# Final exam for GRA4153

2025

This exam is to be completed individually. The page limit is 15 pages; this is an exaggerated upper limit and not a suggestion. You may freely refer to results from the lecture notes, slides and exercises (including solutions). You may also use other resources (e.g. books, internet resources), but – similarly – you *must* cite them if you do so.[1] All code must be in Python.

Each question is worth 1/3 of the total grade. Questions 1 & 2 are equally weighted. The individual parts of Question 3 display their relative weight (e.g. part (i) is worth 5% of the total grade for Question 3). Partial credit is always given if a partially correct answer is provided. Please provide a typed set of solutions as a pdf.[2] Your solutions must be submitted online by 12:00 on 5 December 2025.

## Question 1

Consider the linear regression model in matrix form:

$$y = X\beta + \epsilon,$$

where we assume that $X$ has full column rank and

$$\mathbb{E}[\epsilon|X] = 0, \qquad \mathbb{E}[\epsilon\epsilon'|X] = \sigma^2 I.$$

Often the data we observe is measured with error; in this question you will explore the effect of this on the OLS estimator.

(i) Suppose that instead of $(y, X)$ you observe $(y^*, X)$ where $y^* = y + \nu$ for a random vector $\nu$ independent of $(X, \epsilon)$ with $\mathbb{E}[\nu] = 0$ and $\text{Var}(\nu) = \varsigma^2 I$. Is the OLS estimator $\hat{\beta}^* = (X'X)^{-1}X'y^*$ unbiased? What is the variance of $\hat{\beta}^*$?

(ii) Suppose that instead of $\beta$, $y = X\beta^* + \epsilon$, where $\beta^* = \beta + \nu$, with $\nu$ a random vector independent of $(X, \epsilon)$ and $\mathbb{E}\nu = 0$, $\text{Var}(\nu) = \varsigma^2 I$. Calculate $\mathbb{E}[y|X]$ and $\text{Var}(y|X)$.

Now we will consider some large sample (asymptotic) results. Instead of the assumptions above now assume that $(y_i, X_i)$ for $i = 1, \ldots, n$ are i.i.d. and such that

$$\mathbb{E}\|X_i\epsilon_i\| < \infty \quad \text{with} \quad \mathbb{E}[X_i\epsilon_i] = 0 \quad \text{for each } i = 1, \ldots, n,$$

[1] Of course, if you are asked to "derive", "show", "demonstrate" or "prove" something (or similar) you have to do so.

[2] How you produce the document is up to you. One good option is LaTeX; overleaf is a collaborative web based LaTeXeditor. Another option is to use a jupyter notebook and use markdown cells to write your solutions; a markdown cheat sheet is here and a tutorial about how to write mathematics in markdown is here.

$$Q := \mathbb{E}X_1X_1' \text{ and } \Sigma := \mathrm{Var}(X_i\epsilon_i) \text{ exist and are positive definite.}$$

Suppose that instead of $(y_i, X_i)$ you observe $(y_i^*, X_i)$ where $y_i^* = y_i + \nu_i$ for a random variable $\nu_i$ where $(\nu_i)_{i\in\mathbb{N}}$ is i.i.d. and independent of $(X_i, \epsilon_i)$ with $\mathbb{E}[\nu_i] = 0$ and $\mathrm{Var}(\nu_i) = \varsigma^2$. Let $\hat{\beta}^*$ be the OLS estimator based on $(y_i^*, X_i)$.

(iii) Is $\hat{\beta}^*$ consistent for $\beta$?

(iv) Derive the asymptotic distribution of $\hat{\beta}^*$. Comment on the result.

## Question 2

This question reviews the concept of stationarity of a time series.

(i) Let $(X_t)_{t\in\mathbb{Z}}$ be given by $X_t = \epsilon_t\epsilon_{t-1}$ where $\epsilon_t \overset{iid}{\sim} \mathrm{WN}(0, \sigma^2)$. Is $(X_t)_{t\in\mathbb{Z}}$ covariance stationary? Is it white noise?

(ii) Let $(X_t)_{t\in\mathbb{Z}}$ be given by $X_t = Y\cos(at) + Z\sin(at)$ where $Y, Z$ are uncorrelated random variables with mean zero and variance $\sigma^2$ and $a \in \mathbb{R}$ is non-random. Is $(X_t)_{t\in\mathbb{Z}}$ covariance stationary? Is it white noise?[3]

(iii) Let $(X_t)_{t\in\mathbb{Z}}$ be given by $X_t = Z_t\cos(at) + Z_{t-1}\sin(at)$ where $(Z_t)_{t\in\mathbb{Z}}$ is a covariance stationary process and $a \in \mathbb{R}$ is non-random. Is $(X_t)_{t\in\mathbb{Z}}$ covariance stationary? Is is white noise?

## Question 3

In this exercise you will analyse a time series of industrial production and forecast its future values. The data set is `INDPRO.csv` and is available on wiseflow.

(i) (5%) Select the sample from 1970-01-01 to 2015-12-01 and plot the sample.

(ii) (15%) Test for a unit root (i.e. nonstationarity) using `sm.tsa.stattools.adfuller`; what is your conclusion?

(iii) (5%) Based on the result of the test in part (ii), you may or may not want to transform the series in some way. Let $X_t$ be the series following any (or no) transform you perform.

(iv) (15%) Plot the sample ACF and PACF for $X_t$. Can you conclude anything about an appropriate ARMA model order?

(v) (20%) Fit ARMA$(p, q)$ models for $p$, $q$ up to 4 to $X_t$ and select an appropriate model order using AICC and BIC. What order do you find? Do the two criteria lead to the same conclusion? If not, pick a model and justify your choice.

(vi) (20%) Estimate the model you picked in the preceding section. Interpret the results.

(vii) (20%) Forecast $X_t$ 4 periods ahead and plot $X_t$ and your forecast.

[3] Hint: $\cos(a)\cos(b) + \sin(a)\sin(b) = \cos(a - b)$.