Midterm Sample Paper · Answer Key

MATH60033A · Quantitative Methods in International Business

Thierry Warin, PhD · HEC Montréal · A2026

How the marks are awarded

Three principles, applied throughout.

1. **Precision beats length.** A correct two-sentence answer scores full marks. A page that circles the right idea without stating it does not.
2. **A justified refusal earns full marks.** Particularly in Parts C and D. *"This cannot be claimed, and here is what would be needed"* demonstrates the thing the course is trying to build.
3. **Your own words.** The answers below are model answers. Any precise formulation earns the marks; in the critiques of Parts I and II there is no single right list, and any two distinct, precise critiques earn the marks.

Part I — The first reading *(15)*

Fraiberger, Lee, Puy and Ranciere (2021), *Media sentiment and international asset prices.*

I(a) *(3)* — Does the tone of the news about a country move its stock market: is a better-than-usual news day followed by higher returns in that country, and does **global** news matter more than **local** news? *(1.5)* It matters because international investors price countries partly on what they read: if sentiment moves prices beyond the fundamentals, a country that depends on foreign capital is exposed to swings in mood, not only to swings in its economy *(1.5)*.

I(b) *(4)* — Outcome: the country's stock-index return over the following days *(1)*. Unit: a country-day, for 25 countries over more than two decades *(0.5)*. The main variable is a **news-sentiment index** built from Reuters articles: the share of positive minus negative words, standardised within each country, so it measures a day that is better or worse than usual for that country *(1)*. Controls: past returns, trading volume, volatility and global risk (the VIX) *(0.5)*. Method: a linear regression of returns on lagged sentiment and the controls *(1)*.

I(c) *(4)* — Any two distinct critiques, 2 each:

- **Reverse causality**: journalists write about markets, so a rising market can make the news more positive; lagging sentiment by a day helps, but not if both respond to the same slow trend.
- **A common cause**: good economic news can raise both the tone of the articles and the returns; the coefficient then mixes sentiment with fundamentals.
- **Measurement**: counting positive and negative words is a constructed measure; it misses negation, irony and context, and the dictionary was not written for every country's news.
- **Size, not significance**: with over a hundred thousand country-days almost any effect is significant; the effect on a single day's return is small, and R^2 is tiny. What matters is whether it is economically large.

I(d) *(4)* — A **binary logit** of whether the market rises (1) or falls (0) the next day, on the sentiment index and the VIX *(1.5)*. It would report an **odds ratio** — how the odds of an up day change for one standard deviation more of sentiment — and **predicted probabilities** of an up day after bad, neutral and good news *(1.5)*. What it adds: a direct answer to the question an investor asks, "how much more likely is a rise?", without assuming that the effect on returns is a straight line *(1)*. (An ordinal logit of down, flat or up, or a lasso choosing among many text measures, is also accepted if justified.)

Part II — The second reading *(10)*

Saganowski, Bródka, Koziarski and Kazienko (2019), *Analysis of group evolution prediction in complex networks.*

II(a) *(2)* — Can a group's current structure in a social network predict **what happens to it next** — whether it dissolves, shrinks, grows, continues, merges or splits? *(1)* It matters for anyone who manages a community — a platform, a firm's internal network, a team: knowing which groups are about to break up lets them act before it happens *(1)*.

II(b) *(3)* — Outcome: the next event of the group, **six unordered categories** — a categorical, not an ordered, variable *(1)*. Unit: a group in a time window *(0.5)*. Predictors: about 90 network measures of the group — its

size, density, centrality, how long its members have been around *(0.5)*. The authors compare 15 classifiers, and judge them by the **average of the per-class F-measures**, not accuracy, because accuracy rewards predicting the common events *(1)*. Tree-based classifiers (bagging, random forest) do best.

II(c) *(3)* — Any two distinct critiques, 1.5 each:

- **Class imbalance**: merging and splitting are rare, so the classifiers learn little about them, and a good average can hide poor performance on exactly the events one may care about.
- **Correlated features**: 90 measures of the same groups overlap heavily, so the ranking of which measure matters is unstable, and should not be read as what *drives* a group's evolution.
- **Generalisation**: the results vary across data sets; a classifier tuned on one network may not travel to another.
- **Prediction, not explanation**: a good classifier says which groups will dissolve, not why.

II(d) *(2)* — A **multinomial logit** — the outcome has six categories with no natural order, so one category is the baseline *(1)*. It would report, for each event against the baseline, a **relative risk ratio** for each predictor, and the **predicted probability** of each event for a given group *(1)*. (A lasso penalty to handle the 90 correlated features earns credit as an addition.)

Part A — Multiple choice *(30)*

1.5 points per correct answer.

Q	Answer	Session	Why
1	**B**	S2	The SD is in the units of the variable; the variance is in squared units.
2	**C**	S2	The deviations from $\bar{x}$ sum to zero; dividing by n would underestimate the variance.
3	**D**	S2	Symmetric is not normal; heavy tails put more mass far from the mean.
4	**A**	S2	The middle of the sorted values; the 40 moves the mean (14), not the median.
5	**A**	S3	OLS minimises $\sum_i (y_i - \hat{y}_i)^2$.
6	**C**	S3	The intercept is the prediction at $x = 0$, a value no observation is near.
7	**D**	S3	A large sample makes the estimate precise; it does nothing about causality.
8	**B**	S3	Multicollinearity: the coefficient is hard to pin down and its sign can be unstable.
9	**B**	S3	BIC's penalty per parameter, $\ln n$, exceeds AIC's 2 as soon as $n > 7$.
10	**A**	S3	A log compresses the long right tail; slopes then read as proportional changes.
11	**C**	S4	Odds $= p/(1-p) = 3$ gives $p = 3/4$.
12	**D**	S4	It chooses the coefficients that make the observed 0s and 1s most probable.

Q	Answer	Session	Why
13	**B**	S4	The S-curve is steepest at 0.5: the slope is $p(1-p)\beta$.
14	**A**	S4	When misses are costly, flag at lower probabilities: $\tau^\star = c_{FP}/(c_{FP}+c_{FN})$.
15	**C**	S4	AUC is discrimination (ranking), not calibration and not accuracy.
16	**D**	S4	Precision $= TP/(TP+FP)$; recall is the share of failures that were flagged.
17	**A**	S5	More shrinkage: further from the truth on average, but more stable.
18	**C**	S5	$X^\top X + \lambda I$ is invertible for any $\lambda > 0$, even when $p > n$.
19	**B**	S5	Without it, a variable in euros and one in thousands of euros would be penalised differently.
20	**D**	S5	Each fold is held out once; each observation sits in one fold.

Part B — Short answers *(18)*

B1 *(3)* — The description shows what a model will have to cope with: skew, outliers, implausible values, missing data *(1)*. It can change the analysis by leading you to report the median rather than the mean, to model the logarithm of a skewed variable, to investigate or correct an error, or to treat extreme observations with care *(2, any two)*.

B2 *(3)* — It plots each residual (what the model missed) against its fitted value *(1)*. When the model is adequate, the points scatter at random around zero, with the same spread everywhere *(1)*. Problems: a **curve** (the relationship is not linear — try a squared term or a transformation) and a **fan** (the spread grows with the fitted values) *(1)*.

B3 *(3)* — When a variable that affects y and is correlated with x is left out, the coefficient on x absorbs part of its effect *(1)*. Example: regressing firms' exports on productivity without management quality *(1)*. Good management raises both exports and productivity, so the productivity coefficient is pushed **upward** — it takes credit for management *(1)*. (The direction is the sign of the omitted effect times the sign of its correlation with x.)

B4 *(3)* — The cutpoints are the thresholds that separate the ordered categories, one between each pair of neighbouring categories; they play the role of intercepts *(1.5)*. Proportional odds: each predictor has **one** coefficient, the same at every cut — a predictor shifts the whole distribution up or down the scale by the same amount in log-odds *(1.5)*.

B5 *(3)* — Leakage is when information from the held-out fold reaches the model during training *(1)*. Example: standardising, imputing or selecting variables on **all** the rows before splitting into folds; or shuffling time-series data so the model trains on the future *(1)*. It makes the cross-validation error **optimistic**: the model looks better than it will on genuinely new data *(1)*.

B6 *(3)* — To let the effect of x change with its level — to accelerate, level off or reverse — when the residual plot shows a curve *(1)*. With $y = \beta_0 + \beta_1 x + \beta_2 x^2$, the effect of one more unit of x is $\beta_1 + 2\beta_2 x$: it depends on x, so β_1 alone must not be read *(1)*. The turning point is the value of x where that effect is zero, $x^\star = -\beta_1/(2\beta_2)$, where the curve peaks (if $\beta_2 < 0$) or bottoms out *(1)*.

Part C — Reading output *(15)*

C1 *(5)*

C1(a) *(2)* — "Comparing firms of the same size, a one-point higher productivity index is associated with an export share 0.29 percentage points higher." Units *(1)*; "holding size fixed" and "associated with" *(1)*.

C1(b) *(1.5)* — Larger firms are both more productive and more export-oriented. In Model 1, productivity took part of the credit for size (omitted-variable bias); Model 2 compares firms of the same size, so the coefficient answers a different, narrower question.

C1(c) *(1.5)* — No. It is the predicted export share of a firm with productivity 0 and log employees 0 (one employee) — far outside the data, and a negative share is impossible. It anchors the line; it is not a finding.

C2 *(5)*

C2(a) *(2)* — "A large firm has three times the **odds** of adopting AI of a small firm with the same R&D status."

C2(b) *(2)* — An odds ratio multiplies the **odds**, not the probability *(1)*. The same multiplier moves the probability more where it is closer to 0.5: from 0.25 a large firm has more room to move than a small firm starting from 0.10 *(1)*.

C2(c) *(1)* — No. Without R&D, 0.25 against 0.10 is 2.5 times as likely; "three times" is the **odds** ratio, read as if it were a ratio of probabilities.

C3 *(5)*

C3(a) *(2)* — $\lambda = 0.1$: it has the lowest **cross-validation** error (1.05), the error on data the model did not see *(1)*. The training error must not be used to choose: it always prefers the least penalised model *(1)*.

C3(b) *(2)* — A larger penalty restrains the model, so it fits the training rows less closely: training error rises *(1)*. On new data, a small λ overfits (low bias, high variance) and a large one underfits (high bias): the cross-validation error is lowest in between — the bias–variance trade-off *(1)*.

C3(c) *(1)* — With no predictor kept, it predicts the **mean** R&D intensity for every country.

Part D — Judgement *(12)*

- **What the numbers are (3):** an AUC of 0.93 says that, on past data, the model ranks risky markets above safe ones well. The feature ranking says which indicators the model leans on for those predictions.
- **What they do not license (3):** AUC is not calibration — it does not say the predicted probabilities are right, nor which threshold to use; it says nothing about costs or about how the model performs on future markets. And an important **predictive** feature is not a **cause**: political instability may stand in for income, conflict or currency risk, and the model was never designed to identify causes.
- **Evidence needed (3):** a calibration check and a validation on later years (out-of-time); a threshold set from the costs of a missed failure and of a refused contract; for any causal claim, a design that isolates the effect of instability — not a feature ranking.
- **Meanwhile (3):** use the model as a **screening tool** with human review, not an automatic ban; set the threshold from the costs; report the predictive performance without "causes"; and monitor how the flagged markets actually perform.

Full marks for a well-argued refusal that covers the four elements.

Score yourself by part, not in total

The parts diagnose different failures, and the total hides which one is yours. Add up each part separately and read the row that applies:

Weak in	What it means	What to do next
I or II	You read papers for their results, not their design	Take any paper and write its question, model, two critiques and an alternative in ten lines
A or B	You are working from impressions, not definitions	Re-do the pre-session self-checks in writing
C	You accept output as given	Interpret one of your own practice outputs in words, sentence by sentence
D	You over-claim	Re-read your Session 1 memo. This is the row that matters most for the final paper

The paper: the sample paper · more practice: the practice questions and its answer key · the midterm summary