A Beginner's Introduction

To

Box-Jenkins Models

I. Introduction

In their seminal work, Time Series Analysis: Forecasting and Control (1970, Holden Day), Professors Box and Jenkins introduced a set of time series models, later called Box-Jenkins models, that revolutionized the analysis of time series data - observations on a variable observed over time at regular time intervals. An example would be real Gross Domestic Product produced by the Commerce Department of the U.S. Government that is published quarterly. See below for a graph of the real Gross Domestic Product observed from 1947Q1 to 2012Q3.


Real GDP Time Series
X=Date Y=RGDP Series
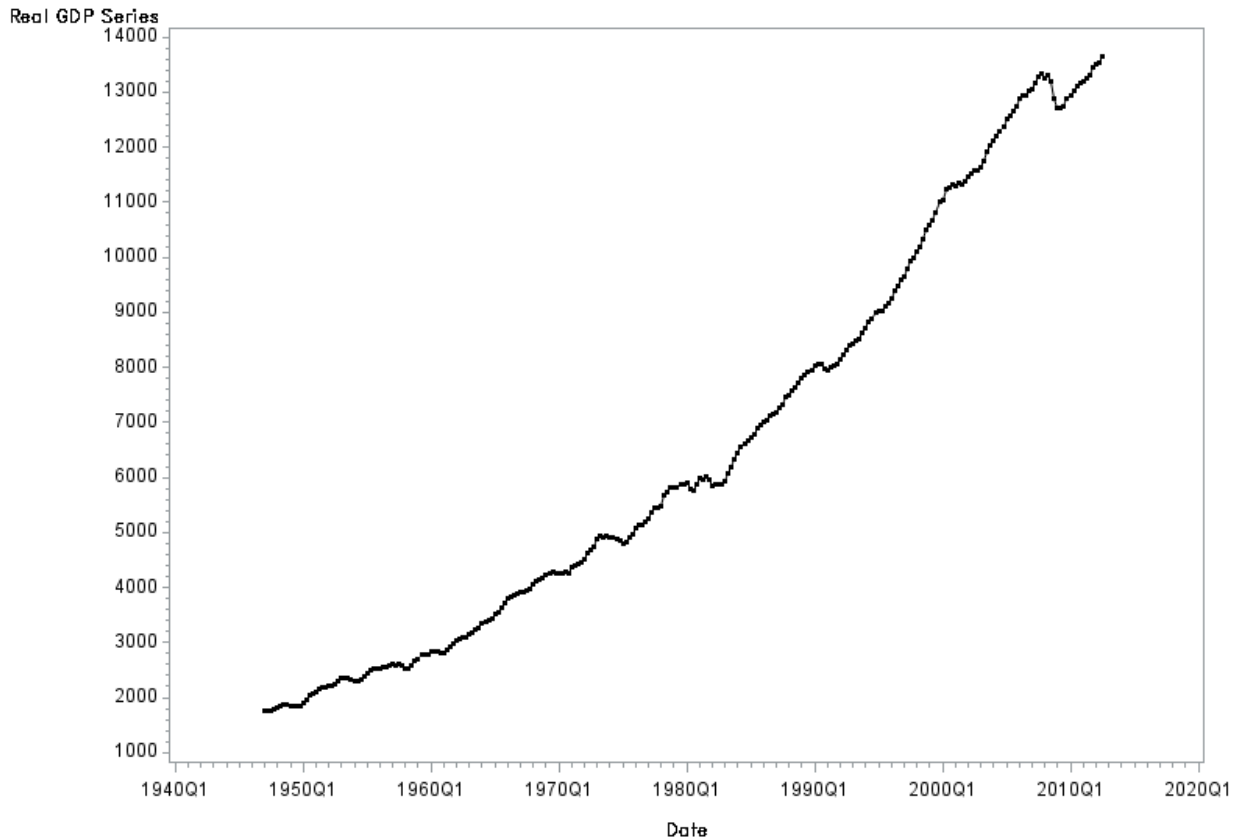
In a SAS program titled "Real GDP.sas" I have estimated a Box-Jenkins model for the quarterly, annualized growth rates of this data for the span of time 1980Q1 to 2012Q3. Below you will find the

growth rate of Real GDP plotted for this period. The growth rate at time t is calculated as $y_t = (\log(RGDP_t) - \log(RGDP_{t-1})) \cdot 400$. This is the quarterly annualized growth rate in percentage terms.

Annual Growth Rate of Real GDP - 1980Q1 - 2012Q3
X=Date Y=Growth in RGDP
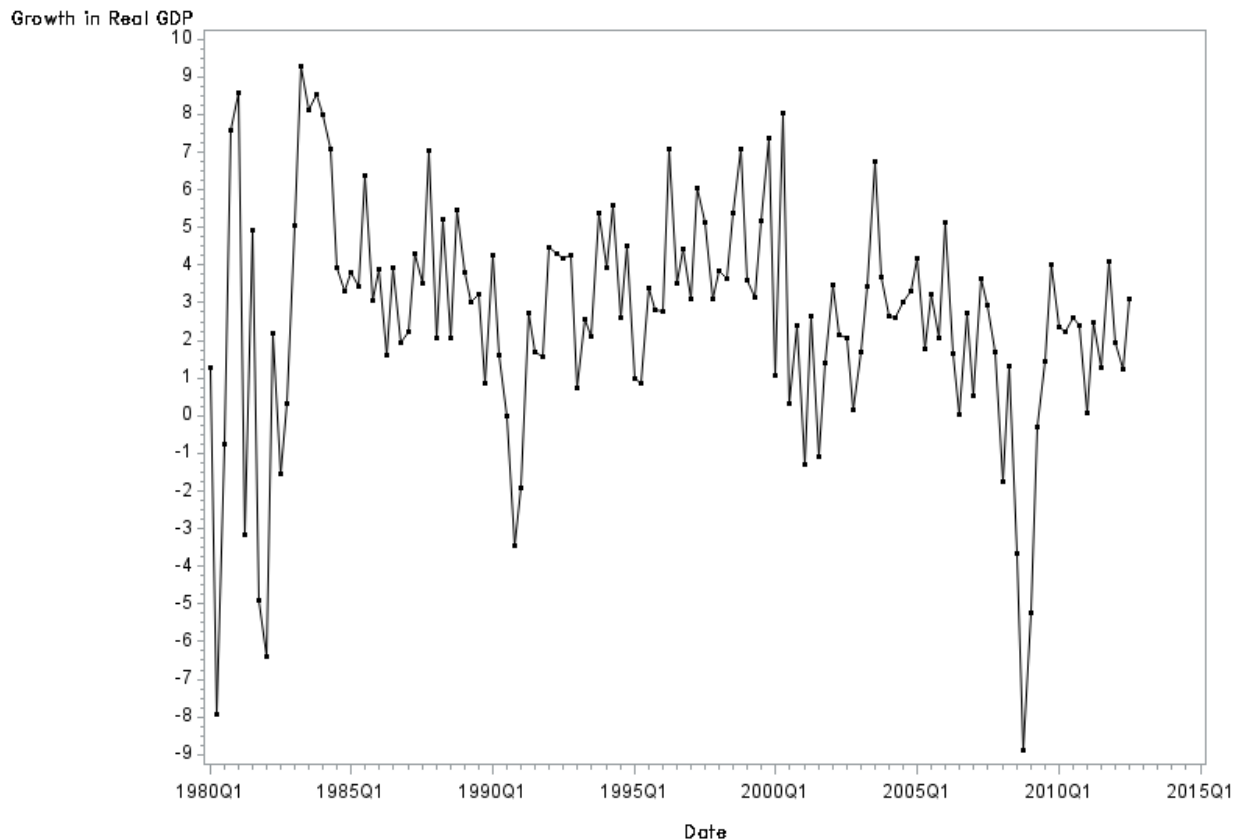
Following the traditional approach of Box-Jenkins modeling, a very good model for this data is as follows:

$$y_t = 1.48 + 0.42y_{t-1} + \hat{a}_t \qquad (1)$$

Equation (1) is obtained by using the growth data and applying least squares to the general Autoregressive Model of order one model, denoted AR(1),

$$y_t = \emptyset_0 + \emptyset_1 y_{t-1} + a_t . \qquad (2)$$

The quarterly per annum growth rate observed at time t is represented by y_t while last period's growth rate is denoted by y_{t-1}. Therefore, the least squares estimate of $\emptyset_0$, denoted $\widehat{\emptyset}_0$, is 1.48. This is the so-called constant term of the model. The least squares estimate of the autoregressive coefficient $\emptyset_1$, denoted $\widehat{\emptyset}_1$, is 0.42. The residuals of this model are represented by $\hat{a}_t$ and represent the lack of perfect

fit of the model to the y_t observations. The estimated mean of this data is 1.48/(1 − 0.42) = 2.56% quarterly per annum growth rate. If one were standing at the time 2012Q3 and wanted to forecast 4 future growth rates using the Box-Jenkins approach and model (1) you would get

Forecasts for variable grgdp				
Obs	Forecast	Std Error	95% Confidence Limits	
132	2.7712	2.7720	-2.6618	8.2041
133	2.6495	3.0100	-3.2500	8.5489
134	2.5979	3.0506	-3.3812	8.5771
135	2.5762	3.0579	-3.4172	8.5695

The four-step-ahead forecasts are 2.77%, 2.64%, 2.59%, and 2.57%. 2.77% is the one-step-ahead (one quarter ahead) forecast while 2.57% is the four-step-ahead (four quarters ahead) forecast. Notice the pretty wide 95% confidence intervals. Future growth rates could either be positive or negative with the tendency being on the positive growth rate side and the best-guess growth rates being the numbers (point forecasts) listed above. The last available growth rate we had was that of 2012Q3, namely, 3.0587%. Therefore, using Box-Jenkins methods the 4 forecasts above were generated by the formula

$$\hat{y}_{t+h} = 2.56 + (3.0587 - 2.56)(0.42)^h \tag{2}$$

where h denotes the forecast horizon and h = 1, 2, 3, and 4 in the current case. More generally, the forecasting equation for the AR(1) model is

$$\hat{y}_{t+h} = \hat{\mu} + (y_t - \hat{\mu})(\widehat{\emptyset}_1)^h \tag{3}$$

where $\hat{\mu}$ denotes the estimated mean of the y_t data.

Notice that these forecasts revert to the mean (2.56) over time. This reversion to the mean is a hallmark characteristic of what we will eventually denote as a stationary Box-Jenkins model.

The 95% prediction confidence intervals above are calculated as

$$\hat{y}_{t+h} \pm 1.96 se(\hat{y}_{t+h}) \tag{4}$$

where

$$se(\hat{y}_{t+h}) = \sqrt{\hat{\sigma}_a^2 (1 + \widehat{\emptyset}_1^2 + \widehat{\emptyset}_1^4 + \cdots + \widehat{\emptyset}_1^{2(h-1)})} . \tag{5}$$

From our computer printout we have as the estimate of the sample variance of the errors, $\hat{\sigma}_a^2 = 7.68$ and the autoregressive coefficient, $\widehat{\emptyset}_1 = 0.42$. Therefore, as the above SAS Proc Arima output indicates $se(\hat{y}_{t+1})$ = 2.7720, $se(\hat{y}_{t+2})$ = 3.0100, $se(\hat{y}_{t+3})$ = 3.0506, and $se(\hat{y}_{t+4})$ = 3.0579. As we are dealing

with a stationary time series here we not that as $h \to \infty$, $se(\hat{y}_{t+h})$ approaches the unconditional standard deviation of the data, $\sqrt{\hat{\sigma}_a^2/(1 - \hat{\emptyset}_1^2)} = \sqrt{7.68/(1 - 0.42^2)} = 3.056$.

Of course, if one wants to convert these forecasts into the "level" of real GDP four periods (quarters) ahead we should use the following formulas:

$$\hat{l}_{t+1} = 13652.51 \cdot (1 + \hat{y}_{t+1}/400)^{0.25} = 13676.09 \text{ (2012Q4 forecast)}$$

$$\hat{l}_{t+2} = \hat{l}_{t+1}(1 + \hat{y}_{t+2}/400)^{0.25} = 13698.68 \text{ (2013Q1 forecast)}$$

$$\hat{l}_{t+3} = \hat{l}_{t+2}(1 + \hat{y}_{t+3}/400)^{0.25} = 13720.87 \text{ (2013Q2 forecast)}$$

$$\hat{l}_{t+4} = \hat{l}_{t+3}(1 + \hat{y}_{t+4}/400)^{0.25} = 13742.91 \text{ (2013Q3 forecast)}$$

Here, 13652.51 is the last available real GDP figure as of 2012Q3.

So what is the logic of all of this number crunching? To get a sense of how Box-Jenkins modeling works we need to derive some results for the AR(1)model of equation (2).

II. The AR(1) Box-Jenkins Model

To get a beginning understanding of Box-Jenkins models let us consider the AR(1) model of equation (2) but, for now, we will let $\emptyset_0 = 0$. Thus, the model we are considering is

$$y_t = \emptyset_1 y_{t-1} + a_t . \qquad (6)$$

The error terms of this model, a_t, are assumed to have the following "White Noise" properties:

$E(a_t) = 0$	for all t	(zero mean assumption)	(6i)
$E(a_t^2) = \sigma_t^2$	for all t	(constant variance assumption)	(6ii)
$E(a_s a_t) = 0$	for all s $\neq$ t	(independence assumption)	(6iii)
a_t is normally distributed for all t		(Normality assumption)	(6iv)

In order for model (6) to have a constant mean, variance, and covariances we require that $|\emptyset_1| < 1$.

The so-called moving average representation of model (6) is derived by recursive substitution as follows:

$$y_t = \emptyset_1(\emptyset_1 y_{t-1} + a_{t-1}) + a_t$$

$$= \emptyset_1^2 y_{t-1} + \emptyset_1 a_{t-1} + a_t \quad .$$

Substituting s times produces

$$y_t = \emptyset_1^{s+1} y_{t-(s+1)} + a_t + \emptyset_1 a_{t-1} + \emptyset_1^2 a_{t-2} + \cdots + \emptyset_1^s a_{t-s} \quad .$$

Letting $s = t - 1$ produces

$$y_t = \emptyset_1^t y_0 + a_t + \emptyset_1 a_{t-1} + \emptyset_1^2 a_{t-2} + \cdots + \emptyset_1^{t-1} a_1 \qquad (7)$$

where y_0 denotes the initial value of the time series. Obviously, as time (t) goes to infinity, the initial value, y_0, has a negligible effect on the series since $|\emptyset_1| < 1$. Therefore, without loss of generality, we can just set $y_0 = 0$ and letting t be large we have the basic description of y_t as

$$y_t = a_t + \emptyset_1 a_{t-1} + \emptyset_1^2 a_{t-2} + \cdots \qquad (8)$$

Now we are, in turn, going to derive the mean, variance, covariances, and autocorrelations of the AR(1) process (6) above.

The Mean

Using the linearity of the expectation operator E(.) we have

$$E(y_t) = E(a_t + \emptyset_1 a_{t-1} + \emptyset_1^2 a_{t-2} + \cdots + \emptyset_1^{t-1} a_1)$$

$$= E(a_t) + \emptyset_1 E(a_{t-1}) + \emptyset_1^2 E(a_{t-2}) + \cdots + \emptyset_1^{t-1} E(a_1) = 0 \qquad (9)$$

because of the White Noise assumption (6i). Therefore, the mean of model (6) is zero.

The Variance

Using the White Noise independence property (6iii) it is straight forward to derive the variance of the process (6). Namely,

$$Var(y_t) = E(y_t - E(y_t))^2 = E(y_t^2) = E(a_t + \emptyset_1 a_{t-1} + \emptyset_1^2 a_{t-2} + \cdots)^2$$

$$= E(a_t^2) + \emptyset_1^2 E(a_{t-1}^2) + \emptyset_1^4 E(a_{t-2}^2) + \cdots$$

$$= \sigma_a^2 + \emptyset_1^2 \sigma_a^2 + \emptyset_1^4 \sigma_a^2 + \cdots$$

$$= \sigma_a^2 (1 + \emptyset_1^2 + \emptyset_1^4 + \cdots)$$

Then since the expression in the parentheses above represents a geometric progression with radius of $\emptyset_1^2 < 1$, have as the unconditional variance of (6)

$$Var(y_t) = \sigma_a^2 (1 + \emptyset_1^2 + \emptyset_1^4 + \cdots) = \frac{\sigma_a^2}{1 - \emptyset_1^2}. \qquad (10)$$

The Auto-Covariance Function

Now let us consider the autocovariance between an observation at time t (y_t) and an observation the period before (y_{t-1}). By definition we have the first-order autocorrelation as

$$\gamma_1 = E[(y_t - E(y_t))(y_{t-1} - E(y_{t-1}))] = E(y_t y_{t-1})$$

$$= E[(a_t + \emptyset_1 a_{t-1} + \emptyset_1^2 a_{t-2} + \cdots)(a_{t-1} + \emptyset_1 a_{t-2} + \emptyset_1^2 a_{t-3} + \cdots)]$$

$$= E(\emptyset_1 a_{t-1}^2 + \emptyset_1^3 a_{t-2}^2 + \emptyset_1^5 a_{t-3}^2 + \cdots) = \emptyset_1 E(a_{t-1}^2) + \emptyset_1^3 E(a_{t-2}^2) + \emptyset_1^5 E(a_{t-3}^2) + \cdots$$

$$= \emptyset_1 \sigma_a^2 (1 + \emptyset_1^2 + \emptyset_1^4 + \cdots) = \emptyset_1 \sigma_a^2 / (1 - \emptyset_1^2). \tag{11}$$

In general the j-th order autocovariance between an observation at time t (y_t) and an observation j-periods before (y_{t-j}) is

$$\gamma_j = E\left[(y_t - E(y_t))\left(y_{t-j} - E(y_{t-1})\right)\right] = E(y_t y_{t-j})$$

$$= E[(a_t + \emptyset_1 a_{t-1} + \emptyset_1^2 a_{t-2} + \cdots)(a_{t-j} + \emptyset_1 a_{t-j-1} + \emptyset_1^2 a_{t-j-2} + \cdots)]$$

$$\emptyset_1^j \sigma_a^2 (1 + \emptyset_1^2 + \emptyset_1^4 + \cdots) = \emptyset_1^j \sigma_a^2 / (1 - \emptyset_1^2). \tag{12}$$
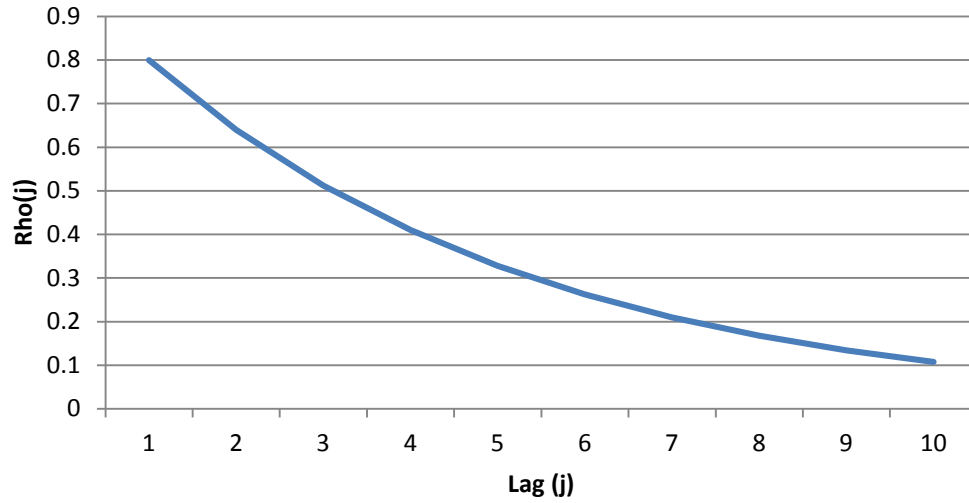
The Autocorrelation Function

The autocorrelation function is just a conversion of the autocovariance function into a measurement of association between time series observations that is not dependent on the scale of measurement of the time series (units, hundreds, thousands, millions, etc.).

The autocorrelations between time series observations that are j periods apart is

$$\rho_j = \frac{Cov(y_t, y_{t-j})}{\sqrt{var(y_t)Var(y_{t-j})}} = \frac{\gamma_j}{Var(y_t)} = \emptyset_1^j \quad . \tag{13}$$

Consider, for example, when $\emptyset_1 = 0.8$ in model (6) the autocorrelation function is

ACF for AR(1) Model with Phi1 = 0.8

But, consider when $\emptyset_1 = -0.8$ in model (6). The autocorrelation function is

ACF for AR(1) Model with Phi1 = -0.8