



Journal of Statistical Software

July 2008, Volume 27, Issue 1.

<http://www.jstatsoft.org/>

Econometrics in R: Past, Present, and Future

Achim Zeileis

Wirtschaftsuniversität Wien

Roger Koenker

University of Illinois at Urbana-Champaign

Abstract

Recently, computational methods and software have been receiving more attention in the econometrics literature, emphasizing that they are integral components of modern econometric research. This has also promoted the development of many new econometrics software packages written in R and made available on the Comprehensive R Archive Network. This special volume on “Econometrics in R” features a selection of these recent activities that includes packages for econometric analysis of cross-section, time series and panel data. This introduction to the special volume highlights the contents of the contributions and embeds them into a brief overview of other past, present, and future projects for econometrics in R.

Keywords: econometrics, open-source software, R.

1. Introduction

Stigler’s (1994) study of citation patterns in statistical research documents an impressive inflow of ideas from econometrics journals. Among the journals he surveys, *Econometrica* is the most influential according to his “balance of trade” measure. And yet econometrics has lagged behind other statistical fields in embracing computational methods as an intrinsic part of the research process. One can attribute this reticence to many factors: failure of disciplinary mechanism design, excessive faith in market solutions, general indolence, and so forth. But we have detected a recent change in the attitude of econometricians toward computational methods, an increasing willingness especially among younger researchers to regard software development and dissemination as a natural concomitant of econometric research. This change in attitude is also reflected in various journals publishing papers in the field of computational econometrics—e.g., in *Computational Economics* or several special issues of *Computational Statistics & Data Analysis*—and econometric software—e.g., the software review section of the *Journal of Applied Econometrics*.

Certainly, recent developments in computer hardware, software and networking have dramat-

ically lowered the costs of such activities and increased their potential benefits. New methods that used to wait for years for commercial implementations can now be shared almost immediately over the Web. Writing software for general use (and abuse) is, of course, quite different than writing code for one-off research projects, but it has its own charms and rewards. The R project (R Development Core Team 2008) has made it considerably easier to undertake such development efforts, providing standards and templates for documentation, version control, and consistency checking. R is still relatively obscure in the econometrics community, but has a devoted and growing following.

2. Past

As a language with a rich toolbox for matrix-based computations R is a natural candidate for writing econometric software (Cribari-Neto and Zarkos 1999) and its open-source license makes it attractive for teaching econometrics (Racine and Hyndman 2002). Nevertheless, only relatively few packages available from the Comprehensive R Archive Network (CRAN, <http://CRAN.R-project.org/>) are devoted to econometrics explicitly which is also reflected in the fact that only few such packages have been published in the *Journal of Statistical Software* (JSS). Notable exceptions include the packages **strucchange** (Zeileis, Leisch, Hornik, and Kleiber 2002), **sandwich** (Zeileis 2004, 2006), and **systemfit** (Henningesen and Hamann 2007).

In the last couple of years, however, the number of econometrics-related packages on CRAN grew (see e.g., the econometrics CRAN task view at <http://CRAN.R-project.org/view=Econometrics>) and several of these packages were presented by their authors at econometrics sessions of the useR! conferences in 2004 (<http://www.R-project.org/conferences/useR-2004/>) and 2006 (<http://www.R-project.org/conferences/useR-2006/>). At the latter conference, held at Wirtschaftsuniversität Wien in Vienna, Austria, the JSS editors decided to encourage the authors from several sessions to put together “Foometrics in R” special volumes—with values of “foo” including “chemo” (volume 18), “psycho” (volume 20), “environ” (volume 22), and “econo” (volume 27), among others. We hope that the packages/papers in this special volume will stimulate further interest in R development in econometrics.

3. Present

This special volume features some recent efforts to complement the econometrics functionality available in R. Some of the packages/papers fill “gaps” in the R toolbox for econometrics by providing standard tools previously unavailable in R, others present original implementations of cutting edge methods.

In the first paper, Croissant and Millo (2008) describe their **plm** package for analysis of panel data using linear models. It starts out with basic panel models estimated via ordinary least squares (OLS)—including the familiar within, between, and random effects models—and continues to present some extensions including models estimated by generalized methods of moments (GMM) or feasible generalized least squares (FGLS).

The two subsequent papers are devoted to the analysis of time series. Hyndman and Khandakar (2008) focus on automatic forecasting of univariate time series using their **forecast** package. Specifically, functions for exponential smoothing methods are presented in a general

state space model framework, as well as tools for automatically selecting ARIMA models. Pfaff (2008) contributes software for modeling and forecasting multivariate time series using (structural) vector autoregressive (VAR and SVAR) and structural vector error correction (SVEC) models in package **vars**. In addition to standard fitting and prediction functionality, a set of diagnostic procedures—based on both, significance tests and visualizations—is discussed.

The remaining contributions to the special volume present various packages for regression, including methods for binary, count, and censored data which are probably particularly useful in microeconometrics. Hayfield and Racine (2008) implement a rich collection of tools for nonparametric and semiparametric econometrics using a kernel-based approach with data-driven bandwidth selection. In particular, their **np** package includes functions for regression (nonparametric, partially linear, single index, quantile, among others) and density estimation for a mix of continuous, discrete, and categorical data. Koenker (2008) introduces an extension of his **quantreg** package: The new function `crq()` implements quantile regression for censored data using various estimation techniques which are assessed with asymptotic and simulation comparisons. The package **sampleSelection** of Toomet and Henningsen (2008) provides a class of sample selection models, ranging from the classic Heckman model to other so-called “generalized tobit models”. The special volume is concluded by Zeileis, Kleiber, and Jackman (2008) who provide an overview of count data regression in R, dealing both with overdispersion and excess zeros. Starting out from Poisson and negative binomial regression using previously available functionality in R, new functions `hurdle()` and `zeroinfl()` in package **pscl** are introduced for fitting hurdle and zero-inflated versions of the classic count regression models. Of course, the packages/papers in this special volume are only a selection from the many current projects doing econometrics in R. For somewhat broader overviews (without making claims to be complete) of what is available presently we refer the reader to Farnsworth (2006) and Kleiber and Zeileis (2008) as well as the econometrics task view mentioned above.

4. Future

We hope that the current trend of increasing interest in R from econometricians continues; not least due to the exciting collection of new tools presented in this special volume which should help researchers, practitioners and teachers alike to use R and its contributed packages in their work. But further innovations are on their way: At the upcoming useR! 2008 conference (<http://www.R-project.org/useR-2008/>), held at Technische Universität Dortmund, Germany, there will be several sessions related to econometrics that feature packages with new instruments for the econometric toolbox. All of this should help to spawn many new tools for econometrics written in R. Information about these will hopefully be discussed on the mailing list of the related special interest group “R-SIG-Finance” (see <https://stat.ethz.ch/mailman/listinfo/R-SIG-Finance>), made available on CRAN, listed in the econometrics CRAN task view at <http://CRAN.R-project.org/view=Econometrics>, and published in future issues of the *Journal of Statistical Software*.

References

Cribari-Neto F, Zarkos SG (1999). “R: Yet Another Econometric Programming Environment.”

Journal of Applied Econometrics, **14**, 319–329.

- Croissant Y, Millo G (2008). “Panel Data Econometrics in R: The **plm** Package.” *Journal of Statistical Software*, **27**(2). URL <http://www.jstatsoft.org/v27/i02/>.
- Farnsworth GV (2006). *Econometrics in R*. Contributed documentation, URL <http://CRAN.R-project.org/doc/contrib/Farnsworth-EconometricsInR.pdf>.
- Hayfield T, Racine JS (2008). “Nonparametric Econometrics: The **np** Package.” *Journal of Statistical Software*, **27**(5). URL <http://www.jstatsoft.org/v27/i05/>.
- Henningsen A, Hamann JD (2007). “**systemfit**: A Package for Estimating Systems of Simultaneous Equations in R.” *Journal of Statistical Software*, **23**(4). URL <http://www.jstatsoft.org/v23/i04/>.
- Hyndman RJ, Khandakar Y (2008). “Automatic Time Series Forecasting: The **forecast** Package for R.” *Journal of Statistical Software*, **27**(3). URL <http://www.jstatsoft.org/v27/i03/>.
- Kleiber C, Zeileis A (2008). *Applied Econometrics with R*. Springer-Verlag, New York. ISBN 978-0-387-77316-2, URL <http://CRAN.R-project.org/package=AER>.
- Koenker R (2008). “Censored Quantile Regression Redux.” *Journal of Statistical Software*, **27**(6). URL <http://www.jstatsoft.org/v27/i06/>.
- Pfaff B (2008). “VAR, SVAR and SVEC Models: Implementation Within R Package **vars**.” *Journal of Statistical Software*, **27**(4). URL <http://www.jstatsoft.org/v27/i04/>.
- Racine J, Hyndman RJ (2002). “Using R to Teach Econometrics.” *Journal of Applied Econometrics*, **17**, 175–189.
- R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
- Stigler SM (1994). “Citation Patterns in the Journals of Statistics and Probability.” *Statistical Science*, **9**(1), 94–108.
- Toomet O, Henningsen A (2008). “Sample Selection Models in R: Package **sampleSelection**.” *Journal of Statistical Software*, **27**(7). URL <http://www.jstatsoft.org/v27/i07/>.
- Zeileis A (2004). “Econometric Computing with HC and HAC Covariance Matrix Estimators.” *Journal of Statistical Software*, **11**(10). URL <http://www.jstatsoft.org/v11/i10/>.
- Zeileis A (2006). “Object-oriented Computation of Sandwich Estimators.” *Journal of Statistical Software*, **16**(9). URL <http://www.jstatsoft.org/v16/i09/>.
- Zeileis A, Kleiber C, Jackman S (2008). “Regression Models for Count Data in R.” *Journal of Statistical Software*, **27**(8). URL <http://www.jstatsoft.org/v27/i08/>.
- Zeileis A, Leisch F, Hornik K, Kleiber C (2002). “**strchange**: An R Package for Testing for Structural Change in Linear Regression Models.” *Journal of Statistical Software*, **7**(2). URL <http://www.jstatsoft.org/v07/i02/>.

Affiliation:

Achim Zeileis
Department of Statistics and Mathematics
Wirtschaftsuniversität Wien
Augasse 2–6
A-1090 Wien, Austria
E-mail: Achim.Zeileis@R-project.org
URL: <http://statmath.wu-wien.ac.at/~zeileis/>

Roger Koenker
Department of Economics
University of Illinois
Champaign, IL 61820, United States of America
E-mail: rkoenker@uiuc.edu
URL: <http://www.econ.uiuc.edu/~roger/>