

Sample Exam for ECON434 Spring 2025

Master of Quantitative Economics

Time Limit: 2 hours

This is the sample exam for the ECON 434 Spring 2025 final. Students are expected to complete all problems within 2 hours. The actual final exam will follow a similar format. It will include at least one reasoning question, one Python coding question, and several theoretical questions. The number of questions for each type may vary.

1 Reasoning Question

1.1 Basic Concept

Please explain the definition of the important term in Machine Learning: *Sparcity*

1.2 ECON434 Final Exam, 06/09/2024, Question 2

A big problem in the US is health of US citizens. To deal with this problem, some politicians argue that we need to provide universal health insurance like many other countries do. However, it is actually questionable whether providing health insurance improves health, and the costs of insurance may outweigh the benefits. Suppose that to study this issue, a researcher collects a data on a random sample of US citizens, some of which are insured and others are not. Suppose that the researcher then runs a regression of a particular health outcome (say blood pressure) on the dummy for insurance (including the constant, of course). Provide at least two mechanisms why the result may be biased (one mechanism should lead to over-estimation of the insurance health benefit, and the other one to under-estimation). Explain formally the direction of the bias for each mechanism using the selection bias formula. (You will get some extra points if you provide more than two mechanisms.)

2 Coding Question

In lecture 12, we discussed the Random Forest Method and how to apply this method via python code. In the lecture slides, there is a line of code writing:

```

rf = RandomForestRegressor(n_estimator = 100,
                           max_depth = 5,
                           random_state = 42)

rf.fit(X,Y)

```

Please explain the meaning of the following two arguments in the function:

```
n_estimator = 500 & max_depth = 5
```

3 Theoretical Questions in Machine Learning

3.1 ECON434 Final Exam, 06/09/2024, Question 1

Suppose that we have a pair of random variables, X and Y , where X can take three values: -1, 0, and 1, all with equal probabilities. Suppose that we run an OLS regression of Y on X (including the constant). Derive the probability limit of the slope coefficient of such a regression in terms of the conditional mean function $E[Y|X = x], x = -1, 0, 1$.

3.2 ECON434 Final Exam, 06/09/2024, Question 3

Consider random variables $Y \in \mathbb{R}$, $X \in \mathbb{R}$, and $Z \in \mathbb{R}$. Suppose we are interested in estimating

$$\theta = \mathbb{E}[Zh(X)^4] \quad (1)$$

where

$$h(x) = \mathbb{E}[Y|X = x], \quad x \in \mathbb{R}.$$

(a). Show that θ in (1) can be alternatively written as

$$\theta = \mathbb{E}[Zh(X)^4 + 4(Y - h(X))h^3(X)p(X)], \quad (2)$$

where

$$p(x) = \mathbb{E}[Z|X = x], \quad x \in \mathbb{R}$$

(b). Show that the estimating equation (2) satisfies the Neyman orthogonality condition with respect to both $h(x)$ and $p(x)$.

(c). Show that the original estimating equation (1) does not satisfy the Neyman orthogonality condition.