

<b>Institution: Heriot-Watt University</b>
<b>Unit of Assessment: C19: Business and Management Studies</b>
<b>Title of case study:</b> Enabling The Econometric Analysis Of Policy Interventions
<p><b>1. Summary of the impact</b> (indicative maximum 100 words)</p> <p>The research created programs for the Stata statistical software environment that are used by thousands of researchers in economics and other fields around the world in academia, the private sector, government and quasi-governmental organisations, with approximately 400,000 downloads in the REF 2014 period. The core programs enable researchers to rigorously analyse the causal impact of a policy in settings where an experiment is infeasible and for experiments where take-up of treatment is incomplete, i.e. for the settings in which the vast majority of empirical work is done. The programmes are used to analyse complex data to establish causal links across a broad range of policy areas.</p>
<p><b>2. Underpinning research</b> (indicative maximum 500 words)</p> <p>Research in the area of econometrics in the past 30 years has moved very quickly. Many advances have proven to be very useful for applied research. An important obstacle to wide adoption of these techniques has been the availability of new estimation and testing methods in commercially-available software packages. Such techniques are typically not introduced by commercial providers until the technique is generally accepted. Implementation of these sophisticated techniques in a software environment is, moreover, a challenging exercise that itself requires a substantial research investment.</p> <p>Instrumental variables (IV) estimators have followed this pattern. They are widely used by economists and other empirical researchers when the econometric problem of endogeneity is present. Endogeneity means that standard methods of estimating the causal impact of a policy cannot be used because the policy or “treatment” is non-random. The importance of the problem, and the centrality of IV methods in addressing them, is reflected in the substantial coverage given to IV estimation and its generalisation, the Generalised Method of Moments (GMM), in undergraduate and especially graduate econometrics textbooks. But because developments in the theory of IV estimation have been rapid, availability of cutting-edge methods has been patchy, and many recent advances are simply unavailable to empirical researchers who wish to make use of them.</p> <p>Starting in 2003, Schaffer, working primarily in collaboration with Christopher Baum (Boston College) and also with Steven Stillman (University of Otago, New Zealand) and Frank Kleibergen (Brown University) amongst others, developed and implemented a comprehensive set of IV/GMM estimation and testing procedures for the Stata statistical software package. These routines include single-equation and panel data IV/GMM estimation, robust covariance estimation (heteroskedastic, autocorrelation and 1- and 2-way cluster-robust), specification tests for IV estimation (heteroskedasticity, autocorrelation, RESET, overidentification, endogeneity, weak identification), and weak-instrument-robust inference. These procedures brought the range of methods available to researchers working in the Stata environment to the frontier of econometric “best-practice”. The capabilities of the software and the range of packages have been and continue to be regularly extended; recent examples are the facilities for weak-instrument-robust inference, the introduction of 2-way cluster-robust covariance estimation (a technique that was published in the econometric literature only several years ago) and IV estimation using heteroskedasticity-based instruments (the core theory for which was published only in 2012).</p>
<p><b>3. References to the research</b> (indicative maximum of six references)</p> <p>[1] Schaffer, M. E., Baum, C., Finlay, K., Kleibergen, F., Magnusson, L. &amp; Stillman, S. 2013 <u>Stata software for econometric estimation and testing; avar, weakiv, actest, ivreg2h, ranktest, ivreg2</u></p> <p>[2] Christopher F Baum &amp; Mark E. Schaffer &amp; Steven Stillman, 2003. "<u>Instrumental variables and GMM: Estimation and testing.</u>" <i>Stata Journal</i>, StataCorp LP, vol. 3(1), pages 1-31, March. Available on request</p>

## Impact case study (REF3b)

[3] Christopher F Baum & Mark E. Schaffer & Steven Stillman, 2007. "Enhanced Routines for Instrumental Variables/Generalized Method of Moments Estimation And Testing," Stata Journal, StataCorp LP, vol. 7(4), pages 465-506, December. Available on request.

#### 4. Details of the impact (indicative maximum 750 words)

Stata is a statistical package used worldwide with a large user community. An important feature of Stata is that StataCorp supports integration of user-written software. This has meant that the programs and supporting documentation developed by Schaffer et al. can be and have been easily found and installed by users.

Take-up of the routines by Schaffer et al. has been very substantial. The software is stored on a service run by Boston College, "Statistical Software Components" (SSC), in collaboration with RePEc (Research Papers in Economics, [www.repec.org](http://www.repec.org)). SSC tracks all software downloads/installations; these details are available via the SSC server. (NB: downloads reported on the RePEc website include only those installations done "by hand" rather than automatically by the Stata package, and understate the number of installations by roughly a factor of 8-10.) As of August 2013, the software described above had been downloaded 470,000 times, with over 400,000 of these downloads during the REF period [S5]. The most widely-used package is the core estimation routine, "ivreg2", downloaded over 130,000 times during the REF period. These downloads include both first-time installations and downloads by existing users of updates to routines.

The programs written by Schaffer et al. are now well established in non-academic institutions, particularly in governmental and quasi-governmental organisations in the US, as well as in academia. A partial list of institutions where use has been documented:

- Federal Reserve Board
- International Monetary Fund (IMF)
- Mathematica Policy Research, Inc.
- RAND Corporation
- Urban Institute (Washington, DC)
- US Census Bureau
- US Department of Agriculture (USDA)
- US Federal Trade Commission (FTC)
- US Government Accountability Office (US GAO)
- World Bank

A Lead Economist in the World Bank's Research Department explained how he uses the programs: "*In my work where I have to deal either with very large cross-sections, or less frequently, with panel data, several softwares developed by Mark were absolutely indispensable. ... [Our] paper investigated prevalence of corruption in transition economies and argued that it should be less if there were more frequent changes in government (alternation in government). We faced a tough problem of how to find instruments for alternation in power because corruption could in turn affect alternation (ruling party losing election because it is corrupted) and thus the causation would run the other way round. In (another) paper, the problem was similar: here I looked at decile income shares in practically all countries of the world (ten regressions per each country; one per decile) and tried to determine how they were affected by trade/GDP ratios. I used GMM-IV approach, again as developed by Mark. I also used the same software in a paper written with [a colleague] that looked at decile shares in transition countries and the effect that privatization (among other variables) had on them*" [S2]

A Senior Research Associate from the Urban Institute described the extensive use they make of the software. "*We have completed hundreds of studies where our analytical methods have been significantly affected by your work. The software has been used to analyze the impact of mental health treatment on work, and the impact of housing vouchers on family economic status and health. The current frontier of policy analysis quite simply would be far inside its current boundaries without the development of the IV regression technology at Heriot-Watt University.*" [S3]

## Impact case study (REF3b)

**5. Sources to corroborate the impact** (indicative maximum of 10 references)

[S1] Professor, Department of Economics, Boston College will confirm that the software is stored on a service run by Boston College, "Statistical Software Components" (SSC), in collaboration with RePEc (Research Papers in Economics, [www.repec.org](http://www.repec.org)). SSC tracks all software downloads/installations; these details are available via the SSC server.

[S2] Lead Economist, Development Research Group, World Bank will describe the application of the softwares in a broad range of complex policy environments with very large cross-sections, or with panel data. The softwares make the difficult analysis possible with a degree of confidence.

[S3] Senior Research Associate, Urban Institute will confirm that they have completed hundreds of studies where analytical methods have been significantly affected by the work. The software has been used to analyze the impact of mental health treatment on work, impact of housing vouchers on family economic status and health.

[S4] Senior Survey Statistician, Abt SRBI <http://srbi.com> The senior survey statistician will confirm the statistics cited.

[S5] Statistics for verification:

Summary statistics of Stata downloads are regularly announced on the Statalist discussion site, e.g., <http://www.stata.com/statalist/archive/2013-01/msg00114.html>

Raw data on all downloads can be obtained in Stata format files for January 2008 onwards from the RePEc website. URLs are <http://repec.org/docs/sschotPXXX.dta>, where XXX=576 for January 2008, 577 for February 2008, 642 for July 2013.

(Note: the download statistics cited above include *all* downloads, including those from within the Stata software environment using the "ssc install" and "adoupdate" commands.

Download statistics listed on the LogEc website of RePEc

(<http://logec.repec.org/scripts/itemstat.pf?type=redif-software>) cover only those download done "by hand" and do *not* include those from within the Stata software environment.)

[S6] <http://www.rand.org/pubs/monographs/MG829.html> Nancy Nicosia, Rosalie Liccardo Pacula, Beau Kilmer, Russell Lundberg, James Chiesa (2009), "The Economic Cost of Methamphetamine Use in the United States, 2005", Rand Drug Policy Research Center. "This study was sponsored by the Meth Project Foundation and the National Institute on Drug Abuse and was conducted under the auspices of the Drug Policy Research Center, a joint endeavour of RAND Infrastructure, Safety, and Environment and RAND Health." Citation on p. 112: "All of these instruments were assessed in terms of their validity, exogeneity, and exclusion criteria using linear probability models estimated using ivreg2 in Stata 8.2."