README Document for Using SAS Macro `JMFit`

The `SAS` macro `JMFit` as well as the five submacros, `SPM1L`, `SPM1Q`, `SPM2L`, `SPM2Q`, and `TVC`, should be stored in a folder named "jmfit". Then `JMFit` can be accessed by including the following lines:

```
filename jmfit "directory of the file JMFit";
%include jmfit(JMFit.sas);
%JMFit(LONG=, SURV=, MODEL=, TS=, TMAXI=, WEIGHT=, NPIECES=, PARTITION=, OPTIONS=,
 INITIAL=, OUTPUT=);
```

Inputs for `JMFit`

`LONG`: Data set with the first three columns, SID (subject ID), Y (longitudinal measure), A (time at which Y was taken), and additional columns for covariates (X_{L1}-X_{Lp}), where SID, Y, and A should be arranged in the first, second, and third columns, and $X_{L1}, \cdots, X_{Lp}$ should be placed after column 3, which can be enforced in `SAS` by using the "**retain**" command. Note that $X_{L1}, ..., X_{Lp}$ can be time-dependent or baseline covariates. Required.

`SURV`: Data set with the first three columns, SID, survival time (T), censoring indicator (delta) (1 = death and 0 = censored), and additional columns for covariates (X_{S1}-X_{Sq}), where SID, T, and delta should be arranged in the first, second, and third columns, and $X_{S1}, \ldots, X_{Sq}$ should be placed after column 3. Required.

`MODEL`: Model specification. Required. One of

1. `SPM1L`: Shared Parameter Model 1 with Linear trajectory.
2. `SPM1Q`: Shared Parameter Model 1 with Quadratic trajectory.
3. `SPM2L`: Shared Parameter Model 2 with Linear trajectory.
4. `SPM2Q`: Shared Parameter Model 2 with Quadratic trajectory.
5. `TVC`: Time-Varying Covariates Model.

`TS`: Indicates whether to implement the model specified in the `MODEL` argument or the corresponding two-stage model. If `0` (default), the model specified in the `MODEL` argument will be fit. If `1`, the corresponding two-stage model will be fit instead. It only works for `SPM1L`, `SPM1Q`, `SPM2L`, and `SPM2Q`.

`TMAXI`: $t_{\max,i}$ adjustment to the model specified in the `MODEL` argument. If `0` (default), no $t_{\max,i}$ adjustment will be applied. If `1`, the trajectory will become flat after $t^*_{\max,i} = t_{\max,i} +$ `WEIGHT`$\times\max(t_i - t_{\max,i}, 0)$. If `2`, starting at $t^*_{\max,i}$, the trajectory will linearly go down to 0 at the last follow-up survival time. It only works for `SPM1L` and `SPM1Q`.

WEIGHT: The proportion ($\in [0, 1]$) of $\max(t_i - t_{\max,i}, 0)$. If 0 (default), the starting point of the modified extrapolation of the trajectory is $t_{\max,i}$. If 1, the trajectory extends to t_i with no $t_{\max,i}$ adjustment. It only works when TMAXI=1 or TMAXI=2.

NPIECES: Number of intervals J (≥ 1) for the piecewise constant baseline hazard function. Required.

PARTITION: Algorithm for constructing the partition of the time axis. Required. One of

    (i) 1: Equally-Spaced Quantile Partition (ESQP).

    (ii) 2: Left Bi-Sectional Quantile Partition (LBSQP).

    (iii) 3: Middle Bi-Sectional Quantile Partition (MBSQP).

    (iv) 4: Right Bi-Sectional Quantile Partition (RBSQP).

OPTIONS: Allows users to specify options that are available in the PROC NLMIXED statement. For example, "OPTIONS=%str(QPOINTS=5 TECH=CONGRA ABSGCONV=0.0001)" specifies Gaussian quadrature with five quadrature points for approximating the integral of the likelihood over the random effects, the conjugate-gradient optimization, and an absolute gradient convergence criterion of 0.0001.

INITIAL: Allows users to set their own initial values. JMFit will automatically generate the starting values for the model parameters and these initial values will be stored in the data set "_initial". Since the order of the parameters is very important when calculating AIC_{Long} and BIC_{Long}, users are recommended to change the initial values in "_initial" and then rename "_initial".

OUTPUT: Name of the output rich text file (rtf). One can also specify the directory in which the file will be put. For example, the output file named "myoutput" will be stored in "C:\...\" by "%JMFit(..., OUTPUT = C:\...\myoutput);" If OUTPUT is not specified, the file will be indexed by the name of Y from LONG and the model's name.

Note 1: (i) The name of the SID variable in LONG should be the same as that of the SID variable in SURV;

(ii) A and T should be in the same unit of time (month preferred);

(iii) the categorical covariates must be coded as dummy variables; and

(iv) the SAS macro allows for any numbers of covariates for both components of the joint model and the covariates for the longitudinal component can be totally different from those for the survival component.

Note 2: (i) "ERROR: Not enough memory to generate code." This might occur if J is too big;

(ii) a too long path for the OUTPUT may lead to an error;

(iii) "ERROR: A character operand was found in the %EVAL function or %IF condition where a numeric operand is required. The condition was: &output =

ERROR: The macro SPM1L will stop executing." This error is due to the use of a math operator such as "+", "−", or "∗" in the output file name or in the local folder name for the output file. Please avoid using any math operators in folder names or output file names.

(iv) the macro is assuming "options validvarname=v7;" for valid variable names that can be processed in SAS; and

(v) the calculations of AIC_{Long} and BIC_{Long} require the IML Procedure.

Note 3: No missing values are allowed in both data sets.

Output for `JMFit`

The macro automatically produces an rtf file indexed by the name of Y from `LONG` and the model's name. The rtf file includes five tables: (i) Number of Subjects; (ii) Fit Statistics; (iii) Survival Parameter Estimates (Survival Alone); (iv) Parameter Estimates; and (v) Hazard Ratios & λ Estimates.

> Note: The construction of the Parameter Estimates table is different for each model. For SPM1L, SPM1Q, SPM2L, and SPM2Q, it consists of three subtables: "Covariance Parameter Estimates", "Longitudinal Parameter Estimates", and "Survival Parameter Estimates"; for the TS model corresponding to SPM1L, SPM1Q, SPM2L, or SPM2Q, there are two tables: "Longitudinal Parameter Estimates (Stage I)" and "Survival Parameter Estimates (Stage II)"; and for the TVC model, there is only one table named "Parameter Estimates".

The IBCSG data example

The datasets `os`, `appet`, `coping`, `mood`, and `phys` represent the overall survival and the four QOL indicators, appetite, coping, mood, and physical well-being from the IBCSG trial described in Section 4.1 of the manuscript. The `SAS` code `The_IBCSG_data` which includes the following lines will produce four rtf files, appet_tm2_w05_np9_p2.rtf, coping_tm2_w05_np9_p2.rtf, mood_tm2_w05_np9_p2.rtf, and phys_tm2_w05_np9_p2.rtf. Figures 2-6 can be found in coping_tm2_w05_np9_p2.rtf and Figure 7 refers to the other three rtf files. Additional rtf files are included to support Tables 1 and 2.

```
filename jmfit "C:\...\jmfit";
%include jmfit(JMFit.sas);

%JMFit(LONG=appet, SURV=os, MODEL=SPM1L, TS=0, TMAXI=2, WEIGHT=0.5, NPIECES=9,
 PARTITION=2, OUTPUT=C:\...\appet_tm2_w05_np9_p2);

%JMFit(LONG=coping, SURV=os, MODEL=SPM1L, TS=0, TMAXI=2, WEIGHT=0.5, NPIECES=9,
 PARTITION=2, OUTPUT=C:\...\coping_tm2_w05_np9_p2);

%JMFit(LONG=mood, SURV=os, MODEL=SPM1L, TS=0, TMAXI=2, WEIGHT=0.5, NPIECES=9,
 PARTITION=2, OUTPUT=C:\...\mood_tm2_w05_np9_p2);

%JMFit(LONG=phys, SURV=os, MODEL=SPM1L, TS=0, TMAXI=2, WEIGHT=0.5, NPIECES=9,
 PARTITION=2, OUTPUT=C:\...\phys_tm2_w05_np9_p2);
```

A simulated data example

The datasets `surv`, `long`, `long1`, `long2`, and `long3` represent the simulated datasets D_{Surv}, D_{Long}, D_{Long1}, D_{Long2}, and D_{Long3} used in Section 4.2 of the manuscript. The `SAS` code `A_simulated_data_example` which includes the following lines will produce four rtf files, long.rtf, long1.rtf, long2.rtf, and long3.rtf. Figure 8 in Section 4.2 of the manuscript refers to all the four rtf files. Figures 9, 10, 11, and 12 in Appendix B of the manuscript can be found in long.rtf.

```
filename jmfit "C:\...\jmfit";
%include jmfit(JMFit.sas);
```

```
%JMFit(LONG=long, SURV=surv, MODEL=SPM2L, TS=0, NPIECES=1, PARTITION=2,
 OUTPUT=C:\...\long);

%JMFit(LONG=long1, SURV=surv, MODEL=SPM2L, TS=0, NPIECES=1, PARTITION=2,
 OUTPUT=C:\...\long1);

%JMFit(LONG=long2, SURV=surv, MODEL=SPM2L, TS=0, NPIECES=1, PARTITION=2,
 OUTPUT=C:\...\long2);

%JMFit(LONG=long3, SURV=surv, MODEL=SPM2L, TS=0, NPIECES=1, PARTITION=2,
 OUTPUT=C:\...\long3);
```