A Code Archive for Economics and Econometrics

Dirk Eddelbüttel[1]
GREQAM-EHESS, Marseilles, France

October 1996

Abstract. This paper introduces CodEc, an Internet repository for computer programs written by, and for, economists and econometricians. CodEc provides a means to aid communication among researchers by using the Internet to share implementations of algorithms. After a description of the current content of CodEc, we discuss the use of CodEc to store and retrieve code from published articles.

Key words: Internet, code repository, computer programs.

1 Introduction

> The value of an exchange of computing techniques has long been realized, and various arrangements for program-sharing and for the publication of algorithms have arisen.
>
> Editorial Note,
> *J. R. Statist. Soc. C 17(2)*, 1968

Almost three decades after the first publication of this citation, it still remains difficult to obtain programs, or implementations of algorithms, that other researchers have made available. But with a steadily rising number of computers connected to the Internet, and an increasing number of sites of interest to economists as listed for example in the exhaustive compilation by Goffe (1996), it is clear that the economics profession no longer has to rely exclusively on printed journals as a means of transmission for scientific work as was the case when this citation marked the start of the series of algorithms that are published in *Applied Statistics*.

This paper describes CodEc, the first Internet archive for programs for economists and econometricians. It was started in the spring of 1994 using the `gopher` protocol, and converted to the `http` protocol of the world wide web (WWW) in the summer of 1994. CodEc is part of the NetEc volunteer project started by Thomas Krichel in February 1993.[2] Apart from CodEc, NetEc comprises several other services: BibEc, a large bibliography of working papers in economics, WoPEc, a database about downloadable papers, and WebEc, a collection of starting points for the WWW. CodEc, which can be found in several different places on the Internet,[3] aims to complement these other services by offering an archive for programs written in any programming or application language of use in economics.

2 Description of CodEc

CodEc provides an Internet repository for computer programs that a researcher wishes to make available. These programs are classified according to the programming or application language in the case of source code, or by hardware platform in case of binary executables. Currently, the following sections exist:

C and C++ contains two C++ matrix classes written by Chris Birchenhall and Robert B. Davies; C code for the test for independence of a time-series by Brock, Dechert, Scheinkman, and LeBaron (1996); and further links.

Dos and Windows contains several executable programs: XploRe, a package for non-parametric regression (Härdle 1990, appendix 1); BCI Manager, a data managing program for the data provided by the U.S. Department of Commerce; programs for count and duration regression; and a data-bank manager for time-series.

Fortran contains source code from articles by Goffe, Ferrier, and Rogers (1994) and Kalaba and Tesfatsion (1990, 1992).

Gauss which contains three sets of programs for maximum likelihood estimation, for OLS, Arima and nonlinear regression and for logit and probit estimation; a complete mirror of the Gauss archive at American University, and more links.

Limdep contains a link to the on-line version of the current Limdep manual.[4]

Mathematica contains code that accompanies Varian (1984); and more links.

Matlab contains an implementation by Michael Gordy of the article by Dorsey and Mayer (1995); code from Hansen and Sargent (1990) and more links.

Rats contains five packages of code made available by the producer of Rats that implement various time series methods; an implementation of a covariance matrix estimator; and another link.

Shazam contains links to the different Shazam archives at the University of British Columbia.

Xlisp-Stat contains a link to the Xlisp-Stat (Tierney 1991) archive at UCLA.

Within each of these directories, CodEc is organised by using software information files that have the extension `.si`; Table 1 lists their entries.[5] The software information files are typically modifications of the original README file provided by the author. It is usually a good idea to start with the software information file if one wants to obtain more information about a particular package.

CodEc also contains a section of links to other sites on the WWW that are of similar or related interest. Currently, there are links to

Other archives as ELSA, GAMS, NetLib, QM&RBC, Statlib, StatMath and UICSTAT.

Commercial packages as Gempack, Maple, Mathematica, Matlab, NAG, Rats, SAS, Shazam, Stata and TSP International.

Finally, CodEc comprises a section of programs that deal with different file archive and compression formats such as `zip`, `tar` and `gzip`. Some documentation is provided to help novice users to unpack the files provided in any of these formats, as well as links to freely

Table 1: Format of the software information file

|TI|	title of the program or package
|DS|	description of the package
|AU|	author's name
|EM|	email of author, if available
|RQ|	requirements *i.e.* C compiler for C source code
|SY|	system it is running *i.e.* Dos, Mac, ...
|CO|	code format, mostly `zip` or `tar.gz`
|SZ|	length in bytes

available executables that can perform the unpacking.

In summary, CodEc offers the service of a repository for programs written in different programming languages and for different application programs that are frequently used by economists and econometricians. It thereby provides a central archive where researchers can retrieve implementations of algorithms that can aid them in their own work. The mirroring structure of CodEc allows for convenient and fast access for users in Europe, North America and the Far East. Access has been increasing dramatically: in the six months from April to September 1996, an average of 2457 WWW pages have been requested per month at the mirror at Washington University alone; this compares very favourably with the preceding six months period from October 1995 to March 1996 when on average 943 WWW pages were requested per month.

Authors who wish to submit their programs to CodEc should, at least until an automated procedure for uploads is installed, contact the archive maintainer by email at the address `edd@qed.econ.queensu.ca`. Submissions can be accepted in two ways. The first, and preferred form, consists of a `zip` or `tar.gz` compressed file containing the sources which is accompanied by a README file that gives a brief overview of the package. Alternatively, authors can submit a uniform resource locator (url) that points to another site, for example on the personal or departmental workstation of the person making the submission. The first format is preferred and recommended as, through the mirroring of the hosts in England, the United States and Japan, it fully utilises the distributed structure of CodEc which gives better service to users in different parts of the globe.

3 Proposal for CodEc

Increasingly, academic journals demand that the data for a published article be made available by the author.[6] For example, both the *Journal of Business and Economic Statistics* and the *Journal of Applied Econometrics* have data archive sites on the Internet.[7] These sites enable authors to make their data available so that readers of their articles can replicate the published findings. But this still leaves the reader with the task of re-implementing the methodology used by the author. The next logical step up from the data archives is to also make the code available that was used to conduct the published research.

Reporting on the *JMBC Project*,[8] Dewald, Thursby, and Anderson (1986) conclude: *"On the basis of our findings, we recommend that journals require the submission of programs and data at the time empirical papers are submitted."* Anderson and Dewald (1994), reporting on the JMBC as well as a similar project at the Federal Reserve Bank of St. Louis *Review*, reconfirm that authors are much more likely to respond favourably to a request for programs and data corresponding to an article if only they know ahead of time that these will be archived.

CodEc could very easily provide such an archive for journals using the existing infrastructure at the hosts in England, the USA and Japan. New programs and routines could be distributed via the WWW interface that is already in place, and that have been used successfully over a two-year period. Organised as a central archive, albeit with local mirrors, CodEc offers 'one-stop-shopping' for programs and code stemming from different areas of research within economics. Further, new sections could easily be added to CodEc on a per-journal basis, with links into the language directories. Thus, journals can suggest to their authors that they can provide their computer code via CodEc. An automated procedure could be installed by using a scripting language such as `perl` so that submission can be made via email, or through a WWW form, with little or no human involvement for each submission in order to minimise costs. The present paper hopes to have offered a basis for further discussion towards wider use of economics code archives.

Notes

[1]The author gratefully acknowledges helpful comments from Bill Goffe, Thomas Krichel, Bob Parks and Lisa Powell.

[2]The original host is provided by the Manchester Computing Centre on the MIDAS server at `http://netec.mcc.ac.uk/NetEc.html`. Mirrors are available at Washington University in St. Louis, USA, under the address `http://netec.wustl.edu/NetEc.html` and at Hitotsubashi University in Tokyo, Japan, under `http://netec.ier.hit-u.ac.jp/NetEc.html`.

[3]CodEc can be reached via the server at the Manchester Computing Centre under the address `http://netec.mcc.ac.uk/CodEc.html`. While the system is geared towards the WWW, access is also provided at `ftp://netec.mcc.ac.uk/pub/NetEc/CodEc` to allow mirroring through anonymous ftp. CodEc is also accessible at the NetEc mirror at Washington University in St. Louis under `http://netec.wustl.edu/CodEc.html` and `ftp://netec.wustl.edu/pub/NetEc/CodEc`, and at the NetEc mirror at Hitotsubashi University at `http://netec.ier.hit-u.ac.jp/NetEc.html` and `ftp://netec.ier.hit-u.ac.jp/pub/NetEc/CodEc`.

[4]Providing more Limdep content, as *e.g.* the examples from the Limdep manual, is currently under discussion with Bill Greene, the author of Limdep.

[5]The software information file format shown in Table 1 will at some point be changed to a new scheme which supports metadata and searches better than the current one. The new scheme will presumably use a variant of the `whois++` service described by Deutsch, Schoultz, Faltstrom, and Weider (1995).

[6]Of course, exceptions can be made if the data is either subject to confidentiality agreements or of a proprietary nature.

[7]The address for the *Journal of Business and Economic Statistics* archive is `ftp://ftp.duke.edu/jbes/` and the archive of the *Journal of Applied Econometrics* can be accessed via `http://qed.econ.queensu.ca/jae` and `ftp://qed.econ.queensu.ca/jae`.

[8]The *Journal of Money, Credit and Banking* (JMBC) undertook the National Science Foundation funded *JMBC Data Storage and Evaluation Project* which consisted of an editorial policy of requesting from authors the programs and data that were used in the research that led to articles published in the JMBC. The programs and data were also made available to other researchers upon request.

References

Anderson, Richard G. and William G. Dewald (1994). Replication and scientific standards in applied economics a decade after the *Journal of Money, Credit and Banking* project. *Federal Reserve Bank of St. Louis Review 76(6)*, 79–83.

Brock, William A., W. Davis Dechert, José A. Scheinkman, and Blake LeBaron (1996). A test for independence based on the correlation dimension. *Econometric Reviews 15(3)*, 197–235.

Deutsch, Peter, Richard Schoultz, Patrik Faltstrom, and Chris Weider (1995). Architecture of the `WHOIS++` service. Request for Comments 1835, `ftp://ds.internic.net/rfc/rfc1835.txt`.

Dewald, William G., Jerry G. Thursby, and Richard G. Anderson (1986). Replication in empirical economics: The *Journal of Money, Credit and Banking* Project. *American Economic Review 76(4)*, 587–603.

Dorsey, Robert E. and Walter J. Mayer (1995). Genetic algorithms for estimation problems with multiple optima, nondifferentiability, and other irregular features. *Journal of Business & Economic Statistics 13(1)*, 53–66.

Goffe, William L. (1996). Resources for economists on the internet. Vol. 2, No. 1, `http://econwpa.wustl.edu/EconFAQ/EconFAQ.html`.

Goffe, William L., Gary D. Ferrier, and John Rogers (1994). Global optimization of statistical functions with simulated annealing. *Journal of Econometrics 60*, 65–99.

Hansen, Lars Peter and Thomas J. Sargent (1990, October). Recursive linear models of dynamic economics. NBER Working Paper 3479, NBER, Cambridge, Ma.

Härdle, Wolfgang (1990). *Applied Nonparametric Regression*. Econometric Society Monographs No. 19. Cambridge: Cambridge University Press.

Kalaba, Robert and Leigh Tesfatsion (1990). Flexible least squares for approximately linear systems. *IEEE Transactions on Systems, Man, and Cybernetics 20*, 978–989.

Kalaba, Robert and Leigh Tesfatsion (1992). Nonlocal automated comparative static analysis. *Computer Science in Economics and Management 5*, 313–331.

Tierney, Luke (1991). *Lisp Stat: An Object Oriented Environment for Statistical Computing and Dynamic Graphics*. Wiley Series in Probability and Mathematical Statistics. New York: John Wiley & Sons.

Varian, Hal R. (1984). *Microeconomic Analysis* (Second ed.). New York: Norton.