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.*

MONASH 5/93

MONASH UNIVERSITY

AUSTRALIA

PANEL DATA MODELLING: A SOFTWARE REVIEW

Pierre Blanchard and László Mátyás

Working Paper No. 5/93

June 1993

LIBRARY
WITHDRAWN
AUG 10 1993
GIANNINI FOUNDATION OF
AGRICULTURAL ECONOMICS

DEPARTMENT OF ECONOMETRICS

ISSN 1032-3813

ISBN 0 7326 0373 0

PANEL DATA MODELLING: A SOFTWARE REVIEW

Pierre Blanchard and László Mátyás

Working Paper No. 5/93

June 1993

DEPARTMENT OF ECONOMETRICS, FACULTY OF ECONOMICS COMMERCE & MANAGEMENT

MONASH UNIVERSITY, CLAYTON, VICTORIA 3168, AUSTRALIA.

PANEL DATA MODELLING: A SOFTWARE REVIEW

Pierre Blanchard* and László Mátyás**

* Universite de Paris Val de Marne, France
** Monash University, Australia

Abstract

The use of panel data is more and more popular in economic and econometric modelling. The purpose of this paper is to help potential user to find the right computing tools for their analysis.

Keywords

Panel data, Econometrics, Time series, Cross section, Software

Our ability to manipulate large data sets rapidly on personal computers and the increasing availability of panel data on individuals, firms and households has revolutionised economic modelling and empirical econometrics. (Panel data refers to cross-section data that has been pooled over time.)

Several attractive features of such data suggest themselves (see *Mátyás–Sevestre* [1992]):

1. Time dynamics can be explored without suffering aggregation bias common in usual time series studies.
2. The consistency of the parameter estimators can often be proved for short time series provided the cross section dimension is large enough.
3. The identification of economic models and discrimination among competing economic and statistical hypotheses is much more easier.
4. The problem of multicollinearity can practically be eliminated.

The increasing use of panel data sets is very closely linked with the increasing performance of personal computers (Apple MacIntosh and IBM–PC and compatibles) due to three main factors: faster microprocessors, more and cheaper memory available (RAM memory and hard disk) and new versions of operating systems together with continuous fall in price.

This has led econometric software editors to offer their software packages on different hardware with the same capabilities in terms of econometric methods.

Most econometric software packages are regularly reviewed in various publications,[1] and so, the purpose of this paper is not to do this work again but to present the specific functions of some econometric software in terms of panel data management and relevant econometric methods. This can help potential panel data users to make an appropriate choice in terms of computer software.

It must be noted that the volume of data, its particular organisation (by time period or by individual) and the need for specific econometric methods implies the use of software especially designed, in part at least, for the treatment of panel data sets. We analyze here five different packages which have been selected because they have at least one of the three following important characteristics:

— They have data management capabilities (like merging, matching various data sets) which are extremely useful for panel data;

[1] For instance, in *Applied Statistics, Economic Journal, Journal of Applied Econometrics, Journal of Economic Surveys*, etc.

— Some econometric methods which are relevant[2] for panel data (like fixed effects or error components models) are available with a minimum of programming and for large data sets;

— They offer a powerful programming language which allows one to apply all the estimation (and hypothesis testing) methods theoretically available.

The five selected[3] packages in this review are: GAUSS, LIMDEP, RATS, TSP and SAS.

It is our belief that the five most important features required for a software package to use panel data effectively are:

1) — The software must be available for different hardware (mainframe, minicomputers, microcomputers) with similar capabilities. It may happen that due to the availability of the data, the heterogeneity of the hardware, work organisation, etc. econometric analysis has to be performed on different computers.

2) — It must have powerful features of panel database management (taking into account the double individual/time dimension of the data) like merging matching and sorting,[4] etc. Furthermore, import/export of data in and from various file formats must be as simple as possible.

3) — Basic econometric methods for panel data: estimation of fixed effects and error components models, and tests for the presence of specific effects, must be available without programming.

4) — As the data volume in panel data analyses is usually large, it is essential to have good numerical accuracy and speed. And, above all, it should be able to manage large RAM capacity as well as large virtual memory (on hard disk). This last aspect is central but concerns exclusively some package versions running on IBM PC and compatibles, particularly those using MS–DOS 3.xx and above due to this RAM limitation.

5) — Panel data econometrics is a field which is developing quite rapidly. Consequently, the software must offer a powerful programming language which allows the user to program sophisticated econometric methods for panel data

---

[2] In a sense, as panel data estimation procedures quite often consist of applying OLS on transformed data, ostensibly all types of the software packages can be used. Nevertheless, we have considered that, at least, these transformations must be simple to carry out. Any econometric estimation and inference can be performed using a programming language. Once again, we would insist, that some basic estimation methods should be available with little programming.

[3] Two classes of softwares are excluded from this study in spite of their intrinsic qualities:
- General statistical packages: Minitab, BMDP, SPSS, SYSTAT, SST, STATA.
- Specialized econometric packages in other areas than panel data estimation (for instance TROLL, AREMOS, SORITEC).

[4] On this point, it can be noted that using database management software, like dBASE or PARADOX for instance, is not per se a solution because this implies learning one more new language and, above all, requires a lot of programming, these packages being ill adapted to the double individual/time dimension of the data.

such as incomplete panel data estimation, duration models, non linear models, simultaneous equations models, dynamic models, etc.

The econometric software packages reviewed in this paper can be divided into two groups. In the first are LIMDEP, RATS, SAS, and TSP, which offer a broad range of econometric methods as they can be applied to time–series, cross–sections or panel data, whereas in the second is GAUSS, which has a very specific objective of offering a matrix programming language. In this sense, these two categories are difficult to compare because the range of the econometric methods available and the nature of the user interface (menu–driven, command language, programming language) are meant to satisfy different users.[5]

Panel data software: an overview

GAUSS (Version 3.0 — 1992)

GAUSS is an econometrics and statistics oriented programming language designed to perform statistical and mathematical (especially matrix) calculus. It also includes an excellent graphical module.

GAUSS 386 is only available on IBM–PC microcomputers and compatibles (with a mathematical coprocessor).[6] The RAM limitation is the effective RAM memory even if it exceeds 640 Kb and the virtual memory (hard disk).

GAUSS can directly read/write only GAUSS and ASCII files. With the STAT TRANSFER external module, it is possible to convert a GAUSS file into LOTUS, DIF, and other formats as well.

With its set of powerful instructions, data transformations and estimation methods can be programmed in a very compact way. Using GAUSS for large econometric problems is effective as GAUSS is fast both in program's interpretation/execution and reading large GAUSS datasets. A good internal program editor is provided.

Moreover, GAUSS has some (for an extra charge[7]) applications modules and add–in programs. The applications modules are source programs written in GAUSS which perform some statistical or econometric tasks. As GAUSS is an interpreter, the user can run (and modify) these programs. There are, so far, nine application modules:

[5] This review was made using an 80386 (25 Mhz) IBM–PC compatible with a VGA colour screen, 4 Mb RAM memory, a 150 Mb hard disk and a 80387 coprocessor. For doing serious work on panel data with an IBM PC, we can consider this configuration (except for the colour screen) as minimal for all the softwares reviewed here.

[6] Requiring a micro–computer with a microprocessor (and a coprocessor) of the 80386 or 80486 type.

[7] $ 75 in average.

1) Descriptive Statistics
2) Quantal Response Models
3) Loglinear Analysis
4) Optimisation
5) Nonlinear Systems
6) Time Series and Cross Sections & Autoregressive Models
7) Simplex Method
8) Maximum Likelihood
9) Linear Regression.

The 6th module contains source programs for the estimation of fixed effects and error components models with balanced and unbalanced data. Modules 1, 2, 4, and 8 can be very useful because they contain source programs for calculating basic statistics, doing data transformations (including sort and merge) and nonlinear estimation.

Moreover, there are also some add–in programs like GAUSSX, MARKOV, DATAWIZ, etc. which are written in GAUSS language and require GAUSS to work, but which are simpler to use, *i.e.,* with a more easy command language or/and are menu–driven. For instance, DATAWIZ is a spreadsheet for GAUSS data sets.

More useful from our point of view, GAUSS users have written many application programs which can be obtained on request, and very often free of charge. Yet, in the absence of a Users' club, information on available programs is difficult to obtain. We can mention the DPD (Dynamic Panel Data) program, written by *Arellano and Bond* [1988], which computes estimators for dynamic panel data models. This program allows the application of various estimators and provides robust test statistics, Sargan test, tests for serial correlation, etc., even with unbalanced data. By using the DPD program and the 6th module (for balanced or unbalanced data sets) numerous estimation methods are available: OLS, GLS, IV estimators, Generalised Method of Moments estimators, etc.

With some work, the users not familiar with structured programming languages can program all the linear or nonlinear econometric methods for panel data or adapt the source program already written.

GAUSS has the pros and cons of a programming language. Requiring high human capital investment, it can be applied, for a very reasonable cost,[8] to almost any econometric problem.

[8] GAUSS 386 plus all application modules cost about $ 1300.

SAMPLE PROGRAM FOR PANEL DATA ANALYSIS USING GAUSS

The following GAUSS program does the same thing as the RATS, TSP and LIMDEP programs presented later on. It was shortened by suppressing some statitics.
Moreover, we assume that the GAUSS data file is balanced and organized by observation (not necessary).
As all GAUSS programs, it can (easily ?) be adapted for managing an unlimited number of time periods and/or individuals (with a DO-LOOP on data file reading) and for allowing estimation on unbalanced data.

```
/* Procedure for the estimation and editing of results */

PROC (3) = ESTIM(Met,Exo,Corddl,Nobs,X,Y) ;

LOCAL d, Mask, fmt, Edx, K, Df, Beta, Ssr, S, Vbeta, T, Vy, R2 ;

  K=ROWS(Exo) ; Df = Nobs-K-Corddl ; Beta = INVPD(X'X)*X'Y ;
  Ssr = (Y-X*Beta)'*(Y-X*Beta) ; S = SQRT(Ssr/(Df)) ;
  Vbeta = S*S*INV(X'X) ; Vy = (Nobs-1)*VCX(y) ;
  R2 = 1 - Ssr/Vy ; T = Beta./SQRT(diag(Vbeta)) ;

  LET Mask[1,3] = 0 1 1 ;
  LET fmt[3,3] = "-*.*s " 8 8  "*.*lf " 12 3  "*.*lf " 12 3 ;
  ? Met ; ? ; ? "VARIABLE        COEFF.      T-STUDENT" ;
  Edx = Exo~beta~T ; d = PRINTFM(Edx,Mask,Fmt) ; ? ; ? ;
  ? " R2    = " ;; ? R2   ;; ? " s     = " ;; ? s    ;
  ? " SSR   = " ;; ? ssr  ;; ? " Nobs  = " ;; ? Nobs ;
  ? " Df    = " ;; ? Df   ;; ? " Var Y = " ;; ? Vy   ;
  IF (Met .$/="FGLS") ; ? ; ? "More ?" ;; WAITC ; CLS ; ENDIF ;

RETP(Beta,Vbeta,Ssr) ;

ENDP ;

/* Initializations, Filename, exog. & endog. variable list */
closeall f1 ; n = 800 ; t = 10 ; nt = n*t ; FILE = "RANDATA" ;
LET Pexo  = TREND X1 X1L X2 X2C X3 ; LET Pendo = Y ;
Ptot  = Pexo|Pendo ; Pexo1 = Pexo|"CONSTANT" ;

/* Reading the Gauss data file and variable selection */

OPEN F1 = ^File FOR READ ;
Vnoms = GETNAME(File) ; Xdat = READR(F1,nt) ;
Lvexo  = INDCV(Pexo,Vnoms)  ; Lvendo = INDCV(Pendo,Vnoms) ;
Lptot  = INDCV(Ptot,Vnoms)  ;

/* Descriptive statistics and correlation matrix */

? "Descriptive Statistics" ; ? ; X=SUBMAT(Xdat,0,Lptot) ;
__altnam = vnoms[lptot] ;
{v,m,va,std,min,max,val,mis}=DSTAT(0,X) ;
? ; ? "More ?" ;; WAITC ; CLS ;
? "Correlation matrix" ; ? ; ? $Ptot' ; ? ;
Crx = CORRX(X) ; ? Crx ; ? ; ? "More ?" ;; WAITC ; CLS ;
CLEAR X, Crx ;

/* OLS */

X = SUBMAT(Xdat,0,Lvexo)~ONES(nt,1) ; Pexo1 = Pexo|"CONSTANT" ;
```

```
Y = SUBMAT(Xdat,0,Lvendo) ;
{Bmco,Vmco,Ssrmco} = ESTIM("OLS",Pexo1,0,nt,X,Y) ;

/* calculing Xi., Yi., Xit-Xi. (Xi) and Yit-Yi. (Yi) */

I = 1 ; X = 0 ; Y = 0 ; Xi = 0 ; Yi = 0 ; Xim = 0 ; Yim = 0 ;
e1 = SEQA(0,t,n) ; e2 = SEQA((n-1)*t,-t,n) ;
DO WHILE I <= n ;
 X = TRIMR(SUBMAT(Xdat,0,Lvexo),e1[i],e2[i]) ;
 Y = TRIMR(SUBMAT(Xdat,0,Lvendo),e1[i],e2[i]) ;
 Xim = MEANC(X) ; Yim = MEANC(Y) ;
 Y = Y-Yim ; X = X-Xim' ;
 IF i==1 ; Xm = Xim' ; Xi = X ; ELSE ; xm = Xm|Xim' ; Xi = Xi|X ; ENDIF ;
 IF i==1 ; Ym = Yim ; Yi = Y ; ELSE ; Ym = Ym|Yim ; Yi = Yi|Y ; ENDIF ;
 I = I + 1 ;
ENDO ;

/* Between estimation  */

 {Binter,Vinter,Ssrinter} = ESTIM("INTER",Pexo,0,n,Xm,Ym) ;

/* Within  estimation  */

{Bintra,Vintra,Ssrintra} = ESTIM("INTRA",Pexo,n,nt,Xi,Yi) ;

/* calculating Sig2eps and Sig2ind */

Sig2eps = SsrIntra/(nt-n-rows(pexo)) ;
Sig2ind = SsrInter/(n-rows(pexo1)) - (1/t)*Sig2eps ;
Theta = 1 - (SQRT(Sig2eps)/(SQRT(Sig2eps+t*Sig2ind))) ;

/* Calculating Xit-theta*Xi. (Xg) and Yit-theta*Yi. (Yg) */

I = 1 ; Xi = 0 ; Yi = 0 ; Xm = 0 ; Ym = 0 ;
X = 0 ; Y = 0 ; Xg = 0 ; Yg = 0 ;
e1 = SEQA(0,t,n) ; e2 = SEQA((n-1)*t,-t,n) ;
DO WHILE I <= n ;
 X = TRIMR(SUBMAT(Xdat,0,Lvexo),e1[i],e2[i]) ;
 Y = TRIMR(SUBMAT(Xdat,0,Lvendo),e1[i],e2[i]) ;
 Y = Y-Theta*MEANC(Y) ; X = X-Theta*MEANC(X)' ;
 IF i==1 ; Xg = X ; Yg = Y ; ELSE ; Xg = Xg|X ; Yg = Yg|Y ; ENDIF ;
 I = I + 1 ;
ENDO ;

/* Estimating FGLS */

IF Theta >= 0 ;
   Xg = Xg~(ones(nt,1)*(1-Theta)) ;
   {Bqgls,Vqgls,Ssrqgls} = ESTIM("FGLS",Pexo1,0,nt,Xg,Yg) ;
ELSE ; ? "FGLS impossible, Vu < 0" ; GOTO HALT ; ENDIF ;
? ; ? "Theta = " ;; ? Theta ; ? "Var e = " ;;
? Sig2eps ; ? "Var u = " ;; ? Sig2ind ; ? ;
Bqgls=SUBMAT(Bqgls,ROWS(Pexo),0) ;
Vqgls=SUBMAT(Vqgls,ROWS(Pexo),ROWS(Pexo)) ;
Hauss = (Bintra-Bqgls)'*(INV(Vintra-Vqgls))*(Bintra-Bqgls) ;
? "Hausman Statistics = " ;; ? Hauss ; Prob = CDFCHIC(Hauss,ROWS(Pexo)) ;
? "Probability        = " ;; ? Prob ;
HALT:
? ; ? ; ? "END OF PROGRAM" ; End;
```

LIMDEP (Version 5.1 — 1991)

LIMDEP[9] (LIMited DEPendent variable) was initially designed, as its name indicates, for estimating models with limited or qualitative dependent variables. With the last release, its scope is much greater, allowing one to estimate Box Jenkins ARIMA and ARMAX models, as well as simultaneous equations models.

LIMDEP is available on two types of hardwares: mainframes having a FORTRAN 77 compiler and microcomputers IBM–PC and compatibles. The latter version seems more complete in terms of estimation methods than the mainframe version and encompasses in fact two sub–versions: a standard version for MS–DOS 3.xx and above and a 386 version which is absolutely necessary when working with panel data. It requires a RAM memory of at least 4 Mb and a coprocessor.

As a command language, it works either in interactive or in batch mode. It can read/write various file formats: binary, DIF, WKS, ASCII. Its essential characteristic is the great number of non elementary econometric methods easily available (TOBIT, PROBIT, LOGIT, GMM, MLE etc.). The methods available for panel data are estimation methods of fixed effects and error components models with balanced or unbalanced data and with or without autocorrelation correction. For more complex problems, LIMDEP offers a very complete set of matrix instructions.

The new version 6.0 (not reviewed here) seems to be still more powerful and allows the estimation of random effects probit models, random effects stochastic frontier models and random coefficients models. The graphic module with a graph editor is now also quite good.

The instructions for data management are not as developed as the econometric instructions. There is no MERGE instructions and the SORT command allows for only one sort key. The programming language is unusual in the way it treats IF..THEN..ELSE loop instructions over procedures, over variables and observations. This can be a problem for complex applications.

In spite of the presence of an internal program's editor, we found the interface and the command language a little bit severe. Moreover, the LIMDEP syntax is very specific without any efficiency gains, particularly for data manipulation instructions (extensive use of $ and ;).

We must recognize that these criticisms are of minor importance compared to the variety of sophisticated econometric methods which are quite easily available with little programming and for a very low cost ($ 600 for the 386 versions).

As the enclosed program shows, little programming is needed indeed.

[9] Version 6.0 is now available.

LIMDEP needs a stratification variable for individuals and/or periods: this is done by the option STR=IDENT and PERIOD=TIME. The series IDENT and TIME can be in the data file or can be created by specialised instructions (if the data is balanced).

CREATE ; IDENT = TRN(10,0) $ generates a variable which contains (1,1, ... 1,2,2, ... 2, ..., n, n, ..., n), each number being repeated 10 times for the n individuals.

CREATE;TIME=TRN(-10,0)$ creates a serie like (1,2,3, ..., 10,1,2, ..., 10, ..., 1,2, ..., 10) each sequence of time variable repeated n times.

The output given by LIMDEP is very detailed, particularly for test statistics: F–tests, Hausman test and Breusch–Pagan test. However, in the presence of unbalanced panel data, groupwise heteroscedasticity is not taken into account for the fixed effects model but is correctly treated for the error components model. With a 4 Mb RAM 386 microcomputer, four limitations are important with LIMDEP: the data must have less than 400,000 data points, less than 2000 individuals; the model must have less than 100 exogeneous variables and less than 21000 observations (exogenous variables times the number of individuals).

SAMPLE PROGRAM FOR PANEL DATA ANALYSIS USING LIMDEP

```
? Reading an ASCII data file

READ ; NOBS = 8000 ; NVAR = 12 ;
FILE = RANDATA.ASC ;
NAMES = IDENT, TIME, CONSTANT, TREND, X1, X1L, X2, X2C, X3, X3L,
        ALEA, Y $

? descriptive statistics

DSTAT ; RHS = X1, X1L, X2, X2C, X3, X3L, Y ; OUTPUT = 3 $

? OLS, Fixed effects and GLS

CRMODEL ; LHS = Y ;
          RHS = X1, X1L, X2, X2C, X3, X3L ;
          STR=IDENT ; PANEL $


? Fixed effects, GLS estimation  with autocorrelated errors.
? The option PANEL can be repleced by FIXED or RANDOM option.

CRMODEL ; LHS = Y ;
          RHS = ONE, X1, X1L, X2, X2C, X3, X3L ;
          STR=IDENT ; PANEL ; AR1 $

STOP $
```

RATS (Version 4.01 — 1992)

RATS (Regression Analysis of Time Series) is mainly designed for time series analysis but it also provides instructions to estimate fixed effects or error components models.

RATS works in interactive mode (with the editor RATSEDIT) as well as in batch mode and is available in many versions: a mainframe version (VAX/VMS, IBM/CMS and IBM/MVS), a minicomputer version (APOLLO, SUN, DEC, IBM RS/6000...) and a microcomputer version. For this latter version, different releases exist for IBM–PC and compatibles, one for MS–DOS 3.xx and above (with the memory limit of 640 Kb) and two other RATS 386 and RATS OS/2, and a version for the Apple MacIntosh.

RATS can read and write various file formats (ASCII, LOTUS, DIF, TROLL, BINARY) and also provides databank management facilities. But, the same remarks can be made as for LIMDEP: no MERGE instruction is available and the SORT instruction admits only one sort key. Cross–tabulation instructions are unfortunately not available.

RATS has several interesting features:

— The variety of available econometric methods: time series analysis, logit and probit models, SUR models, and simultaneous equation models. It is important to note that maximum likelihood estimation method is now available with Version 4.01.

— Existence of a complete programming language with matrix instructions and macro–instructions (string substitution facility).

— Availability of a good graphic module (GRAFEDIT) with Postscript and HPGL device.

In the field of panel data estimation, RATS provides several interesting features. It deals explicitly with the double dimension of data (individual–time), each observation being referred by i:j where (for instance) i refers to the i-th individual and j to the j-th year. Data may be stored in a direct access database (with unlimited number of series of various periodicities). And with simple instructions (PANEL and PSTAT), one can transform easily the original data and estimate linear fixed or random effects models.

This software is cheap, easy to use and represents a good compromise between the variety of available econometric methods and their performances. The documentation, although an improvement on the previous versions, is not sufficiently explicit. The program is longer than in TSP or LIMDEP, but it is quite clear. All lines with an * at the beginning are comments. A $ is the continuation lines symbol.

Concerning the enclosed program, we can make the following remarks:

— The program is relatively long because with RATS we need to transform all the variables (*e.g.*, $Y_{it} - Y_{i.}$ and $X_{it} - X_{i.}$ for the Within estimator).

— It is possible to include individual and/or time effects in the PANEL and PSTATS instruction.

— In the presence of an unbalanced panel data, it is possible to adapt the WITHIN estimator to obtain unbiased estimators of the true standard errors. For example, if the model contains only individual effects, we can create a series containing for each individual $\sqrt{\frac{T_i}{T_i-1}}$, where T_i is the number of periods for which individual i is observed. Then we use the option SPREAD of LINREG instruction to perform weighted least squares. We assume that a variable IDENT exists in the data containing the RATS missing value code (NA$), if a variable of the model is missing for the individual i period t. This can be done for the Within estimator as:

```
SET IDENT 1:1 800:10 = IF$(IDENT(T).EQ.NA$,0,1)
PANEL CORHET 1:1 800:10 IDENT 1:1 ENTRY 0.0 INDIV 10.
SET CORHET 1:1 800:10 = SQRT(CORHET(T)/(CORHET(T)--1))

LINREG(DFC=800,SPREAD=CORHET) YFE 1:1 800:10
# X1FE X1LFE X2FE X2CFE X3FE X3LFE
```

However, for the (F)GLS estimator, the procedure PSTATS produces incorrect results in the case of unbalanced panel data.

— If we use the instruction AR1 instead of LINREG, it is possible to estimate fixed and error components models with first–order serially correlated errors. The coefficient ρ can be the same for all cross–section units (by default) or can be different in each if we use the option DIFFERING. RATS drops automatically missing values due to lags or leads in a LINREG or AR1 instruction, but not with the transformation instruction SET.

With 4 Mb RAM, the memory limitation is about 310,000 numbers, *e.g.*, 1000 individuals, 10 periods and 31 variables (series).

SAMPLE PROGRAM FOR PANEL DATA ANALYSIS USING RATS

```
* Panel data structure declaration and reading a RATS data file.
* The sample description 1:1 800:10 may be replaced by / (default)
* or by 1//1979:1 800:1988:1 for annual data on individuals
* if we modify the calendar and allocate instructions.

CALENDAR(PANELOBS=10) 1978:1

ALLOCATE 0 800//1987:1

OPEN DATA RANDATA.RAT
DATA(UNIT=DATA,ORG=OBS,FORMAT=RATS) 1//1978:1 800//1987:1  $
     IDENT TIME TREND X1 X1L X2 X2C X3 X3L Y
CLOSE DATA

* Descriptive statistics

CMOMENT(PRINT,CORR) 1//1978:1 800//1987:1
#  X1 X1L X2 X2C X3 X3L Y

TABLE

* Estimation by OLS. The residuals are saved in a variable
* (a series in RATS terminology) named RESIDS for latter use.

LINREG Y  1//1978:1 800//1987:1 RESIDS
#  CONSTANT X1 X1L X2 X2C X3 X3L

* Data transformation (Yit-Yi. and Xit-Xi.) for computing the WITHIN
* estimator.

PANEL Y   1//1978:1 800//1987:1 YFE   1//1978:1 ENTRY 1.0 INDIV -1.0
PANEL X1  1//1978:1 800//1987:1 X1FE  1//1978:1 ENTRY 1.0 INDIV -1.0
PANEL X1L 1//1978:1 800//1987:1 X1LFE 1//1978:1 ENTRY 1.0 INDIV -1.0
PANEL X2  1//1978:1 800//1987:1 X2FE  1//1978:1 ENTRY 1.0 INDIV -1.0
PANEL X2C 1//1978:1 800//1987:1 X2CFE 1//1978:1 ENTRY 1.0 INDIV -1.0
PANEL X3  1//1978:1 800//1987:1 X3FE  1//1978:1 ENTRY 1.0 INDIV -1.0
PANEL X3L 1//1978:1 800//1987:1 X3LFE 1//1978:1 ENTRY 1.0 INDIV -1.0

* Fixed effects estimation. Note that the correction for the degrees
* of freedom  since we compute and substract 800 individuals means.

LINREG(DFC=1000) YFE  1//1978:1 800//1987:1
#   X1FE X1LFE X2FE X2CFE X3FE X3LFE

* Data transformation (Yi. and Xi.) for computing the BETWEEN
* estimator.
* We can use the same variable transformed names (suffixed FE)
* and the option COMPRESS for saving memory space.

PANEL(COMPRESS) Y   1//1978:1 800//1987:1 YB   1//1978:1 INDIV 1.0
PANEL(COMPRESS) X1  1//1978:1 800//1987:1 X1B  1//1978:1 INDIV 1.0
PANEL(COMPRESS) X1L 1//1978:1 800//1987:1 X1LB 1//1978:1 INDIV 1.0
PANEL(COMPRESS) X2  1//1978:1 800//1987:1 X2B  1//1978:1 INDIV 1.0
PANEL(COMPRESS) X2C 1//1978:1 800//1987:1 X2CB 1//1978:1 INDIV 1.0
PANEL(COMPRESS) X3  1//1978:1 800//1987:1 X3B  1//1978:1 INDIV 1.0
PANEL(COMPRESS) X3L 1//1978:1 800//1987:1 X3LB 1//1978:1 INDIV 1.0

* Between estimator on 8000-7200 = 800 individuals means.
```

```
LINREG YB  1//1978:1 80//1987:1
#   CONSTANT X1B X1LB X2B X2CB X3B X3LB

* GLS estimation

* Variance analysis of Uit = mui + nit. Sig2n and Sig2mu are stored
* in VRANDOM and VINDIV.

PSTATS(TESTS,EFFECTS=INDIV) RESIDS 1//1978:1 800//1987:1

* Compute    theta = ....

COMPUTE THETA = 1.0 - SQRT(%VRANDOM/(%VRANDOM + 10*%VINDIV))

* Data transformation (Yit-theta*Yi. and Xit-theta*Xi.) for
* computing the GLS  estimator.

PANEL Y    1//1978:1 800//1987:1 YRD   1//1978:1 ENTRY 1.0 INDIV -THETA
PANEL X1   1//1978:1 800//1987:1 X1RD  1//1978:1 ENTRY 1.0 INDIV -THETA
PANEL X1L  1//1978:1 800//1987:1 X1LRD 1//1978:1 ENTRY 1.0 INDIV -THETA
PANEL X2   1//1978:1 800//1987:1 X2RD  1//1978:1 ENTRY 1.0 INDIV -THETA
PANEL X2C  1//1978:1 800//1987:1 X2CRD 1//1978:1 ENTRY 1.0 INDIV -THETA
PANEL X3   1//1978:1 800//1987:1 X3RD  1//1978:1 ENTRY 1.0 INDIV -THETA
PANEL X3L  1//1978:1 800//1987:1 X3LRD 1//1978:1 ENTRY 1.0 INDIV -THETA

SET CONSTRD 1//1978:1 800//1987:1 = 1.0 - THETA

LINREG YRD  1//1978:1 800//1987:1
#   CONSTRD X1RD X1LRD X2RD X2CRD X3RD X3LM

* Hausman test. Modifying IEVAL PVAR=6, this program, using
* RATS matrix calculus, works no matter what the number of
* independent variables is.
* But the intercepts must be in last position in the
* independent variables list.

DECLARE INTEGER PVAR
COMPUTE PVAR = 6
DECLARE SYMM VB1(PVAR,PVAR) XXDOUBLE(PVAR+1,PVAR+1) VB2(PVAR,PVAR)
DECLARE VECT BETA1(PVAR) BETADOUB(PVAR+1) BETA2(PVAR)

LINREG(NOPRINT) YRD  1//1978:1 800//1987:1
#   X1RD X1LRD X2RD X2CRD X3RD X3LRD CONSTRD

COMPUTE XXDOUBLE = %XX*%SEESQ
COMPUTE BETADOUB = %BETA
*
OVERLAY XXDOUBLE(1,1) WITH VB1(PVAR,PVAR)
OVERLAY BETADOUB(1) WITH BETA1(PVAR)

LINREG(DFC=1000,NOPRINT) YFE  1//1978:1 800//1987:1
#   X1FE X1LFE X2FE X2CFE X3FE X3LW

COMPUTE VB2 = %XX*%SEESQ
COMPUTE BETA2 = %BETA

COMPUTE HS = (TR((BETA2-BETA1))*(INV(VB2-VB1)))*(BETA2-BETA1)
COMPUTE HHS = HS(1,1)
CDF CHISQUARRED HHS (PVAR+PVAR-1)
```

SAS (Version 6 — 1990)

SAS is probably the best known of all statistical software packages. Earlier, available only for mainframes (IBM, DEC...), SAS is now available for mini-computers, UNIX workstation and personal computers IBM and compatibles (for MS-DOS version 3.xx and above, for Windows 3, for OS/2) as well.[10] Microcomputer and mainframe versions have (or will have very soon) the same features except for the user's interface. SAS Micro provides a very good interface with multiple windows, interactive and batch mode.

SAS covers a wide range of statistical and econometric methods. It works in interactive and batch mode and consists of basic and many application modules which can be bought separately. In the PC version, various file formats are supported (ASCII, DBASE, LOTUS, etc.).

A SAS program consists of a series of DATA steps which create a SAS table followed by PROCEDURE steps to carry out statistical, graphical and other types of analysis. It is not possible to do loops in the data step and in the procedures step, which means that SAS programs are often quite long unless a powerful but complex macro-language (SAS-MACRO) is used. In addition, users can use a sophisticated programming language (SAS-IML), can interface programs written in C, and many others facilities are available.

The econometric part (SAS/ETS) is often criticized for its outdated features which do not support econometric testing. Moreover, it offers only one specific procedure for panel data (TSCSREG) in the supplementary library (in version 5 but not in version 6). There is no reason to believe that this situation will get better in the future as the development effort is concentrated on other parts of this software. Estimating panel data models either requires a lot of work or the use of previously written macro instructions.

As noted by *Kőrösi et al.* [1992], "The basic idea of SAS is that those who bought all modules and learnt to use them properly... should not encounter problems which they are unable to solve". Nevertheless, the great power of these data management instructions justifies its use by itself. With SAS, the most difficult operations on panel data sets like merging, complex sort, matching, etc., become virtually instantaneous with very little programming. For users of large databases SAS is a very adequate solution,[11] available for a substantial rental fee and requiring a large disk space, but it is not a good choice for the estimation and testing of panel data models.

One can compare the length and the complexity of the following SAS program calculating OLS, WITHIN and (F)GLS estimators with those of the other packages. The program for the Hausman test is not reported here due to its length.

---

[10] It is unlikely that a version for MacIntosh will be created.

[11] Note that other packages, like SYSTAT for example, are good and cheaper (but less complete) alternatives to SAS.

SAMPLE PROGRAM FOR PANEL DATA ANALYSIS USING SAS

```
/* READING DATA FILE */

LIBNAME IN '\SAS' ;
DATA TAB0 ;
SET IN.RANDATA ;

/* DESCRIPTIVE STATISTICS */

PROC MEANS DATA = TAB0 ;
     VAR Y X1 X1L X2 X2C X3 X3L ;

/* ESTIMATION BY OLS */

PROC REG DATA = TAB0 ;
     MODEL Y = X1 X1L X2 X2C X3 X3L ;

/* ESTIMATION OF SIG2e ans SIG2u */

PROC MEANS DATA = TAB0 MEAN NOPRINT ; BY IDENT ;
     VAR Y X1 X1L X2 X2C X3 X3L ;
     OUTPUT OUT = MOYTAB0 MEAN = MY MX1 MX1L MX2 MX2C MX3 MX3L ;

PROC REG DATA = MOYTAB0 OUTEST = VARB ;
     MODEL MY = MX1 MX1L MX2 MX2C MX3 MX3L ;

DATA VARB ; SET VARB ; SIGMA2U = (_RMSE_)**2 ; DUMVAR = 1 ;

PROC PRINT ; VAR SIGMA2U _RMSE_ ;

DATA INTRA ; MERGE TAB0 MOYTAB0 ; BY IDENT ;

PROC SORT DATA = INTRA ; BY IDENT TIME ;

DATA INTRA (DROP = _TYPE_ ) ; SET INTRA ;
DUMVAR = 1 ;
WY = Y - MY ;
WX1 = X1 - MX1 ;
WX1L = X1L - MX1L ;
WX2 = X2 - MX2 ;
WX2C = X2C - MX2C ;
WX3 = X3 - MX3 ;
WX3L = X3L - MX3L ;

PROC REG DATA = INTRA OUTEST = VARW ;
     MODEL WY = WX1 WX1L WX2 WX2C WX3 WX3L /NOINT ;

DATA VARW (DROP = _TYPE_) ; SET VARW ;
SIGMA2W = (_RMSE_)**2 ; DUMVAR = 1 ;
PROC PRINT ; VAR SIGMA2W _RMSE_ ;

DATA VARIANCE ; MERGE VARB VARW ; BY DUMVAR ;
DUMVAR = 1 ; THETA = SIGMA2W/(10*SIGMA2U) ;
LAMBDA = 1 - SQRT(THETA) ;

DATA FGLS ; MERGE  INTRA VARIANCE ; BY DUMVAR ;
GY     = Y -  LAMBDA*MY ;
```

```
GX1    = X1 - LAMBDA*MX1 ;
GX1L   = X1L - LAMBDA*MX1L ;
GX2    = X2 - LAMBDA*MX2 ;
GX2C   = X2C - LAMBDA*MX2C ;
GX3    = X3 - LAMBDA*MX3 ;
GX3L   = X3L - LAMBDA*MX3L ;
GCONST = 1.0 - LAMBDA ;

PROC REG DATA = FGLS ;
     MODEL GY = GX1 GX1L GX2 GX2C GX3 GX3L GCONST / NOINT ;
```

TSP (Version 4.2A — 1991)

The latest TSP release has many new features which are of special interest and which represent substantial improvements on the previous versions. A PANEL instruction is now available for the estimation of linear models with balanced or unbalanced panel data, including the Between, Within and (F)GLS estimators. In addition to time series and nonlinear estimation, GMM estimators can be computed as well with the GMM instruction. The sequence of instructions for applying the Chamberlain's PI matrix method is also provided in the user's guide.

Many improvements have been for this new version: the programming language has been enhanced and now offers many matrix operations with clear and standard syntax. Unfortunately, a matrix cannot have double dimension subscript (i,j) and must be treated as a vector.

Like other packages, when working on IBM–PC and compatibles, the TSP 386 version must be preferred to the others due to the usual memory limitations. With this version, if your 386 computer has 3 Mb of extended memory in addition to the 640 Kb of base memory, it can estimate models on panel data sets containing at most 130000 figures *e.g.,* 15 variables, 860 individuals and 10 periods (460000 with 7 Mb with a maximum of 15 Mb and a minimum of 2 Mb) and 25000 data points for a series[12] (2500 individuals over 10 years for instance). Like other packages, the estimation of models with large panel data sets requires substantial RAM memory: a model with 2500 firms on 10 years observed for 20 (endogenous and exogenous) variables needs about 8 Mb. A virtual memory version is available which can overcome the 130000 figure limitation. TSP can run on various type of hardware (IBM VM/CMS, VAX/VMS, UNIX machines, APPLE McIntosh).

It must be noted that the documentation is excellent (particularly the TSP User's Guide), even if no example files are provided.

Taking into account the limits, the graphic module (on micro–computers) is simple but correct (few customizing options). For the 386 version, the graph (scatter plot, x–y graph or histogram) cannot be printed which is not the case for the standard PC version. This is a real problem because graphs can neither be saved nor stored in a graphic file for further printing.

TSP can run in interactive or batch mode but an integrated text editor is not provided. In addition to databank management instructions, TSP can read/write different file formats (formatted or unformatted ASCII, LOTUS, BINARY). Cross tabulations instructions are unfortunately not provided. As with many packages, because TSP was initially oriented towards time series analysis, panel data is treated as undated data which implies that transformations, sorting instructions and databank capabilities have limited use and

[12] In addition, it is possible to modify some of the non virtual memory limitations (*e.g.,* 25000 observations for a series) on request.

require a lot of manipulation. For instance, using the instructions AR1(TSCS) with numerous cross section units requires writing a tedious SMPL instruction. For an unbalanced panel dataset, correction for heteroskedasticity in the fixed effects model is not provided.

In general, a TSP program is quite concise and clear. The option T=... should be used to define the structure of the panel data set. However, TSP cannot estimate an error components model with both individual and time effects.

SAMPLE PROGRAM FOR PANEL DATA ANALYSIS USING TSP

```
? Reading an ASCII data file

OPTIONS CRT ;

FREQ N ; SMPL 1 8000 ;

READ(FILE='RANDATA.ASC',FORMAT=FREE)
     IDENT, TIME, CONST, TREND, X1, X1L, X2, X2C, X3, X3L, ALEA, Y ;

? Descriptive statistics

MSD(CORR)
    X1, X1L, X2, X2C, X3, X3L, Y ;

? OLS and Between estimation

PANEL(T=10,TOTAL,BETWEEN,NOVAR,NOWITH) Y
   C, X1, X1L, X2, X2C, X3, X3L ;

? Fixed effects and random components estimation

PANEL(T=10,NOTOT,VARCOM,WITHIN,NOBET) Y
    C, X1, X1L, X2, X2C, X3, X3L ;

? Hausman test

PANEL(T=10,NOTOT,SILENT,VARCOM,WITHIN,NOBET) Y
   X1, X1L, X2, X2C, X3, X3L, C ;

MFORM(NROW=6,NCOL=6,TYPE=SYM) VCOVV=@VCOVV ;
MFORM(NROW=6) COEFV=@COEFV ;
MAT DVAR = @VCOVW - VCOVV ;
MAT K = RANK(DVAR) ;
MAT Htest = (@COEFW-COEFV)'YINV(DVAR)*(@COEFW-COEFV) ;
CDF(CHISQ,DF=K) Htest ;

END ;
```

Evaluation by Criteria

For a general overview of the analysed software packages see, for example *Körösi–Mátyás–Székely* [1993].

In order to present the specific characteristics of the packages, we use the following notation:

code	signification
n.a.	non communicated
–	not concerned
Y	present
N	absent
R	recommended
1,2,3 or4	quality apppreciation: insufficient, acceptable, good, very good

Table 1: Data Manipulation, Descriptive Statistics Programming Language and Graphics

	GAUSS	LIMDEP	RATS	SAS	TSP
Data transformation					
Changing periodicity	–	N	Y	Y	Y
Seasonal adjustment	–	N	Y	Y	Y
Numerical integration & differentiation	Y	N	N	N	Y[13]
Use of wildcards names	N	Y	N	N	Y[14]
Use of predefined list of variables	Y	Y	Y	Y	Y
Calc. of Capital stock	N	N	N	N	Y
Price index	N	N	N	N	Y
Data file management					
Controled merge	Y[15]	N	N	Y	N
Sort (Nb of Key)	Y	Y(1)	Y(1)	Y[16]	Y(1)
String management	Y	N	N	Y	N
Macro–substitution	N	Y	Y	Y	N[17]
Programming language					
Procedure	Y	Y	Y	Y	Y
Global/local variables	Y	N	Y	Y	Y
Library facility	Y	Y	Y	Y	N
Loop over variable	Y	Y	Y	Y	Y
Loop over observation	Y	N	Y	Y	Y
Loop over file	Y	N	Y	N[18]	N
Loop over procedure	Y	Y	Y	N[19]	Y
Others language interface	Y	N	N	Y	N
Kronecker product	Y	Y	Y	Y	Y

[13] Differentiation only.
[14] With DO–loops
[15] With DATA Transformation module.
[16] No limit.
[17] Only with variable names in DO–loops.
[18] With macro–instructions and IML modules.
[19] With macro–instructions and IML modules.

Data manipulation, etc. (continued)

	GAUSS	LIMDEP	RATS	SAS	TSP
Presentation results facility	Y	N	Y	Y	N
Quality appreciation	4	2	3	3	2[20]
Graphs (in graph mode only 386 PC version)					
Multiple series	Y	Y	Y	Y	Y
Mixed scale	Y	N	Y	Y	N
Scatter plot	Y	Y	Y	Y	Y
Histogram	Y	Y	N	Y	Y
Piecharts	Y	N	N	Y	N
3D	Y	N	N	Y	N
Hardcopy	N	Y	Y	N	N
Graph file saving & printing	Y	N	Y	Y	N
Quality appreciation	4	2	3	4	1

[20] A two dimensional matrix cannot have a variable double subscript.

Table 2: Estimation Methods

	GAUSS[21]	LIMDEP	RATS	SAS	TSP
General statistics					
Descriptive statistics	Y	Y	Y	Y	Y
Cross tabulation	–	Y	N	Y	N
Principal components	–	N	N	Y	Y
Non-elementary non-panel data methods (without programming)					
Distributed lags	–	N	N	Y	Y
Inst.var., 2SLS & 3SLS	–	Y	Y	Y	Y
ARIMA	–	Y	Y	Y	Y
FIML, LIML	–	N	N	Y	Y
Kalman Filter	–	N	Y	N	Y
Vector autoregression	–	N	Y	N	Y
SURE	–	Y	Y	Y	Y
Nonlinear estimation	–	Y	Y	Y	Y
TOBIT	–	Y	N	Y	Y
LOGIT/PROBIT	–	U/M	U	U	M
GMM	–	Y	Y	N	Y
ARCH	–	N	N	N	Y
Panel data methods without programming					
Fixed effects models					
– with balanced data	Y	Y	Y	N	Y
– with unbalanced data	Y	Y	N	N	Y

[21] All these methods are available with programming or with buying supplementary modules. We get more details for panel data features with TSCS supplementary module and *Arellano Bond's* [1988] program (DPD).

Estimation methods (continued)

	GAUSS	LIMDEP	RATS	SAS	TSP
Error components models					
– with balanced data	Y	Y	Y	N	Y
– with unbalanced data	Y	N	N	Y	Y
Kmenta's models	Y	Y	N	N	Y
Hypothese tests directly available					
Breush Pagan test	N	Y	N	N	N
Hausman test	N	Y	N	N	N

Conclusion

We can conclude that none of the reviewed software packages completely satisfies the five criteria specified in the introduction. Nevertheless, criteria 1, 4, 5 are almost met by several reviewed packages. From this point of view, with the exception of GAUSS and SAS, we can note the great convergence in performance. One problem, however, remains: most of the packages were originally written for time series or cross section analysis, never explicitly for panel data. Thus, surprisingly, it is easier with many of them to apply sophisticated estimation methods than to make elementary operations such as sorting, merging or matching two or more panel data sets.

Ideally, the perfect software package would have the database management facilities of SAS, the availability of general econometric methods of RATS or TSP, the variety of panel pre–program econometric instructions of LIMDEP, a complete and fast programming language like GAUSS, and finally, a user's interface like SAS micro.

As no software is conceived explicitly for panel data even if the need for such a product exists, the choice is a second–best: If we want to one use without too much programming of sophisticated econometrics methods, LIMDEP is the best choice or GAUSS (with more programming but increased flexibility). SAS can be a good choice if one is mainly concerned with database management.

Package's Editor References

GAUSS Version 3.0	APTECH SYSTEMS, INC 26250 196TH PLACE SE, KENT, WA 98042 USA PHONE (206)631--6679, FAX (206)630--1220
LIMDEP Version 6.0	ECONOMETRIC SOFTWARE, INC. 43 MAPLE AVE., BELLPORT, NEW--YORK 11713 USA PHONE (516)286--7049, FAX (516)938--2395
RATS Version 4.0	ESTIMA P.O. BOX 1818, EVANSTON, IL 60204--1818 USA PHONE (708)864--8772 FAX (708)864--621
SAS Version 6	SAS INSTITUTE INC. SAS CIRCLE, BOX 8000, CARY, NC 27512--8000 USA PHONE (919) 467--8000,
TSP Version 4.2 A	TSP INT. P.O. BOX 61015 STATION A, PALO ALTO CA 94306 USA PHONE (415) 326--1927, FAX (415) 328--4163

References

Arellano M. and S. Bond [1988]: "Dynamic Panel Data Estimation Using DPD — A Guide for Users", *Technical document*, p. 11

Kőrösi G., Mátyás L., and Székely I. [1992]: *Practical Econometrics*, Avebury – Gower.

Kőrösi G., Mátyás L., and Székely I. [1993]: Comparative Review of Some Econometric Software Packages, *Journal of Economic Surveys*, (forthcoming) .

Mátyás L. and Sevestre, P. [1992]: *The Econometrics of Panel Data*, Kluwer Academic Publishers.