Econometrics Continued ...

The Sources of Data

The data used in empirical analysis may be collected by a governmental agency (e.g., the Department of Commerce), an international agency (e.g., the International Monetary Fund (IMF) or the World Bank), a private organization (e.g., the Standard & Poor's Corporation), or an individual. Literally, there are thousands of such agencies collecting data for one purpose or another.

The Internet:

The Internet has literally revolutionized data gathering. If you just "surf the net" with a keyword (e.g., exchange rates), you will be swamped with all kinds of data sources. There are some of the frequently visited web sites that provide economic and financial data of all sorts. Most of the data can be downloaded without much cost. We may bookmark the various web sites that might provide us with useful economic data.

The data collected by various agencies may be 'experimental' or 'non-experimental'. In experimental data, often collected in the natural sciences, the investigator may want to collect data while holding certain factors constant in order to assess the impact of some factors on a given phenomenon. For instance, in assessing the impact of obesity on blood pressure, the researcher would want to collect data while holding constant the eating, smoking, and drinking habits of the people in order to minimize the influence of these variables on blood pressure.

In the social sciences, the data that one generally encounters are non-experimental in nature, that is, not subject to the control of the researcher. For example, the data on GNP, unemployment, stock prices, etc., are not directly under the control of the investigator. As we shall see, this lack of control often creates special problems for the researcher in pinning down the exact cause or causes affecting a particular situation. For example, is it the money supply that determines the (nominal) GDP or is it the other way round?

The data collected by various agencies may be experimental or non-experimental. In experimental data, often collected in the natural sciences, the investigator may want to collect data while holding certain factors constant in order to assess the impact of some factors on a given phenomenon.

For instance, in assessing the impact of obesity on blood pressure, the researcher would want to collect data while holding constant the eating, smoking, and drinking habits of the people in order to minimize the influence of these variables on blood pressure.

In the social sciences, the data that one generally encounters are non-experimental in nature, that is, not subject to the control of the researcher. For example, the data on GNP, unemployment, stock prices, etc., are not directly under the control of the investigator. As we shall see, this lack of control often creates special problems for the researcher in pinning down the exact cause or causes affecting a particular situation. For example, is it the money supply that determines the (nominal) GDP or is it the other way round?

The Accuracy of Data:

Although plenty of data are available for economic research, the quality of the data is often not that good. There are several reasons for that.

1) First, as noted, most social science data are non-experimental in nature. Therefore, there is the possibility of observational errors, either of omission or commission.

2) Second, even in experimentally collected data errors of measurement arise from approximations and round offs.

3) Third, in questionnaire-type surveys, the problem of non-response can be serious; a researcher is lucky to get a 40 percent response to a questionnaire. Analysis based on such partial response may not truly reflect the behavior of the 60 percent who did not respond, thereby leading to what is known as (sample) selectivity bias. Then there is the further problem that those who respond to the questionnaire may not answer all the questions, especially questions of financially sensitive nature, thus leading to additional selectivity bias.

4) Fourth, the sampling methods used in obtaining the data may vary so widely that it is often difficult to compare the results obtained from the various samples.

5) Fifth, economic data are generally available at a highly aggregate level. For example, most macro-data (e.g., GNP, employment, inflation, unemployment) are available for the economy as a whole or at the most for some broad geographical regions. Such highly aggregated data may not tell us much about the individual or micro-units that may be the ultimate object of study.

6) Sixth, because of confidentiality, certain data can be published only in highly aggregate form. Therefore, if one wants to find out how much individuals with a certain level of income spent on health care, one cannot do that analysis except at a very highly aggregate level. But such macro-analysis often fails to reveal the dynamics of the behavior of the micro-units. Similarly, the Department of Commerce, which conducts the census of business every 5 years, is not allowed to disclose information on production, employment, energy consumption, research and development expenditure, etc., at the firm level. It is therefore difficult to study the inter-firm differences on these items.

Because of all these and many other problems, the researcher should always keep in mind that the results of research are only as good as the quality of the data. Therefore, if in given situations researchers find that the results of the research are "unsatisfactory," the cause may be not that they used the wrong model but that the quality of the data was poor. Unfortunately, because of the non-experimental nature of the data used in most social science studies, researchers very often have no choice but to depend on the available data. But they should always keep in mind that the data used may not be the best and should try not to be too dogmatic about the results obtained from a given study, especially when the quality of data is suspect.

Measurement Scales of Variables:
The variables that we will generally encounter fall into four broad categories: *ratio scale, interval scale, ordinal scale,* and *nominal scale.* It is important that we understand each.

Ratio Scale:
For a variable X, taking two values, X_1 and X_2, the ratio X_1/X_2 and the distance $(X_2 - X_1)$ are meaningful quantities. Also, there is a natural ordering (ascending or descending) of the values along the scale.
Therefore, comparisons such as $X_2 \leq X_1$ or $X_2 \geq X_1$ are meaningful. Most economic variables belong to this category. Thus, it is meaningful to ask how big is this year's GDP compared with the previous year's GDP.

Interval Scale:
An interval scale variable satisfies the last two properties of the ratio scale variable but not the first. Thus, the distance between two time periods, say (2000–1995) is meaningful, but not the ratio of two time periods (2000/1995).

Ordinal Scale: A variable belongs to this category only if it satisfies the third property of the ratio scale (i.e., natural ordering). Examples are grading systems (A, B, C grades) or income class (upper, middle, lower). For these variables the ordering exists but the distances between the categories cannot be quantified. Students of economics will recall the indifference curves between two goods, each higher indifference curve indicating higher level of utility, but one cannot quantify by how much one indifference curve is higher than the others.

Nominal Scale:
Variables in this category have none of the features of the ratio scale variables. Variables such as gender (male, female) and marital status (married, unmarried, divorced, separated) simply denote categories. Question: What is the reason why such variables cannot be expressed on the ratio, interval, or ordinal scales?

As we shall see, econometric techniques that may be suitable for ratio scale variables may not be suitable for nominal scale variables. Therefore, it is important to bear in mind the distinctions among the four types of measurement scales discussed above.

* To be continued ...

Tuesday, 29 June 2021