponentially-weighted series of random shocks. So, an AR(1) process

- is like an integrated process, in that the effects of a shock *persist* over time; however,

- the *size* of those effects *decay* over time.

- In the limit, the effect of any one shock very much later is zero.

These results point out another important fact about the AR(1) process: **An AR(1) process is mean-stationary iff** $|\phi| < 1$. Similarly, we can write the variance of a stationary AR(1) process as a function of (22):

$$
\begin{aligned}
Var(Y_t) \equiv E(Y_t^2) &= E(\sum_{j=0}^{\infty} \phi^j u_{t-j})^2 \\
&= \sum_{j=0}^{\infty} \phi^{2j} E(u_{t-j}^2) \\
&= \sigma^2 \sum_{j=0}^{\infty} \phi^{2j} \\
&= \frac{\sigma^2}{1-\phi^2} \qquad (23)
\end{aligned}
$$

and so can see that the variance is defined only for $|\phi| < 1$.

3.3 Higher-order AR Models

A more general AR(p) model can be written:

$$
Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + u_t \qquad (24)
$$

To discuss these models, let's introduce a little notation. Let L be the "lag" operator; treat this as any other algebraic operation. So, we can write:

$$
\begin{aligned}
LY_t &= Y_{t-1} \\
L^2 Y_t &= Y_{t-2} \\
L^s Y_t &= Y_{t-s} \\
L^0 Y_t &= Y_t
\end{aligned}
$$

Likewise, its easy to see that the relationship between L and the difference operator Δ is simply $\Delta = 1 - L$, e.g.:

$$
\begin{aligned}
\Delta Y_t &= Y_t - Y_{t-1} \\
&= (1-L)Y_t
\end{aligned}
$$

Using this notation, we can write the AR(1) equation (18) as:

$$(1 - \phi L)Y_t = u_t \tag{25}$$

Similarly,

$$\Delta^2 Y_t = (1 - L)^2 Y_t = (1 - 2L - L^2)Y_t = Y_t - 2Y_{t-1} + Y_{t-2}$$

This notation is useful because it allows us to write the more general AR(p) series in (24) as:

$$(1 - \phi L - \phi_2 L^2 - ... - \phi_p L^p)Y_t = u_t \tag{27}$$

We'll use this notation more when we get to general ARIMA(p,d,q) series, below.

Next time: MA models, and identifying and estimating ARIMA models in practice...