

GLS user guide

PsN 4.6.0

Revised 2015-03-30

1 Overview

During the modelling process one may encounter scenarios where first-order conditional estimation with $\epsilon - \eta$ interaction (FOCEI) is indicated but may be prohibitively time consuming or unstable. Also, it may be a concern that a misspecification of the residual error model may translate to bias in structural or inter-individual variability parameters. These two problems may be overcome by using a generalised least squares (GLS) type approach where dependent variable predictions are obtained from a previous model fit and then used in the residual error model. The `gls` program automates this procedure, which is further described in [1] Using the `gls` program requires that the model is encoded in "the Uppsala way", see below.

In summary the GLS method amounts to the following: Run the original model. In a second step IPRED is replaced with GLSP in the definition of W where GLSP is computed as

$$\text{PRED}_{orig} \cdot \text{iwres_shrinkage} + (1 - \text{iwres_shrinkage}_{orig}) * \text{IPRED}_{orig},$$

where PRED_{orig} and IPRED_{orig} are obtained from running the original model and `iwres_shrinkage` is either a population shrinkage obtained from the original model run or a per-observation shrinkage based on simulations. The modified `gls` model is then estimated. Example

```
gls run1.mod
```

Uppsala way

The "Uppsala way" of encoding residual error is to express the standard deviation of the residual error as a parameter W and setting SIGMA 1 FIX.

Examples:

Additive error

```
W      = THETA(5)
Y      = IPRED+ERR(1)*W
IRES   = DV-IPRED
IWRES  = IRES/W
```

Proportional error

```
W      = THETA(5)*IPRED
Y      = IPRED+ERR(1)*W
IRES   = DV-IPRED
IWRES  = IRES/W
```

Additive plus proportional error

```
W      = SQRT(THETA(5)**2*IPRED**2+THETA(6)**2)
Y      = IPRED+ERR(1)*W
IRES   = DV-IPRED
IWRES  = IRES/W
```

2 Input and options

2.1 Required input

A model file is required on the command-line.

2.2 Optional input

-ind_shrinkage

Default not set. Compute per-observation iwres-shrinkage based on simulations.

-samples = N

Default not used. Only relevant if -ind_shrinkage is set and

-gls_model is not set. Creates N copies of input model with different seeds in \$SIM. Run to get N IWRES values for each data point y_{ij}. Compute

-reminimize

Default not set. Only relevant if -ind_shrinkage is set and -gls_model is 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 the same \$EST as in the input model will be used.

-additive_theta = XX

Default not used. In gls model, add a small and fix additive error in W. The error is added by changing $W = \text{SQRT}(\langle \text{expression} \rangle)$ to $W = \text{SQRT}(\text{THETA}(T)**2 + \langle \text{expression} \rangle)$ in the gls model, where T is the order number of new \$THETA XX FIX added to the model.

-iwres_shrinkage = X

Default not used. Forbidden in combination with -ind_shrinkage. If the population iwres shrinkage from the input model run is already available, or if a special values such as 0 or 1 is desired, the user can give the value as input on the command-line. Important note: PsN reports shrinkage in percent in the raw_results file, so if using the value from raw_results as input that value must be divided by 100.

-gls_model

Default not set. Only possible together with option -iwres_shrinkage or -ind-shrinkage. This option is to be used when a datafile with all data needed for the gls model run is already available, i.e. all input for the original model plus columns with PRED and IPRED from the original model run, and if -ind_shrinkage is set also a ISHR column with per observation shrinkage values. The option indicates that \$DATA specifies the file with the gls input data, and that \$INPUT lists the parameters in the datafile. In \$INPUT the columns PPRE and PIPR must be present as headers for PRED and IPRED values, plus ISHR for the shrinkage column if -ind_shrinkage is set.

-sim_table

Default not set. Only relevant if `-ind_shrinkage` is set and `-gls_model` is not set. PsN will delete all existing `$TABLE` in the simulation models before adding a `$TABLE` for per-observation IWRES values, but if option `-sim_table` is set then an extra `$TABLE` with diagnostic output is added to each simulation model.

-set_simest

Default not set. Only relevant if `-gls_model` is not set. This option can be used to set different `$EST` for original, simulation (if used) and final gls models, and to set a custom `$SIM` for the simulation model. When this option is set, PsN will look for lines starting with certain tags in the input model. All lines starting with the tag

```
;gls-final
```

will be collected, the tag will be removed, and then PsN will check that the lines define a single `$EST` record. This `$EST` record will be set in the final model, instead of the one in the input model. If no `;gls-final` tag is found the `$EST` record in the final model will be the same as in the input model. All lines starting with the tag

```
;gls-sim
```

will also be collected, the tag will be removed, and then PsN will check that the lines define either a single `$EST` or a single `$SIM`, or one of each. If a `$SIM` record is defined this will be used in the simulation model instead of any `$SIM` found in the input model. If a `$EST` record is defined this will be used in the simulation model instead of the `$EST` found in the input model.

2.3 Some important common PsN options

For a complete list see `common_options.pdf`, or `psn_options -h` on the commandline.

-h or -?

Print the list of available options and exit.

-help

With `-help` all programs will print a longer help message. If an option name is given as argument, help will be printed for this option. If no option is specified, help text for all options will be printed.

-directory = *'string'*

Default `gls_dirN`, where N will start at 1 and be increased by one each time you run the script. The `directory` option sets the directory in which PsN will run NONMEM and where PsN-generated output files will be stored. You do not have to create the directory, it will be done for you. If you set `-directory` to a the name of a directory that already exists, PsN will run in the existing directory.

-seed = *'string'*

You can set your own random seed to make PsN runs reproducible. The random seed is a string, so both `-seed=12345` and `-seed=JustinBieber` are valid. It is important to know that because of the way the Perl pseudo-random number generator works, for two similar string seeds the random sequences may be identical. This is the case e.g. with the two different seeds 123 and 122. Setting the same seed guarantees the same sequence, but setting two slightly different seeds does not guarantee two different random sequences, that must be verified.

-clean = *'integer'*

Default 1. The `clean` option can take four different values:

0 Nothing is removed

1 NONMEM binary and intermediate files except INTER are removed, and files specified with option `-extra_files`.

2 model and output files generated by PsN restarts are removed, and data files in the `NM_run` directory, and (if option `-nmqual` is used) the xml-formatted NONMEM output.

- 3** All `NM_run` directories are completely removed. If the PsN tool has created `modelfit_dir:s` inside the main run directory, these will also be removed.

-nm_version = *'string'*

Default is 'default'. If you have more than one NONMEM version installed you can use option `-nm_version` to choose which one to use, as long as it is defined in the `[nm_versions]` section in `psn.conf`, see `psn_configuration.pdf` for details. You can check which versions are defined, without opening `psn.conf`, using the command

```
psn -nm_versions
```

-threads = *'integer'*

Default 5 (if default PsN config file is used). Use the `threads` option to enable parallel execution of multiple models. This option decides how many models PsN will run at the same time, and it is completely independent of whether the individual models are run with serial NONMEM or parallel NONMEM. If you want to run a single model in parallel you must use options `-parafile` and `-nodes`. On a desktop computer it is recommended to not set `-threads` higher the number of CPUs in the system plus one. You can specify more threads, but it will probably not increase the performance. If you are running on a computer cluster, you should consult your system administrator to find out how many threads you can specify.

-version

Prints the PsN version number of the tool, and then exit.

3 Output

The results are in the `raw_results` file. If `-ind_shrinkage` is set the individual shrinkage values are in `ind_iwres_shrinkage.dta`. All `lst`-files, model files and table files are in the `m1` subdirectory.

4 Procedure overview

4.1 Original input model (only done if `-gls_model` is not set)

1. If option `-additive_theta=XX` is set, add `$THETA XX FIX` as last `$THETA`. Store order number `T` of new theta.
2. Remove `MSFO` option from `$EST`, if present.
3. If the model has both `$PRIOR` and `$SIM` then set option `TRUE=PRIOR` in `$SIM`.
4. Remove `$COVARIANCE`, if present.
5. If a `lst`-file is found for the input model, update initial estimates in input model based on `lst`-file.
6. Copy undropped `$INPUT` variables to new `$TABLE`. In `$TABLE` add `IPRED PRED`. Add `NOPRINT ONEHEADER NOAPPEND FILE=glsinput.dta`
7. Run modified original input model. Let `PsN` compute `iwres-shrinkage` for this model regardless if option `-ind_shrinkage` or `-iwres_shrinkage` is set.

4.2 Simulation models (only done if option `-ind_shrinkage` is set and `-gls_model` is not set)

1. Create 'samples' copies of modified original input model after modifications step 1-4 above.
2. If `$SIM` not present, create simple `$SIM (1234 NEW)`. Seed and `NSUB` will be set below.
3. If option `-set_simest` is set and a simulation record behind the tag `;gls-sim` is found, use this new simulation record instead.
4. In each copy set unique seed in `$SIM` and set `NSUB=1`.
5. Set `IGNORE=@` since datafile will get a header during copying. Keep any `IGNORE=(...)`.

6. Unless option `reminimize` is set, set `MAXEVAL=0` (or corresponding for non-classical estimation methods).
7. If option `-set_simest` is set and an estimation record behind the tag `;gls-sim` is found, use this new estimation record instead. Do not change `MAXEVAL` in this new estimation record.
8. If `ONLYSIM` is found in `$SIM` then remove the `$EST` record.
9. Update initial estimates with output from running modified original input model (step 7 above).
10. Remove existing `$TABLE`.
11. In each sim model, set `$TABLE IWRES ID NOPRINT ONEHEADER NOAPPEND FILE=iwres_⟨order number⟩.dta`.
12. If option `-sim_table` is set: In each sim model set an extra `$TABLE` with `ID TIME IPRED W IWRES NOPRINT ONEHEADER FILE=sdtab-sim⟨order number⟩.dta` to be used e.g. for diagnostics.
13. Run 'samples' sim models. Let PsN compute `iwres-shrinkage` for all of them.
14. Read all `iwres_⟨order number⟩.dta` files, storing all `IWRES` values per data point. Compute, per data point, `ISHR_ij=1-stdev(IWRES_ij)`. Open `glsinput.dta` and append `ISHR` column with computed values. Print also shrinkage column to new file `ind_iwres_shrinkage.dta`

4.3 GLS model if option `-gls_model` is not set

1. Copy modified original input model after modifications step 1-3 in original input model section.
2. Remove `DROP` columns completely from `$INPUT`. Add variables `PIPR` and `PPRE`. If `-ind_shrinkage` is set add variable `ISHR`.
3. remove `$SIM` if present
4. If `$PRIOR` is set, remove option `PLEV`

5. Update initial estimates with output from running modified original input model.
6. change filename in \$DATA to glsinput.dta.
7. If option -set_simest is set and an estimation record is found behind the tag ;gls-final then use this estimation record instead of the old one.
8. If \$TABLE is present: append -gls to filename set with FILE to distinguish from table output from original model.
9. Set IGNORE=@ in \$DATA. Skip all old IGNOREs (glsinput.dta is filtered).
10. In \$DATA add option IGNORE=(PIPR.LE.0.000000001)
11. add code in the very beginning of \$PRED/\$ERROR


```
SHRI = <iwres_shrinkage from input model run or iwres_shrinkage
option or ISHR if -ind_shrinkage is set>

IF(SHRI.LE.0) SHRI = 0
GLSP = SHRI*PPRE + (1-SHRI)*PIPR
```
12. check that W definition is of the form $W=\text{SQRT}(\dots\text{IPRED}\dots)$ and replace IPRED with GLSP in W definition. If option -additive_theta=XX is set, prepend $\text{THETA}(T)**2+$ to the expression inside parentheses of $W=\text{SQRT}(\dots)$ where T is order number of new THETA.
13. Do 12) for every line of the form $W=\text{SQRT}(\dots\text{IPRED}\dots)$.
14. Run gls model. Let PsN compute iwres shrinkage for this model. Append raw_results to raw_results of original model, if original model was run.

4.4 GLS model if option -gls_model is set

1. If option -additive_theta=XX is set, add \$THETA XX FIX as last \$THETA. Store order number T of new theta.
2. Remove MSFO option from \$EST, if present.

3. If a lst-file is found for the gls model, update initial estimates in gls model based on lst-file.
4. In \$DATA add option IGNORE=(PIPR.LE.0.000000001)
5. add code in the very beginning of \$PRED/\$ERROR


```
SHRI = <iwres_shrinkage from input model run or iwres_shrinkage
option or ISHR if -ind_shrinkage is set>

IF(SHRI.LE.0) SHRI = 0
GLSP = SHRI*PPRE + (1-SHRI)*PIPR
```
6. check that W definition is of the form $W=\text{SQRT}(\dots\text{IPRED}\dots)$ and replace IPRED with GLSP in W definition. If option -additive_theta=XX is set, prepend $\text{THETA}(T)**2+$ to the expression inside parentheses of $W=\text{SQRT}(\dots)$ where T is order number of new THETA.
7. Do 6) for every line of the form $W=\text{SQRT}(\dots\text{IPRED}\dots)$.
8. Run gls model. Let PsN compute iwres shrinkage for this model. Append raw_results to raw_results of original model, if original model was run.

References

- [1] V. Ivaturi, Andrew C. Hooker, and Mats O. Karlsson. “Individual Observation-Shrinkage Scaling of Residuals in a Generalized Least Square Type of Estimation in NONMEM”. In: *PAGE 21 Abstract 2586* (2012).