## An Overview of the Data Sets

The data sets are intended to provide continuity in the practical work undertaken by students. There are three major data sets with a suite of exercises to be done by a group of students working in parallel but independently and obtaining differing results. In addition there are some minor data sets.

### Major data sets

The Educational Attainment and Earnings Functions cross-section data set, an extract from the US National Longitudinal Survey of Youth, is accompanied by a student manual containing 40 exercises fitting educational attainment functions and earnings functions, starting with simple regression models and gradually improving the specification as progress is made through the text (multiple regression, nonlinear models, dummy variables, etc), dealing with econometric problems as they are encountered, and eventually reaching such topics as binary choice models and sample selection bias. The data set actually consists of 20 subsets with the same variables but different observations so that a group of students performing the same exercises will obtain varying results.

The Demand Functions time-series data set, an extract from the US National Income and Product Accounts, is accompanied by a student manual containing 28 exercises fitting demand functions and earnings functions, again starting with simple regression and developing a progressively more sophisticated specification, ultimately reaching an error-correction model. This data set is primarily intended for use with the last three chapters of the text and hence the exercises are mostly addressed to the time-series topics in those chapters. The data set contains 19 categories of expenditure so that a group of students performing the same exercises will obtain varying results.

The Consumer Expenditure Survey cross-section data set, an extract from the US Consumer Expenditure Survey, is accompanied by a student manual containing 20 exercises fitting demand functions, again starting with simple regression and developing a progressively more sophisticated specification, ultimately reaching a tobit model. The data set contains 21 categories of expenditure so that a group of students performing the same exercises will obtain varying results.

### Minor data sets

The minor data sets are intended for individual, rather than group, use.

The Labor Force Participation data set provides exercises with binary choice and sample selection bias models. Like the Educational Attainment and Earnings Functions data sets, it is an extract from the US National Longitudinal Survey of Youth.

The School Costs data set is the data set that provides examples of the use of dummy variables in Chapter 6 of the text.

The OECD data set contains data for 26 OECD countries on the following variables, all defined as average annual growth rates over the period 1988-1997: wages, prices, GDP, employment, and money supply. In addition it contains the average rate of unemployment over the same time period. The data set is intended to provide a resource for investigating simple macroeconomic relationships, but no

specific exercises are proposed other than those that replicate exercises in the text (Exercises 2.1, 3.13, 3.20, and 5.5).

The UNIDO data set contains data on manufacturing output, GDP, and population for 28 countries in 1994. It is used in the text to provide examples of heteroscedasticity and methods of eliminating it. The exercises replicate the illustrative examples in the text.

The UNESCO data set contains data on recurrent public expenditure on education, GDP, and population for 38 countries in 1994. Like the UNIDO data set, it provides examples of heteroscedasticity. The exercises replicate the exercises in the text.