THE NEWTON UNIVERSITY

SEMESTER 2 2014/2015

TIME SERIES ECONOMETRICS

Question 1

$$Pr(Y = y) = \frac{e^{-\lambda} \lambda^{y}}{y!} \quad \lambda > 0, \quad \mathrm{y} = 0,1,2, \ldots$$

a) *This is Poisson distribution with parameter λ.*

   It is applicable when:

   - The occurrence is independent
   - The event is something countable in whole numbers
   - There is a known average frequency of occurrence for the time period

In earth quake engineering the probability of an earth quake occurring at a particular fault within a specific time interval follows a Poisson distribution.

b) *The likelihood function*

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

$$= \prod_{i=1}^{n} \frac{e^{-\lambda} \lambda^{y}}{y!} = \frac{e^{-\lambda \mathrm{n}} \lambda^{\sum_{i=1}^{n} y_i}}{\prod_{i=1}^{n} y_i!}$$

The log likehood function is

$$l(y|\lambda) = -\lambda \mathrm{n} + \sum_{i=1}^{n} y_i \ln \lambda - \sum_{i=1}^{n} \ln y_i!$$

c) ***Score***, *$s(y_i|\lambda)$ is the derivative of l with respect to λ*

$$\frac{\partial l}{\partial \lambda} = -n + \frac{\sum_{i=1}^{n} y_i}{\lambda}$$

The Hessian function is the derivative of the score with respect to λ

$$= -\left[-\frac{\sum_{i=1}^{n} \boldsymbol{y_i}}{\Lambda^2}\right]$$

The information function is the expected value of the Hessian function

$$= \left[-\frac{n}{\Lambda^2} E(y_i)\right]$$

But $E(y_i)$ is Λ hence

$$I(\Lambda) = \left[\frac{n}{\Lambda}\right]$$

d) Proof of E(score) is zero

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

$$= -n + \frac{n}{\Lambda}.\Lambda = 0$$

e) Solution of the MLE

We equate the score to zero and solve for Λ

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

$$\hat{\Lambda} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

But $E(y_i) = \bar{Y}$

Hence

$$\hat{\Lambda} = \boldsymbol{\bar{Y}}$$

f) The C-R lower bound

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

$$Var\left(\hat{\Lambda}\right) = \frac{1}{I(\Lambda)} = \frac{\Lambda}{n}$$

g) The MLE $\widehat{\Lambda}$ reaches the C-R lower bound because the C-R regularities are met.

h) Wald and score test

$$n = 20; \quad \sum_{i=1}^{20} y_i = 20$$

$$\chi^2{}_{n,\alpha} = \chi^2{}_{20,\ 0.05} = 31.41$$

The Wald statistic is given by

$$\frac{\left(\widehat{\Lambda} - \Lambda\right)^2}{var\left(\widehat{\Lambda}\right)} \underline{\Lambda = 2}\ \chi^2{}_1$$

$$\widehat{\Lambda} = \bar{Y} = \frac{\sum_{i=1}^{n} y_i}{n} \frac{20}{20} = 1$$

$$var\left(\widehat{\Lambda}\right) = \frac{\Lambda}{\mathrm{n}} = \frac{1}{20} = 0.05$$

The Wald statistic

$$\frac{(1-2)^2}{0.05} = 20 \qquad 20 < 31.41\ \ so\ we\ accept\ H_0$$

The score test

$$s(\theta_0) = \frac{U(\theta_0)^2}{I(\theta_0)} \qquad where{:}\ \ U(\theta_0)\ is\ the\ score\ and\ I(\theta_0) is\ the\ information\ function$$

For our case;

$$\frac{\left[-20 + \frac{20}{2}\right]^2}{\frac{20}{2}} = \frac{(-10)^2}{10} = 10 \quad 10 < 31.41\ so\ we\ accept\ H_0$$

QUESTION 2

a) *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.

b) *Difference between the autocorrelation function (acf) and sample autocorrelation function (sample acf)*

Autocorrelation function determines correlation at two stationary points in time series while sample autocorrelation function deals with the correlation between points that are not stationary within the population.

c) *i) Which of these sample autocorrelations and sample partial autocorrelations are significant at a 5% significance level?*

The lower and upper limits at 5% significance level is given by,

$$\pm {}^{1.96}/_{\sqrt{n}} \quad for\ our\ case, n\ is\ 225$$

Thus the limits are $[-0.13067,\ 0.13067]$

This comes with an implication that all the *sample acf* are significant, since we fail to accept them because they are not within the acceptance interval.

For the *sample pacf*, we fail to reject Lags 3, 5, 6, 7 and 8. This is because they fall within the acceptance interval. Their values therefore are not significant.

ii) Good time series for the data.

Sample acf would produce the best time series since all its Lag values are significant.

iii) Check whether the first two sample autocorrelation coefficients are jointly significantly different from zero using the Ljung-Box test at a 5 % significance level.

The test statistic for the Ljung-Box test is $Q = n(n+2)\ \sum_{k=1}^{h} \frac{\hat{\rho}^2 k}{n-k}$

$$where\ n\ is\ the\ sample\ size, \hat{\rho}_k\ is\ the\ sample\ autocorrelation\ at\ lag\ \text{k}, and\ h\ is\ the\ number\ of\ lags\ being\ tested.$$

The confidence interval is given by $Q > \chi^2_{1-\alpha,h}$= 5.99

From the Ljung-Box test, $\left[225(223)\frac{0.61^2}{224}\right] + \left[225(223)\frac{0.61^2 * 2}{223}\right] = 250.7937.$

We fail to accept the null hypothesis since the value formed falls in the rejection region.

The two sample autocorrelation are therefore jointly significant.

d) $y_t = \frac{1}{2} y_{t-1} + \varepsilon_t + \frac{1}{8}\varepsilon_{t-1} + \frac{1}{2}\varepsilon_{t-2}$ $\varepsilon t \sim^{iid}$ N(0; σ2).

$i)\ Is\ the\ model\ stationary$?

The model is not stationary since the co-efficient of L(φ) is 0.5 which is less than 1

ii) The one-step-ahead forecasts is given by,

$ut = {}^{\varepsilon_t}\sqrt{\alpha_0 + \alpha_1 u_{t-1}^2}$

$E_{t-1}[u_t] = E_{t-1}\ {}^{\varepsilon_t}\sqrt{\alpha_0 + \alpha_1 u_{t-1}^2}$

$E_{t-2}E_{t-1}[u_t] = 0$

The two-step-ahead forecasts is given by $\sigma_t^2 = \alpha_0 + \alpha_1 u^2{}_{t-1}$.

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

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

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

$= \sigma_t^2$

iii) The one-step-ahead is given by,

$Et\text{-}1[utut\text{-}1] = ut\text{-}1Et\text{-}1[ut] = 0$

The two-step-ahead is given by,

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

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

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

iv) Are the forecast errors correlated?

Without laws of generality, we can state that forecast errors are not correlated since one-step-ahead already yields zero as its value.

e) *Derive $V(\omega_t)$ from the equation $yt = \beta yt$-1 + εt; εt iid* $\sim$ N(0; σ2) with -1 < β < 1. Let $wt = \Delta yt = yt - yt$-1.
[Hint: V(A - B) = V(A) + V(B) - 2 Cov(A; B)]

(Xt α) = 1 (Xt1 α) + … + p (Xtp α) + εt.

When α = 0, if we take an initial condition having zero average (this is needed if we want stationarity),

then E [Xt] = 0 for all t. We may escape this restriction by taking α 6= 0. The new process Zt = Xtα

has zero average and satis…es the usual equation

Zt = 1Zt1 + ... + pZtp + εt.

But Xt satis…es

Xt = 1Xt1 + ... + pXtp + εt + (α 1α ... pα)

= 1Xt1 + ... + pXtp + εt + α. e

i

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 $y_t = \hat{C} + \sum_{i=1}^{k} \hat{a}_i y_{t-i} + \hat{\epsilon}_t$.

This model is mostly useful when the goal of the study is to analyze and forecast volatility.