coefficients on labor and capital \( (a \text{ and } b) \) are interpreted as the marginal products of labor and capital. Suppose one wishes to analyze the effects on \( A \) of explicitly treating energy as a factor of production. Explain how to do that.

6. Determine an appropriate theory to apply to your research question. Use either a common economic theory or modify an existing theory from the literature on your topic. Explain why the theory is appropriate for your research question.

---

**Chapter 8**

Locating (and Collecting) Economic Data

"It is a capital mistake to theorize before one has data. Insensibly one begins to twist facts to suit theories, instead of theories to suit facts."

SIR ARTHUR CONAN DOYLE

This is the first of two chapters on locating and collecting economic data. In this chapter, we focus on how data are constructed and where they may be found, both in primary and secondary sources. In the next chapter, we explain how to compile a data set for your own research project.

A key part of any empirical research project is collecting and manipulating data. It is never too early to begin looking for potential data sources. In fact, it is not uncommon for novice researchers to invest a great deal of time and effort developing a research project only to discover that data aren’t available to adequately test the hypothesis. Don’t let this happen to you.

Even though this book describes the research process sequentially, in practice, many steps can be performed in parallel or at least in overlapping steps. Once you have settled on a research question, it is wise to begin looking for usable data. To begin, you should look for data in the general area your research (e.g., macro data, international trade, financial data, etc.). As you review the literature on your topic, pay attention to the data sources previous studies have used. Think about their strengths and weaknesses. Would they work for your research?
Once you have sketched out your conceptual or theoretical model, you should delve into your data search in earnest. What variables do you need data for in order to test your research hypothesis? Often times, researchers must simplify their theoretical model because the data are inadequate. Either there are insufficient data points for one or more variables, or a specific variable they need is not available. Before we get into the details of developing your own data set, to be discussed in Chapter 9, we first need to examine the process of data creation and the structure of published data.

**Data Creation**

In Chapter 1, we noted that many people think that knowledge consists of facts that can be plucked from a tree in the forest, once that tree is discovered. By contrast, we have argued that knowledge is created; this knowledge is the interpretation of facts.

In the same way, people tend to think of statistics as facts. This is misleading or wrong. The vast majority of data are constructed rather than collected. Government or private organizations that do a serious job constructing statistics make decisions throughout the process that affect the quality of that data for any given purpose.

**Steps in Data Construction**

Best (2001) identifies three steps in the construction of a data series:

1. Defining the concept;
2. Deciding how the concept will be measured; and
3. Determining how to define the sample on which the data will be based.

Let’s examine these steps to get a better sense of what published data really mean.

Suppose we are interested in finding income so we can test the theory of consumer behavior. What is the basic consuming unit? There are commonly used possibilities: households and families. If we choose families, our concept of income will be average family income. If we choose households, it will be average household income. If you compare these variables, they will be similar but not the same.

This illustrates an important point about data. Every data series is constructed for a specific purpose. You should realize that that purpose may not be the same as yours. As a consequence, a given data series may not be defined or measured in a way that best matches your needs. Unless you are going to construct your own data, you will have to live with this problem.

However, you can minimize this problem by doing your best to discover exactly where the data come from and what they actually measure. Suppose we’ve chosen to define our variable as average family income. How is that to be measured? What will be included? Should we include labor income only, for example, wages and salaries, or capital income, for instance, interest, dividends, capital gains, too? Should we include in-kind payments, such as food grown on farms and consumed by the farmers? It’s not sufficient to decide you want “total” income. You still need to define what is included.

How are we going to measure average family income? What measure of the average should we use? Unless the data are approximately normally distributed, which income data tend not to be, you can get substantially different values for the mean, median, and mode. Thus, for income data one tends to use the median rather than the mean as a measure of central tendency.

**Sample Data**

Most social science statistics are based on sample data rather than populations. Therefore, the data that are published are extrapolated from samples. For example, average family income is not, in fact, the average income of all families; rather, it is the average income of the families in the sample.

Only if the sample is random, and thus truly representative of the population, will the sample statistics correctly measure the population. A truly random sample is difficult and expensive to acquire, so typically samples are less than truly random. As a consequence, bias creeps into the results. Such, you should think of all data as estimates rather than facts.

Best (2001, 161) sums it up well: “Every statistic must be created, and the process of creation always involves choices that affect the resulting number. . . . a statistic is perfect, but some are less imperfect than others.”

**The Structure of Economic Data**

When we discussed surveying the literature on a research topic, we mentioned that knowing how the economics profession organizes its literature can help scholars find information quickly and efficiently. The same thing can be said for data.

It is important to differentiate between those organizations that collect or store data and those that publish it. This is the same distinction as primary versus secondary sources of information. Let’s talk first about the sources of data, and then we can figure out the best places for users to obtain that data.
Statistics are not collected at random. Rather, they are typically the result of a specific data collection effort or process. The product of this process is a specific data set that includes a collection of certain variables. The characteristics of data sets are described in the next section. The section that follows provides an overview of the various agencies that collect the data and the major data sets they produce. Then, we will explain the most important of those data sets in detail.

**Characteristics of Data Sets**

Survey data comes in three forms: time series, cross-section, and longitudinal (or panel) data. **Time-series data** gives different observations of data points on the same variable at different points across time. For example, a time series could consist of annual data for U.S. GDP over the time period 1950-2000. **Cross-section data**, by contrast, gives different observations of a comparable variable at the same point in time. For example, a researcher could obtain a cross section of data that consisted of average disposable personal income across the fifty states in the United States. Finally, **longitudinal data** take a cross-section sample and follow over time. For example, a sample of family income for the same ten families over five years would be a longitudinal data set. Note that all three of these examples have a sample size of fifty observations, but the basis of these observations is different. Longitudinal data are an example of a macro data set, meaning that the data points or observations are of individual economic agents such as individuals, households, or firms. Contrast this with macro data compiled at the national level, which might measure national income or total expenditure by consumers.

Time-series data are available in different frequencies. Frequency means how often the concept is measured. The U.S. Census is conducted every ten years. U.S. GDP is measured quarterly. Unemployment is available monthly. Interest rates are available daily, and stock prices may be available hourly. Each of these variables is also available for longer time periods. For example, U.S. GDP, unemployment, interest rates, and stock prices are available at annual rates, too. Figure 8.1 lists a selection of commonly used time-series variables and the frequencies with which they are measured.

Cross-section data have a comparable concept to data frequency that involves the "unit of analysis." With time series data, the observations are taken at different points in time. By contrast, with cross-section data, observations are measured across different nations, states, or other units of analysis. Suppose you need data for income. Average disposable personal income, for example, can be measured in each of the fifty states...
Organizations That Collect and Publish Data

A number of U.S. governmental, international, and private organizations gather economic and social statistics. These include the Census Bureau, the Bureau of Economic Analysis (BEA), the Bureau of Labor Statistics (BLS), the Federal Reserve, the International Monetary Fund, the World Bank, and the United Nations, as well as private organizations such as the Conference Board. Let's survey some of these.

Census Bureau

The Census Bureau (www.census.gov) constructs general and population statistics, as well as foreign trade data on goods and services imported into or exported from the United States. An excellent place to start looking for general statistics is the Statistical Abstract of the United States. Though the Statistical Abstract may not have the specific data you need, the references (source information) for each of its tables can give you clues about where that specific data may be found. The Census Bureau publishes several other useful references, such as the State & Metropolitan Area Data Book, USA Counties, and the County & City Data Book. The Census Bureau conducts the decennial U.S. Census of Population, which includes detailed socioeconomic data. It also conducts several lesser-known surveys, such as the Economic Census (every five years), Annual Survey of Manufactures, Current Industrial Reports (annual/quarterly/monthly), and the American Housing Survey. Finally, the Census Bureau conducts a number of surveys on behalf of other agencies (e.g., BLS). These will be discussed later.

Bureau of Economic Analysis

The Bureau of Economic Analysis (www.bea.doc.gov) constructs major U.S. macroeconomic indicators. The BEA also produces estimates of private and public capital stocks (under the heading "Fixed Assets") and GDP by industry, U.S. balance of payments, U.S. international investment position, and regional and state data. The BEA has several major data collections, most prominently the National Income & Product Accounts (www.bea.doc.gov/bea/dn/nipaweb). It includes detailed estimates of GDP and its components (e.g., personal consumption expenditure, government expenditure, and net exports), as well as price indices for each. The report also provides a detailed breakdown of national income. The BEA publishes the monthly Survey of Current Business (http://www.bea.doc.gov/bea/pubs.htm#SCB%20Table), which provides notice of revisions of each of its products as well as detailed discussion of data sources and methodology.

Bureau of Labor Statistics

The Bureau of Labor Statistics (www.bls.gov) constructs data related to employment issues, productivity, and prices (e.g., the Consumer Price Index, the Producer Price Index, the Employment Cost Index, and components of each). The BLS produces a number of major data collections, which will be discussed later in this chapter, including the Current Population Survey, Current Employment Statistics, the Consumer Expenditure Survey, as well as the National Longitudinal Surveys. The BLS also publishes the Monthly Labor Review, which plays a role for the BLS similar to that of the Survey of Current Business for the BEA.

Federal Reserve

The Federal Reserve (www.federalreserve.gov) constructs principally financial data: interest and exchange rates, money stock and components (e.g., bank reserves), debt measures, bank assets and liabilities, official reserves, household assets and liabilities, and corporate debt. These data are available at http://www.federalreserve.gov/press.htm. The Fed also constructs the U.S. Flow of Funds Accounts and the Survey of Consumer Finances. Though much of the data from the Federal Reserve System is produced by the Board of Governors (BOG), researchers should not neglect the data available from the regional Federal Reserve banks, especially if one is interested in regional economic research. Some of these data are not available from the BOG, for example, the Atlanta Federal Reserve bank computes a weighted foreign exchange value for the U.S. dollar, based on the ten largest trading partners of the United States. Also, the St. Louis Fed has an easy-to-access collection of U.S. macro data (called FRED II).

International Agencies

A number of international agencies also collect economic and financial data of use to researchers. These agencies include the International Monetary Fund, the World Bank, the Organization for Economic Cooperation and Development, EuroStat, the Asian Development Bank, and the Inter-American Development Bank. A number of United Nations' agencies also publish data. Most of this data, however, is available in electronic formats via expensive subscriptions. However, it is fairly widely available in print formats in many college and university libraries.

Major Primary Data Collections

In this section, you can find more detailed explanations of the major data collections mentioned previously, as well as other major primary data collections. A more comprehensive survey of the major data collections can be found in Maier (1999).
chapter 8 • Locating (and Collecting) Economic Data

U.S. National Income and Product Accounts
The National Income and Product Accounts (NIPA) are the official national accounts (i.e., macroeconomic statistics) of the United States and can be found at [http://www.bea.doc.gov/bea/dn/nipaweb/index.asp](http://www.bea.doc.gov/bea/dn/nipaweb/index.asp). This is the source of all U.S. macroeconomic data. The accounts include real and nominal data on national income and product, personal income and expenditures; government receipts and expenditures at the federal, state, and local levels; foreign transactions; saving and investment; income and employment by industry; and price indices for nearly all these categories. The data set offers an amazing amount of detail. For example, Table 2.6, "Personal Consumption Expenditure by Type of Product," indicates that U.S. residents spent $80.6 billion on higher education in 2000, but only $62.1 billion on dental services. Annual data are available back to 1930, and quarterly data (for the major series) are available back to 1972. Much of the data is available online and is easily downloadable in spreadsheet format. For additional information, see Mauer (1999), Chapter 7.

U.S. Flow of Funds Account
The Flow of Funds Account ([http://www.federalreserve.gov/releases/Z1/](http://www.federalreserve.gov/releases/Z1/)) collects data on financial flows across the U.S. economy. It includes detailed information on levels of and changes in assets and liabilities by each sector of the economy. Annual data are available back to 1945, quarterly data are available back to 1952, and monthly data for the major components of domestic nonfinancial debt are available back to 1955.

U.S. Balance of Payments Accounts and International Investment Position of the U.S.
The international accounts of the U.S. ([http://www.bea.doc.gov/bea/d1.htm](http://www.bea.doc.gov/bea/d1.htm)) are compiled in two sets of data that correspond to a flow account and a stock (or asset) account. The flow account measures exports and imports of goods and services (the Current Account) as well as capital inflows and outflows (the Capital Account). These data are available quarterly for aggregate data that measure U.S. balance of payments vis-à-vis the rest of the world. Quarterly data are also available for U.S. international transactions with Western Europe, the European Union, the United Kingdom, Eastern Europe, Canada, Latin America, Japan, and Australia. Annual data are available for trade with Belgium-Luxembourg, the Netherlands, many, France, Italy, Mexico, Venezuela, and South Africa. The aggregate stock account data, which are collected annually, summarize U.S. holdings of assets from and liabilities to the rest of the world. More detail is available (by country and industry) for data on U.S. direct investment abroad and foreign direct investment in the United States.

U.S. Census of Population and Integrated Public Use Microdata Series
The U.S. Census ([http://www.census.gov/main/www/cen2000.html](http://www.census.gov/main/www/cen2000.html)) is the decennial measurement of the U.S. population. In addition to counting the total number of Americans, the Census collects a wide range of socioeconomic information that is disaggregated by location, from the national level all the way down to zip codes. The Minnesota Population Center's Integrated Public Use Microdata Series (IPUMS, [http://www.ipums.umn.edu/](http://www.ipums.umn.edu/)) is comprised of twenty-five "high-precision" samples of the U.S. population from U.S. censuses since 1850. The data include economic and social variables. For additional information, see Mauer (1999), Chapter 2.

Current Population Survey
The Current Population Survey (CPS, [http://www.bls.gov/cps/home.htm](http://www.bls.gov/cps/home.htm)) is a monthly survey of a sample of U.S. households designed to collect data on employment, unemployment, and individuals not in the labor force. Its data are classified by a number of demographic characteristics, including age, gender, race, occupation, industry, and hours of full- or part-time employment. Some data are available on the work experience and educational background of workers. For additional information, see Mauer (1999), Chapter 9.

Current Employment Statistics
Current Employment Statistics (CES, [http://www.bls.gov/ces/home.htm](http://www.bls.gov/ces/home.htm)) is a monthly survey of payroll records of a large sample (more than 100,000) of U.S. businesses. It collects data similar to that of the CPS, but the source is employers rather than households. Specifically, the CES collects detailed data on employment (i.e., number of employees), weekly work hours, and hourly and weekly earnings of nonfarm workers by industry. The data are organized in two ways: Ownership (private versus public employer, and if the latter, then federal, state, or local government) and industry (using the North American Industry Classification System (NAICS). Data are available at the national level, as well as by state and metropolitan area. National data for employment are available back to 1999. National data for work hours are available back to 1947; national data for hourly earnings are available back to 1962. Data for states and metropolitan areas are available back to the 1980s.
The Economic Census

The Economic Census (http://www.census.gov/epcd/www/econ.html) provides a detailed snapshot of the production side of the U.S. economy. It covers all the production sectors of the economy, except for agriculture and government (which are addressed in separate surveys). These include: mining; utilities; construction; manufacturing; wholesale and retail trade; transportation and warehousing; information; finance; insurance; real estate; rental and leasing; professional, scientific, and technical services; management; educational services; health care; entertainment; food services; and others. Beginning with the 1997 Economic Census, data are organized under the NAICS industry classification system rather than the previous Standard Industrial Classification (SIC) system. The Economic Census is a survey of “establishments,” which are defined as individual business units at a single location. Thus, a single company (e.g., Pizza Hut) may include multiple establishments. Key data collected include the number of establishments, number of employees, payroll, measures of sales or output, costs, and assets. Data are available (in decreasing detail) at the national, state, metropolitan area, county, and zip code levels, and also by specific industry.

Annual Survey of Manufactures

The Annual Survey of Manufactures (ASM) is a survey of sample manufacturing establishments organized according to the NAICS industry classification system. The survey can be found at http://www.census.gov/econ/overview/ma0300.html. The survey includes data on output, value added, employment, labor costs, materials costs, capital expenditures, energy consumption, and inventories.

Current Industrial Reports

The Current Industrial Reports (CIR, http://www.census.gov/www/index.html) provide annual, quarterly, and monthly data on production and shipments of manufactured goods in selected industries including aerospace equipment, chemicals, computer and electronic components, consumer goods, industrial equipment, primary metals, textiles, and apparel. The most disaggregated data are available at the annual level; the least at the monthly level.

American Housing Survey

The American Housing Survey (AHS, http://www.census.gov/hhes/www/ahs.html) is a panel survey conducted biannually at the national level and every four years for major metropolitan areas. The national panel includes approximately fifty-five thousand housing units, including single-family homes, apartments, and mobile homes. The survey includes data on housing and neighborhood characteristics, housing costs, and income of the residents.

Consumer Expenditure Survey

The Consumer Expenditure Survey (CES, http://www.bls.gov/cex/home.htm) is a survey of the spending behavior of a sample of American households. The data are derived from a quarterly interview and a weekly diary that respondents fill out. The result is annual data on income and expenditures disaggregated by age, household size, and other demographic characteristics. Expenditures are broken down by major classes, including food, housing, apparel and services, transportation, health care, entertainment, personal care products and services, reading, education, tobacco products and supplies, and personal insurance and pensions. Data are available back to 1980 for some series, and 1984 for the rest. For a similar data set, see Tables 2.6 and 2.7 of the NIPA.

National Longitudinal Surveys

The National Longitudinal Surveys (NLS, http://www.bls.gov/nls/home.htm) are a collection of surveys, conducted by the Bureau of Labor Statistics over the past thirty years, to develop data on the labor market activities, broadly defined, of several representative groups of Americans. These are probably the longest running and certainly the best-known panel data sets in the United States. The original surveys consisted of four cohorts: Young Men, Young Women, Mature Women, and Older Men. All cohorts were constructed to be nationally representative samples of their respective groups. The surveys include data on education, training, marriage, fertility, health, income, and assets. The Young Women’s survey includes women who were ages fourteen through twenty-four when first interviewed in 1968. The Mature Women’s survey includes women who were ages thirty through forty-four when first interviewed in 1967. These surveys are now conducted simultaneously in beef-numbered years. The Young Men’s survey, which was discontinued in 81, includes men who were ages fourteen through twenty-four when first interviewed in 1966. The Older Men’s survey, which was discontinued in 1990, includes men who were ages forty-five through fifty-nine when first interviewed in 1966.

There have been three subsequent cohorts: “Youth 1979,” “Children of Youth 1979 - Supplement to NLSY79,” and “Youth 1997.” The NLSY79 is a survey of men and women born in the years 1957-64; respondents were fourteen through twenty-two when first interviewed in 1979. The NLSY79 “Children and Young Adults” is a survey of the children of the men in the NLSY79. The NLSY97 is a survey of young men and women.
born in the years 1980-84; respondents were ages twelve through seventeen when first interviewed in 1997.

The NLS79 includes data on wages, work hours, industry, and occupation. It also includes some information on spouse employment; highest level of education completed (by grade); high school GPA and DOD aptitude test (ASVAB) scores; job training; family relationships (marriage, number of children, etc.). The “Children of Youth 1979” includes data on cognitive, socio-emotional, and physiological assessments; household income; net worth; and health status. For a similar data set, see the Panel Study of Income Dynamics (PSID).

Panel Study of Income Dynamics

The Panel Study of Income Dynamics (PSID, http://www.isr.umich.edu/src/psid/) is a longitudinal study of a representative sample of U.S. families. The study includes data on the families as well as the individual members—men, women, and children in them. Begun in 1968, the data were collected annually through 1997, and biennially starting in 1999. The sample size began as forty-eight hundred families. By 2001, it had grown to over seven thousand families. Though similar to the NLS in terms of content, the PSID focuses more explicitly on families.

The PSID includes data on income sources and amounts; poverty status; public assistance in the form of food or housing; other financial matters (e.g., taxes, inter-household transfers); family structure and demographic measures (e.g., marital events, birth and adoptions, children forming households); labor market work (e.g., employment status, weekly work hours, unemployment/vacation/sick time, occupation, industry, work experience); housework time; housing (e.g., owner/renter, house value/rent payment, size); geographic mobility (e.g., when and why moved; where family head grew up; all states head has lived in); socioeconomic background (e.g., education, ethnicity, religion, military service; parents’ education, occupation, poverty status); and health (e.g., general health status, disability).

Beginning in 1997, the PSID includes a Child Development Supplement. This survey uses a subsample of the PSID consisting of 2,394 households and 3,563 children up to twelve years old. Its purpose is to collect data, every five years, on a variety of issues impacting the cognitive and emotional development of children.

Surveys of Consumers

The Survey Research Center (http://www.sca.isr.umich.edu) at the University of Michigan publishes a number of indices that measure household expectations. The most prominent of these is the Index of Consumer Sentiment. These data are accessible through the “guest” login at the Survey Research Center. They are also available from the FRED II database at the St. Louis Fed.

The Conference Board (http://www.conference-board.org) publishes the monthly Consumer Confidence Index. The most recent monthly data are available free from their webpage, but past data require a subscription.

Survey of Consumer Finances

The Survey of Consumer Finances (SCF, http://www.federalreserve.gov/pubs/oss/oss2/sfindex.html) is a survey conducted by the Federal Reserve Board of Governors of selected demographic characteristics of U.S. families, including their income, balance sheets, and use of financial services. The survey has been triennial since 1983. For similar data on the U.S. economy as a whole, see the U.S. Flow of Funds Accounts.

Major Secondary Data Collections

In the previous section we identified the major producers of U.S. economic and social data. You can certainly obtain data from these sources, especially if you are willing to dig for it. However, it is often easier to go to secondary sources of data. Secondary sources are usually more user-friendly than primary sources, which tend to be designed for expert users. Additionally, just as it may be helpful to begin a literature search with secondary sources, the same is true for data searches, especially if you are uncertain about what data are available on your topic or where those data may be found. What follows is meant to be an exhaustive list but rather a sample of what is available. It also includes a sample of the major secondary sources of international economic and social data.

The Economic Report of the President

The Economic Report of the President (ERP, http://w3.access.gpo.gov/erp/) provides easy access to nearly all the macroeconomic data undergraduates need for their courses (though not for their research papers). This is a good first place to look for macro data.

Econographic

Econographic (www.econographic.com) is an excellent source of U.S. macroeconomic regional as well as some foreign data. The advantages of Econographic are, first, that it is a very large site, covering data for more than one hundred thousand variables, and second, that it provides a single, easy-to-use interface for accessing all that data, which is produced by a variety of government agencies. The data are available in multiple formats from spreadsheets to PDF files. The site also allows users to do some data manipulation.
chapter 8 • Locating (and Collecting) Economic Data

before downloading. For example, if the data available focus on GDP but you need growth rates, Economagic will compute them for you. More advanced features are available by subscription.

**FRED II (Federal Reserve Economic Data)**

FRED II (http://research.stlouisfed.org/fred2/) is an excellent source for U.S. macroeconomic and financial data. It also provides regional data for the region of the country serviced by the St. Louis Federal Reserve Bank. The data are easy to download.

**STAT-USA/State of the Nation**

STAT-USA (http://www.stat-usa.gov/) is a U.S. federal government-sponsored data site. It provides a single interface for obtaining U.S. macroeconomic, financial, and industry data. The primary data are obtained from the U.S. Bureau of Economic Analysis, the U.S. Bureau of Labor Statistics, and the Federal Reserve. The data are available by subscription or for free through federal depository libraries, which includes many universities. STAT-USA is not as user friendly as Economagic or FRED II. The main advantage of this site is that it includes some specialized data that are unavailable from those other sources.

**Inter-university Consortium for Political and Social Research**

Inter-university Consortium for Political and Social Research (ICPSR, http://www.icpsr.umich.edu/) is a major archive of data for social science research. It covers data on a broader range of topics, for example, survey and panel data, than the other (i.e., economics) sources discussed here; it is a good source of data for microeconomic research. ICPSR is a subscription service, but it is available to researchers at many universities.

**International Financial Statistics**

International Financial Statistics (IFS) is the principal data set of the International Monetary Fund (IMF). It is an excellent source of macroeconomic, international, and domestic financial data for IMF member countries. Data include exchange rates, international liquidity, international banking, money supplies, interest rates, price levels, international trade, government accounts, and GDP and its major components. Data are available in monthly, quarterly, and annual frequencies. Not all data sets are available for all countries or for all frequencies. The main negative feature (from an undergraduate's point of view) is access. IFS is widely available in hard copy, but it is not available online. It is possible to purchase the data on a single-user CD-ROM, but the price ($350 for the academic rate) makes it prohibitive for many colleges. For annual data, the IFS book, even in hard copy, can provide a quick fix. One additional caveat that since the data are supplied from member-country governments, you should not assume that they are directly comparable. For such comparisons, a better data source might be the Penn World Tables (discussed later in this section).

**World Economic Outlook Database**

The World Economic Outlook is a biannual (May and October) survey of world economic conditions published by the IMF. The database (http://www.imf.org/external/np/staff/cms/oeaction=FilterSearch&filter=squery.htm&QueryText=weodb) is designed to complement the survey. It provides annual data (back to 1970) for roughly fifteen macro series, for IMF member countries. These include GDP (various measures), inflation, government budget balance (actual and structural), output gaps, net capital flows, and external debt and debt service. There are also half a dozen "world" macro aggregates, including trade volume.

**Penn World Tables**

The Penn World Tables (http://datacentres.chass.utoronto.ca/pwt/index.html) provide annual data for more than 150 countries on twenty-nine macro and related aggregates. The strength of this data set is that it offers the ability to make meaningful international comparisons between each country's data. The range of data available is 1950 through 1992. Data are not available, however, for all countries over the full range.

**Joint BIS-IMF-OECD World Bank Statistics on External Debt**

The Joint BIS-IMF-OECD World Bank Statistics on External Debt database (http://www.oecd.org/statistics/jointdebt) provides definitive data on nations' external debt. It includes data on debt balances and changes in debt.

Eurostat

Eurostat (http://epp.eurostat.ec.eu.int/portal/page?_pageid=1090,37978&_dad=portal&_schema=PORTAL) is the official statistical agency for the European Union. As such, it provides a large amount of high-quality data on a wide array of subjects, including national accounts, foreign trade, and finance. Current data are available for free, but historical time series are fee-based.

**CD Main Economic Indicators and National Accounts**

The Organization for Economic Cooperation and Development (OECD), http://www.oecd.org/statsportal/0,2639,en_2825_203564_1_1_1,100.html) compiles a number of excellent databases, including Main Economic Indicators and National Accounts of OECD Countries. These
cover the major macroeconomic and financial data for the thirty member countries. The major negative feature of these data sets is accessibility. Though they are widely available in hard copy, the online versions are fee-based.

A final note: If you don’t find what you need in a secondary source, consider the primary sources!

SUMMARY

- Data are constructed rather than collected.
- Most social science statistics are based on sample data.
- Survey data come in three forms: time series, cross-section, and longitudinal.
- The majority of U.S. data are produced by the Census Bureau, the Bureau of Economic Analysis, the Bureau of Labor Statistics, and the Federal Reserve System.
- Secondary data sources are often easier to use than primary data sources.
- An overview of primary and secondary data sources is given in Appendix 8A.

NOTES

1. Strictly speaking, a panel is more than that; in other words, it is possible to create a “pool” of cross-sections, which is not a panel per se.
2. When we talk about micro or macro data sets, we are describing the source of the observations, not the type of analysis the data are used for. One can test a macro hypothesis (e.g., a consumption function) with micro data, and vice versa (e.g., a demand function for an aggregate consumption good or service, using the decomposition of quantities and price data for personal consumption expenditures).
3. For nations other than the United States, there are comparable government statistical agencies. See, for example, the country entries under “World and Non-U.S. Data” at Resources for Economists on the Internet (http://www.rfe.org/Data/World/index.html). A similar listing is available at (http://www.ceps.lu/statisticsframe.cfm).
4. Data are also available for Washington, D.C.; Puerto Rico; and the Virgin Islands.

SUGGESTIONS FOR FURTHER READING

Bat (2001)—Very thoughtful monograph on how data series are constructed, and the ways they can distort what they purport to measure.

Gow et al. (1996)—Very readable introduction to the Panel Study of Income Dynamics. Well worth reviewing if you are considering using the PSID.

Clayton and Giesbrecht (2001)—Thoughtful handbook on how the major macroeconomic statistics are computed, what they measure, and how to correctly use them. Discusses statistics on output, production and growth; investment and capital expenditures; employment, earnings, and profits; spending, sales, and expectations; the price level, money, and interest rates; and financial markets, international trade, and foreign exchange rates.

Faff (1954)—Classic monograph on the potential problems with commonly reported statistics. Very readable and intuitive.

Gajer (1999)—Excellent guide to the major social science statistics. Includes information on how to obtain the data online, as well as commentary on how the data are obtained and on their reliability.


Ogami et al. (2001)—Helpful introduction to the National Longitudinal Surveys. Worth checking out if you are considering using the NLS.

Shephard (2000)—Decent introduction to the National Income and Product Accounts. Worth reading if you are considering using the NIPA.

Whybark (2001)—Very good introduction to the Flow of Funds Accounts. Well worth reading if you are considering using the Flow of Funds.

EXERCISES

Go to the FRED II database at the St. Louis Federal Reserve Bank online at http://research.stlouisfed.org/fred2/. Download the monthly unemployment rates for the past five years for the states in the St. Louis Federal Reserve region. Record the complete URL for each data series, compare and contrast the state unemployment rates.

Go to Economagic.com. Collect data for 2000 per capita disposable personal income for your home state and all the contiguous states. Rank them from highest to lowest. Report the complete URLs.
chapter 8 • Locating (and Collecting) Economic Data

3. Find the primary source for data gross state product. Collect the data for 2002 for any ten states in the United States. Report the complete URLs.

4. Suppose you need annual data for GDP for Germany, France, and the United Kingdom. What source would you go to and why would that be the best?

5. Suppose you are asked to obtain annual data for an aggregate production function for the United States: \( Q = f(L, K) \). Track down source(s) for these three variables and report the complete URLs. Explain why you chose each data source.

Overview of Data Sources

Places to Start

Resources for Economists on the Internet (http://rfe.org)
WebEc (http://netec.wustl.edu/WebEc)

U.S. Macroeconomic Data
National Income and Product Accounts—Economagic, FRED, STAT-USA
U.S. Flow of Funds Accounts (macro/financial data)

Other Nations’ Macroeconomic Data
National Statistical Agencies for Selected Countries (http://rfe.org/data/world/index.htm)
EC Member Countries—EuroStat
OECD Member Countries—OECD Main Economic Indicators and OECD National Accounts
IMF Member Countries—IMF International Financial Statistics
Other Nations—World Economic Outlook, Penn World Tables, U.N. System of Accounts

U.S. Labor Market Statistics
Current Population Survey
Current Employment Statistics
Economagic

U.S. Microeconomic Data
Agricultural products—U.S. Department of Agriculture
Production of Manufactured Goods—Survey/Census of Manufactures
Demand for Consumer Goods and Services—Survey of Personal Consumption Expenditures (NIPA Table 2.4—PCE by Type of Expenditure), or Consumer Expenditure Survey
Demand for individual companies’ products—Hoover’s Online and Edgar

S. Industry Data
U.S. Economic Census
Annual Survey of Manufactures
Current Industrial Reports
chapter 8 • Locating (and Collecting) Economic Data

U.S. Micro Data Sets
National Longitudinal Surveys
Panel Study of Income Dynamics
IPUMS

International/Trade Data
U.S. Balance of Payments Accounts
U.S. imports and exports—Bureau of Economic Analysis, Customs Bureau
IMF International Financial Statistics.
Eurostat
International Investment Position of the U.S.
World Bank Statistics on External Debt

chapter 9

Putting Together Your Data Set

"Economists . . . may have a tendency to place more faith in the accuracy of data than is sometimes warranted . . . We should be diligent to remind ourselves that the mere existence of data in numerical form does not, in itself, make it accurate or error-free.”

DON ETHERIDGE

In Chapter 8 we observed that data collection is a major part of any empirical research project. Empirical research can be divided into two types: experimental and survey (or nonexperimental). In experimental research, the data come from performing the experiment. Thus, in experimental research collecting the data is, by definition, significant part of the project. Unfortunately, in survey research where one uses preexisting data, data collection is not always given the same respect; researchers often don’t put the same care and effort into it. This is a serious mistake! If you think I’m wrong here, ask yourself the following: you use the most advanced statistical techniques but your data are skewed, what kind of inferences can you draw from the results?

In this chapter we will explain what you need to do at a minimum to construct a satisfactory data set. We will discuss how to develop a search strategy for finding appropriate data with a minimum of time and effort, using both primary and secondary data sources. We will also discuss how to convert the data into a useable form and to construct a data appendix for your research.