EBE_NPDE userguide

2013-04-09

The ebe_npde script is work in progress.

Overview

D

Example

ebe_npde run1.mod

Input and options

<u>Required input</u>

A model file is required on the command-line.

<u>Optional input</u>

-lst_file=filename	Default not set. By default PsN will, before copying the input model to the simulation models, look for an output file with final estimates for the input model and if found update the initial estimates in the simulation models. If option -lsf_file is set Psn will instead use the final estimates in the given file. If option is not set and no output file is found PsN will not update the estimates but instead rerun the input model, if option -estimate_input is set.
-estimate_input	Default set. By default, PsN will rerun the input model to get parameter estimates unless an output file is found or option -lst_file is set. But if option -estimate_input is unset with -no-estimate_input and no lst-file is found then the parameter estimates set in the input model will be used for simulations.
-samples=N	Default 300. The number of simulations and reestimations to run.
-reminimize	Default not set. By default, simulated datasets will be run with MAXEVAL=0 (or equivalent for non-classical estimation methods). If option -reminimize is set then $EST will be the same as in the input model.
-gls_data_file=filename	Default gls_data.dta. A file with input data for the gls program is always generated. This option changes the name of that file. After this file is created, the gls program can be run with option -gls_model and -ind_shrinkage with minimum runtime, see gls_userguide.pdf

<u>Some common PsN-options useful with ebe_npde</u>

For a complete list of common options see common_options_defaults_versions.pdf, or psn_options -h on the commandline.

-directory=ebe_npde_dirN   The run directory can be named.

-help   With -help ebe_npde will print a longer help message.

Output

The results are in eta_npde.csv and iwres_npde.csv. Also there are files decorrelated_original_eta.csv and decorrelated_original_iwres.csv with decorrelated original values. Original values are in raw_original_eta.csv and raw_original_iwres.csv. npd of raw values (normalized pd, not decorrelated) are in eta_npd.csv and iwres_npd.csv.

Algorithm overview

1. Remove MSFO option from $EST.

2. Add code in $PK or $PRED for printing IDs and individual ETAs to file.

3. If $SIM not present, create simple $SIM as in sse. Create 'samples' copies of modified input model with order number of copy indicated in filename. In each copy set unique seed in $SIM and set NSUB=1, set unique filename for ETA output in $PK/$PRED, remove $TABLE, if option reminimize is not set then set MAXEVAL=0, and add $TABLE IWRES ID NOPRINT ONEHEADER NOAPPEND FILE=iwres-<ordernum>.dta.

4. In 'original' input model: Remove $SIM if present. Add $TABLE <all undropped items in $INPUT> IPRED PRED IWRES NOPRINT ONEHEADER NOAPPEND FILE=orig_pred.dta. Add $TABLE IWRES ID MDV NOPRINT ONEHEADER NOAPPEND FILE=original_iwres.dta.

5. run modified original input model and 'samples' sim models.

6. Read all iwres_<order number>.dta files, storing IWRES values per data point. Compute, per data point, ISHR_ij=1-sd(IWRES_ij). Open orig_pred.dta, append ISHR column with computed values, and print to gls_input.dta. Print also shrinkage column to new file ind_iwres_shrinkage.dta. File gls_input.dta can be used as input when running gls program, see gls_userguide.pdf.

7. Read all eta files. For each individual do the following: Collect the simulated ETA values in a matrix A, where rows go from simulation 1 to simulation 'samples' and columns go from 1 to 'number of ETA'. Compute the empirical variance-covariance matrix of A.

(The matlab code for this would be cov(A), while PsN computes $R^T R * diag(1/(samples-1))$ where R is the upper triangular matrix from the QR-decomposition of a matrix B which is matrix A where the column mean has been subtracted from each column, and $diag(1/(samples-1))$ is a diagonal matrix where all elements are equal to 1/(samples-1)).

The original and simulated ETA vectors are all decorrelated by subtracting the mean and multiplying them with the inverse of the Cholesky factor of the variance-covariance matrix. (In PsN the system $R*diag(1/sqrt(samples-1))*vec\_decorr=vec$ is solved by back substitution followed by multiplication with sqrt(samples-1))

Compute per individual the order numbers of the decorrelated original ETA values in a sorted array of all the decorrelated simulated values. In file eta_rank.csv, print values ID,rank where rank =ordernum/ (samples+1). In file decorrelated_eta.csv print the decorrelated original ETA. These two files will have

one line per individual.

8. (Not necessary, division preserves rank) For each data point do the following: Compute the variance of the simulated IWRES_ij values (normalize with samples-1). Divide the original IWRES value minus the mean and all simulated values minus the mean with the variance. Compute the order number of the transformed original IWRES value in a sorted array of all the transformed simulated values. In file iwres_rank.csv, print values ID,MDV,rank for each data point where rank =ordernum/(samples+1).