

This syllabus is intended to provide an explicit list of all the mathematical formulae and proofs that you are expected to know for the examination at the end of the year. You are warned that a knowledge of the formulae and proofs is taken for granted in the examination and that relatively little credit is given for it in the marking scheme. The examination is intended to be an opportunity for you to display your understanding of the material, rather than to reproduce standard items.

### **Review: Random variables and sampling theory**

Probability distribution of a random variable. Expected value of a random variable. Expected value of a function of a random variable. Population variance of a discrete random variable and alternative expression for it. Expected value rules. Independence of two random variables. Population covariance, covariance and variance rules, and correlation. Sampling and estimators. Unbiasedness. Efficiency. Loss functions and mean square error. Estimators of variance, covariance and correlation. The normal distribution. Hypothesis testing. Type II error and the power of a test.  $t$  tests. Confidence intervals. One-sided tests. Convergence in probability and plim rules. Consistency. Convergence in distribution (asymptotic limiting distributions) and the role of central limit theorems.

*Formulae and proofs:* This chapter is concerned with statistics, not econometrics, and is not examinable. However you are expected to know the results in this chapter and to be able to use them.

### **Chapter 1 Simple regression analysis**

Simple regression model. Derivation of linear regression coefficients. Interpretation of a regression equation. Goodness of fit.

*Formulae and proofs:* You are expected to know, and be able to derive, the expressions for the regression coefficients in a simple regression model, including variations where either the intercept or the slope coefficient may be assumed to be zero. You are expected to know the definition of  $R^2$  and how it is related to the residual sum of squares. You are expected to know the relationship between  $R^2$  and the correlation between the actual and fitted values of the dependent variable, but not to be able to prove it.

### **Chapter 2 Properties of the regression coefficients**

Types of data and regression model. Assumptions for Model A. Regression coefficients as random variables. Unbiasedness of the regression coefficients. Precision of the regression coefficients. Gauss–Markov theorem.  $t$  test of a hypothesis relating to a regression coefficient. Type I error and Type II error. Confidence intervals. One-sided tests.  $F$  test of goodness of fit.

*Formulae and proofs:* You are expected to know the regression model assumptions for Model A. You are expected to know, though not be able to prove, that, in the case of a simple regression model, an  $F$  test on the goodness of fit is equivalent to a two-sided  $t$  test on the slope coefficient. You are expected to know how to make a theoretical decomposition of an estimator and hence how to investigate whether or not it is biased. In particular, you are expected to be able to show that the OLS estimator of the slope coefficient in a simple regression model can be decomposed into the true value plus a weighted linear combination of the values of the disturbance term in the sample. You are expected to be able to derive the expression for the variance of the slope coefficient in a simple regression model. You are expected to know how to estimate the variance of the disturbance term, given the residuals, but you are not expected to be able to derive the

expression. You are expected to understand the Gauss–Markov theorem, but you are not expected to be able to prove it.

### Chapter 3 Multiple regression analysis

Multiple regression with two explanatory variables. Graphical representation of a relationship in a multiple regression model. Properties of the multiple regression coefficients. Population variance of the regression coefficients. Decomposition of their standard errors. Multicollinearity.  $F$  tests in a multiple regression model. Hedonic pricing models. Prediction.

*Formulae and proofs:* You are expected to know how, in principle, the multiple regression coefficients are derived, but you do not have to remember the expressions, nor do you have to be able to derive them mathematically. You are expected to know, but not to be able to derive, the expressions for the population variance of a slope coefficient and its standard error in a model with two explanatory variables. You are expected to be able to perform  $F$  tests on the goodness of fit of the model as a whole and for the improvement in fit when a group of explanatory variables is added to the model. You are expected to be able to demonstrate the properties of predictions within the context of the classical linear regression model. In particular, you are expected to be able to demonstrate that the expected value of the prediction error is 0, if the model is correctly specified and the regression model assumptions are satisfied. You are not expected to know the population variance of the prediction error.

### Chapter 4 Transformation of variables

Linearity and nonlinearity. Elasticities and double-logarithmic models. Semilogarithmic models. The disturbance term in nonlinear models. Box–Cox transformation. Models with quadratic and interactive variables. Nonlinear regression.

*Formulae and proofs:* You are expected to know how to perform a Box–Cox transformation for comparing the goodness of fit of alternative versions of a model with  $Y$  and  $\log Y$  as the dependent variable. You are expected to be able to provide interpretations of the coefficients in models containing quadratic or interactive terms.

### Chapter 5 Dummy variables

Dummy variables. Dummy classification with more than two categories. The effects of changing the reference category. The dummy variable trap. Multiple sets of dummy variables. Slope dummy variables. Chow test. Relationship between Chow test and dummy group test.

*Formulae and proofs:* You are expected to be able to provide interpretations of the coefficients in models containing slope dummy variables. You are expected to be able to perform a Chow test and a test of the explanatory power of a group of dummy variables, and to understand the relationship between them.

### Chapter 6 Specification of regression variables

Omitted variable bias. Consequences of the inclusion of an irrelevant variable. Proxy variables.  $F$  test of a linear restriction. Reparameterization of a regression model.  $t$  test of a restriction. Tests of multiple restrictions. Tests of zero restrictions.

*Formulae and proofs:* You are expected to be able to derive the expression for omitted variable bias when the true model has two explanatory variables and the fitted model omits one of them. You are expected to know how to perform an  $F$  test on the validity of a linear restriction, given appropriate data on the residual sum of squares. You are expected to understand the logic behind the  $t$  test of a linear restriction and to be able to reparameterize a regression specification to perform such a test in a simple context. You are expected to be able to perform  $F$  tests of multiple linear restrictions.

### Chapter 7 Heteroscedasticity

Meaning of heteroscedasticity. Consequences of heteroscedasticity. Goldfeld–Quandt and White tests for heteroscedasticity. Elimination of heteroscedasticity using weighted or logarithmic regressions. Use of heteroscedasticity-consistent standard errors.

*Formulae and proofs:* You are expected to know how to perform the Goldfeld–Quandt and White tests for heteroscedasticity.

### Chapter 8 Stochastic regressors and measurement errors

Stochastic regressors. Assumptions for models with stochastic regressors. Finite sample and asymptotic properties of the regression coefficients in models with stochastic regressors. Measurement error and its consequences. Friedman's Permanent Income Hypothesis. Instrumental variables (IV). Asymptotic properties of IV estimators, including the asymptotic limiting distribution of  $\sqrt{n}(b_2^{IV} - \beta_2)$  where  $b_2^{IV}$  is the IV estimator of  $\beta_2$  in a simple regression model. Use of simulation to investigate the finite-sample properties of estimators when only asymptotic properties can be determined analytically. Application of the Durbin–Wu–Hausman test

*Formulae and proofs:* You are expected to be able to demonstrate that, in a simple regression model, the OLS estimator of the slope coefficient is inconsistent when there is measurement error in the explanatory variable. You should know the expression for the bias and be able to derive it. You should be able to explain the consequences of measurement error in the dependent variable. You should know the expression for an instrumental variable estimator of the slope coefficient in a simple regression model and be able to demonstrate that it yields consistent estimates, provided that certain assumptions are satisfied. You should also know the expression for the asymptotic population variance of an instrumental variable estimator in a simple regression model and to understand why it provides only an approximation for finite samples. You are not expected to know the formula for the Durbin–Wu–Hausman test.

### Chapter 9 Simultaneous equations estimation

Definitions of endogenous variables, exogenous variables, structural equations and reduced form. Inconsistency of OLS. Use of instrumental variables. Exact identification, underidentification, and overidentification. Two-stage least squares (TSLS). Order condition for identification. Application of the Durbin–Wu–Hausman test.

*Formulae and proofs:* You are expected to be able to derive an expression for simultaneous equations bias in a simple regression equation and to be able to demonstrate the consistency of an IV estimator in a simple regression equation. You are expected to be able to explain in general terms why TSLS is used in overidentified models.

### Chapter 10 Binary choice models and maximum likelihood estimation

Linear probability model. Logit model. Probit model. Maximum likelihood estimation of the population mean and variance of a random variable. Maximum likelihood estimation of regression coefficients. Likelihood ratio tests. The following sections are not part of the syllabus: Tobit model. Selection bias model.

*Formulae and proofs:* You are expected to know the expression for the probability of an event occurring in the logit model, and to know the expressions for the marginal functions in the logit and probit models. You would not be expected to calculate marginal effects in an examination, but you should be able to explain how they are calculated and to comment on calculations of them. You are expected to be able to

derive a maximum likelihood estimator in a simple example. In more complex examples, you would only be expected to explain how the estimates are obtained, in principle. You are expected to be able to perform, from first principles, likelihood ratio tests in a simple context.

### **Chapter 11 Models using time series data**

Static demand functions fitted using aggregate time series data. Lagged variables and naive attempts to model dynamics. Autoregressive distributed lag (ADL) models with applications in the form of the partial adjustment and adaptive expectations models. Error correction models. Asymptotic properties of OLS estimators of ADL models, including asymptotic limiting distributions. Use of simulation to investigate the finite sample properties of parameter estimators for the ADL(1,0) model. Use of predetermined variables as instruments in simultaneous equations models using time series data. (Section 11.7 of the text, *Alternative dynamic representations ...*, is not in the syllabus)

*Formulae and proofs:* You are expected to be able to analyze the short-run and long-run dynamics inherent in ADL(1,0) models in general and the adaptive expectations and partial adjustment models in particular. You are expected to be able to explain why the OLS estimators of the parameters of ADL(1,0) models are subject to finite-sample bias and, within the context of the model  $Y_t = \beta_1 + \beta_2 Y_{t-1} + u_t$  to be able to demonstrate that they are consistent.

### **Chapter 12 Autocorrelation**

Assumptions for regressions with time series data. Assumption of the independence of the disturbance term and the regressors. Definition of autocorrelation. Consequences of autocorrelation. Breusch–Godfrey lagrange multiplier, Durbin–Watson  $d$ , and Durbin  $h$  tests for autocorrelation. AR(1) nonlinear regression. Potential advantages and disadvantages of such estimation, in comparison with OLS. Cochrane–Orcutt iterative process. Autocorrelation with a lagged dependent variable. Common factor test and implications for model selection. Apparent autocorrelation caused by variable or functional misspecification. General-to-specific versus specific-to-general model specification.

*Formulae and proofs:* You are expected to know how to perform the tests for autocorrelation mentioned above and to know how to perform a common factor test. You are expected to be able to explain why the properties of estimators obtained by fitting the AR(1) nonlinear regression specification are not necessarily superior to those obtained using OLS.

### **Chapter 13 Introduction to nonstationary processes**

Stationary and nonstationary processes. Granger–Newbold experiments with random walks. Unit root tests. Akaike Information Criterion and Schwarz’s Bayes Information Criterion. Cointegration. Error correction models.

*Formulae and proofs:* You are expected to be able to determine whether a simple random process is stationary or nonstationary. You would not be expected to perform a unit root test in an examination, but you are expected to understand the test and to be able to comment on the results of such a test.

### **Chapter 14 Introduction to panel data models**

This chapter is not part of the syllabus.