Imperial College London

MATH70131

BSc, MSci and MSc EXAMINATIONS (MATHEMATICS)
May 2023

This paper is also taken for the relevant examination for the Associateship of the Royal College of Science

Consumer Credit Risk Modelling

Date: 25 May 2023

Time: 10:00 – 12:00 (BST)

Time Allowed: 2 hrs

This paper has 4 Questions.

Please Answer All Questions in 1 Answer Booklet

Candidates should start their answers to each question on a new sheet of paper.

Supplementary books may only be used after the relevant main book(s) are full.

Any required additional material(s) will be provided.

Allow margins for marking.

Credit will be given for all questions attempted.

Each question carries equal weight.

DO NOT OPEN THIS PAPER UNTIL THE INVIGILATOR TELLS YOU TO

1. (a) Explain the differences between a *generic* and a *custom* credit score. (2 marks)

Let $Y \in \{0, 1\}$ indicate outcome: 0 if borrower does not default, 1 if borrower does default.

The following logistic regression scorecard model was built on a data set of 50,000 personal loans. The predictor variables are Age, Employment status (employed, unemployed), Annual Income, and Marital status (single, married, divorced, widow).

Variable	**Coefficient**	**Estimate**	**95% confidence interval**
(Intercept)	β_0	$+0.173$	
X_1: Age	β_1	-0.005	(-0.0461, 0.0361)
X_2: Employed	β_2	$+0.312$	(0.284, 0.340)
X_3: Monthly Income	β_3	$+0.0019$	(0.0002, 0.0035)
X_4: Single	β_4	$+0.076$	(-0.007, 0.160)
X_5: Married	β_5	$+0.214$	(0.196, 0.232)
X_6: Divorced	β_6	-0.438	(-0.462, -0.413)

(b) Is this an application or a behavioral scorecard? What are the differences between them? (2 marks)

(c) Explain why one of the categories of "Marital Status" is excluded from the model specification. (3 marks)

(d) If X_4 was "Widow", X_5 and X_6 remain as "Married" and "Divorce" respectively, and the excluded category was "Single", what would be the estimates of β_4, β_5 and β_6 ? (3 marks)

(e) Determine which risk factors are significant working with $\alpha = 5\%$. For the significant risk factors, describe how each is associated with default. (5 marks)

(f) As a credit risk analyst, what transformations of the predictor variables prior to inclusion in the model do you think could build a better model? Justify your answer. (2 marks)

(g) What is the probability of default of a divorced 45-years-old applicant who works as a waiter and makes 24,000 pounds annually? Leave your answer in terms of the exponential function. (3 marks)

(Total: 20 marks)

2. (a) Express the following concepts as probabilities as a function of an applicant's score ($S(x)$), the cut-off score (c) and the outcome (Y).

   (i) Specificity. (2 marks)

   (ii) Sensitivity. (2 marks)

In what follows we are going to focus on performance measures, in particular, in those that consider how good the model is at classifying borrowers. To that end, we are going to reason about a theoretical ROC curve, defined by the relationship between F_0 and F_1, throughout.

(b) Given the following receiver operating characteristic (ROC) curve:

$$F_1(c) = \sqrt{1-(F_0(c)-1)^2}, \qquad 0 \le F_0(c) \le 1, \quad 0 \le F_1(c) \le 1.$$

Compute the area under the curve (AUC).

Hint:

Probability density function of Beta Distribution:

$$f_X(x;\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \quad 0 < x < 1.$$

Where:

$$\Gamma(y) = \int_0^\infty s^{y-1}e^{-s}ds$$

In particular:

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}, \quad \Gamma\left(\frac{3}{2}\right) = \frac{\sqrt{\pi}}{2}, \quad \text{if } n \in \mathbb{Z}^+ \text{ then } \Gamma(n) = (n-1)!$$

(5 marks)

(c) Interpret the AUC in probabilistic terms.

(2 marks)

(d) Explain how to use the trapezoid rule to obtain an estimate of the AUC working with 5 trapezoids of equal base.

(3 marks)

(e) Compute the Kolmogorov-Smirnoff (KS) statistic.

(2 marks)

(f) Compute the Accuracy Rate (AR).

(1 mark)

(g) Prove that AUC=0.5 for a model with no classification power.

(3 marks)

(Total: 20 marks)

3. Consider the data set below. This shows 12 credit card customers with the following variables:

- X_1: Home owner (1=yes, 0=no)
- X_2: Age
- Y (Outcome): At least one failure to make a payment in a 12-month period (1=yes, 0=no)

Applicant No.	X_1	X_2	Y	Applicant No.	X_1	X_2	Y
1	1	38	0	7	1	51	1
2	0	42	0	8	1	32	0
3	0	27	0	9	1	44	0
4	1	49	0	10	0	35	1
5	0	29	1	11	1	56	0
6	0	34	1	12	0	36	0

(a) Transform Age into a binary variable X_2' using the indicator function $I_{\{X \leq 35\}}$. (2 marks)

(b) If you construct a regression tree for the outcome using the variables X_1 and X_2' with the square error as penalty measure (P_D), $n_{min} = 4$ and $P(X) \geq 0.75 \times P_D$ as stopping criteria, what would be the root node of the tree? Justify your answer. (4 marks)

(c) How is the probability of delinquency estimated at the leaf nodes? (2 marks)

(d) (i) Why is it considered a good practice to *prune* a decision tree? (2 marks)

(ii) Give an example of how to *prune* a decision tree. (2 marks)

(e) A range of techniques have been developed under the heading of *Reject Inference*.

(i) State briefly the question that these techniques aim to answer. (1 mark)

(ii) What is the strategy implied by the *augmentation* method? (3 marks)

(f) When it comes to asymmetric information, there is an important concept called *adverse selection*.

(i) Briefly explain the concept applied to the credit risk industry. (2 marks)

(ii) What would be the concept applied to the insurance industry? (2 marks)

(Total: 20 marks)

Mastery Question. This question is based on *Chapter 11: Capital Requirements and the Basel Accords* of the book *Credit Scoring and Its Applications* by Jonathan Crook, David Edelman, Lyn Thomas. Second Edition (2017).

4. (a) The Accord contains three sets of requirements known as pillars. What is covered by each of them? (3 marks)

   (b) Briefly describe how the minimum regulatory capital requirement (MRCR) for credit risk is determined in Basel I and Basel II. (4 marks)

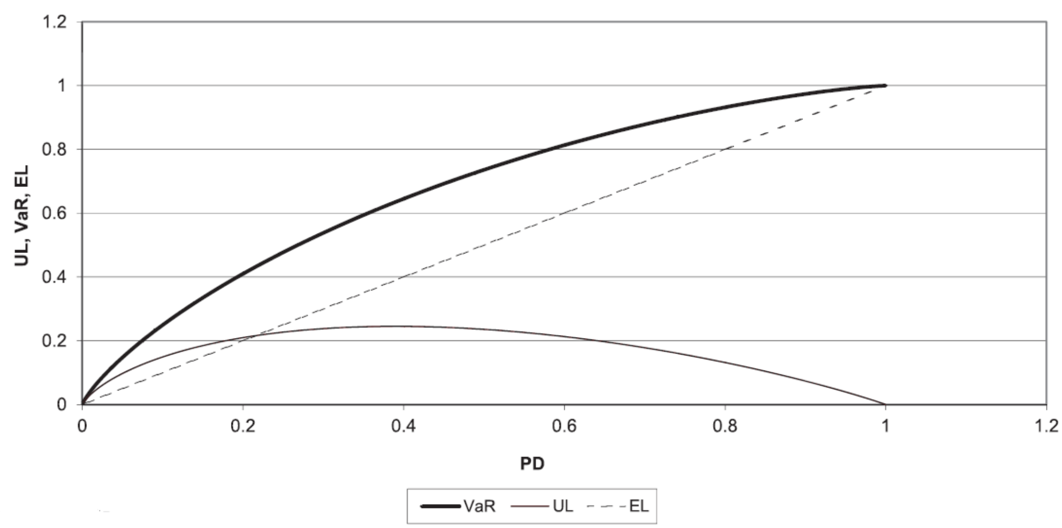
   (c) Let L_N be the total loss for a portfolio of N loans. Define the following concepts:

   (i) Value at Risk: $\text{VaR}_q(L_N)$. (2 marks)

   (ii) Expected Loss (EL), using any additional concepts you consider relevant. (3 marks)

   (iii) Unexpected Losses (UL). (2 marks)

   (d) Given the following graph for a retail portfolio, where PD stands for probability of default:



   Assuming that the loss given default (LGD) equals the exposure at default (EAD), what information can you extract from the graph? (3 marks)

   (e) Why is it crucial to count on this set of standards known as Basel Accords? (3 marks)

(Total: 20 marks)

BSc and MSci EXAMINATIONS (MATHEMATICS)

May 2023

This paper is also taken for the relevant examination for the Associateship.

MATH60131/MATH70131

Consumer Credit Risk Modelling (Solutions)

Setter's signature	Checker's signature	Editor's signature
	. . . .	. . . .

1. (a) Generic credit score seen ⇓

* Generic scores are often provided by credit bureaus.
* These are general scores about individuals regarding their individual characteristics (age, income, home ownership, etc) and credit history (number of times they have defaulted in the past, etc).
* Generic scores are not linked to particular loan products.

Custom credit score

* These are built by banks and financial institutions to compute creditworthiness for specific loan products.
* The credit scores will be computed from past data about borrowers with the same (or similar) loan product.
* For a specific product, custom credit scores are likely to be more accurate than generic scores.
* Generic scores are often used as a variable within a custom credit scoring model. 2, A

(b) It is an application scorecard as it is built on application data, as opposed to behavioural scorecards which assess the risk of existing customers based also on their recent accounting transactions, financial information, repayment performance, delinquencies, etc. part seen ⇓ 2, A

(c) At least one category value must be excluded since the combination of all categories is colinear. That is, the sets of events expressed by indicator (dummy) variables X_i^I are mutually exclusive and exhaustive. In particular: $\sum_{i=1}^n X_i^I = 1$. Including all the indicator variables would lead to infinite maxima in the likelihood function, and therefore, the model cannot be estimated with a unique solution. seen ⇓ 3, B

(d) $\hat{\beta}_4 = -0.076$
$\hat{\beta}_5 = 0.138$
$\hat{\beta}_6 = -0.514$ part seen ⇓ 3, D

(e) If the CI contains the 0, it provides no evidence that the true value is not zero, and therefore we consider there is not significant association between the j^{th} variable and default (and vice versa). part seen ⇓

* Age
  Not a significant variable to predict default.
* Being employed
  It is a statistically significant risk factor and has a negative association with default risk.
* Monthly income
  It is statistically significant and has a negative association with default.
* Being single
  It has not significant association with default risk relative to being widow.

* Being married
  It is a statistically significant risk factor and it has a negative association with default risk relative to being widow.
* Being divorced
  It is a statistically significant risk factor and it has a positive association with default risk relative to being widow.

5, A

(f)
* Age should be categorized, since relationship to credit score may not be linear.
* Annual income should be categorized or log transform applied since it usually has a right-skewed distribution.

meth seen ⇓

2, B

(g)

seen ⇓

$$S(x) = 0.173 - 0.005 \times 45 + 0.312 + 0.0019 \times 2,000 - 0.438 = 3.622.$$

Then,

$$P(Y = 1|X = x) = \frac{1}{1 + e^{S(x)}} = 0.026$$

3, A

2. (a) (i) Specificity (True negative rate) $\Rightarrow P(S(x) > c|Y = 0)$. seen $\Downarrow$

(ii) Sensitivity (True positive rate) $\Rightarrow P(S(x) < c|Y = 1)$. 2, A

(b) 2, A

unseen $\Downarrow$

$$AUC = \int_0^1 \sqrt{1-(x-1)^2}\,dx$$

$$= \int_{-1}^0 \sqrt{1-u^2}\,du$$

Let $y = 1 - u^2 \Rightarrow dy = -2udu$

Let $u = \sqrt{1-y} \Rightarrow du = -\frac{1}{2}(1-y)^{-\frac{1}{2}}dy$

Then,

$$AUC = \int_{-1}^0 \sqrt{y}\,du = \int_{-1}^0 -\frac{1}{2}y^{\frac{1}{2}}(1-y)^{-\frac{1}{2}}dy = \frac{1}{2}\int_0^1 y^{\frac{1}{2}}(1-y)^{-\frac{1}{2}}dy$$

$$= \frac{1}{2}\int_0^1 y^{\frac{3}{2}-1}(1-y)^{\frac{1}{2}-1}dy = \frac{1}{2}\frac{\Gamma(\frac{3}{2})\Gamma(\frac{1}{2})}{\Gamma(2)} = \frac{1}{2}\frac{\sqrt{\pi}}{2}\sqrt{\pi} = \frac{\pi}{4}$$

Alternatively, AUC is the area of a fourth of a circle of radius 1, and therefore is equal to $\frac{\pi}{4}$. 5, D

(c) AUC can be interpreted as the probability that a randomly drawn positive case has a lower score than a randomly drawn negative case. unseen $\Downarrow$

2, D

(d) First, we compute the 6 boundaries of the bases and the heights of the trapezoids as follows: seen $\Downarrow$

i	$F_0(c)$	$F_1(c)$
0	0	0
1	0.2	$\sqrt{1-(0.2-1)^2} = 0.6$
2	0.4	$\sqrt{1-(0.4-1)^2} = 0.8$
3	0.6	$\sqrt{1-(0.6-1)^2}$
4	0.8	$\sqrt{1-(0.8-1)^2}$
5	1	1

The area of each trapezoid is:

$$a_i = \frac{1}{2}\left[F_1(S_{i-1}) + F_1(S_i)\right]\left[F_0(S_i) - F_0(S_{i-1})\right]$$

Finally, the AUC estimate is:

$$\hat{A} = \sum_{i=1}^{5} a_i$$

3, B

(e) KS $= \text{max}_c |F_1(c) - F_0(c)|$ $\boxed{\text{part seen} \Downarrow}$

This is equivalent to the maximum length of the vertical line between the ROC curve and the base line.

Therefore, we need to maximize the following function:

$$f(x) = \sqrt{1-(x-1)^2} - x = (2x - x^2)^{\frac{1}{2}} - x$$

$$\frac{d}{dx}f(x) = \frac{1}{2}(2x-x^2)^{-\frac{1}{2}}(2-2x) - 1 = 0$$
$$(2x-x^2)^{-\frac{1}{2}}(1-x) = 1$$
$$(1-x) = (2x-x^2)^{\frac{1}{2}}$$
$$(1-x)^2 = 2x - x^2$$
$$1 - 2x + x^2 = 2x - x^2$$
$$2x^2 - 4x + 1 = 0$$
$$x_1 = 1 + \sqrt{\frac{1}{2}}\text{(rejected)}, x_2 = 1 - \sqrt{\frac{1}{2}}$$

Then, KS=1-$\sqrt{\frac{1}{2}}$ $\boxed{2, \text{C}}$

(f) Accuracy Rate = Gini = 2×AUC-1 = $2\frac{\pi}{4} - 1. = \frac{\pi}{2} - 1.$ $\boxed{\text{seen} \Downarrow}$

(g) If a model has no classification power, then $F_0(c) = F_1(c)$ for all c. $\boxed{1, \text{A}}$
Then,

$\boxed{\text{part seen} \Downarrow}$

$$AUC = \int_c F_0(c)F_0'(c)dc = \int_c F_0(c)\frac{dF_0(c)}{dc}dc$$

Integrating by parts,

$$= \left[F_0(c)^2\right]_c - \int_c F_0(c)dc = \left[\frac{1}{2}(F_0(c)^2)\right]_c = \frac{1}{2}(1-0) = \frac{1}{2}$$

$\boxed{3, \text{C}}$

3. (a) The values of X_2' are shown below: seen ⇓

Applicant no.	X_1	X_2'	Y	Applicant no.	X_1	X_2'	Y
1	1	0	0	7	1	0	1
2	0	0	0	8	1	1	0
3	0	1	0	9	1	0	0
4	1	0	0	10	0	1	1
5	0	1	1	11	1	0	0
6	0	1	1	12	0	0	0

2, A

(b) meth seen ⇓

$$\bar{y} = \frac{1}{12}\sum_{i=1}^{12} y_i = \frac{1}{3}$$

$$Var(Y) = \frac{1}{11}\sum_{i=1}^{12}(y_i - \bar{y})^2 = \frac{1}{11}\left(\sum_{i=1}^{12} y_i^2 - 12 \times \bar{y}^2\right) = \frac{1}{11}\left(4 - 12 \times (\frac{1}{3})^2\right) = \frac{24}{99}$$

For root node,

$$0.75 \times Var(Y) = \frac{3}{4} \times \frac{24}{99} = \frac{2}{11}$$

Then for X_1,

$$d(0) = \frac{3}{6} = \frac{1}{2}$$

$$d(1) = \frac{1}{6}$$

So

$$P(X_1) = \frac{1}{12}\sum_{i=1}^{12}(y_i - d(x_i))^2$$

$$= \frac{1}{12} \times \left(5 \times (0 - \frac{1}{2})^2 + 3 \times (0 - \frac{1}{6})^2 + 3 \times (1 - \frac{1}{2})^2 + 1 \times (1 - \frac{1}{6})^2\right) = \frac{25}{108}$$

and for X_2',

$$d(0) = \frac{1}{6}$$

$$d(1) = \frac{3}{6} = \frac{1}{2}$$

So

$$P(X_2') = \frac{1}{12}\sum_{i=1}^{12}(y_i - d(x_i))^2$$

$$= \frac{1}{12} \times \left(6 \times (0 - \frac{1}{6})^2 + 2 \times (0 - \frac{1}{2})^2 + 1 \times (1 - \frac{1}{6})^2 + 3 \times (1 - \frac{1}{2})^2\right) = \frac{19}{108}$$

Therefore X_2' gives the lowest penalty value and it is also lower than 0.75 $\times$ Var(Y), hence X_2' is chosen as root node. 4, C

(c) By computing the ratio of number of cases where Y=1 to the total number of cases in the leaf node. seen ⇓

2, A

(d)

(i) There is some evidence that simply using the stopping criteria to control the size of the decision tree and its accuracy as an estimator is not sufficient. Decision tree modelling algorithms (such as CART) also include a pruning stage to generate a **parsimonious** tree. seen ⇓ 2, B

(ii) One way to prune is to include a penalty term for the size of the decision tree (for example, considering the number of leaf nodes) and modify the penalty measure accordingly. 2, B

seen ⇓

(e)

(i) "*If* the rejected applicants had been given a loan, would they have defaulted?" 1, A

(ii) The idea is that categories of observations that are under-represented within the accepted applicants are given greater weight in the estimation. This would mean that the distribution of rejects is better represented during the estimation 3, B

(f)

(i) Those applicants who take the offer of a loan are likely to be a higher risk than the scorecard may suggest. part seen ⇓ 2, A

(ii) Those who buy insurance products are likely to be a higher risk (expected loss) than the company assesses (value of the premium). 2, D

4.

(a) This is an open-ended question, but candidates are expected to address the following:

• Pillar 1 relates to minimum capital requirements and specifies the methodology for computing the minimum amount of capital a bank is to hold in relation to credit risk, operational risk, and market risk.

• Pillar 2 specifies requirements of the supervisory review process and also introduces some other risks not included in Pillar 1, for example, credit concentration risk.

• Pillar 3 seeks to promote market discipline through regulatory disclosure requirements. $\boxed{3, \text{M}}$

(b) This is an open-ended question, but candidates are expected to address the following:

• Basel I: MRCR for credit risk is 8% of the risk-weighted assets (RWA).

• Basel II: There are 3 approaches: standardized approach, foundation and advanced internal ratings-based (IRB) approaches

– For standardized approach, it is similar to Basel I. Treatments are different for banks, corporations and sovereigns.

– For internal ratings-based (IRB) approaches, the capital charge for credit risk is the relative VaR or unexpected loss on the portfolio.

· The expected loss and the worst case default rate (WCDR) are calculated where their difference is the capital charge needed.

· The foundation IRB approach allows internal estimation of PD only.

· The advanced IRB approach allows internal estimation of PD, EAD, LGD and maturity, but not default correlation which must be specified by a standard formula depending on the exposure type. $\boxed{4, \text{M}}$

(c)

(i) $\text{VaR}_q(L_N)$ the qth quantile of the portfolio of loans loss distribution. $\boxed{2, \text{M}}$

(ii) Given the following:

· EAD_i: Exposure at default for loan i.

· LGD_i: loss given default for loan i, which is the proportion of the balance of a loan that is outstanding at the time of default that is never recovered by the bank.

· PD_i: Probability of default for loan i.

Assuming that PD_i, EAD_i, and LGD_i are uncorrelated, then for a portfolio of N loans the expected loss is given by

$$E(L_N) = \sum_{i=1}^{N} EAD_i \times LGD_i \times PD_i$$

$\boxed{3, \text{M}}$

(iii) The unexpected losses are the difference between the $\text{VaR}_{99.9}(L_N)$ and the expected loss. $\boxed{2, \text{M}}$

(d) • The expected loss, given the assumed values of LGD and EAD, is proportional to the value of the PD. In the hypothetical case that PD is equal to 1, the expected loss is 100%.

• The VaR increases monotonically with PD.

• The unexpected loss, which gives the amount of capital to be retained by the bank, increases, reaches a maximum, and then declines as PD increases. In the hypothetical case that PD was equal to 1, the unexpected loss is 0. $\boxed{3, \text{M}}$

(e) This is an open-ended question, but there are some key points (in bold) that candidates are expected to mention:

The Accords are designed to ensure that banks and financial institutions maintain **enough capital** on account to **meet their obligations** and also **absorb unexpected losses**. The purpose is to count on an **international standard** that **banking regulators** can use when creating regulations about how much capital banks need to put aside to guard against the **different types of risks banks face** (credit, operational, market, etc), while setting an upper bound to the **bankruptcy probability**, aiming to ensure the **solvency of the financial system** as a whole. $\boxed{3, \text{M}}$

Review of mark distribution:

Total A marks: 24 of 32 marks

Total B marks: 15 of 20 marks

Total C marks: 9 of 12 marks

Total D marks: 12 of 16 marks

Total marks: 80 of 80 marks

Total Mastery marks: 20 of 20 marks

If your module is taught across multiple year levels, you might have received this form for each level of the module. You are only required to fill this out once for each question.

ExamModuleCode	**QuestionNumber**	**Comments for Students**
MATH60131/70131	1	This question was generally well addressed. Some common mistakes were:- In the logistic regression model, to determine that risk factors are significant "because the point estimate is contained in the confidence interval". Note that the point estimate is always contained in the confidence interval (by construction of the interval). The justification is that those in which the 0 is not contained in the interval are the ones that are significant.- In the logistic regression model, to justify that one of the categories of "Marital Status" is excluded because "if we know the value of the other categories, then we can know the value of the excluded one". This statement is correct. But that's not the reason to exclude it from the model. Statistically speaking, a category must be excluded because otherwise there is a problem of multicollinearity and therefore the solution is not unique.- Not knowing the differences between a generic and a custom credit score. - Not knowing the differences between an application and a behavioural scorecard. - Not providing good / enough transformations on the predictor variables to build a better model.
MATH60131/70131	2	Overall, the question was well answered. Some common mistakes were:- Not knowing what the concepts "Specificity" and "Sensitivity" represent.- Difficulties in computing the integral to get the AUC (a significant number of students computed the integral well, while a few mentioned that the AUC is a quarter of a circle of radius 1, so the AUC is equal to pi/4, which was also a valid answer and didn't require for them to solve any integral).- Interpreting the AUC in probabilistic terms (most strudents explained what the AUC is, said that the higher its value is, the better the model, etc, but didn't addressed what was being asked).- Not knowing how to compute the KS statistic.- Showing that the AUC is equal to 0.5 for a model with no classification power, but not proving it analytically, as it was asked. In such cases partial credit was given.
MATH60131/70131	3	Overall, the question was well answered. Some common mistakes found:- Difficulties in building a decision tree. In some cases, procedures were correct but stopping criteria were not applied.- Not knowing how to estimate the probability of delinquency at the leaf nodes of a decision tree.- Not having a clear idea of what reject inference techniques aim to do.- Confusing the methods "augmentation" and "experimentation".- Not having a clear idea of the concept of "adverse selection". A number of students have confused "adverse selection" with "fraud" or some sort of "bad intention", which is not the same.- Difficulties to transfer the concept of "adverse selection" and apply it to a different industry.<u>Important:</u> Part (b) involved computations with fractional numbers and calculators were not allowed. Miscalculations were not penalised. As long as the decisions made were consistent with the results obtained, the corresponding credit was granted.
MATH70131	4	No Comments Received