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

1. Give two reasons by which the return series of an asset tend to contain outliers.
   A: (1) Jumps in price and (2) the distribution of returns has heavy tails (high excess kurtosis).

2. Describe two differences between an AR(1) model and an MA(1) model of a time series.
   A: (1) ACF of MA(1) model has only one non-zero lag, but all ACF of an AR(1) model are non-zero. (2) MA(1) model is always stationary, whereas AR(1) may be unit-root non-stationary. [You may also state that MA(1) model has finite memory.]

3. Give two characteristics of the return $r_t$ if it follows the model $r_t = 0.05 + a_t$, $a_t = \sigma_t \epsilon_t$, where $\epsilon_t$ are iid $N(0, 1)$ and $\sigma_t^2 = 0.02 + 0.4a_{t-1}^2$.
   A: (1) Volatility clustering, (2) heavy tails because excess kurtosis is positive.

4. (Questions 4 to 6): Suppose that the asset return $r_t$ follows the model
   \[
   r_t = a_t, \quad a_t = \sigma_t \epsilon_t, \quad \epsilon_t \sim \text{iid } t_6^* \\
   \sigma_t^2 = 0.09 + 0.145a_{t-1}^2 + 0.855\sigma_{t-1}^2.
   \]
   Does the unconditional variance of $r_t$ exist? Why?
   A: No, it does not exist.

5. Suppose that $r_{100} = -0.05$ and $\sigma_{100} = 0.3$. Compute 1-step and 2-step ahead volatility forecasts at the forecast origin $t = 100$. (Note that it is volatility, not $\sigma^2$.)
   A: Since $r_t = a_t$, $r_{100}(1) = 0$ and $r_{100}(2) = 0$. For volatility, $\sigma_{101}^2 = 0.09 + 0.145r_{100}^2 + 0.855(0.3)^2 = 0.167$ so that $\sigma_{100}(1) = 0.409$. Note that the model is an IGARCH(1,1) model with drift in volatility, we have $\sigma_{100}^2(2) = 0.09 + \sigma_{100}^2(1) = 0.257$ so that $\sigma_{100}(2) = 0.507$. 

Booth School of Business, University of Chicago
Business 41202, Spring Quarter 2016, Mr. Ruey S. Tsay

Solutions to Midterm
6. Compute the 22-step ahead mean and volatility forecasts (one month ahead).
   A: The 22-step ahead forecast for mean is zero and the volatility \[ \sqrt{21 \times 0.09 + \sigma^2_{101}} = 1.434. \]

7. Give an advantage of Spearman’s \( \rho \) over the Pearson correlation.
   A: (1) Robust to outliers or (2) can show some nonlinear correlation.

8. Give a feature that GARCH-M models have, but the GARCH models do not.
   A: Ability to assess the risk premium of asset returns.

9. Suppose that \( r_t \) follows the model
   \[ r_t = r_{t-1} + a_t - 0.9a_{t-1}, \]
   and we have \( r_{1001} = 1.2 \) and \( r_{1000}(1) = 1.0 \), where \( r_{1}(1) \) denotes the 1-step ahead prediction of \( r_{t+1} \) at the forecast origin \( t \). Compute \( r_{1001}(1) \).
   A: \( a_{1001} = r_{1001} - r_{1000}(1) = 0.2 \). \( r_{1001}(1) = r_{1001} - 0.9a_{1001} = 1.2 - 0.9 \times 0.2 = 1.02. \)

10. Why is the usual \( R^2 \) measure not proper in time series analysis?
    A: The \( R^2 \) fails when the time series has a unit root.

    A: (1) Modeling and forecasting quarterly earnings, (2) Handle energy (oil) related time series. [You may use handling intraday seasonality too.]

12. (Questions 12-13) Suppose that the daily simple returns of an asset in week 1 were -0.5\%, 1.2\%, 2.5\%, -1.0\%, and 0.6\%. What are the corresponding daily log returns?
    A: Log returns are -0.501\%, 1.193\%, 2.469\%, -1.005\% and 0598\%.

13. What is the weekly simple return of the asset?
    A: The weekly log return is 2.754\% so that the weekly simple return is 2.792\%.

14. (Questions 14-15): The summary statistics of daily simple returns of an asset are given below:

   ```r
   > basicStats(rtn)
   
   nobs   2515.000000
   Mean   0.000410
   ```
SE Mean  ????????
LCL Mean -0.000257
UCL Mean  0.001077
Stdev  0.017060
Skewness  0.517184
Kurtosis  6.661044

What is the standard deviation of the mean? Is the expected return of the asset significantly different from zero? Why?

A: Standard deviation of the mean is 0.00034. No the expected return is not significantly different from zero because the 95% confidence interval contains zero.

15. Based on the summary statistics, are the returns normally distributed? Perform a statistical test to justify your conclusion.

A: No, it is not normally distributed. An obvious test is to use the excess kurtosis. The test statistic is \( 6.661 / \sqrt{24/2515} = 68.187 \), which is highly significant.

**Problem B.** (23 points) Consider the monthly U.S. unemployment rates from January 1947 to March 2016. Due to strong serial dependence, we analyze the differenced series \( x_t = r_t - r_{t-1} \), where \( r_t \) is the seasonally adjusted unemployment rate. Answer the following questions, using the attached R output. Note: A fitted ARIMA model should include **residual variance**.

1. (2 points) The `auto.arima` command specifies an ARIMA(2,0,2) model for \( x_t \). The fitted model is referred to as \textbf{m1} in the output. Write down the fitted model.

   A: \( x_t = 1.655x_{t-1} - 0.775x_{t-2} + a_t - 1.629a_{t-1} + 0.844a_{t-2}, \quad \sigma^2_a = 0.0384. \)

2. (3 points) Model checking shows two large outliers. An ARIMA(2,0,2) model with two outliers are then specified, \textbf{m3}. Write down the fitted model.

   A: \( (1 - 1.69B + 0.791B^2)(x_t + 1.53I_t^{(22)} - 1.147I_t^{(21)}) = a_t - 1.613a_{t-1} + 0.801a_{t-2} \) with \( \sigma^2_a = 0.0337. \)

3. (3 points) Model checking shows some serial correlations at lags 12 and 24. A seasonal model is then employed and called \textbf{m4}. Write down the fitted model.

   A: \( (1 - 0.554B^{12})(1 - 1.236B + 0.361B^2)x_t = (1 - 1.235B + 0.515B^2)(1 - 0.822B^{12})a_t, \) where \( \sigma^2_a = 0.0354. \)

4. (3 points) The outliers remain in the seasonal model. Therefore, a refined model is used and called \textbf{m5}. Write down the fitted model.
A: \[(1 - 0.549B^{12})(1 - 1.574B + 0.659B^2)(x_t + 1.476I_t^{(22)} - 1.144I_t^{(21)}) = (1 - 1.487B + 0.672B^2)(1 - 0.821B^{12})a_t, \quad \text{where } \sigma_a^2 = 0.0306.\]

5. (2 points) Based on the model checking statistics provided, are there serial correlations in the residuals of model m5? Why?

A: Yes, the model is adequate. The Ljung-Box statistics show that there are no serial correlations in the residuals.

6. (2 points) Among models m1, m3, m4 and m5, which model is preferred under the in-sample fit? Why?

A: Model m5. It has the smallest AIC value.

7. (2 points) If root mean squares of forecast errors are used in out-of-sample prediction, which model is preferred? Why?

A: Model m5. It has the smallest RMSE.

8. (2 points) If mean absolute forecast errors are used in out-of-sample comparison, which model is selected?

A: Model m5.

9. (2 points) Consider models m1 and m3. State the impact of outliers on in-sample fitting.

A: The outliers reduce \(\sigma_a^2\) markedly.

10. (2 points) Again, consider models m1 and m3. State the impact of outliers on out-of-sample predictions.

A: The impact of outliers is rather small. In fact, model m3 fares worse than model m1.

**Problem C.** (27 points) Consider the daily log returns of Amazon (AMZN) stock obtained via `quantmod`. Statistical analysis is included in the attached R output. Answer the following questions. Note, a model should include both mean and volatility equations and the innovation distribution used.

1. (2 points) Are there serial correlations in the daily log returns? Why?

Write down the proper null hypothesis for testing.

A: No, there are no serial correlations. The null hypothesis is \(H_0 : \rho_1 = \cdots = \rho_{10} = 0\) and the alternative hypothesis is \(H_a : \rho_i \neq 0\) for some \(1 \leq i \leq 10\), where \(\rho_i\) is the lag-i ACF of the return series. The Ljung-Box statistics give \(Q(10) = 14.02\) with \(p\)-value 0.172.

2. (3 points) A standard GARCH(1,1) model is fitted. Write down the fitted model.
A. The model is \( r_t = 0.00136 + a_t, \) \( a_t = \sigma_t \epsilon_t \) with \( \epsilon_t \sim_{iid} N(0, 1) \). The volatility equation is
\[
\sigma_t^2 = 2.0 \times 10^{-6} + 0.00816 \alpha_{t-1}^2 + 0.989 \sigma_{t-1}^2.
\]

3. (3 points) Model checking shows the normality is rejected. A skew standardized Student-\( t \) distribution is used. Write down the fitted model. Model \textbf{m3}.

A: Model is \( r_t = 0.00146 + a_t, \) \( a_t = \sigma_t \epsilon_t, \) where \( \epsilon_t \sim_{iid} t^{*}_{3.574}(1.077), \) where 1.077 and 3.574 denote the skew and degrees of freedom, respectively. The volatility equation is
\[
\sigma_t^2 = 4.0 \times 10^{-6} + 0.0227 \alpha_{t-1}^2 + 0.971 \sigma_{t-1}^2.
\]

4. (2 points) Based on the fitted model \textbf{m3}. Does the model support that the innovation is skewed? Perform a test to support your conclusion.

A: Yes. Consider \( H_0 : \text{skew} = 1 \) vs \( H_a : \text{skew} \neq 1. \) The test statistic is \( t = \frac{1.0766 - 1}{0.0316} = 2.432, \) which is greater than 1.96. Therefore, the distribution is skewed.

5. (2 points) Compute the 95\% interval forecasts for 1-step and 2-step ahead predictions using model \textbf{m3}.

A: 1-step ahead: (-0.0474, 0.0503). 2-step ahead (-0.0474, 0.0504).

6. (2 points) An IGARCH(1,1) model is also entertained. Write down the fitted model. Model \textbf{m4}.

A: \( r_t = 0.00153 + a_t, \) \( a_t = \sigma_t \epsilon_t, \) where \( \epsilon_t \sim_{iid} t^{*}_{3.28}(1.080). \) The volatility equation is
\[
\sigma_t^2 = 2.0 \times 10^{-6} + 0.0243 \alpha_{t-1}^2 + 0.9757 \sigma_{t-1}^2.
\]

7. (2 points) Why are the 1-step to 5-step ahead volatility forecasts of the IGARCH(1,1) model not constant?

A: Because the drift \( 2.0 \times 10^{-6} \) is in the equation.

8. (2 points) An EGARCH model is also entertained. Write down the fitted model? Model \textbf{m5}.

A: The model is \( r_t = 0.00181 + a_t, \) \( a_t = \sigma_t \epsilon_t, \) where \( \epsilon_t \sim_{iid} N(0, 1). \) The volatility equation is
\[
\ln(\sigma_t^2) = -0.456 - 0.0498 \epsilon_{t-1} + 0.128(|\epsilon_{t-1}| - 0.8) + 0.936 \ln(\sigma_{t-1}^2).
\]

9. (2 points) Based on the fitted EGARCH model, is the leverage effect significant? Why?

A: Yes, the estimated leverage parameter \( -0.0498 \) is significantly different from zero.
10. (3 points) The lag-1 VIX index is used as an explanatory variable for volatility. Write down the fitted model. Model m6

A: The model is \( r_t = 0.00191 + a_t \), \( a_t = \sigma_t \epsilon_t \), where \( \epsilon_t \sim_{iid} t^{*}_{3.958}(1.119) \). The volatility equation is

\[
\sigma_t^2 = 0.000026 \text{vix}_{t-1} + 0.0899 \sigma_{t-1}^2 + 0.000173 \sigma_{t-1}^2.
\]

Model checking indicates that the model may need further improvement.

11. (2 points) Based on the fitted model, does the lag-1 VIX index affect significantly the AMZN volatility? Why?

A: Yes, the coefficient of \( \text{vix}_{t-1} \) is significant at the 5% level. Thus, it contributes to the modeling of AMZN volatility.

12. (2 points) Among all volatility models entertained, which model provides best in-sample fit? Why?

A: Based on AIC, model m6 fits the data best.

---

**Problem D.** (10 points) Consider the monthly log returns of Procter and Gamble stock from January 1960 to March 2015. Use the R output to answer the following questions.

1. (2 points) An IGARCH(1,1) model is entertained. Write down the fitted model.

A: The model is \( r_t = a_t \), \( a_t = \sigma_t \epsilon_t \), where \( \epsilon_t \sim_{iid} N(0,1) \). The volatility equation is

\[
\sigma_t^2 = 0.084a_{t-1}^2 + 0.916\sigma_{t-1}^2.
\]

2. (2 points) Based on the statistics provided, is the model adequate? Why?

A: The Ljung-Box statistics of the standardized residuals and squared standardized residuals give \( Q(12) = 10.68 \) and 5.71, respectively. Both have high \( p \)-values. Thus, the mean equation and the volatility equation are adequate.

3. (4 points) Based on the fitted IGARCH(1,1) model, compute the 1-step and 2-step ahead forecasts for mean and volatility of the log returns.

A: 1-step ahead forecast: mean = 0, volatility = 0.0395, which is obtained by \( \sqrt{0.084(-0.03819212)^2 + 0.916(0.03961325)^2} \). 2-step ahead forecast: mean = 0, volatility = 0.0395 (the same as 1-step ahead).
4. (2 points) A GARCM-M model is entertained. Based on the fitted model, is the risk premium statistically significant? Perform a test to justify your answer.

A: The null hypothesis is $H_0 : \gamma = 0$ vs $H_a : \gamma \neq 0$. The test statistic is $t = 0.7036/1.5749 = 0.449$ with $p$-value 0.653. Thus, the null hypothesis cannot be rejected. That is, the risk premium is not significantly different from zero.

**Problem E.** (10 points) Consider the monthly log returns of value-weighted index and the S&P composite index from January 1960 to March 2015. Our goal is to study the relationship between the volatility of the two market indexes. Based on the output provided, answer the following questions:

1. (1 points) A GARCH(1,1) model with skew standardized Student-t innovations is employed for the S&P index returns. Does the fitted model support the use of skew innovations? Why?

A: The null hypothesis is $H_0 : \text{skew} = 1$ vs the alternative hypothesis $H_a : \text{skew} \neq 1$. The test statistic is $t = -5.157$, which is greater than 1.96 in absolute value. Thus, the null hypothesis is rejected. The model supports the skew innovations.

2. (2 points) A similar GARCH(1,1) model is also employed for the value-weighted index returns. Let the resulting volatility be $\text{vol}_{vw,t}$. Let $\text{vol}_{sp,t}$ be the corresponding volatility of the S&P index return. Write down the fitted simple linear regression model for the dependent variable $\text{vol}_{sp,t}$. Is this simple linear regression model adequate? Why?

A: The model is

$$\text{vol}_{sp,t} = 0.000792 + 0.949\text{vol}_{vw,t} + \epsilon_t,$$

where $\epsilon_t$ is the error term with $\sigma_\epsilon = 0.00189$. This model is inadequate as its residuals follow an AR(1) model.

3. (2 points) A refined model is employed. Write down the fitted linear regression model with time series errors.

A: The model is $(1-0.886B)(\text{vol}_{sp,t}-0.0012-0.939\text{vol}_{vw,t}) = a_t$, where $\sigma_a^2 = 7.603 \times 10^{-7}$.

4. (3 points) Alternatively, one can use $\text{vol}_{vw,t}$ as an explanatory variable in volatility modeling of the S&P index return. Write down the fitted volatility model.

A: The model is $\text{sp}_t = 0.00592 + a_t$, $a_t = \sigma_t \epsilon_t$, where $\epsilon_t \sim_{iid} t^*_{7.983}(0.77)$. The volatility equation is

$$\sigma_t^2 = 0.00492\text{vol}_{vw,t} + 0.116a_{t-1}^2 + 0.771\sigma_{t-1}^2.$$
5. (2 point) Does volvw, significantly contribute to the volatility modeling of the S&P index returns? Why?

A: The coefficient of volvw, has a $p$-value of 0.0613. Thus, the null hypothesis of no effect cannot be rejected at the 5% level. However, the volatility of VW index contributes significantly at the 10% level.