Midterm Sample Paper · 3 hours

MATH60033A · Quantitative Methods in International Business

Thierry Warin, PhD · HEC Montréal · A2026

Format	**Paper. Handwritten, in pen.**
Duration	3 hours
Total	100 points
Permitted	a pen, and a **non-programmable calculator** (you will barely need it)
Not permitted	any computer, phone, tablet or smartwatch · any networked device · **no internet** · notes, textbooks, formula sheets
Answer	in the spaces provided · for multiple-choice questions, tick one box

No machine, and no assistant. Every other assessment in this course assumes you have a language model beside you, because that is how the work is now done. This one deliberately does not. It is the only instrument that isolates what **you** understand — which is the thing that lets you tell when the assistant is wrong for the rest of your career.

What this paper is testing, and why

The paper tests your **understanding** of the methods: what a number means, and what it does not license. There is very little arithmetic. The six parts test five different things:

Parts I and II — the readings. Can you state a paper's question, describe its model, criticise it, and name another way to answer it?

Part A — multiple choice. Do you hold the core ideas of sessions 2 to 5 precisely?

Part B — short answers. Can you state a concept in two or three exact sentences?

Part C — reading output. Given a table, can you say what it shows — and nothing more?

Part D — judgement. Knowing what a number does *not* license. The hardest part, and the one that separates an analyst from a technician.

Part	Points	Suggested time
I · The first reading	15	25 min
II · The second reading	10	20 min
A · Multiple choice	30	30 min
B · Short answers	18	25 min
C · Reading output	15	30 min
D · Judgement	12	20 min
Total	**100**	**150 min**, leaving 30 to check

Part I — The first reading

15 points · your Session 2 reading · answer in words

Fraiberger, Lee, Puy and Ranciere (2021), *Media sentiment and international asset prices* (Journal of International Economics). Answer from your reading of the paper.

I(a). *(3)* What is the **research question** of the paper? Why does it matter for an investor, or for a country that depends on foreign capital?

I(b). *(4)* Describe the **model**: what is the outcome, what is the unit of observation, how is the main explanatory variable built, what does it control for, and what method do the authors use?

I(c). *(4)* Give **two distinct critiques** a referee could raise about what the results allow the authors to conclude.

I(d). *(4)* Which **other model** from this course could answer a version of the same question? Say what its outcome would be, what it would report, and what it could tell you that the authors' model does not.

Part II — The second reading

10 points · your Session 4 reading · answer in words

Saganowski, Bródka, Koziarski and Kazienko (2019), *Analysis of group evolution prediction in complex networks* (PLoS ONE). Answer from your reading of the paper.

II(a). *(2)* What is the **research question** of the paper, and why does it matter in practice?

II(b). *(3)* Describe the **model**: what is the outcome, and what kind of variable is it? What is the unit of observation, what are the predictors, and how do the authors judge their classifiers?

II(c). *(3)* Give **two distinct critiques** a referee could raise about what the results allow the authors to conclude.

II(d). *(2)* Which **model from Session 4** fits this outcome? Say what it would report.

Part A — Multiple choice

30 points · 20 questions × 1.5 · tick one answer per question · no penalty for a wrong answer

1. *(S2)* The standard deviation of household income in a sample is 12,000. In what unit is it expressed?

- ☐ **A.** percent
- ☐ **B.** euros
- ☐ **C.** euros squared
- ☐ **D.** it has no unit

2. *(S2)* Why does the sample variance divide by $n-1$ rather than n?

- ☐ **A.** to make the variance larger, as a safety margin
- ☐ **B.** because one observation is always an outlier
- ☐ **C.** because the mean was estimated from the same data, so only $n-1$ deviations are free
- ☐ **D.** because the first observation is dropped

3. *(S2)* A variable is symmetric but has much heavier tails than a normal distribution. The 68–95–99.7 rule:

- ☐ **A.** applies exactly, because the variable is symmetric
- ☐ **B.** applies to the median only
- ☐ **C.** applies once the outliers are removed
- ☐ **D.** may fail: it holds for a roughly normal variable

4. *(S2)* The median of 3, 7, 8, 12 and 40 is:

- ☐ **A.** 8
- ☐ **B.** 14
- ☐ **C.** 12
- ☐ **D.** 7

5. *(S3)* Ordinary least squares chooses the line that:

- ☐ **A.** minimises the sum of squared vertical distances between the points and the line
- ☐ **B.** passes through the largest number of points
- ☐ **C.** maximises the slope
- ☐ **D.** minimises the number of outliers

6. *(S3)* A regression of GDP per capita on a productivity index (observed between 72 and 129) has an intercept of −86,000. You should:

- ☐ **A.** report that a country with no productivity has negative income
- ☐ **B.** drop the intercept from the model
- ☐ **C.** not interpret it: it is a prediction far outside the data
- ☐ **D.** conclude the model is wrong

7. *(S3)* Exports and productivity are positively related across firms. Which of these is **not** a reason to doubt that productivity causes exports?

- ☐ **A.** exporting may raise productivity (reverse causality)
- ☐ **B.** management quality may raise both (a common cause)
- ☐ **C.** only successful firms survive into the sample (selection)
- ☐ **D.** the sample has 5,000 firms

8. *(S3)* A predictor has a variance inflation factor (VIF) of 15. This means:

- ☐ **A.** it has no effect
- ☐ **B.** it is almost a combination of the other predictors, so its coefficient is unstable
- ☐ **C.** its coefficient is biased upward
- ☐ **D.** the model's R^2 is 15%

9. *(S3)* Two models are compared with AIC and BIC on the same rows. Which criterion favours the **smaller** model more strongly?

- ☐ **A.** AIC
- ☐ **B.** BIC
- ☐ **C.** both equally
- ☐ **D.** neither: they only measure fit

10. *(S3)* Why is GDP per capita often entered in logarithms?

- ☐ **A.** to reduce its right skew and the pull of a few very rich countries
- ☐ **B.** to make all coefficients positive
- ☐ **C.** because logarithms raise R^2
- ☐ **D.** because GDP cannot be negative

11. *(S4)* The odds of an event are 3. Its probability is:

- ☐ **A.** 0.33
- ☐ **B.** 0.25
- ☐ **C.** 0.75
- ☐ **D.** 3

12. *(S4)* A logistic regression is estimated by:

- ☐ **A.** minimising the squared residuals
- ☐ **B.** taking the mean of y in each group
- ☐ **C.** dropping the observations with $y = 0$
- ☐ **D.** maximum likelihood

13. *(S4)* In a logit, one more unit of x moves the **probability** the most when the predicted probability is near:

- ☐ **A.** 0
- ☐ **B.** 0.5
- ☐ **C.** 1
- ☐ **D.** the base rate

14. *(S4)* A missed default costs a bank far more than a false alarm. The threshold for flagging a borrower should be:

- ☐ **A.** below 0.5
- ☐ **B.** exactly 0.5
- ☐ **C.** above 0.5
- ☐ **D.** the base rate, always

15. *(S4)* A classifier has an AUC of 0.90. This measures:

- ☐ **A.** the share of correct predictions
- ☐ **B.** how close its probabilities are to observed frequencies
- ☐ **C.** how well it ranks positives above negatives
- ☐ **D.** its precision at the 0.5 threshold

16. *(S4)* Of the firms a model flags as "will fail", the share that actually fail is called the:

- ☐ **A.** recall
- ☐ **B.** accuracy
- ☐ **C.** base rate
- ☐ **D.** precision

17. *(S5)* As the penalty λ of a ridge regression increases:

- ☐ **A.** bias rises and variance falls
- ☐ **B.** bias and variance both fall
- ☐ **C.** bias falls and variance rises
- ☐ **D.** nothing changes on new data

18. *(S5)* There are more predictors than observations. Which method still gives a unique solution?

- ☐ **A.** OLS
- ☐ **B.** OLS with all the predictors standardised
- ☐ **C.** ridge
- ☐ **D.** none

19. *(S5)* Why must predictors be standardised before a lasso?

- ☐ **A.** to make the coefficients positive
- ☐ **B.** because the penalty depends on the units in which each predictor is measured
- ☐ **C.** to remove outliers
- ☐ **D.** to raise R^2

20. *(S5)* In 5-fold cross-validation, each observation is used to **test** the model:

- ☐ **A.** five times
- ☐ **B.** never
- ☐ **C.** four times
- ☐ **D.** exactly once

Part B — Short answers

18 points · 6 questions × 3 · two or three sentences each · be precise, not lengthy

B1. *(S2)* Why should you **describe** a variable — its centre, spread and shape — before you model it? Give two ways the description can change the analysis that follows.

B2. *(S3)* What does the **residuals-versus-fitted** plot show? What should it look like when the model is adequate, and what are two patterns that signal a problem?

B3. *(S3)* What is **omitted-variable bias**? Give an international-business example, and say in which direction it would push the estimate.

B4. *(S4)* In an **ordinal logit**, what are the **cutpoints**, and what is the **proportional-odds** assumption, in words?

B5. *(S5)* What is **data leakage** in cross-validation? Give an example, and say what it does to the reported error.

B6. *(S6)* Why would you add a **squared term** to a regression? How do you then read the effect of x, and what is a **turning point**?

Part C — Reading output

15 points · 3 items × 5 · answer in words · no calculation beyond reading the table

C1. Two regressions *(5 points)*

The export share of 400 manufacturing firms (in % of sales) is regressed on a productivity index, then on the same index and firm size (log of employees). One year, one country.

	Model 1		Model 2	
	coef	p	coef	p
const	-3.8	0.21	-12.4	0.001
productivity_idx	0.48	0.000	0.29	0.000
log_employees	-	-	2.10	0.000
R^2	0.21		0.30	

C1(a). *(2)* Write the sentence a reader needs for the productivity coefficient in Model 2.

C1(b). *(1.5)* Why does the productivity coefficient fall from 0.48 to 0.29 when firm size is added?

C1(c). *(1.5)* Model 2's intercept is −12.4. Should you interpret it? Explain.

C2. A logistic regression *(5 points)*

Whether a firm has adopted AI (1) or not (0), on two characteristics.

	coef (log-odds)	odds ratio e^coef
large_firm	1.10	3.00
does_rd	0.70	2.01

Predicted probability of adoption	no R&D	R&D
small firm	0.10	0.18
large firm	0.25	0.40

C2(a). *(2)* Write a correct sentence using the odds ratio 3.00.

C2(b). *(2)* Doing R&D adds 8 percentage points to the probability of adoption for a small firm, but 15 points for a large one, although its odds ratio is the same. Why?

C2(c). *(1)* A report says "large firms are three times more likely to adopt AI". Is that what the table shows? Explain.

C3. Choosing the penalty *(5 points)*

A lasso is fitted to predict R&D intensity from 40 candidate indicators, for five values of the penalty λ.

	0.001	0.01	0.1	1	10
error on training rows	0.80	0.85	0.95	1.20	1.60
cross-validation error	1.65	1.30	1.05	1.25	1.62
predictors kept	40	28	12	4	0

C3(a). *(2)* Which λ would you choose, and why?

C3(b). *(2)* The training error rises steadily with λ, while the cross-validation error first falls and then rises. Explain both, using the bias–variance trade-off.

C3(c). *(1)* What does the model at $\lambda = 10$ predict?

Part D — Interpretation and judgement

12 points · one paragraph · a well-argued refusal earns full marks

An export-credit agency has trained a classifier that flags "high-risk" export markets from 30 country indicators. On past data its **AUC is 0.93**, and political instability is its most important feature. An analyst proposes to **stop insuring all exports to the flagged countries**, and writes that "the model proves that political instability causes export failure".

Write the paragraph you would send in reply. It must state: **what the AUC and the feature ranking are** (3), **what they do not license**, and why (3), **what evidence would be needed** before acting on or explaining the model this way (3), and **what you would recommend the agency do in the meantime** (3).

End of examination.

Worked solutions, with the marks shown per sub-part: the answer key

More to train on: the practice questions, questions by session and by part, with an answer key · the midterm summary.

Sit the paper first, closed book and timed. The solutions are only worth reading against an attempt you have already made.