paramsurv SAS macro

Fitting flexible parametric survival models for economic modeling, including the generalized gamma and generalized F distributions and standard models in the generalized gamma and generalized F family, like exponential, Weibull, log-logistic, etc.

Macro Parameters

Parameter	Description	Default	Required	Valid Values	Comments
data	Dataset name		Y		Dataset that contains covariates and response, and/or weight, and/or strata
t1	Starting / observed time variable		Y		Variable name in &data that corresponds to starting time of the interval, or observed time if &t2 is not supplied and &censor is supplied
t2	Ending time variable		N		Variable name in &data that corresponds to ending time of the interval. Either &t2 or &censor must be supplied
censor	Censoring indicator variable		N		Variable name in &data that corresponds to censoring status. Either &t2 or &censor must be supplied
censval	Censoring value	0	N		Value that corresponds to censoring in &censor
covars	Covariates		N		Variable names in &data that correspond to covariates associated with location parameter, delimitated by space
anc	Covariates on ancillary parameters		N		Covariates that are associated with ancillary parameters. Different parameters are delimitated by comma. Covariate name should be given inside the parentheses that follow the corresponding parameter. For example, anc=%str(sigma(x1 x2), lambda(x2))
class_cov	Classification covariates		N		Covariates in &covars that are classification covariates. No need to specify if the covariate is already non-numeric in &data
refgrp	Reference group		N		Levels of reference groups for &class_cov, delimitated by comma. If not given, the first group in alphabetical order is selected
weight	Observation weights		N		Variable name in &data that corresponds to case weights
strata	Stratification variables		N		Variable name in &data that corresponds to stratification variables
optim_method	Optimization method	nlpnra	N		Optimization method used in PROC IML
dist	Distribution		N	Gengamma, genf, genf_orig, exp / exponential, Weibull, gamma, lnorm, llogis, gompertz	Distribution name (case insensitive). If not supplied, then a valid self-defined distribution must be supplied
init	Initial values		N		Initial values for all parameters to be optimized. If not supplied, automatic initial values will be assigned according to range of parameters
					Amount of output printed by the NLP subroutine. The higher the

nlp_print	Amount of output	0	N	0, 1, 2, 3, 4, 5	value, the more printed output is produced. By default, no output is produced by NLP.
log_result	Indicator of printed results in log scale	no	N	yes, no	Indicator of whether to print estimation results in the log scale, default is not.
alpha	Confidence level	0.05	N		Confidence level for inference
robust	Robust estimator	no	N	yes, no	Yes indicates sandwich estimator for standard error, no indicates regular estimator
lower	Lower bounds	{. . . . . .}	N		Lower bounds for parameters, missing for negative infinity
upper	Upper bounds	{. . . . . .}	N		Upper bounds for parameters, missing for infinity
density	Custom density		N		User-defined density function, use 'time' to denote time-to-event variable. For custom distribution, at least two of &density, &survival and &hazard need to be defined
survival	Custom survival		N		User-defined survival function, use 'time' to denote time-to-event variable. For custom distribution, at least two of &density, &survival and &hazard need to be defined
hazard	Custom hazard		N		User-defined hazard function, use 'time' to denote time-to-event variable. For custom distribution, at least two of &density, &survival and &hazard need to be defined
custom_prep	Custom parameter preparation		N		Statements to state relationships between parameters in custom distributions and optimization parameters (theta), used to prepare variables for optimization
location	Primary parameter name	beta	N		Name of primary parameter that is associated with covariates in &covars in custom distribution, default is beta
param_anc	Ancillary parameter names (log-transformed)		N		Ancillary parameter names after log transformation
log_transf_param	Parameter name for log transformation		N		Names of positive parameters, e.g. sigma for Generalized Gamma
log_density	Custom log density		N		User-defined log density function, can be used to simplify calculation and avoid numerical issues
log_survival	Custom log survival		N		User-defined log survival function, can be used to simplify calculation and avoid numerical issues
pred	Prediction dataset		N		Prediction dataset that contains time (column named 'time') and covariates (with same column name with the original data) to predict upon, each row is a prediction case
pred_max_time	Furthest time for prediction		N		Longest time point for prediction shown in survival and hazard plots, for example 10 years
pred_plot_cl	Confidence bands in prediction plots	yes	N	yes, no	Indicator of whether confidence bands are shown in predicted survival and hazard plots
res_print	No results printed	yes	N	yes, no	If no, no results will be printed

Examples of Macro Call:

Assume we have a dataset called *data* stored in the current workspace, whose first 10 observations are shown below. The *time* column represents the observed time and the *delta* column represents the censoring status where 0 is censored and 1 is observed. Age is a continuous covariate that has been centered and scaled, while sex is a character covariate with two categories. Several examples are presented below to call the macro with this dataset.

Obs	time	delta	age	sex
1	0.50007	1	-1.03832	female
2	2.92097	0	0.45638	female
3	2.94312	0	1.25355	male
4	0.21046	1	-2.13443	female
5	0.05548	1	0.55603	female
6	2.92410	0	0.65567	female
7	0.43712	1	-0.44044	male
8	0.44384	0	-0.34079	male
9	0.82329	0	-0.44044	male
10	0.27194	1	-0.73938	female

Example 1: Standard models

We can fit standard models with different distributions on this dataset. To do that, we can specify the dataset name (`data=data`) and variable names that correspond to the observed time (`t1=time`), censoring status (`censor=delta`), and covariates (`covars=age sex`). Then we can use any of the built-in distributions. Exponential and Weibull models are shown as below. Only the location parameters will depend on age and sex in these models.

```
%paramsurv(data=data, t1=time, censor=delta, covars=age sex, dist=exp)
%paramsurv(data=data, t1=time, censor=delta, covars=age sex, dist=Weibull)
```

Example 2: Covariates on ancillary parameters

We can fit more complex models where ancillary parameters also depend on certain covariates. For example, in a generalized gamma (GG) distribution, we would like the scale parameter σ to be associated with age and sex and the shape parameter associated with sex. If we denote the age variable by x_1 and the sex variable by x_2. The density can be expressed by $f(t|\beta(x_1, x_2), \sigma(x_1, x_2), \lambda(x_2))$. Since σ in GG is constraint to be positive, we can link the covariates to the parameter with the log link function. For β and λ, the identity link function will be used. We can fit the model by specifying the `anc` argument to be `anc=%str(sigma(age sex), lambda(sex))`, as in the following macro call.

```
%paramsurv(data=data, t1=time, censor=delta, covars=age sex, anc=%str(sigma(age sex),
lambda(sex)), dist=gengamma)
```

Example 3: Custom distribution

In this example, we fit a model with a user-defined distribution which is essentially a Weibull model with the proportional hazards parameterization. In this parameterization, the hazard and survival functions are expressed by $h(t) = \mu\alpha t^{\alpha-1}$ and $S(t) = \exp\left(-\mu t^{\alpha}\right)$, respectively, where $\mu = \exp\left(-\beta\right)$ is a function of the location parameter β and α is the new scale parameter. We can fit the model by specifying the custom hazard and log survival functions as `hazard=%str(mu*alpha*time**(alpha-1))` and `log_survival=%str(-mu*time**alpha)`. Note that the time variable must be denoted by *time*. The ancillary parameter name is `alpha` and `log_alpha` in the original and log scales, respectively. Since α is constraint to be positive, we transform α to the log scale by specifying `log_transf_param=alpha`. We still use `beta` to be the location parameter which is the default option and needs not to be explicitly specified. In the `custom_prep` argument, we need to define the relationship between the undefined parameter in the distribution, μ, and built-in defined parameter β by providing the argument of `custom_prep=%str(mu=exp(-beta);)`.

```
%paramsurv(data=data, t1=time, censor=delta, covars= age sex, hazard=%str(mu*alpha*time**
(alpha-1)), log_survival=%str(-mu*time**alpha), param_anc=alpha, log_transf_param=alpha,
custom_prep=%str(mu=exp(-beta);))
```

Example 4: Long-term prediction

We fit a generalized gamma model to the dataset and made predictions for cases with certain covariate values. We used the robust estimator for standard errors by specifying `robust=t`. The prediction dataset called *pred* stored in the current workspace is shown below. We would like to obtain the predicted survival and hazard for males and females at average age (scaled age = 0) at time point 1, respectively. For predicted survival and hazard curves over time, we can specify the longest prediction time, which is set to 5 here (`pred_max_time=5`).

Obs	sex	time	age
1	female	1	0
2	male	1	0

```
%paramsurv(data=data, t1=time, censor=delta, covars=age sex, dist=gengamma, robust=t,
pred=pred, pred_max_time=5)
```