ARMA Handout

Jialin Yu

1 Linear Difference Equations

- First order systems

Let $\{\varepsilon_t\}_{t=1}^{\infty}$ denote an "input sequence" and $\{y_t\}_{t=1}^{\infty}$ denote an "output sequence" generated by

$$y_t = \phi y_{t-1} + \varepsilon_t \quad t = 1, 2, ... \text{ with } y_0 \text{ given}$$

The equation is a first order difference equation. It is easily solved by recursive substitution

$$y_t = \phi^t y_0 + \sum_{i=0}^{t-1} \phi^i \varepsilon_{t-i}$$

- p-th order systems

Consider a p-th order system

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + \varepsilon_t$$

To solve this system it is useful to write it in vector first order form (called *companion form*) as

$$\begin{bmatrix} y_t \\ y_{t-1} \\ y_{t-2} \\ ... \\ y_{t-p+1} \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 & \phi_3 & ... & \phi_p \\ 1 & 0 & 0 & ... & 0 \\ 0 & 1 & 0 & ... & 0 \\ ... & ... & ... & ... & ... \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ y_{t-2} \\ y_{t-3} \\ ... \\ y_{t-p} \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ 0 \\ 0 \\ ... \\ 0 \end{bmatrix}$$

or

$$Z_t = \Phi Z_{t-1} + e_t$$

with

$$Z_t = \begin{bmatrix} y_t \\ y_{t-1} \\ y_{t-2} \\ \cdots \\ y_{t-p+1} \end{bmatrix}$$

etc. By recursive substitution,

$$Z_t = \Phi^t Z_0 + \sum_{i=0}^{t-1} \Phi^i e_{t-i}$$

so that a solution is obtained given $Z_0 = (y_0, y_{-1}, ..., y_{t-p+1})$.

- Stability:
  - The linear difference equation $Z_t = \Phi Z_{t-1} + e_t$ is stable if, when $e_t = e$ for all t and Z_0 is an arbitrary constant, them $\lim_{t\to\infty} Z_t = Z^*$ where Z^* does not depend on Z_0
  - In the first-order case, stability is achieved if $|\phi| < 1$
  - In the higher order case, what is required is that

$$\lim_{t\to\infty} \Phi^t = 0 \text{ and } \sum_{i=0}^{t-1} \Phi^i \text{ converges}$$

These two conditions are equivalent to the conditions that all eigenvalues of Φ are less than one in modulus (absolute value). To see this, assume for the moment that the eigenvalues of Φ are distinct. In this case we can decompose

$$\Phi = P\Lambda P^{-1}$$

where the columns of P are eigenvectors of Φ and Λ is a diagonal matrix with the eigenvalues of Φ on the diagonal (see Hamilton).

Thus

$$\begin{aligned}\Phi^2 &= P\Lambda P^{-1}P\Lambda P^{-1} \\ &= P\Lambda^2 P^{-1}\end{aligned}$$

and

$$\Phi^t = P\Lambda^t P^{-1}$$

When the eigenvalues of Φ are not distinct, a similar argument can be applied to the Jordan decomposition of Φ (see Hamilton).

2 Stochastic Processes

In stochastic models, the input sequence $\{Y_t\}_{t=1}^{\infty}$ of random variables form an example of stochastic processes. Some useful definitions:

- Stochastic process: the probability law governing $\{Y_t\}_{t=1}^{\infty}$
- Realization: One "draw" from the process, i.e. $\{y_t\}_{t=1}^{\infty}$. Note that $\{y_t\}_{t=1}^{\infty}$ is only ONE observation instead of an infinite number of observations of the *stochastic process*. What is the implication and how to deal with this problem?
- Strict Stationarity: The process is strict stationary if the probability distribution of $(Y_t, Y_{t+1}, ..., Y_{t+k})$ is identical to the probability distribution of $(Y_\tau, Y_{\tau+1}, ..., Y_{\tau+k})$ for all t, τ, k. (Thus, all joint distributions are time invariant)
- Autocovariances: The autocovariances are $\lambda_{t,k} = cov\left(Y_t, Y_{t+k}^T\right)$
- Autocorrelations: The autocorrelations are $\rho_{t,k} = corr\left(Y_t, Y_{t+k}^T\right)$. If Y_1 correlates with Y_2, Y_2 correlates with Y_3, does Y_1 correlate with Y_3?
- Covariance Stationarity: The process is covariance stationary if

$$\begin{aligned}\mu_t = E\left(Y_t\right) = \mu \quad &\text{for all t} \\ \lambda_{t,k} = \lambda_k \quad &\text{for all t}\end{aligned}$$

Thus the means and autocovariances do not depend on time. When Y is scalar, covariance stationary implies

$$\lambda_k = \lambda_{-k}$$

and when Y is a vector, covariance stationarity implies

$$\lambda_k = \lambda_{-k}^T$$

Why? consider the example where

$$\underbrace{\begin{pmatrix} x_{t+1} \\ y_{t+1} \end{pmatrix}}_{Z_{t+1}} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \underbrace{\begin{pmatrix} x_t \\ y_t \end{pmatrix}}_{Z_t} + \varepsilon_{t+1}$$

Assuming Z_t is i.i.d. standard bi-variate normal. In this case,

$$\lambda_1 = Cov\left(Z_t, Z_{t+1}\right) = \begin{pmatrix} a & c \\ b & d \end{pmatrix}$$
$$\lambda_{-1} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

- Trend Stationary: The process is trend stationary if

$$Y_t - f\left(t\right)$$

is stationary, for appropriately chosen $f\left(.\right)$. In many cases, $f\left(t\right)$ is taken to be $f\left(t\right) = \alpha + \beta t$. In the case of stock price, the trend can represent for example equity premium.

- Difference Stationary: A process is difference stationary if

$$Y_t - Y_{t-1}$$

is stationary. For example, if Y_t is stock price, does difference stationary mean the stock price follows a random walk? (No, there might be other variables that can predict stock return, see the various notions of market efficiency.) What is the implication for proving stock market efficiency? It is really equivalent to prove a negative statement – there does not exist a variable that can predict stock return. Let Δ denote the differencing operator so that $\Delta Y_t = Y_t - Y_{t-1}$, then sometimes a series is said to be *Integrated of order* d written as $I(d)$ if

$$\Delta^d Y_t$$

is stationary, but $\Delta^{d-1} Y_t$ is not stationary.

- White noise: A process is called white noise if it is covariance stationary, $\mu = 0$ and $\lambda_k = 0$ for $|k| > 0$
- Martingale Process: Y_t is a martingale if

$$E\left[Y_t \left| \Omega_{t-1}\right.\right] = Y_{t-1}$$

where $\Omega_{t-1} \sqsubseteq \Omega_t$ is the time $t-1$ information set. Often $\Omega_t = \{Y_\tau\}_{\tau=0}^{t}$.

- Martingale Difference Process: Y_t is a martingale difference process if

$$E\left[Y_t \left| \Omega_{t-1}\right.\right] = 0$$

- Markov process: $\{X_t\}$ is Markov if, given X_t, the distribution of X_s for $s > t$ does not depend on X_u for $u < t$. Is a Markov process martingale? Is a martingale a markov process?

3 Autoregressive Processes

An example of stochastic process is

$$Y_T = \phi_1 Y_{T-1} + \phi_2 Y_{T-2} + ... + \phi_p Y_{T-p} + \varepsilon_t$$

where $\varepsilon_t \sim iid\left(0, \sigma^2\right)$ and $(Y_0, Y_{-1}, ...Y_{-p+1})$ is independent of $\{\varepsilon_t\}_{t=1}^{\infty}$ with mean μ_0 and variance σ_0^2. This process is called *Autoregressive Process of*

Order p, abbreviated as AR(p).

In the AR(1) model

$$Y_t = \phi Y_{t-1} + \varepsilon_t \quad \text{implies } Y_t = \phi^t Y_0 + \sum_{i=0}^{t-1} \phi^i \varepsilon_{t-i}$$

It can be verified that

$$\begin{aligned} \mu_t &= EY_t = \phi^t \mu_0 \\ \lambda_{t,0} &= VarY_t = \phi^{2t}\sigma_0^2 + \frac{1-\phi^{2t}}{1-\phi^2}\sigma^2 \text{ for } |\phi| < 1 \\ &= \sigma_0^2 + t\sigma^2 \text{ for } |\phi| = 1 \end{aligned}$$

And these are time invariant if

1. $|\phi| < 1$
2. $\mu_0 = 0$, $\sigma_0^2 = \frac{\sigma^2}{1-\phi^2}$

It then can be verified that

$$\lambda_{t,k} = \frac{\phi^k}{1-\phi^2}\sigma^2 \overset{def}{=} \lambda_k$$

which does not depend on t. Hence condition 1 and 2 are necessary and sufficient for covariance stationarity. Often only 1 is mentioned because, given 1, Y_0 only has a transitory effect on the process. But if you want to simulate a truely stationary process, the initial observation of an AR(1) model needs to be drawn from its stationary distribution. For AR(p) model, there are similar restrictions that lead to covariance stationarity. Like in AR(1), there are two sets of restrictions. The first is that the process is stable, that is all the eigenvalues of Φ are less than 1 in modulus. The second involves the mean and variance of the initial conditions. These are worked out in detail in Hamilton.

4 The Lag Operator

A useful notational device is the *lag operator*, denoted by L (some authors use B). In general, L is an operator that maps the sequence $\{y_t\}_{t=-\infty}^{\infty}$ into another sequence $\{x_t\}_{t=-\infty}^{\infty}$. Specifically, L "lags" the sequence one period. Thus

$$Ly_t = y_{t-1}, L^2 y_t = y_{t-2}, ..., L^p y_t = y_{t-p}$$

If b denotes a constant

$$bLy_t = by_{t-1} = Lby_t$$

We can use this operator to write AR(p) model as

$$\begin{aligned} Y_t &= \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + \varepsilon_t \\ &= \phi_1 L Y_t + \phi_2 L^2 Y_t + ... + \phi_p L^p Y_t + \varepsilon_t \end{aligned}$$

or

$$\left(1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p\right) Y_t = \varepsilon_t$$

or

$$\phi(L) Y_t = \varepsilon_t$$

with $\phi(L) = \left(1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p\right)$. Notice that the operator $\phi(L)$ is a $p-$th order polynomial in the lag operator L. This polynomial is called the *autoregressive polynomial.*

As an exercise, you should show that the roots of the autoregressive polynomial are the reciprocals of the eigenvalues of the companion matrix Φ (the zeros of the autoregressive polynomial are the values of z that makes $\phi(z) = 0$). Thus, often the conditions for covariance stationary are stated as "The roots of AR polynomial are greater than 1 in modulus".

The lag operator is useful because it can be manipulated in familiar algebraic ways to simplify calculations. For example, in the AR(1) process, we can write each observation as the sum of lagged shocks

$$Y_t = \phi^t Y_0 + \sum_{i=0}^{t-1} \phi^i \varepsilon_{t-i}$$

It is quite tedious to do this for AR(p) process. With the lag operator, this task can be accomplished in a less tedious way. To begin, for AR(1) process, one can write

$$(1-\phi L)\,Y_t=\varepsilon_t$$

and we want to find the operator that maps $\{\varepsilon_t\}$ into $\{Y_t\}$, that is we seek a $c(L)$ such that

$$Y_t=c(L)\,\varepsilon_t$$

Since $(1-\phi L)\,Y_t=\varepsilon_t$, it is natural to write $c(L)=(1-\phi L)^{-1}$. What does this mean? Recall

$$(1-x)^{-1}=1+x+x^2+x^3+...$$

when $|x|<1$. This suggests writing

$$(1-\phi L)^{-1}=1+\phi L+\phi^2L^2+\phi^3L^3+...$$

which yields the solution

$$Y_t=\sum_{i=0}^{\infty}\phi^i\varepsilon_{t-i}$$

which is the same one would obtain from infinite recursive substitution. Thus, inverting $(1-\phi L)$ algebraically gives the "right answer".

A word of warning however: The algebraic inverse of the polynomial $(1-\phi z)$ where z is a variable, is not unique. To see this, note

$$(1-\phi z)\,\phi^{-1}z^{-1}=\phi^{-1}z^{-1}-1$$

so that

$$1-\phi z=\frac{\phi^{-1}z^{-1}-1}{\phi^{-1}z^{-1}}$$

and thus

$$\begin{aligned}(1-\phi z)^{-1}&=-\phi^{-1}z^{-1}\left(1-\phi^{-1}z^{-1}\right)^{-1}\\&=-\phi^{-1}z^{-1}\left(1+\phi^{-1}z^{-1}+\phi^{-2}z^{-2}+...\right)\end{aligned}$$

which suggests writing

$$(1-\phi L)^{-1} = -\phi^{-1}L^{-1}\left(1+\phi^{-1}L^{-1}+\phi^{-2}L^{-2}+...\right)$$

which implies the solution

$$Y_t = \sum_{i=1}^{\infty} \phi^{-i}\varepsilon_{t+i}$$

Where does this come from?

$$Y_{t+1} = \phi Y_t + \varepsilon_{t+1}$$

hence

$$\begin{aligned} Y_t &= \phi^{-1}Y_{t+1} + \phi^{-1}\varepsilon_{t+1} \\ &= \phi^{-2}Y_{t+2} + \phi^{-2}\varepsilon_{t+2} + \phi^{-1}\varepsilon_{t+1} \\ &= ... \end{aligned}$$

This is the "forward" solution to the difference equation which could be deduced, for example, by forward recursive substitution.

How do we choose between the two solutions

$$Y_t = \sum_{i=0}^{\infty} \phi^{i}\varepsilon_{t-i}$$

and

$$Y_t = \sum_{i=1}^{\infty} \phi^{-i}\varepsilon_{t+i}$$

One approach is to impose a side condition on the solution: bounded input sequences $\{\varepsilon_t\}$ must lead to bounded output sequences. This would rule out the forward solution if $|\phi| < 1$ since ϕ^{-t} will explode as t grows large. Analogously, it would rule out the backward solution if $|\phi| > 1$. This yields a solution rule: "Solve stable roots backwards and unstable roots forward".

5 Moving Average Models

We say that Y_t follows a moving average process of order q if

$$Y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - ... - \theta_q \varepsilon_{t-q}$$

where $\varepsilon_t \sim iid\left(0, \sigma^2\right)$.

Example: where might you see moving average models. Market microstructure models. E.g., assume the observed price equals the true fundamental (which follows random walk) plus a bid-ask bounce

$$p_t = p_t^* + \varepsilon_t$$

then the return

$$\begin{aligned} r_t &= p_t - p_{t-1} \\ &= p_t^* - p_{t-1}^* + \varepsilon_t - \varepsilon_{t-1} \end{aligned}$$

which is in the form of a MA model (for now, this is easy to see by assuming $p_t^* - p_{t-1}^* = 0$, later we'll show that this assumption does not matter).

Some properties

- Covariance Stationary
  - MA(1) model: $Y_t = \varepsilon_t - \theta \varepsilon_{t-1}$
    * $E\left(Y_t\right) = 0$
    * $Var\left(Y_t\right) = \sigma^2\left(1 + \theta^2\right)$
    * $Cov\left(Y_t, Y_{t-1}\right) = -\theta\sigma^2$ (given this, is stock return positively or negatively auto-correlated if there is bid-ask bounce?)
    * $Cov\left(Y_t, Y_{t-k}\right) = 0$ for $|k| > 1$
    * Hence the process is covariance stationary
  - MA(q) model: $Y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - ... - \theta_q \varepsilon_{t-q}$
    * $E\left(Y_t\right) = 0$
    * $Var\left(Y_t\right) = \sigma^2\left(1 + \sum_{i=1}^{q} \theta_i^2\right)$

* $Cov\,(Y_t, Y_{t-k}) = \sigma^2 \left(-\theta_k + \sum_{j=1}^{q-k} \theta_j \theta_{k+j}\right)$ for $k \leq q$
* $Cov\,(Y_t, Y_{t-k}) = 0$ for $|k| > q$
* Hence the process is covariance stationary

- Invertibility
  - Motivation: Consider forecasting in MA(1) model $Y_t = \varepsilon_t - \theta\varepsilon_{t-1}$. Thus, the forecast of Y_t constructed at time $t-1$ would be $-\theta\varepsilon_{t-1}$. The problem is that ε_{t-1} is not directly observed; it must be constructed from the lagged values of Y_t. How can this be done? Note

$$\varepsilon_t = \theta^t \varepsilon_0 + \sum_{i=0}^{t-1} \theta^i Y_{t-i}$$

    Thus, the sequence of $\varepsilon' s$ could be formed from present and lagged $Y' s$ if ε_0 were known. But of course ε_0 is unknown. The process is invertible (meaning that the $\varepsilon' s$ can be determined from lagged $Y' s$) if

$$\widehat{\varepsilon_t} = \sum_{i=0}^{t-1} \theta^i Y_{t-i} \overset{qm}{\to} \varepsilon_t \quad \text{as } t \to \infty$$

    i.e., assuming $\varepsilon_0 = 0$ has no lasting effect on the quality of the forecasts.
  - For the MA(1) model, invertibility requires $|\theta| < 1$. Equivalently, writing $Y_t = (1 - \theta L)\,\varepsilon_t$, invertibility requires that the roots of $\theta\,(z) = 1 - \theta z$ are greater than 1 in modulus
  - For the MA(q) model, we get an analogously result, namely that the process is invertible if the roots of $\theta\,(z) = 1 - \theta_1 z - \theta_2 z^2 - ... - \theta_q z^q$ are greater than 1 in modulus
  - Another motivation for the invertibility restriction comes from considering the autocovariances of the MA(1) model

$$\lambda_0 = \sigma^2 \left(1 + \theta^2\right)$$
$$\lambda_1 = -\theta\sigma^2$$

Note we could also describe these two autocovariances as

$$\lambda_0 = \widetilde{\sigma}^2 \left(1 + \widetilde{\theta}^2\right)$$
$$\lambda_1 = -\widetilde{\theta}\widetilde{\sigma}^2$$

with $\widetilde{\theta} = \theta^{-1}$ and $\widetilde{\sigma}^2 = \sigma^2 \left(1 + \widetilde{\theta}^2\right)^{-1} \left(1 + \theta^2\right)$. That is, the two MA(1) processes

$$Y_t = \varepsilon_t - \theta \varepsilon_{t-1} \text{ with } Var\left(\varepsilon_t\right) = \sigma^2$$

and

$$Y_t = \widetilde{\varepsilon}_t - \widetilde{\theta} \widetilde{\varepsilon}_{t-1} \text{ with } Var\left(\widetilde{\varepsilon}_t\right) = \widetilde{\sigma}^2$$

have exactly the same autocovariances. Thus, given data on Y_t, we can't tell the processes apart, at least using the first two moments of the data. These two models are *Observationally Equivalent* for their first two moments. Since $\widetilde{\theta} = \theta^{-1}$, one way to (arbitrarily) choosen between them is to impose the restriction that the MA parameter is ≤ 1 in absolute value.

6 Autocovariance Generating Functions

The autocovariance generating function for a covariance stationary process is given by

$$\lambda\left(z\right) = \sum_{j=-\infty}^{\infty} \lambda_j z^j$$

so that the autocovariances are given by the coefficients on the argument z^j. It's purpose (or one of the purposes) is the same as the moment generating function - namely it is a convenient way to "store" the autocovariances of a covariance stationary stochastic process.

For the MA process, the ACGF is particularly easy to construct. Suppose

$$Y_t = \theta\left(L\right)\varepsilon_t$$

then

$$\lambda(z) = \sigma^2 \theta(z) \theta\left(z^{-1}\right)$$

We will verify this formula for MA(1) process - you should verify it for higher order MA processes

For an MA(1) process $\theta(z) = (1 - \theta z)$, so that

$$\theta(z) \theta\left(z^{-1}\right) = (1 - \theta z)\left(1 - \theta z^{-1}\right) = -\theta z^{-1} + \left(1 + \theta^2\right) - \theta z$$

which implies autocovariances

$$\begin{aligned} \lambda_{-1} &= -\theta \sigma^2 \\ \lambda_0 &= \sigma^2 \left(1 + \theta^2\right) \\ \lambda_1 &= -\theta \sigma^2 \end{aligned}$$

with all other autocovariances equal to 0. Thus, the formula yields the correct answer for an MA(1) process.

Now, return to the issue of invertibility. Consider the MA polynomial for the MA(q) model

$$\theta(z) = 1 - \theta_1 z - \theta_2 z^2 - ... - \theta_q z^q$$

Suppose that this polynomial has zeros at $z = \gamma_1^{-1}, \gamma_2^{-1}, ..., \gamma_q^{-1}$. In this case, we can factor the polynomial as

$$\theta(z) = (1 - \gamma_1 z)(1 - \gamma_2 z) ... (1 - \gamma_q z)$$

and so that ACGF is given by

$$\lambda(z) = \sigma^2 (1 - \gamma_1 z)(1 - \gamma_2 z) ... (1 - \gamma_q z)\left(1 - \gamma_1 z^{-1}\right)\left(1 - \gamma_2 z^{-1}\right) ... \left(1 - \gamma_q z^{-1}\right)$$

But since $(1 - \gamma z)\left(1 - \gamma z^{-1}\right)$ is proportional to $\left(1 - \gamma^{-1} z\right)\left(1 - \gamma^{-1} z^{-1}\right)$ (recall the discussion of invertibility in the MA(1) model), we can "flip" or invert the roots of the MA polynomial, change σ^2 to adjust for the factor of porportionality, and obtain the same ACGF – hence a model with the same autocovariances.

Thus, for example, in the MA(2) model, if

$$\theta(z) = (1 - \gamma_1 z)(1 - \gamma_2 z)$$

then models with

$$\theta_1(z) = \left(1 - \gamma_1^{-1} z\right)(1 - \gamma_2 z)$$
$$\theta_2(z) = (1 - \gamma_1 z)\left(1 - \gamma_2^{-1} z\right)$$
$$\theta(z) = \left(1 - \gamma_1^{-1} z\right)\left(1 - \gamma_2^{-1} z\right)$$

are observationally equivalent.

7 Autoregressive-Moving Average (ARMA) Models

- Autoregressive-Moving Average models combine the simple AR and MA models. The ARMA(p,q) model is

  $$Y_t = \phi_1 Y_{T-1} + \phi_2 Y_{T-2} + ... + \phi_p Y_{T-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - ... - \theta_q \varepsilon_{t-q}$$

  or

  $$\phi(L) Y_t = \theta(L) \varepsilon_t$$

  with

  $$\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p$$

  and

  $$\theta(L) = 1 - \theta_1 L - \theta_2 L^2 - ... - \theta_q L^q$$

- Conditions for covariance stationary and invertibility are just the same as in the simple models: the roots of $\phi(z)$ and $\theta(z)$ are greater than 1 in modulus.

- The ACGF for the ARMA model can be derived as follows. Since

$$\phi(L)\,Y_t = \theta(L)\,\varepsilon_t$$

then

$$Y_t = c(L)\,\varepsilon_t$$

with $c(L) = \phi(L)^{-1}\theta(L)$ which is a well defined (mean square convergent) polynomial in positive powers of L (i.e. backward looking) if the roots of $\phi(z)$ are greater than 1 in absolute value. Thus Y_t has the MA representation $Y_t = c(L)\,\varepsilon_t$ so that

$$\begin{aligned}\lambda(z) &= \sigma^2 c(z)\,c\left(z^{-1}\right) \\ &= \sigma^2\phi(z)^{-1}\theta(z)\,\phi\left(z^{-1}\right)^{-1}\theta\left(z^{-1}\right)\end{aligned}$$

8 ARIMA Models

Suppose Y_t is integrated of order d so that Y_t must be differenced d times to be stationary. Let $X_t = (1-L)^d\,Y_t$ and suppose that X_t follows an ARMA(p,q) model. Then we say that Y_t follows an ARIMA(p,d,q) model. (The extra "I" in ARIMA stands for "integrated").