AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
http://ageconsearch.umn.edu
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

The Stata Journal (2010)
10, Number 1, pp. 3–8

A conversation with Kit Baum

Nicholas J. Cox
Department of Geography
Durham University
Durham, UK
n.j.cox@durham.ac.uk

Abstract. As explained in the *Editorial announcements* at the start of this issue, the *Stata Journal* will be featuring various interviews with selected members of the Stata community in 2010, Stata's 25th anniversary year. In this issue, we start with an online interview, dated October 14, 2009, of Christopher F. "Kit" Baum, an economics professor at Boston College. Kit has been highly active as a Stata user since the 1990s, contributing as an author and associate editor to this journal and its predecessor, the *Stata Technical Bulletin*; as a much-downloaded Stata program author; as an author of two notable Stata-based texts; as a frequent participant on Statalist and at Users Group meetings in both the United States and several European countries; and as founder and maintainer for more than a decade of the SSC archive, which now contains many hundred user-written Stata packages. In this interview, he comments on how he got into Stata and adds his own speculations for the future.

Cox: Tell us a bit about your background. How did you get into economics as a student?

Baum: I became interested in economics as a freshman in college, around the time I figured out that I was not cut out to be a physics major. I ended up with an elective course in economics in the third quarter of that academic year, taught by Professor Phil Thomas, who became my senior thesis advisor. Despite an unimpressive start, I never looked back.

Cox: Do you see yourself as an economist or an econometrician?

Baum: I am an economist who spends a lot of time doing applied econometrics and programming, both statistical/econometric and database applications. I used to teach monetary theory and policy, macroeconomics, and financial derivatives regularly, but in the last several years, I have been teaching nothing but various flavors of econometrics to undergraduates and PhD students. Despite the fact that my most highly cited article (by a mile) is Baum, Schaffer, Stillman, *Stata Journal* (SJ) 2003[1], most of my recently published research has been oriented toward the effects of uncertainty on firms' behavior, banks' behavior, and international trade flows.

1. Baum, C. F., M. E. Schaffer, and S. Stillman. 2003. Instrumental variables and GMM: Estimation and testing. *Stata Journal* 3: 1–31.

© 2010 StataCorp LP

Cox: How did you get started in computing? What hardware and software were you using?

Baum: Early in my college career, I took a computer science course—the only course offered in the late 1960s!—and became intimately acquainted (as did Bill Gould) with an IBM 1620 Model II. I learned Fortran II (the only thing beyond assembler and machine language which that machine supported), worked in the computer center as my work-study job, and did quite a bit of programming. My senior thesis in economics involved developing some interactive programs for teaching economic models. As "dumb terminals" were only then becoming available, this was quite impractical, as one had to run the modeling programs on the 1620s console typewriter, but I did get top marks for effort.

When I became a PhD student at Michigan–Ann Arbor, I earned my keep by programming for two years as Professor Bob Stern's assistant, working on various data-management and modeling exercises related to international trade and tariffs. I put those programming skills to good use in my doctoral dissertation, written with Professor Phil Howrey, applying Gregory Chow's techniques for optimal control of discrete-time models to macroeconomic modeling in the presence of parameter uncertainty (a computationally burdensome problem on mainframes of the mid-1970s). That work was all done in Fortran as well. Some estimation was done in mainframe TSP.

Cox: How did you get into Stata? What other econometrics or statistics software were you using at the time?

Baum: Several years after I joined the Boston College economics faculty, one of my colleagues was working on a research project with a coauthor at Harvard or MIT which made heavy use of panel data. He came to me and insisted that we purchase and support this package called Stata, which I had not heard of before. At the time, I was teaching the first-year PhD econometrics course using Ken White's Shazam package, requiring the students to do a number of exercises and an empirical research paper. Shazam had a number of good features, but also was quite clumsy in some ways, so I migrated to the RATS package. It wasn't much more popular with the students but provided tools like vector autoregressions (VARs) and impulse response functions (IRFs) that no other package had at the time. That first version of Stata, which the department purchased for a Unix server, was version 5. I switched to using Stata in the first-year PhD econometrics course. The next time I taught the undergraduate econometrics course, I used Stata. Since that time, I have used Stata exclusively in teaching, and it is used by my colleagues in all sections of undergraduate econometrics. I still turn to RATS occasionally for specialty capabilities such as multivariate GARCH, but it may be that Stata's new commands in version 11 can do what I need in that area as well.

Cox: A very big contribution you have made to the Stata community is setting up and maintaining the SSC archive over several years. Tell us how that came about.

Baum: My involvement here came as a side effect of becoming involved with RePEc, Research Papers in Economics, a volunteer organization that has become the largest col-

lection of economics-related bibliographic information. RePEc was started by Thomas Krichel to enable the sharing of working papers (preprints), which are the predominant mode of rapid communication of research results in economics. Boston College Economics was one of the earliest participants in this effort, which was an outgrowth of earlier efforts in the United Kingdom. It became clear that we could not only document (and provide downloads for) working papers, but we could also document journal articles, monographs—and in a blinding flash, it came to me: why not software?

This was a controversial notion, and one that was somewhat of a hard sell to the rest of the RePEc team: why would we create software archives and commingle them in a sense with working paper and journal article archives? If it were not for Stata, this concept would probably not have emerged. Many other kinds of software have their own repositories (CPAN for Perl, for instance). Some software vendors host archives (e.g., MATLAB routines at the MathWorks or RATS routines at Estima). Stata software, in the form of ado-files and help files, comprised complete routines: if well-written, far more useful than most software circulating for other statistical packages and languages, in that Stata user-written commands are "first-class citizens" once installed on the adopath.

At the time (1998), Stata routines were "published" as electronic supplements to the *Stata Technical Bulletin* (STB) and disseminated on 3.5-inch floppy disks mailed with the hardcopy STB. The latest user-written routines were included as in-line messages to Statalist, but the code was often mangled by line wraps and the like. It seemed that having an archive from which these files could be downloaded would make sense, and the RePEc architecture supported that as soon as the other team members grudgingly accepted my proposals to create a "software template" that could provide bibliographic information about software. That was not limited to Stata, of course; any source code can be housed in the archive, and there is a nontrivial number of components in the SSC archive in other languages (however, binary code in the form of `.exe` is banned for security reasons).

The popularity of the SSC archive as a single source for user-written Stata commands was obviously not what StataCorp intended when it created the `net` commands, but since then the market has spoken. Most user-programmers find it a useful alternative to maintaining their own `net` site. Nick Cox and I wrote `archutil` (STB, 1999,[2] 2000[3]) to implement `net` for the SSC archive, and StataCorp responded with the full-featured `ssc` command. More recently, the development of `adoupdate` has made it as easy to stay up to date with user-written software as `update` does with official Stata. It has been a very fruitful collaboration.

2. Baum, C. F., and N. J. Cox. 1999. ip29: Metadata for user-written contributions to Stata programming language. *Stata Technical Bulletin* 52: 10–12. Reprinted in *Stata Technical Bulletin Reprints*, vol. 9, pp. 121–124. College Station, TX: Stata Press.
3. Cox, N. J., and C. F. Baum. 2000. ip29.1: Metadata for user-written contributions to Stata programming language: Extensions. *Stata Technical Bulletin* 54: 21–22. Reprinted in *Stata Technical Bulletin Reprints*, vol. 9, pp. 124–126. College Station, TX: Stata Press.

Cox: Do you view yourself as a gatekeeper in any sense over SSC?

Baum: I would have reacted to this question "of course not", but on this very day, I found myself looking at the ado-file of a submission and composing an email to the author indicating that without several changes I could not accept the routine (e.g., thou shalt not wantonly strew global macros about the user's workspace). The author kindly complied and provided a new version, dealing with this and a couple of other issues, the same day. But generally, I do not read the code, especially that of well-known authors. It is clearly stated that the material in the archive bears no warranty by virtue of its inclusion. If clear flaws are pointed out, and an author does not respond to a request to make corrections to the routine, I will remove it—but that has very rarely happened.

Cox: You have contributed several much-used Stata commands yourself. Do you have any particular favorites?

Baum: Of course, I am very fond of `ivreg2`, although Mark Schaffer has contributed much more of the inventive programming to that routine as it has become more and more complex. I am proud of the fact that StataCorp was compelled to improve official Stata's capabilities in this area, developing the `ivregress` routine. Between `ivregress` and `ivreg2`, I believe Stata has the best feature set for dealing with instrumental-variables estimation of any available econometric software. Apart from `ivreg2`, my major contributions have been in the time-series area: the first implementation of vector autoregressions (`vecar`, made wholly obsolete by Stata's `var` suite); rolling regressions (`rollreg`, largely superseded by the `rolling:` prefix); `dfgls`, with Richard Sperling, adopted by StataCorp; `kpss`, `modlpr`, and `roblpr` tests for long memory; and several unit-root tests, now available in official Stata 11. I believe that the popularity of several of these routines has had a real influence on the course of development at StataCorp and, in particular, on the strengthening of its capabilities in the time-series domain. That makes Stata much more competitive with those packages and languages which have traditionally been used for time-series work. Being able to do that work while remaining in the Stata environment is very helpful.

Cox: You have presented at many Stata Users Group meetings. Perhaps you could say something to those who have never been about what they are like.

Baum: The Stata Users Group meetings are always enjoyable experiences for new users and experienced users alike. The user presentations are an eclectic mix of topics from a broad set of disciplines, which is very interesting for those of us who only attend economics and finance seminars. The techniques of interest in medical fields, for instance, often provide inspiration for adapting some of that methodology to good advantage in my own research, although I do not routinely peruse the biostatistics literature. Ample time is given in coffee breaks, luncheons, and dinners for participants to chat with speakers, both users and StataCorp staff, about their work. Last summer's DC09 Conference was quite large, with over 100 present each day; but even in that venue, there were many opportunities for fruitful interactions.

Cox: You have recently written two fat books for Stata Press...

Baum: Well, not fat in comparison to other recent Stata Press offerings, but long enough to cover the subject in reasonable detail, I believe. I am exceedingly pleased and humbled by the popularity of *An Introduction to Modern Econometrics Using Stata*.[4] As Austin Nichols' excellent SJ review[5] indicates, it is not a textbook from which one learns econometrics, but a guide for those who know some econometrics and need guidance beyond theory in applying it to their research. Those of us who do empirical research realize that a small fraction of the research effort involves actual estimation, with a sizable set of data-management tasks preceding estimation and analysis of econometric results. Likewise, a research project often entails a lot of work in producing tables, graphs, and the like, and that should be automated. The book stresses how both of those important ancillary activities are important adjuncts to performing the proper estimation and conducting the appropriate diagnostic tests. I look forward to reworking the book someday to include discussion of the new features added in Stata 11, such as factor variables and margins, that further enhance researchers' efforts.

I am also very pleased to see that *An Introduction to Stata Programming*[6] has been quite popular. I have always thought that StataCorp did not make it easy to get into Stata programming, as most users did not have the *Stata Programming Reference Manual* available, and numerous important details were only available in that manual. With the enlightened change in Stata 11, making all the manuals available to every Stata user in very convenient PDF, the user experience has radically changed. Users are now much more able to use do-file programming to good advantage with the help of the manuals and a copy of ISP.

Cox: By the way, we gather that you are a Macintosh user...

Baum: I would say that I am a Unix user. During my 30+ years at Boston College (my one and only academic employer), I have always used Unix machines for serious computing. In the early years, that meant Digital Equipment Corporation's flavor of Unix known as VMS. When we first acquired a "compute server" for academic computing, it was IBM's AIX. Later, we acquired a multiprocessor Sun server, and that provided access to Stata with a useful amount of memory and disk space. When that machine became obsolete, we replaced it with an Apple xServe. In addition to that machine, we have a Mac Pro with four-core Intel Xeon to run Stata/MP 11, and the university's research computing unit has a Stata/MP license for their Linux cluster.

The point here is not that I use Macintosh: the point is that I use Unix tools (including the console version of Stata) heavily, and I find that Mac OS X is a convenient way to use those tools. Linux would be a good choice too, although not quite as well supported in

4. Baum, C. F. 2006. *An Introduction to Modern Econometrics Using Stata*. College Station, TX: Stata Press.
5. Nichols, A. 2007. Review of An Introduction to Modern Econometrics Using Stata by Baum. *Stata Journal* 7: 131–136.
6. Baum, C. F. 2009. *An Introduction to Stata Programming*. College Station, TX: Stata Press.

terms of software. But the innate stability and professional quality of Unix/Linux/Mac OS X operating systems stands in stark contrast to the various versions of Windows. I've heard Bill Gould tell us that Windows Vista is actually a great OS, but his beliefs do not seem to have been shared by the marketplace. Windows 7 may be a great choice, but I'd rather rely on an operating system that has evolved over several decades, whose flaws are found by programmers rather than end-users. But one interesting feature of Apple hardware these days is that it can run any flavor of Windows (and Linux, for that matter) in addition to Mac OS X. Its industrial design is also highly innovative, and many of the features common to all machines these days were pioneered by Apple. So I am quite happy to use Apple hardware and software, and never averse to moving my compute-bound tasks over to Linux if I need more horsepower. Like about half of my colleagues in the economics department, I just don't use Windows.

Cox: How do you see Stata use developing on your side of statistical science?

Baum: I hope that with the aggressive development of new econometric capabilities, evident in the long list of enhancements to Stata 11, more researchers in economics and finance will find Stata to be a very useful platform. Those who use Monte Carlo simulation regularly are often unaware that Stata makes that very straightforward and efficient. My colleagues who use vector autoregressions and related techniques are not fully aware that Stata is very competitive in this area. The addition of state-space modeling, multivariate ARCH, and related time-series techniques should make reliance on Stata more workable for those of us with needs in that area, and the addition of the `gmm` command is an important step forward for our field of research. Nevertheless, I still believe that several important capabilities are lacking: for instance, estimation and forecasting of nonlinear simultaneous models (NL3SLS or FIML), support for contour plots/likelihood profile plots, and support for some forms of 3D graphics. These are areas where one must be apologetic when colleagues ask "can Stata do ...". I hope to see developments in these areas, as well as official support of some capabilities for publication-quality output. On the whole, though, StataCorp has been highly responsive to researchers' expressed needs, and that explains a lot of its growing popularity among researchers in economics, finance, and related disciplines. As an economist, I must point out that the reasonable price tag certainly doesn't hurt. I look forward to further enhancements that will further improve my ability to perform research tasks efficiently and effectively.

Cox: Thank you, Kit, for your time and thoughts.