The prediction of price of commodity 9 using Autoregressive Model and Random Walk Model

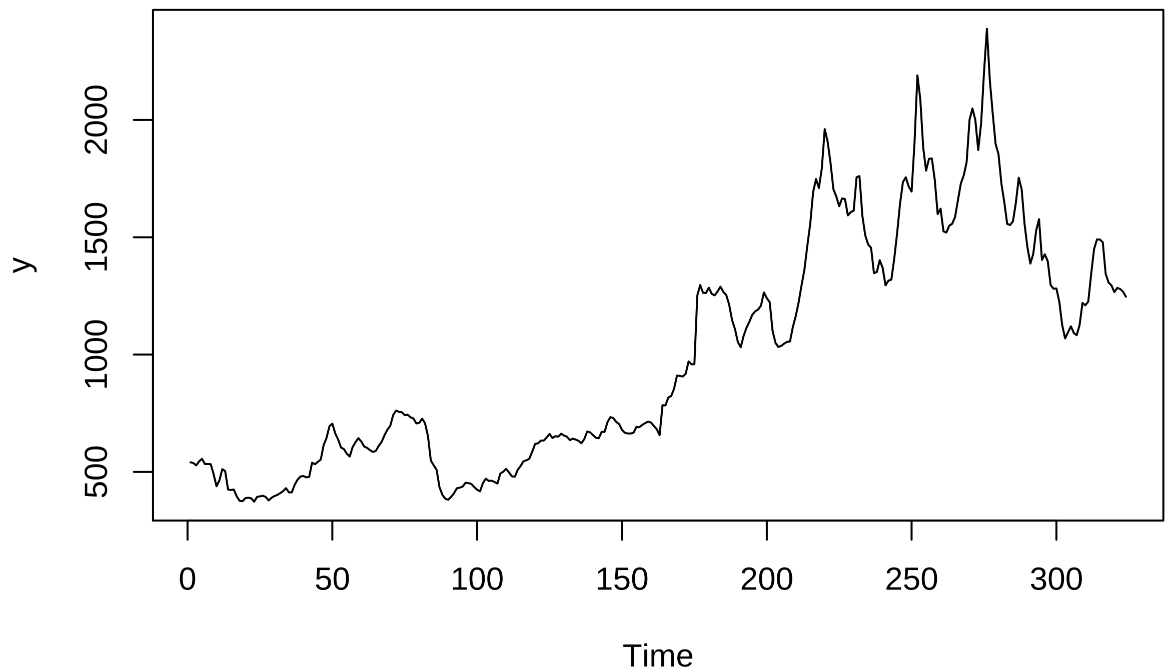
ECMT3130 Individual Report

SID: 470518795

University of Sydney | November 2020

Contents

Data Processing



The data used in this report is the price of commodity 9. The first three rows are `Commodity Name`, `Description of Commodity` and `Data Type`, which can be deleted. Due to first three rows, R recognize the price of commodity 9 to be `character` data type. Thus, it is transformed to `numeric`. This analysis uses log transformed price. The monthly price data begins at 1992-01-01 and ends at 2018-12-01.

Augmented Dickey-Fuller Test

An augmented Dickey-Fuller (ADF) test is test for the null hypothesis that a unit-root is presented in time series data. A unit root process is a nonstationary process.

An ADF test of order p Autoregressive time series data is of the form: $\Delta y_t = \alpha + \gamma y_{t-1} + \cdots + \delta_{p-1} \Delta y_{t-p+1} + \epsilon_t$ where $\epsilon_t \sim IID\left(0, \sigma_\epsilon^2\right)$ The null hypothesis is $H_0 : \gamma = 0$ which means the series is nonstationary. The alternative hypothesis is $H_a : \gamma < 0$ which means the series is stationary. The test statistic is calculated as: $DF = \frac{\hat{\gamma}}{SE(\hat{\gamma})}$

Test statistics and p-value are -2.6903 and 0.2854 respectively. It can be seen that we do not reject null hypothesis that the series is nonstationary. Our series is nonstationary.

Model Selection

The analysis use SIC to choose how many lags in the AR Model. The formula of SIC is $SIC = \ln\left(\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2\right) + \ln T \frac{k}{T}$ where T is the total time period of data observations, k is number of parameters estimated in the model and ε_t is the error of the estimate at time t.

The SICs for models with different lags are calculated. The criteria for model is the less SIC the better, by calculating SIC for Autoregressive (AR) (p) model where p is from 1 to 12, it can be seen that the AR(2) model has the lowest SIC which is -6.131075. Therefore, the following forecasting will use AR(2) model.

Forcasting using AR(2) Model and Random Walk Model

Iterated Approach: one-step and twelve-step forecasting

For 1-step ahead forecast, firstly, the iterated method use the data of log price to estimate the coefficient of α, β and γ of the linear regression model $log(y_t) = \alpha + \beta log(y_{t-1}) + \gamma log(y_{t-2})$ Then, use the data of log price with no lag and log price with lag 1 multiplied with estimated coefficients to estimate the one step ahead forecast.

For 12-step ahead forecast, it is needed to compute all h-step forecast where h is less then 12. The first step is to use the data of log price to estimate the coefficient of α, β and γ of linear regression $log(y_t) = \alpha + \beta log(y_{t-1}) + \gamma log(y_{t-2})$. Then use 11-step ahead forecast $\hat{l}og(y_{t+11})$ and 10-step ahead forecast $\hat{l}og(y_{t+10})$ multiplied with the estimated coefficients to estimate the twelve-step ahead forecast.

Direct Approach: one-step and twelve-step forecasting

The one step ahead forecast in direct method is the same as in iterated method. Estimating the coefficients of $log(y_t) = \alpha + \beta log(y_{t-1}) + \gamma log(y_{t-2})$ model based on data with lag 1 and lag 2. Then use the estimated coefficients combined with data of lag 1 and lag 2 to get the one-step ahead forecasts.

The twelve-step ahead forecast in direct forecast is easier than iterated method since it is not needed to get all h-step ahead forecast where h is less than 12. The first step is to regress $log(y_{t-12})$ and $log(y_{t-13})$ on $log(y_t)$, get the coefficients of the regression. Multiplied with these estimated coefficients with the $log(y_t)$ and $log(y_{t-1})$, the result of twelve-step direct forecasts can be obtained.

Random Walk: one-step and twelve-step ahead forecasting

The twelve-step and one-step ahead forecast of random walk model are the same, both are $log(y_t)$. For one-step ahead of random walk, the $log(y_{t+1}) = log(y_t) + \varepsilon_{t+1}$ thus $log(y_{t+1|t}) = E(log(y_t)) = y_t$. For twelve-step ahead forecast, $log(y_{t+12}) = log(y_{t+11}) + \varepsilon_{t+12}$. $log(y_{t+11|t}) = E(log(y_{t+10})) = log(y_{t+10})$. By doing this iteratively, it can be obtained that $log(y_{t+12|t}) = log(y_t)$.

Combined forecast

Since we forecast using log price, to obtain nominal price forecast, an exponential function is applied to all the forecasts. The combined 12-step ahead forecast of nominal price is using both direct and iterated method with equal weighting scheme. So, the combined forecast is obtained by $1/2 * \hat{y}_{it12} + 1/2 * \hat{y}_{di12}$.

Comparison of RMSFEs of forecasts

Now, we have 1-step ahead random walk forecast, 1-step ahead forecast both iterate and direct, 12-step iterate forecast, 12-step direct forecast, 12-step random walk forecast and 12-step combined forecast. To measure the accuracy, the Root Mean Square Forecast Error (RMSFE) is used. The formula is given by:

$RMSFE = \sqrt{\frac{1}{T-R}\sum_{t=R}^{T-1}\left(\hat{e}_{t+1|t}\right)^2}$ where $\hat{e}_{t+1|t} = \hat{y}_{t+1|t} - y_{t+1}$ is the difference between one step ahead forecast and the actual price value for a given time period. The RMSFE for these six forecasts are below in the table.

From the RMSFEs, one-step ahead forecast using direct or iterated method is better than that using random walk method. However, random walk 12-step ahead forecast is better than direct, iterated and combined 12-step ahead forecast.

Test Equal Forecast Accuracy to Random Walk

Diebold-Mariano test

To compare the forecast accuracy of two different forecasts, Diebold-Mariano test can be used. The null hypothesis of equal predictive ability can be given in terms of unconditional expectation of the loss difference: $H_0 : E\left[\Delta L\left(e_{t+h|t}\right)\right] = 0$ where $\Delta L\left(e_{t+h|t}\right) = L\left(e_{i,t+h|t}\right) - L\left(e_{j,t+h|t}\right)$ That is there is no difference in forecast accuracy between these two forecasts. The alternative hypothesis thus is $E\left[\Delta L\left(e_{t+h|t}\right)\right]! = 0$ there is difference between two forecast accuracy. The test statistics is calculated by $\frac{\bar{d}}{\sqrt{\sigma_{\bar{d}}^2/P}} \sim N(0,1)$ where $\bar{d} = P^{-1}\sum_{t=1}^{P} d_t$ and $d_t \equiv \Delta L\left(e_{t+h|t}\right)$.

In practice, the test of equal accuracy can be applied in a regression model as follows: $\Delta L\left(e_{t+h|t}\right) = \beta + v_{t+h} \quad t = R, \ldots, T-h$ The null hypothesis is $H_0 : \beta = 0$ vs the alternative hypothesis $H_a : \beta! = 0$.

By implementing Diebold-Mariano test, all of the p-value are greater than 0.05 which indicates that our forecasts are not better than random walk forecast, i.e. it is not rejected that the difference between random walk forecast accuracy and our forecast accuracy is zero. All the test statistics are not significant, however, from observing RMSFEs, the one-step ahead forecast direct or iterate are better than random walk. This may be caused by serially correlated difference between two different forecast. Therefore, another test with consistent standard error is used for inference.

Diebold-Mariano test with consistent standard errors

By doing Diebold-Mariano test with consistent standard errors, it can be seen that the p-value for 1-step ahead forecast is less than 0.05, which indicates that the 1-step ahead forecast using direct or iterated method is more accurate than random walk method. However, for other forecast, 12-step ahead direct, iterated and combined are not better than random walk 12-step ahead forecast, this is consistent with our RMSFEs' results.

Forecast	RMSFEs	RMSFE ratios (RW:AR(2))	DM Test p-value	DM Test (with consistent SE) p-value
1-step ahead AR (2)	91.09723	1.069473	0.248	0.0112*
12-step ahead (direct)	374.0468	0.9917476	0.515	0.07897
12-step ahead (iterated)	394.3699	0.9406397	0.55	0.5836
12-step ahead (combined)	383.3629	0.9676471	0.112	0.1426
1-step ahead RW	97.42602	-	-	-
12-step ahead RW	370.96	-	-	-