Solutions to Midterm

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

1. Describe the leverage effect of stock returns and give two classes of volatility models that can model the leverage effect.
   A: (1) post return and volatility has a negative correlation or past negative returns have a stronger impact than positive returns on volatility. (2) Any two of EGARCH, TGARCH or APARCH models.

2. Describe two ways that a GARCH model can induce heavy tails.
   A: (1) Innovational distributions and (2) the ARCH parameter of a GARCH model.

3. Write down an airline model for a quarterly time series. Why is the model useful in practice?
   A: \( (1 - B)(1 - B^4)r_t = (1 - \theta_1 B)(1 - \theta_4 B^4)a_t \), where \( a_t \) is an sequence of serially uncorrelated random variables with mean zero and variance \( \sigma_a^2 \). It is useful because it consists of a regular and a seasonal exponential smoothing model.

4. For questions 4 to 8, consider the daily adjusted closing prices of Johnson and Johnson (JNJ) stock from January 3, 2005 to April 26, 2011. We compute the daily log returns of JNJ from the price. Some summary statistics of the log returns are attached. Let \( p_t \) be the log closing price of JNJ. Is there evidence to support the hypothesis that \( p_t \) follows a random walk model with drift? Why?
   A: No, because the mean of the log returns is not significantly different from zero. The 95% C.I. is \([-0.000403, 0.000662]\).

5. Use Jacque-Bera statistics to test the null hypothesis that the log returns are normally distributed versus the alternative that they are not. What is the value of the test statistic and draw your conclusion.
   A: Use skewness and kurtosis. \( (0.700451/\sqrt{6/1589})^2 + (15.266/\sqrt{24/1589})^2 = 15559.02 \), which is highly significant. Thus, the log returns are not normally distributed.

6. What is the standard deviation of the mean, i.e. SE Mean, for the log returns?
   A: SE mean = 0.010822/\sqrt{1589} = 0.000271.

7. Let \( \rho_i \) be the lag-i ACF of the log returns. Based on the output, test \( H_0 : \rho_2 = 0 \) versus \( H_a : \rho_2 \neq 0 \). What is the test statistic and draw your conclusion.
   A: \( t = \sqrt{1589} \times (-.13225) = -5.272 \), which implies that the null hypothesis is rejected.
8. Test \( H_0 : \rho_1 = \rho_2 = \cdots = \rho_{10} = 0 \) versus \( H_a : \rho_i \neq 0 \) for some \( 1 \leq i \leq 10 \). What is the test statistic and draw your conclusion.

A: \( Q(10) = 66.24 \) with p-value close to zero. Thus, there exist serial correlations in the log returns.

9. For Questions 9 to 11, consider the quarterly AR(2) model

\[
r_t = -0.7 + 1.1r_{t-1} - 0.4r_{t-2} + a_t, \quad \text{Var}(a_t) = 16.
\]

Is the model weakly stationary? Why?

A: Yes, because the two roots of the characteristic equation \( 1 - 1.1x + 0.4x^2 = 0 \) are greater than 1 in absolute value.

10. Does the model imply the existence of business cycles? If yes, what is the average length of the cycles?

A: Yes, the characteristic equation has two complex roots. The average length of the cycles is 12.17 quarters.

11. Suppose that \( r_{300} = 7 \) and \( r_{299} = 5 \). What is the 1-step ahead prediction of \( r_{301} \) at the forecast origin \( T = 300 \)? Is zero in the 95\% confidence interval of the prediction?

A: \( \hat{r}_{300}(1) = -0.7 + 1.1 \times 7 - 0.4 \times 5 = 5 \). The 95\% interval is \( 5 \pm 2 \times \sqrt{16} \), which is \([-3,13]\). Thus zero is in the interval.

12. Describe two methods that can be used to test the hypothesis that a return series follows a skew distribution. Hint: one uses data directly and the other uses volatility modeling.

A: (1) Use skewness test of the returns. (2) Use a skew distribution in GARCH modeling and test the skew parameter being 1.

13. Why is the usual \( R^2 \) (coefficient of determination) not a good measure of the goodness of fit in a linear regression model?

A: The \( R^2 \) is not informative when there are serial correlations in the residuals, especially when the residuals have a unit root.

14. Consider the CAPM model \( r_t = \alpha + \beta r_{m,t} + \epsilon_t \), where \( r_t \) and \( r_{m,t} \) are the excess returns of a stock and market, respectively. Based on the volatility modeling discussed in class, describe a method to estimate the time-varying beta.

A: (1) Estimate the volatility of \( r_{m,t} \) via GARCH models. (2) Estimate the covariances between \( r_t \) and \( r_{m,t} \) using the property \( \text{cov}(r_t, r_{m,t}) = \mathbb{I} \left[ \text{cov}(r_t + r_{m,t}) - \text{cov}(r_t - r_{m,t}) \right] \).

15. Describe two methods that can be used to identify an AR model for a time series.

A: (1) PACF and (2) information criteria of AR fits.

**Problem B.** (30 pts) Consider the daily log returns of Google stock.
1. (2 points) Is there any ARCH effect in the daily log returns? Why?
A: Yes, because Q(10) of the squared returns is 87.27 with p-value close to zero.

2. (3 points) A volatility model, called \texttt{m1} in R, is entertained. Write down the fitted model, including the mean equation. Is the model adequate? Why?
A: The model is
\[ r_t = 7.33 \times 10^{-4} + a_t, \quad a_t = \sigma_t \epsilon_t, \quad \epsilon_t \sim N(0, 1) \]
\[ \sigma_t^2 = 4.586 \times 10^{-6} + 0.0384 a_{t-1}^2 + 0.953 \sigma_{t-1}^2. \]

Even though the model fits the data well, the normality assumption is clearly rejected.

3. (3 points) Another volatility model, called \texttt{m2} in R, is fitted to the returns. Write down the model, including the degrees of freedom.
A: The model is
\[ r_t = 1.074 \times 10^{-3} + a_t, \quad a_t = \sigma_t \epsilon_t, \quad \epsilon_t \sim t^*_{3.709} \]
\[ \sigma_t^2 = 2.498 \times 10^{-5} + 0.0476 a_{t-1}^2 + 0.951 \sigma_{t-1}^2. \]

This model is very close to an IGARCH model because \( \hat{\alpha}_1 + \hat{\beta}_1 = 0.997 \).

4. (2 points) Based on the fitted model \texttt{m2}, test \( H_0 : \nu = 5 \) versus \( H_a : \nu \neq 5 \), where \( \nu \) denotes the degrees of freedom of Student-t distribution. Perform the test and draw a conclusion.
A: The t-ratio is \((3.709 - 5)/0.436 = -2.96\), which is smaller than \(-2\). Thus, the null hypothesis is rejected. That is, the degrees of freedom is not 5.

5. (3 points) A third model, called \texttt{m3} in R, is also entertained. Write down the volatility equation, including the distribution parameters. Is the model adequate? Why?
A: The model is
\[ r_t = 1.203 \times 10^{-3} + a_t, \quad a_t = \sigma_t \epsilon_t, \quad \epsilon_t \sim t^*_{(3.66, 1.025)} \]
\[ \sigma_t^2 = 2.453 \times 10^{-5} + 0.0486 a_{t-1}^2 + 0.951 \sigma_{t-1}^2, \]

where \( t^*_{(3.66, 1.025)} \) denotes a skew Student-t distribution with 3.66 degrees of freedom and skew parameter 1.025. This model fits the data well as the residuals and their squared series have no serial correlations, but the model might over-fit the data because the skew parameter is close to 1.

6. (2 points) Let \( \xi \) be the skew parameter in model \texttt{m3}. Test \( H_0 : \xi = 1 \) versus \( H_a : \xi \neq 1 \). Perform the test and draw a conclusion.
A: The t-ratio is \( t = (1.025 - 1)/0.0429 = 0.583 \), which is less than 1.96. Thus, we cannot reject the null hypothesis. That is, there is no evidence to reject the hypothesis that the returns are symmetric.
7. (3 points) Again, focus on model m3 and use the last data point as the forecast origin. Compute 95% interval forecasts for the 1-step and 2-step ahead prediction of the log return.

A: 1-step ahead: $0.0012 \pm 2 \times 0.02103$. 2-step ahead: $0.0012 \pm 2 \times 0.02109$

8. (3 points) A fourth model, called m4 in R, is also fitted. Write down the fitted model, including the distribution of innovations.

A: The model is an APARCH model given by

$$
\begin{align*}
    r_t &= 5.803 \times 10^{-4} + a_t, \quad a_t = \sigma_t \epsilon_t, \quad \epsilon_t \sim N(0, 1) \\
    \sigma_t^2 &= 3.937 \times 10^{-6} + 0.013(|a_{t-1}| - 1.0a_{t-1})^2 + 0.965\sigma_{t-1}^2.
\end{align*}
$$

9. (2 points) Based on model m4, is the leverage effect significant? Why?

A: The t-ratio of leverage parameter is 3.301 with p-value 0.00096. Thus, the leverage effect is statistically significant at the 5% level.

10. (2 points) Among the volatility models entertained so far, which model is preferred? State the criterion used in your choice.

A: Based on the AIC, m2 is preferred. The same selection holds for BIC.

11. (3 points) Since the estimates $\hat{\alpha}_1 + \hat{\beta}_1$ is close to 1, we consider an IGARCH(1,1) model. Write down the fitted IGARCH(1,1) model.

A: The model is

$$
\begin{align*}
    r_t &= 6.902 \times 10^{-4} + a_t, \quad a_t = \sigma_t \epsilon_t, \quad \epsilon_t \sim N(0, 1) \\
    \sigma_t^2 &= 2.691 \times 10^{-6} + 0.0451a_{t-1}^2 + 0.9549\sigma_{t-1}^2.
\end{align*}
$$

12. (2 points) Use the IGARCH(1,1) model and the information provided to obtain 1-step ahead prediction for the log return and its volatility at the forecast origin $t = 1086$.

A: $r_{1086}(1) = 6.902 \times 10^{-4}$ and $\sigma_{1086}^2 = 2.691 \times 10^{-6} + 0.0451(0.1469 - 0.00069)^2 + 0.9549(0.02114)^2 = 0.000439$. Thus, the 1-step ahead forecast for the mean is essentially zero and the volatility is $\sqrt{0.000439} = 0.021$.

**Problem C.** (18 pts) Consider the quarterly earnings per share of Wal-Mart Stores from 1992 to 2010 for 76 observations.

1. (3 points) Write down the model m1 in R, including residual variance.

A: Let $r_t$ be the quarterly log earnings per share. The model is

$$
(1 - B)(1 - B^4)r_t = (1 - 0.337B)(1 - 0.354B^4)a_t, \quad \sigma_a^2 = 0.00297.
$$

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

A: Yes, the model is adequate. The model checking statistics all look fine.
3. (3 points) Write down the fitted model \( m2 \) in R, including residual variance.
   A: The model is
   \[
   (1 + 0.365B^4)(1 - B)(1 - B^4)r_t = (1 - 0.308B)a_t, \quad \sigma_a^2 = 0.00293.
   \]

4. (2 points) Is the model adequate? Why?
   A: Yes, the model checking statistics also indicate that the model is adequate.

5. (2 points) Compare the two models. Based on in-sample fit, which model is preferred? Why?
   A: The AIC for \( m2 \) is smaller so that it is the preferred model.

6. (2 points) Based on 1-step ahead out-sample prediction, which model is preferred? Why?
   A: \( m2 \) has a smaller RMSE. Thus, it is preferred. The difference is small, however.

7. (2 points) Based on 2-step ahead out-sample prediction, which model is preferred? Why?
   A: \( m2 \) has a smaller RMSE. Thus, it is preferred. The difference is also small.

Problem D. (22 pts) Consider the daily natural gas demand for a major northern city in the U.S. in one winter. It is well-known that gas demand depends on weather, such as temperature and windspeed, and weekly activities. Let \( y_t \) be the gas demand, \( x_{1t} \) be the degrees of heating days, defined as \( x_{1t} = 65 - \frac{\text{daily high temperature} + \text{daily low temperature}}{2} \), \( x_{2t} = x_{1,t-1} \) (the lagged value of \( x_{1t} \)), \( x_{3t} \) the windspeed and \( x_{4t} \) the dummy variable for weekend. Preliminary analysis indicates starting with a linear regression model using all four explanatory variables.

1. (4 points) Write down the fitted linear regression model, (called \( m1 \) in R), including R-square and the residual standard error.
   A: The linear regression model is
   \[
   y_t = 1.858 + 5.874x_{1t} + 1.405x_{2t} + 1.315x_{3t} - 15.857x_{4t} + e_t,
   \]
   where the standard error of \( e_t \) is 18.32 and the \( R^2 \) is 95.21%.

2. (2 points) Is the linear regression model adequate? Why?
   A: No, the linear regression model is not adequate. It’s residuals have serical correlations as shown by the residual ACF and PACF.

3. (5 points) Based on the ACF of the residuals, we employ an MA(7) model for the residuals. Initial estimates show some insignificant parameters in the model so that we further refine the model by removing insignificant parameters. Write down the refined model, called \( m3 \) in R, including residual variance.
   A: The model is
   \[
   y_t = 5.942x_{1t} + 1.296x_{2t} + 1.375x_{3t} - 12.175x_{4t} + (1 + 0.707B + 0.247B^2 + 0.358B^4 + 0.271B^6 + 0.279B^7)a_t,
   \]
   where \( \sigma_a^2 = 155.5 \).
4. (2 points) Based on the output provided, are there any serial correlations in the residuals of model m3? Why?
   A: The null hypothesis is $H_0 : \rho_1 = \cdots = \rho_{14} = 0$. No, the Ljung-Box statistics of the residuals give $Q(14) = 7.596$ with p-value 0.909. Thus, we cannot reject the hypothesis of no residual serial correlations.

5. (5 points) As an alternative, we employ a seasonal model for the residuals of linear regression in question 1. This model is called m4. Write down the fitted model, including residual variance.
   A: The fitted model is
   
   $$(1 - 0.536B)(1 - 0.368B^7)(y_t - 5.765x_{1t} - 1.473x_{2t} - 1.282x_{3t} + 9.595x_{4t}) = a_t,$$

   where $\sigma_a^2 = 189.5$.

6. (2 points) Between the two models (m3 vs m4), which one is preferred? Why?
   A: m3 is preferred based on the AIC criterion. Also, m3 has a smaller residual variance.

7. (2 points) Does the demand for natural gas decrease over the weekend? Why?
   A: Yes. Based on the preferred model m3, the weekend effect is statistically significant with t-ratio $-12.175/5.046 = -2.41$. Thus, we reject the null hypothesis of no weekend effect.