**Mock Exam Answers**

**Question 1**

a) *Definition of a random sample*

This is a portion of the total population that has an equal probability of being chosen as the rest of the population. A random sample contains random variables.

b) *Likelihood function*

$$Pr(Y = y) = (1-p)^{y-1}p \qquad for\ 0 < p < 1 \quad and \quad y = 1,2,3,\ldots$$

The likehood function,

$$L(y_i|p) = \prod_{i=1}^{n} f(y,p) = \prod_{i=1}^{n} (1-p)^{y-1}p$$

$$= \boldsymbol{p^n(1-p)^{\sum_{i=1}^{n} y_i - n}}$$

The log likehood function; $l(y_i|p)$ is the log of $L(y_i|p)$

$$= \boldsymbol{log\left[p^n(1-p)^{\sum_{i=1}^{n} y_i - n}\right] = nlnp + \left(\sum_{i=1}^{n} y_i - n\right) ln(1-p)}$$

c) *Score, Hessian and information matrix.*

**Score**, $s(y_i|p)$ is the derivative of $l$ with respect to $p$

$$\frac{\partial l}{\partial p} = \frac{n}{p} - \frac{\boldsymbol{(\sum_{i=1}^{n} y_i - n)}}{1-p}$$

The **Hessian matrix** in case is a $1 \times 1$ matrix whose element is the second derivative of the log likelihood function.

$$\frac{\partial^2 l}{\partial p^2} = \frac{\partial}{\partial p}\left(\frac{n}{p} - \frac{(\sum_{i=1}^{n} y_i - n)}{1-p}\right) = \frac{n}{p} - \frac{(\sum_{i=1}^{n} y_i)}{1-p} + \frac{n}{1-p}$$

$$= -\left[\boldsymbol{\frac{n}{p^2} - \frac{\sum_{i=1}^{n} y_i}{(1-p)^2} + \frac{n}{(1-p)^2}}\right]$$

The **information matrix** $I(p)$ is the expected value of the Hessian matrix

$$= -\left[-\frac{n}{p^2} - \frac{n}{p(1-p)^2} + \frac{n}{(1-p)^2}\right] = \frac{-n(1-p)^2 - np + np^2}{p^2(1-p)^2}$$

Thus

$$= \left[\boldsymbol{\frac{n}{p^2(1-p)}}\right]$$

*d) Proof of the $E\{s(y_i|p)\}$ being zero*

$$E\left[\frac{n}{p} - \frac{(\sum_{i=1}^{n} y_i - n)}{1-p}\right] = \frac{n}{p} - \frac{n}{1-p}\left(\frac{1}{p}\right) + \frac{n}{1-p}$$

$$= \frac{n - np - n + np}{p(1-p)} = \frac{0}{p(1-p)} = 0$$

*e) Finding the MLE*

We equate the score to zero and solve for p

$$\frac{n}{p} - \frac{\sum_{i=1}^{n} y_i - n}{1-p} = 0$$

$$p\left(\sum_{i=1}^{n} y_i - n\right) = n(1-p)$$

$$\hat{p} = \frac{n}{\sum_{i=1}^{n} y_i} = \boldsymbol{\frac{1}{\bar{Y}}}$$

*f) Deriving the C-R lower bound*

If the variance of the estimator $\hat{p}$ attains the C-R lower bound then

$$Var(\hat{p}) = \frac{1}{I(p)} = \boldsymbol{\frac{p^2(1-p)}{n}}$$

*g) The Wald test statistic*

$$\frac{(\hat{p} - p_0)^2}{var(\hat{p})} \underline{H_0}\ \chi^2{}_1$$

$$n = 10; \sum_{i=1}^{n} y_i = 15; \quad at\, H_0 = \frac{1}{2}$$

From this, $\hat{p} = \frac{1}{\bar{Y}}$

$$\bar{Y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{15}{10} = 1.5$$

$\hat{p} = 0.667$

$$Var(\hat{p}) = \frac{p^2(1-p)}{n} = \frac{\frac{1}{4}\left(\frac{1}{2}\right)}{10} = \mathbf{0.0125}$$

Hence the Wald statistic is

$$\mathbf{2.23112} \quad \underline{\boldsymbol{H_0}} \quad \boldsymbol{\chi^2}_{\mathbf{1}}$$

**QUESTION 2**

*a) Checking whether GARCH model is correctly specified.*

These models are especially useful when the goal of the study is to analyze and forecast volatility.

If the GARCH model effectively performs its goal analyzing and forecasting volatility, then it can be argued out that it is correctly specified.

If the GARCH model satisfies both the sufficient and the necessary sufficient conditions, their conditional variance to be positive and the probability to be equal to 1, then the model is correctly specified.

*b) Leverage effect.*

This refers to an observed tendency of the volatility of an asset being negatively correlated with the returns of the asset. This can also be defined as the separation between the X-value and the mean of x-value. Leverage has a small residual since it does not always change the regression line. In most cases, it strengthens the correlation and the $R^2$ value. The stochastic volatility model accounts for leverage. It is described as follows:

$$\delta X_t = \left(\mu - \frac{v_t}{2}\right) dt + {v_t}^{\frac{1}{2}} dA_t$$

$$\delta v_t = k(\alpha - v_t)dt + Y {v_t}^{\frac{1}{2}} dB_t$$

$$where\ A\ and\ B\ are\ two\ standard\ Brownian\ motions$$

$$with\ E(dA_t dB_t) = dpt\ \ and\ \mu, \alpha, k, Y\ and\ p\ are\ constants$$

*c)*

$$\varepsilon_t = v_t \sqrt{h_t}; \quad v_t\ \ \underline{iid}\ N(0,1)$$

$$h_t = \omega + \alpha_1 {\varepsilon_{t-1}}^2 + \alpha_2 {\varepsilon_{t-2}}^2$$

Since $v_t$follows a normal distribution with $N(0,1)$ it follows that;

$$v_t = \frac{\sqrt{h_t}}{\varepsilon_t} \sim N(0,1)$$

$$L(t, \mu, \delta^2) = \frac{\sqrt{h_t}}{\varepsilon_t} \cdot \left(\frac{1}{2\pi}\right)^{\frac{h}{2}} e^{-\frac{1}{2}\varepsilon(t_1)^2}$$

*d) Derive var(εt) and var(εt|εt-1, εt-2…). Which assumption(s) do you need to make to ensure that var(εt) exists?*

The conditional variance, is given by $\sigma_t^2 = \alpha_0 + \alpha_1 u^2_{t-1}$.

$u_t^2 = \varepsilon^2_t [\alpha_0 + \alpha_1 u^2_{t-1}]$

$E_{t-1}[u^2_t] = \sigma\varepsilon^2[\alpha_0 + \alpha_1 u^2_{t-1}]$

$= 1[\alpha_0 + \alpha_1 u^2_{t-1}]$

$= \sigma_t^2$

The unconditional variance is given by,

$E_{t-2}E_{t-1}[u^2_t] = E_{t-2}[\alpha_0 + \alpha_1 u^2_{t-1}]$

$= \alpha_0 + \alpha_1 E_{t-2}[u^2_{t-1}]$

$= \alpha_0 + \alpha_0\alpha_1 + \alpha_1^2 u^2_{t-2}$

$E_{t-3}E_{t-2}E_{t-1}[u^2_t] = E_{t-3}[\alpha_0 + \alpha_0\alpha_1 + \alpha_1^2 u^2_{t-2}]$

$= \alpha_0 + \alpha_0\alpha_1 + \alpha_1^2 E_{t-3}[u^2_{t-2}]$

$= \alpha_0 + \alpha_0\alpha_1 + \alpha_0\alpha_1^2 + \alpha_1^3 u_{t-3}$

(...)

$E_0E_1E_2(\ldots)E_{t-2}E_{t-1}[u^2_t] = \alpha_0(1 + \alpha_1 + \alpha_1^2 + \ldots + \alpha_1^{t-1}) + \alpha_1^t u^2_0$

$= \frac{\sigma_0}{1-\alpha_1}$

$= \sigma^2$

The assumption for the existence of var($\varepsilon_t$) is that the mean $u_t$ *is* zero

e) The process followed by the sequence $\{\varepsilon_t^2\}$ is the white noise stochastic process.

**QUESTION 3**

Using the ARMA (2, 1) model:

$$y_t = y_{t-1} - \frac{1}{4}y_{t-2} + \frac{1}{2}\varepsilon_{t-1}$$

$$where\ \{\varepsilon_t\}\ is\ the\ white\ noise\ with\ mean\ 0\ and\ variance\ \delta^2$$

a) *Is the model stationary*

We write the process using lag operators:

$$\varphi_2(L)y_t = \vartheta_2(L)\varepsilon_t$$

Where $\varphi_2(L) = 1 - L - 0.25L^2\ and\ \vartheta_2(L) = 0.5L$

Solving for the roots of quadratic equation $1 - L - 0.25L^2$ we get the roots to be 4.828 and -0.828

Since $|L_1| > 1$ and $|L_2| <1$, we conclude that the model is weakly stationary

b) *Is the model invertible*

The model is invertible since the co-efficient of $0.5L$ in $\vartheta_2(L)$ is $\frac{1}{2}\ which\ is\ less\ than\ 1$

c) *Definition of autocorrelation function (acf) and partial autocorrelation function (pacf)*

Autocorrelation function is defined as the measure of the correlation between observations of a time series which are separated by *n* time units ie$(x_m - x_{m-n})$

The autocorrelation function does not control other lags

Partial autocorrelation function is the correlation between two variables, assuming that other values and set of variables are known and taken into account.

d) *Show that $\rho_k = \rho_{k-1} - \frac{1}{4}\rho_{k-2}$ for $k > 1$.*

$\varphi_2(L)y_t = \vartheta_2(L)\varepsilon_t$ Where $\varphi_2(L) = 1 - L - 0.25L^2\ and\ \vartheta_2(L) = 0.5L$

Solving for the roots of quadratic equation $1 - L - 0.25L^2$ we get the roots to be 4.828 and -0.828

Now that this model is at least weakly stationary, it implies that the condition in question above is true for all k>1

e) *Derive the two-step-ahead forecast.*

This is given by $\sigma_t^2 = \alpha_0 + \alpha_1 u^2{}_{t-1}$.

$u_t^2 = \varepsilon^2{}_t\,[\alpha_0 + \alpha_1 u^2{}_{t-1}]$

$E_{t-1}[u^2{}_t\,] = \sigma\varepsilon^2[\alpha_0 + \alpha_1 u^2{}_{t-1}]$

$= 1[\alpha_0 + \alpha_1 u^2{}_{t-1}]$

$= \sigma_t^2$

f) *Derive the two-step-ahead forecast error*

$x_1 = \mu + \epsilon_1 \qquad \epsilon_1 \sim N(0, \sigma_w^2/(1 - \varphi^2))$ .

For our case, we have $h_1 = 1/(1 - \varphi^2)$ and $h_t = 1$ for $t \geq 2$. Thus, the unconditional sum of squares is now

$S(\mu, \varphi) = (1 - \varphi\ 2)(x1 - \mu)2 + {}^n\sum_{t=2}[(x_t - \mu) - \varphi\ (x_{t-1} - \mu)]^2$ .

g) *Model selection criteria*

*Akaike Information Criterion (AIC)*

This is an estimator of the relative quality of statistical models for a given set of data. Whenever a models' collection for the data is given, AIC is then useful in estimating the quality of each model in relation to each of the other models. This implies that AIC provides a means for model selection. Suppose a statistical model of some data is given .

We let k to be the number of estimated parameters in the model.

We further let $\hat{L}\ be\ the\ maximum\ value\ of\ the\ likelihood\ function\ for\ the\ model.$

Then it follows that

$$AIC = 2k - 2ln(\hat{L})$$

*Bayesian Information Criterion*

This is a model selection criterion among a finite set of models. It is preferable to select models with the lowest BIC.

It is partially based on the likelihood function.

BIC is defined as follows

$$BIC = \ln(n)\, k - 2ln(\hat{L}) \quad where, x\ is\ the\ observed\ data$$

*n is the number of observations*

*k is the no. of parameters estimated by the model*

$\hat{L}$ *is the maximum likelihood function of the model* $M$

The properties of BIC are as follows;

It penalizes the number of parameters in a model.

It does not depend on the prior.

It is approximately equal to the minimum description length criterion with a negative sign

*h) Which model do you choose based on AIC?*

For easier interpretation, choose a model with higher AIC. This for our case, model, m1 is the best.

*i) Assume BIC/SBC selects a different model than AIC. What do you do?*

Depending on what you want to achieve, AIC is normally considered for situations when a false negative finding would be considered more misleading than a false positive. On the other hand, BIC is considered better for in situations that a false positive is as misleading as, or even more than a false negative.

**QUESTION 4**

GARCH stands for Generalized Autoregressive Conditional Heteroskedastic Model

It is represented as GARCH (p, q)

GARCH is obtained by expanding the residual term from white nose to an ARMA (p, q)

It is represented as $\varepsilon_t \sqrt[v_t]{h_t}$ where $v_t$ is the white noise term and

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^{p} \beta\ \ {}_i h_{t-i}$$ defines the conditional variance.

In estimation of GARCH model with parameters, with *k,q,p* we have

$$y_t = C + \sum_{i=1}^{k} a_i y_{t-i} + \epsilon_t$$

$$\epsilon_t = v_t \sqrt{h_t}$$

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^{p} \beta\ \ {}_i h_{t-i}$$

where $v_t$ represents the white noise term. Here, ${}_t$ is normally distribution with mean zero and conditional variance $h_t$, i.e

$$p\left(\epsilon_t \,|\epsilon_{t-1}, \cdots, \epsilon_0\right) = \frac{1}{\sqrt{2\pi h_t}} e^{-\frac{\epsilon_t^2}{2h_t}}.$$

The log-likelihood function of parameter vector $\theta = (\alpha_0, \alpha_1, \cdots, \alpha_q, \beta_1, \cdots, \beta_p)^T$ is

$$L(\theta) = \sum_{t=q+1}^{n} l_t(\theta) = \sum_{t=q+1}^{n} \left(-\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln h_t - \frac{\epsilon_t^2}{2h_t}\right)$$

Thus the gradient will be given by $\nabla L(\theta) = \frac{1}{2}\sum_{t=q+1}^{n}\left(\frac{\epsilon_t^2}{h_t^2} - \frac{1}{h_t}\right)\frac{\partial h_t}{\partial \theta}$

For the Fisher Information matrix, we have

$$J = \sum_{t=q+1}^{n} E\left[\left(\frac{\epsilon_t^2}{2h_t^2} - \frac{1}{2h_t}\right)\frac{\partial^2 h_t}{\partial\theta\partial\theta^T} + \left(\frac{1}{2h_t^2} - \frac{\epsilon_t^2}{h_t^3}\right)\frac{\partial h_t}{\partial\theta}\frac{\partial h_t}{\partial\theta^T}\right]$$

$$= -\frac{1}{2}\sum_{t=q+1}^{n} E\left(\frac{1}{h_t^2}\frac{\partial h_t}{\partial\theta}\frac{\partial h_t}{\partial\theta^T}\right).$$

When dealing with GARCH models, it is common and convenient to work with the likelihood function.

A local quadratic approximation can be used to obtain results of optimization problems. For the multidimensional optimization, we seek a zero of the gradient.

Thus, for the maximum likelihood problem

$$\hat{\theta} = \text{argmax} L(\theta),\ \theta \in \Theta$$

Fisher Information matrix in this case becomes.

$J = E\left(\frac{\partial^2 L}{\partial\theta\partial\theta^T}\right)$.

For its algorithm, given observations $\{y_t\}_{t=1}^{n}$, we may obtain $C, \hat{a}_1, \cdots, \hat{a}_k$ from best fitting autoregressive model $AR(k)$ and $y_t = \hat{C} + \sum_{i=1}^{k} \hat{a}_i y_{t-i} + \hat{\epsilon}_t$.