LECTURE - APPLIED Regression

Time Series

An Example

Probelm: Monthly gold prices. The fluctuation of gold prices is a reflection of the strength or weakness of the U.S. dollar. The data file shows monthly gold prices from January 2001 to December 2008. Suppose we want to model the monthly price, y_t , as a function of t , where t represents month (i.e., t = 1, 2, 3, . . ., 96).

(a) Construct a scatter-plot of the data. Do you observe a long-term trend?

(b) Propose a time series model that includes a long-term quadratic trend and autocorrelated residuals.

(c) Fit the time series model, part b. Identify and interpret

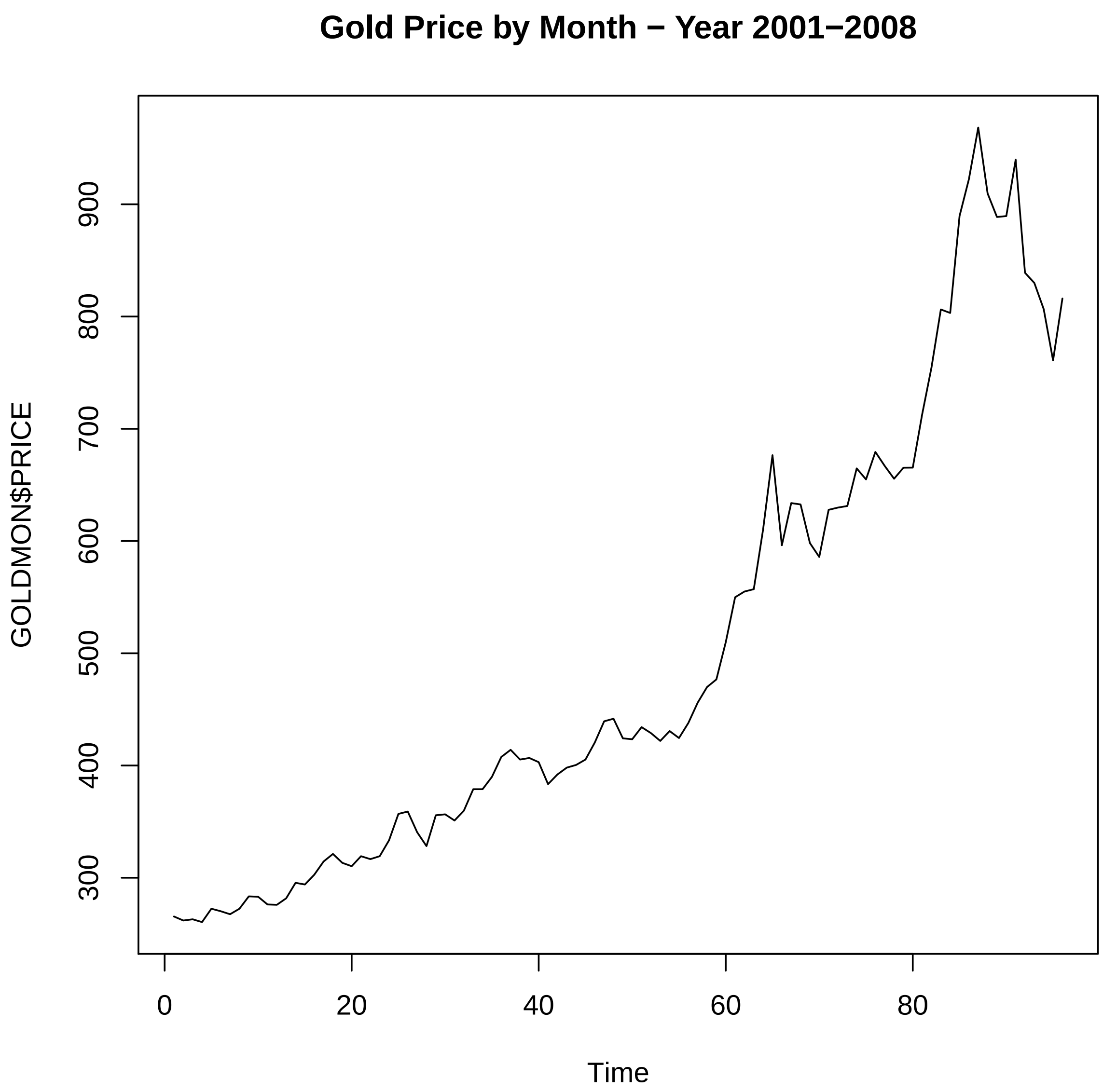
(i) the estimates of the model parameters,

(ii) the value of R^2, and

(iii) the test for a quadratic long-term trend.

(d) Find the forecast for the next time period(t=97) and its 95% interval estimate.

(a) Below is the plot of the data: It seems that price has a quadratic relationship with price.



Data For the Problem:

Original Data

	YEAR	MONTH	PRICE
1	2001.00	Jan	265.50
2	2001.00	Feb	261.90
3	2001.00	Mar	263.00
4	2001.00	Apr	260.50
5	2001.00	May	272.40
6	2001.00	Jun	270.20

Original Data with Time (tm)

	YEAR	MONTH	PRICE	tm	tm2
1	2001.00	Jan	265.50	1.00	1.00
2	2001.00	Feb	261.90	2.00	4.00
3	2001.00	Mar	263.00	3.00	9.00
4	2001.00	Apr	260.50	4.00	16.00
5	2001.00	May	272.40	5.00	25.00
6	2001.00	Jun	270.20	6.00	36.00

Model: $$Y_t = \beta_0 + \beta_1 \times t + \beta_2 \times t^2 + \epsilon_t$$

In this problem $$Price = \beta_0 + \beta_1 \times tm + \beta_2 \times tm2 + \epsilon_t$$

where $\epsilon_t = \phi \times \epsilon_{t-1} + a_t$ and $a_t \sim N(0, \sigma_a^2)$

(c) As part the parameter estimates is the estimation of autocorrelation from the Step-1 MLR model we run the MLR model and estimate the autocorrelation (as $\hat{\phi}$). Result from MLR model:

Code: mdl1=lm(PRICE tm + tm2, data=gld); summary(mdl1) ; anova(mdl1)

ar=acf(resid(mdl1),lag.max=1); ar

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	275.3876	13.9806	19.70	0.0000
tm	0.1738	0.6653	0.26	0.7945
tm2	0.0681	0.0066	10.26	0.0000

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
tm	1	3392731.45	3392731.45	1697.04	0.0000
tm2	1	210246.83	210246.83	105.17	0.0000
Residuals	93	185926.16	1999.21		

The autocorrelation $\hat{\phi} = 0.792$.

In the next step, the parameter can be estimated using many methods and they vary to some extent(See the R-Program), we will the result from the Cochran-Orcutt output as it has answers for all the questions.

The output is as follows:

Parameter	Estimate	Std. Error	t-value	$Pr(>\|t\|)$
(Intercept)	252.252591	69.120347	3.649	0.0004358
tm	1.735612	2.855033	0.608	0.5447428
tm2	0.049888	0.025558	1.952	0.0539811

Residual standard error: 25.8046 on 92 degrees of freedom

Multiple R-squared: 0.5943 , Adjusted R-squared: 0.5855

F-statistic: 67.4 on 2 and 92 DF, $p - value :< 9.484e - 19$

Hence the final estimated model is:

$$\hat{y}_t = \hat{\beta}_0 \ + \ \hat{\beta}_1 \times t + \hat{\beta}_2 \times t^2 \ + \ \hat{\phi} \times \hat{\epsilon}_{t-1}$$

Using the fact that $\hat{\epsilon}_{t-1} = y_{t-1} - \hat{\beta}_0 \ - \ \hat{\beta}_1 \times (t-1) - \hat{\beta}_2 \times (t-1)^2$

it also can be written as

$$\hat{y}_t = \hat{\phi} \times y_{t-1} + (1-\hat{\phi}) \times \hat{\beta}_0 \ + \ \hat{\beta}_1 \times (t - \hat{\phi}.(t-1)) + \hat{\beta}_2 \times (t^2 - \hat{\rho}.(t-1)^2)$$

For this example, we have

$$\hat{y}_t = 252.25 \ + \ 1.736 \times t + 0.04989 \times t^2 \ + \ 0.792 \times \hat{\epsilon}_{t-1}$$

The adjusted R-square reported is 58.55%.

To the quadratic part we set $H_0 : \beta_2 = 0 \ vs \ H_1 : \beta_2 \neq 0$.

t-stat = 1.952 and p-value for that is 0.0539. Hence at 5% level we do not reject the null hypothesis.

Now to forecast for next (97th period) we use the equation

$$\hat{y}_{97} = 252.25 \ + \ 1.736 \times 97 + 0.04989 \times 97^2 \ + \ 0.792 \times \hat{\epsilon}_{97-1}$$

So we need to fist find $\hat{\epsilon}_{97-1}$ as

$$
\begin{aligned}
\hat{\epsilon}_{96} &= y_{96} - \hat{\beta}_0 \; - \; \hat{\beta}_1 \times 96 - \hat{\beta}_2 \times 96^2 \\
&= 816.1 - 252.25 \; - \; 1.736 \times 96 - 0.04989 \times 96^2 \\
&= -62.5922
\end{aligned}
$$

$$
\begin{aligned}
\text{Now} \quad \hat{y_{97}} &= 252.25 \; + \; 1.736 \times 97 + 0.04989 \times 97^2 \; + \; 0.792 \times (-62.5922) \\
&= 840.484
\end{aligned}
$$

Now to compute the 95% interval estimate of the forecast, we use the formula,

$$
\hat{y}_t \; \pm \; z_{\alpha/2}.\hat{\sigma}_a
$$

From the R-output we get $\hat{\sigma}_a = 25.80$, as results the 95% interval estimate of the forecast is

$$
\hat{y}_{97} \; \pm \; z_{\alpha/2}.\hat{\sigma}_a \; = \; 840.484 \qquad \pm 1.96 \times 25.80 = (789.916, 891.052)
$$