

Enhancing access to statistical software tools and datasets for research and instruction

Christopher F Baum

Department of Economics, Boston College

May 2000

Prepared for UK Stata Users' Group Meeting

<http://fmwww.bc.edu/RePEc/docs/UKSUG2K.pdf>

Statistical software tools have become more extensible, readily permitting their users to extend functionality, while widespread access to the Internet has made it possible to exchange those materials within the research community.

Stata has become particularly supportive of these trends with its `.ado` architecture, in which user commands properly installed are indistinguishable from built-in commands, and recent enhancements: the “net-aware” facilities for installation and archive access (such as *net describe* and *webseek*).

This paper describes an initiative to enhance information flow in the discipline of economics – the **RePEc** project – which has been expanded from its original focus on preprints and published articles to incorporate “metadata,” or bibliographic information, on “software components” such as user-authored additions to Stata.

It also describes preliminary efforts to extend this system to include metadata describing statistical datasets.

Given Stata’s ability to access datasets directly over the Web, a tool that locates and describes Stata-format datasets should be very helpful in instruction and research.

The use of a RePEc archive to house meta-data describing software components provides greater visibility for these materials, and integrates them into a broader set of software components that may be referenced to enhance Stata's facilities.

The SSC-IDEAS archive provides Web browser access to information about more than 400 Stata components, incorporating those published on StataList, and is mirrored by the new *webseek* facility. The archive's Stata-oriented contents are accompanied by automatically generated package (.pkg) files that render them installable in "web-aware" Stata.

This file is part of [IDEAS](#), which uses [RePEc](#) data

[[Series](#) | [JEL Classification](#) | [Search](#) | [Viewers](#) | [Help!](#)]

If a paper is available in full text, a red [Downloadable!] appears just below and download information is at the bottom of its page between two lines. If not, see [Help!](#)

Boston College Department of Economics

Statistical Software Components

Publisher-Postal: Boston College, Carney Hall, Chestnut Hill MA 02467 USA

Publisher-Fax: +1-617-552-2308

Publisher-Homepage: <http://fmwww.bc.edu/EC/>

For information about this series, please contact baum@bc.edu.

- 1.- [RePEc:boc:bocode:S375501](#) **ARCHUTIL:** Stata modules to work with contents of SSC-IDEAS Archive *by* Nicholas J. Cox & Christopher F Baum [[Downloadable!](#)]
- 2.- [RePEc:boc:bocode:S411301](#) **COPYDESC:** Stata module to copy description of variable *by* Nicholas J. Cox [[Downloadable!](#)]
- 3.- [RePEc:boc:bocode:S411201](#) **SHOWGPH:** Stata module to to show graphs previously saved *by* Jan Brogger & Nicholas J. Cox [[Downloadable!](#)]
- 4.- [RePEc:boc:bocode:S411101](#) **RECODE2:** Stata module to provide extended recode capabilities *by* John Hendrickx [[Downloadable!](#)]
- 5.- [RePEc:boc:bocode:S411002](#) **MODLPR:** Stata module to estimate long memory in a timeseries *by* Christopher F Baum [[Downloadable!](#)]
- 6.- [RePEc:boc:bocode:S411001](#) **ROBLPR:** Stata module to estimate long memory in a set of timeseries *by* Christopher F Baum [[Downloadable!](#)]
- 7.- [RePEc:boc:bocode:S410801](#) **MATVSORT:** Stata module to sort vector *by* Nicholas J. Cox [[Downloadable!](#)]
- 8.- [RePEc:boc:bocode:S410701](#) **SMHSIAO:** Stata module to conduct Small-Hsiao test for IIA in multinomial logit *by* Nick Winter [[Downloadable!](#)]
- 9.- [RePEc:boc:bocode:S410601](#) **IVGMM0:** Stata module to perform instrumental variables via GMM *by* Christopher F Baum & David M. Drukker [[Downloadable!](#)]
- 10.- [RePEc:boc:bocode:S410501](#) **METATRIM:** Stata module to perform nonparametric analysis of publication bias *by* Thomas Steichen [[Downloadable!](#)]
- 11.- [RePEc:boc:bocode:S410401](#) **KPSS:** Stata module to compute Kwiatkowski-Phillips-Schmidt-Shin test for **stationarity** *by* Christopher F Baum [[Downloadable!](#)]
- 12.- [RePEc:boc:bocode:S410301](#) **ENLARGE:** Stata module to propagate constant values through group *by* Stanislav Kolenikov [[Downloadable!](#)]
- 13.- [RePEc:boc:bocode:S410201](#) **ICOMP:** Stata module to calculate model selection information criteria *by* Stanislav Kolenikov [[Downloadable!](#)]
- 14.- [RePEc:boc:bocode:S410101](#) **CHARUTIL:** Stata module of utilities for working with characteristics *by* Nicholas J. Cox [[Downloadable!](#)]
- 15.- [RePEc:boc:bocode:S410001](#) **DFGLS:** Stata module to compute Dickey-Fuller/GLS unit root test *by* Christopher F Baum [[Downloadable!](#)]
- 16.- [RePEc:boc:bocode:S409901](#) **XRIGLS:** Stata module to calculate reference intervals via generalised least squares *by* Patrick Royston & Eileen Wright [[Downloadable!](#)]
- 17.- [RePEc:boc:bocode:S409902](#) **CENTCALC:** Stata module to calculate distribution-based centiles *by* Patrick Royston

Linked Metadata Archives

The **RePEc (Research Papers in Economics)** project is an international effort to enhance the information freely available to economics researchers through the development of linked metadata archives.

Metadata are the bibliographic details, similar to the information content of a library catalog, which describe a particular archive component and permit it to be located by its title, author(s), keywords, or words in its description (abstract). It extends the library card catalog model to hyperlinks, which might allow you to read a working paper, download a journal article, install a software component on your desktop computer, or access a statistical dataset.

RePEc archives and services

A RePEc archive is a set of **templates**: ASCII files, which may be produced manually or automatically from other information sources—which are automatically processed on a daily basis by the software underlying RePEc services. These templates may be thought of as “card catalog” entries that concisely describe the material in a standardized form.

A RePEc service provides a user interface to these metadata. There are many archives (over 100 at this time), but only one “virtual database” containing the compendium of all archives’ contents. The underlying model:

- Many archives
- One database
- Many services

Template-Type: ReDIF-Software 1.0
Title: SHOWGPH: Stata module to to show graphs previously saved
Author-Name: Jan Brogger
Author-Email: jan.brogger@med.uib.no
Author-WorkPlace-Name: University of Bergen, Norway
Author-Name: Nicholas J. Cox
Author-Email: N.J.Cox@durham.ac.uk
Author-WorkPlace-Name: University of Durham
Programming-Language: Stata
Abstract: showgph shows graphs previously saved as .gph files and present in the same directory or folder. Graphs are drawn one at a time in alphanumeric filename order. This is version 1.0.10 of the software.
Keywords: graphics
Requires: Stata version 6.0
Note: This module may be installed from within 'net-aware' Stata by typing "archinst showgph" (q.v. STB-54 archutil).
Series: Statistical Software Components
Number: S411201
Creation-Date: 20000428
File-URL: <http://fmwww.bc.edu/repec/bocode/s/showgph.ado>
File-Format: text/plain
File-Size: 102 lines
File-Function: program code
File-URL: <http://fmwww.bc.edu/repec/bocode/s/showgph.hlp>
File-Format: text/plain
File-Size: 81 lines
File-Function: help file
Handle: RePEc:boc:bocode:S411201

d 'SHOWGPH': module to to show graphs previously saved
d
d showgph shows graphs previously saved as .gph files and present
d in the same directory or folder. Graphs are drawn one at a time
d in alphanumeric filename order. This is version 1.0.10 of the
d software.

d
d Author: Jan Brogger, University of Bergen, Norway
d Support: email jan.brogger@@med.uib.no

d
d Author: Nicholas J. Cox, University of Durham
d Support: email N.J.Cox@@durham.ac.uk

f showgph.ado
f showgph.hlp

Open access for non-commercial use

The data contained in RePEc are freely accessible to any who wish to make use of it, or repackage it via a RePEc “service” for any non-commercial purpose. The inclusion of meta-data in RePEc does not preclude a charge being made for the components themselves: e.g. a journal article may be downloadable only by subscribers, or a NBER working paper or an institute’s research report may require payment. It does imply that the summary information is available without charge.

Contrast this with existing services such as *EconLit*, which are not freely available over the Web. You must access them through a library which has subscribed to web access, and deal with firewall and authentication issues.

RePEc versus individual Web servers

Each research institution, or individual, could make their work freely available via the Web. General-purpose search engines could in principle find it by reference to its title, keywords, etc. Why do we need metadata archives?

- Search engines are looking for the “needle in the haystack”
- A metadata archive is searchable without reference to the entire WWW
- Items classified by subject code (e.g. JEL) can be grouped

RePEc versus individual Web servers

The individual archives of materials—e.g. working papers or research reports from a single department or research center—are useful as a method of disseminating those materials. If a user knows that Dr. John Doe is a member of the BC Economics Department, she can find his papers on the BC Economics web site.

But what if that affiliation is not known? Or if she is a Ph.D. candidate searching for recent working papers on a particular subject? Then the existence of hundreds or thousands of web servers without a unifying framework becomes a serious detriment.

RePEc versus monolithic archives

For more specialized materials—e.g. statistical software tools—a generalized search is almost worthless. One approach involves the creation and maintenance of a “monolithic” archive. But that requires substantial hardware and software resources, and funding to maintain the archive, which for a collection of thousands of items will be nontrivial.

Furthermore, the monolithic archives may not be unique. If competing archives emerge, contributors will question whether they should participate in several to maximize visibility of their materials. This is cumbersome and inefficient to say the least.

The RePEc concept

RePEc addresses these concerns via its design concepts. It is not a monolithic archive, but a decentralized network of archives. Each archive is maintained by local information providers or volunteers who serve to “catalog” materials of one or more institutions. The set of archives can be scaled to any size, with minimal central administration; the only coordination required is the allocation of archive names.

Via decentralization, each institution can choose to provide information about those research materials which they are interested in sharing via RePEc. This information generally overlaps with the information compiled for local web pages, and RePEc templates may be generated mechanically (or manually, with minimal effort) from the details already “published” on the Web.

RePEc for software distribution

The RePEc framework was designed to provide metadata for “papers” (working papers, or preprints), “articles” (published articles), and other print materials such as “chapters” or “books.” But the framework may be used for the storage of metadata of any sort, and potentially to provide access to an item stored in virtually any format (HTML, PDF, ZIP file, tar.gz archive, etc.)

These characteristics make the RePEc framework very useful for the dissemination of information about **software components**, and for the delivery of those components themselves, as well as sample programs, data, and documentation.

What are **software components**?

Many statistical packages (e.g. Stata, SAS, RATS, Ox) or specialized programming languages (Mathematica, GAUSS, MATLAB, S+) support the use of **components**: specialized functions, or procedures, or modules that add functionality to the package. These “components” are not programs per se, but rather components of programs: i.e. functions that may be called within a user program, or additional commands that may be invoked by the user of that package or language. Some degree of generality is implied, in that useful components are not completely specific to their author’s task at hand, but have the capability to perform some function based on their argument list.

RePEc and software components

Two years ago, I proposed an extension to the RePEc standard: a ReDIF-Software template, and the first RePEc series containing software templates was established at the Boston College Department of Economics. This series, the **SSC** (Statistical Software Components) archive, was designed to provide users of statistical packages and specialized programming languages with a way of making their public-domain contributions accessible via the Web.

Many users of RePEc metadata employ the IDEAS service. IDEAS, maintained by Christian Zimmermann at UQAM (Montreal), is accessible at <http://ideas.uqam.ca>.

SSC-IDEAS

The SSC archive, when accessed in IDEAS, provides access to over 450 software components, for Stata, MATLAB, Mathematica, RATS, GAUSS, and Ox. Contributions in any language (including code in standard languages such as FORTRAN and C) are welcomed. The vast majority of the components are Stata code, most having appeared on the StataList listserv within the last two years. These components have been contributed by a wide range of authors from the US, the UK, Europe and Australia. Some components have been written or coauthored by StataCorp staff, and represent “preview” versions of software that may eventually become available in “official” Stata.

“Net-aware” Stata

With the advent of Version 6, Stata is “net-aware”: that is, the program may enquire via HTTP to determine whether there are updated elements of official Stata available. Likewise, Stata components associated with the Stata Technical Bulletin (STB) may be accessed via “net” commands from the StataCorp web site.

“Net-awareness” also allows for any Stata user to share his or her Stata code with other users. The Stata documentation contains detailed instructions on creating your own web site of materials; all that is needed is access to a Web server, and the ability to place text files on that server.

What is missing from this model?

Just as each department or individual can establish a web page for dissemination of their publications or research findings, they may establish a Stata web site from which Stata components may be downloaded. Stata will dutifully retrieve those materials, as long as the site's design follows a few modest requirements.

But how are you to find the materials if you don't know where to look?

SSC-IDEAS and “net-aware” Stata

If a Stata component accessible to “net-aware” Stata is described in a ReDIF-Software template and included in SSC-IDEAS, that component may be located via the IDEAS *eXcite* search facility by:

- author
- word in title
- keywords
- “abstract”: text description of functionality

SSC-IDEAS and “net-aware” Stata

The Stata code itself (the `.ado`) may be viewed in the web browser, as may the help file (`.hlp`). But if you’re using “net-aware” Stata, the best way to install the code in your copy of Stata is via the “net install” command.

Additional utilities are now available from SSC-IDEAS (and described in STB-52 and STB-54) to facilitate this process of:

- locating the component you need
- accessing the SSC-IDEAS archive
- installing the component

SSC-IDEAS “archutil” package

“archutil” contains utilities for SSC-IDEAS archive access. The “archdesc” command produces a list of all Stata components on the archive, organized by first letter of package name, with a short title (that appearing on the web page listing).

“archdesc *letter*” will provide a listing of those packages with names beginning with *letter*. Furthermore, the “archinst *package*” command may be given directly, to access the archive and install a specific package, or component, with a single command.

SSC-IDEAS “archutil” package

“archtype *file*” allows you to see the text of an .ado or .hlp file on the SSC-IDEAS archive without actually installing it. This mirrors the capability of viewing these files from the web browser accessing the SSC-IDEAS archive, but works within “net-aware” Stata.

“archcopy *file*” will copy that file to the appropriate directory on your computer. This should not be used as an alternative to “archinst” or “net install”, but enhances the functionality of Stata’s native “copy” command by prespecifying the web site from which the file is to be copied.

By combining the metadata archive of Stata components produced by dozens of authors with the “net-aware” facilities built in to the application, the whole is more than the sum of the parts.

The metadata archive allows the Stata user to do a “fuzzy search” for available components without having to know their names, who wrote them, or when they were announced or updated on Statalist. Simply specifying keywords that appear in the package title or description in a SSC-IDEAS search will suffice.

The “net-aware” facilities of Stata ensure that an entire SSC-IDEAS-accessible package will be properly installed, if that is possible, so that the application’s functionality is protected.

The “archutil” facilities are integrated with Bill Gould’s new ‘webseek’ facility; as described in STB-54, ‘webseek’ searches a database of many sites, among which is the entire SSC-IDEAS archive. Thus, any terms that appear in the .pkg file will be located by “webseek”, and within desktop versions of Stata, one may just click on the link to visit the archive and install the package.

SSC-IDEAS Stata modules created or revised 2000-04-12 to 2000-05-12

[COPYDESC](#) module to copy description of variable
Author: Nicholas J. Cox
Created: 2000-05-10

[SHOWGPH](#) module to to show graphs previously saved
Authors: Jan Brogger Nicholas J. Cox
Created: 2000-04-28

[RECODE2](#) module to provide extended recode capabilities
Author: John Hendrickx
Created: 2000-04-28

[MODLPR](#) module to estimate long memory in a timeseries
Author: Christopher F Baum
Created: 2000-04-24

[ROBLPR](#) module to estimate long memory in a set of timeseries
Author: Christopher F Baum
Created: 2000-04-24

[MATVSORT](#) module to sort vector
Author: Nicholas J. Cox
Revised: 2000-04-20

[LABUTIL](#) modules for managing value labels
Author: Nicholas J. Cox
Revised: 2000-05-10

[MATSELR](#) module for selection of rows and/or columns from matrix
Author: Nicholas J. Cox
Revised: 2000-04-20

[GPHUDAK](#) module to estimate long memory in a timeseries
Authors: Christopher F Baum Vince Wiggins
Revised: 2000-04-21

[OUTREG](#) module to format regression output for published tables
Author: John Luke Gallup
Revised: 2000-05-03

End of recent additions and updates

Extension to datasets

Although RePEc does not yet have a “template type” for the description of statistical datasets, development of such a metadata model is underway. Anyone teaching statistics, quantitative methods, or econometrics will appreciate the convenience of making datasets available for student use over the Web. In the remainder of this presentation, we will examine the first attempts to use a RePEc metadata archive as a tool for making datasets available for instructional use. Such methods could also be used in the future to provide organized access to research datasets, as well as their supporting materials such as codebooks.

In the course of preparing materials for graduate econometrics this spring, I have constructed web pages describing three textbooks' supplementary materials. Jeffrey Wooldridge's newly published *Introductory Econometrics* text contains a link to his website, from which over 70 datasets are available – some of them quite sizable. Codebook files listing each dataset's variables are also provided. I have constructed a web page from which these datasets' descriptions can be accessed, on which the Stata command needed to “use” each one is given.

From the HTML code on that page, it is quite simple—using perl, as the next slide shows—to transform the information on each dataset into a “ReDIF template” (ReDIF: Research Document Information Format) which may then be assembled in a RePEc series. That has been done for these datasets, as well as for 27 datasets associated with the fourth edition of Bill Greene’s *Econometric Analysis*, and nine datasets associated with the second edition of Terry Mills’ text *The Econometrics of Financial Time Series*. Although the Greene datasets are generally quite small, some of the Mills datasets are sizable (e.g. 15,000 days of stock index data).

```
#!/usr/bin/perl
# message Wooldridge html page
# 0508 cfb
print "\nwool v1.0\n\n";
print "C F Baum May 2000\n\n";
use Text::Wrap qw(wrap $columns);
$columns=68;
open(IN,"Wooldridge.datasets") || die "Cannot open Wooldridge.datasets for input";
$outfilename="Wooldridge.rdf";
unlink $outfilename;
open(OUT,">$outfilename") || die "Cannot open $outfilename for output";
chop(@lines=<IN>);
$lines=@lines;
$nrtem=0;
foreach $line (@lines) {
    if($line =~ /<li>/ig){
        ($url,$title,$n,$desc) = ($line =~ /<li><big><A HREF="(.*?)">(.*?)</a>: N=(.*?),(.+
#
        print "$url $title $n $desc\n"; }
        print OUT "\nTemplate-Type: ReDIF-Paper 1.0\n";
        print OUT "Title: $title\n";
        print OUT "Author-Name: Jeffrey M. Wooldridge\n";
        print OUT "Series: Wooldridge instructional datasets\n";
        print OUT "Abstract: Instructional dataset, N=$n, $desc\n";
        print OUT " Accompanying Introductory Econometrics: A Modern Approach, \n";
        print OUT " Jeffrey M. Wooldridge, South-Western College Publishing, (c) 2000\n";
        print OUT " Datasets also accessible from http://wooldridge.swcollege.com.\n";
        print OUT "File-URL: http://fmwww.bc.edu/ec-p/data/wooldridge/$url\n";
        print OUT "File-Format: text/plain\n";
        print OUT "File-Function: codebook for dataset\n";
        print OUT "File-URL: http://fmwww.bc.edu/ec-p/data/wooldridge/$title.dta\n";
        print OUT "File-Format: application/x-Stata\n";
        print OUT "File-Function: Stata v6.0 web-accessible dataset\n";
        print OUT "Handle: RePEc:boc:bocins:$title\n";
        $nrtem++;}
    }
close OUT;
print "\ntotal templates created: $nrtem\n\n";
exit;
__END__
```

These dataset templates—currently numbering 107—have been placed in a RePEc “paper series” associated with the Boston College Department of Economics. This series can then be accessed from its descriptor page, or its contents searched from within the IDEAS search facility. Here is the descriptor page, and a search example.

This file is part of [IDEAS](#), which uses [RePEc](#) data

[[Series](#) | [JEL Classification](#) | [Search](#) | [Viewers](#) | [Help!](#)]

If a paper is available in full text, a red [Downloadable!] appears just below and download information is at the bottom of its page between two lines. If not, see [Help!](#)

Boston College Department of Economics

Instructional Stata datasets for econometrics

Publisher-Postal: Boston College, Carney Hall, Chestnut Hill MA 02467 USA

Publisher-Fax: +1-617-552-2308

Publisher-Homepage: <http://fmwww.bc.edu/EC/>

For information about this series, please contact baum@bc.edu.

- 1.- [RePEc:boc:bocins:TBL4-1](#) **TBL4-1** by William H. Greene [\[Downloadable!\]](#)
- 2.- [RePEc:boc:bocins:TBL5-1](#) **TBL5-1** by William H. Greene [\[Downloadable!\]](#)
- 3.- [RePEc:boc:bocins:TBL6-1](#) **TBL6-1** by William H. Greene [\[Downloadable!\]](#)
- 4.- [RePEc:boc:bocins:TBL6-2](#) **TBL6-2** by William H. Greene [\[Downloadable!\]](#)
- 5.- [RePEc:boc:bocins:TBL6-3](#) **TBL6-3** by William H. Greene [\[Downloadable!\]](#)
- 6.- [RePEc:boc:bocins:TBL6-4](#) **TBL6-4** by William H. Greene [\[Downloadable!\]](#)
- 7.- [RePEc:boc:bocins:TBL7-1](#) **TBL7-1** by William H. Greene [\[Downloadable!\]](#)
- 8.- [RePEc:boc:bocins:TBL7-2](#) **TBL7-2** by William H. Greene [\[Downloadable!\]](#)
- 9.- [RePEc:boc:bocins:TBL8-1](#) **TBL8-1** by William H. Greene [\[Downloadable!\]](#)
- 10.- [RePEc:boc:bocins:TBL9-1](#) **TBL9-1** by William H. Greene [\[Downloadable!\]](#)
- 11.- [RePEc:boc:bocins:TBL10-1](#) **TBL10-1** by William H. Greene [\[Downloadable!\]](#)
- 12.- [RePEc:boc:bocins:TBL13-1](#) **TBL13-1** by William H. Greene [\[Downloadable!\]](#)
- 13.- [RePEc:boc:bocins:TBL14-1](#) **TBL14-1** by William H. Greene [\[Downloadable!\]](#)
- 14.- [RePEc:boc:bocins:TBL15-1](#) **TBL15-1** by William H. Greene [\[Downloadable!\]](#)
- 15.- [RePEc:boc:bocins:TBL15-2](#) **TBL15-2** by William H. Greene [\[Downloadable!\]](#)
- 16.- [RePEc:boc:bocins:TBL15-3](#) **TBL15-3** by William H. Greene [\[Downloadable!\]](#)
- 17.- [RePEc:boc:bocins:TBL16-1](#) **TBL16-1** by William H. Greene [\[Downloadable!\]](#)
- 18.- [RePEc:boc:bocins:TBL16-2](#) **TBL16-2** by William H. Greene [\[Downloadable!\]](#)
- 19.- [RePEc:boc:bocins:TBL17-1](#) **TBL17-1** by William H. Greene [\[Downloadable!\]](#)
- 20.- [RePEc:boc:bocins:TBL17-2](#) **TBL17-2** by William H. Greene [\[Downloadable!\]](#)
- 21.- [RePEc:boc:bocins:TBL18-1](#) **TBL18-1** by William H. Greene [\[Downloadable!\]](#)
- 22.- [RePEc:boc:bocins:TBL18-2](#) **TBL18-2** by William H. Greene [\[Downloadable!\]](#)
- 23.- [RePEc:boc:bocins:TBL19-1](#) **TBL19-1** by William H. Greene [\[Downloadable!\]](#)
- 24.- [RePEc:boc:bocins:TBL19-2](#) **TBL19-2** by William H. Greene [\[Downloadable!\]](#)

- 25.- [RePEc:boc:bocins:TBL19-3](#) **TBL19-3** by William H. Greene [[Downloadable!](#)]
- 26.- [RePEc:boc:bocins:TBL20-1](#) **TBL20-1** by William H. Greene [[Downloadable!](#)]
- 27.- [RePEc:boc:bocins:FAIR_PT](#) **FAIR_PT** by William H. Greene [[Downloadable!](#)]
- 28.- [RePEc:boc:bocins:rs-r20](#) **rs-r20** by Terence C. Mills [[Downloadable!](#)]
- 29.- [RePEc:boc:bocins:rsq-20-r](#) **rsq-20-r** by Terence C. Mills [[Downloadable!](#)]
- 30.- [RePEc:boc:bocins:fta](#) **fta** by Terence C. Mills [[Downloadable!](#)]
- 31.- [RePEc:boc:bocins:exchd](#) **exchd** by Terence C. Mills [[Downloadable!](#)]
- 32.- [RePEc:boc:bocins:exchq](#) **exchq** by Terence C. Mills [[Downloadable!](#)]
- 33.- [RePEc:boc:bocins:sp500a](#) **sp500a** by Terence C. Mills [[Downloadable!](#)]
- 34.- [RePEc:boc:bocins:sp500d](#) **sp500d** by Terence C. Mills [[Downloadable!](#)]
- 35.- [RePEc:boc:bocins:ft30d](#) **ft30d** by Terence C. Mills [[Downloadable!](#)]
- 36.- [RePEc:boc:bocins:stockw](#) **stockw** by Terence C. Mills [[Downloadable!](#)]
- 37.- [RePEc:boc:bocins:ADMNREV](#) **ADMNREV** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 38.- [RePEc:boc:bocins:APPLE](#) **APPLE** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 39.- [RePEc:boc:bocins:ATHLET1](#) **ATHLET1** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 40.- [RePEc:boc:bocins:ATHLET2](#) **ATHLET2** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 41.- [RePEc:boc:bocins:ATTEND](#) **ATTEND** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 42.- [RePEc:boc:bocins:AUDIT](#) **AUDIT** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 43.- [RePEc:boc:bocins:BARIUM](#) **BARIUM** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 44.- [RePEc:boc:bocins:BWGHT](#) **BWGHT** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 45.- [RePEc:boc:bocins:CARD](#) **CARD** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 46.- [RePEc:boc:bocins:CEMENT](#) **CEMENT** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 47.- [RePEc:boc:bocins:CEOSAL1](#) **CEOSAL1** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 48.- [RePEc:boc:bocins:CEOSAL2](#) **CEOSAL2** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 49.- [RePEc:boc:bocins:CONSUMP](#) **CONSUMP** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 50.- [RePEc:boc:bocins:CORN](#) **CORN** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 51.- [RePEc:boc:bocins:CPS78_85](#) **CPS78_85** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 52.- [RePEc:boc:bocins:CRIME1](#) **CRIME1** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 53.- [RePEc:boc:bocins:CRIME2](#) **CRIME2** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 54.- [RePEc:boc:bocins:CRIME3](#) **CRIME3** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 55.- [RePEc:boc:bocins:CRIME4](#) **CRIME4** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 56.- [RePEc:boc:bocins:FAIR](#) **FAIR** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 57.- [RePEc:boc:bocins:EARNS](#) **EARNS** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 58.- [RePEc:boc:bocins:EZANDERS](#) **EZANDERS** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 59.- [RePEc:boc:bocins:EZUNEM](#) **EZUNEM** by Jeffrey M. Wooldridge [[Downloadable!](#)]

- 60.- [RePEc:boc:bocins:FERTIL1](#) **FERTIL1** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 61.- [RePEc:boc:bocins:FERTIL2](#) **FERTIL2** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 62.- [RePEc:boc:bocins:FERTIL3](#) **FERTIL3** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 63.- [RePEc:boc:bocins:FRINGE](#) **FRINGE** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 64.- [RePEc:boc:bocins:GPA1](#) **GPA1** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 65.- [RePEc:boc:bocins:GPA2](#) **GPA2** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 66.- [RePEc:boc:bocins:GPA3](#) **GPA3** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 67.- [RePEc:boc:bocins:HPRICE1](#) **HPRICE1** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 68.- [RePEc:boc:bocins:HPRICE2](#) **HPRICE2** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 69.- [RePEc:boc:bocins:HPRICE3](#) **HPRICE3** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 70.- [RePEc:boc:bocins:HSEINV](#) **HSEINV** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 71.- [RePEc:boc:bocins:INFMRT](#) **INFMRT** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 72.- [RePEc:boc:bocins:INJURY](#) **INJURY** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 73.- [RePEc:boc:bocins:INTDEF](#) **INTDEF** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 74.- [RePEc:boc:bocins:INTQRT](#) **INTQRT** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 75.- [RePEc:boc:bocins:INVEN](#) **INVEN** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 76.- [RePEc:boc:bocins:JTRAIN](#) **JTRAIN** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 77.- [RePEc:boc:bocins:KIELMC](#) **KIELMC** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 78.- [RePEc:boc:bocins:LAWSCH85](#) **LAWSCH85** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 79.- [RePEc:boc:bocins:LOANAPP](#) **LOANAPP** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 80.- [RePEc:boc:bocins:LOWBRTH](#) **LOWBRTH** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 81.- [RePEc:boc:bocins:MEAP93](#) **MEAP93** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 82.- [RePEc:boc:bocins:MLB1](#) **MLB1** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 83.- [RePEc:boc:bocins:MROZ](#) **MROZ** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 84.- [RePEc:boc:bocins:MURDER](#) **MURDER** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 85.- [RePEc:boc:bocins:NYSE](#) **NYSE** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 86.- [RePEc:boc:bocins:OPENNESS](#) **OPENNESS** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 87.- [RePEc:boc:bocins:PHILLIPS](#) **PHILLIPS** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 88.- [RePEc:boc:bocins:PNTSPRD](#) **PNTSPRD** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 89.- [RePEc:boc:bocins:PRISON](#) **PRISON** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 90.- [RePEc:boc:bocins:PRMINWGE](#) **PRMINWGE** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 91.- [RePEc:boc:bocins:RDCHEM](#) **RDCHEM** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 92.- [RePEc:boc:bocins:RECID](#) **RECID** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 93.- [RePEc:boc:bocins:RENTAL](#) **RENTAL** by Jeffrey M. Wooldridge [[Downloadable!](#)]

- 94.- [RePEc:boc:bocins:RETURN](#) **RETURN** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 95.- [RePEc:boc:bocins:SAVING](#) **SAVING** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 96.- [RePEc:boc:bocins:SLEEP75](#) **SLEEP75** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 97.- [RePEc:boc:bocins:SLP75_81](#) **SLP75_81** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 98.- [RePEc:boc:bocins:SMOKE](#) **SMOKE** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 99.- [RePEc:boc:bocins:TRAFFIC1](#) **TRAFFIC1** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 100.- [RePEc:boc:bocins:TRAFFIC2](#) **TRAFFIC2** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 101.- [RePEc:boc:bocins:VOLAT](#) **VOLAT** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 102.- [RePEc:boc:bocins:VOTE1](#) **VOTE1** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 103.- [RePEc:boc:bocins:VOTE2](#) **VOTE2** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 104.- [RePEc:boc:bocins:WAGE1](#) **WAGE1** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 105.- [RePEc:boc:bocins:WAGE2](#) **WAGE2** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 106.- [RePEc:boc:bocins:WAGEPAN](#) **WAGEPAN** by Jeffrey M. Wooldridge [[Downloadable!](#)]
- 107.- [RePEc:boc:bocins:WAGEPRC](#) **WAGEPRC** by Jeffrey M. Wooldridge [[Downloadable!](#)]

This information is provided to you by [IDEAS](#) at UQAM using [RePEc](#) data



IDEAS

RePEc

Search in IDEAS

Just put some keywords and Excite will show you... 1 to be in the searched
 file, type +word, while typing -word returns papers about inflation containing the word "money", but not "wages". As the entire html files are searched, keywords can also be
 papers about inflation containing the word "money", but not "wages". As the entire html files are searched, keywords can also be
 author or institution names. Excite also searches for synonyms or similar words. For example, house will also return entries about
 housing. Excite will even correct some spelling mistakes. Unfortunately, this makes some searches more difficult. For example,
 searches for stata will look for state. Try something more specific, for example stata module.

The following boolean operators are supported: AND, NOT, OR, AND NOT. Use parentheses to regroup them.

Accented characters are not equivalent to accent free characters. If you are looking for literature in a language other than English, especially French and German, use accents and umlauts.

Once you have search results, you can search for similar entries by clicking on the little icon next to a paper. Clicking on the paper title will give you the full bibliographic information of the paper.

If you have a preference for downloadable papers, enter downloadable as the first keyword.

Please note that the index for the search among all documents sometimes does not build correctly overnight. If you find abnormally few search results, please try the other searches.



Database	Archive	Series	Item
Fixed	3 letters	6 characters	Any (none if looking for series)
RePEc :	<input type="text"/>	:	<input type="text"/>
<input type="button" value="Search among Handles"/>			

IDEAS search results

[\[New search\]](#)

 - high similarity,  - low similarity; click on the icon to search for similar documents

Excite: documents trouvés: **murder and data** [\[Nouvelle recherche\]](#)

 - haute similitude,  - basse similitude; cliquer sur l'icône pour rechercher les documents similaires

 78% **MURDER**

Sommaire : [Downloadable!] Author(s): Jeffrey M. Wooldridge Abstract: Instructional dataset, N=153, longitudinal state murder rate data Accompanying Introductory Econometrics: A Modern Approach, Jeffrey M. Wooldridge, South-Western College Publishing, (c) 2000 Datasets also accessible from <http://wooldridge.swcollege.com>.

 74% **Estimating the Economic Model of Crime in the Presence of Organised Crime : Evidence from Italy.**

Sommaire : Author(s): Marselli, R & Vannini, M Abstract: The economic model of crime is tested using a panel dataset of Italian regions for the period 1980-1989. Four different crimes are considered : murder, theft, robbery and fraud.

 70% **The Determinants of Punishment: Deterrence, Incapacitation and Vengeance**

Sommaire : [Downloadable!] Author(s): Edward L. Glaeser & Bruce Sacerdote Abstract: Does the economic model of optimal punishment explain the variation in the sentencing of murderers? As the model predicts, we find that murderers with a high expected probability of recidivism receive longer sentences. Sentences are longest in murder types where apprehension rates are low, and where deterrence elasticities appear to be high. However, sentences respond to victim characteristics in a way that is hard to reconcile with optimal punishment. In particular, victim characteristics are important determinants of sentencing among vehicular homicides, where victims are basically random and where the optimal punishment model predicts that victim characteristics should be ignored. Among vehicular homicides, drivers who kill women get 56 percent longer sentences. Drivers who kill blacks get 53 percent shorter sentences.

 70% **Crime and Social Interactions**

Sommaire : [Downloadable!] Author(s): Edward L. Glaeser & Bruce Sacerdote & Jose A. Scheinkman Abstract: The high degree of variance of crime rates across space (and across time) is one of the oldest puzzles in the social sciences (see Quetelet (1835)). Our empirical work strongly suggests that this variance is not the result of observed or unobserved geographic attributes. This paper presents a model where social interactions create enough covariance across individuals to explain the high cross-city variance of crime rates. This model provides a natural index of social interactions which can compare the degree of social interaction across crimes, across geographic units and across time. Our index gives similar results for different data samples and suggests that the amount of social interactions are highest in petty crimes (such as larceny and auto theft), moderate in more serious crimes (assault, burglary and robbery) and almost negligible in murder and rape. The index of social interactions is also applied to non

 02% **Evaluating Real Business Cycle Models: the Data Transformation Problem**

Sommaire : [Downloadable!] Author(s): John Landon-Lane Abstract: There are many methods used for evaluating and comparing models found in the Real Business Cycle (RBC) literature. One major problem faced is how to transform the data into a form acceptable to the model being evaluated, the most commonly used being a filter such as Hodrick-Prescott. Models are evaluated by comparing the filtered data that are observed with filtered data that are simulated. A number of authors have criticised the use of the Hodrick-Prescott filter for distorting the data, leading to other suggestions. This paper formally compares and evaluates competing methods for transforming the observed data using a likelihood-based approach. The paper also describes a method for transforming the data directly to a form that is similar to that predicted by a model. Thus a direct evaluation of a model with the data can be obtained. Since the method of comparison is likelihood based, the competing methods are compared across the full dimension of t

 02% **Validation of Simulation, with and without Real Data.**

Sommaire : Author(s): Kleijnen, J.P.C. Abstract: This paper gives a survey on how to validate simulation models through the application of mathematical statistics. The type of statistical test actually applied, depends on the availability of data on the real system: (i) no data, (ii) only output data, (iii) both input and output data.

 02% **Use of Bad Training Data**

Here is the dataset descriptor page for the Wooldridge “MURDER” dataset, and its Web-accessible codebook. There are no codebooks for the Mills datasets, but each “abstract” contains a listing of the variables in the dataset.

This file is part of [IDEAS](#), which uses [RePEc](#) data

[[Series](#) | [JEL Classification](#) | [Search](#) | [Viewers](#) | [Help!](#)]

If a paper is available in full text, a red [Downloadable!] appears just below and download information is at the bottom of its page between two lines. If not, see [Help!](#)

[Downloadable!]

Are you an author of this paper? Please take the time and register at our new [HoPEc service](#) and link yourself with your work listed here. Note that you do not need to register in order to use the search service!

Do you want more of your research listed here? See how you can [participate!](#)

To be notified about new online papers in your field of interest, subscribe to the free [NEP reports!](#)

MURDER

Jeffrey M. Wooldridge

Instructional dataset, N=153, longitudinal state murder rate data Accompanying Introductory Econometrics: A Modern Approach, Jeffrey M. Wooldridge, South-Western College Publishing, (c) 2000 Datasets also accessible from <http://wooldridge.swcollege.com>.

Handle: RePEc:boc:bocins:MURDER

Paper provided by [Boston College Department of Economics](#) in its series [Instructional Stata datasets for econometrics](#)

If you experience problems downloading a file check if you have the proper [application](#) to view it first. Information about this may be contained in the File-Format links below. In case of further problems read [the IDEAS help file](#)

[Download this File!](#)

Address: <http://fmwww.bc.edu/ec-p/data/wooldridge/MURDER.des>

Format: text/plain

Function: codebook for dataset

Size:

Restrictions (if any):

[Download this File!](#)

Address: <http://fmwww.bc.edu/ec-p/data/wooldridge/MURDER.dta>

Format: application/x-Stata

Function: Stata v6.0 web-accessible dataset

Size:

Restrictions (if any):

This information is provided to you by [IDEAS](#) at UQAM using [RePEc](#) data

MURDER.DES

id	state	year	mrdrte	exec	unem	d90	d93
cmrdrte	cexec	cunem	cexec_1	cunem_1			

Obs: 153

1. id	state identifier
2. state	postal code
3. year	87, 90, or 93
4. mrdrte	murders per 100,000 population
5. exec	total executions, past 3 years
6. unem	annual unem. rate
7. d90	=1 if year == 90
8. d93	=1 if year == 93
9. cmrdrte	mrdrte - mrdrte[t-1]
10. cexec	exec - exec[t-1]
11. cunem	unem - unem[t-1]
12. cexec_1	cexec[t-1]
13. cunem_1	cunem[t-1]

Having decided that we want to use this MURDER dataset in a class problem set, we need merely click upon the link at the foot of the dataset descriptor page. The web browser will launch Stata and 'use' the selected dataset with one click. If you have not downloaded .dta files in your browser, you may have to instruct the browser to deal properly with a .dta file by handing it off to Stata.

```
running :Keewaydin:Rumelihisari:Stata:profile.do ...
```

```
. use ":Keewaydin:Desktop Folder:MURDER.dta"
```

```
. desc
```

```
Contains data from :Keewaydin:Desktop Folder:MURDER.dta
```

```
obs:          153  
vars:          13          26 Jan 2000 12:16  
size:          8,568 (100.0% of memory free)
```

```
-----  
 1. id         float   %9.0g  
 2. state      float   %9.0g  
 3. year       float   %9.0g  
 4. mrdрте     float   %9.0g  
 5. exec       float   %9.0g  
 6. unem       float   %9.0g  
 7. d90        float   %9.0g  
 8. d93        float   %9.0g  
 9. cmrdрте    float   %9.0g  
10. cexec      float   %9.0g  
11. cunem      float   %9.0g  
12. cexec_1    float   %9.0g  
13. cunem_1    float   %9.0g  
-----
```

```
Sorted by:
```

Naturally, it would be helpful to have metadata for statistical datasets separated from metadata on working papers in a RePEc service such as IDEAS, and separately searchable in that service. That enhancement will be feasible when a separate ReDIF template type is devised and added to the RePEc system and its services. Design of the elements of such a template involves classification issues: what attributes of a dataset ought to be captured in keyword form? Study of existing dataset archives elsewhere on the Web may assist us in this effort.

Just as the SSC-IDEAS archive contains references to “local” software modules (those housed on the Boston College server) it may also contain links to those modules wherever on the Web they may be located. In this manner, authors may choose to house their software on their own web page, and use links from SSC-IDEAS to publicize and describe their software. One limitation: for security reasons, a .pkg file cannot reference software on a different server. Thus, those modules described in SSC-IDEAS which are not housed locally cannot be installed via “archinst” .

By the same logic, a metadata archive for statistical datasets need not house the datasets themselves; it need only contain references to their location (their URLs). Although each of the datasets currently included in the Instructional Stata datasets for econometrics series are held locally, that is not in any way necessary—especially since Stata can “use” a dataset anywhere on the Web. Thus, the further development of a collection of freely available statistical datasets useful for instruction and/or research need only involve the creation of templates describing the datasets’ characteristics and specifying their locations.

How can you participate in RePEc?

You may make use of SSC-IDEAS to access the latest developments in user-written components for Stata and several other programming languages.

If you use Stata in your work and develop materials that you want to share with other Stata users, you may contribute them to the SSC-IDEAS archive, and note their availability on StataList.

The latter part of this presentation has argued that the RePEc system may prove equally useful for the description, classification and dissemination of statistical datasets. I welcome your suggestions for the design of a ReDIF “template type” that would contain standard classifying features of a dataset. Any existing collection of datasets might be incorporated into an expanded series of instructional datasets. It would be very simple to create a number of series, each containing certain types of data, to further categorize collections of statistical datasets. If collections of datasets are already described by Web pages, those pages can generally be transformed into ReDIF templates mechanically with a minor effort of customization in perl.

For more information on the RePEc system—which currently contains metadata on working papers, published journal articles, software components, personal data on many authors, and, now, information on a number of statistical datasets for instructional use—please visit **<http://repec.org>**.

Kit Baum for the RePEc team