

Time Series Analysis

Advanced Assignment

1. When modeling $\ln Y_t$ using a time trend model, what is the relationship between $\exp E_T[\ln Y_{T+h}]$ and $E_T[Y_{T+h}]$ for any forecasting period h ? Are these ever the same? Assume that the error terms are normally distributed around a mean of zero.

2. Why does a unit root with a time trend, $Y_t = d_1 + Y_{t-1} + \epsilon_t$ not depend explicitly on t ?

3. The Yule-Walker (YW) equations provide a set of expression that relate the parameters of an AR to the autocovariances of the AR process. This approach uses $p + 1$ equations to solve for the long-run variance g_0 and the first p autocorrelations. Autocovariances (or autocorrelations) at lags larger than p are then easily computed with a recursive structure starting from the first p autocovariances. The equations are:

$$\text{Cov}[Y_t, Y_t] = \text{Cov}[d + f_1 Y_{t-1} + g + f_p Y_{t-p} + P_t, Y_t] \quad \text{Cov}[Y_t,$$

$$Y_{t-1}] = \text{Cov}[d + f_1 Y_{t-1} + g + f_p Y_{t-p} + P_t, Y_{t-1}]$$

.

..

$$\text{Cov}[Y_t, Y_{t-p}] = \text{Cov}[d + f_1 Y_{t-1} + g + f_p Y_{t-p} + P_t, Y_{t-p}]$$

Excluding the first equation, dividing each row by the long run variance g_0 produces a set of equations that relate the autocorrelations:

$$\begin{aligned} \rho_1 &= \phi_1 + \phi_2 \rho_1 + \dots + \phi_p \rho_{p-1} \\ \rho_2 &= \phi_1 \rho_1 + \phi_2 \rho_2 + \dots + \phi_p \rho_{p-2} \end{aligned}$$

Compute the first three autocorrelations for the AR(2) process as shown below:

$$Y_t = 1.4Y_{t-1} - 0.45Y_{t-2}$$

4. (Python programming assignment)

a. Download the "S&P Dividend Yield by Month" and the "S&P Dividend Yield by Year". The data can be accessed using the Quandl/NASDAQ data link. (Sample [here](#)).

b. Download the series MULTPL/SP500_DIV_YIELD_MONTH and MULTPL/SP500_DIV_YIELD_YEAR.

- Plot and compare the autocorrelation function (ACF) and partial autocorrelation function (PACF) for the monthly and yearly series. Take the log of each series and plot the ACF and PACF of the log series. How are the ACF/PACFs different for the log series and the raw series (without log).

- Perform the [Box-Ljung](#) test for the first 5 autocorrelation for each of the 4 series from part a (annual, monthly) * (log, without log). Report the test statistics and p values. What can you conclude based on these observations?

- Perform the [ADF test](#) for each of the 4 series from part a (annual, monthly) * (log, without log). Report the test statistics and p values. What can you conclude based on results of these tests?

ANSWERS

Q1. Modeling $\ln Y_{T+h}$ as a time trend model:

$$\ln Y_{T+h} = \alpha + \beta^*(T+h) + \varepsilon_{T+h}, \text{ where } \varepsilon_{T+h} \sim N(0, \sigma^2) \text{ (mean given to be 0)}$$

$$\text{Let } X \equiv \ln Y_{T+h}$$

Then taking expectation of X we obtain $ET[\ln Y_{T+h}] = ET[X] = \alpha + \beta^*(T+h) + 0$

$$\Rightarrow X \sim N(\mu, \sigma^2) \text{ with } \mu = \alpha + \beta^*(T+h).$$

Thus the term $\exp ET[\ln Y_{T+h}]$ is $\exp \mu$.

$$\text{Now, } ET[Y_{T+h}] = ET[\exp X] = \int_{-\infty}^{\infty} e^x f_X(x) dx = \int_{-\infty}^{\infty} e^x \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx$$

Inside the integrand we have two exponentials, so combining them $\Rightarrow e^x \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) = \exp\left(x - \frac{(x-\mu)^2}{2\sigma^2}\right)$

Now completing the square in exponent:

We need to rewrite $x - \frac{(x-\mu)^2}{2\sigma^2}$ as a perfect square plus constant:

$$x - \frac{(x-\mu)^2}{2\sigma^2} = \frac{2\sigma^2 x - (x-\mu)^2}{2\sigma^2} = \frac{2\sigma^2 x - x^2 - \mu^2 + 2x\mu}{2\sigma^2}$$

$$\text{Now, } -x^2 - \mu^2 + 2x\mu + 2\sigma^2 x = -[x^2 - 2(\mu + \sigma^2)x + (\mu + \sigma^2)^2] + (\mu + \sigma^2)^2 - \mu^2$$

Computing the term out of square: $(\mu + \sigma^2)^2 - \mu^2 = \sigma^4 + 2\mu\sigma^2$

$$x - \frac{(x-\mu)^2}{2\sigma^2} = -\frac{(x-(\mu+\sigma^2))^2}{2\sigma^2} + \frac{\sigma^4 + 2\mu\sigma^2}{2\sigma^2} = -\frac{(x-(\mu+\sigma^2))^2}{2\sigma^2} + \left(\frac{\sigma^2}{2} + \mu\right)$$

$$\text{Now, } ET[\exp X] = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(x - \frac{(x-\mu)^2}{2\sigma^2}\right) dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-(\mu+\sigma^2))^2}{2\sigma^2} + \left(\frac{\sigma^2}{2} + \mu\right)\right) dx$$

The remaining integral is exactly the integral of a normal density with mean $\mu + \sigma^2$ and variance σ^2 :

$$\Rightarrow ET[\exp X] = \exp\left(\mu + \frac{\sigma^2}{2}\right) * \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \exp\left(-\frac{(x-(\mu+\sigma^2))^2}{2\sigma^2}\right) dx, \text{ as } \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \exp\left(-\frac{(x-(\mu+\sigma^2))^2}{2\sigma^2}\right) dx = 1$$

$$\Rightarrow \mathbf{ET[\exp X] = \exp\left(\mu + \frac{\sigma^2}{2}\right)}.$$

On comparing $ET[\exp X]$ and $\exp ET[\ln Y_{T+h}]$, **It becomes a comparison between $\exp\left(\mu + \frac{\sigma^2}{2}\right)$ and $\exp \mu$**

And essentially,

if $\sigma^2 > 0$ then $\exp \mu < \exp\left(\mu + \frac{\sigma^2}{2}\right)$ as exponent function is an increasing function

if $\sigma^2 = 0$, then $\exp \mu = \exp\left(\mu + \frac{\sigma^2}{2}\right)$

Hence, if $\sigma^2 > 0$, $\exp E\mathcal{T}[\ln Y_{T+h}] < E\mathcal{T}[\exp \ln Y_{T+h}]$
 if $\sigma^2 = 0$, $\exp E\mathcal{T}[\ln Y_{T+h}] = E\mathcal{T}[\exp \ln Y_{T+h}]$

Q2.

An AR(1) process $Y_t = \phi Y_{t-1} + \varepsilon_t$ has a unit root when $\phi = 1$,

In lag operator form that is $(1 - L)Y_t = \varepsilon_t$ and the polynomial $1 - z = 0$ has root $z = 1$

A unit root makes the series non stationary as expectation of above AR(1) process is constant, but variance has t in it and is not constant (because of the shocks ε_t).

The expression $Y_t = \phi Y_{t-1} + \varepsilon_t + d_1$ is just adding a constant drift d_1 to each period

$$Y_t = Y_{t-1} + \varepsilon_t + d_1 \Leftrightarrow (1 - L)Y_t = \varepsilon_t + d_1$$

There is no explicit ' βt ' term so this does not depend on t explicitly,

Now, to check any other dependence implicitly we iterate the one step equation from some initial Y_0 :

$$\begin{aligned} Y_1 &= Y_0 + \varepsilon_1 + d_1, \\ Y_2 &= Y_1 + \varepsilon_2 + d_1 = Y_0 + \varepsilon_1 + \varepsilon_2 + 2d_1, \\ &\dots\dots\dots \\ Y_t &= Y_0 + td_1 + \sum_{i=1}^t \varepsilon_i \end{aligned}$$

The term td_1 is deterministic trend in the mean: $E[Y_t] = Y_0 + td_1$,

The shocks or error terms which are white noise result in the term $\sum_{i=1}^t \varepsilon_i$, because of which

the variance of Y_t becomes $0 + t\sigma^2$ which is $t\sigma^2$, $\text{Var}(Y_t) = t\sigma^2$ which also depends on t.

Solving the equation shows a dependence on t for the recursed version considering an initial point along with linear dependence on t for both mean and variance.

- The model's form is recursive, not a direct regression on t.
- The **dependence on t** appears only when recursion is unwind. We add d_1 each of the t steps,

giving td_1

- The unit root ($\phi=1$) is what makes the effect of each drift and shock permanent— keep carrying it forward—so both the mean and variance grow with t .

Thus, In conclusion:

even without an explicit t in the initial time trend $Y_t = \phi Y_{t-1} + \varepsilon_t$, which is unlike the regression style trend model with explicit t dependence shown as $Y_t = \alpha + \beta t + u_t$, in the unit-root-with-drift formulation, the dependence on t is built into the recursive accumulation of the constant drift d_1 .

Adding d_1 every period naturally produces a $t^* d_1$ term once the recursion is solved.

Q3.

For an AR(2) process: $Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \varepsilon_t$ where ε_t is white noise

The Yule-Walker equations for autocorrelations $\rho_k = \text{Corr}(Y_t, Y_{t-k})$ are:

$$\rho_1 = \phi_1 + \phi_2 \rho_1$$

$$\rho_2 = \phi_1 \rho_1 + \phi_2$$

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} \quad \text{for } k \geq 3$$

Given AR(2) process: $Y_t = 1.4Y_{t-1} - 0.45Y_{t-2} + \varepsilon_t$

So for our process $\phi_1 = 1.4$, $\phi_2 = -0.45$

Solving for ρ_1 and ρ_2

Equation 1:

$$\rho_1 = \phi_1 \rho_0 + \phi_2 \rho_1$$

But $\rho_0 = 1$, so:

$$\rho_1 = 1.4 \times 1 + (-0.45) \rho_1$$

$$\rho_1 + (0.45) \rho_1 = 1.4$$

$$1.45 \rho_1 = 1.4$$

$$\rho_1 = \frac{1.4}{1.45} \approx 0.9655$$

Equation 2:

$$\rho_2 = \phi_1 \rho_1 + \phi_2 \rho_0$$

$$\rho_2 = 1.4 \times 0.9655 + (-0.45) \times 1$$

$$\rho_2 = 1.3517 - 0.45 = 0.9017$$

Solving for ρ_3 :

$$\rho_3 = \phi_1 \rho_2 + \phi_2 \rho_1$$

$$\rho_3 = 1.4 \times 0.9017 + (-0.45) \times 0.9655$$

$$\rho_3 = 1.2624 - 0.4345 = 0.8279$$

Hence the first 3 autocorrelations of the AR(2) process are : **0.9655, 0.9017, 0.8279**

Q4-

Here is the notebook where we have executed all the commands (all interpretations are also written as comments in this notebook)-

 time_series.ipynb

1. Plot Comparison: ACF & PACF for Monthly Dividend Yield-

```
# Plot Comparison: ACF & PACF for Monthly Dividend Yield

#1. ACF (Autocorrelation Function)
# Raw & Log Series:

# Slowly decaying ACF with very high autocorrelation at lag 1 and gradually
decreasing values beyond that.

# This is a classic indicator of non-stationarity – specifically a unit root
or persistent trend.

# The similarity between raw and log ACF curves confirms that log
transformation had little impact.

# 2. PACF (Partial Autocorrelation Function)
# Raw & Log Series:
```

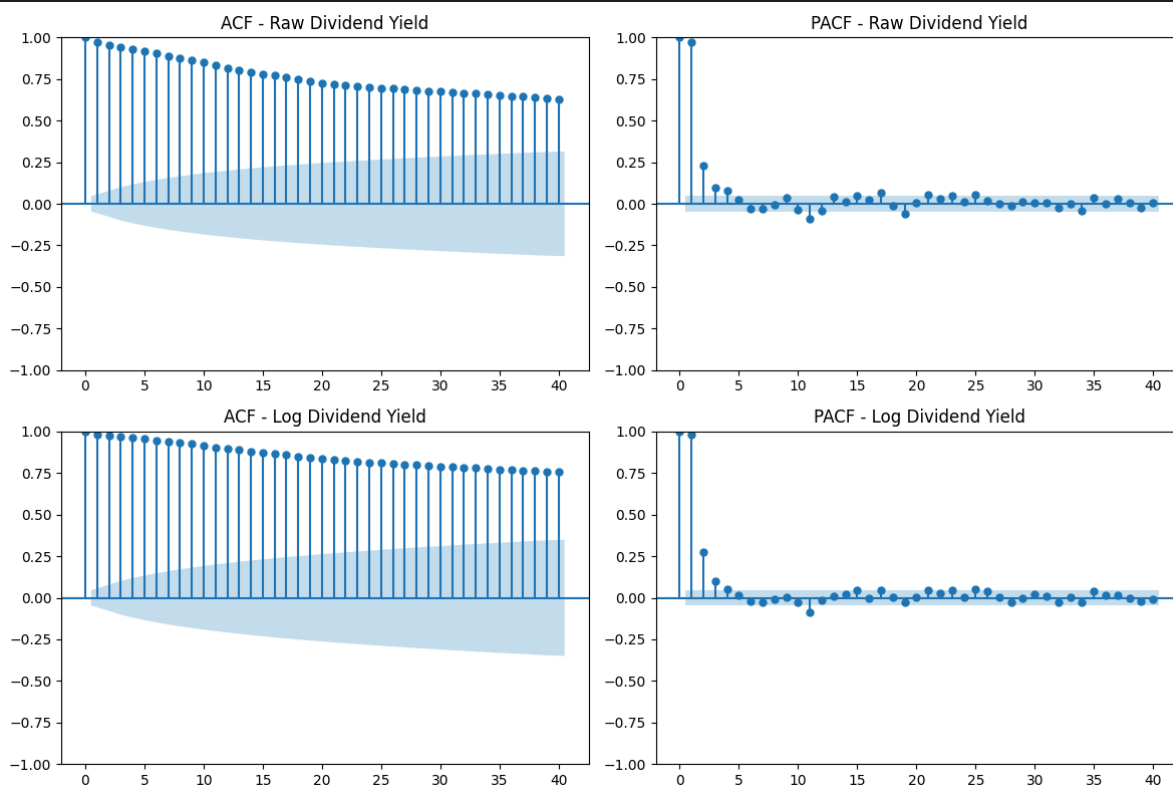
```
# PACF shows a strong spike at lag 1, and smaller spikes at lag 2 (and
possibly 3).

# After that, PACF values drop closer to zero.

# This structure suggests:

# Possible AR(1) or AR(2) behavior,

# Once differenced, the series may become stationary and suitable for ARIMA
modeling.
```



2. Why log and raw series look similar-

```
# Why Log and Raw Series Look Similar
# Log transformation has minimal impact on values that are already small and
positive (like the Dividend_Yield values, which are around 0.05-0.06).

# Since the range of values is narrow, log(x) behaves nearly linearly – so
the correlation structure (ACF/PACF) is preserved.
```

```
# Hence, both the raw and log series are non-stationary and exhibit strong autocorrelation (slowly decaying ACF, significant PACF spikes).

# Interpretation of Plots
# ACF:
# Both raw and log series show strong persistence (high values) – the ACF tails off slowly, indicating non-stationarity (likely a trend or unit root).

# PACF:
# Both show significant spikes at lags 1-2, suggesting short-term dependency that might be captured by an AR(1) or AR(2) process after differencing.
```

Results of the hypothesis test-

Box-Ljung Test (lag=5):

Dividend Yield: Statistic = 8133.1721, p-value = 0.0000

Log Dividend Yield: Statistic = 8592.5017, p-value = 0.0000

Augmented Dickey-Fuller (ADF) Test:

Dividend Yield: ADF Statistic = -2.4718, p-value = 0.1225

Log Dividend Yield: ADF Statistic = -1.3980, p-value = 0.5832

Interpretation:

The null hypothesis of the Box-Ljung test is that the data are independently distributed (i.e., no autocorrelation).

Since p-values < 0.05, we reject the null hypothesis for both raw and log series.

Conclusion: Significant autocorrelation exists in both series → the data are not white noise.

Interpretation:

The null hypothesis of the ADF test is that the series has a unit root (i.e., non-stationary).

```
# Since p-values > 0.05, we fail to reject the null hypothesis for both series.
```

```
# Conclusion: Both raw and log Dividend Yield series are non-stationary.
```

```
# Final Conclusion-
```

```
# 1. Autocorrelation is present (Box-Ljung Test).
```

```
# 2.Both series are non-stationary (ADF Test).
```

```
# 3.Log transformation does not induce stationarity, as ACF/PACF and ADF test show similar results.
```