EC 313 Spring 2000 Part A

Christopher F. Baum

January 19, 2000

Introduction

What is computational economics?

To quote `Hans Amman's` 1997 editorial in the journal *Computational Economics*, "Computational Economics is a new methodology for solving economic problems with the help of computing machinery. Hence, Computational Economics is not restricted to a specific branch of economics. The only restriction we have to make is that this new methodology has a value added in terms of (economic) problem solving."

Within the last three to five years, two sizable changes have taken place in economic research, both of which have computational aspects.

- The World Wide Web access available on every researcher's desktop has the potential and unprecedented capability to create a "virtual research community." Researchers' awareness of colleagues' areas of interest and expertise, and educators' access to their recent findings can be dramatically enhanced in such a community. It is quite feasible for collaborators on different continents to work together and develop their joint research–even if they rarely meet face to face! In fact, collaborators separated by several time zones may be able to work together very efficiently, handing off their work for review by their coauthor at the end of the day, local time. The ability to produce and disseminate research materials via the Web has greatly enhanced the potential for communication of scholarly findings. At the same time, developments in "metadata archiving" technology have made it possible to generate subject-oriented, searchable collections of those materials without reliance on a central archive.

- The advent of ubiquitous high-performance desktop computers and the availability of shared workstations that are as powerful as latter-day supercomputers, coupled with accessible programming languages, have made it feasible for many researchers to use computation to address issues, simulate models, or estimate parameters of interest by "bootstrapping" techniques, using computing power that was inordinately costly and scarce a decade ago. Likewise, new programming languages provide researchers with the tools to not only perform computations and visualize the results, but also to produce reusable software components, or "building blocks", that can be later reassembled to address a slightly different question.

Outline and purpose of the course

What, then, does a survey course in computational economics entail?

In the first section of the course, we will discuss some of the mechanisms by which economic researchers are better able to locate materials of interest, and share the materials they produce

with others. This will include surveying some of the major information resources in economics. We will then discuss a major initiative for the sharing of economic information: the RePEc database, and the IDEAS service built upon it, and explore the tools used to produce these metadata. The first section will end with a discussion of statistical databases, with a case study of a cross-country database being accessed via SQL (structured query language) and the PHP3 web interface.

The second section of the course will discuss programming languages in economics. Recent trends in the development of computational tools have made more and more powerful languages accessible to a broad audience, and greatly reduced the "startup cost" of becoming proficient in a programming language that can be used to perform economic research. In that context, we will discuss traditional "low-level" languages, the evolution toward high-level languages, and the advent of matrix languages, database languages, and general-purpose languages such as Mathematica.

In the third and largest section of the course, we will develop Mathematica programming skills and apply them to a broad variety of economic and financial problems. The tools of Mathematica programming: in particular, the emphasis on functional programming rather than procedural programming, will be shown to have general use in other contexts, and provide a paradigm in which a computer language may be used to address any issue of interest, from the most mundane housekeeping or data-organizational details to sophisticated applications such Monte Carlo simulation, agent-based computation, option valuation, or the simulation of dynamic and stochastic models. The course will conclude with selected applications of computational economics using Mathematica.

You will be responsible for using several computational tools in the preparation and documentation of a computational research project on an economic or financial topic of your choice. The project will entail preparing a set of web pages on your `www2.bc.edu` webspace describing the issue at hand, with hyperlinks to relevant literature and methodology, as well as a Mathematica notebook containing a full description of the issue, model or technique being considered (using appropriate mathematical notation), as well as the Mathematica code to execute and, where appropriate, graph and/or tabulate results. The grade on the research project will be based on both the economic importance of the topic and the quality of your documentation, evaluation, and program illustrating the topic.

Access to Economic Information

Economics Information Resources

You should become familiar with the BC Economics Department's `Economics Information Resources` page. This page, which is frequently updated, contains links to:

- general economic information: economics resources on the Internet, access to the full text of major journals, access to the *EconLit* bibliographic database, links to various Federal agencies' economic information, and links to working paper archives and the Web of Science citation index;

- professional organizations and conference web sites: the American Economic Association, the Econometric Society, the Society for Computational Economics and others;

- economics software sources, including the Boston College SSC-IDEAS software component archive, and links to sites for the Stata, RATS, Ox, MATLAB, Mathematica and GAUSS programming languages.

Although this page is no model of organization, it brings together many useful links for economic research and access to software tools. There is a counterpart "Economics Data Resources" page that will be described below. You are encouraged to explore the links, and suggest any additional links to me for inclusion at `email to baum@bc.edu`.

Resources for Economists on the Internet

The single most useful site for economists seeking information, data, or links to published materials is `Bill Goffe's Resources for Economists on the Internet`. This pioneering collection of economic information has been steadily enhanced and improved by the suggestions of many researchers, and includes a Yahoo-style hierarchical organization of a vast quantity of information about the discipline of economics. It is an annotated list of more than 1,000 Internet resources, most of which are described in *RFE*. To quote Goffe, "In selecting these resources, we exercise some editorial judgment and select items that either offer a substantial amount of information, or are specialized to a specific area. A particularly good place to look for a broader array of business and economic resources is WebEc."

WebEc, another attempt to collect and categorize a vast quantity of economics-related information, is available at `http://netec.wustl.edu/WebEc/` with mirrors in Finland, UK, and Japan. WebEc provides a useful list of journals in economics, with links to their home pages where available. Since an increasing number of journals provide online access to at least their table of contents and abstracts pages–and possibly full-text, as noted in *Economics Information Resources*–this is a very useful compendium. The headings of WebEc are categorized in subject fields, broadly similar to the *Journal of Economic Literature* classification scheme, allowing you to narrow your search for information by field.

If you're interested in contacting someone at another academic institution or learning more about their graduate programs, `Christian Zimmermann's EDIRC`, Economics Departments, Institutes and Research Centers in the World is very helpful.

These compendia of economic information are often far more fruitful than general-purpose search engines such as AltaVista, Lycos or Yahoo. They require some effort to locate the appropriate category of information, but will return far fewer extraneous references than a search through the entire Web.

The RePEc concept

The major challenge facing the development of a virtual research community in any field is the necessity of creating a single virtual archive of metadata. For example, the advent of the Web has made it possible for each institution to create a Web-accessible archive of materials, such as working papers and course material. Without a consistent mechanism for locating institutions, departments, and researchers, the mere accessibility of materials does not translate into efficient use, despite the efforts of researchers and librarians to produce metadata on the "most useful sites" in a given field, as described above. With current technology, we can greatly improve upon the status quo, in which the Web contains hundreds of sites related to a certain body of knowledge, but lacks an integrating framework for collection of that information into a readily accessible corpus.

The impediments to creation of that framework are not technical, but economic in nature. There is little incentive for any single institution to bear the cost of establishing such a digital library, cataloging and housing a vast and rapidly-expanding collection of research materials, and providing the technical means for its dissemination throughout the Web. However, since every institution will

benefit from participation in such an effort, we may solve this incentive problem by creating a virtual collection via a network of linked metadata archives. Each institution need only maintain their own collection of metadata describing research and instructional materials using a set of standardized templates–a modest and affordable effort. This idea is not new; library scientists developed such methods (such as the MARC record format) decades ago, and the concept underlies the worldwide interchange of information on cataloged books, journals and media.

The innovation in the `RePEc: Research Papers in Economics` approach comes from the focus on the nontraditional materials to be cataloged in metadata archives: researchers' contact information, working papers, syllabi, statistical datasets and software components. These materials, crucial to the everyday practice of economic research, are made available to colleagues via metadata archives assembled by automated processes. These processes create a single virtual archive from the disparate collections of metadata at individual institutions' archives, which is then searchable by any of a number of "user services." That is, the design of the Web-based metadata archive network need not specify a single access mechanism by which it may be searched; its format is in the public domain, and anyone who wishes to provide an alternate *user service* to extract information from the virtual archive may do so. This implies that the latest developments in search engine technology, or platform-specific Web access tools may be applied as they become available. This "client-server" model allows a variety of clients to access the standardized information served to them, just as competing firms' Web browsers can all access the same HTML-coded Web pages. Like the Web itself, the virtual archive is an open system, with metadata stored as plain text on HTTP or FTP sites, and as such accessible by a variety of common tools.

The metadata structures to be defined in this research are templates of fielded *attribute:value* bibliographic data, conceptually similar to the MARC record format of the Library of Congress. In this simple data structure, certain fields are mandatory, certain fields may appear one or more times, and certain fields' contents are governed by lists of admissible values. The data in this structure may be manipulated by automated processes and routinely checked for validity. These validation processes, written as Perl scripts that may be executed on diverse platforms, greatly reduce the labor involved with producing valid metadata. Simplicity of these data structures makes it possible for researchers or their assistants without formal training in cataloging procedures to construct valid metadata, assisted by automated validation processes.

The individual metadata templates–for example, one "Article" template per working paper, or one "Software" template per software component–are then assembled into plain-text archives. These archives may then be integrated by automated processes into a single virtual archive, and the data they contain will be accessible by any "user service" searching the archive. The software tools for archive handling are freely available to individual institutions' archive maintainers, as well as complete documentation on the template design specifications.

The RePEc system, focused primarily on metadata describing working papers in a format known as "`ReDIF`", has grown dramatically from its modest beginnings in May 1997. The Boston College Department of Economics was one of the founding archives of RePEc. By January 2000, over 100 archives have joined RePEc, including prestigious research centers such as the National Bureau of Economic Research, the Federal Reserve System, the Centre for Economic Policy Research, the Bank of England, and many of the world's leading university economics departments. The metadata describe both printed and electronic documents, with over 20,000 downloadable papers, articles and software components accessible at present, as well as bibliographic information on over 40,000 additional documents.

The IDEAS user service

The RePEc effort generates a single virtual archive, access to which may be provided by any interested party in the form of a "user service". In this sense, there can be multiple access tools that work with the archived data, competing for users' interest. One of the most successful user services has been Christian Zimmermann's `IDEAS`, which recently recorded its 9 millionth download. IDEAS works with a static set of HTML pages which are generated nightly from the collective contents of contributing archives. The pages, grouped into `series` of working papers, published articles and software components, may be accessed either by series or via a search engine. For instance, you can examine the `Boston College Working Papers in Economics` series; each link on that page leads to a web page describing the paper, giving (where provided) its abstract, JEL classification, appropriate keywords, and if possible a link from which the full text of the paper may be downloaded (usually in Adobe PDF format). You may view some of the underlying "`templates`" for those papers online. The same logic is used to provide access to bibliographic information on articles in a number of leading journals, such as *Econometrica*, *Journal of Applied Econometrics*, the Federal Reserve System's publications, and the International Monetary Fund's *Staff Papers*. In the latter case, the links to individual articles also provide access to datasets, where provided by the authors, used to prepare the paper.

Access to software components

Although the use of these Web-based tools to provide an "electronic card catalog" listing working papers and published articles might seem a straightforward enhancement of an on-line library card catalog, it should also be noted that the same metadata tools may be used to organize and catalog a variety of economic software. The pioneering effort in this vein is Baum's SSC-IDEAS `Statistical Software Components archive`, which contains links to over 400 "software components". What is a software component? Many programming languages, as we shall see, are extensible via user-written functions, procedures, or applications that may be of general use. For instance, a procedure may add a new statistical test to a econometrics package, such as `Stata`, that does not contain that test, or provide a "canned" model of a macroeconomic phenomenon for use in `MATLAB`. These are files that may be described, cataloged, searched and served over the web in the same manner as publications. The usefulness of archiving software tools reflects the same rationales developed above: although every user of this package could place her contributions on her web site, how would one ever find them? Where is the "card catalog" for Stata-based materials, for instance? In some cases, the individual software vendors provide such an archive (e.g. Wolfram Research's `MathSource` for Mathematica materials) but materials stored there are out of the author's control, making updates for corrections and enhancements more cumbersome. In contrast, an IDEAS-based software archive may contain "pointers" to the author's own web site, so that any changes will be reflected in the material to be accessed. You may view some of the underlying `templates` for those software components online. At present, the majority of software components in the SSC-IDEAS archive are actually housed on the BC ftp server, but for Stata components that is technically most useful. A description of the rationale for the SSC-IDEAS archive in the context of Stata components is available on line at `http://fmwww.bc.edu/RePEc/docs/RePEc.AUBER.html` while a recent report on the popularity of this archive for Stata components is also `accessible`. An illustration of a dynamic report detailing recent additions to the SSC-IDEAS archive may also be `accessed` (at present rather slowly). This link runs a perl script, or CGI, that searches through over 200 RePEc templates for recent Creation-Date or Revision-Date fields. You may view the source of this `script`.

Extensions of RePEc

Further developments of the RePEc system include an attempt to separate the details of an individual author's affiliation and contact information from the templates describing her contributions, placing the personal information in a separate database with links to authored items. In a decentralized system, adoption of this scheme is hindered by the lack of any central authority to ensure that these links are made; at present we ask authors to register with HoPEc (Home Pages in Economics) and subsequently identify their works, but this requires the cooperation of many individuals.

The potential for RePEc template formats to be extended to the cataloging of datasets has also been proposed. RePEc already contains links to a number of datasets associated with articles or published papers, or those provided as test data for software components. However, datasets might be cataloged in their own right, and series of available datasets made accessible for browsing and searching just as other RePEc contents are today.

Economics Data Resources

You should become familiar with the BC Economics Department's `Economics Data Resources` page. This page, which is frequently updated, contains links to:

- Boston College Access to Economic and Financial Data, described below,

- A variety of U.S. government data statistical agencies;

- Each of the regional Federal Reserve Banks and the Board of Governors–notably, `FRED` at the St. Louis Fed;

- A number of other data sources.

There are, as noted in *RFE* and *WebEc*, a wide variety of sites on the Internet that provide access to data, in more or less friendly forms. An important adjunct to acquiring such data is the documentation of the procedure you have used to obtain it–ensuring that if you must revisit the site and re-extract the data, you may retrace your steps. And, of course, you will need the URL of any site when you cite the data source in written work.

Boston College Access to Economic and Financial Data

This `page` is an attempt to collect information on a wide variety of economic and financial datasets that are accessible at Boston College. They are grouped into five headings:

- Macroeconomic data

- Firm and stockmarket data

- Bond market and interest rate data

- International data

- Exchange rate data

Many datasets may fall into more than one of these categories; some are cross-classified. The ease of use of these datasets varies widely; some require that you acquire an account on a separate computer system and write an SQL (Structured Query Language) program to retrieve the data. Others provide access to datasets in `Stata` format (the format used by a popular statistical package, available in the SLSC), or in Excel-compatible format. Other datasets may be accessed online–e.g. those of the NBER, or the Michigan Panel Study of Income Dynamics (PSID), producing downloads that may then be read into the program of your choice.

An Experimental Interface to Online Data via Linux

A fully 'open source' experimental interface to data owned by the University is available for two datasets at present:

- World Bank *World Development Indicators*

- DRI *Basic Economics*

The platform housing these data is a "commodity" desktop PowerMacintosh G3 system, running a flavor of the popular `Linux` operating system rather than Apple's MacOS. Web services are provided by the ubiquitous `Apache` webserver, with the `mySQL` database engine being used to store and retrieve the data. The user interface is written in `PHP3`, a Perl-like language that makes it very easy to produce web pages with dynamic content, without a program or "plug-in" running in the user's browser.

You may use the interface to browse the data available in either of these databases. For the `World Development Indicators` *(WDI), documented on the web*, you may select from a number of categories of available data. When you have done so, the PHP3 script underlying the first page will produce a customized list of the variables in the selected categories from among the 503 annual series available, and allow you to select any number of them for retrieval. The following screen will allow you to specify the countries and/or country groupings for which data are desired, and limit the years for which data are to be retrieved. The system will then produce a list of the variables, countries (and/or groupings) and years specified, and allow you to either display the data in the web browser (for up to 500 rows) or download the datafile. If downloaded, the file should merely be saved to the desktop; it may be read into Excel as a tab-delimited text file, or read into Stata with "insheet". A Stata program that performs a "reshaping" step on the data is provided. Although the program does not run very quickly, it must be considered that the main data table for this database contains over 1.65 million rows.

A similar interface is provided for the `DRI Basic Economics` database. This is a more complex data structure, in that DRI contains data at annual, quarterly, and monthly frequencies–although a given series is only available at one of those frequencies (the highest available, so that e.g. consumer price data are available monthly, GDP data are available quarterly, and capital stock data are only available at the annual frequency). This interface provides access to 6,053 timeseries in total: 1,601 annual series, 1,835 quarterly series, and 2,617 monthly series, most pertaining to the U.S. economy. The database is `documented` on the web.