INSTITUTE AND FACULTY OF ACTUARIES

EXAMINATION

19 April 2022 (am)

Subject CS2 – Risk Modelling and Survival Analysis Core Principles

Paper A

Time allowed: Three hours and twenty minutes

In addition to this paper you should have available the 2002 edition of the Formulae and Tables and your own electronic calculator.

If you encounter any issues during the examination please contact the Assessment Team on T. 0044 (0) 1865 268 873.

1 (i) Write down the two-factor Lee–Carter model, clearly defining each of the terms you use. [3]

A statistician is using the two-factor Lee–Carter model to project future mortality rates and has fitted the model to a set of mortality data. The statistician has observed that the fitted forces of mortality do not vary regularly with each calendar year but vary more regularly with age.

Therefore, the statistician suggests that, before projecting future mortality rates, the *k* parameters in the Lee–Carter model should be smoothed using penalised regression splines but the *a* and *b* parameters should be kept as separate, unsmoothed parameters for each age.

(ii) Discuss the statistician's suggestion by considering each of the three parameters of the Lee–Carter model, *k*, *a* and *b*, in turn. [3]

[Total 6]

2 An international pensions provider is interested in quantifying the force of mortality at certain ages for a particular country, for the period from 1 January 2017 through to 1 January 2020.

In this country, deaths are recorded by calendar year and classified by age at last birthday on the date of death, and the annual population censuses are completed on 30 June each year, classifying the population by age at last birthday.

Derive an expression for the crude force of mortality, $\hat{\mu}_{x+\frac{1}{2}}$, for a given age interval $[x, x + 1]$, for the period from 1 January 2017 to 1 January 2020 using the population and death data available, clearly stating all assumptions and defining each of the terms you use. [7]

3 During a football match, the referee can caution players if they commit an offence by showing them a yellow card. If a player commits a second offence that the referee deems worthy of a caution, they are shown a red card, and are sent off the pitch and take no further part in the match. If the referee considers a particularly serious offence to have been committed, a red card will be shown to a player who has not previously been cautioned, and the player will be sent off immediately.

The football team manager can also decide to substitute one player for another at any point in the match so that the substituted player takes no further part in the match. Due to the risk of a player being sent off, the manager is more likely to substitute a player who has been shown a yellow card. Experience shows that players who have been shown a yellow card play more carefully to try to avoid a second offence.

The rate at which uncautioned players are shown a yellow card is 1/8 per hour.

The rate at which players who have already been shown a yellow card are shown a red card is 1/12 per hour.

The rate at which uncautioned players are shown a red card is 1/20 per hour.

The rate at which players are substituted is 1/10 per hour if they have not been shown a yellow card, and 1/5 if they have been shown a yellow card.

An actuary decides to model a player's state during a match using a Markov jump process and the following four states:

- State U: uncautioned
- State Y: yellow card shown
- State R: red card shown
- State S: substituted.

(i) Write down the generator matrix of the Markov jump process that is used in the compact form of Kolmogorov's forward equations. [2]

A football match lasts 1.5 hours.

(ii) Solve Kolmogorov's forward equation for the probability that a player who starts the match remains in the game for the whole match without being substituted and without being shown a yellow card or a red card. [3]

(iii) Determine the probability that a player who starts the match is sent off during the match without having previously been cautioned. [4]

[Total 9]

4 An investigation was carried out into the effects of a newly developed medication to treat a potentially fatal illness, including an analysis of the effects of delaying the treatment after symptoms are first reported. A placebo that was designed to look like the real medication but that contained no active treatment was also used among the patients in the study to act as a control to test the effectiveness of the newly developed medication.

A Cox proportional hazards model was used to model the rate at which patients recovered from the illness. The following two covariates were used in the model:

X_1 = Medication indicator (0 = Placebo administered, 1 = Medication administered)
X_2 = Treatment delay indicator (0 = Treated on first day after symptoms developed
1 = Treated on second day after symptoms developed
2 = Treated after second day after symptoms developed)

The investigation followed a sample of 600 patients for 10 days after treatment. The number of patients in each of the six covariate groupings was as follows:

	$X_1 = 1$	$X_1 = 0$
$X_2 = 0$	100	100
$X_2 = 1$	100	100
$X_2 = 2$	100	100

The first two patients to recover were both treated with medication on the second day after symptoms developed. The first individual recovered 2 days after treatment and the second individual recovered 3 days after treatment. There were no censoring events prior to the second individual recovering.

(i) State the term in the partial likelihood expression that relates to the second individual recovering, clearly defining each of the terms you use. [4]

Following completion of the investigation, the coefficients for the two covariates were estimated as follows:

Covariate	*Coefficient*	*Standard error*
X_1	+0.15	0.02
X_2	–0.02	0.02

(ii) Comment on the impact of the two covariates implied by these results. [5]

[Total 9]

5 A banking regulator in a developing country is assessing the probability of insolvencies in the country's banking sector. Of the country's ten banks, two banks, A and B, are considered to be 'very large', and two banks, C and D, are considered to be 'large'.

The regulator wants to understand the likelihood that both of the two largest banks, A and B, will become insolvent in the next 12 months. Bank A has calculated its probability of insolvency in the next 12 months to be 10% while Bank B has calculated this to be 5%. A banking analyst at the regulator has commented that the probability of both becoming insolvent in the next 12 months is therefore 1 in 200.

(i) Discuss the analyst's comment. [4]

The regulator is now analysing the probability of insolvency of all four of the largest banks, A–D, in the next 12 months. Banks C and D have calculated their individual probabilities of insolvency over the next 12 months to be 3% and 8%, respectively. The regulator decides to model the dependency between the insolvency of the banks with the lower tail of a Clayton copula with a parameter value, $\lambda = 6$.

The generator function for the Clayton copula is:

$$\Psi(t) = \frac{1}{\lambda}\left(t^{-\lambda} - 1\right) \quad \text{where } -1 \leq \lambda < \infty, \lambda \neq 0$$

(ii) Determine, by first deriving a formula for the four-dimensional Clayton copula $C(u_1, u_2, u_3, u_4)$ in terms of the generator function and its pseudo-inverse function, the probability that all four banks will become insolvent in the next 12 months. [7]

[Total 11]

6 An actuary is working on a machine learning modelling project involving a very large number of covariates. A colleague suggests that the best approach is to include all the covariates in the model and use the method of maximum likelihood to estimate the regression parameters.

(i) Explain why this approach may not work well. [3]

The actuary consults a data scientist who suggests using penalised regression. This involves maximising a penalised log-likelihood of the form:

$$\log L\left(\beta_0, \beta_1, \ldots, \beta_p \mid x_1, \ldots, x_p\right) - \lambda P\left(\beta_1, \ldots, \beta_p\right)$$

where L is the standard likelihood function, P is the penalty function and λ is the regularisation parameter.

(ii) Explain the consequences of using too large or too small a value of lambda in the penalised log-likelihood function when estimating the regression parameters. [3]

The actuary decides to use the following penalty function P:

$$P\left(\beta_1, \ldots, \beta_p\right) = a\sum_{i=1}^{p} \beta_i^2 + (1-a)\sum_{i=1}^{p} \left|\beta_i\right|$$

where a is a hyperparameter with $0 \le a \le 1$.

(iii) State the type of model obtained by the actuary if $a = 0$ and if $a = 1$. [2]

The actuary fits the model to two data sets:

- Data set A, in which the actuary believes that only a small number of covariates have a non-negligible impact on the response
- Data set B, in which the actuary believes that most of the covariates have a non-negligible impact on the response.

(iv) Explain which data set the actuary should use a higher value of a for. [3]

[Total 11]

7 The annual aggregate claim amount, S, arising on a short-term insurance portfolio follows a compound Poisson distribution with parameter 5. Individual claim amounts follow a two-parameter Pareto distribution with parameters α and λ. A sample of individual claim amounts was taken and the sample mean and standard deviation were 10,000 and 15,000, respectively.

(i) Estimate the parameters of the Pareto distribution of the individual claim amounts using the method of moments. [5]

(ii) Determine the variance and the third central moment of S, using the estimated Pareto parameters from part (i). [6]

[Total 11]

8 The number of customers, N_t, in a queue at each integer time t is modelled using a Markov chain model.

At the start of each time interval, a number of customers following a Poisson distribution with parameter p join the queue, subject to the constraint that the number of customers in the queue can be no more than N_{max}.

At the end of each time interval, a number of customers following a Poisson distribution with parameter q, where $q > p$, are served and leave the queue, subject to the constraint that the number of customers in the queue cannot be negative.

(i) Comment on the limitations of this model. [2]

An analyst sets the model's parameters to $N_{max} = 2$, $p = 0.5$ and $q = 1$.

(ii) Verify, by separately considering the two transition matrices of customers joining the queue and customers leaving the queue, respectively, that the transition matrix of N_t is:

$$\begin{array}{c} 0 \\ 1 \\ 2 \end{array} \begin{pmatrix} 0.82207 & 0.14475 & 0.03318 \\ 0.48737 & 0.36788 & 0.14475 \\ 0.26424 & 0.36788 & 0.36788 \end{pmatrix}$$

[8]

(iii) Determine the stationary distribution of N_t. [8]

[Total 18]

9 A zero-mean, first-order moving average process is defined by the following equation:

$$X_t = e_t + be_{t-1}$$

where e_t is a sequence of independent and identically distributed $N(0, \sigma^2)$ random variables.

(i) Derive, in terms of b, the value of p that minimises:

$$E[(X_t - pX_{t-1})^2]$$ [7]

(ii) Comment on your answer to part (i) in the case where $b = 1$. [3]

(iii) Determine, in the case where $b = 1$, the values of q and r that minimise:

$$E[(X_t - qX_{t-1} - rX_{t-2})^2].$$ [8]

[Total 18]

END OF PAPER

INSTITUTE AND FACULTY OF ACTUARIES

EXAMINERS' REPORT

April 2022

CS2 – Risk Modelling and Survival Analysis
Core Principles
Paper A

Introduction

The Examiners' Report is written by the Chief Examiner with the aim of helping candidates, both those who are sitting the examination for the first time and using past papers as a revision aid and also those who have previously failed the subject.

The Examiners are charged by Council with examining the published syllabus. The Examiners have access to the Core Reading, which is designed to interpret the syllabus, and will generally base questions around it but are not required to examine the content of Core Reading specifically or exclusively.

For numerical questions the Examiners' preferred approach to the solution is reproduced in this report; other valid approaches are given appropriate credit. For essay-style questions, particularly the open-ended questions in the Specialist Advanced (SA) and Specialist Principles (SP) subjects, the report may contain more points than the Examiners will expect from a solution that scores full marks.

The report is written based on the legislative and regulatory context pertaining to the date that the examination was set. Candidates should take into account the possibility that circumstances may have changed if using these reports for revision.

Sarah Hutchinson
Chair of the Board of Examiners
July 2022

A. General comments on the *aims of this subject and how it is marked*

The aim of the Risk Modelling and Survival Analysis Core Principles subject is to provide a grounding in mathematical and statistical modelling techniques that are of particular relevance to actuarial work, including stochastic processes and survival models.

Some of the questions in this paper admit alternative solutions from those presented in this report, or different ways in which the provided answer can be determined. All mathematically correct and valid alternative solutions received credit as appropriate.

In cases where the same error was carried forward to later parts of the answer, candidates were given full credit for the later parts.

In higher order skills questions, where comments were required, well-reasoned comments that differed from those provided in the solutions also received credit as appropriate.

Candidates are advised to take careful note of all instructions that are provided with the exam in order to maximise their performance in future CS2A examinations. The instructions applicable to this diet can be found at the beginning of the solutions contained within this document.

B. Comments on *candidate performance in this diet of the examination.*

Performance was generally satisfactory. Most candidates demonstrated a reasonable understanding and application of core topics in mathematical and statistical modelling techniques.

The most poorly answered question on this paper was Question 9, on Time Series. Lack of time may have affected candidates' performance on this question. Candidates are reminded of the need to plan their time so that they are able to make a reasonable attempt at all the questions.

Question 2, on Exposed to Risk, was also relatively poorly answered. Candidates are reminded that examination questions will often test their ability to apply the concepts in the Core Reading to unfamiliar situations.

It is important that candidates heed all of the instructions provided with the examination paper. A number of candidates lost marks because they did not include workings for numerical questions despite being forewarned about this in the instructions.

Higher order skills questions were generally answered poorly. Candidates should recognise that these are generally the questions which differentiate those candidates with a good grasp and understanding of the subject.

The comments that follow the questions in the marking schedule below, concentrate on areas where candidates could have improved their performance. Candidates approaching the subject for the first time are advised to concentrate their revision in these areas.

C. Pass Mark

The Pass Mark for this exam was 57
1175 presented themselves and 452 passed.

Solutions for Subject CS2A – April 2022

Q1

(i)

The two-factor Lee-Carter model may be written:
ln $m(x,t) = a\_x + b\_x * k\_t + epsilon(x, t)$ [1]

where:
$m(x,t)$ is the central mortality rate at age x in year t [1]
a_x describes the general shape of mortality at age *x* / *a_x* is the mean of the time-averaged logarithms of the central mortality rate at age *x* [½]
b_x measures the change in the rates in response to an underlying time trend in the level of mortality of *k_t*. [½]
epsilon(x, t) are independently distributed normal random variables with means of zero and some variance to be estimated. [½]

[Marks available 3½, maximum 3]

(ii)

k parameters
The lack of regularity by calendar year is likely to be a genuine feature of the underlying process [1]
Mortality is subject to fluctuations from year to year, due to factors such as epidemics and harsh winters [½]
If the *k* parameters are smoothed using penalised spline regression and the penalty is used to project future mortality, then the projection is likely to place too much weight on the last few years of data [½]

a parameters
The *a* parameters represent the general shape of the mortality rates, which would be expected to vary in a regular manner with age [½]
This would suggest that smoothing might be required [½]
However, in this case, the fact that the fitted mortality rates vary regularly with age suggests that the *a* parameters also vary regularly enough to obviate the need for smoothing [½]
Smoothing of the *a* parameters would be more likely to be necessary for a small data set [½]

b parameters
The *b* parameters represent the variation of mortality improvements with age, which would be expected to be regular [½]
This would suggest that smoothing might be required [½]
Although the fitted mortality rates vary regularly with age in this case, it cannot be assumed that the *b* parameters also vary regularly enough to obviate the need for smoothing [½]
The *b* parameters require a higher volume of data to estimate reliably than the *a* parameters, increasing the likely need for smoothing [½]

If the b parameters are not smoothed, then the projected mortality rates at successive ages may cross over in the later years of the projection, which is not intuitively reasonable [½]

[Marks available 6½, maximum 3]

[Total 6]

Part (i) was very well answered.

Part (ii) was extremely poorly answered. Many candidates commented on the advantages and disadvantages of smoothing in general, without specific reference to the model in the question. Alternative comments that were clear, distinct and relevant to the context of the question were also awarded credit.

Q2

Let *d_x* = the total number of deaths aged *x* last birthday over the period 1st January 2017 to 1st January 2020 [½]

d_x = deaths aged *x* last birthday in 2017
+ deaths aged *x* last birthday in 2018
+ deaths aged *x* last birthday in 2019 [½]

Crude force of mortality (*mu_x*+1/2)^*hat* = *d_x* / (*c_E_x*) [½]

where *c_E_x* is the central exposed to risk corresponding to the deaths data [½]

Since deaths are classified by age last birthday at date of death,

c_E_x = INT(0,3):[*P*(*x*,1/1/2017+*t*)] *dt* [½]

where *P*(*x*, *t*) denotes the population aged *x* last birthday measured at time *t* [½]

This integral can be broken into the relevant four pieces as follows:

c_E_x
= INT(0,1/2):[*P*(*x*,1/1/2017+*t*)] *dt* + INT(1/2,3/2):[*P*(*x*,1/1/2017+*t*)] *dt*
+ INT(3/2,5/2):[*P*(*x*,1/1/2017+*t*)] *dt* + INT(5/2,3):[*P*(*x*,1/1/2017+*t*)] *dt* [½]

Assuming that population varies linearly between census dates [½]

we can use the trapezium rule which gives: [½]

c_E_x
= (1/4) * (*P*(*x*,1/1/2017) + *P*(*x*,30/6/2017))
+ (1/2) * (*P*(*x*,30/6/2017) + *P*(*x*,30/6/2018))
+ (1/2) * (*P*(*x*,30/6/2018) + *P*(*x*,30/6/2019))
+ (1/4) * (*P*(*x*,30/6/2019) + *P*(*x*,1/1/2020)) [1]

But *P*(*x*,1/1/2017) and *P*(*x*,1/1/2020) are unknown. They can be approximated by linear interpolation as follows:

P(*x*,1/1/2017) = (1/2) * (*P*(*x*,30/6/2016) + *P*(*x*,30/6/2017)) [½]

P(*x*,1/1/2020) = (1/2) * (*P*(*x*,30/6/2019) + *P*(*x*,30/6/2020)) [½]

Therefore,

c_E_x
= (1/8) * *P*(*x*,30/6/2016) + (7/8) * *P*(*x*,30/6/2017) + *P*(*x*,30/6/2018)
+ (7/8) * *P*(*x*,30/6/2019) + (1/8) * *P*(*x*,30/6/2020) [½]

[Total 7]

This question was poorly answered. Some candidates may not previously have encountered a question on Exposed to Risk where the census dates, between which the trapezium rule was to be applied, were different from the ends of the annual periods over which mortality rates were to be estimated. Candidates are reminded that examination questions will often test their ability to apply the concepts in the Core Reading to unfamiliar situations.

Q3
(i)
Generator matrix *A* of Markov jump process =

	U	**Y**	**R**	**S**
U	– 11 / 40	1 / 8	1 / 20	1 / 10
Y	0	– 17 / 60	1 / 12	1 / 5
R	0	0	0	0
S	0	0	0	0

[2]

(ii)
d*P_UU* (*t*) / d*t* = (– 11 / 40) * *P_UU* (*t*) [1]
P_UU (*t*) = exp ((– 11 / 40) *t*) * *constant* [½]
= exp ((– 11 / 40) *t*) since *P_UU* (*0*) = 1 q [½]

At the end of the match *t* = 3/2, [½]
so *P_UU* (3/2) = exp (−33/80) = 66.20% [½]

(iii)
Prob[sent off without being cautioned]
= INT(0,3/2):[*P_UU* (*s*) * (1 / 20)] d*s* [1]
= INT(0,3/2):[exp ((– 11 / 40) *s*) * (1 / 20)] d*s* [½]
= [-(2/11) * exp ((– 11 / 40) *s*)]:(0,3/2) [1]
= (2/11) * (1 – exp (−33/80)) [1]
= 6.15% [½]
[Total 9]

Parts (i) and (ii) were very well answered.

Answers to part (iii) were generally satisfactory. Candidates who calculated P_UR(3/2) were not awarded credit since this does not take account of the requirement that the player must not have been cautioned before being sent off.

Q4
(i)
exp(*Beta*_1 + *Beta*_2) / [½]
(100 * [exp(*Beta*_1 + 2 * *Beta*_2) + exp(*Beta*_1) + exp(*Beta*_2)
+ exp(2 * *Beta*_2) + 1] + 99 * exp(*Beta*_1 + *Beta*_2)) [2½]

Where:
*Beta*_1 = coefficient for the Medication Indicator covariate [½]
*Beta*_2 = coefficient for the Treatment Delay Indicator covariate [½]

(ii)
Treating with the medication increases the rate at which patients recover [½]
by exp (0.15) – 1 = 16.2% [1]
Given that the coefficient of the Medication Indicator covariate is not within two standard errors of the estimated coefficient [½]
we have sufficient evidence to reject the hypothesis that the medication has no impact on the recovery rate [½]
Delaying treatment reduces the rate at which patients recover [½]
by around 2% each day (1 – exp (-0.02)) [1]
However, given that the coefficient of the Treatment Delay Indicator covariate is within two standard errors of the estimated coefficient [½]
we have insufficient evidence to reject the hypothesis that delaying the medication has no impact on the recovery rate [½]

[Total 9]

Part (i) was poorly answered. Common errors were:

- *Including a separate covariate for treatment after the second day, whereas the question specifies that this scenario is to be dealt with by setting the covariate X_2 equal to 2.*
- *Using a numerator that did not reflect the fact that the second individual recovering had X_1 = X_2 = 1.*
- *Using a denominator that did not reflect the fact that the first individual recovering had X_1 = X_2 = 1, so that the term exp(Beta_1 + Beta_2) should have a coefficient of 99 rather than 100.*

Part (ii) was very poorly answered. Many candidates lost marks for:

- *Referring to the impacts of the covariates as impacts on the hazard rate, without specifying what the hazard rate represents in the context of the question, i.e. the recovery rate.*
- *Not quantifying the percentage impacts on the recovery rate implied by the parameter estimates.*
- *Stating that the covariate X_1 is significant and X_2 is not significant without justifying this conclusion by reference to the parameter estimates and their standard errors.*

Alternative comments that were clear, distinct and relevant to the context of the question were also awarded credit.

Q5

(i)
1 in 200 (or 0.5%) is a correct estimate of the probability of both banks becoming insolvent IF the two insolvency events are independent
(0.1*0.05 = 0.5%) [1]

If the two insolvency events were perfectly positively dependent / correlated then the correct estimate would take its maximum value of min(10%, 5%) = 5% [½]
If the two insolvency events were perfectly negatively correlated then the correct estimate would take its minimum value of max(10% + 5% -1, 0) = 0% [½]
It is most likely that the true estimate is greater than 0.5% [½]
as it is very likely that the insolvency of one of the "very large" banks would increase the probability of insolvency of the other "very large" bank [½]
because both banks are likely to be impacted by wider economic conditions [½]
and the strength of this positive association is likely to increase during times of financial crisis or poor economic conditions [½]
We need to understand more of the level of dependency / correlation before we can determine a more robust estimate [1]

[Marks available 5, maximum 4]

(ii)
C(*u*_1, …, *u*_4) = *psi*^[-1] (*psi*(*u*_1) + *psi*(*u*_2) + *psi*(*u*_3) + *psi*(*u*_4)) [1]
Where *psi*^[-1] is the pseudo-inverse function of the generator function *psi* [1]
Now, *psi*^[-1] (*t*) = (*lambda* * *t* + 1)^(-1 / *lambda*) [1]

Here,
t = (*u*_1^(-*lambda*) – 1 + *u*_2^(-*lambda*) – 1 + *u*_3^(-*lambda*) – 1
+ *u*_4^(-*lambda*) – 1) / *lambda* [1]

Therefore,
C(*u*_1, …, *u*_4) = (*u*_1^(-*lambda*) + *u*_2^(-*lambda*) + *u*_3^(-*lambda*)
+ *u*_4^(-*lambda*) – 3)^(-1 / *lambda*) [1]
Using the values from the question:
C(*u*_1, …, *u*_4)
= (0.1^(-6) + 0.05^(-6) + 0.03^(-6) + 0.08^(-6) – 3)^(-1 / 6) [1]
= 2.976% [1]

[Total 11]

Part (i) was poorly answered. Most candidates recognised that the probability of 1 in 200 had been obtained by assuming independence. Many candidates identified reasons why the assumption of independence might not hold, but many did not state that the true probability would be expected to be greater than 1 in 200. Few candidates explained that the strength of the positive association is likely to increase under poor economic conditions, making a copula with positive tail dependence appropriate. Alternative comments that were clear, distinct and relevant to the context of the question were also awarded credit.

Part (ii) was well answered.

Q6

(i)
Including a very large number of covariates in the model could introduce over-fitting into the model [½]

leading to identification of patterns specific to the training data that do not generalize to other data sets [½]
and hence to poor prediction accuracy on other data sets [½]
Too many covariates could cause the model to become complex and computationally expensive to run [1]
The model might become unstable

OR:
The parameter estimates of the model might have high variances [½]
due to high correlation among the variables [½]
In some cases, the optimisation of the likelihood function might not converge [1]
[Marks available 4½, maximum 3]

(ii)
EITHER:
When lambda = 0, the penalised log-likelihood reduces to the standard log-likelihood [1]
Too small a value of lambda can therefore lead to the problems associated with using just the maximum likelihood estimates [1]

OR:
Any 2 marks from the marking schedule for part (i), with reference to small values of lambda [2]

THEN:
Too large a value of lambda means only gross effects of the covariates will be included in the final model (underfitting) and many important effects of the covariates on the outcome may be missed [1]
As lambda tends to infinity, the parameter estimates tend to zero [½]
and so the model captures no relationships at all in the limit [½]
[Marks available 4, maximum 3]

(iii)
If $a = 0$, then the model is a LASSO regression model [1]
If $a = 1$, then the model is a ridge regression model. [1]

(iv)
The actuary should use a higher value of a for data set B [1]
For data set A, the parameters of lower significance will then be forced to zero, since the penalty P will outweigh the improvement in log-likelihood from including them [1]
For data set B, non-zero values will be fitted for the parameters of lower significance as well as those of higher significance, since the squaring of these parameters in the function P results in only a small penalty [1]
[Total 11]

Many candidates did not attempt this question and it was therefore poorly answered overall, apart from part (iii). Candidates are reminded of the need to be familiar with all aspects of the syllabus.

In part (i), alternative comments that were clear, distinct and relevant to the context of the question were also awarded credit.

Part (iv) was particularly poorly answered. Of those candidates who correctly identified that a higher value of a should be used for data set B, few gave an explanation demonstrating an understanding of the issues involved. Partial marks were awarded to candidates who demonstrated some understanding but reached the wrong conclusion.

Q7

(i)

For the sample mean we have

lambda_hat / (*alpha_hat* – 1) = 10,000 [1]

and for the sample variance we have

(*alpha_hat* * *lambda_hat*^2) / (((*alpha_hat* – 1)^2) * (*alpha_hat* – 2))

= 15,000^2 [1]

Substituting gives

(*alpha_hat* * 10,000^2) / (*alpha_hat* – 2) = 15,000^2 [1]

Hence, *alpha_hat* = 3.6 [1]

And, *lambda_hat* = 10,000 * (*alpha_hat* – 1) = 26,000 [1]

(ii)

Let X denote the individual claim amount

Then, we have

E[X^2] = var[X] + E^2[X] = $15,000^2 + 10,000^2$ = 3.25 * 10^8 [1]

Using the formula on page 14 of the Tables,

E[X^3]

= (GAMMA(*alpha_hat* – 3) * GAMMA(1 + 3) * *lambda_hat* ^3)

/ GAMMA(*alpha_hat*) [1]

= (3! * *lambda_hat*^3) / ((*alpha_hat*-1)*(*alpha_hat*-2)*(*alpha_hat*-3)) [1]

= (6 * 26,000^3) / ((2.6)*(1.6)*(0.6)) [½]

= 4.225 * 10^13 [½]

Therefore, using the formulae on page 16 of the Tables,

var[S] = 5 * 3.25 * 10^8 = 1.625 * 10^9 [1]

Third central moment of S = 5 * 4.225 * 10^13 = 2.1125 * 10^14 [1]

[Total 11]

Part (i) was the best answered question part on the whole paper.

*Answers to part (ii) were generally satisfactory, with most candidates determining the variance correctly but making incorrect or incomplete attempts to determine the third central moment. Candidates who evaluated the gamma functions using R or Excel, rather than using the recurrence relation GAMMA(x + 1) = x * GAMMA(x), were not penalised*

Q8

(i)

The Poisson assumption for the distribution of the number of customers joining the queue might be violated e.g. if customers join in groups rather than individually [1]

The Poisson assumption for the distribution of the number of customers being served might be violated e.g. if there is a lower limit on the time taken for a customer to be served [1]

The assumption of time-homogeneity might be violated e.g. if more customers join the queue at certain times of day [1]

Customers might leave the queue without being served [1]

In practice, customers will join and leave the queue continuously rather than at integer times [1]

In practice, there will not be a fixed upper limit to the number of customers in the queue [1]

[Marks available 6, maximum 2]

(ii)

Transition matrix, *A*, for customers joining the queue:

	0	**1**	**2**
0	exp(-0.5) = 0.60653	0.5*exp(-0.5) = 0.30327	1–0.60653–0.30327 = 0.09020
1	0	exp(-0.5) = 0.60653	1–0.60653 = 0.39347
2	0	0	1

[3]

Transition matrix, *B*, for customers leaving the queue:

	0	**1**	**2**
0	1	0	0
1	1–0.36788 = 0.63212	exp(-1) = 0.36788	0
2	1–0.36788–0.36788 = 0.26424	1*exp(-1) = 0.36788	exp(-1) = 0.36788

[3]

Hence, transition matrix for *N_t* = (*A*)(*B*): [1]

	0	**1**	**2**
0	0.82207	0.14475	0.03318
1	0.48737	0.36788	0.14475
2	0.26424	0.36788	0.36788

[1]

(iii)

The probabilities pi(i) for the stationary distribution satisfy:

0.82207 * pi(0) + 0.48737 * pi(1) + 0.26424 * pi(2) = pi(0) (1)

0.14475 * pi(0) + 0.36788 * pi(1) + 0.36788 * pi(2) = pi(1) (2)

0.03318 * pi(0) + 0.14475 * pi(1) + 0.36788 * pi(2) = pi(2) (3) [1]

Also pi(0) + pi(1) + pi(2) = 1 (4) [1]

Substituting (4) into (1) and (2):

0.82207 * (1 – pi(1) – pi(2)) + 0.48737 * pi(1) + 0.26424 * pi(2)
= 1 – pi(1) – pi(2) [½]

0.14475 * (1 – pi(1) – pi(2)) + 0.36788 * pi(1) + 0.36788 * pi(2) = pi(1) [½]

Giving:

$0.66530 * pi(1) + 0.44217 * pi(2) = 0.17793$ (5) [1]
$0.77687 * pi(1) – 0.22313 * pi(2) = 0.14475$ (6) [1]

From $0.22313 * (5) + 0.44217 * (6)$, $0.49196 * pi(1) = 0.10371$ [½]
i.e. $pi(1) = 0.2108$ [½]

Substituting in (5), $0.66530 * 0.2108 + 0.44217 * pi(2) = 0.17793$ [½]
i.e. $pi(2) = 0.0852$ [½]

Substituting in (4), $pi(0) = 1 – 0.2108 – 0.0852 = 0.7040$ [1]
Hence, the required stationary distribution is (0.7040, 0.2108, 0.0852)

[Total 18]

Part (i) was surprisingly poorly answered, despite the fact that a wide range of limitations would have gained credit. Many candidates stated assumptions of the model, with no reason given as to why they may be violated, which did not constitute limitations. Candidates are reminded of the need to read the question carefully. Alternative comments that were clear, distinct and relevant to the context of the question were also awarded credit.

Part (ii) was very poorly answered. Many candidates obtained the correct probabilities from the Poisson distributions but were unable to insert the probabilities into the correct cells of the transition matrices. Many candidates also failed to recognise that, since the rows of a transition matrix must sum to 1, the last column of the matrix A and the first column of the matrix B should be calculated as 1 less the sum of the other columns.

Part (iii) was fairly well answered, although some candidates lost marks for not showing sufficient workings to demonstrate that a valid method had been used.

Q9

(i)

EITHER:

$E[(X\_t – pX\_t\text{-}1)\text{^}2] = E[(e\_t + be\_t\text{-}1 - p(e\_t\text{-}1 + be\_t\text{-}2))\text{^}2]$ [½]
$= E[(e\_t + (b - p)e\_t\text{-}1 – bpe\_t\text{-}2)\text{^}2]$ [½]
$= E[e\_t\text{^}2] + (b - p)\text{^}2E[(e\_t\text{-}1)\text{^}2] + b\text{^}2p\text{^}2E[(e\_t\text{-}2)\text{^}2]$ [1]
where the cross terms vanish since the e_t are uncorrelated [½]

OR:

$E[(X\_t – pX\_t\text{-}1)\text{^}2] = var[X\_t – pX\_t\text{-}1]$ [½]
since $E[X\_t – pX\_t\text{-}1] = 0$ [½]
$= var[e\_t + be\_t\text{-}1 – p(e\_t\text{-}1 + be\_t\text{-}2)]$ [½]
$= var[e\_t + (b – p)e\_t\text{-}1 – bpe\_t\text{-}2]$ [½]
$= var[e\_t] + (b – p)\text{^}2var[e\_t\text{-}1] + b\text{^}2p\text{^}2var[e\_t\text{-}2]$ [½]

THEN:

Hence,
$E[(X\_t – pX\_t\text{-}1)\text{^}2] = (1 + (b - p)\text{^}2 + b\text{^}2p\text{^}2)\ sigma\text{^}2$ [1]
and we need to minimise $f(p) = 1 + (b - p)\text{^}2 + b\text{^}2p\text{^}2$. [1]

$df(p)/dp = -2(b - p) + 2b^2p$ [½]
which is equal to zero at $p = b / (1 + b^2)$. [1]
Since $d^2f(p)/dp^2 = 2(1 + b^2) > 0$, [½]
we have a minimum. [½]
Hence the required value of p is $p = b / (1 + b^2)$

(ii)
When $b = 1$, $p = 1 / (1 + 1^2) = ½$ [½]
This is stating that the best linear predictor (in mean squared error terms) of X_t given X_{t-1} is ½ X_{t-1} [1]
This is intuitively reasonable [½]
since the contribution to X_{t-1} from e_{t-1} will be reflected in X_t, but the contribution from e_{t-2} will not [1]

(iii)
EITHER:
$E[(X_t - qX_{t-1} - rX_{t-2})^2]$
$= E[(e_t + e_{t-1} - q(e_{t-1} + e_{t-2}) - r(e_{t-2} + e_{t-3}))^2]$ [½]
$= E[(e_t + (1 - q)e_{t-1} - (q + r)e_{t-2} - re_{t-3})^2]$ [½]
$= E[e_t^2] + (1 - q)^2E[(e_{t-1})^2] + (q + r)^2E[(e_{t-2})^2] + r^2E[(e_{t-3})^2]$ [1]
where the cross terms vanish since the e_t are uncorrelated [½]

OR:
$E[(X_t - qX_{t-1} - rX_{t-2})^2]$
$= var[X_t - qX_{t-1} - rX_{t-2}]$ [½]
since $E[X_t - qX_{t-1} - rX_{t-2}] = 0$ [½]
$= var[e_t + e_{t-1} - q(e_{t-1} + e_{t-2}) - r(e_{t-2} + e_{t-3})]$ [½]
$= var[e_t + (1 - q)e_{t-1} - (q + r)e_{t-2} - re_{t-3}]$ [½]
$= var[e_t] + (1 - q)^2var[e_{t-1}] + (q + r)^2var[e_{t-2}] + r^2var[e_{t-3}]$ [½]

THEN:
Hence, $E[(X_t - qX_{t-1} - rX_{t-2})^2]$
$= (1 + (1 - q)^2 + (q + r)^2 + r^2)\sigma^2$ [1]
and we need to minimise $g(q, r) = 1 + (1 - q)^2 + (q + r)^2 + r^2$ [½]
The partial derivative of g with respect to q is $-2(1 - q) + 2(q + r)$ [1]
and the partial derivative with respect to r is $2(q + r) + 2r$ [1]
Equating the partial derivatives to zero gives $2q + r - 1 = 0$, $q + 2r = 0$ [1]
Solving these equations simultaneously gives $q = 2/3$, $r = -1/3$ [1]

ALTERNATIVE SOLUTION:
r is the PACF at lag 2 [½]
Using the formula on page 41 of the Tables,
$r = -(1 - b^2) b^2 / (1 - b^6)$ [½]
i.e. $r = -b^2 / (1 + b^2 + b^4)$. [½]

Substituting $b = 1$ gives $r = -1/3$. [½]
$E[(X_t - qX_{t-1} - rX_{t-2})^2]$
$= E[(e_t + e_{t-1} - q(e_{t-1} + e_{t-2}) + 1/3\,(e_{t-2} + e_{t-3}))^2]$ [½]
$= E[(e_t + (1 - q)e_{t-1} - (q - 1/3)e_{t-2} + 1/3\,e_{t-3})^2]$ [½]
$= E[e_t^2] + (1 - q)^2E[(e_{t-1})^2] + (q - 1/3)^2E[(e_{t-2})^2]$

+ 1/9 E[(e_t-3)^2] [1]

where the cross terms vanish since the e_t are uncorrelated [½]
= (1 + (1 - q)^2 + (q – 1/3)^2 + 1/9) $sigma$^2 [1]
and we need to minimise g(q) = 1 + (1 - q)^2 + (q – 1/3)^2 + 1/9. [½]
dg(q)/dq = -2(1 – q) + 2(q – 1/3) = 4q – 8/3 [½]
which is equal to zero at q = 2/3. [½]
Since d^2g(q)/dq^2 = 4 > 0, [½]
we have a minimum. [½]
Hence q = 2/3 and r = -1/3.

[Total 18]

*Part (i) was poorly answered. Many candidates used E(X_t * X_t-1) = E(X_t) * E(X_t-1), which is not correct since X_t and X_t-1 are correlated. Many candidates also failed to check that the stationary point they derived was a minimum. Lack of time may have affected candidates' performance here. Since the command verb was "Derive", candidates who recognised that p is the PACF at lag 1 and used the formula on page 41 of the Tables received very few marks.*

Part (ii) was the most poorly answered question part on the whole paper. Of those candidates who obtained an expression for p in part (i), most evaluated it correctly at b = 1, but very few provided comments demonstrating an understanding of the significance of their result. Alternative comments that were clear, distinct and relevant to the context of the question were also awarded credit.

Part (iii) was very poorly answered. Many candidates who gave satisfactory answers to part (i) received few or no marks on part (iii), despite the similarities between the two parts. Lack of time may again have affected candidates' performance here.

[Paper Total 100]

END OF EXAMINERS' REPORT

