Full credit given to very complete and detailed answers
Full/ near-full credits as long as answers align with some correct subsets of the below content
Full credit given to correct statement without supporting equations
Partial credit if there is obvious mistakes/ wrong statement

Question 1 (10 points)
Define (weak) stationarity and explain why it is important in time series analysis. Explain what is a unit root test and how to set up the null and alternative hypothesis for a unit root test.

Weak stationary is defined by the characteristic statistics, i.e. low order moments, of the stochastic process

The idea is that as long as the low order moments are stationary, the stochastic process can be considered as stationary

A stochastic process $\{X_t\}$ is weak stationary if:

$$(1) \quad EX_t = \mu = constant$$
$$(2) \quad Var(X_t) = constant, (EX_t^2 < \infty)$$
$$(3) \quad Cov(X_t, X_{t+\tau}) \text{ depends only on } \tau \text{ and not on } t$$

It is also called covariance stationary or second-order stationary

Technically, stationarity simplifies statements of the law of large numbers and the central limit theorem, and by appealing to LLN and CLT OLS is argued to be generally justified

On practical level, we need to assume some sort of stability over time. e.g., If we allow the variable x_t to change arbitrarily in each time period, then we cannot hope to learn much about its pattern, since we only observe a single time series realization

It is useful for multivariate regression model for time series data, where we are assuming a certain form of stationarity in β - the coefficient doesn't change over time

We can also use a statistical test to check if a time series process is stationary

- The most popular test is unit root test

- Unit root test: to test whether the characteristic roots are inside or outside the unit circle

Dickey-Fuller tests

Let's use AR(1) model as an example

$$x_t = \phi_1 x_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim N(0, \sigma_\varepsilon^2)$$

- The characteristic equation of the process is : $\lambda - \phi_1 = 0$
- The characteristic root is : $\lambda = \phi_1$

If the characteristic root is within the unit circle, the process is stationary

$$|\phi_1| < 1$$

If the characteristic root is outside or on the unit circle, the process is non-stationary

$$|\phi_1| \geq 1$$

We can therefore conduct tests to check the stationarity of the time series process (by testing on the unit root)

- H_0: the time series process x_t is non-stationary

$$H_0 : |\phi_1| \geq 1$$

- H_1: the time series process x_t is stationary

$$H_1 : |\phi_1| < 1$$

Augmented Dickey-Fuller tests

Let's consider a more general AR(p) model now as

$$\Delta x_t = \mu + \beta t + (\rho - 1)x_{t-1} + \sum_{j=1}^{p-1} \beta_j \Delta x_{t-j} + \varepsilon_t$$

where $\Delta x_{t-j} = x_j - x_{j-1}$ is the differenced series of x_t

- ADF-test $= \dfrac{\hat{\rho} - 1}{SD(\hat{\rho})}$

  $H_0 : \rho = 1$

  $H_1 : \rho < 1$

- The critical value can again be obtained through simulation. DF-test is a special case of ADF-test when p-order=1

Question 2 (10 points)
Discuss the difference between trend-stationary and difference-stationary time series. How can differencing be used to achieve stationarity in a time series? Explain what it means for a time series to be integrated of certain order and discuss how we handle integrated series.

Random walk model

$$x_t = x_{t-1} + \varepsilon_t$$

where x_0 is a real number denoting the starting value of the process and $\{\varepsilon_t\}$ is a white noise series

- The current outcome = the last period's outcome plus white noise

- A special AR(1) process with $\phi_1 = 1$, therefore, not weakly stationary $\rightarrow$ unit-root nonstationary time series

- Time-varying and increasing variance function of time

$$E(x_t) = x_0$$
$$Var(x_t) = t\sigma_\varepsilon^2$$

Take the first difference:

$$\Delta x_t = x_t - x_{t-1} = \varepsilon_t$$

- $E(\Delta x_t) = 0$ and $Var(\Delta x_t) = \sigma_\varepsilon^2$
- The first difference of x_t is a stationary process, with constant mean and constant variance
- Therefore a random walk process (without drift) is also called a difference stationary process

Trend-stationary model

$$x_t = \beta_0 + \beta_1 t + \varepsilon_t$$

- The current outcome grows linearly in time with rate β_1 + stationary process
- $E(x_t) = \beta_0 + \beta_1 t$ and $Var(x_t) = Var(\varepsilon_t)$
- The mean depends on time, and variance is finite and time-invariant
- x_t exhibits behavior similar to that of a random walk with drift

The best way to make the residual series stationary for trend-stationary model is to fit a linear regression, instead of performing differencing. Both ways can achieve trend-stationary, however the latter increases the variance of the residual series

If a non-stationary series x_t must be differenced d times before it becomes stationary, it is integrated of order d: $x_t \sim I(d)$

Therefore if $x_t \sim I(d)$, then $\Delta^d x_t \sim I(0)$

- An I(0) series is stationary
- An I(1) series can contain one unit root
- An I(2) can contain two unit roots and must therefore be differenced twice to become stationary

We can use ARIMA model to estimate difference-stationary time series process

ARIMA model is the combination of ARMA model and differencing operation

If a nonstationary series can be transformed to a stationary series through differencing of d times, then the original series can be estimated by ARIMA(p,d,q)

ARIMA model has non-constant mean and non-finite variance for $d \neq 0$

We can use ARIMA model to conduct forecasting the same way as we use ARMA model

Question 3 (10 points)

Briefly describe the purpose and interpretation of forecast error metrics. List and explain at least two measures of the forecast error metrics we discussed in the lecture.

The forecasts will not be perfectly accurate and the forecast error j periods ahead is defined as

$$e_t(j) \equiv x_{t+j} - E_t(x_{t+j})$$

We use different ways to summarize forecasting error, in order to evaluate the accuracy of the forecast

- Easy to compute and understand: mean absolute error (MAE)
- Widely used: root mean squared error (RMSE)
- Not friendly with 0, mean absolute percentage error (MAPE) = mean($|100 e_j / x_t|$)
- Independent of the scale of the data: mean absolute scaled error (MASE) = $e_j / \frac{1}{T-1} \sum_{t=2}^{T} |x_t - x_{t-1}|$, scaling the errors based on the training MAE

Note there is a typo in the instruction, where says no need to derive in Question 4 and 5 (which actually was meant for Q5 an Q6, multiple choice questions).

Full/near-full credit can be given to correct answers without detailed derivation (in light of the instruction mistake)

Partial credit given to wrong answers

Question 4 (50 points)

(1) Find the mean and the variance of the above AR(1) process.

For a stationary AR(1) processes, the mean:

$$Ex_t = \frac{\phi_0}{1-\phi_1}$$

$E(x_t) = 0$, for demeaned processes

We use the MA representation of AR(1):

$$x_t(1-\phi L) = \varepsilon_t$$

$$x_t = (1-\phi L)^{-1}\varepsilon_t = \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}$$

So

$$\gamma_0 = \left(\sum_{j=0}^{\infty} \phi^{2j}\right)\sigma_\varepsilon^2 = \frac{\sigma_\varepsilon^2}{1-\phi^2} \qquad \rho_0 = 1$$

(2) Derive the autocorrelation function (ACF) and the partial autocorrelation function (PACF) of the above AR(1) process.

$$\gamma_1 = E(x_t x_{t-1}) = E\Big((\phi x_{t-1} + \varepsilon_t)(x_{t-1})\Big) = \phi\sigma_x^2 \qquad \rho_1 = \phi$$

$$\gamma_2 = E(x_t x_{t-2}) = E\Big((\phi^2 x_{t-2} + \phi x_{t-1} + \varepsilon_t)(x_{t-2})\Big) = \phi^2\sigma_x^2 \qquad \rho_2 = \phi^2$$

$$\cdots$$

$$\gamma_k = E(x_t x_{t-k}) = E\Big((\phi^k x_{t-k} + \varepsilon + \cdots)(x_{t-k})\Big) = \phi^k\sigma_x^2 \qquad \rho_k = \phi^k$$

The PAC ϕ_{kk} between x_t and x_{t-k} is the correlation coefficient between x_t and x_{t-k} conditional on, or controlling for, the effects of $x_{t-1}, \cdots, x_{t-k-1}$

$$\phi_{11} = \rho_1$$

The PAC ϕ_{ii} can be found as the coefficient of the final lag in the autoregressions:

$$x_t = \phi_0 + \phi_{11} x_{t-1} + e_{1,t}$$
$$x_t = \phi_0 + \phi_1 x_{t-1} + \phi_{22} x_{t-2} + e_{2,t}$$
$$\cdots$$
$$x_t = \phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_{kk} x_{t-k} + e_{k,t}$$

The ϕ_{ii} measures the correlation between x_t and x_{t-i} after removing the effects of $x_{t-i+1}, x_{t-i+2}, \cdots, x_{t-1}$

- The partial autocorrelation function (PACF) is the ϕ_{ii} plotted against period $i = 1, \cdots, k, k+1, \cdots$
- ACF = PACF at first lag

If the true model is AR(p), the theoretical PACF will be zero after lag p

(3) Suppose now you don't know whether the above AR(1) model is stationary and please find out the necessary and sufficient conditions for this model to be stationary.

$$x_t = \phi_1 x_{t-1} + \varepsilon_t$$

The characteristic equation is $\lambda - \phi_1 = 0$

The characteristic root is $\lambda = \phi_1$

AR(1) is stationary if and only if

$$|\phi_1| < 1$$

The stationary region of $AR(1)$ model is $\{-1 < \phi_1 < 1\}$

(4) Now you are using this AR(1) model to make prediction. Derive the one-step forecast, Et(xt+1). Show that as the forecast horizon goes to infinity, the limit of the conditional mean converges to its unconditional mean.

For AR(1) process:

$$x_t = \phi x_{t-1} + \varepsilon_t$$

$$E_t(x_{t+1}) = \phi x_t$$

$$
\begin{aligned}
E_t(x_{t+1}) &= \phi x_t \\
E_t(x_{t+2}) &= \phi^2 x_t \\
E_t(x_{t+3}) &= \phi^3 x_t \\
&\vdots \\
E_t(x_{t+k}) &= \phi^k x_t
\end{aligned}
$$

As k goes to infinity, the forecast converges to the unconditional mean zero

(5) The process exhibits the feature of mean reversion, and in finance it is often useful to express it as the half-life. Explain what is half-life and derive the half-life of AR(1).

For stationary AR(p) model, the conditional point forecast is shown to converge to its unconditional mean

This property is referred to as the *mean reversion* in the finance literature

For an AR(1) model, the speed of mean reversion is measured by the half-life

$$l = ln(0.5)/ln(|\phi_1|)$$

Partial credit given to partially correct choices in Q5

Question 5: ABCD

Question 6: D