Problem A: (30 pts) Answer briefly the following questions. Each question has two points.

1. Give two reasons that the observed daily asset returns have significant serial correlations, even if the underlying true returns are serially uncorrelated.
   Answer: Any two of (a) nonsynchronous trading, (b) bid-ask bounce, and (c) risk premium.

2. Describe two methods to model seasonality in a financial time series.
   Answer: (a) Use a seasonal time series model and (b) use seasonal dummy variables.

3. For Questions 3 to 6, consider the daily closing values of the S&P 500 index from January 2, 2003 to April 19, 2010. Let sp5 denote the series of the logarithms of closing values. Is there a unit root in the sp5 series? Why?
   Answer: Yes, because the unit-root test ADF has a high p-value so that one cannot reject the unit-root hypothesis.

4. Consider the daily log returns of the S&P 500 index. Based on the observed data, is the mean return significantly different from zero? Why?
   Answer: No, the 95% confidence interval for the mean is $[-0.000469, 0.000769]$ that contains zero.

5. Is the distribution of the daily log returns symmetric with respect to zero? Why?
   Answer: The t-ratio for skewness is $t = -0.269/\sqrt{6/1835} = -4.704$, which is much larger than 1.96. Thus, the distribution is not symmetric with respect to zero.

6. Does the distribution of the daily log returns have heavy tails? Why?
   Answer: Yes, the t-ratio for excess kurtosis is $t = 11.462/\sqrt{24/1835} = 100.22$, which is highly significant.

7. For Questions 7 to 9, consider the simple AR(2) model
   \[ r_t = 0.01 + 1.3r_{t-1} - 0.4r_{t-2} + a_t, \quad \text{Var}(a_t) = 2.0. \]
   What is the mean of $r_t$? That is, find $E(r_t)$.
   Answer: $E(r_t) = \frac{0.01}{1-1.3+.4} = 0.1$.

8. Does the model imply the existence of business cycles? Why?
   Answer: No, the solutions of $(1 - 1.3x + .4x^2) = 0$ are 1.25 and 2.0. They are real numbers.
9. Suppose that \( r_{100} = 1.2 \) and \( r_{99} = 0.5 \). What is the 1-step ahead prediction of \( r_{101} \) at the forecast origin \( T = 100 \)? What is the 95% confidence interval of the prediction.

**Answer:** \( r_{100}(1) = 0.01 + 1.3(1.2) - 0.4(0.5) = 1.37 \). The variance of 1-step ahead prediction error is \( \text{Var}(a_{101}) = 2 \). The 95% confidence interval is \( 1.37 \pm 1.96 \times \sqrt{2} \), i.e. \([-1.402, 4.142]\).

10. Describe two characteristics of the EGARCH models that are not available in the GARCH models.

**Answer:** (a) Allow for asymmetric response to past positive and negative returns (or leverage effect), and (b) use log variance to relax positiveness constraints.

11. For Questions 11 and 13, consider the following ARMA(0,0)-GARCH(1,1) model

\[
\begin{align*}
    r_t &= 0.03 + a_t, \quad a_t = \sigma_t \epsilon_t, \quad \epsilon_t \sim N(0, 1) \\
    \sigma_t^2 &= 0.144 + 0.07a_{t-1}^2 + 0.83\sigma_{t-1}^2.
\end{align*}
\]

What is the unconditional variance of \( a_t \)?

**Answer:** \( \text{Var}(a_t) = E(a_t^2) = \frac{0.144}{1-0.07-0.83} = 1.44. \)

12. Suppose that \( \sigma^2_{100} = 0.6 \) and \( a_{100} = -0.1 \). Obtain the 95% confidence interval for the 1-step ahead prediction of \( r_{101} \) at the forecast origin \( T = 100 \).

**Answer:** First, \( E(r_{101}|F_{100}) = 0.03. \) Second, \( \sigma^2_{100}(1) = 0.144+0.07(-0.1)^2+0.83(0.6) = 0.6427. \) Finally, the 95% confidence interval is \( 0.03 \pm 1.96 \times \sqrt{0.6427}, \) i.e. \([-1.541, 1.601]\).

13. Obtain the 95% confidence interval for the 100-step ahead prediction of \( r_{200} \) at the forecast origin \( T = 100 \).

**Answer:** This is a long-term prediction, the mean is 0.03 and the variance is \( \text{Var}(a_t) = 1.44. \) The 95% confidence interval is \( 0.03 \pm 1.96\sqrt{1.44}, \) i.e., \([-2.322, 2.382]\).

14. Describe two methods to compare different forecasting models.

**Answer:** (a) Use backtest method, and (b) use information criteria.

15. Describe briefly the difference between trend- and difference-stationarity in the long-term prediction.

**Answer:** (a) long-term prediction differ substantially, (b) the confidence intervals of long-term prediction also differ markedly; the length of the interval is fixed for a trend-stationary series, is increasing to infinity for difference-stationary series.

**Problem B.** (37 pts) Consider the monthly log returns, in percentages, of the Coca-Cola stock from January 1960 to December 2009. The relevant R output is attached. Answer the following questions.
1. (2 points) Is the mean of the log return series equal to zero? Why?

**Answer:** No, the t-test has a small p-value so that the null hypothesis of zero mean is rejected.

2. (2 points) Is there any serial correlation in the monthly log returns? Why?

**Answer:** No, there are no serial correlations in the returns because the Ljung-Box test cannot reject $H_0: \rho_1 = \cdots = \rho_{24} = 0$.

3. (2 points) Is there any ARCH effect in the monthly log returns? Why?

**Answer:** Yes, because the Ljung-Box statistic of the squared series $a_t^2$ is highly significant with p-value close to zero.

4. (3 points) A GARCH(1,1) model with Gaussian distribution is used for the volatility equation. Write down the fitted model, including the mean equation. Is the model adequate? Why?

**Answer:** Let $r_t$ be the monthly log returns, in percentages. The fitted model is

\[
\begin{align*}
  r_t &= 1.293 + a_t, \\
  a_t &= \sigma_t \epsilon_t \\
  \sigma_t^2 &= 2.653 + 0.094a_{t-1}^2 + 0.835\sigma_{t-1}^2,
\end{align*}
\]

where $\epsilon_t \sim N(0, 1)$.

The model is not adequate because the normality assumption is clearly rejected.

5. (2 points) Except for the normality, is the fitted GARCH(1,1) model adequate? Why?

**Answer:** Yes, Ljung-Box statistics for the standardized residuals give $Q(10) = 9.62(0.47)$ and $Q(20) = 21.32(0.37)$, where the number in parentheses is the p-value, and those for the squared standardized residuals show $Q(10) = 11.75(0.30)$ and $Q(20) = 13.91(0.83)$. Thus, there are no serial correlations in the standardized residuals and their squared series.

6. (3 points) A GARCH(1,1) model with Student-t innovations is applied. Write down the volatility equation, including the degrees of freedom.

**Answer:** $\sigma_t^2 = 2.306 + 0.100a_{t-1}^2 + 0.841\sigma_{t-1}^2$, where $\sigma_t$ is defined as $a_t = \sigma_t \epsilon_t$ with $\epsilon_t$ being the standardized Student-t with 6.92 degrees of freedom.

7. (3 points) A GARCH(1,1) model with skew Student-t innovations is used. Write down the volatility equation, including the distribution parameters. Is the model adequate? Why?

**Answer:** $\sigma_t^2 = 2.322 + 0.098a_{t-1}^2 + 0.841\sigma_{t-1}^2$, where $\sigma_t$ is defined as $a_t = \sigma_t \epsilon_t$ with $\epsilon_t$ being a skew standardized Student-t distribution with 7.16 degrees of freedom and skew parameter 0.962.

The model checking statistics provided in the output fail to reject that the model is adequate.
8. (2 points) Let $\xi$ be the skew parameter. To check whether the distribution of the innovations is skewed, consider the test $H_0 : \xi = 1$ versus $H_a : \xi \neq 1$. Perform the test and draw a conclusion.

**Answer:** The test statistic is $t$-ratio $= (0.962 - 1)/0.057 = -0.667$, which is smaller than 1.96 in modulus. Therefore, the null hypothesis of $\xi = 1$ cannot be rejected at the 5% significance level.

9. (3 points) Focus on an IGARCH(1,1) model, write down the fitted model.

**Answer:** The fitted IGARCH(1,1) model is

$$
\begin{align*}
    r_t &= 1.318 + a_t, \quad a_t = \sigma_t \epsilon_t, \quad \epsilon_t \sim t_{5.35} \\
    \sigma_t^2 &= 1.153 + 0.141 a_{t-1}^2 + 0.859 \sigma_{t-1}^2.
\end{align*}
$$

10. (3 points) A GJR model is also fitted. Write down the fitted volatility equation.

**Answer:** $\sigma_t^2 = 2.628 + (0.0726 + 0.048 N_{t-1}) a_{t-1}^2 + 0.833 \sigma_{t-1}^2$, where $N_{t-1} = 0$ if $a_{t-1} \geq 0$ and $= 1$ if $a_{t-1} < 0$, and $\sigma_t$ is defined as $a_t = \sigma_t \epsilon_t$ with $\epsilon_t$ being a Student-$t$ distribution with 7.055 degrees of freedom.

11. (2 points) Based on the GJR model, is the leverage effect significant? Why?

**Answer:** The $t$-ratio is 0.896 with $p$-value 0.37. Therefore, one cannot reject the null hypothesis of no leverage effect.

12. (2 points) Among the volatility models entertained so far, which model is preferred? State the criterion used in your choice.

**Answer:** Based on the AIC, the preferred model is the GARCH(1,1) model with Student-$t$ innovations. The model gives the minimum AIC of 6.361.

13. (4 points) An EGARCH(1,1) model is also fitted, but to the log returns instead of the percentage log returns. Write down the fitted volatility equation. For simplicity, you may ignore the ARCH parameter, which is insignificant.

**Answer:** Let $x_t$ be the log return series (not in percentages). The fitted EGARCH(1,1) model is

$$
\begin{align*}
    x_t &= 0.0113 + a_t, \quad a_t = \sigma_t \epsilon_t, \quad \epsilon_t \sim N(0, 1), \\
    \ln(\sigma_t^2) &= -5.614 + \frac{1}{1 - 0.829 B} g(\epsilon_{t-1}), \\
    g(\epsilon_t) &= -0.050 \epsilon_t + 0.156 (|\epsilon_t| - 0.8).
\end{align*}
$$

14. (2 points) Based on the fitted EGARCH(1,1) model, is the leverage effect significant? Why?

**Answer:** The leverage parameter is $\theta_1$, which is not significantly different from zero, because its $t$-ratio is $-1.378$ with $p$-value 0.169.
15. (2 points) Is the EGARCH model adequate in modeling the serial dependence in $r_t$ and $a_t^2$? Why?

**Answer:** Yes, the Ljung-Box statistics of the residuals and the squared residuals fail to reject the null hypothesis of no serial correlations in $r_t$ and $a_t^2$.

**Problem C.** (11 pts) Consider the quarterly earnings per share of the 3M Company from 1992 to 2009. We analyzed the logarithms of the earnings. That is, $x_t = \ln(y_t)$, where $y_t$ is the quarterly earnings per share.

1. (2 points) Test $H_o : \rho_4 = 0$ versus $H_a : \rho_4 \neq 0$, where $\rho_4$ is the lag-4 ACF of the differenced series of $x_t$. Compute the test statistic and draw a conclusion.

**Answer:** The $t$-ratio is $t = -0.119/\sqrt{67} = -0.974$, which is smaller than the 1.96 critical in modulus. Thus, $H_o$ cannot be rejected.

2. (5 points) Write down the final fitted model for $x_t$, including the residual variance.

**Answer:** $(1 - B)(1 - B^4)x_t = (1 + 0.183B - 0.817B^2)a_t$, where the variance of $a_t$ is 0.00693.

3. (2 points) Is the fitted model adequate? Why?

**Answer:** Yes, the model is adequate. Based on the Ljung-Box statistics, we have $Q(20) = 16.47$ with $p$-value 0.69 for the residuals and $Q(20) = 26.29$ with $p$-value 0.156 for the squared residuals.

4. (2 points) Obtain the 1-step and 2-step ahead forecasts of $y_t$ (not the log earnings $x_t$) at the forecast origin December 2009.

**Answer:** The prediction is $\exp[-0.2306 + 0.5 \times 0.0839^2] = 0.80$. That is, 80 cents per share.

5. (2 points) Let $\theta_1$ be the MA(1) coefficient. Test $H_o : \theta_1 = 0$ versus $H_a : \theta_1 \neq 0$. Compute the test statistic and draw a conclusion.

**Answer:** The test statistic is $t$-ratio $= 0.183/0.106 = 1.73$, which is smaller than the critical value 1.96. Thus, we cannot reject the null hypothesis $H_o$.

**Problem D.** (22 pts) Consider the monthly series of the U.S. 30-year fixed mortgage rate from April 1971 to March 2010. The rate is generally believed to be related to the bank prime rate. To understand the relationship between the two rates, we consider some simple analysis. The R output is attached. Answer the following questions:
1. (2 points) Let $Y_t$ and $X_t$ be the monthly mortgage and prime rate, respectively. A simple linear regression model $Y_t = \beta_0 + \beta_1 X_t + \epsilon_t$ is employed. Write down the fitted model, including R-square and the residual standard error.

**Answer:** $Y_t = 2.564 + 0.757 X_t + \epsilon_t$. The R-square of the model is 0.772 and residuals standard error is 1.35.

2. (2 points) Is the simple linear regression model adequate? Why?

**Answer:** No, because the Ljung-Box statistics show strong serial correlations in the residuals of the simple linear regression.

3. (3 points) Let $y_t$ and $x_t$ be the first differenced series of $Y_t$ and $X_t$, respectively. Is the correlation coefficient between $y_t$ and $x_t$ significantly different from zero at the 5% level? Why?

**Answer:** Yes, because the correlation coefficient is related to the coefficient of the linear regression. And the coefficient of the estimate simple linear regression is significant with $t$-ratio 16.79.

4. (4 points) The residuals of the regression $y_t = \beta x_t + \epsilon_t$ show certain serial dependence. An AR(2) model is identified for the residuals, resulting in a regression model with time series errors. Write down the fitted model, including residual variance.

**Answer:** $(1 - 0.40B + 0.26B^2)(y_t - 0.331x_t) = a_t$, where the residual variance is 0.0473.

5. (3 points) Is the model in part (4) adequate? Why?

**Answer:** Yes, the residuals of the model has no serial correlations because $Q(24) = 26.74$ with $p$-value 0.32 for the ACF of the residuals.

6. (3 points) Does the mortgage rate depend on the prime rate? Why?

**Answer:** Yes, because the coefficient of $x_t$ is highly significant in the fitted model.

7. (3 points) If pure AR models are entertained, an AR(3) model is specified for the differenced mortgage rate series. Write down the fitted AR(3) model, including residual variance.

**Answer:** $(1 - 0.579B + 0.386B^2 - 0.11B^3)y_t = a_t$, where the residuals variance is 0.0661.

8. (2 points) Use out-of-sample forecasts to compare the two models for the mortgage rates. Based on the output provided, select a model for the mortgage rate series. Explain your choice.

**Answer:** Based on RMSE, the regression model with time series errors outperforms the pure AR(3) model.