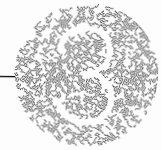
appendix

chapter 

Surveying the Literature on a Topic in Economics

"If I have seen further, it is by standing on the shoulders of Giants."

SIR ISAAC NEWTON

Earlier we noted that scholars engage in a kind of conversation or dialog of competing arguments as they attempt to increase humankind's knowledge about a topic. This conversation manifests itself as published research, which is described as "the literature" on the subject. When you survey the literature on a subject, you are trying to identify the major studies that have been published to date. But more importantly, you are trying to understand what the studies say and how they relate to one another. We will explain how to read and understand scholarly publications in Chapter 6. In this chapter we will explain why and how researchers perform a literature survey as a key component of the research process. In Chapter 12, we will explain how to write a literature survey as part of a research paper.

Why Is a Literature Survey Necessary?

Ethridge (1995, 115) observed that before you can "advance the state of knowledge," you need to know what the state of knowledge is. So when you survey the literature, you are trying to create your own sense of what is known and what is not known about the subject. This is important for several reasons. Early in the research process you can get ideas for possible angles for your research. Later in the process you can show how your research fits into and contributes to the larger conversation on the subject. This serves to establish your credentials as a researcher by showing readers that you are knowledgeable about the field.

There are several additional reasons for surveying the literature. One is to avoid duplicating findings that have already been obtained. In this fashion you can use someone else's prior work as a steppingstone for yours. For example, Ethridge (1995) points out that the literature can help you design your own study by showing how previous approaches either were or were not successful. Similarly, previous studies can suggest ways you might handle problems you run into.

Those new to research sometimes focus their literature survey too narrowly. The literature review should not be limited to searching for previous studies on your *topic*. It should also examine studies of other topics that use conceptual frameworks or testing methodologies that might be applicable to your topic. For example, in a research project on the demand for blood, it could be useful to examine previous demand studies on related products, such as the demand for transplant organs or even on other commodities entirely.

Where to Search: Popular versus Scholarly Literature

Information on a topic can be found in two types of publications: popular and scholarly publications. Scholarly publications are also referred to as scientific or professional publications. What is the difference between popular and scholarly publications? Popular publications are addressed to a general audience, while scholarly ones are addressed to a specialized audience, namely, experts in the field. Scholarly publications are often primary sources of information, while popular publications are almost always secondary sources. A primary source of information is the original publication of a research study, written by an expert and addressed to other scholars in the field. One example would be an article in the *American Economic Review* (AER, a scholarly journal) that explains a new study conducted by the authors of the article on the effects of preschool education on poverty rates. A secondary source is a report based on the primary source, typically addressed to a more general audience. An article in *Time* magazine summarizing the article in the *American Economic Review* would be an example of a secondary source in a popular publication.

One popular publication that undergraduates sometimes mistake for a scholarly journal is the *Economist*. This magazine publishes a regular column entitled "Economics Focus" that does an excellent job of summarizing recent research in language that the educated lay public can understand. In addition, unlike most popular publications, the *Economist* often provides references to the scholarly articles in which the research it reports on was published. Nonetheless, the *Economist* is a popular rather than a scholarly journal.

To begin the literature survey process, it makes sense to consult popular sources because they will give you information on what is generally known about a topic. In addition, since they are addressed to general audiences they are easy to read and assimilate. Popular sources include encyclopedias; news magazines, such as *Time, Newsweek,* and *Business Week*; and newspapers such as the *Wall Street Journal* or the *New York Times*. Note that these popular publications are increasingly available online, often for free. For example, the most recent two weeks of the *Washington Post* are available at **www.washingtonpost.com.**

Internet resources such as directories and search engines are also popular sources of general information. Most undergraduates are familiar with these tools, so we will only discuss them briefly here. Directories are hierarchically organized catalogs on a variety of subjects. Well-known examples include **http://www.galaxy.com, http://www.looksmart.com,** and **http://www.yahoo.com.** Virtual libraries are directories that have been compiled by librarians and other experts. As a result, they tend to include somewhat better resources than ordinary directories. Examples include the WWW Virtual Library (**http://www.vlib.org**) and the Internet Public Library (**http://www.ipl.org**). Each of these has sections on economics that can provide helpful background information.

Another Internet tool is the search engine. Search engines are computer programs that create databases of websites catalogued by topic. Users can perform keyword searches on those databases to obtain information from the websites dealing with their subjects of interest. Popular search engines include **http://www.google.com, http://www.excite.com,** and **http://www.altavista.com.** A particularly useful search engine is MetaCrawler at **http://www.go2net.com.** Instead of searching its own database, MetaCrawler searches the results of several major search engines, including Google, Yahoo, Ask Jeeves, About, LookSmart, and others.

The shortcoming of most general-interest directories and search engines is that they contain a great deal of information that is not useful for scholarly research. This is especially problematic for novice researchers who may have difficulty differentiating between authoritative and nonauthoritative sources. We will discuss some ways in which these tools may be productively used later in this chapter.

Although it makes sense to begin reviewing the literature by consulting popular sources, it is important to realize that the conversation between experts in the field that creates knowledge occurs only in the scholarly literature. These professional publications have undergone a peer referee or review process prior to publication. This review conveys at least a certain degree of validity to these studies. Scientific publications include articles

Table 3.1 Examples of Scholarly Economics Journals

National Journals:	*American Economic Review*
	Journal of Political Economy
	Quarterly Journal of Economics
	Econometrica
	Review of Economics and Statistics
	Review of Economic Studies
Regional Journals:	*Economic Inquiry*
	Southern Economic Journal
	Eastern Economic Journal
Specialized Journals:	*Journal of Urban Economics*
	Journal of Money, Credit and Banking
	Journal of Industrial Economics

in professional journals, scholarly books, monographs, working papers, and some government documents. Monographs are short books, typically no more than one hundred pages, on a single (hence "mono") scholarly topic. Working papers are scholarly articles in progress—prior to final publication. Table 3.1 gives examples of scholarly economics journals published in the United States.[1]

Many professional journals are difficult for undergraduates to understand, and as we noted earlier, Chapter 6 will explain how to decipher them. You should be aware that some journals are more readable than others.[2] For example, each regional Federal Reserve Bank publishes an *Economic Review.* Publications of the Congressional Budget Office are also quite good. In addition, publications of the various think tanks such as the Brookings Institution, the American Enterprise Institute, the Economic Policy Institute, and the Cato Institute tend to be more accessible. However, it is important to remember that think tanks are in business to promote their particular point of view. Thus, one should read their publications more carefully and critically than more objective sources.

How to Search: Developing an Effective Search Strategy

To locate information efficiently, one needs to use a search strategy. This section will explain how to develop one by using a sample research question: "To what extent was the 2001 U.S. economic slowdown caused by the decline in the stock market?"

Table 3.2 AEA/JEL/EconLit Subject Descriptors

A – General Economics and Teaching
B – Schools of Economic Thought and Methodology
C – Mathematical and Quantitative Methods
D – Microeconomics
E – Macroeconomics and Monetary Economics
F – International Economics
G – Financial Economics
H – Public Economics
I – Health, Education, and Welfare
J – Labor and Demographic Economics
K – Law and Economics
L – Industrial Organization
M – Business Administration and Business
N – Economic History
O – Economic Development, Technological Change, and Growth
P – Economic Systems
Q – Agricultural and Natural Resource Economics
R – Urban, Rural, and Regional Economics
Z – Other Special Topics

There are two general approaches to locating information. Ackermann and Hartman (1998) describe them as browsing and keyword searches. You can think of them as emphasizing human capital (or thinking) versus physical capital (or brute force). Neither is superior for all searches; rather a key task in developing a search strategy is to determine the optimal combination of the two for a given search.

Browsing

Browsing means manually examining a document (e.g., a printed bibliography or an online directory) for useful information or references to useful information. Effective browsing requires you to think carefully and logically about how information about a discipline is organized. The American Economic Association (AEA) uses a hierarchical system to classify subject information in economics, an overview of which is shown in Table 3.2.[3]

Table 3.3 Expanded AEA Subject Descriptors for A – General Economics and Teaching

A00 – General
- A1 – General Economics
  - A10 – General
  - A11 – Role of Economics; Role of Economists
  - A12 – Relation of Economics to Other Disciplines
  - A13 – Relation of Economics to Social Values
  - A14 – Sociology of Economics
  - A19 – Other
- A2 – Teaching of Economics
  - A20 – General
  - A21 – Pre-college
  - A22 – Undergraduate
  - A23 – Graduate
  - A29 – Other

Each of these general headings can be further expanded, as illustrated in Table 3.3 for the category "A – General Economics and Teaching." The complete classification system can be reviewed at **http://www.aeaweb.org/journal/elclasjn.html.** By knowing this classification system, you may be able to find useful information on a research topic more easily.[4]

There are two journals you should be particularly aware of as you begin your research. The *Journal of Economic Perspectives* (JEP) is designed as a general-interest journal for economists. As such, the majority of articles it publishes are summaries of current thinking by economists on a given topic. Similarly, the *Journal of Economic Literature* (JEL) primarily publishes survey articles of the literature on specific topics in economics. The American Economic Association publishes both journals, and their tables of contents for the past several years are available at the AEA website (**http://www.aeaweb.org/jep/contents** and **http://www.aeaweb.org/journal/contents**).[5] If you can find an article on your research topic in either of these journals, you will be ahead of the game since they will almost certainly identify and discuss the relevant major studies as of the publication date of the article. You will still need to look for more recent studies, however.

More generally, when you locate a useful journal article, you should consider the sources it references as obvious targets for your review. Bear

in mind, however, that since the article's author may not have pursued the same research question as you, its references should be a complement to, *not* a substitute for, your own literature survey. In sum, browsing is more likely to give you useful information than will simple keyword searches, but your results will depend on the judgment of the author or editor of the document you browsed.

Sample Browse Search

Suppose you are researching **the extent to which the recent U.S. economic slowdown was caused by the decline in the stock market.** You are aware that the life-cycle model of consumption theory suggests that household wealth is an important determinant of consumer spending. A review of the contents of the *Journal of Economic Perspectives* reveals an article by James Poterba titled "Stock Market Wealth and Consumption" in the spring 2000 issue. The article includes two pages of references going back to Ando and Modigliani's classic 1963 paper on which the life-cycle model is based.

Keyword Searching

The alternative way to locate information is by performing a keyword search. Keyword searches use search engines on the World Wide Web or on specialized databases. These latter include bibliographic databases, such as *EconLit* (formerly the *Economics Literature Index*), which consist of citations and abstracts only, as well as full-text databases, such as Dow Jones Interactive, which contain entire documents. Access to specialized databases is provided by a number of commercial vendors, such as OCLC FirstSearch or DIALOG. Researchers can obtain access to many of these databases through their university library or computer network.

Keyword searching allows you to examine far more documents than would be feasible by browsing, even assuming you knew about them. The major web search engines claim to index all the documents on the Internet, while search engines in electronic databases index all the documents in the database.[6] For example, *EconLit* currently indexes more than six hundred economics journals, seventeen hundred new books, and nine hundred new dissertations per year—for nearly a complete record of all English-language publications in economics since 1969. As such, *EconLit* should be one of the first places you start searching when researching in economics.

One exceptional tool for keyword searching is the *Social Science Citation Index (SSCI).* The *SSCI* allows you to search for citations both backward and forward in time. In our earlier discussion of browsing, we noted that when you find a useful publication, say DeLoach (2001), it makes sense to review

NOTES FOR NOVICE RESEARCHERS

Do Not Limit Your Search to Full-Text Databases!

Undergraduates sometimes view the literature search as nothing more than a hurdle on the way to completing a research project. (This is not true—indeed, as we have tried to explain, discovering what other researchers have done in the field can be a major help in completing your own research.) As a consequence, such students may limit their search to full-text databases. After all, they reason, even if you find something in a bibliographic database, it's still possible it would be difficult or impossible to acquire. This is poor judgment! The goal of a literature search is to find the most important studies done on a topic, not merely the easiest ones to obtain. Put another way, if there is a major study that you fail to find because it was not cataloged on a full-text database, your research will have a substantial flaw.[7] Nearly all college and university libraries subscribe to the major economics journals. This means articles from those journals are available in printed form, microfiche, or film at your library. Even if your library does not subscribe to a journal that you need, it's possible that one of your faculty members does. It never hurts to ask. Limiting yourself to full-text databases will also virtually rule out any books as sources. Finally, it is hard to imagine a reference that you could not obtain through interlibrary loan. In short, don't limit yourself to full-text databases! It's false economy—it will cost you in the end more than it's worth.

the earlier studies that it cites. *SSCI* can do this for you electronically, since for every journal article or book it catalogs, it also indexes the publication's citations. Doing this electronically can be a bit faster than reading a study's bibliography, but the real advantage of *SSCI* is in doing forward searches. That is, the *SSCI* database also indexes all the publications that cite DeLoach (2001) in *their* reference lists. As a result, it is very easy to get a complete list of the major studies done on a topic.

An additional advantage of the *SSCI* is that it is interdisciplinary; it covers a variety of social science fields. What this means is that you are less likely to miss a study done on your topic but in a field other than economics, as could occur if you only searched *EconLit*. The main disadvantage of the *SSCI* is that it is quite expensive, so many colleges and universities do not subscribe to it. If your institution has it, you should use it.

Though your scholarly literature survey should concentrate on using specialized databases to find literature, web search engines can also be profitably employed. The best web search engine for academic uses is probably Google (**http://www.google.com**). A criticism of Web-based research is that the majority of what is on the Web is not scholarly in nature. One advantage of Google is that in addition to indexing webpages, it also catalogs a number of other file types, including Adobe document (.pdf) files, Word and WordPerfect documents, Excel and Lotus 123 spreadsheets, PowerPoint presentations, and others. As a result, a researcher could use Google to find scholarly working papers or data sets on researchers' personal webpages, even before they are published.

The principal disadvantage of keyword searches, when using specialized directories or web search engines, is that they may generate many "hits," but only a few useful ones. Imagine doing an electronic search using the phrase *stock market*. Depending on the database, you could get hundreds or even thousands of results. Even if there are many relevant hits, you will need a great deal of time and effort to separate the wheat from the chaff. Consider the way commercial fishermen use a dragnet. The net scoops up everything in the water for several miles around. Then even though the fishermen only want codfish, they need to examine and throw away everything else they catch.

The solution to this quandary is to perform a more advanced keyword search using a combination of Boolean and phrase searching. Boolean searching allows you to focus the search in the way most likely to obtain useful results. There are three Boolean operators: AND, OR, and NOT.

AND is used to **narrow a search.** For example, a search on the keywords *Keynesian* AND *Post* would locate only those items including **both** keywords, such as "Post Keynesian."

OR is used to **widen a search.** A search on the keywords *Monetarist* OR *Keynesian* would locate those items including **either** keyword.

NOT means **exclude.** It includes all items **except** those that contain the NOT keyword. For example, a search on the keywords *Keynesian* NOT *Post* would identify all items with "Keynesian," except those with "Post Keynesian."

You can also fine-tune your search using nested Boolean logic. For example, a search on the keywords *Monetarist* OR (*Keynesian* NOT *Post*) would locate all items including either "Monetarist" or "Keynesian," except those with "Post Keynesian."

Phrase searching looks for the items that contain an exact phrase that you specify within quotes. For example, a search on "*Post Keynesian*" would locate only those items with that exact phrase, but it would exclude items that include those two keywords separately. A wildcard character can allow you to search more efficiently by truncating your search phrase. If the wildcard character for a search engine is *, then searching on the keyword *Keynes** would locate items containing "Keynes" or "Keynesian" or "Keynesians." Similarly, a search on the keyword *wom*n* would yield items containing "woman" or "women."

A Basic Search Strategy

Ackerman and Hartman (1998) propose a basic search strategy, which is shown (in a revised form) in Table 3.4. Though the table gives the impression that this is a straightforward process, in practice it is a great deal more uncertain. For one thing, at the beginning of the process the researcher may not know exactly what to look for. This is especially true for a researcher new to the topic. For another, the process is iterative. You will never find everything in a single search! Rather, you will need to try an initial set of keywords with one database. Once you find a few good sources, read them. As you become more familiar with the literature, other ideas for keywords will come to mind, which you can then search with. If one database is unsuccessful, try another. If one subject approach doesn't yield enough, try a related field. There is an art to keyword searching, and some researchers are better than others at it. Fortunately, anyone can become better with practice.

Table 3.4 A Basic Search Strategy

1. Begin by stating your research topic or question;
2. Identify important concepts relevant to your topic;
3. Brainstorm to create a list of keywords that describe these concepts;
4. Determine any synonyms to these keywords;
5. Choose a subject approach; that is, determine which discipline or disciplines are likely to have literature on your topic; for example, economics, law, public policy, business, sociology, education;
6. Determine which search features may apply; for example, Boolean operators, wildcards;
7. Choose an appropriate database to search the given subject; for example, *EconLit*;
8. Read the search instructions for the database;
9. Create a search expression using the appropriate syntax;
10. View the results;
11. Modify the search if necessary (return to Step 2);
12. Try the same search with another database (return to Step 6).

Sample Keyword Search

Let's perform a keyword search on our sample research topic using the search strategy outlined in Table 3.4.

1. Recall that the research topic is **the extent to which the 2001 economic slowdown was caused by a decline in the stock market.**
2. The research topic can suggest the important concepts to begin with. These could include **stock market decline** and **economic slowdown.**
3. A little brainstorming about possible connections between the two can suggest other keywords to search for: **household wealth, consumer spending,** and **life-cycle model.** Note also that in choosing keywords, one can select the JEL subject classification for one's topic. In this case, we could choose *G100* for "General Financial Markets" or *E200* for "Macroeconomics: Consumption, Saving, Production, Employment, and Investment, General."
4. Synonyms for (stock market) *decline* could be *crash* or *correction* or *bear market.* A synonym for (economic) *slowdown* could be *recession.* Synonyms for *wealth* could be *savings* or *saving.* Synonyms for *consumer spending* could be *consumer expenditure* or *consumption.*
5. The research topic suggests two fields whose literature might prove useful: economics and business.
6. For multiple-word expressions we will need to use phrase searching, for example, "*consumer spending.*" For synonyms we will need to use Boolean operators, for example, "*consumer spending*" OR "*consumer expenditure*" OR "*consumption.*" For relationships we will need to use the Boolean AND operator, for example, "*consumption* AND *wealth.*"
7. When performing economics research, a good database to begin with is *EconLit.* For other fields, the following might be helpful:
   - *Expanded Academic ASAP* – for business academic journals,
   - *Wilson Business Abstracts* – for business academic journals,
   - *Business & Company Resource Center* – for business information at the company level,
   - *ERIC* – for education,
   - *PAIS* – for public affairs and international studies,
   - *Dow Jones Interactive* - for current topics not yet in the scholarly literature, also business academic journals,
   - *SSCI* – for any of the social sciences.

Be aware that the user interface and the format in which search results are provided can differ depending on which vendor supplies your access to the database.

8. The search instructions for *EconLit* are available at **http://www.econlit.org/econlit/hints.html.**[8]
9. An initial keyword search was performed using the expression "*consumption* AND *wealth.*"
10. This search yielded 1,422 hits, which is too many to review meaningfully.
11. A revised search using the expression "*consumption* AND *stock market*" yielded 97 hits, including a promising article by Shirvani & Wilbratte (2000), "Does Consumption Respond More Strongly to Stock Market Declines than to Increases?" A subsequent search using the expression "*consumer spending* (OR "*consumption*") AND *recession*" resulted in 78 hits, including another good one, Blanchard (1993), "Consumption and the Recession of 1990–91." Another search using the expression "*stock market* AND *recession*" yielded 13 hits of which three look useful.
12. Let's try searching another database: Expanded Academic ASAP. A search using the expression "*consumer spending* (OR "*consumption*") AND *recession*" yielded 22 hits, of which several appear interesting. One of these was Blanchard (1993), which we found previously.

You should try all the search expressions that generated good results (i.e., not too many hits, and many of them useful) on *each* of the databases searched. At some point, you will begin to turn up the same sources, like Blanchard (1993). This is a sign that you're near the end of effective searching. Notice that our keyword searches did not turn up the Poterba (2000) paper we discovered earlier by browsing. That is why both types of searches are necessary. After completing the browsing and keyword searches, and examining the references cited by the studies found there, we can conclude that we have found the major published studies to date on our research question.

Obtaining the Resources

Once you have identified potential sources, you need to obtain them. The World Wide Web has revolutionized access to research materials, especially for scholars at smaller institutions whose libraries may have only limited collections. More and more resources are available online, including full-text books and journal articles.

More than a hundred economics journals are available free in full-text format. One list of such journals is available at the *Resources for Economists* webpage at **http://rfe.org.** Another list is maintained at the College of Wooster (**http://www.wooster.edu/economics/archive/journals.html**).

There are also a number of proprietary portals to full-text journals. For example, the American Economic Association hosts a web portal (**http://www.aeaweb.org**) that includes full-text access to the past three years of the journals it publishes: the *AER, JEL,* and *JEP.* Access to the journals' tables of contents and abstracts is available to anyone, but full-text access requires AEA membership. The site also includes Bill Goffe's *Resources for Economists on the Internet,* which will be discussed in detail in Chapter 9. JSTOR is an archive of full-text scholarly journals, including (as of this writing) twenty-four top journals in economics, which are listed in Table 3.5.[9] Many colleges and universities subscribe to JSTOR. AEA members can access JSTOR via a modest annual fee ($10). There are several other proprietary online databases that offer full-text journal articles. One of these is *Expanded Academic ASAP,* which includes the full text of many economics journals that are not included in JSTOR.

In addition to the obvious speed and convenience that online access to journal articles provides, in the near future we will be able to link directly to references cited in those articles (backward linking). It may even be possible to link to articles citing those articles (forward linking).

Government documents are also increasingly available over the Internet. In the past decade the U.S. government has made a significant effort

Table 3.5 Economics Journals Available Through JSTOR

American Economic Review
Brookings Papers on Economic Activity
Canadian Journal of Economics
Econometrica
Economica
Economic History Review
Economic Journal
International Economic Review
Journal of Applied Econometrics
Journal of Economic History
Journal of Economic Literature
Journal of Economic Perspectives
Journal of Finance
Journal of Human Resources
Journal of Industrial Economics
Journal of Money, Credit, and Banking
Journal of Political Economy
Oxford Economic Papers
Quarterly Journal of Economics
Rand Journal of Economics
Review of Economics and Statistics
Review of Economic Studies

to make virtually all print documents accessible online. Moreover, unlike their printed counterparts, they are obtainable for free. A good point to begin searching for government publications is at **www.firstgov.gov.**

The first place to look for books is still your college or university library. To date, the availability of books online is lagging behind the availability of journal articles and government documents. One very promising source for electronic books, however, is NetLibrary (**www.netlibrary.com**), a subsidiary of On-line Computer Library Center (OCLC). NetLibrary offers electronic texts on a subscription basis to libraries. As of this writing, NetLibrary has over fifty-four thousand titles in a variety of fields including economics. Just as you would search for a printed book, you use your college or university library's search engine to find an online text. Indeed, until you find the entry, you won't know whether the book is a hard copy or an electronic one. Once you locate the electronic copy, you can "check it out" just as you would a printed copy. Unless your library subscribes to more than one copy of the e-book, no one else from your institution has access to the book until you relinquish it. An additional benefit is that you can obtain the text twenty-four hours a day without having to physically go to the library to pick it up.

We noted earlier that libraries at smaller institutions may have only limited collections. There are several ways to get around this, such as making use of interlibrary loan services or visiting nearby research university libraries. For example, it is increasingly possible to search such library catalogs remotely to ensure that their collection has what you are looking for before visiting the library.

There are also literally tens of thousands of economics working papers available online from the National Bureau of Economic Research (**www.nber.org**), the Social Science Research Network (**http://papers.ssrn.com**), the Working Papers Archive at Washington University in St. Louis (**http://econwpa.wustl.edu**), and NetEc (**http://netec.wustl.edu/NetEc.html**).

SUMMARY

- In order to perform original research, one needs to determine what is currently known on a topic. Thus, at the beginning of any research project the researcher needs to survey the "literature."
- There are two sources of information on a topic: popular and scholarly. Only the latter is considered part of the literature.
- An effective search strategy includes the following steps:
  - Start with secondary sources in the general topic area.
  - Search for survey articles on the topic, for example, in *JEP* or *JEL*.
  - Move to primary sources: for example, *EconLit* or *Social Science Citation Index*
  - Use a combination of keyword searching and browsing from the references of the useful sources you've located.

NOTES

1. Appendix 10 of Wyrick (1994) provides an extensive listing of economics journals grouped by subject area.
2. Note also that working papers tend to be more readable than published articles because they usually include many more details of the argument—details that are eliminated during the editorial process.
3. This is also referred to as the *EconLit* or *Journal of Economic Literature* classification system.
4. For example, you can do online searches using these classifications as keywords.
5. The AEA website lists the tables of contents for the *Journal of Economic Literature* back to December 1994 and for the *Journal of Economic Perspectives* back to the beginning of 1998. One can easily get the contents from earlier issues of the journals by searching *EconLit,* using the journal name as the "Source" keyword.
6. Note also that you should always use multiple web search engines since they use different search criteria and their databases, while overlapping, are not identical.
7. In my courses, such a flaw would cost at least a letter grade on the final paper.
8. Note that databases like *EconLit* are available through more than one interface. For example, at my institution, we get it from OCLC FirstSearch, a database service provider that provides access to forty-four databases on a wide variety of subjects.
9. Note, however, that articles do not become available through JSTOR until five years after original publication.
10. Note that Turabian's Chapter 8 was published before the 1999 MLA decision to recommend parenthetical references.

SUGGESTIONS FOR FURTHER READING

Ackermann and Hartman (1998)—Excellent, user-friendly guide to online searches and the Internet as a research tool.

Turabian (1996)—Definitive guide to writing and citation styles for research papers and theses. See also Appendix 3A.

EXERCISES

1. Select a research question, preferably one you are actually researching. Write down the research question. Review the tables of contents for the *Journal of Economic Perspectives* and the *Journal of Economic Literature* (provided at **http://www.aeaweb.org/jep/contents** and **http://www.aeaweb.org/journal/contents**). Browse those tables of contents for one or more survey articles that are relevant to your research question. (You should identify at least one article that is relevant to your research.) Write down the complete bibliographic citation information for the survey articles you discover, using the citation style your instructor favors. (See Appendix 3A for information on citation styles.)
2. Select a research question, preferably one you are actually researching. Write down the research question. Use the basic search strategy illustrated in Table 3.4 to perform a keyword search in *EconLit*. Write down a list of "important concepts" that are relevant to your research question. Select and write down appropriate keywords. Create a search expression, and execute the search. Modify the search as necessary to find at least five useful "hits" for your research. Write down the complete bibliographic citation information for the studies you discover, using the citation style your instructor favors.
3. Select a research question, preferably one you are actually researching. Write down the research question. Use the basic search strategy illustrated in Table 3.4 to perform a keyword search using a database *other than EconLit*. Write down a list of "important concepts" relevant to your research question. Select and write down appropriate keywords. Create a search expression, and execute the search. Modify the search as necessary to find at least five useful "hits" for your research. Write down the complete bibliographic citation information for the studies you discover, using the citation style your instructor favors.
4. Select a research question, preferably one you are actually researching. Write down the research question. Use the basic search strategy illustrated in Table 3.4 to perform a keyword search using a web search engine like Google. Write down a list of "important concepts" relevant to your research question. Select and write down appropriate keywords. Create a search expression, and execute the search. Modify the search as necessary to find at least five *scholarly* "hits" for your research. Write down the complete bibliographic citation information for the studies you discover, using the citation style your instructor favors.

appendix

Scholarly References and Citation Styles

An essential part of reviewing literature is keeping careful references for each scholarly work you use. The purpose of citation styles is to provide a standardized, concise way of doing that. References are important for two reasons: first, to give credit to previous authors for their ideas, their intellectual property, and, second, to help readers track down those ideas in full detail. We will expand on these reasons later in Chapters 5 and 6. Here we introduce the styles.

There are three general styles for citations and references: they are known as MLA, the Chicago Style, and APA. Each of these styles explains its preferred way to write scholarly notes in the text (e.g., foot- or endnotes) and scholarly references at the end of the text. All three styles now recommend using the parenthetical form for references—for example, "Turabian (1996)"—instead of citing references in the text through footnotes or endnotes. Of course, for scholarly notes that elaborate on the text, it is still appropriate to use footnotes or endnotes.

An excellent guide to the MLA and Chicago styles is Kate Turabian's *A Manual for Writers of Term Papers, Theses and Dissertations,* Sixth Edition (1996). Ms. Turabian was the graduate secretary at the University of Chicago; she was responsible for accepting all theses and dissertations submitted at that institution and as such was the acknowledged expert on the *Chicago Manual of Style.* Thus, her book is essentially a condensed version of the *Chicago Manual.* Two chapters that are particularly useful are Chapter 11, which shows a comparison of the two citation styles, and Chapter 14, which illustrates the two styles in a large number of sample notes and references. Note that this edition of Turabian is based on the Fourteenth Edition of the *Chicago Manual of Style,* rather than the most recent Fifteenth Edition. The major changes in the new edition have to do with online resources, which were not covered in the Fourteenth Edition, so Turabian is still useful. Hopefully, a new edition of Turabian will be forthcoming. The following paragraphs introduce the three styles and refer you to various guides to each.

MLA is the acronym for the Modern Language Association and is the style favored by scholars in the humanities disciplines. This style is spelled out in Chapter 8 (on footnotes) and Chapter 9 (on bibliographies) of

Turabian (1996).[10] Additional guides to the MLA style are available from the University of Wisconsin Writing Center (**http://www.wisc.edu/writing/Handbook/DocMLA.html**) and the Writer's Workshop at the University of Illinois (**http://www.english.uiuc.edu/cws/wworkshop/MLA/bibliographymla.htm**).

APA is the acronym for the American Psychological Association and is the style used by scholars in psychology and other social sciences, including some economists. Useful APA guides are available from the On-line Writing Lab at Purdue University (**http://owl.English.purdue.edu/handouts/research/r_apa.html**), the University of Wisconsin Writing Center (**http://www.wisc.edu/writing/Handbook/DocAPA.html**), and the Writer's Workshop at the University of Illinois (**http://www.english.uiuc.edu/cws/wworkshop/bibliography_style_handbookapa.htm**).

The Chicago style is the choice of the natural sciences and many social sciences, including economics. The Chicago Style is formally called the author-date system in the *Chicago Manual of Style,* Fifteenth Edition (2003). In this style, references are listed at the end of a work in the "Reference List" (not the bibliography). This style is explained in Chapter 10 of Turabian (1996). An additional guide to the Chicago style is available from the Long Island University library: **http://www.liu.edu/cwis/cwp/library/workshop/citchi.htm.**

Given that most literature surveys are now conducted online, you should be aware that online citation styles may not be adequately spelled out in the guides just mentioned. A good source of information about citation styles for online documents is Andrew Harnock and Eugene Kleppinger's, *Online: A Reference Guide to Using Internet Sources,* available at **http://www.bedfordstmartins.com/online.**

Note that a journal article (or book) found online is still a journal article (or book) and should be cited accordingly. That is, online citation styles should be reserved for documents that have no print analog. For example, an article in the *American Economic Review* that you obtain online from JSTOR should be cited as a journal article rather than as an online document.

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. After all, a research project can be hamstrung by inadequate data just as easily as by a lack of a clearly conceived, testable hypothesis. 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 of 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? Oftentimes, 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; that 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 of 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 interesting in finding income so we can test the theory of consumer behavior. What is the basic consuming unit? There are two 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 the two 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. As 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. . . . No 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 produce 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 or 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 it over time.[1] 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 for these observations is different. Longitudinal data are an example of a **micro 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.[2]

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, the 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. Alternatively, you might be able to obtain family income data across all counties in a particular state, or data on household income for sample households in a particular county. The researcher's decision about what frequency or unit of analysis to select can be guided by economic theory, but is often determined by data availability.

Figure 8.1 Available Frequency of U.S. Time Series Data

Data Available Annually:	Fixed Reproducible Capital—Fixed-Asset Tables
Data Available Quarterly:	Gross Domestic Product and Major Components— Advance Estimate, available with a 1-month lag; Preliminary Estimate, available with a 2-month lag; Final Estimate, available with a 3-month lag. Corporate Profits Productivity Growth Employment Cost Index Flow-of-Funds Accounts U.S. International Transactions (complete balance of payments: current + capital account), available with a 2.5-month lag.
Data Available Monthly:	Personal Income and Outlays, Disposable Personal Income, Personal Consumption Expenditures, available with a 2-month lag Retail Sales Employment and Unemployment Rates Consumer Price Indices Producer Price Indices U.S. International Trade in Goods and Services (i.e., current account), 1.5-month lag. Index of Industrial Production, Capacity Utilization Rate, 1.5-month lag. Consumer Credit, 5-week lag.
Data Available Weekly:	Monetary Base, Monetary Aggregates
Data Available Daily:	Exchange Rates, Selected Interest Rate

Organizations That Collect and Publish Data

A number of U.S. governmental, international, and private organizations gather economic and social statistics.[3] 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), the *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 the major U.S. macroeconomic indicators. The BEA also produces estimates of private and public capital stocks (under the heading "Fixed Assets"), 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, investment 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.

The Federal Reserve

The **Federal Reserve** (**www.federalreserve.gov**) constructs principally financial data: interest and exchange rates, money stocks 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/rnd.htm.** The Fed also conducts 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 trade-weighted foreign exchange value for the U.S. dollar, based on the fifteen-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 only 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 may be found in Maier (1999).

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).** 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. This 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 1952. Much of the data are available online and are easily downloadable in spreadsheet format. For additional information, see Maier (1999), Chapter 7.

U.S. Flow of Funds Account

The Flow of Funds Account (**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 in 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/di1.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 in- and outflows (the Capital Account). These data are available quarterly for the 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, Germany, France, Italy, Mexico, Venezuela, and South Africa. The aggregate data and some of the country data are available back through 1960. The 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**) 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/**) 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 Maier (1999), Chapter 2.

Current Population Survey

The Current Population Survey (CPS, **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 Maier (1999), Chapter 9.

Current Employment Statistics

Current Employment Statistics (CES, **http://www.bls.gov/ces/home.htm**) is a monthly survey of payroll records of a large sample (more than 500,000) 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.[4] National data for employment are available back to 1919. 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/econ97.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 a sample of 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 value of shipments, 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/cir/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 also 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. The 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 odd-numbered years. The Young Men's survey, which was discontinued in 1981, 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 ages fourteen through twenty-two when first interviewed in 1979. The NLSY79 "Children and Young Adults" is a survey of the children of the women 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 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, work/unemployment/vacation/sick time, occupation, industry, work experience); housework time; housing (e.g., own/rent, 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/scfindex.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 those sources, especially if you are willing to dig for it. However, it is often easier to go to a secondary source 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 not 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/eop/**) 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.

Economagic

Economagic (**www.economagic.com**) is an excellent source of U.S. macro and regional as well as some foreign data. The advantages of Economagic 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

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 series 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 Yearbook,* even in hard copy, can provide a quick fix. One additional caveat is

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/search97cgi/s97is_eng.dll/search97cgi/inetsrcheng.ini?action=FilterSearch&filter=spquery.hts&QueryText=weodb**) is designed to complement the survey. It provides annual data (back to 1970) on 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://datacentre.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.cec.eu.int/portal/page?_pageid=1090,1137397&_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.

OECD Main Economic Indicators and National Accounts

The Organization for Economic Cooperation and Development (OECD, **http://www.oecd.org/statsportal/0,2639,en_2825_293564_1_1_1_1_1,00.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, do 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 quantity 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://labour.ceps.lu/statisticsframe.cfm**).
4. Data are also available for Washington, D.C.; Puerto Rico; and the U.S. Virgin Islands.

SUGGESTIONS FOR FURTHER READING

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

Brown 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.

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

Maier (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.

Morgenstern (1979)—Classic critique of the use of macroeconomic statistics. Anyone using macro data should be familiar with this. Reprinted from Morgenstern's *On the Accuracy of Economic Observations,* 1963.

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

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

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

EXERCISES

1. 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.
2. 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.

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 sources for these three variables and report the complete URLs. Explain why you chose each data source.

appendix

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

U.S. Industry Data

The Economic Census

Annual Survey of Manufactures

Current Industrial Reports

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

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, a 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: If you use the most advanced statistical techniques but your data are flawed, 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.

Developing a Search Strategy for Finding Your Data

When you search for data, just as when you do your literature survey, it's a good idea to start with a search strategy. Time spent up front in this activity will almost always pay off by significantly reducing the effort required to find and collect the data you need. The process for developing a search strategy is outlined in Figure 9.1.

Step 1: Before You Search

At least three issues are involved in the construction of a good data set. The first issue is sample size—you need to have enough data points to derive statistically valid empirical test results. You typically need thirty data observations to ensure adequate degrees of freedom, though you can sometimes make do with less. The second issue is that of a random or representative sample. We touched on these two issues in the previous chapter, but we will discuss them in detail in Chapter 10.

The third issue is obtaining data that correctly measure the concepts that your theory deems important. Begin by reviewing your conceptual or theoretical model. What variables do you need data for to test your research hypothesis? If you could obtain the ideal data, what variables would they include? At the highest level of abstraction, these questions should be straightforward to answer. If not, you may need to rethink your research hypothesis to obtain more focus.

Next, come a series of practical questions: What are the different forms in which those data might be available? There are several aspects to this

Figure 9.1 Developing a Search Strategy for Finding Your Data

Step 1: Before You Start Searching

- What are the desired variables?
- How should each variable be defined?
- What data frequency and sample period or what level of analysis?
- What are potential sources for data on each variable?

Step 2: As You Search

- What data are available?
- Are there suitable proxy variables for variables that are unavailable?
- If not, how can the empirical model be modified to use the data available but still test the hypothesis?

query. The first is conceptual. For example, if your theory requires a measure of income, you might find two types: per capita income and family income. Which is the best measure to test your hypothesis ideally?

The second aspect to this question is whether you should seek time-series or cross-section data. If the former, what frequency of data should you look for: annual, quarterly, monthly, weekly, or daily, and over what sample period? If the latter, what unit of analysis is most appropriate for your hypothesis: average income at the state level or the county-level, or even individual family incomes?

Since there may be a variety of ways to measure the variable you are interested in, you should think carefully about what your variable would be ideally, and then what you can live with.

Once you have determined your list of desired variables, the next step is to think about where those data are likely to be found—for example, by using the sources discussed in Chapter 8. What agency collects that data? What data sources were used by previous studies on your topic?

A good place to start your data search is the website, Resources for Economists on the Internet (RFE, http://rfe.org). RFE is an extensive source of a variety of information useful for economists. You may recall that we introduced it in Chapter 3 as part of our discussion on surveying the literature. RFE provides an extensive and annotated list of data sites (including a number of more specialized sources omitted from our discussion in Chapter 8). The annotations are particularly useful since they help you determine if a data site will be useful for your research, without your having to dig into it yourself. RFE's data collection is divided into five sections: U.S. macro and regional data, other U.S. data (e.g., data for microeconomic research), world and non-U.S. data, data from financial markets, and data from published journal articles.

Another useful resource for beginning a data search is the website, World Wide Web Resources in Economics or WebEc (http://netec.wustl.edu/WebEc). This site provides information similar to that of RFE, but with more of an international focus.

Once you have reviewed RFE and/or WebEc, you should have some good ideas about where to look for the data you need. These are likely to include both primary and secondary data sources. You may wish to review Appendix 8A, which outlined the primary and secondary data sources described in Chapter 8.

Step 2: As You Search

If you have completed Step 1, you now have a list of ideal variables for testing your hypothesis. You have also identified several promising data sources, such as Economagic and FRED II. The next step is to begin the search.

As you investigate each data source, you need to ask several questions:

1. **What data are, in fact, available?** Remember that this question involves both correctly defined variables and a sufficient number of data points.
2. **If the data are not the ideal, are they good enough?** In practice, it is uncommon for the ideal data to exist with a sufficient number of observations. After all, the folks who produced the data probably didn't have your exact research in mind. You might find the exact variable you desire but too few data points. Or you might find sufficient data points but for a variable that is not defined exactly as you'd like.
3. **If the data are not acceptable, is there an available proxy that is?** A proxy is a variable that should behave roughly the same as your theoretical variable. For example, it should have the same trends and turning points as the theoretical variable, though it isn't conceptually identical.
4. **If there is no adequate proxy, how can the hypothesis be reformulated to make it testable, given the data available?**

Let's illustrate these issues with a series of examples. Suppose you would like to test the Keynesian consumption function. The theory hypothesizes that consumer spending in the aggregate is a linear function of national income or GDP.

$$C = a + b * GDP$$

If you wanted to test this theory for the U.S. economy using annual or quarterly data, you would have no difficulty obtaining a large number of observations for both C and GDP. Macro data for the U.S. economy are among the best data in the world; they are available in great detail over long time frames.

But suppose you wanted to apply this theory to a foreign country. Such data are often more difficult to obtain, if they are obtainable at all. One complication, especially when doing international comparisons, is that concepts may be measured differently in different countries. For example, the unemployment rate in Japan is not defined in the same way as in U.S. data. Additionally, in many countries published macroeconomic data are thought to be politically manipulated. Thus, the published data, even if available, may not correctly measure the underlying theoretical concepts.

Consider another example closer to home. Suppose you wished to test the Keynesian consumption function over a monthly time frame. Data on personal consumption expenditure and GDP are available in annual or

quarterly frequency, but not monthly. Thus, you would need to find proxy variables in order to do a monthly study. Recall what makes a good proxy. Either the proxy should measure the same underlying economic behavior as the theoretical variable you wish to study, or it should be highly correlated in a statistical sense with the theoretical variable. For example, retail sales are published monthly, and they measure approximately the same thing as consumption expenditure. Finding an adequate proxy for GDP is more problematic. One candidate might be the Fed's index of industrial production, which is available monthly. Industry production measures the manufacturing output of the economy. In the early twentieth century, when the U.S. economy was largely industrial, this would have made a good proxy for GDP. In recent years, however, manufacturing has declined to less than 40 percent of GDP, while the service sector is roughly 60 percent. To see if industrial production is an acceptable statistical proxy, one should examine the correlation between the two variables over the closest frequency available, in this case, quarterly. Examining data for quarterly real GDP and quarterly industrial production indicates a very high correlation of 0.96. Thus, industrial production probably is an acceptable proxy.[1]

These problems are often more acute when you are doing a microeconomic study. For these, you will more likely need to use variables from different sources, and perhaps different definitions. It is also more likely that different variables may come from different populations. Suppose your theory identifies weekly work hours by couples with children between the ages of six and seventeen. The available data may include weekly work hours by gender nationwide as well as employment rates among males and females in couples with children. You can construct a data series that approximates the theoretical concept, by making suitable assumptions. For this example, you could assume that males and females with children have, if employed, the same weekly work hours as males and females in the general population. You still need to be concerned about how well the resulting series serves as a proxy for the desired theoretical variable. Ultimately, to perform the analysis at all, you may have to hope for the best and use the data, but acknowledge its potential weaknesses.

Oftentimes, researchers must simplify their theoretical model because they lack adequate data on the variables they need. Either there are insufficient data points for one or more variables, or the specific variable they need is not available. In our example of the monthly consumption function, if a proxy for monthly GDP cannot be found, you may need to change the study to examine a quarterly consumption function.

Figure 9.2 Different Forms of Data

- Levels
- Per capita
- Changes
- Rates of Change (or Growth Rates)
- Annualized Growth Rates
- Proportions
- Nominal
- Real
- Index Numbers

Data Manipulation

Data for any variable are available in a number of different forms. The most important of these are listed in Figure 9.2. In general, it is possible to convert a variable from one form into another. In this section we summarize some commonly used techniques for performing those conversions. Even if you can find data in the exact form you need so you don't need to perform these calculations, it is nonetheless important that you understand what each form means and how they relate to each other.

Level of Variable

The most basic form of any variable is called the **level.** The level is the actual value or size of the variable being measured. For example, Figure 9.3 shows that the level of U.S. GDP in 2001 was $10,082 billion.

Often, researchers prefer to use the **per capita** form of a variable. Per capita means per person. It is found by dividing the level of the variable by the appropriate population. For example, since the U.S. population in 2001 was 285.318 million, U.S. GDP per capita in 2001 was $10,082

Figure 9.3 Levels, Changes, and Growth Rates of U.S. GDP (in billions of $)

Year	GDP	Change in GDP	Percentage Change
1999	$9,274.3		
2000	$9,824.6	$550.3	5.9%
2001	$10,082.2	$257.6	2.6%
2002	$10,446.2	$364.0	3.6%

NOTES FOR NOVICE RESEARCHERS

Pay Attention to the Units!

Novice researchers often pay little or no attention to the units in which variables are measured. This is a mistake you should avoid! In our previous example of GDP per capita, GDP is measured in billions of dollars, while the population is measured in millions of people. Dividing the two yields units of thousands of dollars per person. Note that it's not enough to know that GDP is measured in dollars—you should know whether it's trillions of dollars or millions of dollars. Suppose you ignored those zeros in GDP and population, and only divided the digits: 10,082 /285.318. You would have obtained an answer of $35.37 per capita. So if you think that the number of zeroes in statistics doesn't matter, ask yourself whether you'd prefer to have an annual income of $35,000 or $35.

billion/285.318 million = $10,082,000 million/285.318 million = $35,336 per person.

One common use of the per capita form of GDP is to compare standards of living between nations. Mexico and Spain have approximately the same size economy: Mexico's 2002 GDP was about $900 billion, while Spain's was $828 billion, so Mexico's was slightly larger. However, Mexico's population is substantially larger than Spain's, so Spain's standard of living is much higher. Mexico's 2002 GDP per capita is about $9,000 per person; by contrast, Spain's 2002 GDP per capita is more than double that, at about $20,700 per person. For similar reasons, when you are performing other types of comparisons, you might use the per family or per household form of income or some other variable.

Change in Variable

For some purposes, researchers find it more useful to examine the **change** in a variable than the level. For example, knowing that the level of U.S. GDP in 2001 was $10,082 billion may be less important than knowing that it increased by $258 billion between 2000 and 2001. In other words, the change in a variable is just the difference in the levels from one time period to another.

Often, it is important to know whether a change was meaningful. Was the increase in GDP between 2000 and 2001 large or small? One way to assess

NOTES FOR NOVICE RESEARCHERS

Caution on International Comparisons

Each country measures its economic data, such as GDP, in its own national currency. In other words, Mexico's GDP is measured in pesos, while Spain's GDP is measured in pesetas. In order to make cross-country comparisons, as we did earlier, the data need to be converted into a common unit of account, a common denominator. Either both GDPs need to be in pesos, both in pesetas, or both in some third currency, like dollars. This conversion can be done in two ways. The most common way is to use an exchange rate, that is, the price of one currency expressed in terms of another. Thus, one could take Spain's GDP (in pesetas) and multiply by the U.S.$-peseta exchange rate to obtain Spain's GDP in terms of dollars. Going through the same process for Mexico, we would multiply its GDP in pesos by the U.S.$-peso exchange rate to yield Mexican GDP in terms of dollars. Then the two GDPs could be compared.

In practice, however, this is not the best way to convert Mexican and Spanish GDP. This is because exchange rates fluctuate regularly, sometimes dramatically in a short period of time, which makes it appear that a country's GDP has increased or decreased when really it has not. For many purposes, then, such as comparing two nations' standards of living, it is preferable to use *purchasing power parity equivalents* (or PPP Equivalent) instead of exchange rates to convert the GDP figures. Indeed, the figures we just cited were computed with this method. The PPP Equivalent is a measure of the relative purchasing power of each country's currency.[2] Essentially, it extracts the volatility from exchange rates to yield a more accurate, long-term measure of the value of a currency in terms of another. Whenever possible, you should use the purchasing power equivalent version of economic data when comparing across countries. This is usually labeled "purchasing power parity $."

Note also that direct comparisons of GDP or income per capita can be misleading when you are making comparisons between economies that have different economic systems or levels of economic development. For example, the reported income per capita of rural residents in a country tends to undervalue their actual standard of living, since they often have unreported income-in-kind such as

produce they grow for their personal consumption. In other words, comparing reported incomes per capita between urban and rural residents would be biased in favor of the former group. The same problem occurs when comparing incomes per capita between developed and developing nations.

this is to look at the **rate of change,** also called the **percentage change** or the **growth rate.** The rate of change is the change as a percentage of the original level. To compute the rate of change for some variable X between an initial period 0 and a subsequent period 1, you can use the following formula:

$$G_{1,0} = \{[X_1/X_0] - 1\} * 100$$

where X_0 is the level (or value) of X in the initial period, X_1 is the level of X in the subsequent period, and $G_{1,0}$ is the rate of change or growth rate between periods 0 and 1.

For example, the growth rate of GDP from 2000 to 2001 was

$$\{[\$10{,}082.2 \text{ billion}/\$9{,}824.6 \text{ billion}] - 1\} * 100$$
$$= \{[1.0262] - 1\} * 100$$
$$= \{.0262\} * 100$$
$$= 2.62\%$$

Growth rates are especially useful for comparing concepts that are different orders of magnitude, like the annual sales of Microsoft versus the annual sales of Joe's Computer Software, Inc.

By convention, researchers usually compare data in terms of annual growth rates, even if the period of time they are examining is more or less than a year. The reason for this is so they have a common standard for comparison. Growth of GDP over a decade is almost certainly greater than growth over a shorter period, but it isn't clear that GDP is growing at a faster or slower rate over the decade until you standardize the two time periods examined. For periods longer than a year, the overall growth rate is simply divided by the number of years involved. Thus, if GDP grew 20 percent over ten years, the average annual growth rate would be 20/10 = 2% per year.

For periods of time shorter than a year, like a month or a quarter, researchers usually follow a different procedure. They ask the question, "If growth was X% for some period less than a year, what would the cumulative rate be if that growth rate continued for an entire year?" This is called an **annualized growth rate.** You might think that you could just multiply

Figure 9.4 Annualized Growth Rates

Beginning:	Quarter I	Quarter II	Quarter III	Quarter IV
	= 1.10*100	=1.10*110	=1.10*121	=1.10*133.1
Sales:				
$100	$110	$121	$133.1	$146.4
				46.4%

quarterly data by four or monthly data by twelve to get annual rates. That would in fact provide a rough approximation, but only a rough one because it ignores the effects of compounding. (You may recall the way that interest on a bank account compounds. Not only do you earn interest on your deposit, but also you earn interest on the interest. That's compounding.)

In the same way, suppose that a company's sales were growing at 10 percent during the first quarter of the year. What would that amount to at an annual rate, assuming that the 10 percent growth rate continued for three more quarters? Figure 9.4 illustrates this question. Suppose the level of sales was $100 at the beginning of the year. At the end of the first quarter, sales would be 10 percent higher or $110. At the end of the second quarter, sales would be 10 percent higher than that, or $111. By the end of the fourth quarter, sales would be 46.4 percent higher than when the year started. In this case, since growth compounded over four quarters, the annualized growth rate (46.4%) is 6.4 percent more than the rough approximation ($4 * 10\% = 40\%$). If you were using monthly data instead of quarterly, the annualized growth rate would be even more than the approximation, since the compounding occurs twelve times. In other words, the more frequently the compounding is calculated, the greater the error in the approximation.

To compute an annualized growth rate for quarterly data, use the following formula:

$$G_q = \{[X_1/X_0]^4 - 1\} * 100$$

where X_0 is the level of X in the initial quarter, X_1 is the level of X in the next quarter, and G_q is the annualized rate of change or growth rate. The formula for annualizing monthly data is the same except that instead of quarterly values of X you would use monthly values. In addition, instead of raising the ratio of quarterly values to the fourth power, you would raise the ratio of monthly values to the twelfth power:

$$G_q = \{[X_1/X_0]^{12} - 1\} * 100$$

Figure 9.5 Proportions of 2002 U.S. GDP

(Values for GDP and components are in billions of $)

GDP	C	I	G	X-M
$10446.2	$7303.7	$1593.2	$1972.9	−$423.6
	=7303.7/	=1593.2/	=1972.9/	=423.6/
	10446.2	10446.2	10446.2	10446.2
	69.9%	15.3%	18.9%	−4.1%

A form of data similar to a growth rate is a **proportion.** A proportion is also called a share or a **percentage** or a fraction. The idea behind proportions is simple: imagine dividing the economic pie into several parts. Each part is a proportion. Thus, labor income (i.e., wages, salaries, etc.) is roughly 75 percent of national income. Similarly, capital income (i.e., interest, rentals, etc.) is roughly 25 percent of national income. A proportion is measured as a fraction or percentage of the whole. Consider the example in Figure 9.5. Any student of macroeconomics knows that GDP can be divided into consumption expenditure, investment expenditure, government expenditure, and net export expenditure. Each of these fractions is a proportion. The proportion of GDP that is consumption expenditure is computed as the level of consumption expenditure divided by the level of GDP, or 69.9 percent.

Proportions are useful for understanding a wide range of issues in economics. For example, Americans are thought to have a low saving rate. The saving rate is just the proportion of income that is saved. Is the amount of poverty in America growing or shrinking? One way to assess that is by looking at the poverty rate, which is the proportion of Americans with income below the poverty line.

Real versus Nominal Magnitudes

A great deal of the available economic data are based on a simple identity:

$$V = P \times Q$$

where V = nominal (or value)
P = price, and
Q = real (or volume).

Nominal data are data measured by using the actual market prices that existed during the time period in question. For example, the value of a firm's output in 1999 is equal to the product of the price the firm charged

in 1999 and the quantity of output it produced in that year. Similarly, nominal GDP in 1999 is the sum of the price times the quantity of every product produced in a nation in 1999. Nominal data are easy to interpret.

Real data are a bit more difficult to understand. At the micro level, real data refer to the actual quantities employed by a firm (e.g., labor hours), produced by a firm (e.g., number of widgets), or sold by a firm (i.e., sales volume). In our example, the firm's real output is the actual quantity of goods or services that it produced.

The story becomes more complex, however, when we move above the level of an individual firm to a more macro perspective. How does one measure real GDP, the actual amounts of goods and services produced by a nation? One could create a list of all the products produced in every industry. But aside from being cumbersome, such a real measure of GDP would make it difficult or impossible to compare GDP from one year to the next. If last year a nation produced 50,000 automobiles and 17,000 new homes, while this year the figures were 52,000 and 14,000, did GDP rise or fall? The solution economists have found to deal with this quandary is to define real GDP as a weighted average of all the goods and services produced, where the weights employed are the prices of each product in a given year.[3] In other words, money provides the common denominator by which to measure diverse goods and services produced in a year.[4]

Index Numbers

Real GDP, as just defined, is an example of what economists call an **index number,** or more specifically, a **quantity index,** since it measures the average quantity of something. The other type of index number is a **price index.** These are more familiar to noneconomists. For example, most people have heard of the consumer price index. What is a price index? A price index is (roughly speaking) a weighted average of the prices of some type of good or service, where the weights are the corresponding quantities in a given year.[5] One could construct an energy price index (the average price of different energy sources), a food price index (the average price of different foods), or a consumer price index (the average price of goods and services consumers typically buy). An index number is a data series that begins at 100 and typically increases in subsequent years, though not necessarily continuously. For example, Figure 9.6 gives the consumer price index for 1984–1993.

How does one interpret an index number? Suppose you learn that in July 2003 the consumer price index had a value of 183.9. What exactly does that mean? The short answer is that 183.9 doesn't mean anything in isolation. Index numbers, unlike most other statistics, have no units. Rather, index numbers are designed for comparison purposes. For example, one

Figure 9.6 Consumer Price Index, 1984–1993

1984	1985	1986	1987	1988	1989	1990	1991	1992	1993
103.9	107.6	109.6	113.6	118.3	124.0	130.7	136.2	140.4	144.0

could use an index number to compare the level of whatever the index is measuring (for example, consumer prices) to some earlier time period known as the base period. Thus, a CPI of 183.9 means that on average consumer prices have risen 83.9 percent since the early 1980s, which was the base period for these figures.

Index numbers are typically constructed by taking the raw index, that is, the actual weighted average just mentioned, and transforming it using the following formula:

$$X_T = [X_t/X_0] * 100$$

The variables in the formula are defined as follows:

X_t is the value of the raw variable in a given time period t in the series,
X_0 is the value of the raw variable in the base period, which is the period to be compared against, and
X_T is the resulting index number.

Note that in the base period $X_t = X_0$. Therefore, an index number always has a value of 100 in the base period.

Index numbers can also be used to compare the relative growth rates of two similar concepts, such as employment in Colorado and Virginia. They are particularly useful when the concepts have very different magnitudes. Imagine trying to graph employment in your local town and employment in the nation as a whole over the past ten years, both on the same graph. The national figures would be so much larger than the local ones as to make it very difficult to compare them.

To make it easier to see what was going on, you could index both data series to a common starting point, say, ten years ago. What you would be doing is transforming each series to have an arbitrary value of 100 in that year, using the formula for computing index numbers given earlier. Then since each would have a common starting point, you could easily see how employment growth in your town compared to the national economy.

Quantity Indices Versus Real Quantities

There are two ways to construct real measurements. The first is to create a quantity index, as described earlier (i.e., a weighted average of the quantities

of the type of good or service one is interested in). The second is to create a price index and then divide the nominal value by the price index. The result is real and follows directly from the identity above: Q = V/P. The data for real GDP in the U.S. National Income and Product Accounts are created using the second approach.

Price Indices Versus Implicit Price Deflators

There are similarly two ways to construct price measurements. The first is to create a price index as described earlier (i.e., a weighted average of the prices of the type of good or service one is interested in). The second is to create a quantity index and then divide the nominal value by the quantity index. The result is called an **implicit price deflator,** since it is the price index implied by the identity, which is expressed in this case as P = V/Q. Note that published price indices are multiplied by one hundred.

You might think that a price index and the corresponding implicit price deflator would be the same. However, though we tend to use them equivalently, only under extreme technical assumptions will that be the case.[6]

How Inflation Distorts Nominal Values

Because prices tend to increase over time, comparing nominal measurements can be misleading. For example, if you receive a 10 percent salary increase from one year to next, while at the same time the prices of the things you typically buy experience 10 percent inflation, you will not enjoy an increase in your standard of living equivalent to your nominal raise. In fact, your living standard, or more precisely your real income, will not have changed at all. This is the concept behind another way to describe real measures: as nominal measurements adjusted (or corrected) for the effects of inflation.

The way this "inflation adjustment" is done is by recalling that GDP is a weighted average of Q, where the Ps are prices in a specific period, called the base year. By using base-year prices but actual-year quantities, real GDP excludes the effects of changing prices over time. (Note that if relative prices change significantly, that can cause a bias in the measure.)

Rebasing Data Series

Periodically, those who collect the data change the base year. One reason they do this involves how one interprets the base year from the point of view of the index number. That is, the base year is the period against which the index measures changes in prices or quantities. We care more about recent economic changes than historic ones, so the base year is changed to keep it "recent." A second reason for changing the base year is to correct the bias mentioned in the last section.

NOTES FOR NOVICE RESEARCHERS

Practical Data Tip

Any nominal variable can be transformed into its corresponding real magnitude by dividing by an appropriate price index: Q = V/P. An easy way to see this is by examining Tables B-1, B-2, and B-3 of the *Economic Report of the President,* which show nominal GDP, real GDP, and the GDP deflator, respectively.[7]

Similarly, if one has data for nominal and real (for example, export sales and export volumes), one can derive a price index: P = V/Q . Note that such an index is called a **unit value.**

Re-basing a data series can cause difficulties for researchers. This is particularly a problem with print data sources. When you compile data from print sources, you should always start with the most recent source and work backward. This will enable you to capture revisions in the data that were not available until a subsequent date. For example, suppose you are compiling an annual data series, and each year your source reports the estimated figure for the current year, and also revised data for the previous two years. If you start from the earliest source and go forward, your data series will consist of the earliest and least reliable estimates of the variable.

Suppose you have a time series of annual CPI data in which the first half of the data uses a base year of 1992, while the second half uses a base year of 1997. There is no perfect way to link the two series! Every way imposes some bias, but that is the nature of working with real-world data. To link the two parts of the series it is necessary to have some overlap, as shown in Figure 9.7 for the year 1995. The first row shows the source data with a base year of 1997. (This should be clear since the 1997 value for the data is 100.) The second row shows the source data with a base year of 1992. Notice in particular that if we merely combined the two series there would be a noticeable jump where the two series connect. In short, the data would show an incorrect change in the CPI in that time period. For example, if you combined the base 1992 data through 1994 with the base 1997 data beginning in 1995, you would see a large apparent decrease in the CPI between 1994 (i.e., 1.30) and 1995 (i.e., 0.85). This would indicate deflation, when in fact prices are rising!

Figure 9.7 Rebasing a Data Series

Year:	1991	1992	1993	1994	1995	1996	1997	1998
Base Year 1997					0.85	0.95	**1.00**	1.15
Base Year 1992	0.95	**1.00**	1.25	1.30	1.40			
Linked Series	0.58	0.61	0.76	0.79	0.85	0.95	**1.00**	1.15

Suppose we want a complete series with a 1997 base year. We will need to transform the values for the observations that are only available with a base year of 1992 so they correctly show the change in CPI between both parts of the source data. The data with the earlier base year need to be reduced in value to correctly link to the data with the later base year. The amount of the reduction is given by the ratio of the two values for 1995, the period of overlap. In other words, dividing the base-year 1997 value into the base-year 1992 value yields the reduction factor: $85/140 = 0.607$.

To obtain the linked series with a 1997 base year, use this factor to multiply the value of each observation with base-year 1992. The result is shown in the third row of Figure 9.7.[8] Notice that there is no longer a drop in the series where the two parts were linked.

Data Smoothing

Data that are volatile are sometimes "smoothed" to better reveal the underlying trends. Consider the monthly income of a salesperson who works entirely on commission. One month his income might be quite high, while the next it could be low. That variability makes it difficult for the salesperson to tell if he's having a good year or not. The problem is similar for researchers using certain types of data.

There are a variety of techniques for smoothing data. We will discuss two such techniques here: moving averages and seasonal adjustment. A **moving average** replaces the actual data point in each period with an average of the n – 1 preceding data points with the nth. The result is that any "abnormal" observations become less important since they are averaged with more normal ones.

Some variables show seasonal patterns, whereby they change predictably in certain months, quarters, or seasons. For example, consumer spending rises in the fourth quarter of every year in anticipation of Christmas. Often, researchers wish to know if the changes are abnormal. Thus, one might ask, did consumers spend more or less than normal this Christmas season? This is a question for which **seasonally adjusted** data are designed. The U.S. Census Bureau has developed techniques that are commonly used to seasonally adjust data. The most recent version is called X12. This version of the earlier X11 technique is widely available on common statistical packages or it can be obtained free from the Census Bureau. Note that many macro data series are only available in seasonably adjusted form. These include GDP, CPI, PPI, retail sales, and housing starts. It is unlikely that you will need to do your own seasonal adjustment of data, but it is important that you understand what the term means. Note the following abbreviations, which are often used with data sources:

NSA: Not seasonally adjusted.
SA: Seasonally adjusted.
SAAR: Seasonally adjusted annual rate.

Appendix 11A discusses some additional ways to manipulate data. Whatever data manipulation you do, once you have your completed data set it is good practice to review the data. Examine each data series by looking at either data tables or charts. Look for observations that appear to be errors. This is easiest with time-series data that show a trend. A good example is the CPI, as shown earlier. Just because data is published does not mean that is error free. Moreover, if you performed any data entry, there is always the possibility you made a keystroke error. For cross-section data you should compute summary descriptive statistics: mean, range, and standard deviation for each variable. You can report these as part of your data appendix. This is not generally done for time-series data, which often contain a time trend. Since the purpose of the descriptive statistics is to give you a rough sense of what the data are like, with time-series data one would be likely to discover that the lowest data point was the earliest, while the highest data point was the latest. With a time trend, you already know that.

Only when you have examined the data and satisfied yourself that they are as accurate as they can be are you ready to begin statistical testing.

Constructing a Data Appendix to Your Research

It is good scientific practice to make your data available so that other researchers may replicate your work. At a minimum, this means explaining the "Sources and Methods" you used in compiling your data set. We can categorize data as **raw (or source) data** or **derived data.** Raw data are the data as you found it at the source. Derived data are the data that you actually used in your empirical testing. This distinction is important because it is not uncommon for the raw data you obtain to be in a form different

from what you need to test your hypothesis. For example, the data available from the source might be the consumer price index, while the data you need might be the inflation rate, that is, the percentage change in CPI. Thus, you may need to manipulate or mathematically transform the data before using it. In this example, the CPI would be the raw data, while the inflation rate would be the derived data.

In explaining your sources and methods you should provide clear citations for exactly which sources provided the raw data, as well as a complete explanation of how you manipulated the raw data to put it in the form you actually used. After reading your documentation, the reader should be able to replicate your exact data with a minimum of digging. Better yet, you should provide a copy of the exact data you used, in addition to documenting where it came from. A number of journals require that authors include copies of their data when they submit their research papers for publication.[9] This copy of your data with documentation is called a data appendix.

SUMMARY

- Since the quality of your research will be no better than the quality of your data set, you should put a substantial effort into constructing a good quality data set.
- To efficiently locate data you should begin with a search strategy. Begin by creating a list of the desired variables for your research. Next, consider likely sources for that data. Think about what proxies might be available for data that are unavailable. Finally, consider how the model might be revised to use the available data.
- Data are available in different forms, including levels, per capita, changes, rates of change, and proportions.
- Growth rates for periods of less than a year are often given in annualized form.
- Nominal data are distorted by inflation; real data are corrected for inflation.
- An index number is essentially a weighted average of whatever the number purports to measure. Thus, the consumer price index (CPI) is a weighted average of the prices of goods and services consumers typically buy. Index numbers are designed to show changes in prices or quantities since some arbitrary base period.
- A data appendix provides the exact data the researcher used as well as a complete description of the sources and methods he or she used to obtain it.

NOTES

1. Monthly values of industrial production were averaged to obtain a quarterly series. Real GDP is available quarterly. Correlation will be explained in detail in Chapter 10.
2. The purchasing power parity (PPP) equivalent gets its name from the fact that it measures currencies in terms of the goods they can buy. PPP is based on the notion that in equilibrium the same product should cost the same amount in two different currencies when adjusted by the exchange rate. This should be true in equilibrium; if these costs are not the same, it would be profitable to export from where it is cheaper and import to where it is more expensive until the costs are equal. Therefore, PPP is a long-term equilibrium measure of the exchange rate.
3. Thoughtful readers may notice that this is the same value as nominal GDP. This point will be clarified shortly.
4. This is exactly what is meant by characterizing money as "a unit of account."
5. For some price indices the weights are budget shares rather than quantities; in other words, the weight of some consumer product price is the proportion of that item in the average consumer's budget.
6. For similar reasons, a quantity index and the corresponding real quantity are rarely equal, though this is easier to see because the units are different. Real quantities are measured in constant (e.g., 1992 dollars), while quantity indices like all index numbers are unit free, that is, they are numbers like 100, 117, 129. This will be explained in more detail later in this chapter.
7. Note, however, that for the math to work out correctly the data for the GDP deflator must be divided by 100.
8. The careful reader will notice that this procedure for rebasing index numbers is a variation on the formula for creating index numbers presented earlier.
9. Recall that the last section of data in the Resources for Economists website was data from published journal articles.

SUGGESTIONS FOR FURTHER READING

DataBasics, Federal Reserve Bank of Dallas website, http://dallasfed.org/data/basics/index.html—A concise summary of the major techniques for manipulating data.

Wyrick (1994)—Offers explanations of a variety of data manipulation techniques.

EXERCISES

1. Suppose you are researching the question of why the U.S. unemployment rate for African Americans is higher than the U.S. average unemployment rate. Create a list of variables that you think might explain that question. For each variable, propose a source for that data. Next, visit each source, and describe the data you found. For example, if there are time-series data, what frequency is available and for what time period? If there are cross-section data, what level of analysis is available? Draw a conclusion as to whether the data is adequate to test the hypothesis empirically.
2. Go to the Bureau of Economic Analysis Fixed Assets webpage at **http://www.bea.doc.gov/bea/dn/faweb/AllFATables.asp#S1.** Download data for Current-Cost Net Stock of Fixed Assets and Consumer Durable Goods (Table 1.1) for the past ten years available. Open the downloaded file in a spreadsheet. Compute the year-to-year growth rate for nonresidential fixed assets. This is the capital stock of private business in the United States, in other words, K in the standard production function Q = f[L, K, t]. Is there a trend in the growth rates? Are they increasing, decreasing, or staying about the same?
3. Go to the Census Bureau's webpage on U.S. Trade Balances by Country at **http://www.census.gov/foreign-trade/balance/index.html.** What were U.S. imports from Mexico, Canada, Japan, and China in 2001? Next, go to **http://www.census.gov/foreign-trade/statistics/historical/gands.txt** to obtain total U.S. imports in 2001. Finally, compute the proportion of total U.S. imports in 2001 that came from Mexico, Canada, Japan, and China.
4. Collect data on quarterly nominal U.S. GDP for the past five years available. One source of this data is the Bureau of Economic Analysis's webpage at **http://www.bea.doc.gov/bea/dn/nipaweb/TableViewFixed.asp?SelectedTable=3&FirstYear=2002&LastYear=2003&Freq=Qtr.** Next, collect data for the U.S. implicit price deflator for GDP for the same time period, at the same webpage. Now, deflate the nominal GDP using the GDP deflator to get real GDP. Please show your work. Finally, to check your work, collect data for real GDP at the same BEA webpage. Compare your computed real GDP with the data you downloaded.
5. For each of the variables used in Exercise 1, provide complete citation information on how you obtained the data.

appendix

What Are Chained-Dollars?

In the chapter, we defined an index number as a weighted average. For example, a quantity index for real GDP would be a weighted average of the quantities of all products in GDP, where the weights are the prices of those products. To be more precise, most index numbers are constructed as the *ratio* of two weighted averages. In other words, a quantity index for real GDP would be the ratio of average production in the current period to average production in the base period. For a quantity index, the natural choice for weights is the corresponding prices, but which prices? Traditionally, there were two choices: the base period prices or the current period prices. If the former was chosen, the result was called a *Laspeyres index;* if the latter was chosen the result was called a *Paasche index.* For example, the consumer price index is computed using base period quantities as weights; thus, the CPI is a Laspeyres index. Each of these choices introduces a bias into the results. The bias of the Laspeyres index is in the opposite direction of the bias in the Paasche index. In recent years, a new weighting scheme has been introduced by the U.S. Bureau of Economic Analysis to deal with this issue. The result is called a *chain-type index,* which is defined as the geometric mean of the Paasche and Laspeyres indices. Note that unlike the fixed weights of a Laspeyres or Paasche index, for a chain-type index the weights change for every observation.

appendix

Example of a Data Appendix

Military Enlistment = f(Unemployment Rate, Military/Civilian Pay Ratio, War Status)

Air Force, Army, Marines, Navy—Number of active duty enlisted soldiers in each of the respective branches of the U.S. military

U-Rate—The national Unemployment Rate averaged from monthly observations to yield annual rates

Pay Ratio—Ratio of military to civilian pay. Military Pay is measured as the monthly basic pay for active duty enlisted men and women of classification E-3 with two or fewer years of service. These monthly observations were divided by 160 to represent a 40-hour workweek. Civilian Pay is defined as the hourly earnings of civilian production workers. Monthly observations were averaged to produce annual averages.

War—This is a dummy variable with 1's for the years during which the United States had a major foreign conflict.

YEAR	AIR FORCE	ARMY	MARINES	NAVY	U-RATE	PAY RATIO	WAR
1981	570,302	781,419	190,620	540,219	7.616667	0.540545	0
1982	582,845	780,391	192,380	552,996	9.708333	0.531489	0
1983	592,044	779,643	194,089	557,573	9.6	0	1
1984	597,125	780,180	196,214	564,638	7.50833	0.512308	1
1985	601,515	780,787	198,025	570,705	7.191667	0.517612	0
1986	608,199	780,980	198,814	581,119	7	0.521581	0
1987	607,035	780,815	199,525	586,842	6.175	0.524918	0
1988	576,446	771,847	197,350	592,570	5.491667	0.518359	0
1989	570,880	769,741	196,956	592,652	5.258333	0.51926	1
1990	535,233	732,403	196,652	579,417	5.616667	0.517419	1
1991	510,432	710,821	194,040	570,262	6.85	0.522679	1
1992	470,315	610,450	184,529	541,886	7.491667	0.531482	1
1993	444,351	572,423	178,379	509,950	6.908333	0.537681	1
1994	426,327	541,343	174,158	468,662	6.1	0.535501	1
1995	400,409	508,559	174,639	434,617	5.591667	0.534311	0
1996	389,001	491,103	174,883	416,735	5.408333	0.529458	0
1997	377,385	491,707	173,906	395,564	4.941667	0.52527	0
1998	367,470	483,880	173,142	382,338	4.5	0.518798	0
1999	360,590	479,426	172,641	373,046	4.216667	0.518653	1
2000	355,654	482,170	173,321	373,193	3.975	0.522991	0
2001	353,571	480,801	172,934	377,810	4.758333	0.522497	1
2002	368,251	486,542	173,733	385,051	5.775	0.544941	1

Data Sources:

Bureau of Labor Statistics, *Data Series LNS 14000000: Seasonally Adjusted National Unemployment Rate,* http://data.bls.gov/PDQ/servlet/SurveyOutputServlet?data_tool=latest_numbers&series_id=LNS14000000 (accessed November 15, 2004).
Used for **U-Rate**

Bureau of Labor Statistics. *Data Series CEU0500000006: Average Hourly Earnings of Production Workers,* http://data.bls.gov/PDQ/servlet/SurveyOutputServlet;jsessionid=f0307d7548c5$3F7$3Ft (accessed November 15, 2004).
Used for **Pay Ratio**

Defense and Accounting Service. *Military Pay Prior Rates,* http://www.dfas.mil/money/milpay/priorpay/ (accessed November 16, 2004).
Used for **Pay Ratio**

Department of Defense. *DoD Active Duty Military Personnel Strength Levels, Fiscal Years 1950–2002,* http://www.dior.whs.mil/mmid/military/ms9.pdf (accessed November 12, 2004).
Used for **Air Force, Army, Marines, Navy**

Military.com. 2004. *Military.com's Military History: The Complete List,* http://www.military.com/Resources/HistorySubmittedFileView?file=history_completelist.htm (accessed November 12, 2004).
Used for **War**

Source: *Data appendix courtesy of David Hutchinson.*