MBAN 5110: Midterm Exam

Deadline for handing in the exam: Sunday Nov 12, 2023 at 11:59 pm.
Please take the following into consideration:

1. **This is an individual take-home midterm exam!**
2. Use Python to answer the questions. You must submit your Python codes in two formats (i.e., ipynb and html formats). Additionally, you must prepare a summary report highlighting main results and key takeaways.
3. Summary report should be in the word format:
   (with the name: "MBAN_midterm_summary.doc").
4. Accordingly, name the Python files as "MBAN_midterm".
5. Create a ZIP-file containing the Python files and the summary report. Put your student number in the name of the ZIP-file. For example, the name of the ZIP-file becomes "123456.zip".
6. Submit the ZIP-file online on the Canvas

PART 1 (20%)

Use "midterm_partone.csv" file that contains the stock-return information of small retailers (the same as the one we used in Session 5). Suppose that an industry expert (e.g., David Berman) claims that there is a bias in moment conditions of instrumental variables such that

$Z^T(Y - XB) = \delta \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$, where δ has a non-zero value.

1. Update the GMM model that we discussed in class by incorporating the δ term to the instrumental-variable moment expressions.
2. By analyzing the GMM summary table and test statistics of coefficients, determine if the industry expert's claim is statistically justified.

PART 2 (80%)

A global bank that started to operate in Turkey in 2000s decided to expand its presence in small cities and even villages. The executives spent long times to develop an expansion strategy by focusing on how to attract customers in small cities, rather than big cities such as Istanbul, because they expected less competition in those regions. The main challenge for such an

expansion was that many people living in small cities did not have even a bank account. According to a World Bank report, for example, 1.7 billion adults in the world do not have any account in a financial institution.

In small cities, people do not need a bank account for two reasons. First, online payments can be made through GSM operators by sending an SMS, which is popular in Turkey (https://www.ft.com/content/ce9227b0-129e-433a-b9fc-0744b163451b). Owing to the SMS payment possibility, residents of a small Turkish city do not need a credit card nor a bank account. The second reason is that credit applications in small cities are processed slowly in comparison to big cities because the documents taken from a customer should be sent by post to main branches in big cities where risk management experts approve or decline the applications. Being aware of long processing times, the residents usually borrow money from their families or friends, instead of going to a bank.

To attract new customers in small cities, the bank decided to automate the assessment of credit applications. A team of experts by the bank used the credit score data and tested some approaches. They helped the bank shorten the processing time for some group of customers. But the involvement of risk management team is unavoidable to assess the application of some high-risk customers. Therefore, the bank attracted new customers for some segments, while many customers still stay unbanked. Use the dataset "midterm_parttwo.csv" to answer the questions. The dataset has 8081 rows with the following columns:

Column Name	**Description**
Credit Rating	It is a binary variable that takes "Positive" if the credit is fully repaid by the customer or "Negative" otherwise.
Requested credit amount	It is a categorical variable that can take three values: "High", "Medium", and "Low".

Marital status	It is a categorical variable that can take three values: “Married”, “Single”, and “Not specified”
Number of Dependents	It is a categorical variable that can take three values: “No dependent”, “Less than two”, and “More than two”
Years of education after high school	This variable can take integer values between 0 to 7.
Monthly income	It is a categorical variable that can take five values: “Very Low”, “Low”, “Moderate”, “High”, “Very High”
Monthly expense	It is a categorical variable that can take five values: “Very Low”, “Low”, “Moderate”, “High”, “Very High”

1. Divide the dataset equally into two as training (50%) and test (50%) sets. Use the training set to fit a logistic regression model, where the credit rating is the dependent variable. Apply the model to the test set, and report the confusion matrix, recall, precision, and F1 score values.
2. Suppose that the bank decided to make the credit approval process more challenging such that only 15% of the applications would be granted. Calculate the threshold value for the prediction probability, so only 15% of the test set would get their applications approved. Then, update your confusion matrix, recall, precision, and F1 scores.