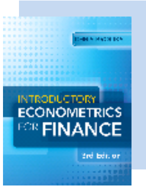

Introductory Econometrics for Finance
Chris Brooks
Solutions to Review Questions - Chapter 1

1. **(a)** Continuous data come from series that can take on any value (possibly within a given range) and can be measured to any arbitrary degree of precision such as the weight of a lump of cheese or the average return on a stock, but discrete data can only take certain specific values – for example, the number of houses in a street.

(b) Ordinal data arise where a variable is limited so that its values define a position or ordering only, and thus the precise values that the variable takes have no direct interpretation – for example, the performance ranking of a mutual fund among a set of 20 such funds. Nominal data by contrast occur when there is no natural ordering of the values at all, so a figure of 12 is simply different to that of a figure of 6, but could not be considered to be better or worse in any sense. Such data often arise when numerical values are arbitrarily assigned, such as telephone numbers or when codings are assigned to qualitative data (e.g. when describing the exchange that a US stock is traded on, '1' might be used to denote the NYSE, '2' to denote the NASDAQ and '3' to denote the AMEX).

(c) Time-series data are data that have been collected over a period of time on one or more variables – for example, a series of house prices, observed monthly for ten years in a particular region. Thus there is no cross-sectional element in this case so there is a single entity being examined – so one country, one stock, one firm, etc. Panel data, by contrast, simultaneously have the dimensions of both time series and cross-sections, e.g. the daily prices of a number of blue chip stocks over two years. Here we have both many time points and (days) many entities (firms) in the sample.

(d) The distinction between noisy and clean data is a subtle one. In general, 'noisy' refers to data that have a large amount of random variation which is considered an uninteresting feature that might get in the way of uncovering the underlying behaviour. The noise might simply be random variation in a series due to its volatility, or it might occur as a result of recording or measurement errors. To a large extent, almost all series that we encounter in economics and finance are noisy. Clean data refers to series where the amount of noise is at a minimal level and the data are at least free of errors.

(e) These are different philosophical approaches to how the data are used in building models. Classical statistics involves building a theoretical model first and then 'showing it to the data' with the parameter estimates being freely determined by the data. Bayesian statistics, on the other hand, involves the data and theory working more closely together. The researcher would start with an assessment of the existing state of knowledge or beliefs, formulated into a set of probabilities. These prior inputs or priors would then be combined with the observed data via a likelihood function. The beliefs and the probabilities would then be updated as a result of the model estimation, resulting in a set of posterior probabilities. Probabilities are thus updated sequentially, as more data become available.

2. There is certainly no shortage of possible examples that could be listed here. It is important to note that in many instances a specific problem could be tackled using time-series or cross-sectional or panel data, although it might be that one approach would be more insightful than the others. The following examples were used in the book:

Problems that could be tackled using time series data:

- How the value of a country's stock index has varied with that country's macroeconomic fundamentals.
- How the value of a company's stock price has varied when it announced the value of its dividend payment.
- The effect on a country's exchange rate of an increase in its trade deficit.

Problems that could be tackled using cross-sectional data:

- The relationship between company size and the return to investing in its shares.
- The relationship between a country's GDP level and the probability that the government will default on its sovereign debt.

Potentially, any of the above issues that could be considered in the time-series or cross-sectional frameworks could also be tackled in a panel context.

3. Asset return time-series have a number of stylised features that are common whether they are referring to stocks, bonds, house prices, etc. The key ones are:

- There is a lot of data available!
- They are noisy and volatile.
- They are leptokurtic and have fatter tails than a normal distribution with the same mean and variance.

- Most such series are negatively skewed, so that large negative returns are more likely than positive returns of the same magnitude.
- They exhibit volatility clustering, so there are bursts where the series is highly volatile for a protracted period and also quiet periods where there is nothing going on for a while.
- They can often best be characterised as a random walk with drift process.

4. These calculations are probably best done in a spreadsheet. If we did so, we would get the following table:

[redacted]

A few notes:

The simple returns are calculated as: $return_t = 100 \times P_t - P_{t-1} / P_{t-1}$
The continuously compounded returns are calculated as: $return_t = 100 \times \log(P_t / P_{t-1})$
The bond prices in 2013 terms are calculated as: $price_t = nominal\ price_t \times CPI_{2013} / CPI_t$
The inflation rate is calculated as $inflation\ rate_t = 100 \times CPI_t - CPI_{t-1} / CPI_{t-1}$
The real returns are calculated as: *nominal* (simple) *returns* (%) – *inflation rate* (%).