Solutions to Midterm

Problem A: (34 pts) Answer briefly the following questions. Each question has two points.

1. Describe two improvements of the EGARCH model over the GARCH volatility model.
   Answer: (1) allows for asymmetric response to past positive or negative returns, i.e., leverage effect, (2) uses log volatility to relax parameter constraint.

2. Describe two methods that can be used to infer the existence of ARCH effects in a return series, i.e., volatility is not constant over time.
   Answer: (1) The sample ACF (or PACF) of the squared residuals of the mean equation, (2) use the Ljung-Box statistics on the squared residuals.

3. Consider the IGARCH(1,1) volatility model: 
   \[ a_t = \sigma_t \epsilon_t \text{ with } \sigma_t^2 = \alpha_0 + \beta_1 \sigma_{t-1}^2 + (1 - \beta_1)a_{t-1}^2. \]
   Often one pre-fixes \( \alpha_0 = 0 \). Why? Also, suppose that \( \alpha_0 = 0 \) and the 1-step ahead volatility prediction at the forecast origin \( h \) is 16.2% (annualized), i.e., \( \sigma_h(1) = \sigma_{h+1} = 16.2 \) for the percentage log return. What is the 10-step ahead volatility prediction? That is, what is \( \sigma_h(10) \)?
   Answer: (1) Fixing \( \alpha_0 = 0 \) based on the prior knowledge that volatility is mean reverting. (2) \( \sigma_h(10) = 16.2 \).

4. (Questions 4 to 8) Consider the daily log returns of Amazon stock from January 3, 2007 to April 27, 2012. Some summary statistics of the returns are given in the attached R output. Is the expected (mean) return of the stock zero? Why?
   Answer: The data does not provide sufficient evidence to suggest that the mean return is not zero, because the 95% confidence interval contains zero.

5. Let \( k \) be the excess kurtosis. Test \( H_0 : k = 0 \) versus \( H_a : k \neq 0 \). Write down the test statistic and draw the conclusion.
Answer: \( t\)-ratio = \( \frac{9.875}{\sqrt{24/1340}} \) = 73.79, which is highly significant compared with \( \chi^2 \) distribution.

6. Are there serial correlations in the log returns? Why?
Answer: No, the Ljung-Box statistic \( Q(10) = 10.69 \) with p-value 0.38.

7. Are there ARCH effects in the log return series? Why?
Answer: Yes, the Ljung-Box statistic of squared residuals gives \( Q(10) = 39.24 \) with p-value less than 0.05.

8. Based on the summary statistics provided, what is the 22-step ahead point forecast of the log return at the forecast origin April 27, 2012? Why?
Answer: The point forecast \( r_T(22) = 0 \) because the mean is not significantly different from zero. [Give students 1 point if they use sample mean.]

9. Give two reasons that explain the existence of serial correlations in observed asset returns even if the true returns are not serially correlated.
Answer: Any two of (1) bid-ask bounce, (2) nonsynchronous trading, (3) dynamic dependence of volatility via risk premium.

10. Give two reasons that may lead to using moving-average models in analyzing asset returns.
Answer: (1) Smoothing (or manipulation), (2) bid-ask bounce in high frequency returns.

11. Describe two methods that can be used to compare different models for a given time series.
Answer: (1) Information criteria such as AIC or BIC, (2) backtesting or out-of-sample forecasting.

12. (Questions 12 to 14) Let \( r_t \) be the daily log returns of Stock A. Assume that \( r_t = 0.004 + a_t \), where \( a_t = \sigma_t \epsilon_t \) with \( \epsilon_t \) being iid N(0,1) random variates and \( \sigma_t^2 = 0.017 + 0.15a_{t-1}^2 \). What is the unconditional variance of \( a_t \)?
Answer: \( \text{Var}(a_t) = \frac{0.017}{1-0.15} = 0.02 \).

13. Suppose that the log price at \( t = 100 \) is 3.912. Also, at the forecast origin \( t = 100 \), we have \( a_{100} = -0.03 \) and \( \sigma_{100} = 0.025 \). Compute the
1-step ahead forecast of the log price (not log return) and its volatility for Stock A at the forecast origin \( t = 100 \).
Answer: \( r_{100}(1) = 0.004 \) so that \( p_{100}(1) = 3.912 + 0.004 = 3.916 \). The volatility forecast is \( \sqrt{\sigma^2_{100}(1)} = \sqrt{0.017 + 0.15(-0.03)^2} = 0.131 \).

14. Compute the 30-step ahead forecast of the log price and its volatility of Stock A at the forecast origin \( t = 100 \).
Answer: \( p_{100}(30) = 3.912 + 0.004 \times 30 = 4.032 \) and the volatility is the unconditional standard error \( \sqrt{0.02} = 0.141 \).

15. Asset volatility has many applications in finance. Describe two such applications.
Answer: Any two of (1) pricing derivative, (2) risk management, (3) asset allocation.

16. Suppose the log return \( r_t \) of Stock A follows the model \( r_t = a_t, a_t = \sigma_t \epsilon_t \), and \( \sigma^2_t = \alpha_0 + \alpha_1 a^2_{t-1} + \beta_1 \sigma^2_{t-1} \), where \( \epsilon_t \) are iid \( \text{N}(0,1) \). Under what condition that the kurtosis of \( r_t \) is 3? That is, state the condition under which the GARCH dynamics fail to generate any additional kurtosis over that of \( \epsilon_t \).
Answer: \( \alpha_1 = 0 \).

17. What is the main consequence in using a linear regression analysis when the serial correlations of the residuals are overlooked?
Answer: The \( t \)-ratios of coefficient estimates are not reliable.

**Problem B.** (30 pts) Consider the daily log returns of Amazon stock from January 3, 2007 to April 27, 2012. Several volatility models are fitted to the data and the relevant R output is attached. Answer the following questions.

1. (2 points) A volatility model, called \textbf{m1} in R, is entertained. Write down the fitted model, including the mean equation. Is the model adequate? Why?
Answer: ARCH(1) model. \( r_t = 0.0018 + a_t, a_t = \sigma_t \epsilon_t \) with \( \epsilon_t \) being iid \( \text{N}(0,1) \) and \( \sigma^2_t = 7.577 \times 10^{-4} + 0.188 a^2_{t-1} \). The model is inadequate because the normality assumption is clearly rejected.

2. (3 points) Another volatility model, called \textbf{m2} in R, is fitted to the returns. Write down the model, including all estimated parameters.
Answer: ARCH(1) model. \( r_t = 4.907 \times 10^{-4} + a_t, \ a_t = \sigma_t \epsilon_t, \) where \( \epsilon_t \sim t_{3.56}^* \) with \( t_{v}^* \) denoting standardized Student-t distribution with \( v \) degrees of freedom. The volatility equation is \( \sigma_t^2 = 7.463 \times 10^{-4} + 0.203a_{t-1}^2. \)

3. (2 points) Based on the fitted model \( \text{m2} \), 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.
Answer: \( t\text{-ratio} = \frac{3.562 - 5}{0.366} = -3.93 \), which compared with 1.96 is highly significant. If you compute the p-value, it is \( 8.53 \times 10^{-5} \). Therefore, \( \nu = 5 \) is rejected.

4. (3 points) A third model, called \( \text{m3} \) in R, is also entertained. Write down the model, including the distributional parameters. Is the model adequate? Why?
Answer: Another ARCH(1) model. \( r_t = 0.0012 + a_t, \ a_t = \sigma_t \epsilon_t, \) where \( \epsilon_t \) are iid and follow a skew standardized Student-t distribution with skew parameter 1.065 and degrees of freedom 3.591. The volatility equation is \( \sigma_t^2 = 7.418 \times 10^{-4} + 0.208a_{t-1}^2. \) Except for the insignificant mean value, the fitted ARCH(1) model appears to be adequate based on the model checking statistics shown.

5. (2 points) Let \( \xi \) be the skew parameter in model \( \text{m3} \). Does the estimate of \( \xi \) confirm that the distribution of the log returns is skewed? Why? Perform the test to support your answer.
Answer: The t-ratio is \( \frac{1.065 - 1}{0.043} = 1.67 \), which is smaller than 1.96. Thus, the null hypothesis of symmetric innovations cannot be rejected at the 5% level.

6. (3 points) A fourth model, called \( \text{m4} \) in R, is also fitted. Write down the fitted model, including the distribution of the innovations.
Answer: a GARCH(1,1) model. \( r_t = 0.0017 + a_t, \ a_t = \sigma_t \epsilon_t, \) where \( \epsilon_t \) are iid and follow a skew standardized Student-t distribution with skew parameter 1.101 and degrees of freedom 3.71. The volatility equation is \( \sigma_t^2 = 1.066 \times 10^{-5} + 0.0414a_{t-1}^2 + 0.950a_{t-1}^2. \)

7. (2 points) Based on model \( \text{m4} \), is the distribution of the log returns skewed? Why? Perform a test to support your answer.
Answer: The t-ratio is \( \frac{1.101 - 1}{0.043} = 2.349 \), which is greater than 1.96. Thus, the distribution is skew at the 5% level.
8. (2 points) Among models \textbf{m1, m2, m3, m4}, which model is preferred? State the criterion used in your choice.
Answer: Model 4 is preferred as it has a smaller AIC value.

9. (2 points) Since the estimates \( \hat{\alpha}_1 + \hat{\beta}_1 \) is very close to 1, we consider an IGARCH(1,1) model. Write down the fitted IGARCH(1,1) model, called \textbf{m5}.
Answer: \( r_t = \alpha_t, \; \alpha_t = \sigma_t \epsilon_t, \) where \( \sigma_t^2 = 3.859 \times 10^{-5} + 0.85 \sigma_{t-1}^2 + 0.15 \alpha_{t-1}^2 \).

10. (2 points) Use the IGARCH(1,1) model and the information provided to obtain 1-step and 2-step ahead predictions for the volatility of the log returns at the forecast origin \( t = 1340 \).
Answer: From the output \( \sigma_{1340}^2(1) = \sigma_{1341}^2 = 3.859 \times 10^{-5} + 0.85 \times (0.02108)^2 + 0.15 \times (1.46)^2 = 0.00361 \). Therefore, \( \sigma_{1340}^2(2) = 3.859 \times 10^{-5} + \sigma_{1340}^2(1) = 0.00365 \). The volatility forecasts are then 0.0601 and 0.0604, respectively.

11. (2 points) A GARCH-M model is entertained for the percentage log returns, called \textbf{m6} in the R output. Based on the fitted model, is the risk premium statistical significant? Why?
Answer: The risk premium parameter is \(-0.112\) with \( t\)-ratio \(-0.560\), which is less than 1.96 in modulus. Thus, the risk premium is not statistical significant at the 5% level.

12. (3 points) Finally, a GJR-type model is entertained, called \textbf{m7}. Write down the fitted model, including all parameters.
Answer: This is an APARCH model. The model is \( r_t = 0.0014 + a_t, \) \( a_t = \sigma_t \epsilon_t, \) where \( \epsilon_t \) are iid and follow a skew standardized Student-\( t \) distribution with skew parameter 1.098 and degrees of freedom 3.846. The volatility equation is
\[
\sigma_t^2 = 7.583 \times 10^{-6} + 0.0362(|a_{t-1}| - 0.478a_{t-1})^2 + 0.953\sigma_{t-1}^2.
\]

13. (2 points) Based on the fitted GJR-type of model, is the leverage effect significant? Why?
Answer: Yes, the leverage parameter \( \gamma_1 \) is significantly different from zero so that there is leverage effect in the log returns.
**Problem C.** (14 pts) Consider the quarterly earnings per share of Abbott Laboratories (ABT) stock from 1984.III to 2011.III for 110 observations. We analyzed the logarithms of the earnings. That is, $x_t = \ln(y_t)$, where $y_t$ is the quarterly earnings per share. Two models are entertained.

1. (3 points) Write down the model $m1$ in R, including residual variance.
   
   Answer: Let $r_t$ be the log earnings per share. The fitted model is
   
   $$(1 - B)(1 - B^4)r_t = (1 - 0.565B)(1 - 0.183B^4)a_t, \quad \sigma_a^2 = 0.00161.$$  

2. (2 points) Is the model adequate? Why?
   
   Answer: No, the Ljung-Box statistics of the residuals give $Q(12) = 25.76$ with p-value 0.012.

3. (3 points) Write down the fitted model $m2$ in R, including residual variance.
   
   Answer: The fitted model is
   
   $$(1 - B)(1 - B^4)r_t = (1 - 0.470B - 0.312B^3)a_t, \quad \sigma_a^2 = 0.00144.$$  

4. (2 points) Model checking of the fitted model $m2$ is given in Figure 1. Is the model adequate? Why?
   
   Answer: Yes, the model checking statistics look reasonable.

5. (2 points) Compare the two fitted model models. Which model is preferred? Why?
   
   Answer: Model 2 is preferred. It passes model checking and has a smaller AIC value.

6. (2 points) Compute 95% interval forecasts of 1-step and 2-step ahead log-earnings at the forecast origin $t = 110$.
   
   Answer: 1-step ahead prediction: $0.375 \pm 1.96 \times 0.038$, and 2-step ahead prediction: $0.0188 \pm 1.96 \times 0.043$. (Some students may use 2-step ahead prediction due to the forecast origin confusion.)

**Problem D.** (22 pts) Consider the growth rate of the U.S. weekly regular gasoline price from January 06, 1997 to September 27, 2010. Here growth rate is obtained by differencing the log gasoline price and denoted by $gt$ in R output. The growth rate of weekly crude oil from January 03, 1997 to September 24, 2010 is also obtained and is denoted by $pt$ in R output. Note that the crude oil price was known 3 days prior to the gasoline price.
1. (2 points) First, a pure time series model is entertained for the gasoline series. An AR(5) model is selected. Why? Also, is the mean of the \( g_t \) series significantly different from zero? Why?
   
   Answer: An AR(5) is selected via the AIC criterion. The mean of \( g_t \) is not significantly different from zero based on the one-sample \( t \)-test. The p-value is 0.19.

2. (2 points) Write down the fitted AR(5) model, called \( m_2 \), including residual variance.
   
   Answer: The fitted model is
   
   \[
   (1 - 0.507B - 0.079B^2 - 0.136B^3 + 0.036B^4 + 0.086B^5)g_t = a_t, \quad \sigma^2_a = 0.000326.
   \]

3. (2 points) Since not all estimates of model \( m_2 \) are statistically significant, we refine the model. Write down the refined model, called \( m_3 \).
   
   Answer: The fitted model is
   
   \[
   (1 - 0.504B - 0.074B^2 - 0.122B^3 + 0.101B^5)g_t = a_t, \quad \sigma^2_a = 0.000327.
   \]

4. (2 points) Is the refined AR(5) model adequate? Why?
   
   Answer: Yes, the Ljung-Box statistics of the residuals give \( Q(14) = 10.27 \) with p-value 0.74, indicating that there are no serial correlations in the residuals.

5. (2 points) Does the gasoline price show certain business-cycle behavior? Why?
   
   Answer: Yes, the fitted AR(5) polynomial contains complex solutions.

6. (3 points) Next, consider using the information of crude oil price. Write down the linear regression model, called \( m_4 \), including \( R^2 \) and residual standard error.
   
   Answer: The fitted linear regression model is
   
   \[
   g_t = 0.287p_t + \epsilon_t, \quad \sigma_\epsilon = 0.0184,
   \]

   and the \( R^2 \) of the regression is 33.66%.

7. (2 points) Is the fitted linear regression model adequate? Why?
   
   Answer: No, because the residuals \( \epsilon_t \) are serially correlated based on the Ljung-Box test.
8. (3 points) A linear regression model with time series errors is entertained and insignificant parameters removed. Write down the final model, including all fitted parameters.
Answer: The model is

\[(1-0.404B-0.164B^2-0.096B^3+0.101B^5)(g_t-0.191p_t) = a_t, \quad \sigma_a^2 = 0.000253.\]

9. (2 points) Model checking shows that the fitted final model has no residual serial correlations. Based on the model, is crude oil price helpful in predicting the gasoline price? Why?
Answer: Yes, because the fitted coefficient of \( p_t \) is significantly different from zero.

10. (2 points) Compare the pure time series model and the regression model with time-series errors. Which model is preferred? Why?
Answer: The regression model with time series error is preferred as it has a smaller AIC criterion.