Revision guide January 2024 exam **SMM248 Data Analysis for Finance**

|  |
| --- |
| **List of Key topics** |
| 1. Understand the concept of a *random* variable. |
| 1. Understand *expected value* and *variance* of a distribution: write down the formulae to estimate them. |
| 1. Understand the concepts of *covariance* and *correlation*: how do they relate? write down the formulae to estimate them. |
| 1. Understand the concepts of *skewness* (S) and *kurtosis* (K) of a distribution |
| 1. What is the skewness (S) and the kurtosis (K) value of any Normally distributed variable? |
| 1. Understand the concepts of *critical value, p-value,* and *significance level* of a test |
| 1. Summarize the key elements of a test in the pertinent probability distribution graph (N(0,1), Student’s t, F distribution, χ2 distribution): test statistic value, critical value(s), p-value |
| 1. Understand the concept of model *predictions* or *fitted values*: how to calculate them. |
| 1. Understand the concept of regression *error*: what does it capture? |
| 1. Understand the concept of *residuals* and *estimated errors*: how to calculate them |
| 1. Understand the properties of a good estimators: unbiasedness, consistency, efficiency. |
| 1. Understand the first OLS assumption E(ε)=0. |
| 1. Understand why the omission of the intercept in a regression model can be problematic. |
| 1. Understand the mechanics of a t-test: two-sided/one-sided, which probability distribution to use (Student’s t or N(0,1)) and when? |
| 1. *Standard error (S.E.) of a parameter estimate* and *S.E. of regression*: what do they measure. |
| 1. Write down the t-statistic expression/how the t-statistic value is calculated. |
| 1. Interpret a *confidence interval* for a parameter: how to use it for hypothesis testing. |
| 1. Conduct a test for a single restriction that is a combination of several parameters (t-test). |
| 1. Conduct a test for the significance of each variable (t-test): write down Ho and Ha. |
| 1. Conduct a test for any set of multiple restrictions (F-test). |
| 1. Understand the concepts of RSS, TSS and ESS and the relationship between them. |
| 1. Coefficient of determination (R2): write down the expression as a correlation measure and as a function of the RSS and TSS. Discuss in which situations the R2 should not be used and why |
| 1. Write down the F-statistic expression/how it is calculated. |
| 1. Understand the mechanics of an F-test: which probability distribution to use. |
| 1. Understand the concepts of Unrestricted Model and Restricted Model |
| 1. Conduct a test for the overall significance of the model (F-test): write down Ho and Ha in terms of the model parameters, and alternatively in terms of the R2. |
| 1. What are the LR (likelihood ratio) statistic, Wald test, and LM (Lagrange multiplier) statistics used for: which probability distribution are they associated with. |
| 1. Adjusted coefficient of determination (adjusted R2): when should we use it? Calculate the adjusted R2 from the R2. |
| 1. Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC): what for? Similarities and differences between these criteria and the R2 and adjusted-R2. |
| 1. How to calculate the marginal effect of a variable (x) on another variable (y) using a linear regression model: what is the difference between *sensitivity* of y to x, and *elasticity*? |
| 1. Understand the problem of omitted variable bias |
| 1. How to calculate the marginal effect of a variable (x) on another variable (y) using a nonlinear-in-variables model: quadratic, reciprocal, exponential or log-linear, log-log, linear-log, log inverse. (Note: a table with these marginal effects will be provided in the exam). |
| 1. What is an interaction variable? Why is it used as an explanatory variable in a regression? |
| 1. Understand why ‘no perfect multicollinearity’ is an assumption in multiple regression. |
| 1. Understand the difference between multicollinearity and perfect multicollinearity |
| 1. Why is multicollinearity a problem in regression analysis. |
| 1. How to detect multicollinearity: correlation matrix, variance inflation factors (VIFs) |
| 1. Understand the mechanics of the Ramsey RESET test for model misspecification. |
| 1. Describe the full set of assumptions on which an OLS regression model is built |
| 1. What are the implications of violation of the no-autocorrelated errors assumption? |
| 1. What are the implications of violation of the homoscedastic errors assumption? |
| 1. Describe the problem of autocorrelation intuitively and with mathematical notation. In which type of regressions is residual autocorrelation likely? |
| 1. Describe the problem of heteroskedasticity intuitively and with mathematical notation. In which type of regressions is residual heteroskedasticity likely? |
| 1. How can we make our regression-based inferences robust to autocorrelated errors? |
| 1. How can we make our regression-based inferences robust to heteroskedastic errors? |
| 1. Understand the mechanics of the Breusch-Godfrey LM test for autocorrelation (Ho and Ha) |
| 1. Understand the mechanics of the Ljung-Box Q for autocorrelation (Ho and Ha) |
| 1. Understand the mechanics of the White LMS test for heteroskedasticity (Ho and Ha) |
| 1. Interpretation of a CS regression model with a dummy variable (0-1) as a regressor: which additional effect can be tested? |
| 1. Interpretation of a CS regression model with a dummy variable (0-1) interacted with a regressor: which additional effect can be tested? |
| 1. Use dummy variables to test for parameter stability in a time-series regression |
| 1. What is a linear probability model? How are the predictions from a logit model calculated? How are the predictions from a logit model interpreted? |
| 1. What is a logit model? How are the predictions from a logit model calculated? How are the predictions from a logit model interpreted? |
| 1. How are the marginal effects of continuous X variables on the probability response (logit model) calculated? |
| 1. Discuss tools to evaluate a logit model: Hit rate, Misclassification rate, Confusion matrix |
| 1. What are the shortcoming of the linear probability model (versus the logit model) |
| 1. Which estimation approach is used for the linear probability model, and the logit model? |
| 1. What are the advantages of panel data versus time-series and cross-section data? (in which real-life scenarios are panel models particularly useful?) |
| 1. What is a Pooled OLS model? Write down the equation with concise notation. How many parameters are estimated? |
| 1. What is a Between OLS model? Write down the equation with concise notation. How many parameters are estimated? |
| 1. What is a 1-way Fixed Effects model (with country effects)? Write down the equation with concise notation. How many parameters are estimated? |
| 1. What is a 2-way Fixed Effects model (with country and time effects)? Write down the equation with concise notation. How many parameters are estimated? |
| 1. What are the assumptions behind the pooled OLS and FE regression models? |
| 1. How can we protect the inferences from our panel regression models against violations of those assumptions? |
| Remember that … |
| * The presentation slides (lectures) serve as a good summary of each topic * Reading through the LAB EXERCISES (follow the discussion) is useful * Reading the weekly TASKS (see answer in Moodle section TASKS TO SUBMIT) is useful * Practicing the end-of-chapter in Lecturer’s book exercises (answers in Moodle section), and prior exams (in Moodle) is useful |

Enjoy this opportunity (revision) to strengthen your data analysis skills! Best wishes. Ana-Maria